International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

Efficient Access of Cloud Storage with Confidentiality by RASP Data Perturbation

Lattupally shruthi
Department of Computer science and Engineering Aurora's Technological and Research Institute
Email id:shruthi.reddy1223@gmail.com
Ms. K. Kavitha
(Associate Professor)
Department of Computer science and Engineering Aurora's Technological and Research Institute
Email id:kavithakbjr@gmail.com

Abstract—
With the wide deployment of public cloud computing infrastructures, using clouds to host data query services has become an appealing solution for the advantages on scalability and cost-saving. However, some data might be sensitive that the data owner does not want to move to the cloud unless the data confidentiality and query privacy are guaranteed. On the other hand, a secured query service should still provide efficient query processing and significantly reduce the in-house workload to fully realize the benefits of cloud computing. We propose the RASP data perturbation method to provide secure and efficient range query and kNN query services for protected data in the cloud. The RASP data perturbation method combines order preserving encryption, dimensionality expansion, random noise injection, and random projection, to provide strong resilience to attacks on the perturbed data and queries. It also preserves multidimensional ranges, which allows existing indexing techniques to be applied to speedup range query processing. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the kNN queries. We have carefully analyzed the attacks on data and queries under a precisely defined threat model and realistic security assumptions. Extensive experiments have been conducted to show the advantages of this approach on efficiency and security.

I INTRODUCTION

Data hosting in the cloud is increasingly because of the unique advantages in scalability and cost-saving. With the help of cloud infrastructures, the service owners can conveniently scale up or down the service and only pay for the hours of using the servers. This has become an attractive feature, since, the workloads of query services are highly dynamic, and is very expensive and inefficient to serve such dynamic workloads with in-house infrastructures. However, data confidentiality and query privacy have become the major concerns, since, the service providers lose the control over the data in the cloud. Adversaries, such as curious service providers, can possibly make a copy of the database or eavesdrop users' queries, which will be difficult to detect and prevent in the cloud infrastructures. While new approaches are needed to preserve data confidentiality and query privacy, the efficiency of query services and the benefits of using the clouds should also be preserved. It will not be meaningful to provide slow query services as a result of security and privacy assurance. It is also not practical for the data owner to use a significant amount of in house resources, because the purpose of using cloud resources is to reduce the need of maintaining scalable in-house infrastructures.

Therefore, there is an intricate relationship among the data confidentiality, query privacy, the quality of service, and the economics of using the cloud. So, here we do summarize these requirements for constructing a practical query service in the cloud as the CPEL criteria: query Privacy, data Confidentiality, Low in-house processing cost and Efficient query processing, Satisfying these requirements will increase the complexity in constructing query services in the cloud. In order to address these aspect of the problems, Some related approaches are proposed. However, they fail to address these problems. For example, the Order Preserving Encryption (OPE) and crypto-index are vulnerable to the attacks.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

The enhanced crypto-index approach puts heavy burden on the in-house infrastructure to improve the security and privacy. Cloaking boxes is used by the New Casper approach to secure data objects and queries, which does affect the efficiency of query processing and the in-house workload. We propose the Random Space Perturbation (RASP) approach to constructing practical range query and k-nearest-neighbor (kNN) query services in the cloud. The proposed approach will address all the four aspects of the CPEL criteria and aim to achieve a good balance on them. The basic idea is to randomly transform the multidimensional datasets with a combination of dimensionality expansion, order preserving encryption, random project, and random noise injection, so that the utility for processing range queries is preserved.

Purpose

In the RASP perturbation the queried ranges are securely transformed into polyhedral in the RASP-perturbed data space, which can be processed with the support of indexing structures in the perturbed space. The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries. The key components in the RASP framework include

(1) the definition and properties of RASP perturbation;
(2) the construction of the privacy-preserving range query services;
(3) the construction of privacy-preserving kNN query services; and
(4) an analysis of the attacks on the RASP-protected data and queries.

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

- The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee.
- The proposed service constructions are able to minimize the in-house processing
- 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 did evaluated our approach with synthetic and real datasets. And the obtained results show its unique advantages on all aspects of the CPEL criteria.

II SYSTEM ANALYSIS

Existing System

With the wide deployment of public cloud computing infrastructures, using clouds to host data query services has become an appealing solution for the advantages on scalability and cost-saving. However, some data might be sensitive that the data owner does not want to move to the cloud unless the data confidentiality and query privacy are guaranteed. On the other hand, a secured query service should still provide efficient query processing and significantly reduce the in-house workload to fully realize the benefits of cloud computing.

Disadvantages

Adversaries, such as curious service providers, can possibly make a copy of the database or eavesdrop users' queries, which will be difficult to detect and prevent in the cloud infrastructures.

Proposed System

We propose the Random Space Perturbation (RASP) approach to constructing practical range query and k-nearest-neighbor (kNN) query services in the cloud. The proposed approach will address all the 2 four aspects of the CPEL criteria and aim to achieve a good balance on them. The basic idea is to randomly transform the multidimensional datasets with a combination of order preserving encryption, dimensionality expansion, random noise injection, and random project, so that the utility for processing range queries is preserved.

The RASP perturbation is designed in such a way that the queried ranges are securely transformed into polyhedra in the RASP-perturbed data space, which can be efficiently processed with the support of indexing structures in the perturbed space. The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries. The key components in the RASP framework include (1) the definition and properties of RASP perturbation; (2) the

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

construction of the privacy-preserving range query services; (3) the construction of privacy-preserving kNN query services; and (4) an analysis of the attacks on the RASP-protected data and queries.

Advantages:

1) The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee.
2) 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.

III RASP: RANDOM SPACE PERTURBATION

RASP is one type of multiplicative perturbation, with a novel combination of OPE, dimension expansion, random noise injection, and random projection. Let's consider the multidimensional data are numeric and in multidimensional vector space1. The database has k searchable dimensions and n records, which makes a d × n matrix X. The *searchable* dimensions can be used in queries and thus should be indexed. Let x represent a d-dimensional record, $\mathbf{x} \in \mathbf{R}^{\mathbf{d}}$. Note that in the d-dimensional vector space Rd, 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.

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, Eope with keys Kope, is applied to each dimension of x: Eope(x,Kope) ∈ Rd 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 1 and the (d + 2)- th dimension, v, is drawn from a random real number generator RNG 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
   $F(x,K = \{A,Kope,RG\}) = A((Eope(x))^T , 1, v)^T$, (1) where A is a (d+2)×(d+2) randomly generated invertible matrix with $a_{ij} \in R$ such that there are at least two non-zero values in each row of A and the last column of A is also non-zero.

Properties of RASP

- RASP does not preserve the order of dimensional values because of the matrix multiplication component, which distinguishes itself from order preserving encryption (OPE) schemes, and thus does not suffer from the distribution-based attack (details in Section 7). An OPE scheme maps a set of single-dimensional values to another, while keeping the value order unchanged. Since the RASP perturbation can be treated as a combined transformation F(G(Eope(x))), it is sufficient to show that F(y) = Ay does not preserve the order of dimensional values, where $y \in R^{d+2}$ and $A \in R^{(d+2)x(d+2)}$. The proof is straightforward as shown in Appendix.
- RASP does not preserve the distances between records, which prevents the perturbed data from distance-based attacks [8]. Because none of the transformations in the RASP: Eope, G, and F preserves distances, apparently the RASP perturbation will not preserve distances. Similarly, RASP does not preserve other more sophisticated structures such as covariance matrix and principal components. Therefore, the PCA-based attack do not work as well.
- The original range queries can be transformed to the RASP perturbed data space, which is the basis of our query processing strategy. A range query describes a hyper-cubic area (with possibly open bounds) in the multidimensional space.

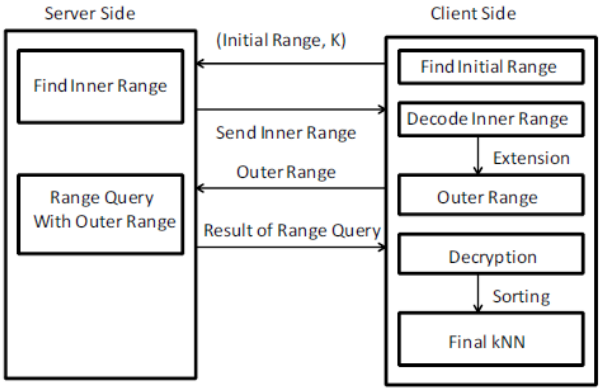
IV KNN QUERY PROCESSING WITH RASP

(A) Finding Inner Range with RASP Perturbed Data

Below algorithm 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

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

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.



Algorithm : Procedure of KNN-R algorithm

(B) Overview of the kNN-R Algorithm

The original distance-based kNN query processing finds the nearest k points in the *spherical range* that s centered at the query point. The basic idea of our algorithm is to use square ranges, instead of spherical ranges, to find the approximate kNN results, so that the RASP range query service can be used. There are a number of key problems to make this work securely and efficiently. (1) How to efficiently find the minimum square range that surely contains the k results, without many interactions between the cloud and the client? (2) Will this solution preserve data confidentiality and query privacy? (3) Will the proxy server's workload increase? to what extent?

The algorithm is based on *square ranges* to approximately find the kNN candidates for a query point, which are defined as follows.

Definition 1*:* A square range is a hyper-cube that is centered at the query point and with equal-length edges.

Figure 1 illustrates the range-query-based kNN processing with two-dimensional data. The *Inner Range* is the square range that contains at least k points, and the *Outer Range* encloses the spherical range that encloses the inner range. The outer range surely contains the kNN results (Proposition 2) but it may also contain irrelevant points that need to be filtered out.

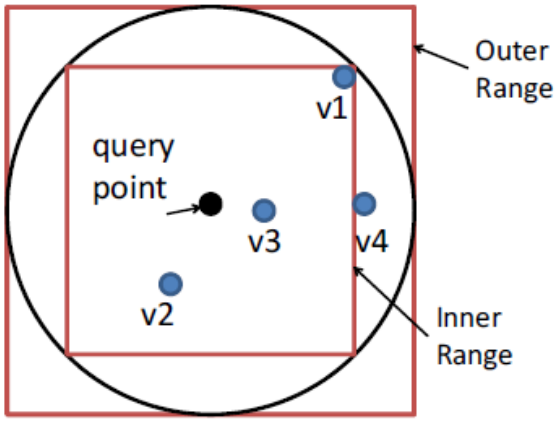


Fig. Illustration for kNN-R Algorithm when k=3.

Private Information Retrieval

Private Information Retrieval (PIR) is a protocol that allows a client to retrieve an element of a database without the owner of that database being able to determine which element was selected. While this problem admits a trivial solution - sending the entire database to the client allows the client to query with perfect privacy - there are techniques to reduce the communication complexity of this problem, which can be critical for large databases. Additionally, Strong Private Information Retrieval (SPIR) is private information retrieval with the additional requirement that the client only learn about the elements he is querying for, and nothing else. This requirement captures the typical privacy needs of a database owner.

It also enables a user to access k replicated copies of a database (k $ 2) and privately retrieve information stored in the database. This means that each individual server (holding a replicated copy of the database) gets no information on the identity of the item retrieved by the user. Our schemes use the replication to gain substantial saving. In particular, we present a two-server scheme with communication complexity O(n1/3).

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

Order Preserving Encryption Scheme

Order Preserving Encryption Scheme allows comparison operations to be directly applied on encrypted data, without decrypting the operands. Thus, equality and range queries as well as the MAX, MIN, and COUNT queries can be directly processed over encrypted data. Similarly, GROUP BY and ORDER BY operations can also be applied. Only when applying SUM or AVG to a group do the values need to be decrypted. OPES is also endowed with the following properties:

- The results of query processing over data encrypted using OPES are exact. They neither contain any false positives nor miss any answer tuple. This feature of OPES sharply differentiates it from schemes such as [13] that produce a superset of answer, necessitating filtering of extraneous tuples in a rather expensive and complex post-processing step.
- OPES handles updates gracefully. A value in a column can be modified or a new value can be inserted in a column without requiring changes in the encryption of other values.
- OPES can easily be integrated with existing database systems as it has been designed to work with the existing indexing structures such as B-trees. The fact that the database is encrypted can be made transparent to the applications.

It also handles updates gracefully and new values can be added without requiring changes in the encryption of other values. It allows standard database indexes to be built over encrypted tables and can easily be integrated with existing database systems.

A privacy-preserving index for range queries

A privacy-preserving index for range queries, enables, an untrusted server to evaluate obfuscated range queries with minimal information leakage. It analyzes the worst-case scenario of inference attacks that can potentially lead to breach of privacy (e.g., estimating the value of a data element within a small error margin) and identify statistical measures of data privacy in the context of these attacks. We also investigate precise privacy guarantees of data partitioning which form the basic building blocks of our index. We then develop a model for the fundamental privacy-utility tradeoff and design a novel algorithm for achieving the desired balance between privacy and utility (accuracy of range query evaluation) of the index.

It also focuses on range queries and a bucketization based approach to support them in the DAS model. In the bucketization approach, an attribute domain is partitioned into a set of buckets each of which is identified by a tag. These bucket tags are maintained as an index (referred to as crypto-index) and are utilized by the server to process the queries. The main goal is to characterize the privacy threats arising from the creation of bucketization-based indices to support range queries.

V EXPERIMENTAL RESULTS

Here, 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 attack? (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

Three datasets are used in experiments.

- A synthetic dataset that draws samples from uniform distribution in the range
- The Adult dataset from UCI machine learning database5. We assign numeric values to the categorical values using a simple one to- one mapping scheme.
- The 2-dimensional NorthEast location data from rtreeportal.org.

(B) Cost of RASP Perturbation

The major costs can be divided into two parts: the OPE and the rest part of RASP. We implement a simple OPE scheme by mapping original column distributions to normal distributions. The sorted original values are proportionally partitioned according to the target bucket distribution to create the buckets for the original distribution. With the aligned original and target buckets, an original value can be mapped to the target bucket and appropriately scaled. Therefore, the encryption cost mainly comes from the bucket search procedure (proportional to logD, where D is the number of buckets).

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

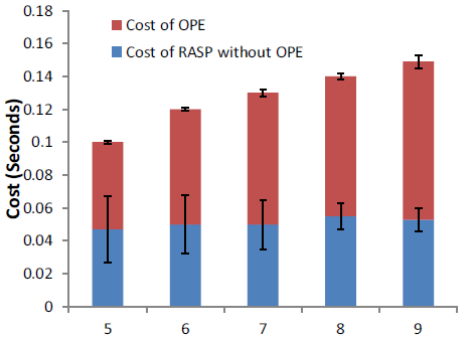


Fig:. The cost distribution of the full RASP scheme. Data: Adult (20K records,5-9 dimensions)

The above figure shows the cost distributions for 20K records at different number of dimensions. The dimensionality has slight effects on the cost of RASP perturbation. Overall, the cost of processing 20K records is only around 0.1 second.

(C) Performance of Two-stage Range Query Processing

The performance aspects of polyhedron-based range query processing is studied in this section. We use the two-stage processing strategy and explore the additional cost incurred by this processing strategy. We implement the two-stage query processing based on an R*tree implementation. The block size is 4KB and we allow each block to contain only 20 entries to mimic a large database with many disk blocks. Samples from the original databases in different size (10,000 − 50,000 records, i.e., 500-2500 data blocks) are perturbed and indexed for query processing. Another set of indices is also built on the original data for the performance comparison with non-perturbed query processing.

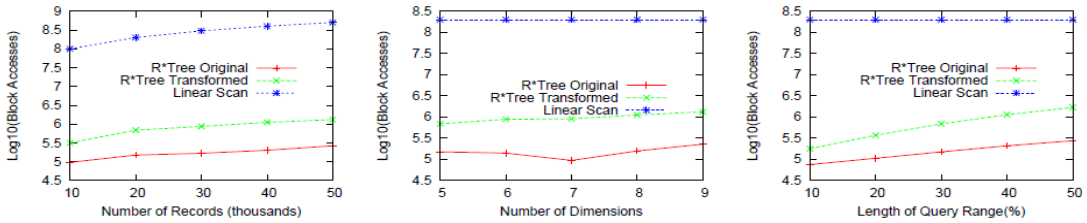


Fig.A Performance comparison on Uniform data. Left: data size vs. cost of query; Middle: data dimensionality vs. cost of query; Right: query range (percentage of the domain) vs. cost of query

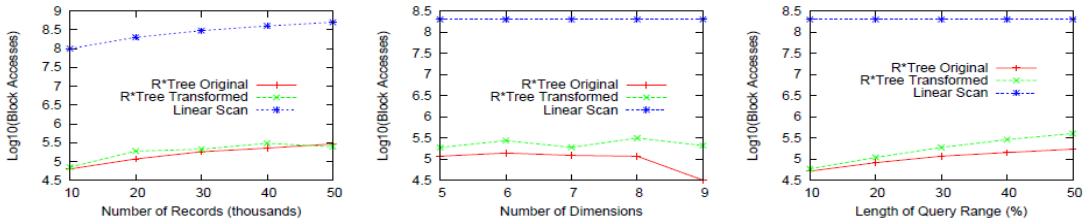


Fig.B Performance comparison on Adult data. Left: data size vs. cost of query; Middle: data dimensionality vs. cost of query; Right: query range (percentage of the domain) vs. cost of query

We will use the number of disk block accesses, including index blocks and data blocks, to assess the performance to avoid the possible variation caused by other parts of the computer system. In addition, we will also show the wall-clock time for some results. Recall the two-stage processing strategy: using the MBR to search the indexing tree, and filtering the returned result with the secured query in quadratic form. We will study the performance of the first stage by comparing it to two additional methods:

- The original queries with the index built on the original data, which is used to identify how much additional cost is paid for querying the MBR of the transformed query;
- The linear scan approach, which is the worst case cost. Range queries are generated randomly within the domain of the datasets, and then transformed

We also control the range of the queries to be [10%,20%,30%,40%,50%] of the total range of the domain, to observe the effect of the scale of the range to the performance of query processing.

Results.

The first pair of figures (the left subfigures of Figure A and B) shows the number of block accesses for 10,000 queries on different sizes of data with different query processing methods. For clear presentation, we use log10(# of block accesses) as the y-axis. The cost of linear scan is simply the number of blocks for storing

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

the whole dataset. The data dimensionality is fixed to 5 and the query range is set to 30% of the whole domain. Obviously, the first stage with MBR for polyhedron has a cost much cheaper than the linear scan method and only moderately higher than R*tree processing on the original data. Interestingly, different distributions of data result in slightly different patterns. The costs of R*tree on transformed queries are very close to those of original queries for Adult data, while the gap is larger on uniform data. The costs over different dimensions and different query ranges show similar patterns.

	Linear Scan	R*Tree-Orig	PrepQ	Stage-1	Stage-2
Uniform5D	21.12	0.27	0.007	4.19	0.01
Adult5D	16.28	0.39	0.007	1.9	0.01

TABLE 1 Wall clock cost distribution (milliseconds) and comparison

We also studied the cost of the second stage. We use "PrepQ" to represent the client-side cost of transforming queries, "purity" to represent the rate (final result count)/(1st stage result count), and records per query ("RPQ") to represent the average number of records per query for the first stage results. The quadratic filtering conditions are used in experiments. Table 1 compares the average wall-clock time (milliseconds) per query for the two stages, the RPQ values for stage 1, and the purity of the stage-1 result. The tests are run with the setting of 10K queries, 20K records, 30% dimensional query range and 5 dimensions. Since the 2nd stage is done in memory, its cost is much lower than the 1st-stage cost. Overall, the two stage processing is much faster than linear scan and comparable to the original R*Tree processing.

VI RELATED WORK

(A) Protecting Outsourced Data

- **Order Preserving Encryption.** Order preserving encryption (OPE) preserves the dimensional value order after encryption. It can be described as a function $\mathbf{y = F(x)}$, $\forall \mathbf{x_i, x_j}$, $\mathbf{x_i < (>,=) x_j \Leftrightarrow y_i < (>,=) y_j}$. A well-known attack is based on attacker's prior knowledge on the original distributions of the attributes. If the attacker knows the original distributions and manages to identify the mapping between the original attribute and its encrypted counterpart, a bucket based distribution alignment can be performed to break the encryption for the attribute
- **Crypto-Index.** Crypto-Index is also based on column-wise bucketization. It assigns a random ID to each bucket; the values in the bucket are replaced with the bucket ID to generate the auxiliary data for indexing. To utilize the index for query processing, a normal range query condition has to be transformed to a set-based query on the bucket IDs. For example, Xi < ai might be replaced with X' i ∈ [ID1, ID2, ID3].
- **Distance-Recoverable Encryption.** DRE is the most intuitive method for preserving the nearest neighbor relationship. Because of the exactly preserved distances, many attacks can be applied. One drawback is the search algorithm is limited to linear scan and no indexing method can be applied.

(B) Preserving Query Privacy

Private information retrieval (PIR) tries to fully preserve the privacy of access pattern, while the data may not be encrypted and these are very costly. Focusing on the efficiency side of PIR, Williams et al. use a pyramid hash index to implement efficient privacy preserving data-block operations based on the idea of Oblivious RAM. It is different from our setting of high throughput range query processing. Hu et al. addresses the query privacy problem and requires the authorized query users, the data owner, and the cloud to collaboratively process kNN queries.

However, most computing tasks are done in the user's local system with heavy interactions with the cloud server. The cloud server only aids query processing, which does not meet the principle of moving computing to the cloud. uses private information retrieval methods to enhance location privacy. Space Twist proposes a method to query kNN by providing a fake user's location for preserving location privacy. But the method does not consider data confidentiality, as well. The Casper approach considers both data confidentiality and query privacy, the detail of which has been discussed in our experiments.

VII CONCLUSION

We introduce the RASP perturbation approach for hosting query services in the cloud, which satisfies the CPEL criteria: data Confidentiality, query Privacy, Efficient query processing, and Low in-house workload. The requirement on low in-house workload is a critical feature to fully realize the benefits of cloud computing, and efficient query processing is a key measure of the quality of query services. RASP perturbation is a unique composition of OPE, random noise injection, random projection, and dimensionality expansion, which provides unique security features.

RASP aims to preserve the topology of the queried range in the perturbed space, and allows to use indices for efficient range query processing. With the topology-preserving features, we are able to develop efficient range query services to achieve sub linear time complexity of processing queries. We then develop the kNN query service based on the range query service. The security of both the perturbed data and the protected queries is carefully analyzed under a precisely defined threat model. We also conduct several sets of experiments to show the efficiency of query processing and the low cost of in-house processing.

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*. Cambridge University Press, 2004.

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

[6] K. Chen, R. Kavuluru, and S. Guo, "Rasp: Efficient multidimensional 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 outsourced data mining," *Knowledge and Information Systems*, 2011.

[8] K. Chen, L. Liu, and G. Sun, "Towards attack-resilient geometric data perturbation," in *SIAM Data Mining Conference*, 2007.

[9] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *ACM Computer Survey*, vol. 45, no. 6, pp. 965–981, 1998.

[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer 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.