ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The Integrated Secure RASE – RASP Query Services

Ponneri Srinivasulu
M.Tech Student,
Audisankara College of Engineering & Technology.

Kadiveti Koteswar Rao, M.Tech
Assistant Professor,
Audisankara College of Engineering & Technology.

Abstract:

The latest developments in cloud technologies have enabled them to host database and data warehouses. This led to the new issues like data confidentiality among the data owners. And, the other such issue is who to retrieve their data from clouded data base or data warehouse. There have been extensive studies to eliminate such problems through different approaches. Here the current study have drawn from those studies and ideas to device an effective system like them. Here the study has been focused on the RASP data perturbation method and its areas of improvement. The kNN-R calculation is intended to work with the Rasp range query algorithm to process the kNN questions. We have precisely investigated the assaults on information and inquiries under an accurately characterized risk model and sensible security suppositions. The Random Space Encryption (RASE) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. The use of RASE to generate index-able auxiliary data that is resilient to prior knowledge enhanced attacks. Range queries are securely transformed to the encrypted data space and then efficiently processed with a two-stage processing algorithm in which studies are made to identify the potential attacks on the encrypted data and queries at three different levels of prior knowledge available to an attacker.

Index terms:

Privacy, Range Query, kNN Query, Query Services in the Cloud, Multidimensional Range Query, RandomSpace Encryption, Attack Analysis, Outsourced Databases.

1 INTRODUCTION:

Hosting data-intensive query services in the cloud isincreasingly popular because of the unique advantagesin scalability and cost-saving.

With the cloudinfrastructures, the service owners can convenientlyscale up or down the service and only pay for thehours of using the servers. This is an attractive featurebecause the workloads of query services are highlydynamic, and it will be expensive and inefficient toserve such dynamic workloads with in-house infrastructures. However, because the service providerslose the control over the data in the cloud, dataconfidentiality and query privacy have become themajor concerns. 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.

This system and its blend give privacy of information and this methodology is mostly used to secure the multidimensional scope of inquiries in secure way, with indexing and proficient inquiry handling. The reach inquiry is utilized as a part of database for recovering the put away data's. It will recover the records from the database where it can mean some worth in the middle of upper and lower limit. The kNN inquiry means k-Closest Neighbor question. K means positive whole number and this question are utilized to discover the estimation of closest neighbor to k. The Scratch bother inserts the multidimensional information into a mystery higher dimensional space, improved with irregular commotion expansion to ensure the privacy of information.

We build up the protected half-space question change strategy that throws any encased range in the first space to an unpredictably moulded range in the irritated space. There- fore, we have the capacity to utilize a two-stage reach inquiry handling technique: a current multidimensional file, for example, R*- Tree in the irritated space is utilized to figure out the records in the jumping box of the sporadically formed extent, which is then sifted with the changed question condition. This preparing system is quick and secure under the security presumption.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Providers, can possibly make a copy of the databaseor eavesdrop users' queries, which will be difficult to-detect and prevent in the cloud infrastructures. Range query is themost frequently used query in online data analytics(OLAP) that requires the service provider to quickly respondto concurrent user queries. To efficiently process range queries, indexingis a necessary step. However, most existing encryption approachesrequire linear scan over the entire database,thus, impractical for OLAP. Fully homomorphic encryption in theory allows any operation on encrypted data that can be tracedback to an equivalent operation on the corresponding plaintexts. However, this is still too expensiveto be practical even for a simple application like encrypted keywordsearch.

2 RASP: RANDOM SPACE PERTURBATION:

In this section, we present the basic definition ofRAndom Space Perturbation (RASP) method and itsproperties. We will also discuss the attacks on RASPperturbed data, based on the threat model given inSection.

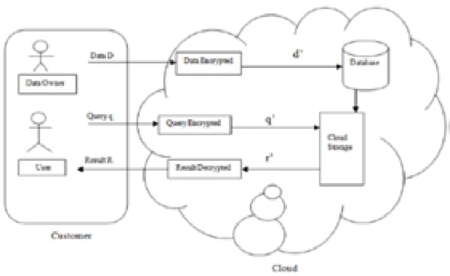

Fig 1:System Architecture for RASP method.

Definition of RASP:

RASP is one type of multiplicative perturbation, witha novel combination of OPE, dimension expansion,random noise injection, and random projection. Let'sconsider the multidimensional data are numeric andin multidimensional vector space1. The RASP perturbation involves three steps. Itssecurity is based on the existence of random invertiblereal-value matrix generator and random real valuegenerator. For each k-dimensional input vector x,

1) An order preserving encryption (OPE) scheme, Eope with keys Kope, is applied to eachdimension of x: Eope(x,Kope) Rd to changethe dimensional distributions to normal distributionswith each dimension's value order stillpreserved.

2) The vector is then extended to d+2 dimensionsas 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 realnumber generator RNG that generates randomvalues from a tailored normal distributions.

3) The (d + 2)-dimensional vector is finally transformedtoF(x,K = {A,Kope,RG}) = A((Eope(x))T , 1, v)T ,(1)1.

Design of OPE and RNG:

the OPE scheme is used to convert all dimensions of the original data to thestandard normal distribution N(0, 1) in the limiteddomain [−β, β]. β can be selected as a value >= 4,as the range [−4, 4] covers more than 99% of thepopulation. This can be done with an algorithm such as the one described in. The use of OPE allowsqueries to be correctly transformed and processed.Similarly, we draw random noises v from N(0, 1) inthe limited domain [−β, β]. Such a design makes theextended noise dimension indifferent from the datadimensions in terms of the distributions.The design of such an extended data vector(Eope(x)T , 1, v) T is to enhance the data and queryconfidentiality. The use of OPE is to transform largescaleor infinite domains to normal distributions,which address the distributional attack. The (d+1)-thhomogeneous dimension is for hiding the query content.The (d+2)-th dimension injects random noise inthe perturbed data and also protects the transformedqueries from attacks. The rationale behind differentaspects will be discussed clearly in later sections.

Properties of RASP:

RASP has several important features. First, RASP doesnot preserve the order of dimensional values becauseof the matrix multiplication component, whichdistinguishes itself from order preserving encryption(OPE) schemes, and thus does not suffer from thedistribution-based attack (details in Section 7). AnOPE scheme maps a set of single-dimensional valuesto another, while keeping the value order unchanged.

Since the RASP perturbation can be treated as acombined transformation F(G(Eope(x))), it is sufficientto show that F(y) = Ay does not preserve theorder of dimensional values, where y Rd+2 andA R(d+2)×(d+2). The proof is straightforward as shown. Second, RASP does not preserve the distances betweenrecords, which prevents the perturbed data2.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Currently, we use a random invertible matrix generator thatdraws matrix elements uniformly at random from the standardnormal distribution and check the matrix invertibility and the nonzeroconditions from distance-based attacks. Because none of thetransformations in the RASP: Eope, G, and F preservesdistances, apparently the RASP perturbation will notpreserve distances. Similarly, RASP does not preserveother more sophisticated structures such as covariancematrix and principal components. Therefore, the PCA-based attacks do not work aswell.Third, the original range queries can be transformedto the RASP perturbed data space, which is the basisof our query processing strategy. A range querydescribes a hyper-cubic area (with possibly openbounds) in the multidimensional space. We will show that a hyper-cubic area in the originalspace is transformed to a polyhedron with the RASP-perturbation. Thus, we can search the points in thepolyhedron to get the query results.

3 RASP RANGE-QUERY PROCESSING:

Based on the RASP perturbation method, we designthe services for two types of queries: range query andkNN query. This section will dedicate to range queryprocessing. We will first show that a range query inthe original space can be transformed to a polyhedronquery in the perturbed space, and then we develop asecure way to do the query transformation. Then, wewill develop a two-stage query processing strategy forefficient range query processing.

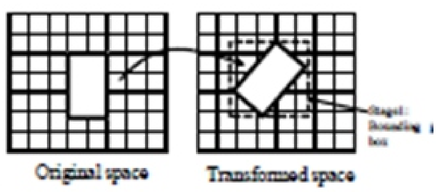

Fig 2: Illustration of the two-stage processing algorithm.

Cost of RASP Perturbation In this experiment, we study the costs of the components in the 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 OPE algorithm partitions the target distribution into Buckets. Then, 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). The following figure shows the cost distributions for 20K records at different number of dimensions.

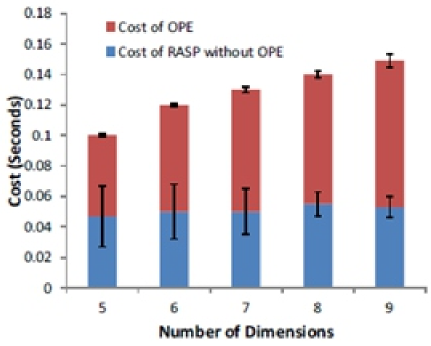

Fig 3:The cost distribution of the full RASP scheme. Data: Adult (20K records,5-9 dimensions).

3. RANDOM SPACE ENCRYPTION:

In this section, we propose the basic Random Space Encryption (RASE) approach for secure range query processing on the encrypted outsourced data. First, we give the system framework and assumptions held for the attack models. Second, we present the definition of the basic random space encryption method and distinguish it from order preserving encryption methods. Finally, we describe how to generate outsourced data and answer queries with the encrypted data.

3.1 System Framework:

System Framework. We assume the outsourced data are multidimensional data and thus the data records can be treated as vectors (or points) in the multidimensional space. The following Figure shows the framework for processing range query services on outsourced data.

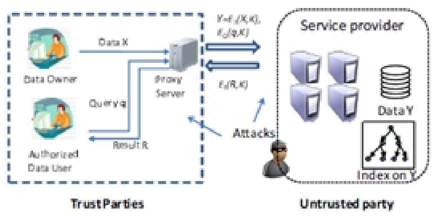

Fig 4:A framework for hosting range query services.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

In the client side, the data owner has all rights to upload/query data, and may also grant the query right to the trusted users. The proxy server receives original data and queries, encrypts and submits them, and decrypts the query results. It keeps the security key, the encryption functions ET (),EQ(), the decryption function D(), and controls the access rights. The traffic between the proxy server and the service provider contains only the encrypted data and queries. Although the proxy server does not handle the large dataset and process queries, it might still become a bottleneck for a large number of users and frequent query submissions. However, the cost to scale the proxy server should be much lower than that to host the entire query processing service.This framework includes several key components.

(1) Encrypted auxiliary data generation. This approach will generate the auxiliary data encrypted with the proposed scheme for indexing purpose through the encryption function ET () in Figure above. It applies a type of multiplicative perturbation on the searchable attributes in the original database to generate the auxiliary data. The goal is to keep the topology of original data vectors in the auxiliary data but obscure the original data values so that they cannot be possibly inferredfrom the auxiliary data.

(2) Query Encryption. A submittedquery should also be appropriately transformed so that the servercan use the index on the encrypted auxiliary data to process thequery. This query transformation should be secure, not reveal anyinformation that helps curious service providers breach privacy. Wedenote it as the EQ() function.

(3) Server side indexing and queryprocessing. The service provider is able to build multidimensionalindex on the auxiliary data. However, processing the transformedqueries requires algorithms different from the existing ones. Ourframework also includes the algorithms for query processing.

4 EFFICIENT RANGE QUERY PROCESSING WITH RASP:

We have shown that the RASP encryption is convexity preserving.This result is closely related to how a query can be transformedand processed. A range query can be represented as a convex setquery. Thus, in the encrypted space there is a unique convex setthat is the answer to the query.

However, there are challenges in efficiently processing it, and making sure query processing doesnot reveal significant information about the encryption key and theoriginal data. One may already notice that the simple query transformationmethod described in this section is vulnerable to attacks. However, in this section, we will focus on the first challenge. Itwill be revisited and significantly improved in security analysis in following Sections.In the encrypted space, a simple dimensional condition in theoriginal space is transformed to a general half space condition (asFigure 2 shows). It would be straightforward to scan each auxiliaryvector with the transformed conditions and return the result. Wewant to explore more efficient index-based processing methods inthis section.

The normal processing strategies are based on multidimensionalindex trees, such as R-Tree [28], that handles axisalignedminimum bounding boxes (MBR). If we still depend onmultidimensional tree indexing to process the transformed queries,the processing algorithm should be slightly modified to handle arbitraryconvex areas, the boundaries of which are not necessarilyaxis-aligned. We will start with the method of query transformation,briefly discuss the normal range query processing algorithmsusing multidimensional indices, and then present the proposed solutionfor processing the transformed queries.

5. ATTACK ANALYSIS:

We categorize the possible attacks into two types: (1) Attackson auxiliary vectors; (2) Attacks based on range queries. Therehas been some related work on attack analysis methods for similarencryption methods, e.g., geometric data perturbation for datamining, which can be migrated to analyze the first type of attacks.However, attacks on range queries are entirely new for ourapproach.

5.1 Attacks on Auxiliary Vectors:

According to the three levels of knowledge the attacker mayhave, we categorize the attacks into three classes: (1) Naive estimation;(2) Distributional Attacks; and (3) Known Input/OutputAttacks. Due to the random component in the RASP encryption,some attacks are actually estimation attacks, i.e., the goal of theattack is to estimate the original values. If the estimation result issufficiently accurate, we say the encryption is broken.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Attack Description and Analysis:

Naive Estimation. With the level 1 knowledge, the attacker observesonly the encrypted data. The only attack is to blindly guessthe matrix A. It has been discussed to find a matrix A to maximizethe difference between the encrypted data and the original data.However, since there is no way to verify how accurate a randomguess is, this type of attack is ineffective, in general.Distributional Attack. With the level 2 knowledge, the attackeralso knows column domains and distributions. This knowledge canbe possibly used to perform more effective attacks. In particular,when the original data have independent columns and no more thanone column having Gaussian distribution, an attack called IndependentComponent Analysis (ICA) can be applied to effectivelyrecover the original data from the perturbed data.Known Input/Output Attack. With the level 3 knowledge, the attackerknows a number of input/output (plaintext/ciphertext) recordpairs.

Countering Attacks on Auxiliary Data

Countering ICA-based Distributional Attack. Since the enumerationbased attack is computationally intractable, we focus on theICA-based attack. We propose two approaches to increase the resilienceto the attack. The first approach is to simulate the ICAattack in sufficient rounds to find a statistically resilient A matrix as the previous work does. However, a more attack-resilientapproach is using the composition encryption scheme (CES) thatconsists of two steps: transforming the original data with an orderpreserving encryption scheme Eo first; then followed by the basicRASP encryption, which can be represented.

6 CONCLUSION:

The RASP perturbation technique was proposed to conduct half spacequeries securely and efficiently on the data hosted in thecloud. The efficient range query processing algorithm has been proposedand evaluated in Chen et al., but its security is not fullyunderstood yet. In this paper we carefully analyze the security ofRASP perturbed data and queries under the three-level adversarialassumptions. The initial analysis shows that the RASP perturbationdoes not satisfy the strong indistinguishability definition on Level2 and 3 assumptions. We notice that the strong indistinguishabilitydefinition might not be necessary for the cloud computing settingand the perturbation techniques in general, where estimation-basedattacks are the typical threats.

Thus, we introduce a weakened definitionon security. This definition is based on statistical learningtheory and information theory, taking the Level 2 and 3 of adversarialknowledge into account. We then analyze a typical estimationattack based on the Level 3 assumption, the regression attack, underthe new security definition. We will continue our study on the securityof RASP perturbed data and queries, and explore more applicationsof the RASP perturbation for secure data intensive computingin the cloud.

REFERENCES:

[1] AGRAWAL, D., AND AGGARWAL, C. C. On the design andquantification of privacy preserving data mining algorithms.In Proceedings of ACM Conference on Principles ofDatabase Systems (PODS) (Madison, Wisconsin, 2002),ACM.

[2] AGRAWAL, R., AILAMAKI, A., BERNSTEIN, P. A.,BREWER, E. A., CAREY, M. J., CHAUDHURI, S., DOAN,A., FLORESCU, D., FRANKLIN, M. J., GARCIA-MOLINA,H., GEHRKE, J., GRUENWALD, L., HAAS, L. M.,HALEVY, A. Y., HELLERSTEIN, J. M., IOANNIDIS, Y. E.,KORTH, H. F., KOSSMANN, D., MADDEN, S.,MAGOULAS, R., OOI, B. C., O'REILLY, T.,RAMAKRISHNAN, R., SARAWAGI, S., STONEBRAKER, M.,SZALAY, A. S., AND WEIKUM, G. The claremont report ondatabase research. SIGMOD Record 37, 3 (2008), 9–19.

[3] AGRAWAL, R., KIERNAN, J., SRIKANT, R., AND XU, Y.Order preserving encryption for numeric data. InProceedings of ACM SIGMOD Conference (2004).

[4] ARMBRUST, M., FOX, A., GRIFFITH, R., JOSEPH, A. D.,KATZ, R., KONWINSKI, A., LEE, G., PATTERSON, D.,RABKIN, A., STOICA, I., AND ZAHARIA, M. Above theclouds: A berkeley view of cloud computing. TechnicalReport, University of Berkerley (2009).

[5] BOLDYREVA, A., CHENETTE, N., LEE, Y., AND O'NEILL,A. Order preserving symmetric encryption. In Proceedingsof EUROCRYPT conference (2009).

[6] BOLDYREVA, A., CHENETTE, N., AND OÏ£¡NEILL, A.Order-preserving encryption revisited: Improved securityanalysis and alternative solutions. In CRYPTO (2011).

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[7] BONEH, D., CRESCENZO, G. D., OSTROVSKY, R., ANDPERSIANO, G. Public-key encryption with keyword search.In Proceedings of Advances in Cryptology, (EUROCRYPT)(2004), Springer.

[8] BOYD, S., AND VANDENBERGHE, L. Convex Optimization.Cambridge University Press, 2004.

[9] CHEN, K., KAVULURU, R., AND GUO, S. Rasp: Efficientmultidimensional range query on attack-resilient encrypteddatabases. In ACM Conference on Data and ApplicationSecurity and Privacy (2011).

[10] COVER, T., AND THOMAS, J. Elements of InformationTheory. Wiley, 1991.