A Comparative Study and Analysis of Query in Encrypted Databases

Atheer Metaab Abbas **Abdul Monem S. Rahma** **Nidaa F. Hassan**

Department of Computer Science, University of Technology, Baghdad, Iraq
etheer_78@yahoo.com
monem.rahm@yahoo.com
110020@uotechnology.edu.iq

ARTICLE INFO
Submission date: 1 / 12/ 2019
Acceptance date: 30/ 12 / 2019
Publication date: 31/ 12 / 2019

Abstract

Data is the main asset of the modern companies and their businesses. Typically, it is stored in the data-base. Every database system has to be capable of responding to information requests from users, which is queries of the process. Encryption mechanisms are required, which give the capability to the query over the encrypted data-base and permit the optimization of data encryption and decryption. However, applying encryption algorithms on Encrypted database then challenge arises that the efficiency of the system degrades on deploying encryption algorithm on the runtime. Therefore, this paper presents most of the recent works that have been conducted on the query preprocessing of encrypted database and analyzes them to clarify the performance analysis, based on different performance metrics in each related work.

Keywords: Query, Database, Security, Encryption, Performance, Metrics.

1. Introduction

Data can be considered as the main element, as the whole organization is dependent on it. This dependency is of high intensity that the organization's objectives success or failure are dependent on the data quantity and quality. As a result, organizations cannot afford losing important data which is available in an institute and its businesses. A big data amount is stored in a repository which is the data-base [1] [2]. Database is a set of correlated data, which may be logged and include implicit meanings. Data is categorized as insensitive or sensitive. The latter is secured in the data-base with the use of encryption schemes. Data which is stored in data-bases is structured and generally stored as relational tables as the majority of establishments utilize relational databases [3]. It is helpful in organizing data for more sufficient performance and more timely retrieval via the maintenance of the locations. It can be helpful in the preservation of transaction logs that help in data recovery. Database management systems perform the concurrency control function. Database management systems can as well perform the operations of database recovery of data. Due to the fact that data which is stored in the data-bases could be critical, there is an importance in securing it. A data-base may be under attack in a variety of ways. It is possible to attack data which is stored in the data-bases due to the fact that the data-bases are interfaced with some applications and via the hampering of applications; there is a possibility in attacking the databases. The algorithms of encryption may be categorized into symmetric encryption and asymmetric encryption. In the symmetric algorithms of encryption, single or shared key is utilized for providing confidentiality for the data-base; whereas in the asymmetric algorithms of encryption two keys are utilized for providing security, one public and one private [4].

Cryptography is widely utilized for supporting the security of databases as the method of encryption offers a sufficient approach for securely storing confidential data. None-the-less, as the cases of addressing information security, the performance is influenced directly. The costs of data encryption and decryption which is inserted or obtained from the data-base increases the regular costs of data storage and retrieval of from the unencrypted data-base [5],[6].

This paper presents numerous encryption techniques that have been recently proposed for query data efficiently.

2. Encrypted Database

For the sake of securing a data-base, encryption is necessary for encrypting sensitive data, and as well for the authentication of users and data integrity too. There are techniques of cryptographic algorithm which were considering the structure and cons points in database and its utilized easily in query processing as follows [7,8] :

- **Privacy homomorphic encryption** permits to perform calculations on ciphertext and get an output which matches the calculations which are carried out on the plaintext. Instances of this type of approaches which allow executing queries of aggregation on enciphered data are found in both [9] and [10], none-the-less, each one of those two solutions has been discovered to include security gaps.
- **Order Preserving Encryptions:** [11] presents an order preserving encryption scheme. Which offers the creation of indices on the ciphertext and for direct comparison on the encrypted data. Which is why, equality, range queries in addition to COUNT, MAX, and MIN queries may be directly carried out on the encrypted data. In addition to that, updating values does not break the model. None-the-less this model showed to be insecure as well.
- **Fast Comparison Encryption:** This model provides fast comparisons between the encrypted data. The processes of encryption and decryption are performed byte-by-byte beginning with the most significant byte, therefore decryption of the two values compared can be implemented with ”early stopping”. That is, the decryption will stop when a difference between the two values has been found in [12].

3. Previous Works in Query Processing on Encrypted Database

The criteria regarded for the encryption of the data-bases for the sake of protecting them from attacks have been proposed by many researchers. In [13, 14] developed an approach which explicitly manages encrypted data with no need to decrypt it, where data are enciphered with the use of algorithms that are based on the homomorphism of the privacy. Song [15] has presented an innovative scheme of encryption which allowed to search encrypted data with no need for the decryption. None-the-less, the scheme of encryption which has been utilized in their method hasn’t adjusted for the data-base. Hankan [16] has suggested a method to execute SQL over encrypted data in the model of data-base service provider. None-the-less, the way is only valid for numerical data, and is impractical for character data. One more limitation of this approach is that it’ll result in many false joining records in the case of the querying over the multi-tables that results in considerably increasing in record decryption costs, as a result, this approach considerably reduces the efficiency of the performance. Zheng-Fei et al. [17], have suggested a function for supporting fuzzy query over encrypted character data. This method performs the conversion of every 2 adjacent characters in a sequence and the conversion of original string a direct way to another string of characters via the hash function. This approach is not capable of dealing with some characters and could be performing poorly for long strings of characters.

Alhanjouri and Al Derawi [18] have suggested using Hash Maps for the improvement of encrypted data-bases performance. They have claimed to have devised an approach of enhancing the speed of response for the queries on the encrypted data-bases. The suggested approach is involved with the construction of an extra layer above the database management system, which is made up of a query processor, meta-data, an encryption/decryption function, and a hash map. The authors didn’t discuss the method that they have suggested for the protection of the actual layer that raises questions about the effectiveness of the presented approach in the preservation of data confidentiality in the first place. The Reverse Encryption Algorithm (REA) is an important enhancement over encrypted data-bases has been suggested by Mousa et al. [19]. The results of this algorithm might decrease the encryption/decryption operations cost time and enhance performance, however, data-base encryption isn’t optimally reliable and requires some additional security via the encryption of data with some other algorithm, for tightening security with no degradation in the performance.

Sharma et al. [20] have provided the users with the ability of directly querying over the encrypted column with no need for the decryption of every record. It enhances system efficiency. The method that they have proposed suggests 2 tables for one main table for the introduction of security in the data-base. The first one of the tables is referred to as the Encrypted_Data_Table that includes the data itself and the second table is referred to as the Query_Search_Table which contains only data on which search query operates. In the case where an authorized user decides to search some of the records from Encrypted_Data_Table and condition of the search is on encrypted column, which is why, the search is going to be carried out on Query_Search_Table. Arasu et al. [21] proposed a data encryption system in which the sensitive columns are encrypted prior to storing them for addressing the security of data.

AL-Saraireh J. [22] presented “An Efficient Approach for Query Processing Over Encrypted Database”. Which is an innovative method that has been proposed in this study for improving query efficiency over the encrypted data-base. It has been modeled on the basis of the use of a hash map function for the generation of a distinct hash value for each sensitive data. In this method, there isn’t any correlation between hashed and encrypted values. This approach may decrease encryption and decryption operations cost and enhance the performance of the cost. Awais Ahmad [23] has suggested an approach of parallel query execution with the use of multi-threading approach up to 6 threads with tests up to 1 million encrypted records. It has utilized AES with 256-bit blocking length.

4. Performance Metrics of Pervious Works

The aim of this study is to evaluate query processing on encrypted databases for the sake of understanding main issues or problems which may overcome in future works, the performance metrics considered in this paper are:

1. Dose query process implies decrypt encrypted database partially?
2. Dose query process implies decrypt encrypted part of the record?
3. Are Metadata used in addition to the query?
4. Are the measures of performance of query processing is computed according to the execution time of the query?
5. Does the problem of Query processing works on heterogeneous sources of data?
6. Do the time of encryption and decryption results have shown that the suggested algorithm of encryption accomplished sufficient performance in comparison with other algorithms of encryption?
7. Does the encryption and decryption of database depend on lightweight procedure?
8. Does the query process running on remote database?

As previous study, Many researchers work on encrypted of database based on query techniques, Table (1) demonstrates wether the above previous works achieve performance metrics or not.

Table (1) : Performance Metrics of Pervious Works

Id	Research Title	Performance Metrics							
		1	2	3	4	5	6	7	8
1	On Data Banks and Privacy Homomorphisms [13]	×	×	×	×	✓	×	×	×
2	Processing Encrypted Data[14]	×	×	×	✓	✓	✓	×	×
3	The execution of SQL over Encrypted Data in the Model of Data-base Server-Provider [15]	×	×	×	×	×	✓	×	✓
4	Storage and query over the encrypted character and the numerical data in the data-base [16]	×	×	×	✓	×	✓	×	×
5	An Innovative approach of query over encrypted data in the data-base with the use of the hash map [17]	✓	×	✓	✓	✓	✓	✓	×
6	The performance of query processing on the encrypted data-bases with the use of REA mechanism [18]	×	×	×	✓	×	✓	✓	×
7	Performance of the Query Processing and Search over Encrypted Data with the use of an Efficient Algorithm [19]	✓	×	✓	✓	×	✓	✓	×
8	Querying encrypted data [20]	×	×	✓	×	✓	✓	×	✓
9	A Sufficient Method for Query Processing Over Encrypted Databases [21]	✓	×	✓	✓	×	✓	×	×
10	Parallel processing of the queries over encrypted data in the data-base as a service (DaaS) [22]	✓	×	✓	✓	×	✓	✓	✓

The best papers of previous works according to above table whose (id=5, id=10), they pass six points from performance metrics.

5. Conclusion

This paper presents most of the recent works that have been conducted in query of encrypted database and analyzes them to clarify the performance metrics in previous works based on set of factors. Some of them, developing methods that pass some of the performance metrics but they have set of issues in other performance metrics. In order to satisfy balance between time and complexity in Query of Encrypted Database, a novel proposal must be suggested to achieve high performance in the encryption and time of Query by pass all performance metrics.

Conflict of Interests.
There are non-conflicts of interest

References

1. E. Bertino and R. sandhu. “Database Security- Concepts, Approaches and Challenges”, IEEE Transactions on Dependable and Secure Computing, Vol 2, No 1, January-March 2005.
2. C. Ribeiro and G. David. “Database Preservation”, briefing paper, Website, June 6, 2012.
3. C. Pfleeger and S. Pfleeger. "Security in Computing", third edition,2004.
4. R. H. Rathod and C. A. Dhote. “A literature survey on performance evaluation of query processing on encrypted database”. Int. J. Eng. Comput. Sci., 3:9637-9642, 2014.
5. M. Nassar et al. “Securing aggregate queries for DNA databases”. IEEE Trans. Cloud Comput. DOI: 10.1109/TCC.2017.2682860, 2017.
6. A. Ali and M. M. Afzal. “Database security: Threats and solutions”. Int. J. Eng. Inventions, 6: 2278-7461. 2017.
7. Mohammed Salih Mahdi and Nidaa Flaih Hassan," A SUGGESTED SUPER SALSA STREAM CIPHER ", Iraqi Journal for Computers and Informatics Vol. [44], Issue [2], Year (2018)
8. H. Hacigumus at el. “Efficient execution of aggregation queries over encrypted relational databases”. DASFAA, 2004.
9. S. Chung and G. Ozsoyoglu. “Anti-tamper databases: Processing aggregate queries over encrypted databases”, Proceedings of the 22nd International Conference on Data Engineering Workshops, Washington, 2006.
10. R. Agrawal et al. “Order Preserving Encryption for Numeric Data”, 2004.
11. T. Ge and S. Zdonik. “Fast, secure encryption for indexing in a column oriented DBMS”, In International Conference on Data Engineering – ICDE 2007, IEEE, 2007.
12. R. L. Rivest et al. “On Data Banks and Privacy Homomorphisms”, in: Foundations of Secure Computation, pp. 169–178, 1978. N.
13. Ahitub et al. “Processing Encrypted Data, Communications of the ACM”., pp. 777–780, September 1987.
14. D. Xiaodong Song et al. “Practical Techniques for Searches on Encrypted Data”, IEEE Symposium on Security and Privacy., pp. 44–55, 2000.
15. H. Hacigumus et al. “Executing SQL over Encrypted Data in the Database-Server-Provider Model”, in: Proc of ACM SIGMOD, pp. 216–227, 2002.

16. W. Zheng-Fei et al. “Storage and query over encrypted character and numerical data in database”. Proceedings of the 5th International Conference on Computer and Information Technology, Sept. 21-23, IEEE Xplore press, Shanghai, China, pp: 77-81. 2005.
17. M. Alhanjouri and A.M. Al Derawi. “A New method of query over encrypted data in database using hash map”. Int. J. Comp. Appl., 41: 975-888. , 2012.
18. A. Mousa et al. “Query processing performance on encrypted databases by using the REA algorithm”. Int. J.Network Security, 14: 280-88. 2013.
19. M. Sharma et al. “Query Processing Performance and Searching over Encrypted Data by using an Efficient Algorithm”, International Journal of Computer Applications (0975 – 8887), Volume 62–No.10, January 2013.
20. A. Arasu et al. “Querying encrypted data”. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 22-27, ACM Press, New York, USA, pp: 1559-1261., 2014.
21. J. AL-Saraireh, “An Efficient Approach for Query Processing Over Encrypted Database”, journal of Computer Science,2017.
22. Awais Ahmad1 ·,"Parallel query execution over encrypted data in database‑as‑a‑service (DaaS)", Springer Science+Business Media, LLC, part of Springer Nature 2019.

الخلاصة

تعد البيانات اليوم هي الموجود الرئيسي للشركات وأعمالها. عادة ما يتم تخزين هذه البيانات في قاعدة البيانات. يجب أن تكون جميع أنظمة قواعد البيانات قادرة على الاستجابة لطلبات الحصول على معلومات من المستخدم وهي استعلامات العمليات. هناك حاجة إلى خوارزميات التشفير التي توفر القدرة على الاستعلام عبر قاعدة البيانات المشفرة وتتيح تحسين تشفير وفك تشفير البيانات. ومع ذلك ، تطبيق خوارزميات التشفير على قاعدة البيانات المشفرة يعتبر التحدي الذي ينشأ هو في انخفاض اداء النظام الناتج عن نشر خوارزمية التشفير في وقت التشغيل. لذلك ، تعرض هذه الورقة معظم الأعمال الحديثة التي تم إجراؤها على معالجة الاستعلام في قاعدة البيانات المشفرة وتحليلها لتوضيح تحليل الأداء ، بناءً على مقاييس أداء مختلفة في كل عمل ذي صلة.

الكلمات الدالة: الاستفسار, قاعدة البيانات, الامنية, التشفير, الكفاءة, المقاييس.