MAT JOURNALS

Journal of Data Mining and Management
Volume 3 Issue 1

Query Processing on Encoded Data using Bitmap

Sarthak Bakshi, Shreyas Chavan, *Anuj, Shyam Hargaonkar
KKWIEER, Nashik
Email:bakshi.sarthak@gmail.com, shreyaschavan1@gmail.com, anuj55f@gmail.com, shyamhargaonkar@gmail.com
*Corresponding Author

Abstract

As database has been highlighted, data encryption schemes are required to protect database from unauthorized access so, to efficiently manage the large amount of encrypted data distributed index is needed with query processing scheme over it. The schemes that are present or exist for processing the query over encrypted data can support limited types of queries. Also data which is in encrypted format need to be decrypted before performing queries as such schemes does not support operations between different columns due to the use of different type of encryption keys (Asymmetric Cryptography). To solve this problem, we propose the Encoded technique which uses bitmap to convert our dataset into encoded dataset and performs query processing on encoded data and also this proposed technique guarantees data privacy preservation and performance improvement for the various types of queries. In addition, it protects our private information from third party to whom we are outsourcing our data. In short we are going to process a query over the encoded data without data decoding. The proposed query processing scheme using bitmap provide both high query performance and accuracy while preserving the data privacy from unauthorized access and also from third party.

Keywords: *Data encryption using Bitmap, Encrypted query processing.*

INTRODUCTION

As a large amount of database has recently been outsourced, data encryption schemes are protect the sensitive data. Accordingly, it is necessary to develop not only a distributed index structure to efficiently manage the large-scale encrypted data, but also a query processing scheme over the encrypted data because to process any query one encrypted data first we have to decrypt the data then only we can perform query processing over data which is not to efficient to use and it is does not have any security. So to solve this, we proposed a bitmap-based query processing scheme which can process a query over the encrypted data without data decryption. Query processing scheme provides both high query processing performance and high query result accuracy while preserving data privacy.

Literature Survey [1][3]

For this work existing query processing schemes decrypt an encrypted database to process a user's query, the original data may be revealed to a malicious attacker. Therefore, query processing schemes over the encrypted database have been actively studied. They can be classified into two categories, SQL-like query processing scheme and aggregation query processing schemes. First proposed method for typical SQL-like query processing scheme over the encrypted database, called CryptDB. CryptDB encrypts data in a column-wise way by considering various query types, i.e., exact-match, range query and so on. To support the exact-match query, for example, it encrypts data using a deterministic encryption method such as AES. However, CryptDB cannot support

operations among data with different columns because they use different types of encryption schemes depending on their attribute types. Therefore, CryptDB has difficulties to support the data analysis or data mining over large sensitive data. Meanwhile, aggregation query processing schemes over the encrypted database are as follows. Second proposed method for query processing scheme is by using an additive homomorphic encryption scheme. The additive homomorphic encryption scheme has a property that the encrypted value of the summation of the original data is the same with the multiplication off their encrypted data. The scheme can calculate the real aggregation result by merging secret values without decrypting data. The existing query processing schemes over the encrypted databases have some problems. First, CryptDB cannot support operations among data with different columns because they use different types of encryption schemes depending on their attribute types. Second, schemes proposed by T. Ge et al and J. Corena et al. cannot support SQL-like queries, such as exact matching, range and join quires. In addition, the query processing cost is high because they use homomorphic encryption. Finally, the scheme proposed by B. Thompson et al. may expose the original data when a service provider holds all secret values.

Encryption [4][2]:

In cryptography, Encryption is the process of encoding a message or information in such a way that only authorized parties can access it and those who are not authorized cannot. Encryption does not itself prevent interference, but denies the intelligible content to a would-be interceptor. In an encryption scheme, the intended information or message, referred to as plaintext, is encrypted using an encryption algorithm that is using a cipher generating Cipher Text that can be read only if decrypted. Unencrypted data which is often referred as plaintext is encrypted using various encryption algorithms and encryption key. This process generates cipher text which is an unreadable text. It is not possible to read this cipher text without decrypting it. Only authorized users and members who possess the key provided by the owner can decrypt the message. Encryption plays vital role in security but also provides Integrity, Authentication and Non repudiation.

Common Encryption techniques

1. Triple DES

Triple DES is an encryption technique where block cipher algorithm is applied three times. It was developed as DES became prone to brute force attack. The key size is increased in this for additional security. In Triple DES approach of original DES is not completely abandoned but modified. Triple DES uses three individual keys with 56 bits each. The total key length adds up to 168 bits, but experts would argue that 112-bits in key strength are more like it. Despite slowly being phased out, Triple DES still manages to make a dependable hardware encryption solution for financial services and other industries.

2. RSA

RSA is a public-key encryption algorithm and the standard for encrypting data sent over the internet. It also happens to be one of the methods used in our PGP and GPG programs. Unlike Triple DES, RSA is considered an asymmetric algorithm due to its use of a pair of keys. RSA uses two different but mathematically linked keys that are One Public key and One Private key. The public key can be shared with

everyone but the private key must be kept secret. Both the keys are used for encryption in RSA. RSA keys are used for Authentications and Verification purpose. The best example for RSA is **SSL certificate**.

Client (Browser) and Server do RSA key exchange during SSL handshake. Today an SSL Certificate uses 2048-Bit RSA Key.

3. Blowfish

Blowfish is yet another algorithm designed to replace DES. This symmetric cipher splits messages into blocks of 64 bits and encrypts them individually. Blowfish is known for both its tremendous speed and overall effectiveness as many claim that it has never been defeated. Blowfish is unpatented, license free and available free for all users. It is significantly faster than DES and one of the flexible encryption techniques available. It is designed with 32 bit processors in mind.

4. Twofish

Computer security expert Bruce Schneier is the mastermind behind Blowfish and its successor Twofish. Keys used in this algorithm may be up to 256 bits in length and as a symmetric technique, only one key is needed. Twofish is one of the fastest encryption techniques. It is said to be efficient for software that runs on a smaller processors such as smart card and for embedding hardware.

5. AES

The Advanced Encryption Standard is a block cipher chosen by US government to protect the classified information and data. National institute of Standard and Technology started developing AES as a successor to DES which became vulnerable to brute force attack in 1997. AES uses 128, 192 and 256 bits for encryption purpose. AES is largely considered impervious to all attacks, with the exception of brute force, which attempts to decipher messages using all possible combinations in the 128, 192, or 256-bit cipher. Still, security experts believe that AES will eventually be hailed the de facto standard for encrypting data in the private sector.

Encryption Use [5]

With more inclusion of smart devices to our daily lifestyle the importance/usage of cryptographic algorithm is also increasing. Applications like WhatsApp they use crypto algorithm for securing your data. Satellite TV and DVDs. they also use cryptography Broadcast encryption. Keyless entry feature of car or Remote keyless system use Rolling Code and are prone to attacks.

Wifi (wireless fidelity) network is mostly present at every residence and at working sector used for various purposes. This network is usually protected with WPA/WPA2 security which is application of cryptography algorithm.

RFID smart cards the ones used in metro or office ID card at office also uses cryptography.

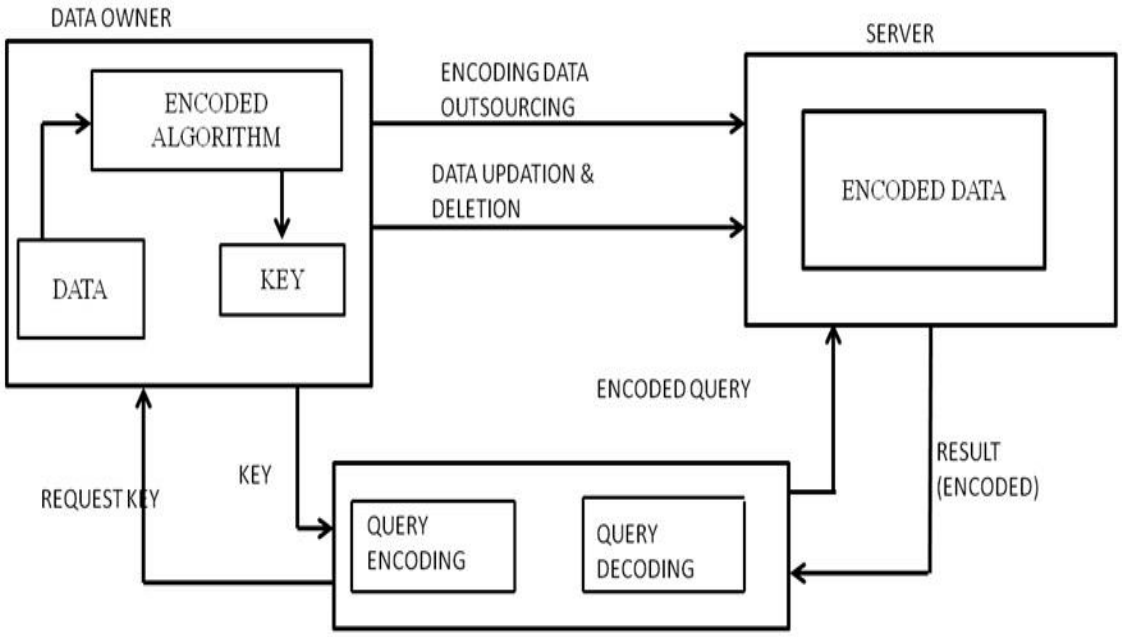
- Your computer (Login, password protecting a file)
- Internet Banking (Login, RSA token)
- Website interaction (Login)
- Secured websites (HTTPS)
- Encrypted secure mail
- Digital currency (Bitcoin)
- Virtual Private Networks (IPsec)
- Banks (3DES, AES (for symmetric encryption) and RSA/DSA/ECC (for key exchange))

Proposed Methodology

As encryption techniques that have been used at various levels now days have lots of overheads like limited query processing which requires decryption of data, decrypting data on the server, time complexity we have proposed an efficient

system which will overcome all this overheads and will be better to secure confidential and important data. Modified encryption technique (i.e. technique we are using) is used for encoding in the following proposed system.

The design and working of the proposed system is given below.



Block Diagram

- Server
  - A Server is nothing but third person storage on which we are store our data.
  - The third person is unknown to us we cannot trust him we encode our data before storing it on server.
  - If anyone wants to get access to the data first he has to verify himself to the data owner then only he can get access to database.
- Data Owner
  - He/she is the owner of the data. Data which we want to protect using Encoding Algorithm so that he/ she can safely store his/her data to any unknown server randomly authorized user can get access to it by using the Key.
- Encoded Algorithm
  - It is the algorithm which is used by the data owner to encode the database before uploading it to the server.
  - It also generates one key which can be used by trusted users only.
- Key
  - It is a key which is generated by Encoded algorithm and it is used for query processing over Encoded database by trusted user.
  - This key can perform to operation one is encoding the query and decoding the result given by that query.
- Authorized user
  - An Authorized user is the trusted user which can access the data by using different types of query. So to get accessed to the data user must have to confirm himself to data owner and data owner provide key to user can use that key to get access to database.

CONCLUSION

Large amount of data has been outsourced so data encryption techniques that are used to protect the sensitive data are also required. To efficiently manage the large amount of encrypted data distributed index is needed with query processing scheme over it. For this we propose a solution that is Bitmap based indexed structure and encoded query processing over big-data (any large data). This proposed structure guarantees data privacy preservation and good performance for dealing various types of queries. Query processing scheme provides both high query performance as well as high result accuracy while protecting sensitive data. This structure also minimizes the time complexity for better performance of the system. We can show from our performance analysis that the proposed structure is suitable for protecting large scale data from attacker in data outsourcing environment.

ACKNOWLEDGEMENT

We are grateful to our guide, Prof. D M Kanade for his valuable guidance for this research work. We would also like to show gratitude towards our institution K. K. Wagh Institute of Engineering Education and Research, Nasik and Head of Computer Engineering Department Prof. Dr. S. S. Sane for providing us the platform required for the work. We would also like to thank our colleagues for their useful insights towards this work.

REFERENCES

1. Lil Marıa Rodrıguez-Henrıquez and Debrup Chakraborty, Using Bitmaps for Executing Range Queries in Encrypted Databases of In IEEE Symposium on Security and Privacy, pages 463–477. IEEE Computer Society.
2. Osama M Ben Omran, Brajendra Panda, Facilitating Secure Query Processing on Encrypted Databases of IEEE International Conference on Smart Cloud
3. Yuhao Wen, Han Wang, Zhen Chen, Junwei Cao, Guodong Peng MASC: A Bitmap Index Encoding Algorithm for Fast Data Retrieval of IEEE ICC 2016 SAC Data Storage
4. Bitmap-based Distributed Index Structure and Encrypted Query Processing Schemes for Outsourcing Mobile Sensitive Data Hyunjo Lee, Hyeong-Jin Kim, Jae-Woo Chang* Dept. of Computer Engineering Chonbuk National University, Republic of Korea {o2near, yeon_hui4,jwchang}@jbnu.ac.kr *Corresponding author Hyeong-Il Kim The 1st Missile Systems PMO Agency for Defense Development, Republic of Korea hikim@add.re.kr
5. https://www.quora.com/in/Where-are-cryptographic-algorithms-used-in-our-daily-applications