A Comparative Study of Encryption Methods for Cloud Query Processing

Mo’ath Naser Magableh and Basel Alshaikhdeeb

1Department of Business Administration, Lincoln University, Oakland, California 94612, USA
2Department of Strategic Information Systems, National University of Malaysia, Bangi, Selangor 43600, Malaysia

Abstract: Query cloud process is an interested research study that caught many researchers’ attentions. Several studies have presented different types of encryption in order to encrypt the data prior to being migrated over the cloud. However, there is an essential demand to balance between the time consumption and encryption security. This paper presented a comparative study of encryption methods for query execution over the cloud. Three common encryption methods have been used including Advanced Encryption Standard (AES), Rivest–Shamir–Adleman (RSA) and Elliptic Curve Cryptography (ECC). A benchmark dataset of queries has been used in the experiments. Based on the time of encryption and decryption along with the secrecy measure, the three methods have been evaluated. Results showed that RSA has the most competitive performance in terms of encryption and decryption time, meanwhile, it has a competitive secrecy measure values. It achieved an average encryption time of 0.57, 1.41 and 0.59 for Delete, Add and Select queries, as well as, it achieved an average decryption time of 2.31, 4.24 and 1.79 for Delete, Add and Select queries. Finally, RSA obtained an average secrecy of 1.10, 1.10 and 1.15 for Delete, Add and Select queries. This emphasis the usefulness of using RSA to maintain both efficiency and security of encryption.

Keywords: Cloud Computing, Query Processing, Encryption Methods, Encryption Time, Encryption Secrecy

Introduction

The last decade has witnessed a growth in the investment of cloud computing technologies where large corporations nowadays are relying on pay-per-usage services rather than creating such services from the scratch (Dillon et al., 2010). Consider a service such as a database editor in which an organization can create, add, update and delete its own data. It is obvious that such a service would require a huge amount of resources including a platform and servers with tremendous capabilities along with database developers and administrators. In addition, this database would indeed require maintenance chronically. Apparently, all these requirements pose a huge increment in the expenses (Zhang et al., 2010). Therefore, organizations recently take the advantage of cloud computing technologies by renting the database service where they can pay per their usage without affording any other expenses such as developing functions, maintenance and other issues (Chang et al., 2008; Cooper et al., 2008).

However, this poses a challenging issue represented by the privacy that could be violated by the cloud service providers (Sammour, 2018). In some industries such as the medical, violating privacy is considered to be intolerant (Curino et al., 2011). Thus, many research studies have addressed this problem by proposing various types of encryption methods. This means that organizations should encrypt their data prior to being migrated over the cloud. Consequently, any query typed after the migration could require both encrypting the query and decrypting the results of such query (Al Shehri, 2013).

Basically, such encryption would significantly impact the efficiency where the query execution time would be longer with such encryption and decryption tasks (Hacigümüş et al., 2002). Therefore, some authors have proposed a classification method prior to the encryption...
in order to clarify which data is considered confidential to be encrypted. Such a classification task is intended to reduce the load of encryption (Albadri and Sulaiman, 2016). Yet, the encryption methods used by such studies still suffer from the trade-off between providing an efficient and secure query execution. Some encryption methods provide high security yet, require longer time in the query execution. Other methods provide efficient query execution yet, it has less secure encryption. Therefore, this paper aims to provide a comparison regarding security and efficiency of query execution over the cloud. A comparison will be accommodated among three common encryption methods including Advanced Encryption Standard (AES), Rivest–Shamir–Adleman (RSA) and Elliptic Curve Cryptography (ECC).

The rest of the paper organized as:

- Related work: Discusses the recent approaches that have been proposed for efficient query execution over the cloud
- The proposed trade-off encryption method: Discusses the proposed encryption method that is resulted from a comparison of three encryption methods
- Experimental results: Highlights the results obtained by the proposed method
- Conclusion: Provides a final summary along with the future directions

Related Work

The literature review has shown numerous concerns regarding the search and query execution over the cloud. For example, Wang et al. (2012) have addressed the problem of providing a secure and efficient search mechanism over the cloud. The authors have examined the difficulty of getting accurate results when accommodate searching over a cloud database. Such difficulty comes from the encrypted data. In this regard, the authors have examined different approaches that are intended to search encrypted data. Yet, most of these approaches are based on a binary search where the result is either found or not. Hence, the authors have presented a ranking technique that has the ability to bring more accurate search results. Their proposed technique was based on term frequency and mutual information. Experimental results showed an enhancement in terms of search result accuracy.

On the other hand, Ren et al. (2013) examined the problem of providing secure and efficient query execution over the cloud. Their proposed method was based on Random Space Perturbation (RASP). Such method was intended to offer random noise in order to provide maximum privacy. Additionally, the authors have used the K-nearest neighbor classification method in order to index the data for improving the retrieval.

Despite the search over the cloud, some research studies have addressed the problem of impacting the efficiency by using overload encryption. For example, Graepel et al. (2012) presented a classification method based on machine learning techniques for categorizing the data into confidential and non-confidential. Such classification was intended to reduce the load of encryption where the encryption process is applied only upon the confidential data.

In the same regard, Zardari et al. (2013) addressed the problem of violating confidential data over the cloud, meanwhile, preserving efficiency. The authors have used a classification technique in order to categorize the data into sensitive and non-sensitive data. Based on such categorization, the encryption will be applied when needed.

Finally, Albadri and Sulaiman (2016) have provided a rule-based classification method for categorizing the data into sensitive and non-sensitive before accommodating the encryption. Within the encryption, the authors have utilized the Advanced Encryption Standard (AES) method for conducting the encryption of queries. Results showed that the query execution time has been reduced based on such categorization process.

Li et al. (2017) have addressed the problem of completeness when querying encrypted data over the cloud. Due to the encryption, sometime the search results or the response results to a particular query seem to be incomplete. Therefore, the authors have proposed a verifiable range query processing scheme by enabling the user to verify the completeness. Their proposed scheme was relying on adding additional information to a complete binary tree in order to verify the indexed elements.

Similarly, Li and Liu (2017) have addressed the problem of integrity and efficiency of querying encrypted data over the cloud. The authors have proposed an Indistinguishable Bloom Filter (IBF) structure in order to index the element. Additionally, the authors have proposed a balanced binary tree also known as indistinguishable Binary Tree (IBtree) in order to insure an efficient query processing.

Sahin et al. (2018) have addressed the problem of accommodating non-aggregate range queries over the cloud be while preserving both efficiency and security. For this purpose, the authors have proposed a differentially private index method for such type of queries. The authors, have intended to enable cleartext index structure for conducting the range queries. After that, the differential method will be applied on both the index and response of the encrypted database.

Examining the literature, one could notice that the prior categorization of data has facilitated toward improving efficiency. However, the encryption methods used are still suffering from the trade-off between
providing optimal privacy, meanwhile, preserving the efficiency. Therefore, this study different encryption techniques in order to find out the most suitable approach that can satisfy the trade-off problem.

**The Proposed Trade-off Encryption Method**

As shown in Fig. 1, the proposed method is composed of multiple phases including queries, encrypted data, encryption methods, evaluation and comparison.

The first and second phases are related to the dataset used in the experiment where it consists of encrypted database and a set of queries. While the third phase is intended to accommodate the encryption over the queries using three encryption methods including AES, RSA and ECC. Finally, the comparison phase the most appropriate encryption method that satisfies the problem of balancing between security and efficiency. In fact, the following subsections will tackle each phase independently.

**Encrypted Database and Queries**

This phase is intended to discuss the queries used in the experiments. In order to facilitate the comparison against the state of art, this paper has used the dataset introduced in (Albadri and Sulaiman, 2016). Such dataset was brought from a university database where some information is sensitive and others are not. The data has been annotated in terms of sensitivity based on four class labels; sensitive, confidential, internal and public. The first class label is composed of basic information about the students including date of birth and median name. The second class label is composed of restricted information that should not be violated such as the successful payments, overdue payments and course assessment. The third class label is composed of the authorized-use information that is being used by the staff of the university including student's progress report. Finally, the fourth class label is composed of the public information that is tolerable to be used or viewed by anyone. Table 1 depicts the details of the dataset.

Table 2 shows a sample of the dataset where the fields and its corresponding classes are depicted.

On the other hand, the dataset consists of a set of queries that will be used to process the encrypted data. Table 3 depicts the details of the query set.

![Fig. 1: The proposed trade-off encryption](image-url)
Encryption

This phase aims to investigate three common encryption methods for the purpose of encrypting the queries. Such encryption methods are composed of AES, RSA and ECC. These encryption methods are being discussed as follows:

**Advanced Encryption Standard (AES)**

This encryption is one of the common methods that have been widely used due to its high efficiency where it has faster performance compared to other methods (Daemen and Rijmen, 2013). It has four stages to be performed including key expansion, SubByte, shift rows and mix columns. The first stage aims to add a round key for a plain text by using the XOR operator as shown in Fig. 2.

The second stage is intended to perform the SubByte where the s-box (shown in Fig. 3) is being used to replace the corresponding items.

The third stage is intended to shift the rows of the resulted state from the previous stage (i.e., SubByte). Such shifting of rows will avoid the first row, shift one item of the second row, shift two items of the third row and finally shift three items of the fourth row as shown in Fig. 4.

<table>
<thead>
<tr>
<th>Table 1: Details of the dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
</tr>
<tr>
<td>Number of tables</td>
</tr>
<tr>
<td>Number of fields</td>
</tr>
<tr>
<td>Number of classes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Sample of dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table</td>
</tr>
<tr>
<td>pass_word</td>
</tr>
<tr>
<td>pass_word</td>
</tr>
<tr>
<td>category</td>
</tr>
<tr>
<td>category</td>
</tr>
<tr>
<td>category</td>
</tr>
<tr>
<td>Religion</td>
</tr>
<tr>
<td>Religion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3: Sample of query set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query set</td>
</tr>
<tr>
<td>Set 1</td>
</tr>
<tr>
<td>Set 2</td>
</tr>
<tr>
<td>Set 3</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

![Fig. 2: Key expansion](image1.png)

![Fig. 3: SubByte](image2.png)
The final stage is the mix columns which is intended to perform a transformation process on the column level. This means that every column will be multiplied by the row of the state's inverse as shown in Fig. 5.

**Rivest–Shamir–Adleman (RSA)**

RSA is one of the encryption methods that use public key and private key in which the public is being used for the encryption and the private is used for the decryption (Zhou and Tang, 2011). It has three stages including key generation, message encryption and message decryption. The first stage is intended to generate the key where both the public and private are being created. To do so, RSA requires identifying two distinct prime numbers for both \( p \) and \( q \). Then, \( n \) will be computed based on the following equation:

\[
n = p \times q
\]  

After that, the Euler \( \varphi \) should be calculated based on the following equation:

\[
\varphi = (p - 1) \times (q - 1)
\]

Hence, a value between 1 and \( \varphi \) will be selected as \( e \) where \( e \) is a number that is not divisible by \( \varphi \). Now, the second stage which is the encryption can be applied using the two keys \( n \) and \( e \) based on the following equation:

\[
Ciphertext = Message^e \pmod{n}
\]  

The third stage is the decryption which can be conducted using the following equation:

\[
Message = Ciphertext^d \pmod{n}
\]

where, \( d \) is the third prime number that is being used by RSA and it can be computed as:

\[
d = e^{-1} \pmod{\varphi}
\]

**Elliptic Curve Cryptography (ECC)**

Similar to RSA, ECC has a public and private key encryption paradigm where the public is used for the encryption and the private is used for the decryption. However, the key distinction between the two methods lies in the capability of ECC to make the guessing of the private key is a way more difficult (Kapoor, 2008). Consider an elliptic curve such as the one in Fig. 6, the possibility of assuming a private key can be expressed as \( A \rightarrow \text{max} \) which reflects \( n \) the key size. Apparently, this would make the possibilities are varied which emphasizes the sophisticated mechanism of finding the private key compared to the RSA which is based on finding three prime numbers.

As shown in Fig. 6, the elliptic curve is symmetric about the x-axis, this due to the equation of ECC is represented as follows:

\[
y^2 = x^3 + ax + b
\]
where, \(4a^3 + 27b^2 \neq 0\), in this case, \(y \) would be resulted as \(y = \pm i\) therefore, the curve of ECC must be symmetric around x-axis.

In order to add two points to be represented on a curve, several approaches have been proposed including projective system, Jacobian system and Lopez Dahab system. The projective system aims to represent the points by three coordinates \((X, Y, Z)\) using the relation of \(x = \frac{X}{Z}\) and \(y = \frac{Y}{Z}\). This can be represented in the following equation:

\[
E \Rightarrow \left(\frac{Y}{Z}\right)^2 + a_i \left(\frac{X}{Z} \left(\frac{Y}{Z}\right) + a_i \left(\frac{Y}{Z}\right)
\right)
\]

\[
= \left(\frac{x}{z}\right)^3 + a_i \left(\frac{x}{z}\right)^2 + a_i \left(\frac{y}{z}\right) + a_i
\]  

Or:

\[
E \Rightarrow y^2z^{-2} + a_i yz^{-2} + a_i y^{-1} 
\]

\[
= x^3 z^{-3} + a_i x^2 z^{-2} + a_i y z^{-1} + a_i
\]

Multiply by \(z^3\):

\[
E \Rightarrow y^2z^{-2} x^3 z^{-3} + a_i yz^{-2} x^2 z^{-2} + a_i y^{-1} x z^{-1} + a_i z
\]

The above mentioned equations can be applied for Lopez Dahab system where the relation is \(x = \frac{X}{Z}\) and \(y = \frac{Y}{Z^2}\), or Jacobian system where the relation is \(x = \frac{X}{Z^2}\) and \(y = \frac{Y}{Z^2}\).

**Comparison**

After applying the three encryption methods, this phase comes to examine both security and efficiency of each encryption method in order to identify the most appropriate method that satisfies the problem of balancing between security and efficiency.

In order to assess the efficiency of encryption, it is possible to consider the execution time of the queries based on such encryption.

However, assessing the security of an encryption method is a challenging task where several research studies have proposed different measurement tools for this purpose. Yet, the most acceptable measure used for evaluating the security is the secrecy (Singh and Maini, 2011). In order to compute the secrecy of encryption, multiple variables should be considered including entropy, uncertainty and equivocation. First, the entropy indicates the extent of information lies on a message \(x\) where the entropy of a message \(H(x)\) refers to the minimum number of bits required to encrypt all potential meanings of such message. It can be denoted by:

\[
H(X) = -\sum_{1 \leq i \leq n} p(x_i) \log p(x_i)
\]

(10)

On the other hand, uncertainty refers to the number of bits required to decrypt a ciphertext \(y\) which is the reverse process of the entropy thus, it is denoted similarly as in the following equation:

\[
H(Y) = -\sum_{1 \leq i \leq n} p(y_i) \log p(y_i)
\]

(11)

Finally, the equivocation is considered to be the uncertainty of a message \(x\) that would be reduced by given additional information. It is the conditional entropy of \(x\) given \(y\) which can be denoted by:

\[
H_x(Y) = -\sum_{X, Y} p(X, Y) \log P(Y|X)
= -\sum_Y p(Y) \sum_{X} p(X | Y) \log_2 \left(\frac{P(X | Y)}{P(Y)}\right)
\]

(12)

After computing all the variables, now the secrecy can be expressed as the key equivocation \(H_e(k)\) of ciphertext \(C\) with a key \(K\), meanwhile, it is the amount of uncertainty in \(K\) given \(C\) which can be denoted as:

\[
H_e(K) = -\sum C p(C) \sum K p(K | C) \log_2 \left[\frac{p(K | C)}{p(K)}\right]
\]

(13)

**Results**

This section will highlight the results of the three encryption methods based on query execution time and secrecy. For this purpose, 150 queries brought from the...
A dataset will be used. These queries are divided into Add queries, Select queries and Delete queries each of which consists of 50 queries. Fig. 7 shows both encryption and decryption time for the Delete queries.

As shown in Fig. 7 both AES and RSA have shown lesser time consumption for both encryption and decryption compared to the ECC. In some queries, AES has shown lesser time than the RSA for both encryption and decryption. However, they still have similar performance.

Figure 8 shows the encryption and decryption time using for Add queries using the three methods.

As shown in Fig. 8, similar to the Delete queries, Add queries have shown lesser encryption and decryption time consumption by AES and RSA compared to the ECC. However, majority of Delete queries have shown superiority for the AES which outperformed the RSA within multiple queries in terms of encryption and decryption time consumption.

Figure 9 shows the time consumption for both encryption and decryption of the Select queries using the three methods.
As shown in Fig. 9 and similar to the previous two set of queries, Delete queries have shown lesser encryption and decryption time consumption for both AES and RSA compared to the ECC. However, AES is still outperforming RSA within some queries in terms of time consumption.

Apart from the encryption and decryption time, Fig. 10 shows the secrecy values for all the queries using the three methods.

As shown in Fig. 10, for the Delete queries, ECC has shown the most secure encryption where the secrecy values for all queries were higher than the ones in both RSA and AES. This has been followed by RSA were the secrecy values for all queries were higher than the ones in AES.

Similarly, for the Add queries, ECC has outperformed AES and RSA by obtaining highest secrecy values for all the queries. Once again, RSA has outperformed AES by attaining higher values of secrecy.

Finally, for the Select queries, ECC was the superior encryption method in terms of secrecy values for all queries and followed by RSA. This means that AES has the lowest secure encryption results.

However, Table 4 depicts the average encryption and decryption time along with the average secrecy for the three methods in accordance to all queries.

As shown in Table 4, the average encryption time for AES, RSA and ECC were 0.64, 0.57 and 2.21 respectively based on the Delete queries. While for the Add queries, the average was 1.49, 1.41 and 2.97 for AES, RSA and ECC. Finally, for the Select query, the average was 0.79, 0.59 and 2.04 for AES, RSA and ECC. Generally, it is obvious that the Add query has the greatest time consumption in terms of encryption.

<table>
<thead>
<tr>
<th>Query type</th>
<th>Avg. encryption time (Seconds)</th>
<th>Avg. decryption time (Seconds)</th>
<th>Avg. secrecy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete</td>
<td>AES: 0.643968 RSA: 0.577902 ECC: 2.216329</td>
<td>AES: 1.931904 RSA: 2.311961 ECC: 13.29797</td>
<td>AES: 0.921687 RSA: 1.104628 ECC: 1.115608</td>
</tr>
<tr>
<td>Add</td>
<td>AES: 1.494973 RSA: 1.413546 ECC: 2.973864</td>
<td>AES: 2.989945 RSA: 4.240639 ECC: 11.89545</td>
<td>AES: 0.912348 RSA: 1.102296 ECC: 1.114341</td>
</tr>
<tr>
<td>Select</td>
<td>AES: 0.797675 RSA: 0.59891 ECC: 2.040183</td>
<td>AES: 1.79673 RSA: 8.160732</td>
<td>AES: 0.967783 RSA: 1.158509 ECC: 1.176321</td>
</tr>
</tbody>
</table>

Table 4: Average of encryption time, decryption time and secrecy.
This is typical since the Add query usually tends to be larger in terms of text. On the other hand, RSA has consumed the lowest time of encryption, while ECC has consumed the most.

Apart from the encryption, the decryption has shown greater consumption of encryption time for all the three methods where the Delete queries show 1.93, 2.31 and 13.29 for AES, RSA and ECC. In addition, the Add queries show 2.98, 4.24 and 11.89 of encryption time for AES, RSA and ECC. Finally, the Select queries show 1.59, 1.79 and 8.16 for AES, RSA and ECC. Such greatness in time within the decryption compared to the encryption is typical since the resulted data is usually larger than the query. However, the lowest consumption of time was achieved by AES. Although RSA showed relatively similar consumption yet, it is known that both RSA and ECC make the identification of private key for the decryption is very difficult. Therefore, most time consumption was achieved by ECC.

Eventually, the secrecy, which is the measure that shows how strong is the encryption, has revealed values for AES, RSA and ECC as 0.92, 1.10 and 1.11 for the Delete queries. While for the Add queries, the results were 0.91, 1.10 and 1.11. Finally, for the Select queries, the results were 0.96, 1.15 and 1.17. Obviously, ECC has the most secure encryption for all the query types. This is due to its sophisticated mechanism to identify the private key. Yet, RSA is also showed a relatively secure encryption even though, it was lower than the ECC. Apparently, AES has the least secure encryption according to its value of secrecy.

In order to find the trade-off encryption method, it is necessary to consider all the results for encryption time, decryption time and secrecy. It is obvious that RSA represents the trade-off encryption method since it has the lowest consumption in terms of encryption. In terms of decryption, RSA showed a competitive performance even though it was not the most efficient. Finally, in terms of security, RAS showed the second secured encryption. Considering the ECC as the most secure encryption, it has the least efficient performance in terms of encryption and decryption times.

Conclusion

This paper presented a comparative study of encryption methods for the task of query processing over the cloud. Three encryption methods have been examined including AES, RSA and ECC. A benchmark dataset has been used the experiments. Based on encryption time, decryption time and secrecy measure, the three methods have been evaluated. Results showed that RSA represents the most accurate trade-off encryption method since it has competitive time consumption in both encryption and decryption, meanwhile, it has also a competitive secrecy value. This demonstrates the usefulness of RSA in terms of efficiency and security.

For future researches, considering different combinations of the three encryption method may yield promising results in terms of efficiency and security. Such combinations would be in parallel mode where multiple encryption methods would be applied at once or in sequential mode where a specific encryption method would encrypt part of the text and another encryption method would encrypt the rest of the text.

Acknowledgment

I would like to thank Dr. Basel for his collaboration.

Author’s Contributions

Mo’ath Naser Magableh: Review the literature, identified the research gap, proposed the potential solution.

Basel Alshaikhdeeb: Implementing the proposed solution by identifying the dataset and applying the method.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References


