NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 1, ISSUE 2 DEC-2014

A survey on approaches to build efficient query services in cloud

Miss. Darpe R. Dipali
Department-Computer Networking/KJ College Of Engineering and Management Research, Kondhawa Saswad Road/ Pune/ Pune University /India.
darpedips12345@gmail.com

Abstract-

Now a days there are low-cost computers, storage devices and high-capacity networks are available which led to a growth in cloud computing. Cloud provides the advantage in scalability and cost-saving. The service owners can get the advantage on scale up or down the service and only pay for the hours of using the servers. The workloads of query services are highly dynamic and serving such workloads within house infrastructure will be inexpensive and inefficient. Data confidentiality and query privacy have become major concerns because the service providers lose the control over the data in the cloud. The data owner does not want to move to the cloud unless the data confidentiality and query privacy are guaranteed for the sensitive data..It is also necessary for a secure query service to provide efficient query processing and also reduce the in-house workload, because the purpose of using cloud resources is to reduce the need of maintaining scalable in-house infrastructure. Data confidentiality, query privacy, efficient query processing, and low in-house processing cost are the requirement for constructing practical query service in the cloud. Satisfying al this requirement will increase the complexity of building query services in cloud. Some related approaches are there which will address some aspects of problem. We give the comparative study of these approaches with their limitations.

Index terms- target distribution, LBS, bucketization, query privacy, data confidentiality

Introduction

Cloud computing is focused on maximizing the effectiveness of the shared resources. Cloud resources are shared by multiple users as well as dynamically reallocated as per user demand Cloud provides advantages in scalability and cost saving. So it becomes popular solution to host data intensive query services in cloud. Cloud provides the service owner an attractive feature to scale up or down the service and only pay for the hours of using the servers. Data confidentiality and query privacy have become major concerns because the service providers lose the control over the data in cloud. For security and privacy assurance it will not be meaningful to provide slow query services. The purpose of using cloud is to reduce the need of maintaining scalable in-house infrastructures. So it is not practical for the data owner to use a significant amount of in-house resources. For building the confidential and efficient query services in the cloud query privacy, data confidentiality and efficient query processing and low in-house processing cost are the requirements. The complexity of building query services in the cloud increases as we consider these requirements. The crypto-index [4] and Order Preserving Encryption (OPE) [1] are vulnerable to the attacks. The enhanced crypto-index approach [3] puts heavy burden on the in-house infrastructure to improve the security and privacy. The New Casper approach [2] uses cloaking boxes to protect data objects and queries, which affects the efficiency of query processing and the in-house workload.

Exiting approaches

Order preserving encryption-

To make query execution more secure one of the solutions is to integrate existing encryption techniques with database systems. But doing so will degrade the performance of query execution. Consider one example in which, if some columns of a table which contains sensitive information are encrypted, and if this columns are used in a query predicate with comparison operator, to evaluate such query an entire table scan would be needed. The reason behind this is the order of values is not preserved and the database indices are not useful. Thus the encrypted database query execution becomes slow. Rakesh Agarwal presented one encryption technique called as Order Preserving Encryption [1] which allows comparison operation to be directly applied on encrypted data. Thus, equality and range queries as well as the MAX, MIN, and COUNT queries can be directly processed over encrypted data. Similarly, GROUP BY and ORDER BY operations can also be applied [1]. OPE scheme is applicable for numeric data only. Non numeric data is not considered in this scheme. It allows comparison operation to be applied directly on encrypted data. The query results obtained with these schemes are very efficient and accurate with no false drops and no false hits. It also handles updating very gracefully. There is no need to make changes in the encryption of other values when adding new values. OPES has been designed to work with the existing indexing structures. This allows OPES to easily integrate with existing database system. User provides target distribution as input to OPES. Transformation is performed on the plaintext values. After transformation transformed values follows target distribution provided by user. In OPE scheme the value order is not changed even though a set of single dimensional values are mapped to another. It means that OPES preservers the order of transformed values. Working of OPES is divided into three stages.

1. Model: - It is the first stage in which it models the target distribution and input as piece-wise linear splines.

2. Flatten:-In this second stage the plain text database is transformed into flat database in which flat database values are uniformly distributed.

3. Transform: - In the last stage the flat database obtained from second stage is transformed into the cipher database. Here the values in cipher database are distributed according to the target distribution provided by user.
In OPES it enables to apply directly comparison operators on encrypted columns which contains numeric data. Query results obtained are very much accurate with no false positive and it does not miss any answer tuple. There is no need to to make changes in the encryption of other values when adding new values .OPES is designed for the environment where attacker does not have information about distribution of values and he is unable to encrypt or decrypt random values even though he will get access to encrypted database.

Casper

It is a privacy aware query processing framework which is useful in Location Based Services. Without providing information about own private location mobile and stationary users can obtain location based services continuously. In this scheme a location database server is embedded with privacy aware query processor to deal with continuous queries based on the knowledge of the users cloaked location. In LBS, to get the service user has to compromise his privacy. The concept of location anonymizer was introduced to preserve the privacy of LBS users. In this framework new privacy aware query types are identified such as private queries over public data, public queries over private data and private queries over private data. It gives framework to support privacy aware query types. Casper contains two components Privacy aware query processor and location anonymizer.

Location anonymizer:

There are two types of mobile users out of which some share their private location information and some wants to protect their private location information. In former case users can register directly with the location

based database server and in later case users register with the location anonymizer by specifying a certain privacy profile. The task of location anonymizer is to receive location changes from mobile users. Then with the help of location anonymization algorithm locations are blurring into clocked areas. This clocked area sent to the location based database server. Before sending a clocked query area to the database server query location information is also blurred by the location anonymizer. Privacy aware query processor provides location anonymizer with the candidate list of answers. Exact answers are computed from the candidate list and send it to the user.

Privacy aware Query Processor:

To deal with the cloaked areas from the location anonymizer location based database server is embedded with privacy aware query processor. It returns back the answer which includes candidate list of answers to the location anonymizer. It is the task of location anonymizer to filter this candidate list of answer throwing out false results and send the exact answer to the mobile users. Depending on users privacy profile candidate list is generated. In Casper approach it is easy for mobile users to obtain location based services without sharing their private location information. In Casper there are two components, the location anonymizer and the privacy aware query processor. The location of task anonymizer, blur the exat location information of user into cloaked area. To deal with these cloaked areas the privacy-aware query processor is embedded into traditional location-based database servers. To deal with privacy aware query candidate list of answer is provided by query processor

Crypto index

Now a day's usage of internet is rapidly increased. Due to the advances in software and networking organizations can share data for various purposes. DAS (Database as Service) is one paradigm where data management is outsourced to a service provider. There are many security and privacy challenges arise due to the unfrosted service provider. In DAS model bucketization approach is used. In this approach, an attribute is partitioned into a set of buckets each of which is identified by a tag [3]. These bucket tags are used by the server to process the queries. DAS model has two parties' i.e. client and servers. The clients are nothing but the data owners and database service providers act as servers. The database service providers are not trusted one. So before storing data at server client encrypt the data contents and data. Due to encryption performed on data stored at server no useful operations can be performed over encrypted data. Crypto index is created over sensitive attributes which are considered important in queries in DAS model. To support different types of queries multiple crypto-indices may be created over each attribute. The objective of Crypto-Index is to minimize work of client and force the server to perform maximum task of query processing. To enhance the privacy of data outsourced at server and to improve query processing Data bucketization techniques is proposed. In case when adversary has partial knowledge about the data buckets there is privacy loss occurs.

HASH MAP

One of the attractive targets for database attacker or hacker is important business data. For such a data to ensure privacy, integrity and data confidentiality becomes major issue. To protect such secure data various encryption techniques are applied. But performance of executing various queries decreases when data is encrypted. There is a need of such a mechanism which will work on many different data-types. The mechanism should also be required to work with any database. In [5] proposed approach based on layering technique. In this select condition is replaced with another faster condition. To perform search operation on encrypted data traditional approach is to first decrypt data execute query and then again encrypt the results again. But doing this will become costly process and time consuming. This will degrade the performance of the system when there are larger records present. To give effective performance on executing query indexing

techniques are also applied. In [5] to solve the problem of database compatibility one layer is added above DBMS. The responsibility of this layer is to handle query execution over encrypted data. But one of the drawback here is addition of such layer will affect the response time. In layered architecture client will send the query to the layer. And the layer is responsible for executing query on any kind of DBMS. Layered architecture has one subsystem which is known as Query processor. The task of Query Processor is to perform check operation on Meta data to check for any query over encrypted data. The Meta data contains an instance of a data structure object called Hash Map [5]. The task of Hash map is to store mapping between encrypted text and plain text. The mapping is stored as KEY: VALUE where plain text is KEY and encrypted value is the VALUE. The Query Processor replaces the client query with 'a plain where' clause on encrypted data value (the where clause is a plain text) with another one with an encryption on the plain searched data [5] . Encryption technique used is AES-256 to encrypt the important data. According to standard needed key is created for AES which will be stored on server side. To make searching operation faster an index is build over the encrypted column which helps to improve the performance. The hash map data structure makes use of hash function to perform mapping between encrypted data and plain text value. The Put Value (KEY, VALUE) function is used to store data on Hash Map where for VALUE, KEY works as an identifier. Another function that is Get Value (KEY): VALUE is used to obtain VALUE by passing the KEY. In Hash Map method is proposed to work to execute the query over encrypted database which can work with many data types. It is implemented as layer above any database. Due to this reason it will not affect in inner structure of the DBMS.

CONCLUSION

In Order Preserving Encryption scheme order of dimensional values is preserved due to which it is suffer from distribution based attack. In case of Casper approach data objects and queries are protected by using cloaking boxes. Due to this in-house workload increases ant it also affects the efficiency of processing queries. In case of Crypto index when adversary has partial knowledge about the data buckets privacy loss occurs there. In hash map layering technique there is extra overhead to maintain layer to support any kind of database. So to build efficient query services in cloud there is need to design such mechanism which will reduce the need of maintaining scalable in-house infrastructure and execute the queries efficiently with the low in-house workload for protected data in 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. 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.
[3] B. Hore, S. Mehrotra, and G. Tsudik, "A privacy-preserving index for range queries," in *Proceedings of Very Large Databases Conference (VLDB)*, 2004.
[4] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, "Executing sql over encrypted data in the database-service-provider model," in *Proceedings of ACM SIGMOD Conference*, 2002.
[5] Mohammed Alhanjouri Asst. Prof. at Islamic university of Gaza Gaza, Palestine Ayman M. Al Derawi Islamic university of Gaza Gaza, Palestine " A New Method of Query over Encrypted Data in Database using Hash Map" *International Journal of Computer Applications (0975 – 8887) Volume 41– No.4, March 2012*

ACKNOWLEDGMENT

I express true sense of gratitude towards my project guide Prof. Rohini V. Agawane , Assistant Prof. of computer engineering department for her invaluable co-operation and guidance that she helping me for my project study. I like to thank her once again for inspiring me and providing me all the lab facilities, which made this survey work very convenient and easy. I would also like to express my appreciation and thanks to Prof. Dipak C. Mehtre Head of Computer Engineering Department and Principal Dr. S. J. Wagh and all my friends who knowingly or unknowingly have assisted me throughout my hard work**.**