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

RASP for Data Perturbation Method to Provide Secure and Efficient Range Query and KNN Query Services for Protected Data in the Cloud

Jadala Mahesh
M.Tech,
Department of CSE,
Global Institute of Engineering and Technology,
Chilkur (V), RR District, Telganana.

Mrs. M.Jhansi Lakshmi
Associate professor,
HOD of CSE,
Global Institute of Engineering and Technology,
Chilkur (V), RR District, Telganana.

Abstract:

With the development of services computing and cloud computing, it has become possible to outsource large databases to database service providers and let the providers maintain the range-query service. 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. We propose the Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. The random space perturbation (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.

Index Terms:

Query services in the cloud, privacy, range query, kNN query.

INTRODUCTION:

Cloud computing is the internet based storage method. It is mainly used for storing the files and applications in it infrastructures. Peoples uses the cloud because of its attractive features like secure service, infinite of storage, it will satisfy the user experience, low cost and multiple user can access the files and applications.

In cloud, the query service process are frequently used because, the user can save their cost. The owners in the cloud will pay the amount only for their using time of server. This is an important feature because, the working time of query service in cloud is very high and it is more expensive. New process are need for the cloud to protect the data and query privacy, so by that new process the query service can be protected. But if the new approaches for providing security will provide sloe query process is not an advantage.

We analyze the CPEL criteria for submit a query in cloud. This CPEL criteria denotes Confidentiality of data, query Privacy, Efficient query processing and Low working cost. This method also used to increase the complexity of query service. We propose the Random space Perturbation (RASP) method to construct the query and here we separate the query as range query and kNN query. The proposed RASP method will use the four concepts of the CPEL criteria and here the multidimensional data can be transformed with the combination of order preserving encryption, random projection and random noise injection.

- The RASP method and its combination provide confidentiality of data and this approach is mainly used to protect the multidimensional range of queries in secure manner, with indexing and efficient query processing.

- The range query is used in database for retrieving the stored data's. it will retrieve the records from the database where it can denotes some value between upper and lower boundary.

- The kNN query denotes k-Nearest Neighbor query. K denotes positive integer and this query are used to find the value of nearest neighbor to k.

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

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.

LITERATURE SURVEY:

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

A)ORDER PRESERVING ENCRYPTION FOR NUMERIC DATA - R. AGRAWAL, J. KIERNAN, R. SRIKANT - 2004

Encryption is a well-established technology for protecting sensitive data. However, once encrypted, data can no longer be easily queried aside from exact matches.

We present an order-preserving encryption scheme for numeric data that allows any comparison operation to be directly applied on encrypted data. Query results produced are sound (no false hits) and complete (no false drops). Our scheme 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. The proposed scheme has been designed to be deployed in application environments in which the intruder can get access to the encrypted database, but does not have prior domain information such as the distribution of values and cannot encrypt or decrypt arbitrary values of his choice. The encryption is robust against estimation of the true value in such environments.

B)ABOVE THE CLOUDS: A BERKELEY VIEW OF CLOUD COMPUTING - M. ARMBRUST, A. FOX, R. GRIFFITH - 2009

Cloud computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service and shaping the way IT hardware is designed and purchased. Developers with innovative ideas for new Internet services no longer require the large capital outlays in hardware to deploy their service or the human expense to operate it. They need not be concerned about overprovisioning for a service whose popularity does not meet their predictions, thus wasting costly resources, or underprovisioning for one that becomes wildly popular, thus missing potential customers and revenue. Moreover, companies with large batch-oriented tasks can get results as quickly as their programs can scale, since using 1,000 servers for one hour costs no more than using one server for 1,000 hours. This elasticity of resources, without paying a premium for large scale, is unprecedented in the history of IT.

C) SECURITY MODELING AND ANALYSIS - J. BAU AND J. C. MITCHELL - 2011

Security modeling centers on identifying system behavior, including any security defenses; the system adversary's power; and the properties that constitute system security. Once a security model is clearly defined, security analysis evaluates whether the adversary, interacting with the system, can defeat the desired security properties.

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

Although the authors illustrate security analysis using model checking, analysts can use various methods and tools to evaluate system security, including manual and automated theorem-proving tools that provide assurance about the absence of attacks in a specified threat model. This article describes a uniform approach for evaluating system security and illustrates the approach by summarizing three case studies. Security modeling and analysis also provides a basis for comparative evaluation and some forms of security metrics.

D) PRIVACY PRESERVING MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA - N. CAO, C. WANG, M. LI - 2011

With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given.

Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

E)RASP: EFFICIENT MULTIDIMENSIONAL RANGE QUERY ON ATTACK-RESILIENT ENCRYPTED DATABASES - K. CHEN, R. KAVULURU - 2011

Range query is one of the most frequently used queries for online data analytics. Providing such a query service could be expensive for the data owner. With the development of services computing and cloud computing, it has become possible to outsource large databases to database service providers and let the providers maintain the range-query service. With outsourced services, the data owner can greatly reduce the cost in maintaining computing infrastructure and data-rich applications. However, the service provider, although honestly processing queries, may be curious about the hosted data and received queries. Most existing encryption based approaches require linear scan over the entire database, which is inappropriate for online data analytics on large databases. While a few encryption solutions are more focused on efficiency side, they are vulnerable to attackers equipped with certain prior knowledge. We propose the Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. We use RASP to generate indexable 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. We thoroughly studied the potential attacks on the encrypted data and queries at three different levels of prior knowledge available to an attacker. Experimental results on synthetic and real datasets show that this encryption approach allows efficient processing of range queries with high resilience to attacks.

F)GEOMETRIC DATA PERTURBATION FOR OUTSOURCED DATA MINING - K. CHEN AND L. LIU - 2011

Data perturbation is a popular technique in privacy-preserving data mining. A major challenge in data perturbation is to balance privacy protection and data utility, which are normally considered as a pair of conflicting factors.

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

We argue that selectively preserving the task/model specific information in perturbation will help achieve better privacy guarantee and better data utility. One type of such information is the multidimensional geometric information, which is implicitly utilized by many data-mining models. To preserve this information in data perturbation, we propose the Geometric Data Perturbation (GDP) method. In this paper, we describe several aspects of the GDP method. First, we show that several types of well-known data-mining models will deliver a comparable level of model quality over the geometrically perturbed data set as over the original data set.

Second, we discuss the intuition behind the GDP method and compare it with other multidimensional perturbation methods such as random projection perturbation. Third, we propose a multi-column privacy evaluation framework for evaluating the effectiveness of geometric data perturbation with respect to different level of attacks. Finally, we use this evaluation framework to study a few attacks to geometrically perturbed data sets. Our experimental study also shows that geometric data perturbation can not only provide satisfactory privacy guarantee but also preserve modeling accuracy well.

EXISTING SYSTEM:

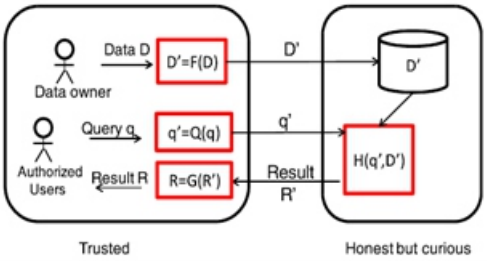
» Requirements for constructing a practical query service in the cloud as the CPEL criteria: data confidentiality, query privacy, efficient query processing, and low in-house processing cost. Satisfying these requirements will dramatically increase the complexity of constructing query services in the cloud. Some related approaches have been developed to address some aspects of the problem.
» The crypto index and order preserving encryption (OPE) are vulnerable to the attacks. The enhanced crypto index approach puts heavy burden on the in-house infrastructure to improve the security and privacy.

DISADVANTAGES:

» Do not satisfactorily addressing all aspects of Cloud.
» Increase the complexity of constructing query services in the cloud.
» Provide slow query services as a result of security and privacy assurance.

PROPOSED SYSTEM:

» We propose the random space perturbation (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.



ADVANTAGES:

» The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee.
» The RASP approach preserves the topology of multi-dimensional range in secure transformation, which allows indexing and efficiently query 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.

IMPLEMENTATION:

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

User Module :

In this module, Users are having authentication and security to access the detail which is presented in the ontology system.

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

Before accessing or searching the details user should have the account in that otherwise they should register first.

Multidimensional Index Tree :

Most multidimensional indexing algorithms are derived from R-tree like algorithms , where the axis-aligned minimum 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. 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.

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)-Range 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 for privacy-preserving kNN search.

Preserving Query Privacy :

Private information retrieval (PIR) tries to fully preserve the privacy of access pattern, while the data may not be encrypted. PIR schemes are normally 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.

CONCLUSIONS:

We proposed RASP method with range query and kNN query. This method mainly used to perturb the data given by the owner and saved in cloud storage it also combines random injection, order preserving encryption and random noise projection and also it has contains CPEL criteria in it. By using the range query and kNN query user can retrieve their data's in secured manner and the processing time of the query is minimized. And also we continue our studies to improve the effect of query.

FURTHER WORK:

We also conduct several sets of experiments to show the efficiency of query processing and the low cost of in-house processing. We will continue our studies on two aspects: 1) further improve the performance of query processing for both range queries and kNN queries; and 2) formally analyze the leaked query and access patterns and the possible effect on both data and query confidentiality.

REFERENCES:

[1] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order Preserving Encryption for Numeric Data," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2004.

[2] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.K. Andy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing," technical report, Univ. 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, May/June 2011.

[4] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge Univ. Press, 2004.

[5] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOMM, 2011.

[6] K. Chen, R. Kavuluru, and S. Guo, "RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases," Proc. ACM Conf. Data and Application Security and Privacy, pp. 249-260, 2011.

[7] K. Chen and L. Liu, "Geometric Data Perturbation for Outsourced Data Mining," Knowledge and Information Systems, vol. 29, pp. 657- 695, 2011.

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

[8] K. Chen, L. Liu, and G. Sun, "Towards Attack-Resilient Geometric Data Perturbation," Proc. SIAM Int'l Conf. Data Mining, 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," Proc. 13th ACM Conf. Computer and Comm. Security, pp. 79-88, 2006.

[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," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 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," Proc. Very Large Databases Conf. (VLDB), 2004.

[15] Z. Huang, W. Du, and B. Chen, "Deriving Private Information from Randomized Data," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2005.