TuijinJishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No.6 (2023),

An Efficient and Novel Approach for Query Processing Over Encrypted Database in Big Data

[1]Dr. Prakasha S, Surekha Pinnapati[2]

[a] *Associate Professor, Department of Computer Science Engineering, R N S I T, Visvesvaraya Technological University, Belagavi, Karnataka-560098*

[b] *Research Scholar, Department of Computer Science Engineering, R N S I T, Visvesvaraya Technological University, Belagavi, Karnataka-560098*

***Abstract*:-** The study proposes an innovative approach to enhance query performance within encrypted atabases, centered on generating unique hash values for individual sensitive data pieces using a hash map function. Notably, the suggested method maintains no correlation between the hashed value and the encrypted value. This process has the potential to heighten query performance while concurrently reducing the encryption and decryption costs. Results from a sequence of experiments showcase accelerated query performance across a spectrum of queries on encrypted data particularly, when the database comprises between 20,000 and 50,000 rows, a comparative analysis with different approaches coducted.

Keywords*: SQL, Database Security, Encryption, Hash Map, query processing*

1. Introduction

A database constitutes an interconnected compilation of information where data, representing factual information imbued with implicit significance, resides. Within this spectrum, data bifurcates into sensitive and insensitive categories. Protection of sensitive data within databases is achieved through encryption mechanisms. Encryption techniques fall into two principal categories: symmetric and asymmetric encryption. Asymmetric encryption employs a pair of keys—the public and private keys—to furnish security, while symmetric encryption relies on a single shared key to ensure confidentiality within database services. Conventional security measures such as access control, physical safeguards, and network security prove inadequate in ensuring the secure handling and storage of sensitive data. The adoption of encryption stands as an effective remedy for safeguarding critical data, as asserted by (Rathod and Dhote, 2014; Nassar et al., 2017; Ali and Afzal, 2017). While cryptography has significantly buttressed database security, the trade-off between security and performance has been an enduring concern. Encrypting and decrypting data for storage or retrieval within an unencrypted database escalates the standard cost associated with storing and accessing data.

The conventional method of decrypting all encrypted data to plaintext for record retrieval is notably time-consuming and exhibits subpar performance, particularly when dealing with a substantial volume of records. This research confronts the challenge of preserving database data security without compromising system speed, aiming to achieve a balance between security and performance. Introducing a strategy that harmonizes security measures with system speed across diverse database types, this study explores the prospect of advancing upon prior approaches.

Objectives

1. To design an efficient searching technique to retrieve relevant data from dataset based on features and attribute range that matched with the query.
2. Develop privacy preservation technique to identify most vulnerable attributes.
3. Design encryption technique to encrypt relevant data.
4. Design compression technique to compress the encrypted data.

5. Develop a technique to store a compressed data on cloud.

Literature survey

Bouganim and Guo (2011)-1 introduced the factors associated with encrypting databases for protection against attacks. They delineated that the first factor encompassed the implementation of three encryption levels: storage, database, and application. Another crucial aspect involved the choice of encryption algorithms such as DES and AES, alongside the key management techniques integrated into each algorithm.

Salama *et al*. (2010) -2 They examined a range of encryption algorithms, assessing their impact on system performance concerning various factors like data block size, data types, battery consumption, key sizes, and encryption/decryption speeds. The algorithms scrutinized in the study encompassed AES (Rijndael), DES, 3DES, RC2, Blowfish, and RC6. The conducted experiments yielded conclusions: AES demonstrated superior speed, whereas Blowfish excelled under varying packet sizes. Surprisingly, the data type utilized for encryption/decryption showcased negligible influence on algorithm performance. Notably, the experiments underscored the pivotal role of key size as a primary factor influencing the power consumption of an encryption algorithm. In our research, we'll integrate these findings when evaluating the efficacy of the methods devised to bolster database system security while maintaining optimal performance levels within the targeted database system.

Sharma *et al*. (2013)-3 Introduced was an approach that aims to achieve data confidentiality while striking a balance between security and performance, prioritizing expedited data retrieval. This methodology entails the utilization of two tables: the Encrypted Data Table, preserving the primary column in an encrypted format, and the Query Search Table, storing the encrypted key alongside the same column from the Encrypted Data Table, albeit in plaintext form. Under this methodology, the key column assumes dual names across both tables, mitigating ambiguity in range and fuzzy match queries while amplifying query processing efficiency. Furthermore, to bolster confidentiality, the sequence of records within the Query Search Table is randomized, contributing to data obfuscation. To obviate relational patterns, the Query Search Table is ensconced within a secure schema and augmented with noise in the records to deter inference, accessible solely to authorized users. An added benefit lies in the execution of query statements, circumventing the need to decrypt all values within the encrypted column.

Alhanjouri and Al Derawi (2012)-4 Suggested was the utilization of Hash Maps to optimize encrypted database performance, purporting an approach to expedite query response times. This method entails constructing a supplementary layer atop the DBMS, comprising metadata, a query processor, a hash map, and encryption/decryption functions. However, the authors' omission regarding the protection of this layer raises concerns about the efficacy of their method in maintaining the initial data confidentiality.

(Mousa et al., 2012)-5 introduced a Reverse Encryption Algorithm (REA) as a substantial enhancement over conventional encrypted databases. The outcomes of employing REA exhibit a reduction in the time cost of encryption/decryption operations, augmenting overall system performance. However, while REA improves efficiency, it falls short in ensuring optimal data integrity. Consequently, an additional layer of security is recommended, involving the encryption of data with an alternate algorithm to fortify security measures without compromising performance.

Zheng-Fei *et al*. (2005; Wang *et al*., 2004) -6 They introduced a function aimed at facilitating fuzzy queries on encrypted character data. Their method involves converting adjacent character pairs within the sequence using a hash function, transforming the original string into another character string directly. However, limitations arise in handling certain characters, particularly with larger character strings, potentially leading to poor performance. This research proposes a novel approach to enhance query performance over encrypted databases. This method generates unique hash values for each sensitive field, translating SQL clauses into a suitable format for execution within an encrypted database. As a result, this approach diminishes the encryption/decryption costs while boosting query performance.

The Casino et al. [1] -7 illustrate blockchain as the Decentralized RS core, seeking to provide it with a wide variety of functionality while protecting consumer privacy. We are implementing a new design focused on decentralized area sensitive hashing classification and a range of suggestion methods, depending on how users

handle data. Extensive study findings show the efficiency and efficacy of our methodology relative to cutting-edge approaches.

Chenyang Ma **et al. [8]** Describes Separating encrypted transactions in the block and calculating only bilinear pairings on node block cypher texts rather than all ciphertexts, which helps to reduce the computing costs that mining operation. Finally, we test the efficacy of our set of rules by conducting statistical analyses and simulator tests on numerical expenses, safety, precision, and time aspects. The results indicate our protocol would produce the right mining outcome and outperform the prior method in terms of performance under a similar protection standard of conditions.

Dongfeng Fang et al. [9] explained IoT network architecture, which contains heterogeneous IoT systems listed. Specific confidence models are developed and evaluated in the IoT framework depending on the confidence relationships between the different actors. We suggest a scalable and robust authentication scheme that considers heterogeneous IoT devices based on a paradigm needed for the least trust. The proposed method offers resource-limited IoT users protection and privacy in a scalable and effective way by using IoT users with improved storage and computational ability.

Anmin Fu et al. [10] Tackle this by introducing a novel Non-negative Matrix Factorization (NMF) outsourced scheme that seeks to reduce the computational pressure of consumers and resolve stable issues posed by NMF outsourcing. O-NMF specifically exploits Paillier homomorphism, focused on two non-collusion servers, to protect data privacy. O-NMF provides a testing system to support clients validate returned findings with a high degree of accuracy probability.

Muhammad Usman et al. [11] This system works in three steps. The first level edge devices affect a lightweight aggregation method to produced data during the first step. This approach limits the size of the data produced and seeks to protect data source privacy. In the 2nd tier, a multi-step process used for linking Level Two Edge Devices (LEDs) with High-Level Edge Devices (LEDs). The validation phase-only valid LEDs can move data to the LEDs, resulting in a reduction in the computational burden on LEDs. In the third level, the LEDs use a convolution neural network to predict the position of touching objects in LED data transmitted.

Jiannan Wei et al. [12] Propose a secure, safe and privacy-conserving IoT Message authentication scheme. Our framework embraces IoT devices with varying cryptographic settings and enables offline and online computing, making them more versatile and powerful than previous systems.

Rong Jiang et al. [13] Big Data processing, delivery, analysis, usage and sharing examined protection and privacy leakage possibilities. They recognized a Clinical Big Data Security and Privacy Leakage Possibility Predictor framework amid four leading indicators and thirty minor indicators. Additionally, weight for every variable was determined using the weight system GI and Entropy. The Fuzzy Method of Structured Analysis was recognized to check the principle of Big Data Medical Protection and Urban Privacy Computing.

Karen R. Sollinset. al. [14] explains requirements and limits and proposes a three-part decomposition of architecture. To arrive at this final analysis, we begin by clarifying the issues in the design space: There is some agreement about what IoT means, particularly on the security and privacy consequences of various definitions. 1. We then consider the requirements and constraints on Big Data resulting from unique IoT system designs.

2. We examine the industry intricacies in parallel. In this sense, we can then break down the set of drivers and data protection/privacy and innovation goals into1) the history of regularity and social policy; 2) the economic and industrial history; and 3) the context of technology and architecture. Ismail Hababehet. al. [15] proposes an advanced approach for classifying and safeguarding extensive data while conducting versatility, duplication, and study. The data classification defines the requirement to protect extensive data accessibility into two categories; confidential and public as per the degree of hazard result of information. The effect of data protection is analyzed and authenticated on the subtle data within the framework of the classification group HDFS cluster.

Si Han et al. [16] proposes a Hidden community exchange key management protocol (SSGK) to avoid unwanted exposure to the contact channel and mutual data. Unlike previous books, the joint data is authenticated with a group key, and a secure distribution mechanism is used to spread the set of keys in SSGK. The detailed protection and efficiency analyses demonstrate that our collection of rules significantly minimizes data contribution's privacy security and hazards in cloud storage and saves

on twelve percentage of storage space.

Xiaodan Yan et al. [17] proposed the GAN Model Attackable of deep learning execution; this approach mainly studies the information protection methods under GAN model attack to find a better way to prevent attacks and effectively protect information. Sometimes medical treatment data may be leaked to third-party organizations. When these essential medical data are illegally used by for-profit organizations or obtained by criminals, it will lead to the disclosure of personal privacy information and cause severe economic losses to the victims. However, the victim cannot delete the leaked report by itself or limit the scope and use of the information that has been revealed.

Gamage Dumindu Samaraweera et al. [18] Proposed Protection and Privacy Effects on Big Data Age Database Systems: A study, his paper analyses protection applications in today's essential database models rely primarily on security and privacy attributes. A collection of standard protection measures is defined and tested based on different protection classifications. This offers a thorough overview and study of the sophistication of protection and privacy frameworks in the database models coupled with possible orders/enhancements. Data owners can agree on the maximum appropriate data storage for their data-driven Big data applications.

Yanan Li et al. [19] Propose a data correlation framework effect on privacy leakage, defined as Previous Differential Privacy (PDP), which is recommended to determine information leakage considering the opponent's specific prior awareness. The model uses two methods for evaluating discrete and continuous results, respectively, the weighted hierarchical graph and the multivariate Gaussian model. This further shows the distinct effect of optimistic, negative, and mixed associations on data leakage. A closed structure definition of privacy leak extracted used for unbroken data considering general associations, and a chain law is introduced for separate data.

Yang Liu et al. [20] Preservation of privacy and aggregation through reinforcement learning suggests a Payment-Privacy Protection Level (PPL) game in which each member submit their sensing data along with a given PPL as a system selects the appropriate payment for the participants. In addition to removing the Nash equilibrium (NE) stage of the game. Consider a payment-PPL system is ambiguous; it uses a reinforcement learning strategy, i.e., Q-learning, to get the payment-PPL method in the complicated payment-PPL game. Here it uses a deep Q network (DQN), which combines deep learning and Q-learning to speed up learning.

Gamage Dumindu Samaraweera et al. [21] Proposed defence design review of leading database models of today, with a greater focus on security and privacy attributes. A set of standard protections is specified and assessed based on specific classifications of security. It offers a thorough summary and systematic analysis of the complexity of security and privacy technologies in the database models, together with potential directions/improvements. Data owners may select the most proper data store for their data-driven Big Data technologies models.

.David Froelicher et al. [22] Drynx, a decentralized data management framework, has been proposed. knowledge of distributed databases as a subject of mathematical study. In order to compute data, such as standard deviation or severity, and to train and test machine-learning models on important and distributed data, the author relies on a network of computer nodes. To provide user anonymity and service provider security, Drynx uses collaborative protocols, homomorphic security, zero information evidence of validity, and differential protection. This ensures audit ability in a transparent adversarial setting where no individual involved needs to be independently dependable by enabling a practical and autonomous scrutiny of the entry data and the method computation.

DatThanh Dang et al. [23] a trust-based Map Reduce solution is suggested for tasks involving large data analysis. In order to minimize slots, we first quantify and propose to distribute the critical values for map data and trust values. Then we assess each tool's level of confidence in performing in-depth data analysis. A work requires a particular amount of faith depending on how vulnerable the data is; for example, low-importance data require servers and slots with a larger degree of confidence. The largest weighted comparable problem of a bipartite graph is devised for the Map Reduce scheduling challenge with the goal of optimising the cumulative confidence factor of all possible assignments according to certain confidence required activities.

Guangquan Guangquan Xu et al. [24] By analysing the data stream connecting the Java and native layers, a novel and adaptable framework called So Protector is proposed as a preventative measure for privacy leaks. A real-time approach for detecting harmful characteristics encoded in SO libraries is thus discovered by the guardian. Using three procedures, we place and extract the malware's features:

1. Presently as native family grayscale picture binary files.

2. Utilising Python to locate the op-code sequence and the ARM instructions package to reverse the SO file code.

3. All files are converted by IDA Pro into assembly language, which also includes.gdl files.

Dharminder et al. [25] introduces an identification-focused sign encryption technique that incorporates encryption and signature and provides a solution for secure and authentic communication in the Big Data environment, known as SFEEC (Safety Mechanism for Energy-Efficient Computing). The goals of less overhead computation and connectivity are met by SFEEC by offering pairing-free calculation at the end of the customer. SEC is also demonstrated to be "stable against chosen post" and "indistinguishable from chosen-ciphertext" under attacks.

Ruiyang Xiao et al. [26] Create a mixing scheme with a single shared signature protocol that is free from transaction costs and third parties. The approach makes use of a participant-supervised negotiating mechanism to maintain the agreement's transparency. The required system also includes a signature process that uses key sharing and the El Gamal signing process.

2. Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore

Privacy Preservation Techniques

The study of privacy preservation in machine learning is a field of study that is now quite popular. Getting the outcomes of a machine learning system without jeopardising the underlying dataset is what privacy preservation is all about. It is examined whether ML algorithms have any negative effects on data privacy. So, the primary goal of protecting privacy during data mining is to develop a method that transforms the original dataset in a way that ensures that even after the mining process, the private of the information is maintained.

Work Description:

The secure query processing for the various application development system getting with different techniques to search the relevant data from the database based on the features and its attribute range that are matched with the query input with encrypted data. To further improve the accuracy and the speed of process, this was enhanced by the cloud based data storage or by the parallel processing or with any other different techniques to reduce the time complexity and space complexity. Due to the size of key selection and the resultant encrypted data, this requires more storage space. Since, in this type of query processing based on the machine learning models, the training for the classification system of ML algorithm needs to improve the number of samples and with other parameter update. This increases the space complexity while the increase in training data for the further updates of features models. This also increases the time complexity for searching. To reduce this, compression techniques are used to store the features of data into the database with better compression rate. Since, the system should not compromise with the data security problem by reducing the key size for encrypting the data. The overall system needs to have the secure data processing with reduced storage space. To achieve this, the data can be encrypted by the lightweight cryptographic system which is to reduce the key size with high data security system. Along with this, the compression technique is to reduce the storage size.

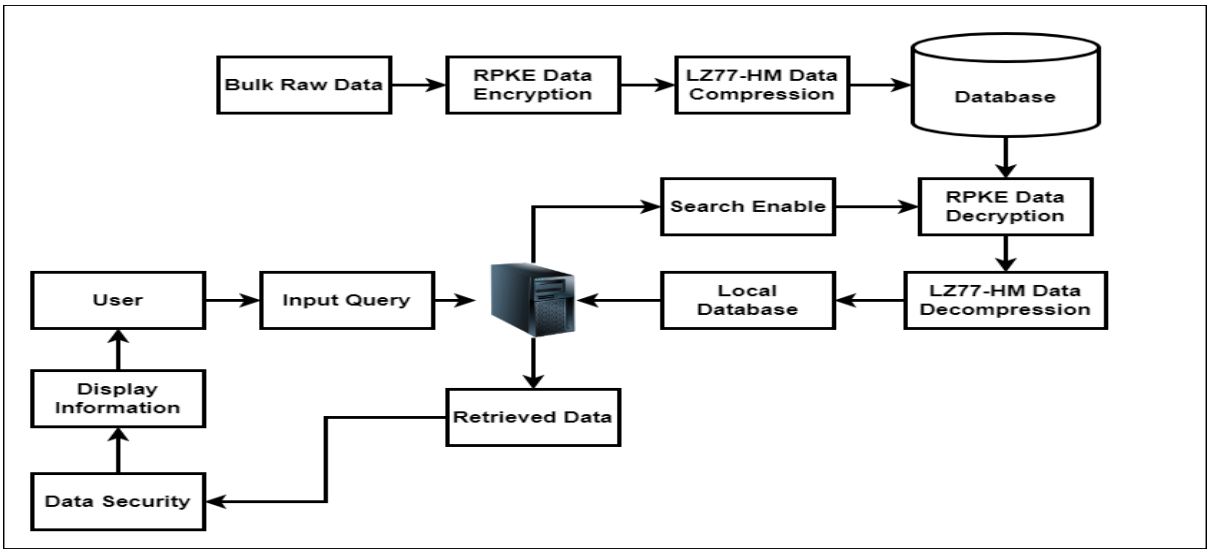


Figure 1: Flow Diagram.

In the compression model of secure query processing system, the cryptographic texture extraction model creates the grid of structure information in the overall data and making the attributes for each blocks to identify and retrieve the matching data from database. In that, the **Regressive Probabilistic Key Encryption (RPKE)** system based lightweight cryptographic model which can reduce the key size with better encryption of data. In this proposed work, the storage system for the database can be managed by using the compression model of data pattern after the encryption process of data to store in the database. This was achieved by the lossless compression technique based on the combination of **LZ77 with Huffman encoding** model. This overall system optimizes the size of data and the key size to reduce the space complexity and time complexity. The proposed method can be implement in the python scripting and validate the performance by using the parameters like, Compression Ratio, Reconstruction rate, database storage size, time complexity, security level, key size, and other related parameters with the reference of Ground truth of database.

Encryption Algorithm

- Start
- Input text data
- Generate permutation based key
- for i to number of words/character
    - convert to ASCII code
- end
- for i to number of ASCII code
    - cipher text=ASCII $\oplus$ key
- end
- for i to number of ASCII code
    - convert to character form
- end
- encrypted text

Results and discussions

3. Results

Using news-related datasets from the breaking news dataset, the suggested technique has been tested. Then, each dataset is divided (80-20%) into training and test subsets.). Four metrics, depending on the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) in the predictions of the binary classifiers, have been used to evaluate the results:

Accuracy: The capacity of the framework to precisely characterize information depends to a vast degree on the illustrations that you give.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \dots (1)$$

Precision also called positive predicted value is the fraction of significant instances among the retrieved instances.

$$Precision = \frac{TP}{TP + FP} \dots \quad (2)$$

Recall also known as sensitivity or True Positive Rate (TPR) is the fraction of significant instances that have been retrieved over the total amount of relevant instances.

$$Recall = \frac{TP}{TP + FN} \dots \quad (3)$$

F1 score also F-score or F-measure is a measure of a test's accuracy for binary classification.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \dots (4)$$

True Negative Rate or Specificity is defined as follows:

$$\text{Specificity} = \frac{TP}{TP + FN} \dots \quad (5)$$

```
            Proposed Algorithm
Accuracy                83.800000
Precision               83.808101
Recall                  83.800000
FScore                  83.797753
==================================================
Input query : marco rubio cruz both good nights bush
==================================================
Encrypted query : =1"3?p"%29?p3"%*p2?$8p7??4p>978$#p2%#
==================================================
Encrypted Retrieve Information : ▯w▯▯▯▯▯▯▯w
==================================================
 Retrieve Information : ['POLITICS']
Compression time: 16.6330947875976 56 seconds
```

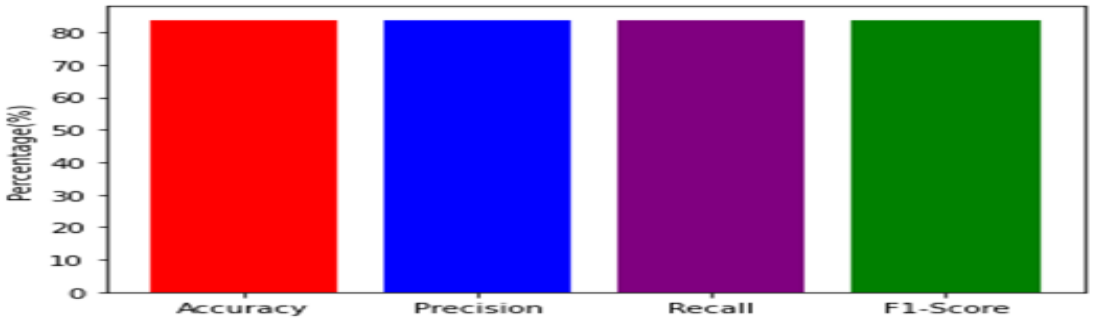
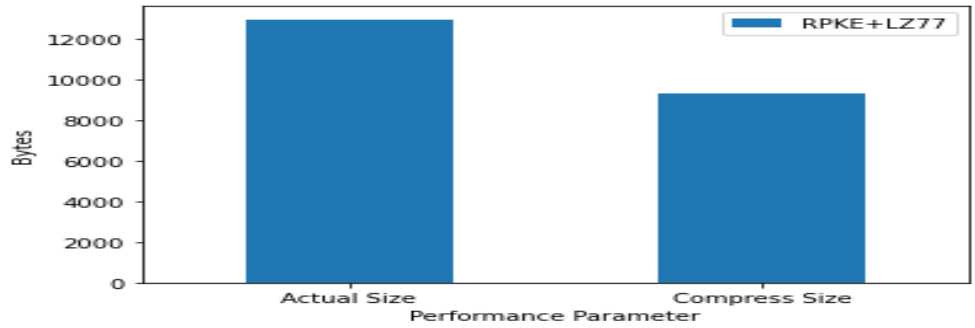


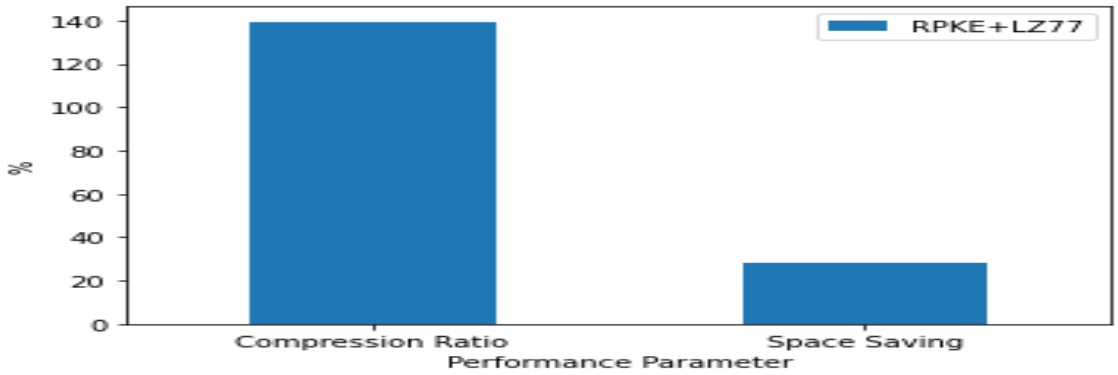
Figure 2: proposed algorithm and performance

```
File was compressed successfully and saved to output path ...
File was decompressed successfully and saved to output path ...
Actual Document Size: 12959 Byte
Compress Document Size: 9292 Byte
Compression Ratio : 1.3946405510116229
Space Saving : 0.2829693649201327
Compression time: 26.92353653907776 seconds
Decompression time: 7.122612476348877 seconds
                RPKE+LZ77
Actual Size         12959
Compress Size        9292
```




```
                  RPKE+LZ77
Compression Ratio  139.464055
Space Saving        28.296936
```

Figure 3: Performance Parameter for compression




```
                    RPKE+LZ77
Compression time    26.923537
Decompression time   7.122612
```

Figure 4: Performance Parameter for Decompression

Discussion

To expedite SQL query execution, this study proposes an enhancement to the basic database encryption strategy. Experimental findings indicate that the adaptive technique significantly amplifies query performance across various record ranges, consistently delivering exceptional query response times. Moreover, the results demonstrate that the suggested methodology exhibits progressive improvements over time concerning response time, compression, and decompression performance.

Refrences

[1] Bouganim, L. and Y. Guo, 2011. Database encryption. Proceedings of the Encyclopedia of Cryptography and Security, Boston, MA: Springer US, pp: 307-12. DOI: 10.1007/978-1-4419-5906-5_677.

[2] Salama, D., A. Elminaam, H. Mohamed, A. Kader and M.M. Hadhoud, 2010. Evaluating the performance of symmetric encryption algorithms. Int. J. Network Security, 10: 213-19.

[3] Sharma, M., A. Chaudhary and S. Kumar, 2013. Query processing performance and searching over encrypted data by using an efficient algorithm. Int. J. Comput. Appli., 62: 975-8887.

[4] Alhanjouri, M. and A.M. Al Derawi, 2012. A New method of query over encrypted data in database using hash map. Int. J. Comp. Appl., 41: 975-888.

[5] Mousa, A., E. Nigm, S. El-Rabaie and O. Faragallah, 2012. Query processing performance on encrypted databases by using the REA algorithm. Int. J. Network Security, 14: 280-88.

[6] Zheng-Fei, W., W. Wang and B.L. Shi, 2005. 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.

[7] Casino F, Patsakis C. An Efficient Blockchain-Based Privacy-Preserving Collaborative Filtering Architecture.IEEE Transactions on Engineering Management.2019 Oct 22.

[8] Chenyang Ma C, Wang B, Jooste K, Zhang Z, Ping Y. Practical Privacy-Preserving Frequent Itemset Mining on Supermarket Transactions. IEEE Systems Journal.2019 Jun 26.

[9] Dong Fang D, Qian Y, Hu RQ. A Flexible and Efficient Authentication and Secure Data Transmission Scheme for IoT Applications, IEEE Internet of Things Journal.2020 Feb 3.

[10] Anmin Fu A, Chen Z, Mu Y, Susilo W, Sun Y, Wu J. Cloud-based Outsourcing for Enabling Privacy-Preserving Large-scale Non-Negative Matrix Factorization. IEEE Transactions on Services Computing.2019 Aug 28.

[11] Muhammad Usman M, Jolfaei A, Jan MA. RaSEC: An Intelligent Framework for Reliable and Secure Multi-Level Edge Computing in Industrial Environments. IEEE Transactions on Industry Applications.2020 Feb 20.

[12] Jiannan Wei J, Phuong TV, Yang G. An Efficient Privacy-Preserving Message Authentication Scheme for Internet-of-Things.IEEE Transactions on Industrial Informatics.2020 Feb 10.

[13] Rong Jiang R, Shi M, Zhou W. A Privacy Security Risk Analysis Method for Medical Big Data in Urban Computing. IEEE Access. 2019 Sep 24; 7:143841-54.

[14] Karen R. Sollins KR. IoT big data security and privacy versus innovation. IEEE Internet of Things Journal. 2019 Feb 15;6(2):1628-35.

[15] Ismail Hababeh I, Gharaibeh A, Nofal S, Khalil I. An integrated methodology for big data classification and security for improving cloud systems data mobility. IEEE Access. 2018 Dec 28; 7:9153-63.

[16] Si Han S, Han K, Zhang S. A Data Sharing Protocol to Minimize Security and Privacy Risks of Cloud Storage in Big Data Era. IEEE Access. 2019 May 3; 7:60290

[17] Xiaodan Shuo Zhang, Yaping Liu, Shudong Li, Zhiyuan Tan, Xiaomeng Zhao, Junjie Zhou, "FIMPA: A Fixed Identity Mapping Prediction Algorithm in Edge Computing Environment", Access IEEE, vol. 8, pp. 17356-17365, 2020.

[18] Gamage Dumindu Samaraweera GD, Chang MJ. Security and Privacy Implications on Database Systems in Big Data Era: A Survey. IEEE Transactions on Knowledge and Data Engineering.2019 Jul 18.

[19] Yanan Y. Li, X. Ren, S. Yang and X. Yang, "Impact of Prior Knowledge, and Data Correlation on Privacy Leakage: A Unified Analysis," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 9, pp. 2342-2357, Sept. 2019.

[20] Yang Liu Y, Wang H, Peng M, Guan J, Xu J, Wang Y. DeePGA: A Privacy-Preserving Data Aggregation Game in Crowdsensing via Deep Reinforcement Learning. IEEE Internet of Things Journal.2019 Dec 3.

[21] Gamage Dumindu Samaraweera GD, Chang MJ. Security and Privacy Implications on Database Systems in Big Data Era: A Survey. IEEE Transactions on Knowledge and Data Engineering.2019 Jul 18.

[22] David Froelicher D, Troncoso-Pastoriza JR, Sousa JS, Hubaux JP. Drynx: Decentralized, Secure, Verifiable System for Statistical Queries, and Machine Learning on Distributed Datasets.

[23] DatThanh Guangquan Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799.

[24] Dang DT, Hoang D, Nguyen D. Trust-based Scheduling Framework for Big Data Processing with MapReduce. IEEE Transactions on Services Computing.2019 Sep 3.

[25] D. Dharminder, M. S. Obaidat, D. Mishra, and A. K. Das, "SFEEC: Provably Secure Signcryption-Based Big Data Security Framework for Energy-Efficient Computing Environment," in IEEE Systems Journal March 2020.

[26] Ruiyang Xiao et al Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.