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

Reena D K[1]

[1] *Final year M.Tech, Department of Computer Science, STJIT Ranebennur, Karnataka, India*

Abstract

With the advent of Cloud computing technology, the cost of computation, application hosting, content storage and delivery is reduced significantly. So now-a-days at an increased speed cloud computing infrastructures are used by the people. Security and Privacy is the biggest concern about cloud computing, since data management and infrastructure management in cloud is provided by third-party, it is always a risk to handover the sensitive information to such providers until and unless data confidentiality and secure query processing is guaranteed by the cloud service provider. To fully realize the benefits of cloud computing the workload must be reduced and efficient query processing must be provided. Therefore, to provide confidential and efficient query service RASP method is proposed, where RASP denotes Random Space Perturbation. Data Perturbation technique allows users to ascertain key summary information about the data that is not distorted and does not lead to a security breach. Exclusive security features are provided by the RASP. The RASP approach satisfies the data Confidentiality, query Privacy, Efficient query processing and Low working cost (CPEL) criteria for hosting queries in the cloud. KNN-R algorithm is used here to process the Range query to the kNN query. Users have been authorized by using the randomly generated product key value provided by the admin after successful registration by the user thus maintaining confidentiality. User queries are retrieved within a very short period of time i.e., less than a second. Outsourcing the management of the data related to organizations and individuals to a service provider are enabled by cloud computing services in order to save on the hardware costs and reducing maintenance cost. Also analyzed how the RASP method will provide confidentiality of data and will increase the working process of query.

Key Words: *query services in the cloud, low in-house processing, RASP perturbation, Range query, KNN query.*

----------***----------

1. INTRODUCTION

Cloud Computing refers to manipulating, configuring, and accessing the applications online. It provides dynamically scalable infrastructure for application, data and file storage. Cloud Computing offers **on-demand self-service**. The resources can be used without interaction with cloud service provider. Cloud Computing is highly cost effective because it operates at higher efficiencies with greater utilization. It just requires an Internet connection. Cloud Computing offers load balancing that makes it more reliable. The billing model is pay as per usage; the infrastructure is not purchased thus lowering maintenance. This is an important feature as in the cloud the working time of the query service is very high and expensive.

For the protection of data and query privacy there is a need for new methods in the cloud. But, if the new methods provide slow query processing than it will be not advantageous. Analyzed the CPEL criteria for submitting a query in cloud. This CPEL criteria denotes Confidentiality of data, query Privacy, Efficient query processing and Low working cost. This method is also used to increase the complexity of query service.

For constructing query services in the cloud Random Space Perturbation (RASP) method is used. Here query is separated as Range query and KNN query. This proposed method uses all the four CPEL criteria concepts. The transformation of the multidimensional data is done with the combination of order preserving encryption, random projection and random noise injection.

- Data confidentiality is provided by the RASP method and its combination. It is mainly used to protect the multidimensional range of queries in secure manner and with efficient query processing.
- The range query is used in database for retrieving the stored data's. It retrieves the record from the database where range denotes some value between upper and lower boundary.
- The kNN query denotes K-Nearest Neighbor query. K denotes positive integer and this query is used to find the k nearest neighbor values.

2. PROPOSED SYSTEM ARCHITECTURE AND IMPLEMENTATION DETAILS

To store large datasets and query services cloud computing infrastructures are mainly used. Fig.1 System architecture shows two main parts in it. The data owner stores a number of data's in the cloud d=n (d, k) here d represents data, n represents normal form of data, k represents key value given by the data owner. This format will be saved in the cloud as perturbed form d=e (d, k) here e represents perturbation. The two separate parties are: Trusted party and the cloud service provider.

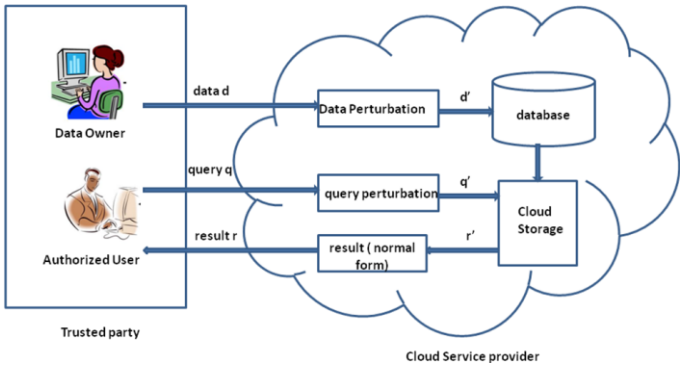

Fig1. System Architecture

2.1 Trusted Party

Trusted Party stores their data in the cloud. The trusted party consists of the data owner and the authorized users. Data is uploaded in the cloud by the authorized users and the data owner and that data will be perturbed and stored in the cloud database. The user's needs to complete the registration procedure first and later can login with valid username and password and the product key value .The product key value will be provided by the admin to the user at their registered email ID. Thus, users are authorized by the unique product key value sent by the admin.

Person who tries to login with invalid product key will be treated as a attacker and his details will be stored in the database in the attackers list. By logging in with valid credentials the user can access various services of the cloud such as search, upload, update logout etc. The query search can be done for either or both the range query and the KNN query. User can only upload the text files. The text file can be viewed or downloaded by the user. After completing the can task he/she can logout.

The admin needs to login as admin. By logging-in with valid credentials the admin can go through the various options made available to him such as users, file logs, downloads, attackers, file updates etc. The user field consists of user details like name, username, password, mail, status, activate and deactivate. Only the admin can activate the user after registration and can deactivate the user based on his/her activities.

2.2 Cloud Service Provider

The cloud service provider stores the data in the cloud database by changing the data to the perturbed form. It is a provider of web hosting services.

Algorithm 1: KNN-R Algorithm

1: The client generates the initial range and sends its Secure perturbed form to the server.
2: The server works on the secure range queries and finds the inner range covering at least k points.
3: The client decodes the secure perturbed inner range from the server and extends it to the outer range, which is sent back to the server.
4: The server returns the points in the outer range.
5: The client decrypts the points and extracts the k nearest points.

2.3 Exploiting User Data

Like hackers cloud service providers want to gain access to their users' data as well, mainly because of achieving profit. In contrast to hackers they have legal possibilities to access this data, mainly due to the terms of service agreement14, a contract which every user has to agree. Many service providers scan user data on tags which are then used to show highly personalized ads. But also more complex data and statistics are recorded, bundled and analyzed (data mining) to be able to do so called user profile marketing, making prediction on what items a user might buy in the near future, what is his next travel destination etc.

The business model of many cloud service providers offering free services is built on highly personalized advertisements. Therefore, they scan the user data on tags and keywords and look for other companies which are offering products matching these tags. Then they charge these companies a small fee for showing the user an advertisement on the web interface. But they are not only scanning for simply keywords, also more sophisticated methods are used to obtain all kinds of statistics from a user's emails, documents etc. This is described under the term data mining, which means extraction, analysis and usage of data in a way that it was not originally stored for. The providers can link several different kinds of user data together to get highly accurate user profiles which can then be used to do behavior prediction. It is obvious that storage encryption is against their business model as long as the provider is not able to decrypt the data.

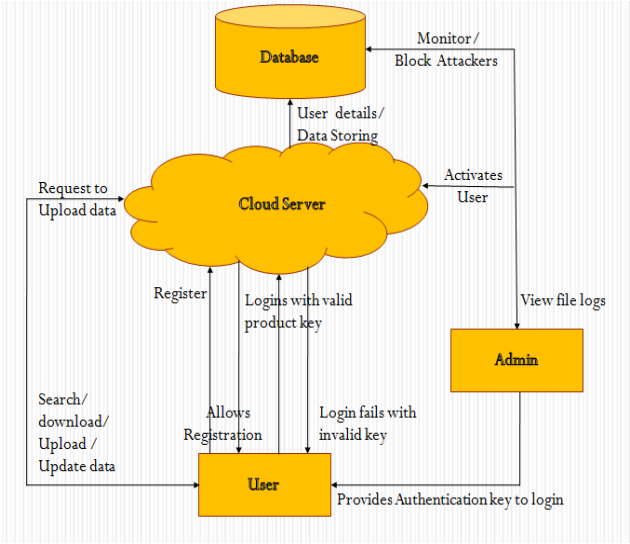

Fig.2 Block diagram

Figure 2 shows the proposed RASP Data Perturbation system which indicates the cloud server interaction with the user and admin respectively.

2.4 Data Leaks through Employees

If the data is stored unencrypted, a displeased employee of the cloud service provider may access the data and disclose it to a third party. As mentioned above storage encryption is an effective countermeasure against this type of data leak, especially if only the user is in possession of the key.

2.5 Attackers

The main interest for attackers in a user's private data is for illegal activities. Therefore, the most interesting data is user confidential details that may be also his/her sensitive data. Hackers may gain a reasonable profit by selling this data. Unlike the service providers attackers have no legal possibilities to access the user's data; instead their activities are also subject to criminal prosecution.

2.6 Implementation Status

Used Java to write and deploy the services. All storage is provided using MySql. All the web services are run on virtual machines powered by Windows 7 and all these machines are hosted within an open source cloud namely DriveHQ first Cloud IT solution Provider.

2.6.1 Drive HQ

DriveHQ File Manager 6.0 - Drag-n-drop manage remote files like local. File Manager makes remote storage as easy as your local storage. With intuitive interface and seamless integration of your local and online storage, it lets you transfer, access, share, sync, collaborate and publish files online with unprecedented easiness and reliability. For businesses, it can replace your file servers for much lower cost and better features. Using DriveHQ's Magic Upload technology, upload speed can be many times faster in certain cases. Encrypted folder support: files are encrypted locally before they are uploaded to DriveHQ; files remain to be encrypted on DriveHQ server. Thus it is extremely secure. It automatically compress data for faster upload and download.

CONCLUSIONS

Proposed RASP perturbation approach to host query services in the cloud with Range and KNN query. This approach is specially used to perturb the data uploaded by the user and save it in the cloud database in a secure form. The approach satisfies the data Confidentiality, query Privacy, Efficient query processing and Low working cost (CPEL) criteria.

Confidentiality of data and low processing cost is a significant feature to fully realize the benefits of cloud computing effectively. Key measure for the quality of query services is secure and efficient query processing.

The proposed approach provides unique security features as it is a unique combination of random noise injection, random projection and dimensionality expansion of order preserving encryption (OPE). The KNN-R algorithm finds the approximate KNN results so that the RASP Range query service can be used. Users have been authorized by using the randomly generated product key value provided by the administrator. Thus, provides data confidentiality. The query either range or KNN query results in not more than a second leading to efficient query processing. Hence, processing time of query is minimized to a larger extent.

The project is a feasibility study that aimed to explore the feasibility and potential for utilizing Cloud capability to address data storage and processing needs. And can also continue studies to provide still better perturbation approach and can improve the effect of the query.

ACKNOWLEDGMENT

The author thanks **Prof. S G Makanur**, Head of the department of Information Science and Engineering for providing sufficient knowledge to do project on this research topic. The author also thanks **Mrs. Poornima D V,** PG- Coordinator of Computer Science and Engineering for her constant support and guidance.

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] J. Bau and J.C. Mitchell, "Security Modeling and Analysis," IEEE Security and Privacy, vol. 9, no. 3, pp. 18-25, May/June 2011.

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

[4] 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.

[5] 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.

[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer-Verlag, 2001.

[7] B. Hore, S. Mehrotra, and G. Tsudik, "A Privacy-Preserving Index for Range Queries," Proc. Very Large Databases Conf. (VLDB), 2004.

[8] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private Information Retrieval," ACM Computer Survey, vol. 45, no. 6, pp. 965-981, 1998.

[9] 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.

[10] N.R. Draper and H. Smith, Applied Regression Analysis. Wiley, 1998.

BIOGRAPHY

Reena D K is a M.Tech Student in Computer Science and Engineering in STJIT Ranebennur affiliated to Visvesvaraya Technological University, Belgaum Karnataka. She completed her B.E in Information Science and Engineering from SDMCET Dharwad in the year 2012.Her area of interest are Data Mining, Cloud Computing Networking etc.