Retrieving Encrypted Query from Encrypted Database Depending on Symmetric Encrypted Cipher System Method

Dr. Ghassan H. Abdul-Majeed [*], Dr. Alaa Kadhim F. [**], Rasha Subhi Ali [**]

[*] Ministry of higher education, Baghdad, Iraq

[**] Computer Sciences Department, University of technology, Baghdad, Iraq

Received 22 November 2015 ; Accepted 12 April 2016

Abstract

More and more data is available on database every day. The greater the amount of data in database led to create a problem in process and retrieving the required data. Security is one of the significant challenges that people are faced over the entire world in every aspect of their lives. Databases are vulnerable to attack from internal and external threats. One of security dialogues is data encryption/decryption whenever data being transmitted over communication lines may be protected by encrypting the data, which can be decrypted only by the authorized person. The retrieval from big encrypted database stills a big problem. The proposed system presents a new method used to retrieve data from (encrypted database; encrypted compressed database or encrypted dynamic clusters). These data retrieved represents the answers to the user query. In this research the retrieving from big encrypted database was processed by matching cipher query with encrypted database. The proposed system uses clustering technique to build block of data according to the encrypted user query (entries or requirements). The comparison includes the retrieving time that was required from matching plain query with plain data and cipher query with cipher data. In traditional system the retrieving was done by decrypting the entire database or decrypting part of it to find the data that matched the user query. This would be consumed too large time. The work of this paper allows to the users to query over encrypted database without decrypting the database, instead of that, it work on comparing cipher query with encrypted database and retrieving the results in cipher form. The data retrieval process is considered the main objective of this research and not the encryption process. So, the simple encryption operation was used to measure the performance of the data retrieving method (by matching encrypted query with encrypted data).

Retrieving Encrypted Query from Encrypted Database Depending on Symmetric Encrypted Cipher System Method

Ghassan H. Abdul-Majeed Alaa Kadhim F. Rasha Subhi Ali

Keywords: Clustering, ICM, Encryption, Decryption, Information Retrieval, Matching cipher query with cipher data.

استرجاع الاستعلام المشفر من قاعدة البيانات المشفرة بالاعتماد على طريقة تشفير متماثلة

أ.د غسان حميد عبد المجيد *، أ.م.د علاء كاظم فرحان **، الطالبة: رشا صبحي علي **

* دائرة البحث و التطوير - وزارة التعليم العالي والبحث العلمي

** قسم علوم الحاسوب – الجامعة التكنولوجية

الخلاصة

في كل يوم يتم توفير المزيد من البيانات المتاحة على قاعدة البيانات. وكلما زادت كمية البيانات في قاعدة بيانات يؤدي إلى خلق مشكلة في عملية واسترجاع البيانات المطلوب من قواعد البيانات الكبيرة .و الأمن هو أحد التحديات الهامة التي تواجه الناس على العالم بأسره في كل جانب من جوانب حياتهم .قواعد البيانات تكن عرضة للهجوم من التهديدات الداخلية والخارجية .واحد اهم مجالات الأمن هو تشفير البيانات / فك التشفير و البيانات التي يتم إرسالها عبر خطوط الاتصالات قد تكون محمية من خلال تشفير البيانات، والتي يمكن فك تشفيرها فقط من قبل الشخص المخول .و الاسترجاع من قاعدة بيانات مشفرة كبيرة يبقى مشكلة كبيرة . والنظام المقترح يعرض طريقة جديدة لحل هذه المشكلة تستخدم هذه الطريقة لاسترداد البيانات من (قاعدة بيانات مشفرة، قاعدة بيانات مضغوطة ومشفرة والعناقيد المشفرة والمبنية ديناميكيا). هذه البيانات التي تم استردادها تمثل إجابات لاستعلام المستخدم. في هذا البحث تم معالجة الاسترجاع من قاعدة بيانات كبيرة ومشفرة عن طريق مطابقة الاستعلام المشفر مع قاعدة بيانات مشفرة. ويستخدم النظام المقترح تقنية التجميع لبناء كتلة من البيانات وفقا للاستعلام المستخدم المشفرة(إدخالات أو متطلبات). وتشمل المقارنة الوقت الاسترجاع المطلوب من مطابقة استعلام صريح مع بيانات صريحة والاستعلام المشفر مع البيانات المشفر. في النظام التقليدي الذي تجرى فيه عملية استرجاع عن طريق فك تشفير قاعدة البيانات بأكملها أو فك تشفير جزء منه للعثور على البيانات التي تطابق استعلام المستخدم .وهذا من شأنه أن تستهلك وقت كبير جدا .العمل هذا البحث يسمح للمستخدمين بالاستعلام من قاعدة بيانات مشفرة دون فك تشفير قاعدة البيانات، وبدلا من ذلك، فإنه يعمل على مقارنة الاستعلام المشفر مع قاعدة بيانات مشفرة واسترجاع النتائج في شكل مشفر. و تعتبر عملية استرجاع البيانات هي الهدف الرئيسي من هذا البحث وليس عملية التشفير .لذلك، تم استخدام عملية تشفير بسيطة لقياس أداء طريقة استرجاع البيانات (عن طريق مطابقة الاستعلام المشفرة مع البيانات المشفرة).

الكلمات المفتاحية: العنقدة ، ICM ، التشفير ، فك التشفير ، مطابقة استعلام مشفر مع قاعدة بيانات مشفرة.

Introduction

Big Data concerns with large-volume, complicated, growing data sets in multiple and independent sources. Data storage and Data collection has become more complex. Big Data is now expanding quickly in all science and engineering domains including physical, biological and biomedical sciences, etc. The most essential challenge for Big Data applications is to search the large amount of data and extract useful information or knowledge for future works [1]. Data Retrieval encompasses extracting the desired data from a database. The two primary forms of the retrieved data are reports and queries. To retrieve the desired data the user offer a set of criteria by a query. The ability to query and retrieve data based on some user defined criteria is a necessary feature of the data storage and retrieval subsystem [2].

The retrieved data may be stored in a file, printed, or viewed on the screen. In traditional database management systems, information retrieval is often performed using keywords contained within fields of each record [3]. So, for faster retrieving the compression methods were used to compress the database files and the dynamic clustering method was used to build clusters in dynamic way, these clusters contain information about the required query data, so the retrieving data become much faster than the retrieving from original and compressed files. The retrieving methods included restoring from plain files and restoring from encrypted files. The study of cryptography has always had interesting research area. It is already known that security of data is the primary interest in the public network. Encryption and decryption is the process of cryptography technique which should be provided secrecy of the data over the network. In the real world there are so many organizations working on large databases over a public network, so the security is of prime concern. Encryption can be an effective process of protecting information, and is widely used for data security in many applications [4]. Data compression seeks to reduce the number of bits used to store or transmit data. It encompasses a wide variety of software and hardware compression techniques which can be so unlike one another that they have little in common except that they compress data [5]. Data Compression methods are divided into two types 1) lossless compression method and 2) lossy compression method [6]. In this paper we used Lossless data compression techniques. In

lossless data compression, the combination of data is preserved without loss any information. In this paper the ICM system results is used to build the dynamic clusters. The Data Mining is defined as an extraction of hidden information from large databases. It has large possibility helps the Libraries and information centers to focus on most important information in their data warehouse. There are several techniques for data mining these are: 1) classification 2) clustering 3) prediction (regression) 4) decision trees 5) sequential patterns and 6) association rules [7]. In this paper the clustering technique was used to solve the problem of accessing big data. the modified k-means clustering methods and its variants (k-means with medium probability and k-means with maximum gain ratio) was used to build clusters depending on special centers, also the dynamic clustering methods was used to build small clusters depending on the user query. The ICM, modified k-means, k-means with medium probability and k-means with maximum gain ratio was mentioned in my paper [6]. This paper has presented the design and implementation of the retrieving methods that was applied on plain and encrypted files. This research is organized as follow. Section one showed the introduction, Section two presents data compression, Section three explains major clustering techniques , Section four explains major data retrieval methods, Section five shows the data security, Section six explains the proposed work, Section seven presents experiments and results and Section eight offers the conclusion.

Data Compression

Data Compression is essentially defined as a technique to reduce the size of data. There are several data compression Techniques available which are used for efficient transmission and storage of the data with less memory space [6]. Data Compression technique takes the advantage of repetition series of data in order to provide a potential cost saving associated with transmitting less amount of data, reduces storage requirement and reduces the probability of transmission errors. Data compression techniques are divided into two main classes. Those are (i) Lossless data Compression and (ii) Lossy data compression. In lossless data compression, the compression Process is carried out without loss of data or Information during compression [8]. Lossy data compression accepts a certain loss of accuracy in

exchange for greatly increased compression. Lossy compression proves effective when applied to graphics images and digitized voice [5]. In this paper the clustering technique (Improved K-means, K-means With Medium Probability and K-means With Maximum Gain Ratio) algorithms were used as lossless compression algorithm and the results have been used to build the dynamic clusters in plain and cipher forms.

Clustering Data

Data mining is the extraction of hidden predictive information from large databases, is a powerful technology with great possibility to help companies focusing on the most important information in their data warehouses [9]. Clustering is data mining technique of grouping objects or data into clusters in which objects within the cluster have high similarity, but are very dissimilar to objects in the other clusters. Similarities and Dissimilarities are measured on the attribute values which describes the objects. Clustering methods are used to formulate and label the data, for data compression and model construction, for detection of outliers etc. Common approach of all clustering methods is to find clusters center which represent each cluster. Based on the similarity metric and input vector cluster center helps in determining which cluster is nearest or most similar one. Many clustering methods have been developed and are categorized from many aspects such as partitioning methods, hierarchical methods, density methods, grid based method, and model based methods. Data set can be numeric or categorical [10]. The two main types of cluster analysis methods these are the nonhierarchical, which divide a dataset of N items into M clusters, and the hierarchical, which output nested dataset in which pairs of items or clusters are successively linked. In the information retrieval (IR) field, cluster analysis has been used to create groups of documents with the goal of benefiting the efficiency and effectiveness of retrieval [11].

K-means method: centroid based method

K-means is one of the most commonly used clustering techniques due to its simplicity and speed. The k-means method takes the input parameter, k, and partitions the data into k clusters by assigning each object to its closest cluster centroid (the mean value of the variables for all

objects in that particular cluster) based on the distance measure used. The k- means method work as follows. Randomly k objects are selected; each object represents a cluster mean or center. object which is most similar or close to cluster mean based on the distance between the object and the cluster is assigned to the cluster . This process will remain continue until the criterion function meets.

Algorithm 1 k-mean [10]

Input: C: the number of cluster and D: A data set containing m objects.

Output: A set of C cluster.

Begin:

1: Choose m objects randomly from dataset as the initial cluster centers;

2: Until there are no changes in the mean values

3: Use the estimated means to classify m objects into k clusters based on similarity measured (Eq 1)

4: For i=1 to k

5: Calculate mean value of the objects for each cluster i and make replacing old mean with new mean

6: End_for

7: End_until

End

$$\mathbf{E} = \sum_{j=1}^{k} \sum_{\substack{i=1 \\ x_i \in c_j}}^{n} \left\| x_{i-} m_j \right\|^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In which, E is total square error of all the objects in the data cluster, xi is the vector of the i-th element of the dataset, mi is mean value of cluster Ci (x and m are both multi-dimensional). K-means is the most important clustering technique that has been used widely in the field of IR. It was grouped data objects into k clusters [12].

Data retrieval

Database is an organized collection of data. More specifically, Databases are electronic collections of information, which allows data to be easily accessed, manipulated and updated. In other words, a database is used by an organization as a method of storing, managing and retrieving information. Each item in a database is a record and each record consists of a set of fields. The database was used to retrieve items in a list or a periodical database [13]. In databases, data retrieval is the process of obtaining and extracting data from a database, based on a query provided by the user or application. It enables the fetching of data from a database in order to store it in a file, print it, viewed on the screen and/or use it within an application [14]. Information retrieval (IR) is finding items (usually documents) of an unstructured nature (usually text) that meets an information need from within large collections (usually stored on computers) [15]. The difference between information retrieval and data retrieval is summarized in the following table:

Table (1): the difference between IR and data retrieval [15]

	Data Retrieval	Information Retrieval
Example	Database Query	WWW Search
Matching	Exact	Partial Match, Best Match
Inference	Deduction	Induction
Model	Deterministic	Probabilistic
Query Language	Artificial	Natural
Query Specification	Complete	Incomplete
Items Wanted	Matching	Relevant
Error Response	Sensitive	Insensitive

Data retrieval typically requests writing and executing data retrieval or extraction commands or queries on a database based on the query provided by the user. The retrieval process has been begun with the user entering a query. The query entered by the user can be a one word or

it can be a sentence [16]. Searching strategies includes: keyword and subject searching, Boolean operators, truncation, phrase searching, search limiting, and nesting [3]. Boolean searching is a method based on logic. Logical conditions return a Boolean result based on an expression supplied by the user. Most online databases and internet search engines based on Boolean searches. The Boolean operators AND, OR, NOT (or AND NOT). Using AND narrow your search. It retrieves records that contain both of the search items or keywords that you specify. Using OR expand your search. It retrieves records that contain either of the search items (terms) or keywords that you specify, but not necessarily both. Using NOT narrows the search. It retrieves records that do not contain a search item (term) in your search [17].An index for a file in a database system works in much the same way as the index in the textbook [18]. Keyword Searching best used method for searching new terms (items), special words, jargon or slang. Phrase searching is a way to retrieve records containing specific phrases. A phrase search will locate only records containing the specified (inputted) words [3]. Keyword query is easy and flexible because it does not require from the database user to know details about the database schema. The goal of information retrieval is to identify documents which best match user needs. While the goal of data retrieval is to identify table records which best match user needs [19].The bellow figure shows the Boolean operations.

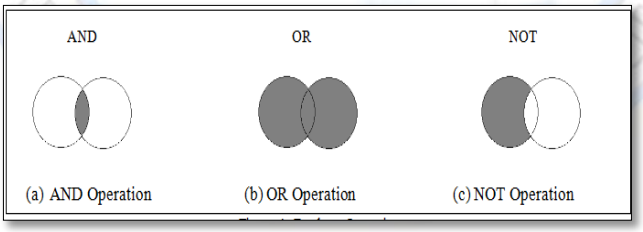

Figure 1: Boolean Operation

The Boolean operations, keyword, indexing and phrase searching methods are used in this paper for purpose of data retrieving.

Data Security

Cryptography has a long and wonderful history. Cryptography is a science of secret writing. It is the technique of protecting the information by transforming it into an unreadable format in

which a message can be hidden from the ordinary reader and only the intended recipient will be able to convert it into original text. Its main goal is to keep the data secure from unauthorized access [20]. Cryptography can reformat and transform the data, making it more secure on its transportation between computers. The technology is based on the principles of secret codes, in addition to the modern mathematics that protects the data in robust ways. To assess the security needs of an organization effectively, there must be consider three aspects of information security: **Security attack**: Any action that compromises the security of information owned by an organization, **Security mechanism**: A mechanism that is designed to detect, prevent or recover from a security attack and **Security service**: A service that promotes the security of the data processing systems and the information transfers of an organization. The services are intended to counter security attacks and it makes use of one or more security mechanisms to provide the service.

The number of keys used [21]:

(1) If the sender and receiver uses same key then it is said symmetric key (or) single key (or) traditional encryption.

(2) If the sender and receiver use different keys then it is said public key encryption.

In this paper the first type (symmetric key) encryption was used to encrypt and decrypt the data by using single secret key shared between the sender and the receiver.

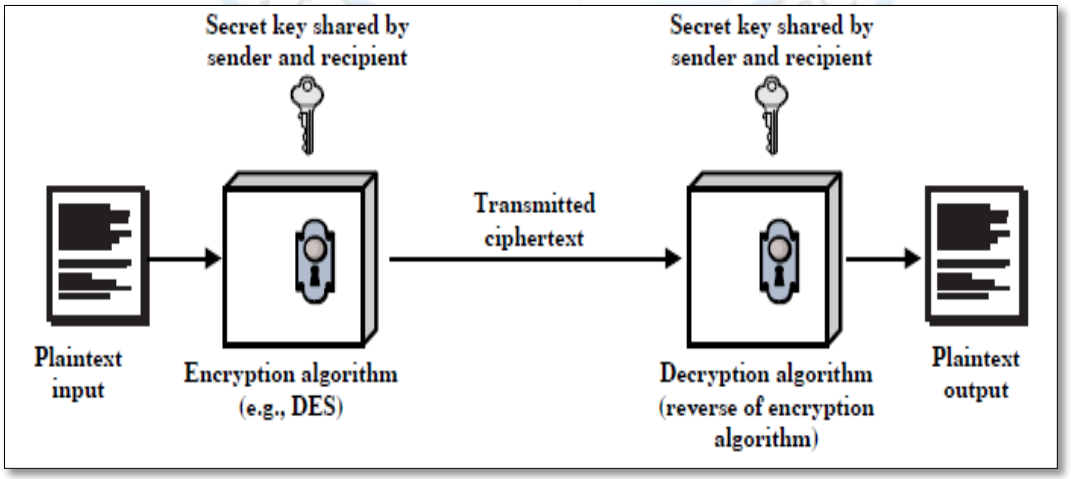

Figure 2: Simplified Model of Symmetric Encryption [24]

Proposal System

Database security is a increasing concern evidenced by an increase in the number of reported incidents from losing data or unauthorized exposure to sensitive data. The encryption/ decryption method was used to keep the confidentiality of personal data. In this research the symmetric cryptography will be used. Symmetric key encryption, usually called secret or traditional encryption. In this type the keys of encryption and decryption have the same values and it was shared between the client and server. The proposed system is dependent on matching cipher query with encrypted database. it was used simple addition method to encrypt the data. The proposed approach solves the problems of consuming too big time when users wanted to retrieve from this encrypted database. In this research many tools were used to help in speeding up the retrieving process. The ICM, improved k-means, k-means with medium probability and k-means with maximum Gain Ratio are used to compressed database and dynamic clustering method also used to build clusters dynamically. The dynamic clustering algorithm builds cluster depending on user entries. The system proposed querying over big encrypted database. In conventional system querying over big encrypted database needs too big time to retrieve from this big encrypted database. While in the proposed system this problem was solved by matching cipher query with encrypted database without decrypting the database. The results are compared based on the retrieval time in two cases: Retrieving from plain database and Retrieving from encrypted database.

1. Retrieving from Plain Database

The retrieving of data in this case is conducted by matching plain query with plain of (database, compressed database and dynamic built clusters). The user (client) is entering plain query, the matching process run at the server with plain database and results returned to the client in the plain form. The clusters built dynamically based on the user query using dynamic clustering algorithm. The proposed system can be explained in figure (3).

Retrieving Encrypted Query from Encrypted Database Depending on Symmetric Encrypted Cipher System Method

Ghassan H. Abdul-Majeed Alaa Kadhim F. Rasha Subhi Ali

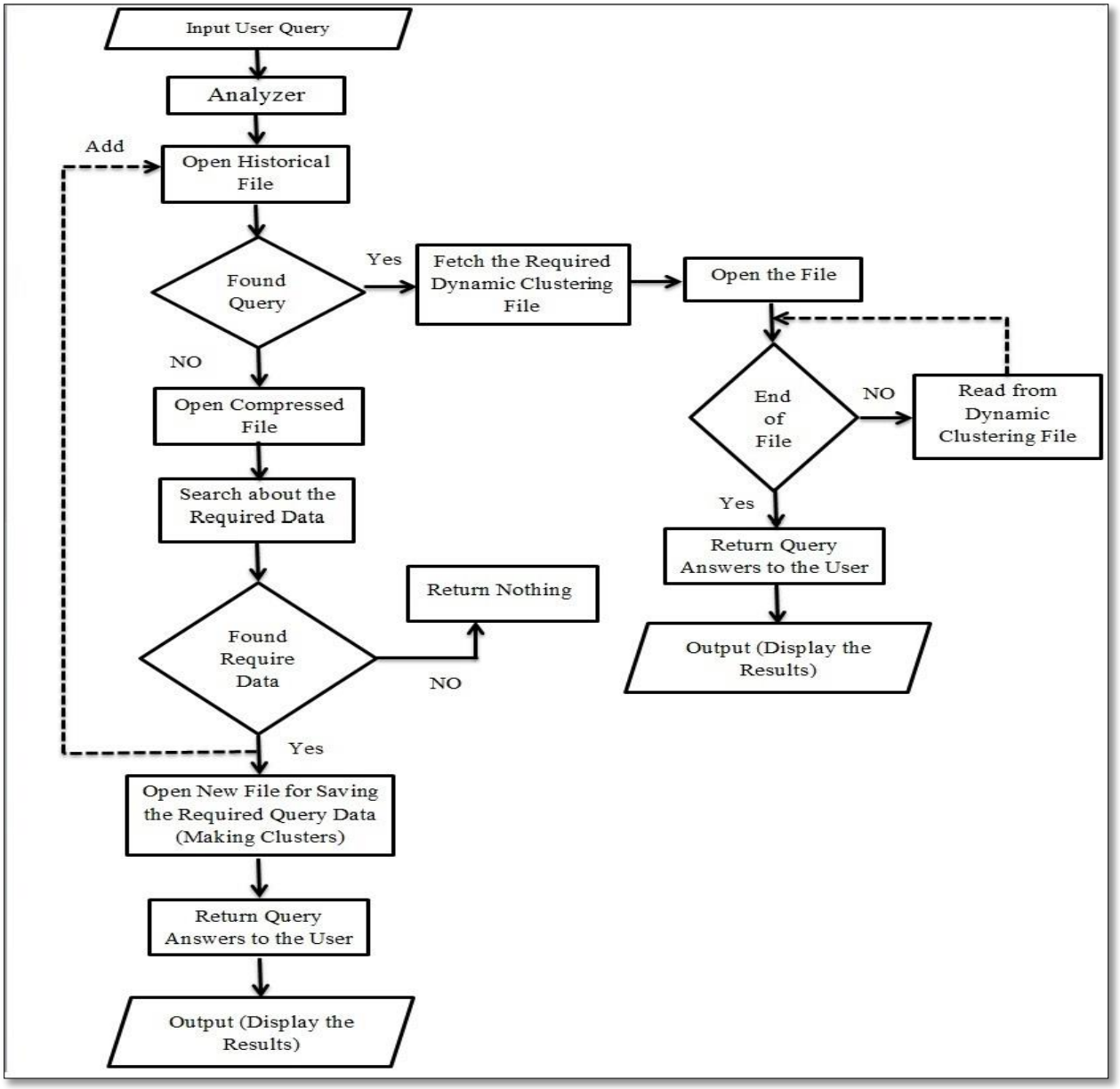

Figure 3: The structure of dynamic clustering method on plain database file

The proposal system for the current state consists of several steps and these are:

1. Input the original database file
2. Apply the ICM algorithm
3. Compress the database file with best compression method
4. Returning compressed file
5. Input the user query to the dynamic clustering algorithm
6. Analyzer: searching the user query if it was existed in the historical file or not and
7. Return the results.

The retrieving and creation of dynamic clustering steps for the current case can be explained in the following algorithm:

Algorithm 2: Retrieving and Creation Dynamic clustering algorithm based on plain database

Input: user query.

Output: answers return the required records that match the user query.

Begin:

1: open the historical file to check if the query exists or not exists in the historical file.
2: if the query exists in the historical file then
3: Fetch the path of the file that contains the data of the entered query and then open this file let it x.
4: While not end of x do
5: Search about the required query using (keyword strategy; indexing strategy or phrase and Boolean operation strategy).
6: Return all the records that match the required query data.
7: End while.
8: Else if the query not exists in the historical file then
9: While not end of the compressed file
10: Search about the required query in compressed file using (keyword strategy; indexing strategy or phrase and Boolean operation strategy).
11: Open the compressed file and match the query data with data in the compressed file.
12: Open new file for saving clusters that was extracted from matching the user query.
13: Save user query in historical file.
14: End while
15: End IF
16: Return the results to the user (client).

End.

2. Retrieving From Encrypted database

Database consists of important information used by enterprises, companies, persons ....etc. Because of the data when transmitted over the communication channels vulnerable to attack from the hackers, therefore the data must be protected. Database Security is the mechanism that protects the database against intentional or accidental threats. So, in this case the clusters created in dynamic way but the encrypted database was used instead of plain databases that were used in the previous case. The clusters were built based on the encrypted user queries (user entries or user requirements) in encrypted form and not in plain form. The proposed system for the current case includes several stages and these are explained in the following steps:

1. Input encrypted compressed database.
2. Return file consists of encrypted data.
3. Input the user query to the dynamic clustering algorithm.
4. Apply an encryption algorithm for the user query (using same encryption algorithm that was used in step5).
5. Analyzer: searching the encrypted user query if it was existed in the historical file or not and
6. Return the results

The retrieving in this case includes three phases and these are:

Phase 1: the first phase in the client, At this stage, the user is entered the query and the query is encrypted using an encryption algorithm with a symmetric key

Phase 2: this phase is working in the server, this phase works on matching encrypted query with encrypted database and also works on the retrieving process.

Phase 3: the third phase is working in the client; this phase includes the decryption process. The decryption is done for the retrieved data only and there is not needed to decrypt entire database. The proposed system for retrieving from encrypted database can be explained in figure (4).

In figure (4) the files (F1 and F2) represent the results of compression operation. The F1 file includes the clusters items and it was larger size than the F2 file. The F2 file contains database information such as (names of columns, data types of columns,…. Etc) and also includes clusters centers. If the encrypted query was not found in the historical file then the searching process would be done. The searching process includes matching the encrypted query with the encrypted compressed database file consequently generating new clusters dynamically. These clusters includes the records have a relation with the encrypted query, the encrypted query would be added to the historical file and the encrypted records are sent to the client. User at the client can do decryption process to the received records. If the encrypted query was found in the historical file then the matching process would be worked directly between the encrypted query and the records which were included in the clusters were generated dynamically. Of course these clusters created using dynamic clustering algorithm. The results would be sent to the client in the encrypted form. The decryption process would be done at the client site using decryption algorithm with shared key. The retrieving from encrypted data steps can be explained in the algorithm3.

Retrieving Encrypted Query from Encrypted Database Depending on Symmetric Encrypted Cipher System Method

Ghassan H. Abdul-Majeed Alaa Kadhim F. Rasha Subhi Ali

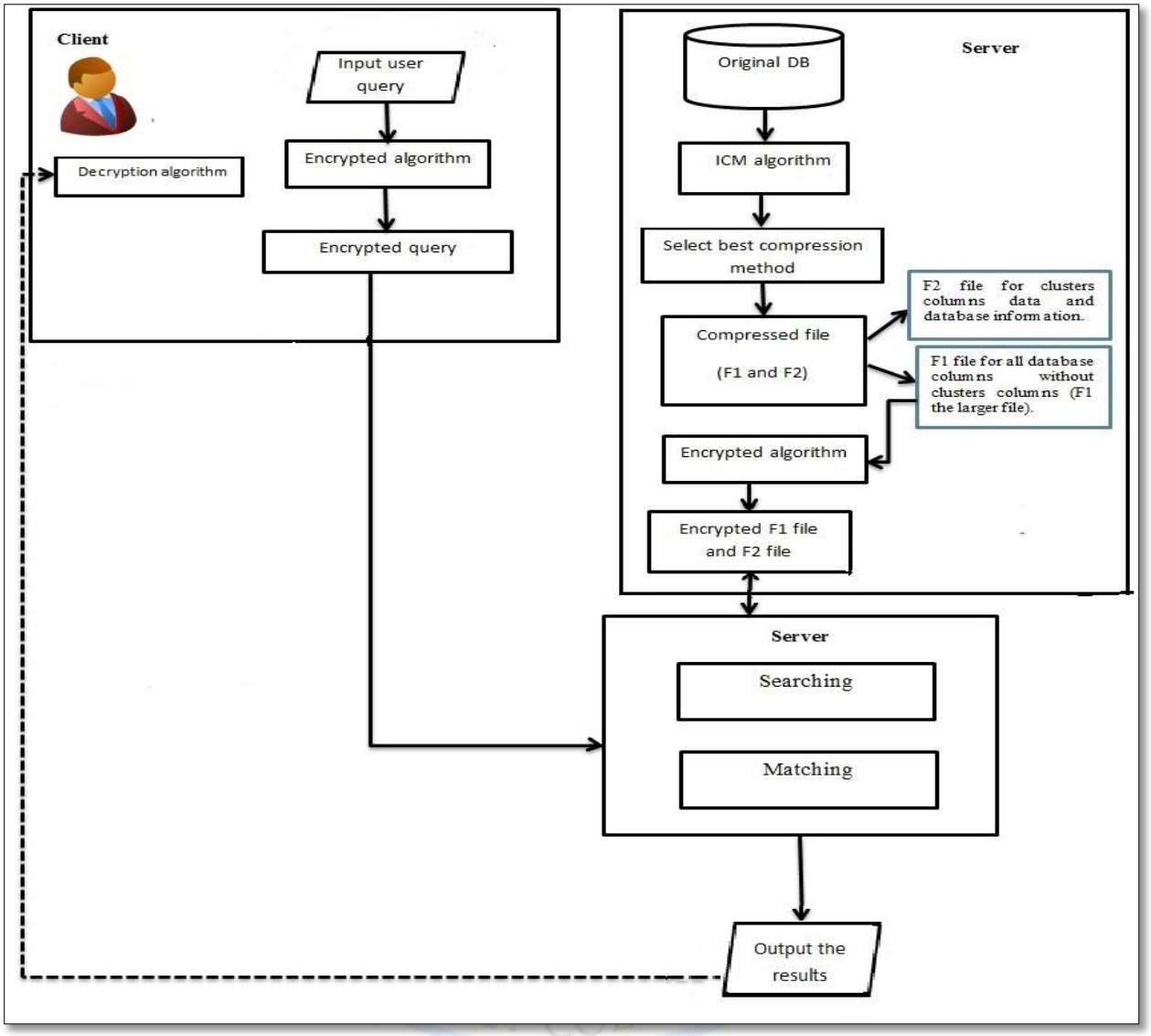

Figure 4: scenario of encrypted query from encrypted database

Algorithm 3: Retrieving and Creation Dynamic clustering algorithm based on encrypted database

Input: user query.

Output: answers return the required records that match encrypted user query.

Begin:

1: Encrypt user query by:
2: Using addition method // The same method which was used encrypt the database//
3: Add secret shared key to Encrypt the query // The key shared between client and server and it was represented same key that was used to encrypt the database//
4: The output is encrypted query let it EQ.
5: Open the historical file to check if the encrypted query exists or not exists in the historical file.
6: If the query exists in the historical file then
7: Fetch the path of the file that contains the data of the EQ and then open this file let it y.
8: While not end of y do
9: Search about the required EQ using (keyword strategy; indexing strategy or phrase and Boolean operation strategy).
10: Return all the records that match the required EQ data.
11: End while.
12: Else If the query not exists in the historical file then
13: While not end of the encrypted compressed file let it EC
14: Search about the required EQ using (keyword strategy; indexing strategy or phrase and Boolean operation strategy).
15: Open EC file and match the EQ data with data in the EC file.
16: Open new file for saving clusters that was extracted from matching the user EQ.
17: Save user EQ in historical file.
18: End while
19: End If
20: For all returned records apply decryption operation by using the following steps:
21: Using subtraction method // The same method which was used for the encryption process//
22: Add secret shared key to decrypt the retrieved records // The key shared between client and server and it was represented same key that was used for the encryption process//
23: Return the results in plain form.
24: End for
25: display the results to the user (client).

End.

In algorithm (3) the steps from 1 to 4 will be in the client site and represent phase 1, the steps from 5 to 20 will be in the server site and assimilate phase 2 and the steps from 21 to 25 will be at the client site and represent phase 3. In this research the Unicode conversion for the characters was used instead of using character code. It was used because it was taken more range than character code. The encryption process for the query in the client site and the decryption process to the returned results at the client site depended on the used encryption method. Always the searching in encrypted database consumes big time because of it needed to decryption either all the database or some database columns and this problem was solved by using cipher with cipher matching. The following figures showing a comparison between these two methods.

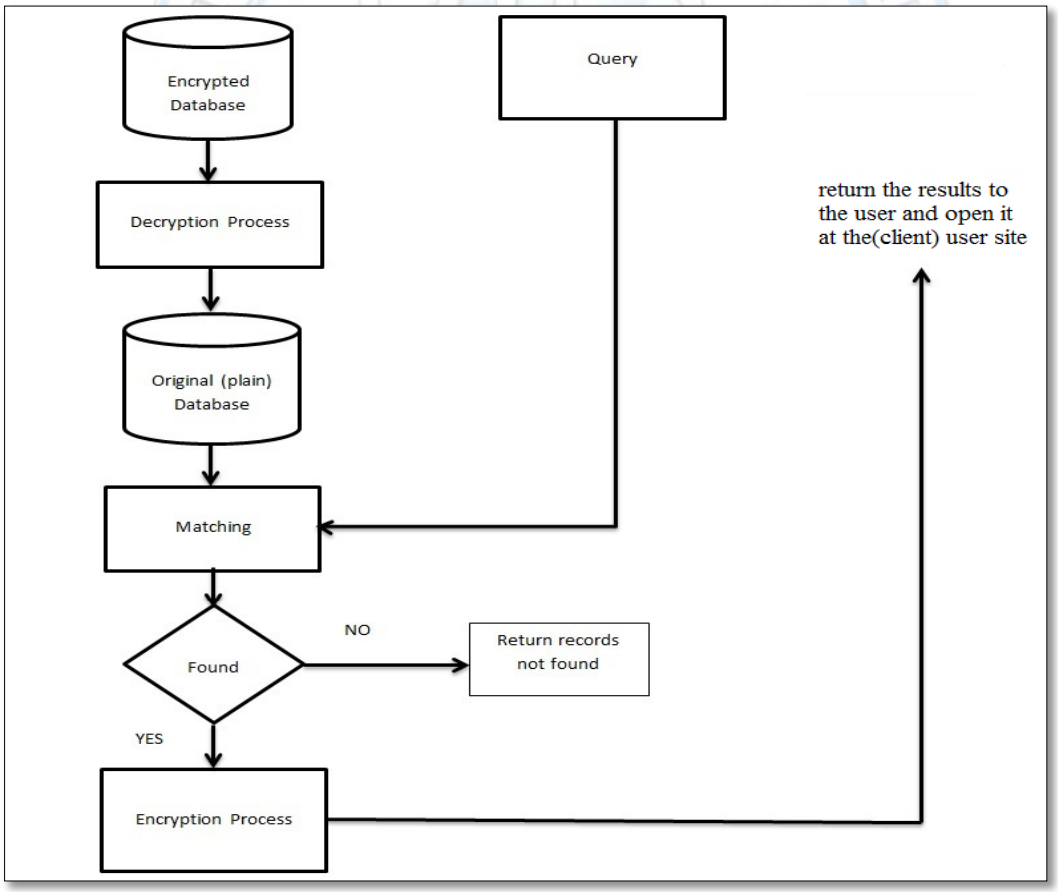

Figure 5: The traditional searching method on encrypted database

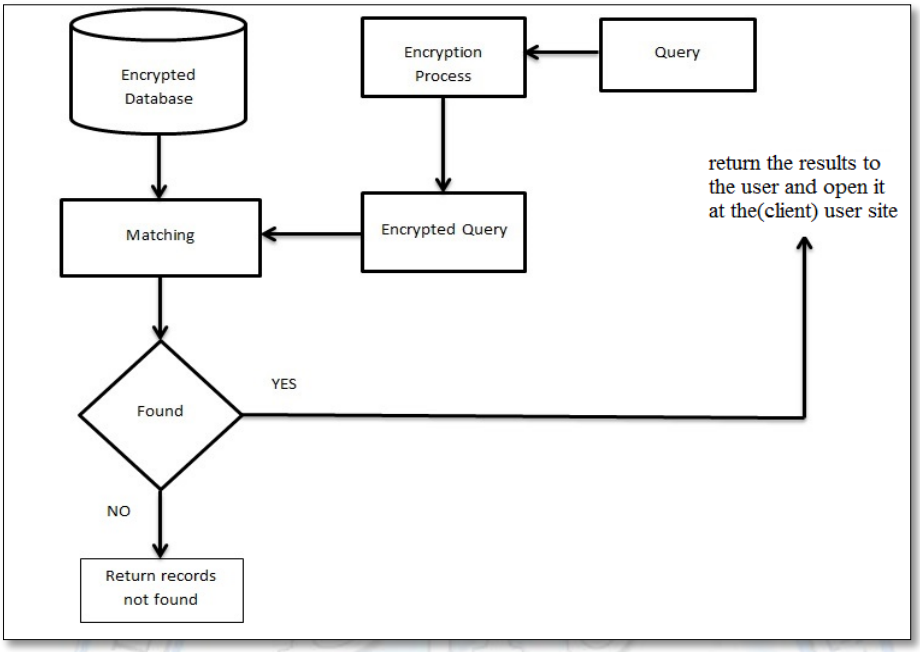

Figure 6: The proposed searching method on encrypted database

Results and Discussion

This section discusses the results which were obtained from the applying query matching on plain and encrypted database. The retrieval method was applied on different data sets for different data types, and was conducted different experiments to specify the performance of the proposed methods (retrieval by matching cipher query with cipher data) which is applied on databases. The experiments have been conducted on different databases and the results are compared based on the consumed time in retrieving from original database file, compressed database file and dynamic clustering file. The proposed method results showed in the following tables. The time was measured in seconds (e.g. 40.11seconds=40110 milliseconds and 0.016 seconds=16 milliseconds). Tables [3, 4 and 5] show the detailed results for the tested databases. From table (2 and 3) we notice that the consumed time to answering about the user query in the state of querying from the dynamic clustering file is too much faster than answering in the situation of querying from the original, compressed database file or encrypted (original and compressed) database file . This is because of the dynamic clustering file contains the data that had related with only user query. Optimization proposal is advances by reducing the required vocabulary check about the user query and this is happening in the case of the clusters for this query have been previously built.

Retrieving Encrypted Query from Encrypted Database Depending on Symmetric Encrypted Cipher System Method

Ghassan H. Abdul-Majeed Alaa Kadhim F. Rasha Subhi Ali

Table (2) Proposed System Answering Time Results

Database information					Retrieval by: plain with plain compression			Retrieval by: cipher with cipher compression			N.R.R	Retrieval by: cipher with cipher compression with decoding time			Decoding time		
						Preprocess DB			Preprocess DB								
DB Name	DB Size	No.of Query	NO. Records	NO. Columns	R.O.DB.T p with p	R.C.T p with p	R.D.T p with p	R.O.DB.T c with c	R.C.T c with c	R.D.T c with c		R.O.DB.T c with c	R.C.T c with c	R.D.T c with c	D.O.DB.T	D.C.T	D.D.T
dept	224 KB	8	100	6	0.98	0.052	0.005	0.623	0.021	0.005	1	0.6231	0.0211	0.0051	0.0001	0.0001	0.0001
					0.47	0.005	0.0001	0.463	0.005	0.0001	1	0.4631	0.0051	0.0002	0.0001	0.0001	0.0001
					0.107	0.0001	0.0001	0.346	0.006	0.0001	1	0.3461	0.0061	0.0002	0.0001	0.0001	0.0001
					0.439	0.0001	0.0001	0.228	0.006	0.0001	1	0.2281	0.0061	0.0002	0.0001	0.0001	0.0001
					0.465	0.005	0.011	0.179	0.003	0.005	2	0.1791	0.0031	0.0051	0.0001	0.0001	0.0001
					0.131	0.0001	0.0001	0.102	0.001	0.0001	1	0.1021	0.0011	0.0002	0.0001	0.0001	0.0001
					0.475	0.052	0.005	0.723	0.011	0.005	3	0.7231	0.0111	0.0051	0.0001	0.0001	0.0001
					0.097	0.0001	0.0001	0.231	0.006	0.0001	1	0.2311	0.0061	0.0002	0.0001	0.0001	0.0001
					1.051	1.903	0.26	0.613	1.793	0.266	257	1.004	2.059	0.532	0.391	0.266	0.266
					0.773	0.113	0.043	0.289	0.125	0.068	16	0.32	0.141	0.0894	0.031	0.016	0.016
					0.586	0.285	0.078	0.794	0.234	0.078	58	0.872	0.312	0.156	0.078	0.078	0.078
					0.798	0.241	0.039	0.584	0.156	0.059	34	0.631	0.203	0.106	0.047	0.047	0.047
					20.775	110.853	2.62	26.81	104.821	2.618	1986	28.753	107.47	5.267	2.943	2.649	2.649
					31.459	18.61	0.203	33.432	13.318	0.163	71	33.557	13.434	0.279	0.125	0.116	0.116
					29.727	16.053	0.078	31.194	12.348	0.026	1	31.21	12.3481	0.0281	0.016	0.0001	0.0001
					29.538	17.673	0.104	31.059	12.171	0.029	1	31.075	12.1711	0.0291	0.016	0.0001	0.0001
					18.503	16.107	0.073	19.037	12.021	0.072	1	19.0371	12.0211	0.0721	0.0001	0.0001	0.0001
					5.322	16.112	0.099	5.024	12.001	0.069	1	5.0241	12.0011	0.0691	0.0001	0.0001	0.0001
					5.16	16.36	0.052	4.908	12.186	0.052	5	4.924	12.1861	0.0521	0.016	0.0001	0.0001
					34.342	19.104	0.322	40.291	14.589	0.31	242	40.887	14.906	0.627	0.396	0.317	0.317
					18.5	16.303	0.016	19.709	11.971	0.026	1	19.713	11.9711	0.0281	0.004	0.0001	0.0001
					18.186	15.747	0.021	19.841	12.06	0.019	1	19.845	12.0601	0.0191	0.004	0.0001	0.0001
					26.09	26.056	0.819	29.53	22.139	0.834	621	30.446	22.971	1.686	0.916	0.832	0.832
					31.429	16.621	0.068	34.091	14.324	0.078	12	34.139	14.34	0.094	0.048	0.016	0.016
					5.305	15.915	0.021	5.14	12.058	0.047	5	5.148	12.062	0.051	0.008	0.004	0.004
					30.236	16.159	0.037	32.432	12.035	0.016	1	32.44	12.0351	0.0161	0.008	0.0001	0.0001
					30.567	16.094	0.026	32.869	11.889	0.016	1	32.8691	11.8891	0.0161	0.0001	0.0001	0.0001
					35.577	31.771	1.014	39.387	27.2	1.018	785	40.529	28.221	2.039	1.142	1.021	1.021
					36.865	16.462	0.026	38.712	11.942	0.022	1	38.7121	11.9421	0.0221	0.0001	0.0001	0.0001
					35.451	17.802	0.443	38.987	13.534	0.45	189	39.393	13.768	0.894	0.406	0.234	0.234
					27.972	18.751	0.385	36.614	15.215	0.365	220	37.067	15.632	0.682	0.453	0.317	0.317
					24.065	25.655	0.717	25.418	22.985	0.833	570	26.309	23.751	1.589	0.891	0.766	0.766
					30.705	16.929	0.021	33.199	12.943	0.016	1	33.1991	12.9431	0.0161	0.0001	0.0001	0.0001
					31.824	16.368	0.167	34.236	13.029	0.075	25	34.314	13.06	0.106	0.078	0.031	0.031
					30.622	16.029	0.021	32.672	12.628	0.015	1	32.6721	12.644	0.031	0.0001	0.016	0.016
DWC	1,360 KB	4	3679	8													
voters	288,304 KB	23	1178528	10													

DB Name (database name), DB Size (database size), NO of Query (number of query), N.R.R (number of retrieving records), NO. Records (number of records in the database), NO. Columns (number of columns in the database), Plain with Plain or P with P (matching plain query with plain data), Cipher with Cipher or C with C (matching encrypted query with encrypted data), Preprocess DB (applying the clustering algorithms on the database either best clustering algorithm only or best clustering algorithm with dynamic clustering algorithm), R.O DB.T p with p (retrieval time from original database file in plain form), R.C.T p with p (retrieval time from the compressed file), R.D.T p with p (retrieval time from built dynamic clustering file), R.O DB.T c with c (retrieval time from original database file in plain form), R.C.T c with c (retrieval time from the compressed file), R.D.T c with c (retrieval time from built dynamic clustering file), D.O DB.T (decoding time for records retrieved from original database), D.C.T (decoding time for records retrieved from compressed database), D.D.T (decoding time for records retrieved from file which was created dynamically (results of dynamic clustering algorithm)), T.T.R.O DB (total time for retrieving from original database file in the plain or cipher form), T.T.R.C DB (total time for retrieving from the compressed database file in the plain or cipher form) and T.T.R.D clus (total time for retrieving from dynamic clustering resulted file in the plain or Cipher form).

Table (3) Proposed System Total time Results

DB Name	plain with plain			cipher with cipher			Retrieval by cipher with cipher comparession with decoding time		
	T.T.R.O DB	T.T.R.C DB	T.T.R.D clus	T.T.R.O DB	T.T.R.C DB	T.T.R.D clus	T.T.R.O DB	T.T.R.C DB	T.T.R.D clus
dept	3.164	0.1144	0.0215	2.895	0.059	0.0155	2.8958	0.0598	0.0163
DWC	3.208	2.542	0.42	2.28	2.308	0.471	2.827	2.715	0.878
voters	588.222	513.894	7.353	644.592	419.407	7.169	652.0626	425.7271	13.4891
Total time	594.592	516.5504	7.7945	649.767	421.774	7.6555	657.7854	428.5019	14.3834

Retrieving Encrypted Query from Encrypted Database Depending on Symmetric Encrypted Cipher System Method

Ghassan H. Abdul-Majeed Alaa Kadhim F. Rasha Subhi Ali

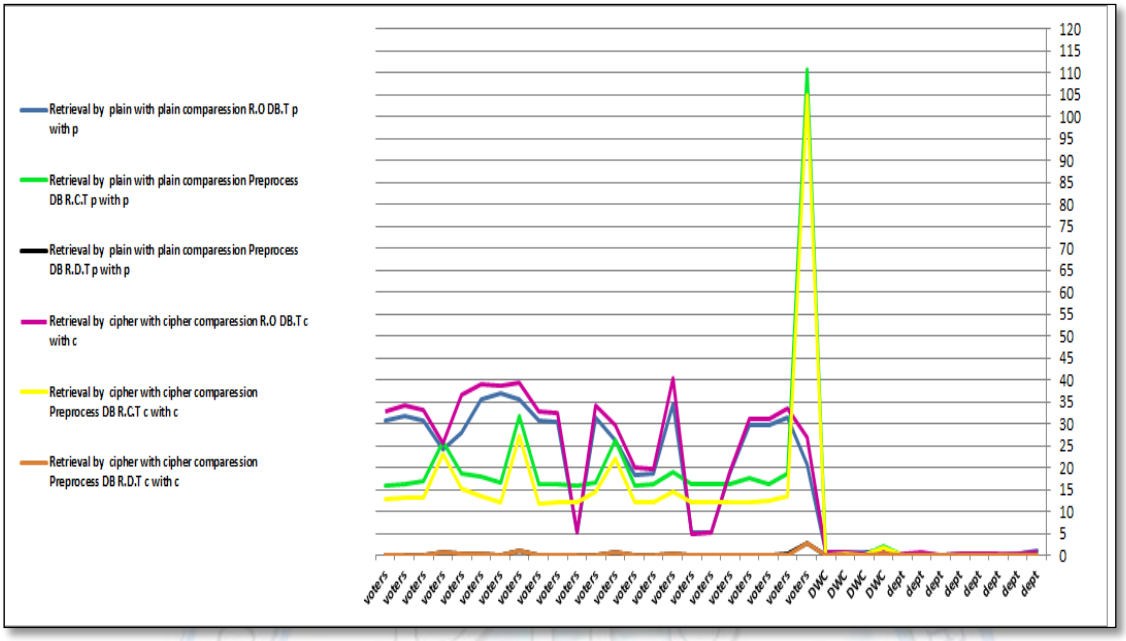

Figure 7: Retrieval Time in Both Cases of Plain and Encrypted data

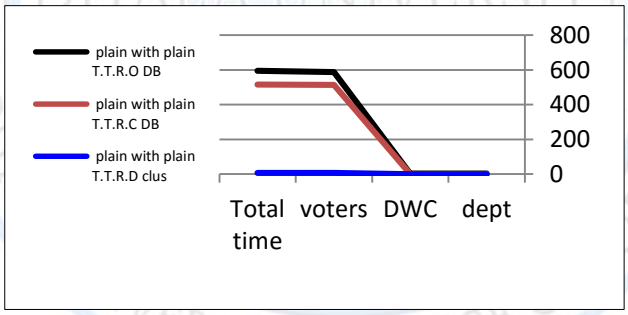

Figure 8: total retrieval time in the case of plain matching

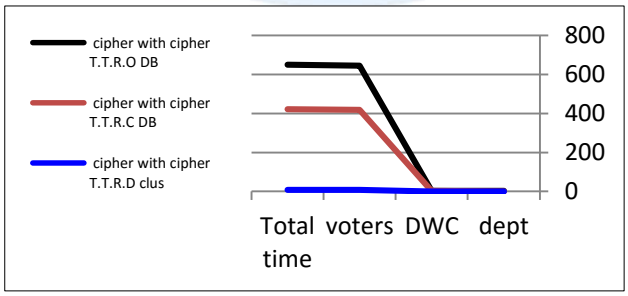

Figure 9: total retrieval time in the case of cipher matching

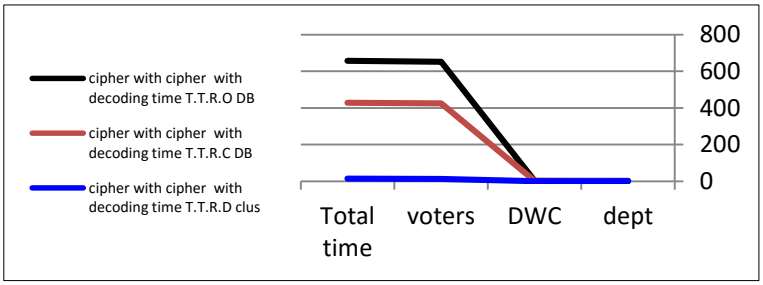

Figure 10: total retrieval time in the case of cipher matching with decoding time

Figures (8,9 and 10) showed total time that is consumed for answering the query from dynamic clustering results much faster than retrieving from original or compressed databases. This is for both cases of matching plain with plain data or cipher with cipher data. The results showed that: 1) the larger retrieving time from the original database file is 36.865 seconds and the smaller retrieving time is 0.097 seconds, 2) the larger retrieving time from the compressed database file is 110.853 seconds and the smaller retrieving time is 0.0001 seconds and 3) the larger retrieving time from the dynamic clustering file is 2.620 seconds and the smaller retrieving time is 0.0001 seconds. These three points in the case of retrieving data by matching plain with plain data. The results of comparison cipher with cipher showed that: 1) the larger retrieving time from the original database file is 40.291 seconds and the smaller retrieving time is 0.102 seconds, 2) the larger retrieving time from the compressed database file is 104.821 seconds and the smaller retrieving time is 0.001 seconds and 3) the larger retrieving time from the dynamic clustering file is 2.618 seconds and the smaller retrieving time is 0.0001 seconds. The results of comparison cipher with cipher with decoding time showed that: 1) the larger retrieving time from the original database file is 40.687 seconds and the smaller retrieving time is 0.1021 seconds, 2) the larger retrieving time from the compressed database file is 107.47 seconds and the smaller retrieving time is 0.0011 seconds and 3) the larger retrieving time from the dynamic clustering file is 5.267seconds and the smaller retrieving time is 0.0002 seconds.

The average time that was consumed for retrieving data of 35 queries is shown below:

Table (4) The average retrieval time

DB Name	Plain with Plain comparission			Cipher with Cipher comparession			Retrieval by cipher with cipher comparession with decoding Time		
	T.T.R. O DB	T.T.R. C DB	T.T.R. D clus	T.T.R. O DB	T.T.R. C DB	T.T.R. D clus	T.T.R.O DB	T.T.R.C DB	T.T.R.D clus
Average Time	16.988	14.759	0.223	18.565	12.051	0.219	18.794	12.243	0.411

Conclusion

The data retrieval process is considered the main objective of this research and not the encryption process. So, the simple encryption operation was used to measure the performance of the data retrieving method (by matching encrypted query with encrypted data).We are in the process of application of this work using the proposed method for retrieving data with the application of one of strong encryption algorithms. Most of the operations occur on the penetration of communication channels on the outgoing records so we need encrypt the database to protect it from attackers. In conventional systems the query process from a large encrypted database needs too large time because it needed to decrypt this database as a whole or a part of it and then recovered records is encrypted and sent to the client. The proposed system improved the performance of the retrieving algorithm by decreasing the consumed time. The proposed system works on matching cipher query with encrypted database consequently gaining time during not decrypting whole database or part of database. Therefore, the proposed system solves this problem is by sending an encrypted query, working encrypted search and returning encrypted results. The results shows that the proposed system for retrieving from big encrypted data get a good results in decreasing the consumed time for retrieving data. Dynamic clustering algorithm has been worked on improving data retrieval time in both cases of 1) The retrieving from plain database and 2) The retrieving from encrypted database.

References

1. S. Parthasarathy, V. Shakila, **"KNOWLEDGE CLUSTERING ON BIG DATA WITH K_MEANS ALGORITHM"**, International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 02, May-2015, e-ISSN: 2395 -0056, p-ISSN: 2395-0072**.**

2. S.Balasubramaniam, Dr.V.Kavitha, **" A Survey on Data Retrieval Techniques in Cloud Computing"**, Journal of Convergence Information Technology(JCIT), Volume8, Number16, November 2013.

3. E. Petraki, C. Kapetis, E. J.Yannakoudakis**, "Conceptual Database Retrieval through Multilingual Thesauri"**, Computer Science and Information Technology 1(1): 19-32, 2013, DOI: 10.13189/csit.2013.010103.

4. Rajni Jain, Ajit Shrivastava, **" Design and Implementation of New Encryption algorithm to Enhance Performance Parameters"**, IOSR Journal of Computer Engineering (IOSRJCE), ISSN: 2278-0661 Volume 4, Issue 5 (Sep.-Oct. 2012), PP 33-39, www.iosrjournals.org.

5. Mark Nelson, **" The Data Compression Book"**, 2nd edition, Jean-Loup Gailly copyright material, (Publisher: IDG Books Worldwide, Inc., ISBN: 1558514341.

6. Assist.Prof.Dr.AlaaKadhim F, Prof. Dr. Ghassan H. AbdulMajeed, RashaSubhi Ali, **"ICM Compression System Depending On Feature Extraction"**, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 4, Issue 3, May-June 2015, ISSN 2278-6856.

7. Simranjit Kaur, RuhiBagga, **"A SURVEY ON DATA MINING AND ITS TECHNIQUES"**, InternationalJournalInAppliedStudiesAndProduction Management, Volume1,Issue 3, 15 May- 15 August2015, ISSN2394-840X.

8. Pooja Jain , Anurag jain, Chetan Agrawal, **"IMPROVING DATA COMPRESSION RATIO BY THE USE OF OPTIMALITY OF LZW & ADAPTIVE HUFFMAN ALGORITHM (OLZWH)"**, International Journal on Information Theory (IJIT),Vol.4, No.1, January 2015.

9. M. Premalatha, G. Baskaran, **" Bootstrap Based Large Scale Data Processing Using Cluster"**, International Journal of Advance Research and Innovation, Volume 3, Issue 2 (2015) 359-361, ISSN 2347 – 3258.

10. Kavita Nagar, **" Data Mining Clustering Methods: A Review"**, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015 ISSN: 2277 128X, Research Paper Available online at: www.ijarcsse.com.

11. Frakes William B. and Yates Ricardo Baeza**," *Information Retrieval: Data Structures &Algorithms"*,*1991*.

12. MansafAlam, KishwarSadaf, **"Web Search Result Clustering based on CuckooSearch and Consensus Clustering"**, Dept. of Computer science, Jamia Millia Islamia, New Delhi, India, 23 Mar 2015.

13. Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, " **DATABASE SYSTEMS The Complete Book"**, Second Edition, Department of Computer Science Stanford University, 2009

14. NamrataGadkari, Sylvester Savio Raj, HarshadRaka, **"Query Subtopic Mining from Search Log Data"**, International Journal of Current Engineering and Technology, Vol.5, No.3 (June 2015), E-ISSN 2277 – 4106, P-ISSN 2347 – 5161.

15. Dr. Pushpak Bhattacharyya, JoydipDatta, **"Ranking in Information Retrieval"**, Department of Computer Science and Engineering, Indian Institute of Technology, Bombay, April 16, 2010.

16. NamrataGadkari, Sylvester Savio Raj, HarshadRaka, **"Query Subtopic Mining from Search Log Data"**, International Journal of Current .

17. Andrew Aksyonoff, **"Introduction to Search with Sphinx"**, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2011.

18. Abraham Silberschatz, Henry F. Korth, S. Sudarshan, **"DATABASESYSTEM CONCEPTS"**, S I XTH E D I T I ON, Published by McGraw-Hill, Copyright © 2011 by The McGraw-Hill Companies,ISBN 978-0-07-352332-3.

19. E. Petraki, C. Kapetis, E. J. Yannakoudakis, **"Conceptual Database Retrieval through Multilingual Thesauri"**, Computer Science and Information Technology 1(1): 19-32, 2013, DOI: 10.13189/csit.2013.010103.

20. Amare Anagaw Ayele, Dr. Vuda Sreenivasarao, **" A Modified RSA Encryption Technique Based on Multiple public keys"**, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 4, June 2013, ISSN (Print) : 2320 – 9798, ISSN (Online): 2320 – 9801.

21. W. Stallings, **"Cryptography and Network Security: Principles and Practices"**, 3rd edition, Prentice Hall, NJ, 2003.