International Journal of Combined Research & Development (IJCRD)
eISSN: 2321-225X;pISSN:2321-2241 Volume: 4; Issue: 6; June -2015

CONSTRUCTING SECURED PRIVATE DATA and SYSTEMATIC QUERY SERVICES in CLOUD

Mr. Nishanth S Airani
M.Tech Scholar
Dept of CSE
VTU RC, P.G studies Mysore
Karnataka-570029,India

Mrs. Pushpalatha R
Assistant Professor
Dept of CSE
VTU RC, P.G studies Mysore
Karnataka-570029,India

Abstract—Cloud computing infrastructure being used widely for the purpose of scalability, cost-saving and sharing by various domains, it has become a tough task for a data owner to place his data inside the cloud because of the inevitable constraints like privacy for both data as well as query, efficiency. Along with the security provided by the cloud service provider (CSP) and by constructing private, secured data and systematic queries the above mentioned constraint can be resolved. Here we also solve the problem of data integrity. We propose a technique called Make-it-simple where we use range query services to construct the systematic query and data confidentiality is done by differentiating the data into data chunks and storing them on data servers in the cloud. Data processing takes place at two stages.

Keywords - Data owner, user, make-it-simple.

I. INTRODUCTION

Introducing the data-comprehensive query services in a cloud is getting popular day-by-day because of the advantages of cloud such as scalability, sharing and cost-saving. Scalability is a beckoning feature as the workloads of query services are influential and it becomes costly and non-systematic to serve such dynamic workloads. As the time dies the service providers may lose their control over the data and thus privacy, efficiency, confidentiality become major concerned areas.

Unique and fresh approaches are required to protect the data privacy, confidentiality, security and efficiency of query services and also there is a need to protect the benefits of cloud. The vital relationship among all the above mentioned constraints exists because the data owner cannot use the resources present inside the cloud significantly as the cloud is meant to reduce the need of maintaining scalable in-house infrastructures.

We summarize these above constraints as our requirements for the construction of query services and to build an architecture t protect the data inside the cloud. Many related approaches have been developed to address some part of the problem. However they do not satisfactorily address all of these aspects. For example, crypto-index [1] and order preserving encryption (ope) [2] are vulnerable to attacks. RASP data perturbation also gives an idea for the architecture.

II. QUERY SERVICES

Range query is one of the vital queries which help us in finding the number of records in the range defined by conditions. A range query looks like this

Select count (*) from T

Where X_i€ [a_i, bi] and X_j€ [a_j, bj] and X_k=AK,

Here T is a table and Xi, Xj and Xk are the real valued attributes in T and a and b are some constants. Range queries may be applied to arbitrary number of attributes and conditions on these attributes combined with conditional operators "and"/"or".

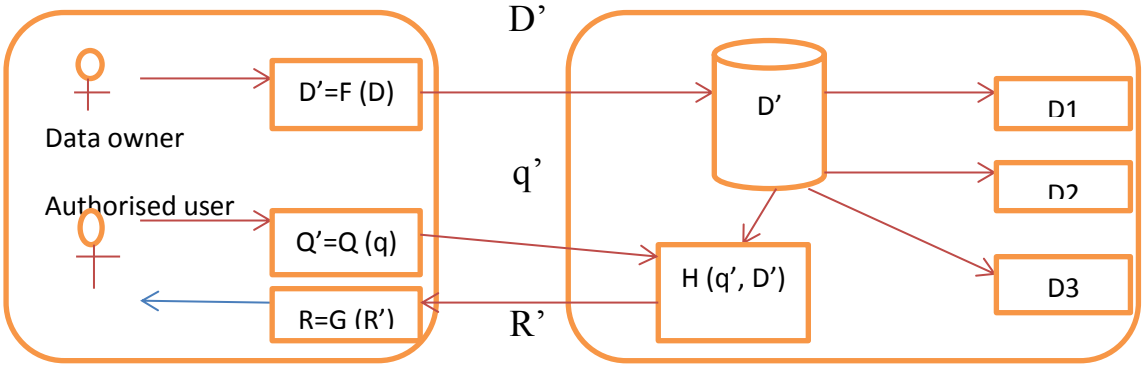
KNN query is to find the closest k records to the query point, where the Euclidean distance is often used to measure the proximity. It is frequently used in location-based services for searching the objects close to a query point, and also in machine learning algorithms such as hierarchical clustering and kNNclassifier. A kNN query consists of the query point and the number of nearest neighbours, k.

III. THREAT MODEL

The goal of the attacker is to recover the original data or identify the exact queries and through that identify the data present inside a database. So this brings in heavy threat to the existing architecture. So with the data being transformed we are not on the safer side, we need to authorise the users and think or assume that authorise users don't leak the information.

When adversary hacks the query and through that the original data, the global information associated with this data base may also get leaked. Example is application of the database. This may cause adverse effect on the other related applications and the systems internal information also.

IV. ARCHITECTURE



Here in the architecture we introduce a technique called Make-it-simple. Here the data owner will transform his data into some other functional form of data and

only he knows that. With this he can store his data in a cloud and make it confidential. The users are authorised using keys. Here the authorised users are less in number.

The data D becomes D' on transforming and later it is divided into data chunks and stored on different servers in cloud. So that even if an adversary hacks the data base he may find it difficult to know what's the original data is. As it needs to be integrated. On the other hand we have the queries transformed into different forms using range queries and using kNN query we can find the nearest neighbour query so that efficiency is reached in query processing. The query processing takes place at two stages. The first stage is where the written query is processed and the different data chunks are integrated to get the original data or file. With this approach the data becomes secure because first we encrypt the data using encryption algorithm and later we encode the data which is obtained as the input from encryption phase. To crack a encrypted-encoded data, it takes thousands of years. We use base64 technique for encoding and AES technique for encrypting. Second step of the processing is done and again the result is transformed into some other format. Thus this two stage processing takes place inside the cloud. At the user end the result is decrypted and original result is got.

This method not only enhances the efficiency of querying services but also protects the data inside the cloud. Differentiating and integrating the data chunks of the original data is effective in terms of privacy and confidentiality.

V. RELATED WORK

Private information retrieval (PIR) [3] 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. [4] 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. [5] 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. Papadopoulos et al. [6] uses private information retrieval methods [7] to enhance location privacy. However, their approach does not consider protecting the confidentiality of data. Space Twist [8] 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 [9] considers both data confidentiality and query privacy.

VI. EXPERIMENTAL RESULTS

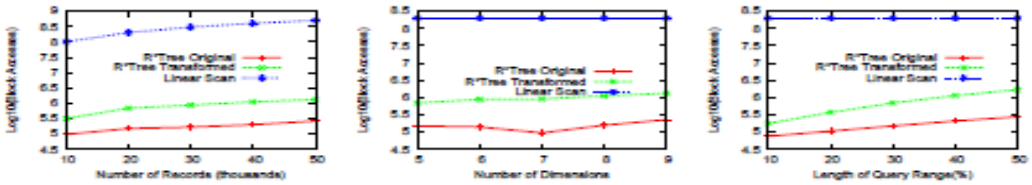


Fig. 8. Performance comparison on Uniform data. Left: data size vs. cost of query; Middle: data dimensionality vs. cost of query; Right: query range (percentage of the domain) vs. cost of query

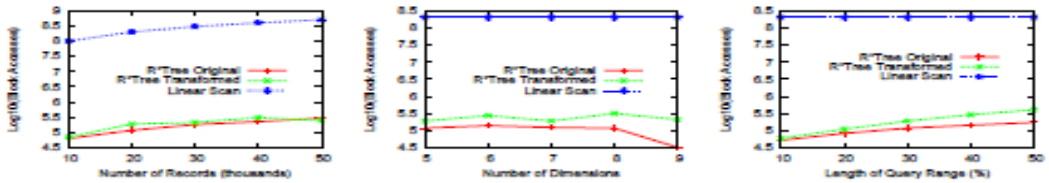


Fig. 9. Performance comparison on Adult data. Left: data size vs. cost of query; Middle: data dimensionality vs. cost of query; Right: query range (percentage of the domain) vs. cost of query

The first pair of figures (the left subfigures of Figure 8 and 9) 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 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.

VII. CONCLUSION AND FUTURE WORK

We propose the Make-it-simple approach to provide the data confidentiality, query privacy, query efficiency. This manner helps us to develop the systematic way of querying and processing the same. It might also reduce the in house cost of the cloud like instead of providing the authentication for every user using password and username, this method reduces that burden on cloud by providing the authentication from the data owner itself i.e. authorised users. Results are also transformed. Data chunks are not stored in one place.

As we know that a work never becomes saturated at one point, there will always be a scope to enhance the work and achieve the greater heights and still better results. Here the future enhancement would be developing a unique technique to repair the lost data chunks, i.e., if data chunks are lost we should develop a new method to repair the original data. Also we can provide cryptography methods to protect the data chunks inside the cloud.

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

[4] 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.

[5] 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.

[6] S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest neighbour search with strong location privacy," in Proceedings of Very Large Databases Conference (VLDB), 2010.

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

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

[9] 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.

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

[11] 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.

[12] 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 Com- munications Security, 2008.

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

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