Volume: 4 Issue: II Month of publication: February 2016

DOI:

*www.ijraset.com* *Volume 4 Issue II, February 2016*
*IC Value: 13.98* *ISSN: 2321-9653*

**International Journal for Research in Applied Science & Engineering Technology (IJRASET)**

# A Novel Approach for Query Services in the Cloud using RASP Data Perturbation

P.T.J.Praveen Raj[1], K.Narayana[1]

[1] *M.Tech Dept of Computer Science and Engineering, Sheshachala Institute of Technology*

[2] *Associate Professor,Dept of Computer Science and Engineering, Sheshachala Institute of Technology*

***Abstract—With the wide sending of open distributed computing foundations, utilizing mists to host information question administrations has turned into an engaging answer for the focal points on adaptability and cost-sparing. On the other hand, some information may be touchy that the information proprietor would not like to move to the cloud unless the information confidentiality and question security are ensured. Then again, a secured question administration ought to still give efficient inquiry handling and significantly decrease the in-house workload to completely understand the benefits of distributed computing. We propose the Scratch information irritation strategy to give secure and efficient range inquiry and kNN question administrations for ensured information in the cloud. The Scratch information annoyance system consolidates request protecting encryption, dimensionality development, irregular clamor infusion, and arbitrary projection, to give solid versatility to assaults on the irritated information and questions. It additionally protects multidimensional reaches, which permits existing indexing strategies to be connected to speedup range question handling. The kNN-R calculation is intended to work with the Scratch range inquiry calculation to prepare the kNN questions. We have painstakingly dissected the assaults on information and questions under an absolutely defined risk model and practical security suppositions. Broad tests have been led to demonstrate the benefits of this methodology on efficienc.***

***Index Terms—query services in the cloud, privacy, range query, kNN query***

## I. INTRODUCTION

Facilitating information escalated question administrations in the cloud is progressively well known due to the exceptional focal points in adaptability and cost-sparing. With the cloud foundations, the administration proprietors can conveniently scale up or down the administration and pay for the hours of utilizing the servers. This is an appealing element in light of the fact that the workloads of question administrations are very dynamic, and it will be costly and inefficient to serve such element workloads with in-house foundations [2]. On the other hand, on the grounds that the administration suppliers lose the control over the information in the cloud, information confidentiality and question pfl@+A è ð+e the significant concerns. Foes, for example, inquisitive administration suppliers, can make a duplicate of the database or listen stealthily clients' questions, which will be difficult to distinguish and anticipate in the cloud foundations. While new approachesare expected to protect information confidentiality and inquiry security, the efficiency of question administrations and the benefits of utilizing the mists ought to likewise be safeguarded. It won't be important to give moderate question administrations as a consequence of security and protection certification. It is additionally not down to earth for the information proprietor to utilize a significant measure of in house assets, on the grounds that the reason for utilizing cloud assets is to diminish the need of keeping up versatile in-house frameworks. In this way, there is a mind boggling relationship among the information confidentiality, inquiry security, the nature of administration, and the financial aspects of utilizing the cloud. We abridge these necessities for developing a functional inquiry administration in the cloud as the CPEL criteria: information Confidentiality, question Protection, Efficient question preparing, and Low in-house handling cost. Fulfilling these prerequisites will significantly build the multifaceted nature of developing question administrations in the cloud. Some relô ôýfl˘ýes have been produced to address a few parts of the issue. On the other hand, they don't attractively address these angles. For instance, the crypto-record [12] and Request Saving Encryption (OPE) [1] are defenseless against the assaults. The upgraded crypto-record approach [14] puts overwhelming weight on the in-house base to enhance the security and protection. The New Casper methodology [24] utilizes shrouding boxes to secure information questions and inquiries, which influences the efficiency of inquiry preparing and the in-house workload. We have outlined the shortcomings of the current methodologies in Area 7. We propose the Irregular Space Irritation (Scratch) way to deal with developing commonsense extent question and k-closest neighbor (kNN) inquiry administrations in the cloud. The proposed methodology will address all the four parts of the CPEL criteria and plan to accomplish a decent adjust on them. The fundamental thought is to haphazardly change the multidimensional datasets with a mix of request safeguarding encryption, dimensionality extension, irregular commotion infusion, and arbitrary venture, so that the utility for handling range questions is protected. The Scratch bother is outlined in such a route, to the point that the questioned reaches are safely changed into polyhedra in the Grate bothered information space, which can be efficiently handled with the backing of indexing structures in the annoyed space. The Scratch kNN question administration (kNN-R) utilizes the Grate range inquiry administration to prepare kNN inquiries. The key parts in the Scratch structure incorporate (1) the definition and properties of Grate annoyance; (2) the development of the security saving extent question administrations; (3) the development of security safeguarding kNN inquiry administrations; and (4) an examination of the assaults on the Scrat.

In summary, the proposed approach has a number of unique contributions.

*A.* The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee.

*B.* The RASP approach preserves the topology of multidimensional range in secure transformation, which allows indexing and efficiently query processing.

*C.* The proposed service constructions are able to minimize the in-house processing workload because of the low perturbation cost and high precision query results. This is an important feature enabling practical cloud-based solutions.

We have precisely assessed our methodology with engineered and genuine datasets. The outcomes demonstrate its exceptional favorable circumstances on all parts of the CPEL criteria. The whole paper is sorted out as takes after. In Segment 3, we define the Grate annoyance strategy, portray its significant properties, and break down the assaults to the Scratch bothered information. We likewise present the system for building the inquiry administrations with the Grate annoyance. In Segment 4 we portray the calculation for changing inquiries and handling range questions. In section 5, the reach question administration is stretched out to handle kNN inquiries. While portraying these two administrations, we additionally investigate the assaults on the question protection. At long last, we display some related methodologies in Segment 7 and break down their shortcomings regarding the CPEL criter

## II. QUERY SERVICES IN THE CLOUD

This section exhibits the documentations, the framework building design, and the danger model for the Scratch approach, and gets ready for the security investigation [3] in later segments. The configuration of the framework construction modeling remembers the cloud financial aspects so that most information stockpiling and registering undertakings will be done in the cloud. The danger model makes sensible security presumptions and unmistakably defines the reasonable dangers that the Scratch methodology will addres

### *A. Definitions and Notations*

In the first place, we set up the documentations. For effortlessness, we consider just single database tables, which can be the consequence of denormalization from different relations. A database table comprises of n records and d searchable traits. We likewise every now and again allude to a property as a measurement or a segment, which are replaceable in the paper. Every record can be spoken to as a vector in the multidimensional space, indicated by low case letters. In the event that a recorddx is d-dimensional, we say $x \in R^d$, where R implies the d-dimensional vector space. A table is likewise regarded as a d×n grid, with records spoke to as section vectors. We utilize capital letters to speak to a table, and filed capital letters, e.g., Xi, to speak to segments. Every section is defined on a numerical area. Straight out information segments are permits in extent inquiry, which are changed over to numerical spaces as we will portray in Segment 3. Range inquiry is a vital kind of question for some information diagnostic assignments from basic conglomeration to more complex machine learning errands. Let T be a table and Xi, Xj, and Xk be the genuine esteemed qualities in T , and an and b be a few constants. Take the numbering question for instance. A normal extent inquiry resemble

$$\textit{select count(*) from T where } X_i \in [a_i, b_i] \textit{ and } X_j \in (a_j, b_j) \textit{ and } X_k = a_k,$$

which ascertains the quantity of records in the extent defined by conditions on Xi, Xj, and Xk. Range inquiries might be connected to subjective number of at tributes and conditions on these traits joined with contingent administrators "and"/"or". We call every part of the question condition that includes stand out quality as a straightforward condition. A simple condition like $X_i \in [a_i, b_i]$ can be described with two half space conditions $Xi \leq bi$ and $-Xi \leq -ai$. Without loss of all inclusive statement, we will talk about how to process half space conditions like $Xi \leq bi$ in this paper. A slight modification will extend the examined calculations to handle different conditions like $Xi < bi$ and $Xi = bi$ . kNN inquiry is to find the nearest krecords to the question point, where the Euclidean separation is frequently used to gauge the nearness. It is as often as possible utilized as a part of area based administrations for seeking the articles near an inquiry point, furthermore in machine learning calculations, for example, various leveled bunching and kNN classifier. A kNN inquiry comprises of the question point and the quantity of closest neighbor.

### B. *System Architecture*

We assume that a cloud computing infrastructure, such as Amazon EC2, is used to host the query administrations and substantial datasets.

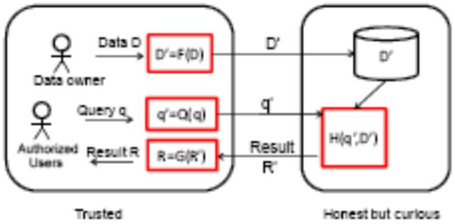

Fig. 1. The system architecture for RASP-based query services.

The reason for this construction modeling is to extend the exclusive database servers to people in general cloud, or utilize a half and half private-open cloud to accomplish versatility and lessen costs while looking after confidentiality. Every record x in the outsourced database contains two sections: the Grate handled properties D′=F (D, K ) and the scrambled unique records, Z= E(D, K′), where Kand K′ are keys for irritation and encryption, individually. The Scratch annoyed information D′ are for indexing and inquiry handling. Figure 1 demonstrates the framework structural planning for both Grate based extent inquiry administration and kNN administration. There are two plainly isolated gatherings: the trusted gatherings and the untrusted parties. The trusted gatherings incorporate the information/administration proprietor, the in-house intermediary server, and the approved clients who can just submit inquiries. The information proprietor trades the irritated information to the cloud. In the mean time, the approved clients can submit range inquiries or kNN questions to learn measurements or find a few records. The untrusted parties incorporate the inquisitive cloud supplier who has the inquiry administrations and the secured database. The Scratch irritated information will be utilized to fabricate lists to backing inquiry preparing.

There are various essential methods in this system: (1) F(D) is the Grate irritation that changes the first information D to the annoyed information D′; (2) Q(q) changes the first inquiry q to the ensured structure q′ that can be handled on the bothered information; (3) H(q′, D′)is the question preparing calculation that profits the outcome R′. At the point when the measurements, for example, Entirety or AVG of a specific measurement are required, Scratch can work with halfway homomorphic encryption, for example, Paillier encryption [25] to figure these insights on the scrambled information, which are then recuperated w

### C. *Threat Model*

*1)* Assumptions. Our security examination is based on the vital elements of the building design. Under this setting, we trust the accompanying presumptions are appropriate.• Just the approved clients can inquiry the exclusive database. Approved clients are not vindictive and won't purposefully break the confidentiality. We consider insider assaults are orthog onal to our exploration; along these lines, we can bar the circumstance that the approved clients connive with the untrusted cloud suppliers to release extra data.

The customer side framework and the correspondence channels are appropriately secured and no ensured information records and inquiries can be spilled.

Adversaries can have the worldwide data of the database, for example, the utilizations of the database, the trait areas, and potentially the property appropriations, by means of other distributed sources (e.g., the dissemination of offers, orpatient illnesses, out in the open reports).

Adversaries can see the annoyed database, the changed questions, the entire inquiry preparing technique, the entrance designs, and comprehend the same question gives back the same arrangement of results, however nothing else.

These suppositions can be kept up and fortified by applying fitting security strategies. Note this model is comparable to the listening in model outfitted with the plaintext distributional information in the cryptographical.

2) *Protected Assets:* Information confidentiality and question security ought to be ensured in the Scratch approach. While the trustworthiness of question administrations is likewise a vital issue, it is orthogonal to our study. Existing respectability checking and anticipating methods [34], [30], [19] can be incorporated into our structure. In this manner, the trustworthiness issue will be avoided from the paper, and we can accept the inquisitive cloud supplier is keen on the information and questions, yet it will genuinely take after the convention to give the framework administration.

3) *Attacker Modeling:* The objective of assault is to recoup (or gauge) the first information from the irritated information, or distinguish the precise questions (i.e., area inquiries) to rupture clients' protection. By level of earlier learning the assailant might have, we arrange the assaults into two classifications.

a) *Level 1:* The assailant knows just the bothered information and changed inquiries, with no other former learning. This relates to the cipertext-just assault in the cryptographic setting.

b) *Level 2:* The assailant likewise knows the first information circulations, including singular property appropriations and the joint dissemination (e.g., the covariance framework) between properties. Practically speaking, for a few applications, whose measurements are fascinating to general society area, the dimensional dispersions may have been distributed by means of different sources.

These levels of information are suitable as indicated by the suppositions we hold. We will break down the security taking into account this risk model.

4) *Security Definition:* Not quite the same as the conventional encryption plans, aggressors can likewise be satisfied with great estimation. Consequently, we will examine two levels of security definitions: (1) it is computationally recalcitrant for the assailant to recuperate the precise unique information in light of the bothered information; (2) the aggressor can't viably evaluate the first information. The viability measure is defined with the NR MSE measure in Are section 3.3

## III. RASP: RANDOM SPACE PERTURBATION

In this section, we present the basic definition of RAndom Space Perturbation (RASP) method and its properties. We will also discuss the attacks on RASP perturbed data, based on the threat model given in Section 2.

### A. *Definition of RASP*

RASP is one sort of multiplicative irritation, with a novel mix of OPE, measurement development, arbitrary commotion infusion, and irregular projection. We should consider the multidimensional information are numeric and in multidimensional vector space1. The database has k searchable measurements and n records, which makes a d × n matrix X . The searchable measurements can be utilized as a part of inquiries and along these lines ought to be listed. Let x speak to a d-dimensional record, $x_d \in R^d$. Note that in the d-dimensional vector space R , the range query conditions are represented as half-space functions and a range query is translated to finding the point set in corresponding polyhedron area described by the half spaces [4].

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

1) An order preserving encryption (OPE) scheme[1], $E_{ope}$with keys $K_{ope}$, is applied to each dimension of x:$E_{ope}(x, K_{ope}) \in R^d$to change the dimensional distributions to normal distributions with each dimension's value order still preserved.

2) The vector is then extended to d + 2 dimensions as $G(x) = ((Eopt(x))^T, 1, v)^T$, where the (d + 1)th dimension is always a and the (d + 2)- th dimension, v, is drawn from a random real number generator RN G that generates random values from a tailored normal distributions. We will discuss the design of RNG and OPE later.

3) The (d + 2)-dimensional vector is finally transformed to

$$F(x, K =_{\{}A, Kope, RG_{\}}) = A((Eope(x))^T, 1, v)^T,$$

where A will be a (d+2)×(d+2) arbitrarily created invertible network with aij∈ R such that there are no less than two non-zero qualities in every line of An and the last segment of An is likewise non-zero2.

Kopeand An are shared by all vectors in the database, but v is haphazardly produced for every individual vector. Subsequent to the Grate bothered information records are utilized for indexing and questioning preparing, there is no compelling reason to recoup the annoyed information. As we said, for the situation that unique records are required, the encoded records connected with the Grate irritated records will be returned. We give the point by point calculation in appendix.

Design of OPE and RNG. We utilize the OPE plan to change over all measurements of the first information to the standard ordinary dissemination N (0, 1) in the constrained space [−β, β]. β can be chosen as a quality >= 4,as the range[−4, 4] covers more than 99%

of the populace. This should be possible with a calculation, for example, the one depicted in [1]. The utilization of OPE permits inquiries to be effectively changed and prepared. Likewise, we draw irregular clamors v from N (0, 1) in the restricted space [−β, β]. Such a configuration makes the broadened commotion measurement impassive from the information measurements as far as the appropriations. The configuration of such an amplified information vector $(\text{Eope}(x)^T, 1, v)$, T is to improve the information and question confidentiality. The utilization of OPE is to change expansive scale or infinite spaces to ordinary conveyances, which address the distributional assault. The (d + 1)- th homogeneous measurement is for concealing the question content. The (d+ 2)- th measurement infuses irregular clamor in the annoyed information furthermore shields the changed inquiries from assaults. The method of reasoning behind various viewpoints will be examined plainly.

### B. *Properties of RASP*

not safeguard the request of dimensional qualities be reason for the grid augmentation part, which separates itself from request saving encryption (OPE) plans, and consequently does not experience the ill effects of the conveyance based assault (points of interest in Area 7). An OPE plan maps an arrangement of single-dimensional qualities to another, while keeping the worth request unaltered. Following the Scratch irritation can be dealt with as a joined change F (G(Eope(x))), it is sufficient to demonstrate that F (y) = Ay does not safeguard the request of dimensional qualities, whereyR$^{d+2}$ and A∈R $^{(d+2)×(d+2)}$ from separation based assaults [8]. Since none of the changes in the Scratch: Eope, G, and F jelly removes, obviously the Grate annoyance won't safeguard separations. So also, Grate does not protect other more modern structures, for example, covariance grid and foremost parts [18]. In this manner, the PCA-based assaults, for example, [16], [20] don't function also.

Third, the first range questions can be changed to the Grate irritated information space, which is the premise of our inquiry handling procedure. An extent question portrays a hyper-cubic region (with conceivably open limits) in the multidimensional space. In Segment 4, we will demonstrate that a hyper-cubic range in the first space is changed to a polyhedron with the Grate annoyance. Consequently, we can look the focuses in the polyhedron to get the question result.

Algorithm 1 RASP Data Perturbation

```
RASP Perturb(X, RN G, RIM G, $K_o$)
Input: X: k × n data records, RN G: random real value
    generator that draws values from the standard normal
    distribution, RIM G : random invertible matrix genera-
    tor, $K_{ope}$: key for OPE $E_{ope}$; Output: the matrix A
A ← 0;
A3 ← the last column of A;
v0 ← 4;
while A3 contains zero do
    generate A with RIM G;
end while
for each record x in X do
    v ← $v_0$− 1;
    while v < v0 do
        v ← RNG;
    end while
    y ← $A((E_{ope}(x, K_{ope}))^T, 1, v)^T$;
    submit y to the server;
end for
return A;
```

### C. *Data Confidentiality Analysis*

As the risk model depicts, aggressors may be keen on finding the careful unique information records or evaluating them taking into account the annoyed information. For estimation assault, if the estimation is sufficiently exact (over certain exactness limit), we say the bother is not secure. Beneath, we define the measure for assessing the adequacy of estimation assault.

1) *Evaluating Effectiveness of Estimation Attacks:* Since assailants should not have to precisely recuperate the first values, an

exact estimation will be sufficient. A measure is expected to define the "precision" or "instability" as we said. We utilize the usually utilized mean-squared-blunder (MSE) to assess the viability of assault. To be semantically reliable, the j-th measurement can be dealt with as test qualities drawn from an arbitrary variable $X_j$. Let $x_{ij}$ be the estimation of the i-th unique record in j-th measurement and ˆij be the evaluated esteem. The MSE for the j-th measurement can be defined as

$$MSE(X_j, \hat{X}_j) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \hat{x}_{ij})^2,$$

which is identical to the difference: var(Xj−Xˆj). The square base of MSE (RMSE) speak to the instability of the estimation - for an expected valueˆ, the first esteem x could be in the reach (ˆ - RMSE,ˆ +RMSE). In this manner, the length of the reach, 2*RMSE, likewise speaks to the precision of the estimation. Then again, this length is liable to the length of the space. In this way, we utilize the standardized square base of MSE (NR MSE).

NR MSE(Xj) = 2qM SE(Xi, $X\hat{}_j$)/domain length,

instead, which is intuitively the rate between the uncertain range and the whole domain.

To think about MSE for various sections, we additionally need to standardize these two series{xij}and {ˆij} to wipe out the distinction on area scales. The standardization method [11] is depicted as takes after. Expect the mean and difference of the arrangement {2xij}is µjand σj, correspondingly. The arrangement is changed by xij← (xij−µj)/σj. A comparative strategy is additionally connected to the series{ˆij}. For the standardized spaces, the extent [−2, 2] just about spreads the entire population3 [11]. In this manner, for standardized arrangement, NR MSE is basically RMSE/2. For an assault that can just result in low-exactness estimation (e.g., NR MSE≥ 20%, the vulnerability is more than 20 % of the space length.), we call the RASP-irritated dataset is versatile to that assault. Naturally, NR MSE higher than 100% won't be exceptionally important. In this manner, we set unquestionably the upper bound to be 100%. We will talk about the specific upper limits as indicated by the level of early knowledge

2) *Prior-Knowledge Based Analysis:* Below, we analyze the security under the two levels of knowledge the attacker may have, according to the two levels of security definitions: exact match and statistical estimation.

a) *Naive Estimation:* We expect every quality in the vector or lattice is encoded with n bits. Let the bothered vector p drawn from an arbitrary variable P , and the first vector x be drawn from an irregular variable X. We demonstrate that gullible estimation is computationally recalcitrant to recognize the accurate unique information with the annoyed information, on the off chance that we utilize an irregular invertible genuine grid generator and an arbitrary genuine quality generator. The objective is to demonstrate the quantity of substantial X jdataset regarding a known annoyed dataset P. Beneath we talk about a simplified variant that contains no OPE part - the OPE rendition has in any event the same level of security.

Recommendation 1: For a known irritated dataset P , there existsO($2^{(d+1)(d+2)}$n)candidate X datasets in the first space.

*Proof:* For a given perturbation P= AZ , where Z is X with the two extended dimensions, we use $B_d$+1 to represent the (d + 1)-th row of A−1. Thus, $B_d$+$1^P$= [1, . . . , 1], i.e., the appended (d+1)-th row of Z. Keeping $B_d$+1 unchanged, we randomly generate other rows of B for a candidate Bˆ . The result Zˆ = ˆBP is a validate estimate of Z if Bˆ is invertible. Thus, the number of candidate X is the number of invertible Bˆ . The total number of Bˆincluding non-invertible ones is $2^{(}$d+1)(d+2)n. Based on the theory of invertible random matrix [28], the probability of generating a non-invertible random matrix is less than exp−c(d+2).

There are a same number of candidate X . Thus, finding the exact X has a negligible probability in terms of the number of bits, n.

As the applicants have an equivalent likelihood over the entire area, as per the definition of NR MSE, the questionable reach is the same as the entire space, bringing about NR MSE = 100%.

b) *Distribution-based Estimation:* With the known distributional data, the aggressor can accomplish more on evaluating the first information. The known most pertinent technique is called Independent Component Analysis (ICA) [17]. For a multiplicative bother P = AX, the essential thought is to find an ideal projection, wP , where w is a d + 2 measurement column vector, to bring about a line vector with its worth dissemination near that of one unique property. It can be reached out to find a network W , so that W P gives free and non-gaussian lines, i.e., a great evaluation of X.

The ICA calculations [17], [13] are streamlining calculations that attempt to find such projections by boosting the non-gaussianity4of the projection wP . The non-gaussianity of the first attributions is vital on the grounds that any projection of a multidimensional ordinary circulation is still a typical appropriation, which takes off no intimation for recuperation.

In this manner, with our outline of OPE and the clamor measurement in Section 3, we have the accompanying result

*Proposition 2:* There are $O(2^{dn})$ candidate projection vectors, w, that lead to the same level of non gaussianity.

*Proof:* The OPE encrypted matrix $X^{-}$ (with the homogeneous dimension excluded, which can be possibly recovered) can be treated as a sample set drawn from a multivariate normal distribution N (μ, Σ). Any invertible transformation $P^{-} = {}^{-}AX^{-}$ will result in an other multivariate normal distribution $N({}^{-}A\mu, A^{-}\Sigma^{-T})$.

Thus, any projection $wP^{-}$ will not change the Gaussianity, and there are $O(2^{dn})$ such candidates of w. Thus, the probability to identify the right projection is negligible in terms of the number of bits n. This shows that any ICA-style stimation that depends on non-guassianity is equally ineffective to the RASP perturbation.

In addition to ICA, Principal Component Analysis (PCA) based attack is another possible distributional attack, which, however, depends on the preservation of covariance matrix [20]. Because the covariance matrix is not preserved in RASP perturbation, the PCA attack cannot be used on RASP perturbed data. It is unknown whether there are other distributional methods for approximately separating X or A from the perturbed data P, which will be studied in the ongoing work

## IV. RASP RANGE-QUERY PROCESSING

In light of the RASP bother system, we plan the administrations for two sorts of inquiries: reach question and kNN inquiry. This segment will commit to range inquiry preparing. We will first demonstrate that a reach inquiry in the first space can be changed to a polyhedron question in the irritated space, and afterward we add to a protected approach to do the question change. At that point, we will add to a two-stage inquiry preparing technique for efficient range question handle.

### A. *Transforming Range Queries*

Let's look at the general form of a range query condition. Let $X_i$be an attribute in the database. A simple condition in a range query involves only one attribute and is of the form "$X_i$<*op*> $a_i$", where $a_i$ is a constant in the normalized domain of $X_i$and op ∈ {<, >, =, ≤, ≥, =}is a comparison operator. For convenience we will only discuss how to process $X_i < a_i$, while the proposed method can be slightly changed for other conditions. Any complicated range query can be transformed into the disjunction of a set of conjunctions, i.e., $S^{n}j{=}1(\cap^{m}i{=}1^{C}i,j)$, where m, n are some integers depending on the original query conditions and Ci,j is a simple condition about Xi. Again, to simplify the presentation we restrict our discussion to a single conjunction condition $\cap^{m}i{=}1^{C}i$ , where Ci is in form of bi ≤Xi ≤ai. Such a conjunction conditions describes a hyper-cubic area in the multidimensional space.

According to the three nested transformations in RASP $F(G(E_{ope}(x)))$, we will first show that an OPE will transform the original hyper-cubic area to another hyper-cubic area in the OPE space.

*Proposition 1:* Order preserving encryption functions transform a hyper-cubic query range to another hyper-cubic query range.

*Proof:* The original range query condition consists of simple conditions like $b_i \le X_i \le a_i$ for each dimension. Since the order is preserved, each simple condition is transformed as follows: $E_{ope}(b_i) \le E_{ope}(X_i) \le E_{ope}(a_i)$, which means the transformed range is still a hyper-cubic query range.

Let y=Eope(x) andci= Eope(ai). A simple condition Yi ≤ ci defines a half-space. With the extended dimensions $z^{T} = (y^{T}, 1, v)$, the half-space can be represented as $w^{T}z \le 0$, where w is a d + 2 dimensional vector with wi = 1, wd+1 = −ci,

### B. *A Two-Stage Query Processing Strategy with Multidimensional Index Tree*

With the transformed queries, the next important task is to process queries efficiently and return precise results to minimize the client-side post-processing effects. A commonly used method is to use multi dimensional tree indices to improve the search performance. However, multidimensional tree indices are normally used to process axis aligned "bounding boxes"; whereas, the transformed queries are in arbitrary polyhedra, not necessarily aligned to axes. In this section, we propose a two-stage query processing strategy to handle such irregular-shape queries in the perturbed space.

1) *Multidimensional Index Tree*: Most multidimensional indexing algorithms are derived from R-tree

$$u^{T}(A{-}1_{)}T_{wq}T_{A}{-}1_{u} \le 0.$$

bounding region (MBR) is the construction block for indexing the multidimensional data. For 2D data, an MBR is a rectangle. For higher dimensions, the shape of MBR is extended to hyper-cube. Figure 2 shows the MBRs in the R-tree for a 2D dataset, where each node is bounded by a node MBR. The R-tree range query algorithm compares the MBR and the queried range to find the answers.

2) *The Two-Stage Processing Algorithm:* The transformed query describes a polyhedron in the perturbed space that cannot be directly processed by multi dimensional tree algorithms. New tree search algorithms could be designed to use arbitrary polyhedron conditions directly for search. However, we use a simpler two-stage solution that keeps the existing tree search algorithms unchanged. At the first stage, the proxy in the client side finds the MBR of the polyhedron (as a part of the submitted transformed query) and submit the MBR and a set of secured query conditions$\{\Theta_1, \ldots, \Theta_m\}$ to the server.

The server then uses the tree index to find the set of records enclosed by the MBR.

The MBR of the polyhedron can be efficiently founded based on the original range. The original query condition constructs a hyper-cube shape. With the described query transformation, the vertices of the hyper cube are also transformed to vertices of the polyhedron. Therefore, the MBR of the vertices is also the MBR of the polyhedron [27]. Figure 3 illustrates the relationship between the vertices and the MBR and the two-stage processing strategy.

At the second stage, the server uses the transformed halfspace conditions to filter the initial result. In most cases of tight ranges, the initial result set will be reasonably small so that it can be filtered in memory by simply checking the transformed half-space conditions. However, in the worst case, the MBR of the polyhedron will possibly enclose the entire dataset and the second stage is reduced to a linear scan of the entire dataset. The result of second stage will return the *exact* range query result to the proxy server, which significantly reduces the post-processing cost that the proxy server needs to take. It is very important to the cloud-based service, because low post-processing cost requires low in-house investment.

Algorithm 3 Two-Stage Query Processing.

```
ProcessQuery(M BR,$\{Q_i\}$)
Input: MBR: MBR for the transformed query; $\{Q_i\}$:filtering conditions; Output: the set of perturbed records satisfying the conditions
Y ←     use the indexing tree to find answers for MBR
Y’
for each record y in Y do
    Success←1
    For each conditionQi do
        ifyTQiY
            Success0tl
            break;
        end if
    end for
    if success = 1 then
        add $y_i$into Y ′;
    end if
end for
return Y ′ to the client;
```

## V. KNN QUERY PROCESSING WITH RASP

Since the RASP bother does not save removes (and separate requests), kNN inquiry can't be straightforwardly prepared with the RASP annoyed information. In this segment, we plan a kNN inquiry handling calculation taking into account range questions (the kNN-R calculation). Therefore, the utilization of list in extent inquiry preparing likewise empowers quick handling of kNN inquiries.

### A. *Overview of the kNN-R Algorithm*

The first separation based kNN question preparing finds the nearest k focuses in the round range that is focused at the inquiry point. The fundamental thought of our calculation is to utilize square ranges, rather than circular reaches, to find the inexact kNN results, so that the RASP range inquiry administration can be utilized. There are various key issues to make this work safely and efficiently. 1) How to efficiently find the base square range that most likely contains the k results, without numerous connections between the cloud and the customer?

2) Will this arrangement safeguard information confidentiality and question protection? (3) Will the intermediary server's workload increment? what exactly degree? The calculation depends on square ranges to roughly find the kNN contender for an inquiry point, which are defined as takes after.

*Definition 1*:A square range is a hyper-shape that is focused at the question point and with equivalent length edges.

Figure 5 shows the reach question based kNN handling with two-dimensional information. The Inner Range is the square range that contains at any rate k focuses and the Outer Range encases the circular range that encases the internal extent. The external range definitely contains the kNN results (Proposition 2) however it might likewise contain superfluous focuses that should be sort out..

*Proposition 2:* The kNN-R algorithm returns results with 100% recall.

*Proof:*The sphere in Figure 5 between the outer range and the inner range covers all points with distances less than the radius r. Because the inner range contains at least k points, there are at least k nearest neighbors to the query points with distances less than the radius r. Therefore, the k nearest neighbors must be in the outer range.

The kNN-R algorithm consists of two rounds of interactions between the client and the server. Figure 4 demonstrates the procedure.

1) The client will send the initial upper-bound range, which contains more than k points, and the initial lower-bound range, which contains less than k points, to the server. The server finds the inner range and returns to the client.

2) The client calculates the outer range based on the inner range and sends it back to the server. The server finds the records in the outer range and sends them to the client. (3) The client decrypts the records and find the top k candidates as the final result. If the points are approximately uniformly distributed, we can estimate the precision of the returned result. With the uniform assumption, the number of points in an area is proportional to the size of the area.

If the inner range contains m points, m >= k, the outer range contains q points, and the dimensionality is d, we can derive $q = 2^{d/}2_m$. Thus, the precision is $k/q = k/(2^{d/}2_m)$. If m ≈ k and d = 2, the precision is around 0.5. When d increases, the precision decreases exponentially due to the curse of dimensionality [23], which suggests kNN-R should not work effectively on high-dimensional data. We will show this weakness in

### B. *Finding Compact Inner Square Range*

A critical stride in the kNN-R calculation is to find the minimal inward square range to accomplish high exactness. In the accompanying, we give the (k, δ)- range for efficiently finding the conservative inward range.

Definition 2: A (k, δ)- reach is any square range focused at the question point, the quantity of focuses in which is in the extent [k, k + δ],δ is an onnegative number.

We plan a calculation like double hunt to efficiently find the (k, δ)- range. Assume a square range focused at the inquiry point with length of L in every measurement is spoken to as S(L). Let the quantity of focuses included by this extent is N(L). On the off chance that a square range S(in) is encased by another square range S(in(out)), we say(out)S(in)⊂ S(out). It specifically takes after that N≤ N , furthermore Corollary 1: If N(1)< N(2), S(1)⊂ S(2). Utilizing this definition and documentation, we can simply build a progression of encased square ranges fixated on the inquiry point:S(L1)⊂S(L2)⊂. . . , ⊂ S(Lm).

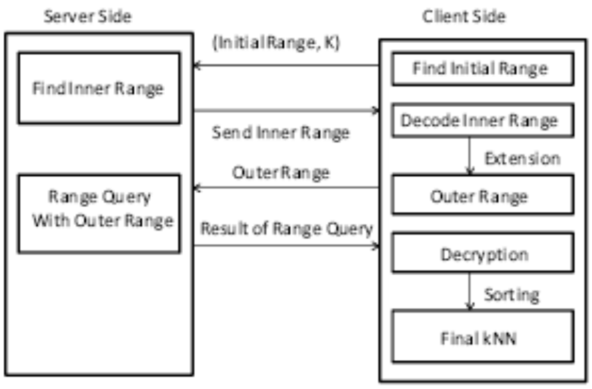

Fig. 2. Procedure of KNN-R algorithm

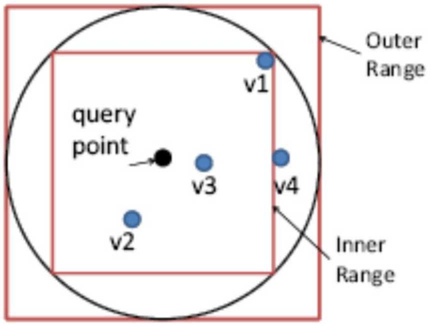

Fig. 3. Illustration for kNN-R Algorithm when k=3

Correspondingly, the numbers of points enclosed by $\{S^{(}L_{i})_{\}}$have the ordering$_{(}$L) $N^{(}L_{1})$ ≤ $N^{(}L_{2})$ ≤ . . . $N^{(}L_{m)}$. Assume that S1is the initial range containing less than k points and $S^{(}L_m)$ is the initial upper bound range; both are sent by the client. The problem of finding the compact inner range S can be mapped to a binary search over the sequence $\{S^{(}L_i)_{\}}$.

In each step of the binary search, we start with a lower bound range, denoted as $S^{(}low)$ and a higher bound range, $S^{(}high)$. We want the corresponding numbers of enclosed points to satisfyN(low)< k(high) N in each step, which is achieved with the following procedure. First, we find the middle square range $S^{(}mid)$, where mid = (low + high)/2. If $S^{(}mid)$

Covers no less than k points, the higher bound: $S^{(}high)$ is updated to $S^{(}mid)$; otherwise,

the lower bound: $S^{(}low)$ is updated to $S^{(}mid)$. At the beginning step $S^{(}low)$ is set to $S^{(}L_1)_{and}S^{(}high)_{is}S^{(}L_m)$. This process where E is some small positive number. Algorithm 4 in Appendix describes these steps.

1) *Selection of Initial Inner/Outer Bounds:* The selection of initial inner bound can be the query point.

If the query point is q(q1, . . . , qd), $S^{(}L_1)$ is a hyper cube defined by (Lm)qi≥Xi≥qi, i= 1. . . d}. The naïve selection of Swould be the whole domain. How ever, we can effectively reduce the range with a coarse density map organized in a tiny flat multidimensional tree, which can be included in the preprocessing step in the client side. The details will be ignored due to the space limitation.

### C. Finding Inner Range with RASP Perturbed Data

Algorithm 4 gives the basic ideas of finding the compact inner range in iterations. There are two critical operations in this algorithm: (1) finding the number of points in a square range and (2) updating the higher and lower bounds. Because range queries are secured in the RASP framework, the key is to update the bounds with the secured range queries, without the help of the client-side proxy server. As discussed in the RASP query processing, a range query such as $S^{(}L)$ is encoded as the $MBR^{(}L)$ of its polyhedron range in the perturbed space and the 2(d+2) dimensional conditions. $y^T\Theta^{(}i^{,L})_y \leq 0$ determining the sides of the polyhedron, and each of the d + 2 extended dimensions gets a pair of conditions for the upper and lower bounds, respectively. The problem of binary range search is to use the higher bound range $S^{(}high)$ and the lower bound range $S^{(}low)$ to derive $S^{(}mid)$. When all of these ranges are secured, the problem is transformed to (1) deriving $\Theta^{(}i^{,mid})$ from $\Theta^{(}ihigh)_{and}\Theta^{(}i^{,low})$; and (2) $S^{(}low)$ is set to $S^{(}L_1)_{and}S^{(}high)_{is}S^{(}L_m)$. This process repeats until $N^{(}mid)_{<\ k}+\delta$ or high− low < E, deriving $MBR^{(}mid)$ from $MBR^{(}high)$ and $MBR^{(}low)$. The following discussion will be focused on the simplified RASP version without the OPE component, which will be extended with the OPE component.

## VI. EXPERIMENTS

In this section, we present four sets of experimental results to investigate the following questions, correspondingly. (1) How expensive is the RASP perturbation?

(2) How resilient the OPE enhanced RASP is to the ICA-based ttack?

(3) How efficient is the two-stage range query processing?

(4) How efficient is the kNN-R query processing and what are the advantages?

### A. Datasets

we skip the details Three datasets are used in experiments. (1) A synthetic dataset that draws samples from uniform distribution in the range [0, 1]. (2) The Adult dataset from UCI machine learning database[5]. We assign numeric values to the categorical values using a simple one- to-one mapping scheme, as described in Section 3. (3) The 2-dimensional NorthEast location data from rtreeportal.org.

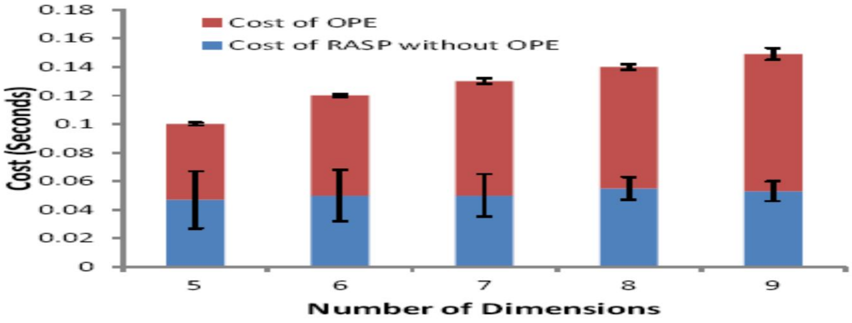

Fig. 4 The cost distribution of the full RASP scheme. Data: Adult (20K records,5-9 dimensions)

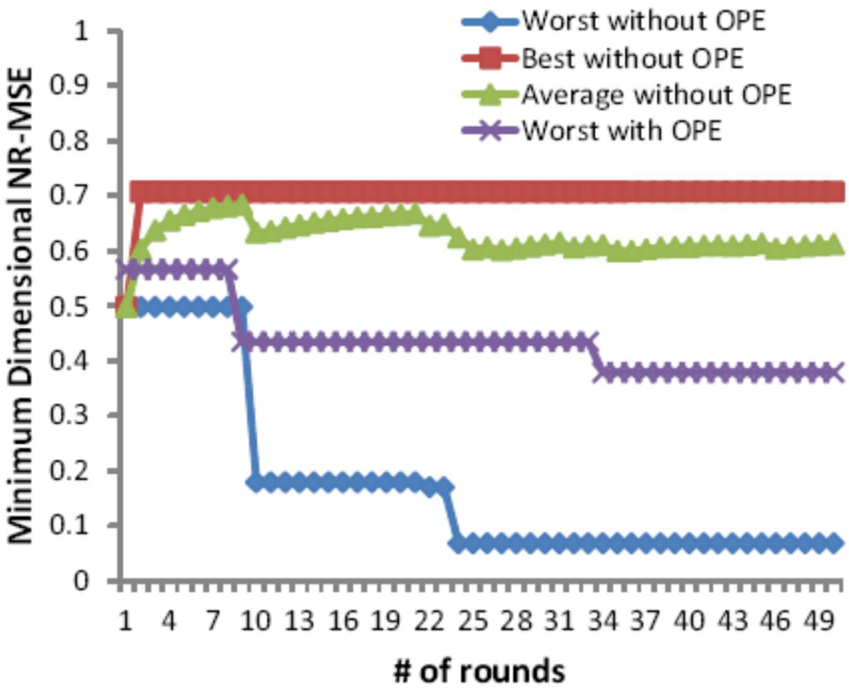

Fig. 5. Randomly generated matrix A and the progressive resilience to ICA attack. Data: Adult (10) dimensions, 10K records)

### B. Cost of RASP Perturbation

In this investigation, we concentrate on the expenses of the segments in the RASP annoyance. The real expenses can be separated into two sections: the OPE and the rest a portion of RASP. We actualize a basic OPE plan [1] by mapping unique section distributions to typical disseminations. The OPE calculation parcels the objective appropriation into basins. At that point, the sorted unique qualities are relatively apportioned by target can conveyance to make the basins for the first circulation. With the adjusted unique and target pails, a unique quality can be mapped to the objective can and properly scaled. In this manner, the encryption cost for the most part originates from the pail seek method (relative to log D, where D is the quantity of containers). Figure 6 demonstrates the cost conveyances for 20K records at various number of measurements. The dimensionality effectsly affects the expense of RASP annoyance. In general, the expense of preparing 20K records is just around 0.1 second.

### C. Resilience to ICA Attack

We have talked about the routines for countering the ICA distributional assault on the annoyed information. In this arrangement of trials, we assess how strong the RASP bother is to the distributional assault.

Results. We mimic the ICA assault for haphazardly picked matrices A. The information utilized as a part of the trial is the 10-dimensional Adult information with 10K records. Figure 7 demonstrates the dynamic results in various arbitrarily picked lattices A. The x-pivot speaks to the aggregate number of rounds for haphazardly picking the matrix A; the y-hub speaks to the base dimensional NR MSE among all measurement. Without OPE, the name "Best-without-OPE" speaks to the strongest An at the round i, "Most exceedingly bad without-OPE" speaks to the A of the weakest versatility, and "Normal without-OPE" is the normal nature of the created A lattices for i adjusts. We see that the best case is now near the upper bound 0.7 (Section 3.3). With the OPE part, the most pessimistic scenario can likewise be significantly moved forward.

### D. Performance of Two-stage Range Query Processing

In this arrangement of examinations, we think about the execution parts of polyhedron-based reach question preparing. We utilize the two-stage preparing procedure portrayed in Section 4, and investigate the extra cost brought about by this handling methodology. We actualize the two-stage inquiry handling in view of a R*tree usage gave by Dr. Hadjieleftheriou at AT&T Lab6. The piece size is 4KB and we permit every square to contain just 20 passages to imitate a substantial database with numerous circle pieces. Tests from the first databases in various size (10,000 − 50,000 records, i.e., 500-2500 information squares) are bothered and listed for inquiry handling. Another arrangement of lists is additionally based on the first information for the execution correlation with non-annoyed question preparing. We will utilize the quantity of circle piece gets to, including list squares and information pieces, to survey the execution to stay away from the conceivable variety created by different parts of the PC framework. Furthermore, we will likewise demonstrate the divider clock time for a few results. Review the two-stage handling system: utilizing the MBR to seek the indexing tree, and filtering the returned result with the secured question in quadratic structure. We will concentrate on the execution of the first stage by contrasting it with two extra routines: (1) the first questions with the list based on the first information, which is utilized to recognize how much extra cost is paid for questioning the MBR of the changed inquiry; (2) the direct sweep approach, which is the most pessimistic scenario cost. Range inquiries are produced arbitrarily inside of the area of the datasets, and after that changed with the technique portrayed in the Section 4. We likewise control the scope of the questions to be [10%,20%,30%,40%,50%] of the aggregate scope of the area, to watch the impact of the size of the reach to the execution of inquiry handling. Results. The first pair of figures (the left subfigures of Figure 8 and 9) demonstrates the quantity of piece gets to for 10,000 questions on various sizes of information with various inquiry handling routines. For clear presentation, we utilize log10(# of square gets to) as the y-pivot. The expense of straight sweep is essentially the quantity of squares for putting away the entire dataset. The information dimensionality is fixed to 5 and the question extent is set to 30% of the entire space. Clearly, the first stage with MBR for polyhedron has a cost much less expensive than the straight output strategy and just reasonably higher than R*tree preparing on the first information. Interestingly, distinctive appropriations of information result in marginally diverse examples. The expenses of R*tree on changed questions are near those of unique inquiries for Adult information, while the crevice is bigger on uniform information. The expenses over various measurements and diverse inquiry ranges show comparable examples. We likewise concentrated on the expense of the second stage. We utilize "PrepQ" to speak to the customer side expense of changing questions, "virtue" to speak to the rate (last result check)/(first stage result tally), and records per inquiry ("RPQ") to speak to the normal number of records per inquiry for the first stage results. The quadratic separating conditions are utilized as a part of analyses. Table 1 thinks about the normal divider clock time (milliseconds) per inquiry for the two stages, the RPQ values for stage 1, and the virtue of the stage-1 result. The tests are keep running with the setting of 10K questions, 20K records, 30% dimensional inquiry range and 5 measurements. Subsequent to the second stage is done in memory, its expense is much lower than the first stage cost. Generally speaking, the two stage preparing is much speedier than direct output and practically identical to the first R*Tree processing.

### *E. Performance of kNN-R Query Processing*

In this set of experiments, we investigate several aspects of kNN query processing. (1) We will study the cost of (k, δ)- ange algorithm, which mainly contributes to the server-side cost. (2) We will show the overall cost distribution over the cloud side and the proxy server. (3) We will show the advantages of kNN-R over another popular approach: the Casper approach [24] for privacy-preserving kNN search.

(k, δ)-Range Algorithms In this set of experiments, we want to understand how the setting of the δ parameter effects the performance and the result precision. Figure 10 shows the effect of δ setting to the (k, δ)-range algorithm. Both datasets are two- dimensional data. As δ becomes larger, both the precision and the number of rounds needs to reach the δ condition decreases. Note that each round corresponds to one server-side range query. The choice of δ represents a tradeoff between the precision and the performance.

As we have discussed, the major weakness with the kNN-R algorithm is the precision reduction with increased dimensionality. When the dimensionality increases, the precision can significantly drop, which will increase the cost of post-processing in the client side. Figure 11 shows this phenomenon with the real Adult data and the simulated uniform data. However, compared to the overall cost, the client-side cost increase is still acceptable. We will show the comparison next.

Overall Costs. Many secure approaches cannot use indices for query processing, which results in poor performance. For example, the secure dot-product approach [33] encodes the points with random projections and recovers dotproducts in query processing for distance comparison. The way of encoding data disallows the index-based query processing. Without the aid of indices, processing a kNN query will have to scan the entire database, leaving many optimization
impossible to implement.

One concern with the kNN-R approach is the work load on the proxy server. Different from range query, the proxy server will need to filter out the points returned by the server to find the final kNN. A reduced precision due to the increased dimensionality will imply an increased burden for the proxy server. We need to show how significant this proxy cost is. We use the database of 100 thousands of data points and 1000 randomly selected queries for the 1NN experiment. The wall clock time (milliseconds) is used to show the average cost per query in Table 2. We also list the cost of the secure dot-product method [33] for comparison. Table 2 shows that the proxy

server takes a negligible pre-processing cost and a very small post-processing cost, even for reduced precision in the 5D datasets.We use 5% domain length to extend the query point to form the initial higher bound. Compared to the dot-product method, the user-specified higher bound setting can cut off uninteresting regions, giving significant performance gain for sparse or skewed datasets, such as Adult5D. This cut-off effect cannot be implemented with the dotproduct method. Furthermore, even for dense cases like the 2D datasets, the overall cost is only about half of the dot-product method.

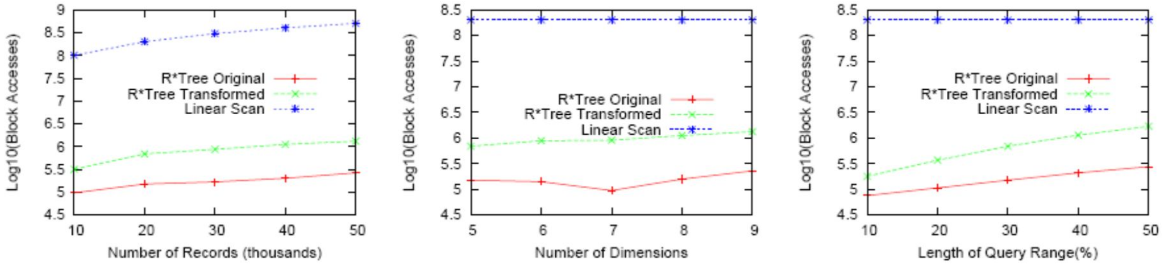

Fig. 6. Performance comparison on Uniform data. Left: data size vs. cost of query; Middle: data dimensionality vs. cost of query; Right: query range (percentage of the domain) vs. cost of query

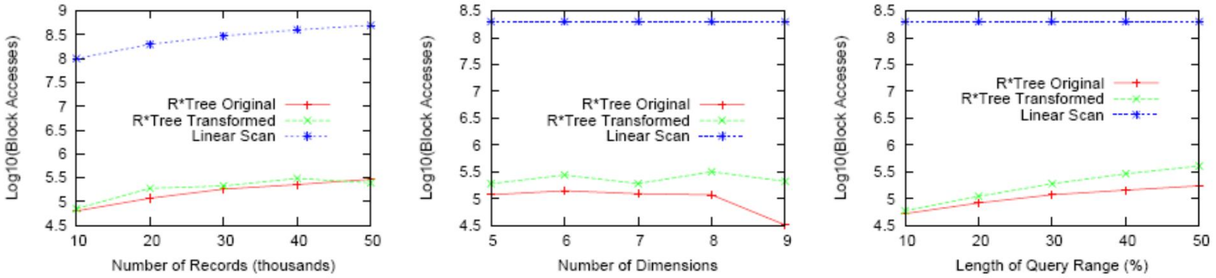

Fig. 7. Performance comparison on Adult data. Left: data size vs. cost of query; Middle: data imensionality vs. cost of query; Right: query range (percentage of the domain) vs. cost of query

Comparing kNN-R with the Casper Approach. In this set of experiments, we compare our approach and the Casper approach with a focus on the tradeoff between the data confidentiality and the query result precision (which indicates the workload of the inhouse proxy). Based on the description in the paper [24], we implement the 1NN query processing algorithm for the experiment. The Casper approach uses cloaking boxes to hide both the original data points in the database and the query points. It can also use the index to process kNN queries. The onfidentiality of data in Casper is solely defined by the size of cloaking box. Roughly speaking, the actual point has the same probability to be anywhere in the cloaking box. However, the size of cloaking box also directly affects the precision

of query results. Thus, the decision on the box size represents a tradeoff between the precision of queryresults and the data confidentiality.

For clear presentation, we assume each dimension has the same length of domain, h and each cloaking box is square with an edge-length e. Assume the whole domain also has a uniform distribution. According to the variance of uniform distribution, the NR MSE measure is $\sqrt{6}e/(3h)$. To achieve the protection of 10% domain length, we have $e \approx 0.12h$. In Figure 12, the x-axis represents NR MSE, i.e., the Casper's relative cloaking-edge length. It shows that when the edge length is increased from 2% to 10%, the precision dramatically drops from 62% to 13% for the 2D uniform data and 43% to 10% for the 2D NE data, which shows the severe conflict between precision and confidentiality. The kNN-R's results are also shown for comparison.

## VII. CONCLUSION

We propose the RASP irritation way to deal with facilitating inquiry administrations in the cloud, which satisfies the CPEL criteria:

information Confidentiality, question Privacy, Efficient inquiry preparing, and Low in-house work load. The necessity on low in-house workload is a basic element to completely understand the benefits of distributed computing, and efficient question handling is a key measure of the nature of inquiry administrations. Scratch irritation is an exceptional piece of OPE, dimensionality development, irregular commotion infusion, and arbitrary projection, which gives novel security highlights. It plans to safeguard the topology of the questioned territory in the bothered space, and permits to utilize files for efficient range inquiry preparing. With the topology-safeguarding highlights, we can create efficient range question administrations to accomplish sublinear time unpredictability of handling inquiries. We then build up the kNN question administration taking into account the reach inquiry administration. The security of both the bothered information and the ensured questions is painstakingly broke down under an accurately defined danger model. We additionally direct a few arrangements of investigations to demonstrate the efficiency of inquiry handling and the minimal effort of in-house preparing. We will proceed with our studies on two viewpoints: (1) further enhance the execution of question preparing for both reach inquiries and kNN inquiries; (2) formally break down the spilled inquiry and access designs and the conceivable impact on both information and query confidentiality

## REFERENCES

[1] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order preserving encryption for numeric data,” in Proceedings of ACM SIGMOD Conference, 2004.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. K. andAndy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the clouds: A berkeley view of cloud computing,” Technical Report, University of Berkerley, 2009.

[3] J. Bau and J. C. Mitchell, “Security modeling and analysis,” IEEE Security and Privacy, vol. 9, no. 3, pp. 18–25, 2011.

[4] S. Boyd and L. Vandenberghe, Convex Optimization. Cam-bridge University Press, 2004.

[5] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy- preserving multi-keyword ranked search over encrypted cloud data,” in INFOCOMM, 2011.

[6] K. Chen, R. Kavuluru, and S. Guo, “Rasp: Efficient mul-tidimensional range query on attack-resilient encrypted databases,” in ACM Conference on Data and Application Security and Privacy, 2011, pp. 249–260.

[7] K. Chen and L. Liu, “Geometric data perturbation for out- sourced data mining,” Knowledge and Information Systems, 2011.

[8] K. Chen, L. Liu, and G. Sun, “Towards attack-resilient geomet-ric data perturbation,” in SIAM Data Mining Conference, 2007.

[9] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, “Privateinformation retrieval,” ACM Computer Survey, vol. 45, no. 6, pp. 965–981, 1998.

[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Search-able symmetric encryption: improved definitions and efficient constructions,” in Proceedings of the 13th ACM conference onComputer and communications security. New York, NY, USA:ACM, 2006, pp. 79–88.

[11] N. R. Draper and H. Smith, Applied Regression Analysis. Wiley,1998.

[12] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, “Executing sql over encrypted data in the database-service-provider model,” in Proceedings of ACM SIGMOD Conference, 2002.

[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer-Verlag, 2001.

[14] B. Hore, S. Mehrotra, and G. Tsudik, “A privacy-preserving index for range queries,” in Proceedings of Very Large Databases Conference (VLDB), 2004.

[15] H. Hu, J. Xu, C. Ren, and B. Choi, “Processing private queries over untrusted data cloud through privacy homomorphism,”Proceedings of IEEE International Conference on Data Engineering(ICDE), pp. 601–612, 2011.

[16] Z. Huang, W. Du, and B. Chen, “Deriving private information from randomized data,” in Proceedings of ACM SIGMOD Conference, 2005.

[17] A. Hyvarinen, J. Karhunen, and E. Oja, Independent Component Analysis. Wiley, 2001.

[18] I. T. Jolliffe, Principal Component Analysis. Springer, 1986.

[19] F. Li, M. Hadjieleftheriou, G. Kollios, and L. Reyzin, “Dynamic authenticated index structures for outsourced databases,” in Proceedings of ACM SIGMOD Conference, 2006.

[20] K. Liu, C. Giannella, and H. Kargupta, “An attacker’s view of distance preserving maps for privacy preserving data mining,” in Proceedings of PKDD, Berlin, Germany, September 2006.

[21] M. L. Liu, G. Ghinita, C. S.Jensen, and P. Kalnis, “Enabling search services on outsourced private spatial data,” The International Journal of on Very Large Data Base, vol. 19, no. 3, 2010.

[22] Y. Manolopoulos, A. Nanopoulos, A. Papadopoulos, and Y. Theodoridis, R-trees: Theory and Applications. Springer- Verlag, 2005.

[23] R. Marimont and M. Shapiro, “Nearest neighbour searches and the curse of dimensionality,” Journal of the Institute of Mathematics and its Applications, vol. 24, pp. 59–70, 1979.

[24] M. F. Mokbel, C. yin Chow, and W. G. Aref, “The new casper: Query processing for location services without compromising privacy,” in Proceedings of Very Large Databases Conference (VLDB), 2006, pp. 763–774.

[25] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in EUROCRYPT. Springer-Verlag, 1999, pp. 223–238.

[26] S. Papadopoulos, S. Bakiras, and D. Papadias, “Nearest neighbor search with strong location privacy,” in Proceedings of Very Large Databases Conference (VLDB), 2010.

[27] F. P. Preparata and M. I. Shamos, Computational Geometry: An Introduction. Springer-Verlag, 1985.

[28] M. Rudelson and R. Vershynin, “Smallest singular value of a random rectangular matrix,” Communications on Pure and Applied Mathematics, vol. 62, pp. 1707–1739, 2009.

[29] E. Shi, J. Bethencourt, T.-H. H. Chan, D. Song, and A. Perrig, “Multi-dimensional range query over encrypted data,” in IEEE Symposium on Security and Privacy, 2007.

[30] R. Sion, “Query execution assurance for outsourced databases,” in Proceedings of Very Large Databases Conference (VLDB), 2005.

[31] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure ranked keyword search over encrypted cloud data,” in Proceedings of IEEE International Conference on

Distributed Computing Systems(ICDCS), 2010.

[32] P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage," in ACM Conference on Computer and Communications Security, 2008.

[33] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proceedingsof ACM SIGMOD Conference. New York, NY, USA: ACM, 2009, pp. 139–152.

[34] M. Xie, H. Wang, J. Yin, and X. Meng, "Integrity auditing of outsourced data," in Proceedings of Very Large DatabasesConference (VLDB), 2007, pp. 782–793.

[35] M. L. Yiu, C. S. Jensen, X. Huang, and H. Lu, "Spacetwist: Managing the trade-offs among location privacy, query perfor mance, and query accuracy in mobile services," in Proceedings of IEEE International Conference on Data Engineering (ICDE), Washington, DC, USA, 2008, pp. 366–375.