

BUILDING CONFIDENTIAL AND EFFICIENT QUERY SERVICES IN THE CLOUD WITH RASP DATA PERTURBATION

Geeta Guttikonda¹, Mediseti Jnana Surya Prakasha Rao²

¹Asst. Professor, Department of IT, VR Siddhartha Engineering College

²M.Tech, Department of IT, VR Siddhartha Engineering College

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 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. 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.

Keywords: *cloud computing, kNN query, encryption.*

INTRODUCTION

HOSTING data-intensive query services in the cloud is increasingly popular because of the unique advantages in scalability and cost-saving. With the cloud infrastructures, the service owners can conveniently scale up or down the service and only pay for the hours of using the servers. This is an attractive feature because the workloads of query services are highly dynamic, and it will be expensive and inefficient to serve such dynamic workloads with in-house infrastructures. However, because the service providers lose the control over the data in the cloud, data confidentiality and query privacy have become the major 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. 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. We summarize these 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. However, they do not satisfactorily address all of these aspects. For example, the cryptindex and order preserving encryption (OPE) are vulnerable to the attacks. The enhanced cryptindex approach puts heavy burden on the in-house infrastructure to improve the security and privacy. The New Casper approach uses cloaking boxes to protect data objects and queries, which affects the efficiency of query processing and the inhouse workload. We have summarized the weaknesses of the existing approaches.

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 data sets 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 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 RASP approach preserves the topology of multidimensional 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.

We have carefully evaluated our approach with synthetic and real data sets. The results show its unique advantages on all aspects of the CPEL criteria. This paper is organized as follows: In Section 3, we define the RASP perturbation method, describe its major properties, and analyze the attacks to the RASP perturbed data. We also introduce the framework for constructing the query services with the RASP perturbation. In Section 4, we describe the algorithm for transforming queries and processing range queries. In Section 5, the range query service is extended to handle kNN queries. When describing these two services, we also analyze the attacks on the query privacy. Finally, we present some related approaches in Section 7 and analyze their weaknesses in terms of the CPEL criteria.

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 OF EXISTING SYSTEM:

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 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 OF PROPOSED SYSTEM:

- 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 workload because of the low perturbation cost and high precision query results. This is an important feature enabling practical cloud-based solutions.

IMPLEMENTATION

Modules:

Admin:

Admin can get logged in by entering a valid username and password. Admin can view all the user details and can activate or deactivate users. Admin can view all the details of users file logs. Admin can view users file download details. Admin can view attacker’s details. Admin can view users file update details.

User:

Users need to get registered with the application by entering all the details in the registration form. User can enter into the application by giving a unique username and password. User will get a secret key to the registered mail id. User can search for the files. User can upload file and can view all the file details.

SCREEN SHORTS

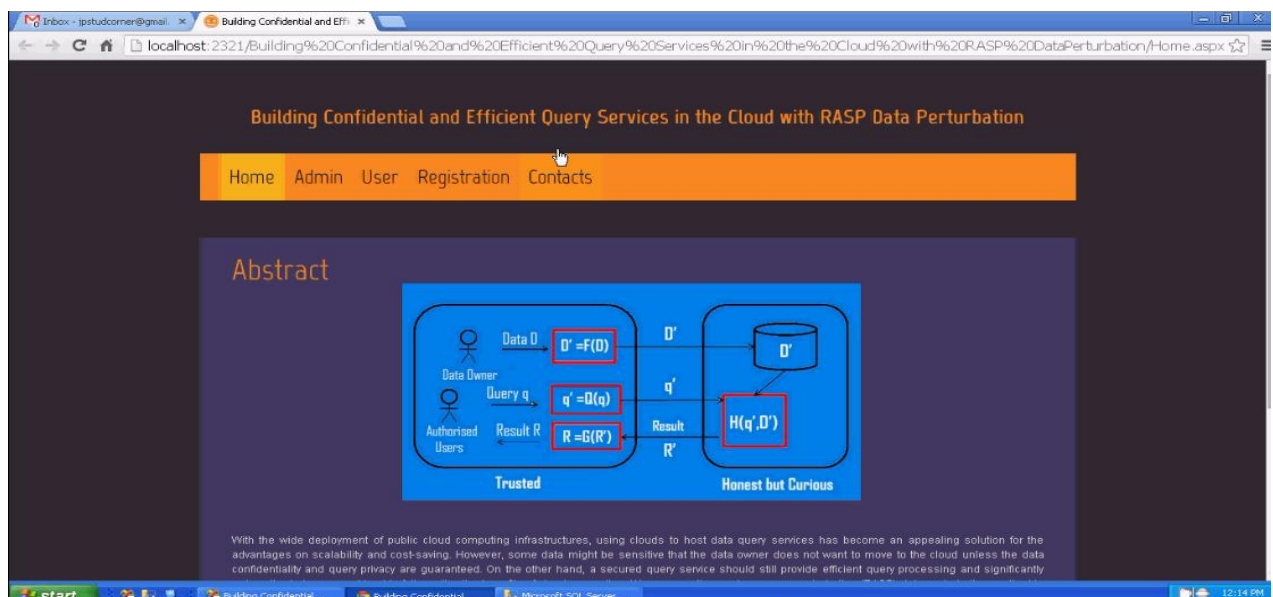


Fig : Home Page

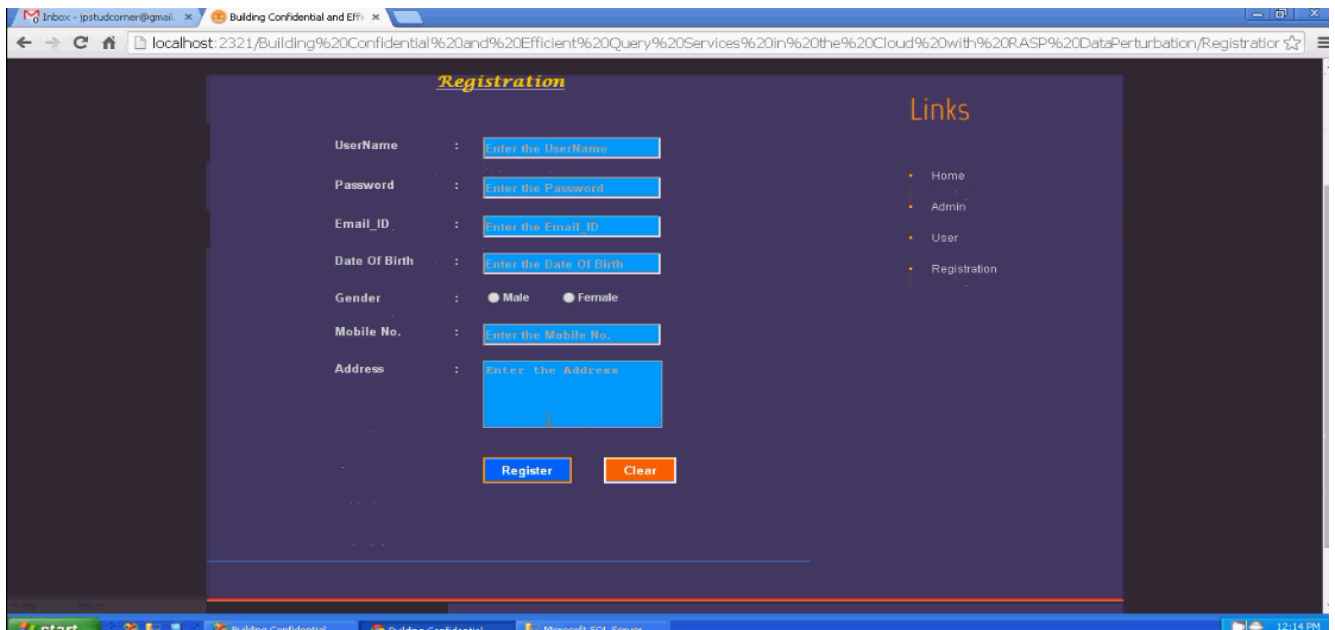


Fig: Registration Page

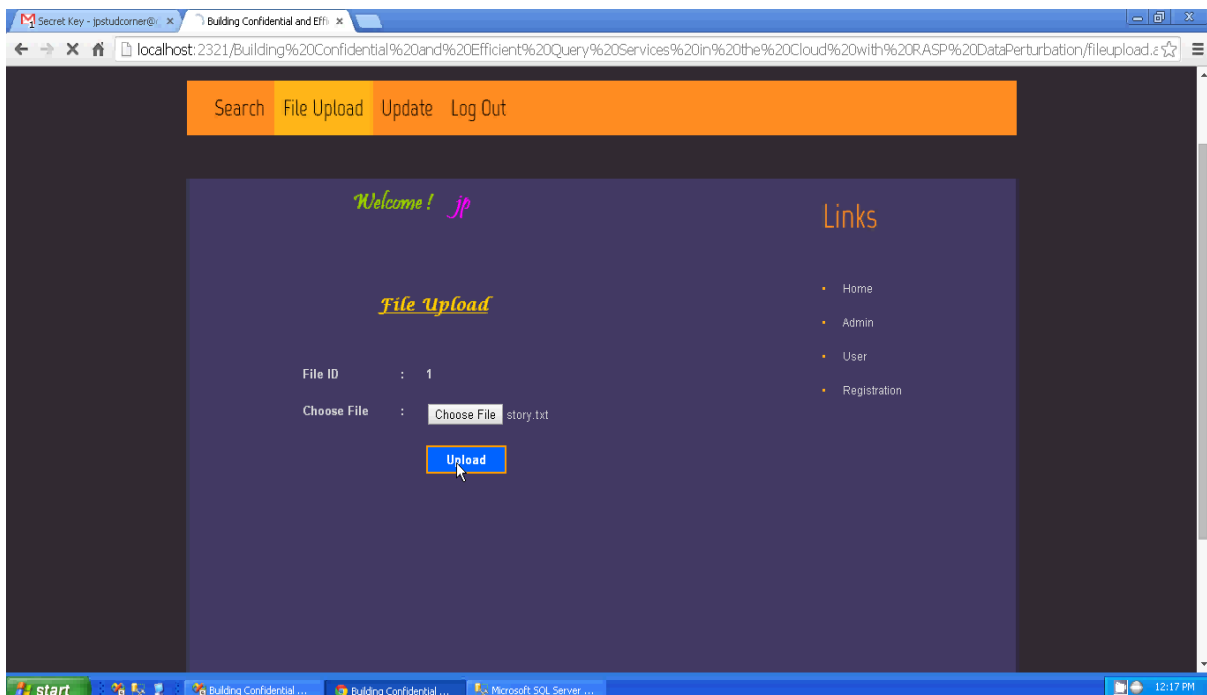


Fig: File Upload

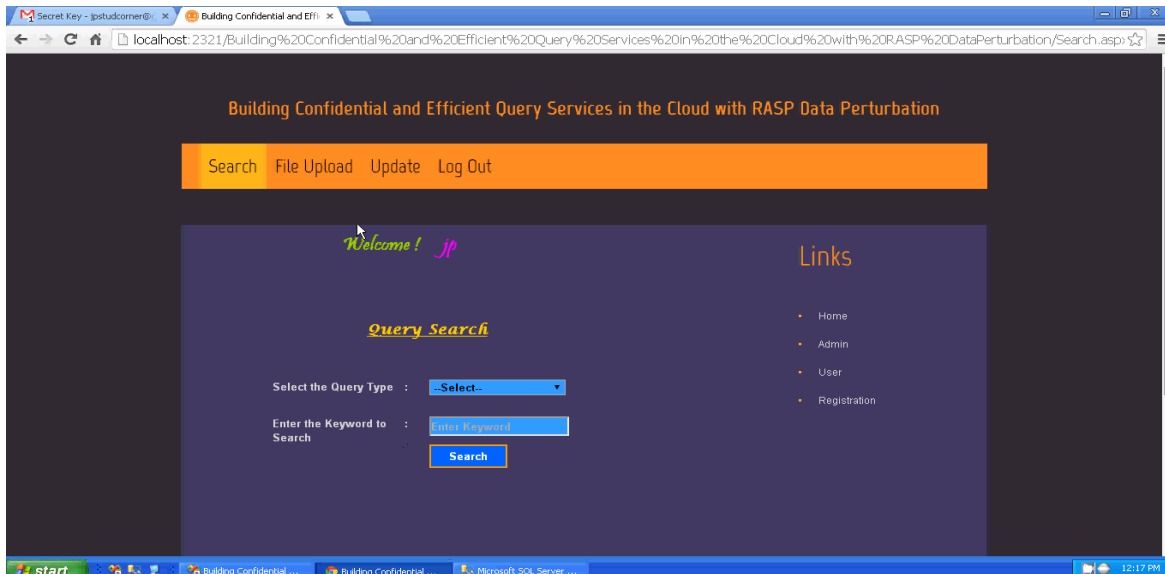


Fig : Query Search

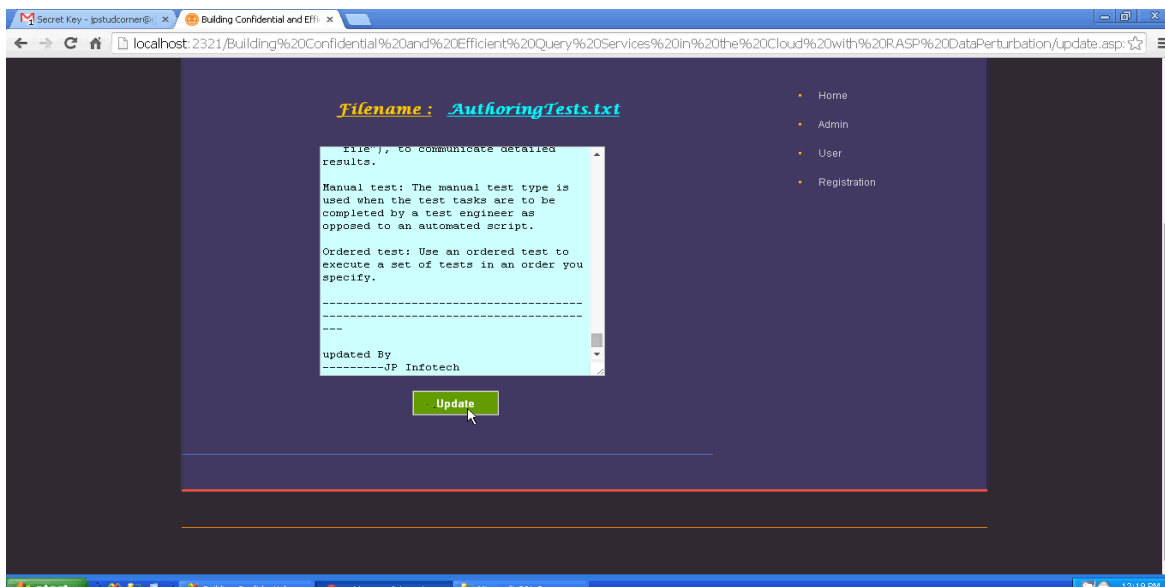


Fig : File Update

CONCLUSION

We propose the RASP perturbation approach to 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, dimensionality expansion, random noise injection, and random projection, which provides unique security features. It 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 sublinear 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. 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 Berkeley, 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.
- [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.