A New Solution to Protect Encryption Keys when Encrypting Database at the Application Level

Karim El bouchti¹, Soumia Ziti², Fouzia Omary³, Nassim Kharmoun⁴
Department of Computer Science, IPPS Team
Faculty of Sciences, Mohammed V University in Rabat
Morocco

Abstract—Encrypting databases at the application level (client level) is one of the most effective ways to secure data. This strategy of data security has the advantage of resisting attacks performed by the database administrators. Although the data and encryption keys will be necessarily stored in the clear on the client level, which implies a problem of trust viz-a-viz the client since it is not always a trusted site. The client can attack encryption keys at any time. In this work, we will propose an original solution that protects encryption keys against internal attacks when implementing database encryption at the application level. The principle of our solution is to transform the encryption keys defined in the application files into other keys considered as the real keys, for encryption and decryption of the database, by using the protection functions stored within the database server. Our proposed solution is considered as an effective way to secure keys, especially if the server is a trusted site. The solution implementation results displayed better protection of encryption keys and an efficient process of data encryption/decryption. In fact, any malicious attempt performed by the client to hold encryption keys from the application level cannot be succeeded since the real values of keys are not defined on it.

Keywords—Database encryption; encryption key protection model; database encryption keys protection; data security

I. INTRODUCTION

Today, Database (DB) security is considered as one of the significant challenges in the computer world. Recently, it has been the subject of several debates and studies by experts and researchers in the data security field. The main purpose is to protect sensitive data against unavailability, leakage, and modification face to attackers’ threats [9], [21], [22].

With the rapid development of technology, the attack scenarios on the DBs have become easy to realize. Currently, a penetration to a DB using SQL injection techniques becomes fast and quite simple thanks to specialized tools [10]. Different threats may come from various sources, some from trusted DB users, others from external ones, and some attacks are performed by the DB administrators [1], [2], [3]. Data theft is a dangerous attack. However, attacks compromising data integrity can generate heavier consequences; they are hardly detectable as the theft of data. For this reason, the implementation of three security levels is necessary: 1 / - Physical security, 2 / - Operating system security, 3 / - Database Management System (DBMS) security [15], [18].

The implementation of these security levels is useful, especially the access control mechanism implemented at the DBMS level. It is considered as the first defense line against unauthorized access [14], [23]. However, this mechanism is not sufficient, it protects only against attacks coming from outside the information system, and some limits can be generated as it has been explained in [11], [16]. Besides the access control mechanism, a security approach based on DB encryption can play an important role; it can be performed on three levels: DB level, hard disk level, and application level [7].

Encrypting DB at the application level consists of delegating encryption and decryption to applications that are connected to it. The DB server processes and manipulates only encrypted data, and the encryption keys are not implemented on the server. This solution provides strong protection of the encryption keys from threats performed by the administrator. However, the keys must necessarily be stored either on the applications or in a place where the applications are managed (application server for example). This approach is limited as the application user (or the application server administrator) are not always trusted sites, they can attack encryption keys at any time, and decrypt all sensitive data without leaving any traces [16].

It is mentioned in [8] that DBs confront several internal threats, especially those coming from legitimate users of the system. These threats can be realized in collaboration with a malicious administrator, another legitimate user, or with a malicious attacker outside the information system [1], [12]. Excessive privilege abuse, legitimate privilege abuse, and elevation of privilege are some models of internal attacks that compromise encryption keys. A legitimate DB user can exploit the privileges one’s to attack encryption keys directly. The user can also exploit uncorrected DB vulnerabilities or DB configuration errors to access illegally to the DB encryption keys location. A mistrust application administrator may reveal the encryption keys to an external attacker in order to attack the DB outside the perimeter of the information system [8], [13]. Trusted users are considered a serious threat to the security of encryption keys when encryption is performed at the application level. Otherwise, the internal attacks are considered dangerous not only on DBs but on Big Data platforms like Hadoop. For instance, the authors of [24] argued that an effective attack launched from the compromised node could degrade the data processing performance of the cluster. Then, they exposed an effective schema that might mitigate the risk of this attack and keep the cluster running efficiently. Also, the
authors of [25] have developed a solution called ROVER, which is an efficient and verifiable Erasure Coding based Storage (ECS) for Big Data platforms. They showed that this solution implementation has good robustness and effectiveness against attacks from compromised nodes.

In order to protect encryption keys within applications, several solutions have been proposed by researchers. Among the most important works, we highlight the work proposed by Ding et al. [19]. They have proposed a new data encryption model implemented at the application level that ensures the confidentiality of sensitive data. It is based on a new method using keys chain to protect encryption keys. The authors of [20] have proposed a special concept called "Encryption as a service" to encrypt DB. Its main goal is to outsource the encryption system outside applications as an encryption service provider unit independently of applications users and the DB server. The protection of encryption keys in this concept is done using a Master key. Bouananim et al. have presented in [17] a solution named "Client HSM". It consists of integrating the module "HW Security Module" at the level of each DB user. This module holds and protects encryption keys to eliminate their exposure.

Despite the efforts made by researchers to improve the protection of encryption keys at the application level, they still need to be developed and improved. The solutions proposed in the literature are not sufficient as well as each solution has its limits, as we have mentioned in our previous works [3], [5], [8]. In addition, the solutions mostly presented are specific in the case when encryption is done at DBMS.

In this context, the present work aims to propose a new solution to protect encryption keys within applications against internal attacks when adopting encryption at the application level. Our solution is original and provides better protection for keys without any requirement of equipment or material. It is based on the outsourcing of the protection of encryption keys defined on the client side to a DB server. The principle is to transform user’s queries holding user’s encryption keys to new queries that remotely call functions stored on the DB server in order to convert these keys to real keys for encryption and decryption.

Our paper will be presented as follows: Section II consists of giving general overview of DB encryption at the application level. Section III describes our proposed solution and explains its implementation; we discuss in this section the test results provided by the implementation. Finally, the article ends with a conclusion in Section IV.

II. DATABASE ENCRYPTION PRELIMINARIES

A. Database Encryption at the Application Level

Encrypting DB at the application level (client side encryption) means that the process of encrypting/decrypting data is done locally on the application before transmitting data to the DB server. This principle is similar to an externalization of the DB in a cloud storage service [6], [16]. The keys in this concept are managed by the application on the client. It holds all the encryption keys, and there is no transmission of data or keys in the clear to the DB server [4] (see Fig. 1).

DB encryption at the application level has many advantages. It resists against attacks coming from administrators. However, it has several disadvantages such as, data and keys are necessarily stored in clear mode on the client, which makes a problematic trust face to the DB users, or even the applications administrator. Both of them cannot always be trusted sites [4], [16].

B. Internal Attack on Encryption at the Application Level

As it has been highlighted by the authors of [4], [5], [8], [16], the DB legitimate operators (including operators that manage applications connected to DB) are considered the main threat compromising the encryption keys when DB is encrypted at the application level. The values of keys and their locations are the vulnerability of this concept. In order to clarify the concept of internal attack when encrypting DB at the application level, we have chosen the attack on DB in both the 2-tier and 3-tier architectures.

In 2-tier client/server architectures, we have 2 essential components [26] including: the client machine (first level) and the DB server (second level). In this client-server model, the DB is encrypted at the client level. The client program accesses DB directly and the encryption keys are in clear on the clients, either in the codes or in the application's configuration files. Thus, they are totally submitted to each user that executes the programs. The users can easily get the encryption keys and attack the DB.

In the 3-tier client-server architectures, we have 3 essential components [26]: the client machine (first level), the application server (second level), and the DB server (third level). In this model, often the keys are stored in clear on the application server and protected using a password in a well-defined server location. Likewise, they are completely submitted to the application server administrator. A simple attack scenario in the 3-tier architecture is a conspiracy attack conducted by a user and the application server administrator. A malicious administrator can reveal all the encryption keys to another user in order to attack DB and decrypt its sensitive data completely without leaving any traces.

III. PROPOSED SOLUTION

The principle of the proposed solution is to outsource the client's encryption key protection to the DB server. In fact,
before sending the user's query, which contains an encryption key (or multiple encryption keys) to the DB server, we transform it into a new query that converts the key (those keys) to the real encryption/decryption key of the DB. For this purpose, we propose and implement two types of solutions. The first one concerns the protection of encryption keys when inserting data into a single sensitive column, while the second concerns the protection of keys when inserting data into multiple sensitive columns.

Let’s consider Table I which has the structure (A) and let us suppose the attribute "Salary" as a sensitive column to be encrypted during insertion and decrypted during the consultation of Table I.

Table I (code, Last_name, First_name, Salary) (A)

A. Solution1: Inserting in a Single Sensitive Column

Let’s consider a user who executes an insert query having the following form:

Insert into Tab1 values ('0001', 'elbouchti', 'karim', AES192 ('Mycolor012345678has@ml@p', 2000)); (B)

The user inserts in Tab1 a data line, the column "Salary" will be encrypted with the algorithm AES192 by using the key K="Mycolor012345678has@ml@p". Before sending this query to the DB server, it will be transformed to another query having the syntax below:

Insert into Table I values ('0001', 'elbouchti', 'karim', AES192(Func ('Mycolor012345678has@ml@p'), 2000)); (C)

Func() is a function called when the user executes an insert query on a DB table, which has a single sensitive column. The description of Func() is not defined in the files of the user's application or the application server. Its main role is to transform the encryption key value defined in the query that comes out of the user's application into a new encryption key whose value is the real encryption/decryption key of the column (see Fig. 2).

We propose to implement the function Func() in the DB server, it will be one of the objects created inside it. Its model follows the algorithm below:

Func (X)

{ K = E (X, H (Nom_table));
Retour (K);}

With:

• E: A symmetric encryption algorithm.
• H: A hash function.
• X: The key defined in the insert statement (user's application files).
• K: the real key to encrypt/decrypt the data of the column.
• Table_name: The table name defined in the insert statement.

The conversion of query (B) into the query (C) is performed by a transformation algorithm (Algorithm 1 and Algorithm 2 defined below in the two implementations of our solution). These algorithms must appear in executable mode among the files of the application as shown in Fig. 3.

B. Implementation of the Solution 1

The implementation of solution 1 has been performed as follows:

• We have developed an application and connect it to a DB named "DB_karim" developed under ORACLE10G.
• We have chosen the AES256 as an encryption algorithm and the MD5 as a hash function to develop the function Func().

When a user executes an insert query, it will be transformed by the algorithm "Algo-client1" into a new query that calls the function Func() of the DB. The code of the "Algo-client1" is defined as follows:

Algorithm 1: Query transformation in solution 1.

```
Algo-client1

Input: executed_query
Output: new_query
Begin
Decompose (executed_query)
   Func(key)
   Generate (new_query)
   Execute (new_query)
End
```

The Procedure Decompose() allows the decomposition of the user’s query and search exactly the key to substitute by the real encryption key.

The function Func() takes the encryption key defined in the user’s query and delivers the real encryption key.

The Function Generate() reformulates the new query with the real encryption key and subsequently send and execute this query using Execute().

The test of "Algo-client1" was performed on the "agent" table which has the following structure:

```
agent (code, first_name, last_name, dose)  (D)
```

In this table, we have defined the column "dose" as a sensitive column. It will be protected by an encryption using the AES256 algorithm and the key k = '@mysonmyson@123'. Table I and Table II show the results obtained before and after the encryption.

C. Solution 2: Insert in Multiple Sensitive Columns

The principle of protecting encryption keys in a query when inserting data in multiple sensitive columns is similar to solution 1. The main goal is to protect the encryption key of each column distinctly from the other columns key.

Let’s consider a user that executes an insert query having the following form:

```
Insert into Table I values ('0001', AES192('Mycolor442266775hrt@HH@T', 'elbouchti'), AES192('Mycoloromybeauty@nn@x', 'karim'), AES192('Mycolor012345678has@ml@p', 2000);  (E)
```

Before sending that query, it will be transformed into the query (F) having the syntax:

```
Insert into Table I values ('0001', AES192(Func_Mc('Mycolor442266775hrt@HH@T', 'elbouchti')), AES192(Func_Mc('Mycoloromybeauty@nn@x', 'karim')), AES192(Func_Mc('Mycolor012345678has@ml@p'), 2000);  (F)
```

In this case, the Func_Mc() function is defined and stored in the DB server. It substitutes the keys defined in the query (E) that comes out from the user's application into the real keys of encryption/decryption. We suggest implementing Func_Mc() in the DB server among its objects. It should follow the algorithm below:

```
Func_Mc (X)
{K = E (X, H ((Nom_table) || (Nom_col)) || (Nom_DB)) || Sum (id_col, Id_tab));
Retour (K) ;}
```

With:

• E: A symmetric encryption algorithm.

• H: A hash function.

• X: The key defined in the insert statement (user's application files)

• K: the real key to encrypt/decrypt the data of the column.

• Nom_col: The column name defined in the insert statement.

• Nom_DB: The name of DB.

• Sum (id_col, Id_tab): the sum of the column and table identifiers.

D. Implementation of the Solution 2

The implementation of the solution 2 was performed using the same tools deployed in the solution 1. We have chosen the AES256 as an encryption algorithm and the MD5 as a hash function to develop the function Func_Mc().
The algorithm "Algo-client2" presented below, transforms the user's query into a new one which calls Func_Mc() before its execution on the DB server.

**Algorithm 2:** Query transformation in solution 2.

```
Input: executed_query
Output: new_query
Begin
Decompose (executed_query)
    Func_Mc(key)
    Generate (new_query)
Execute (new_query)
End
```

We have tested the algorithm “Algo-client2” on the table “travailleur” which has the structure (G) below:

**travailleur** (Matricule, dose_prof, dose_sup, dose_neut, dose_int) (G)

We have considered all the “travailleur” columns as sensitive columns. Thus, these columns will be protected by an encryption according to the elements of the Table III and Table IV:

Table V and Table VI show the results obtained when inserting data in “travailleur” table before and after the encryption.

**E. Results and Discussions**

The results obtained when implementing our two solutions allow us to deduce that both solutions provide better protection of encryption keys, no malicious act performed by the client to hold encryption keys can succeed. The real values of keys are not defined on the client.

Outsourcing the protection of the encryption keys defined on the client side to a DB server is an effective way to secure keys, especially if the server is a trusted site. Our solutions are well adapted to this concept; they are conditioned by a high level of trust at DB server and its operators. Shmuely et al. have mentioned in [11] that the level of trust in the DB server is a fundamental criterion of a DB encryption solution. The partial trust scenario is one of the three levels of trust mentioned in their manuscript. It means that the DB server, together with its memory and the DBMS software are trusted, but the secondary storage that it uses is not. Thus, we choose to integrate our solution into this concept.

**TABLE III. THE ENCRYPTION KEYS ASSOCIATED TO THE “TRAVAILLEUR” COLUMNS**

<table>
<thead>
<tr>
<th>Matricule</th>
<th>Algorithm</th>
<th>Encryption key defined in the user query</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose_prof</td>
<td>AES256</td>
<td>&amp;@I@encrypt@my@</td>
</tr>
<tr>
<td>dose_sup</td>
<td>AES256</td>
<td>&amp;@I@encrypt@he@</td>
</tr>
<tr>
<td>dose_neut</td>
<td>AES256</td>
<td>&amp;@I@encrypt@it@</td>
</tr>
<tr>
<td>dose_int</td>
<td>AES256</td>
<td>&amp;@I@encrypt@we@</td>
</tr>
</tbody>
</table>

**TABLE IV. THE ENCRYPTION KEY GENERATED BY Func_Mc()**

<table>
<thead>
<tr>
<th>Matricule</th>
<th>Encryption key defined in the user query</th>
<th>Encryption key generated by Func_Mc()</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose_prof</td>
<td>&amp;@I@encrypt@he@</td>
<td>B0C849772BA517C 27C5237D2CF06B9</td>
</tr>
<tr>
<td>dose_sup</td>
<td>&amp;@I@encrypt@it@</td>
<td>E03D47BBED9E464 A0B0BD8B69D1F0DA</td>
</tr>
<tr>
<td>dose_neut</td>
<td>&amp;@I@encrypt@we@</td>
<td>F2AB5228D454F263 044CD4D392AD50</td>
</tr>
<tr>
<td>dose_int</td>
<td>&amp;@I@encrypt@it@</td>
<td>16307E27C51A178C B731B6F8BDC29CC8</td>
</tr>
</tbody>
</table>

**TABLE V. THE TABLE “TRAVAILLEUR” BEFORE ENCRYPTION**

<table>
<thead>
<tr>
<th>Matricule</th>
<th>dose_prof</th>
<th>dose_sup</th>
<th>dose_neut</th>
<th>dose_int</th>
</tr>
</thead>
<tbody>
<tr>
<td>A00001</td>
<td>1,2</td>
<td>10,2</td>
<td>1,2</td>
<td>0</td>
</tr>
<tr>
<td>A00002</td>
<td>2,4</td>
<td>5,47</td>
<td>0,2</td>
<td>0</td>
</tr>
<tr>
<td>A00003</td>
<td>6,5</td>
<td>7,21</td>
<td>0,8</td>
<td>0</td>
</tr>
<tr>
<td>A00004</td>
<td>2,8</td>
<td>3,55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A00005</td>
<td>1,1</td>
<td>4,7</td>
<td>0,22</td>
<td>0</td>
</tr>
<tr>
<td>A00006</td>
<td>6,5</td>
<td>8,6</td>
<td>0,14</td>
<td>0</td>
</tr>
</tbody>
</table>

**TABLE VI. THE TABLE “TRAVAILLEUR” AFTER ENCRYPTION**

<table>
<thead>
<tr>
<th>Matricule</th>
<th>dose_prof</th>
<th>dose_sup</th>
<th>dose_neut</th>
<th>dose_int</th>
</tr>
</thead>
<tbody>
<tr>
<td>372E7E32A13F7DB9F7FCB9FBE3CB3ED3D</td>
<td>FAA8DB86C358211C261058A7AEB13CC4</td>
<td>48F105C92DBA3DDC5913C82A9B810BF0</td>
<td>0547D9BD3A57B0C1C9 9EC1805247CE</td>
<td>8B594F1494DFEBB85FE 228B16EAE527F</td>
</tr>
<tr>
<td>676D2CP9Y450ECPBF034CECOA1B6FG66</td>
<td>29C9A2C18AD74B3212B5F952B85C5FEE</td>
<td>0C480ADBEF9CCB14C43977C517F4053C</td>
<td>EF1E90ED97E2112BE602E 1B0E688A09</td>
<td>8B594F1494DFEBB85FE 228B16EAE527F</td>
</tr>
<tr>
<td>6259798233E8455434A204719A83198</td>
<td>BB5F6D4CFD91E1AC436BCA4462DDA</td>
<td>F54676102D30CDD54E824F5D18A4G5SB</td>
<td>BBD0714D3B5D58E39D 9F2E338EBBB</td>
<td>8B594F1494DFEBB85FE 228B16EAE527F</td>
</tr>
<tr>
<td>04AE3E21C176BFCD71EE3BA7D8BCB3</td>
<td>786033AD99863E08F319D8936E5A06B16</td>
<td>5B2FCACOF7418386B5AFB26502096</td>
<td>94AA0C85D49D3DF6CC2F4CB560CD</td>
<td>8B594F1494DFEBB85FE 228B16EAE527F</td>
</tr>
<tr>
<td>4197BE9154135B56BD1B93590731</td>
<td>3F4A4C64C8873E2E01A0C43D24203D6</td>
<td>6DF891B76A0CA0B8851BDFEBFB1B2433</td>
<td>7CF237E8291374EF6980A67E9B0645E</td>
<td>8B594F1494DFEBB85FE 228B16EAE527F</td>
</tr>
<tr>
<td>586F83DB4D84EBB80B2FB</td>
<td>BB5B6D4CFD91E1AC436BCA4462DDA</td>
<td>DAA03A0B62D6A9B9C902788A95152A6D</td>
<td>F75A6E3E9A188328CAB32A1AP9A342D</td>
<td>8B594F1494DFEBB85FE 228B16EAE527F</td>
</tr>
</tbody>
</table>
The proposed solutions are well adapted for 2 and 3 tier architectures. For each architecture, the key protection functions (Func (i) and Func_Mc (i)) are implemented inside the DB server. When we choose to manage the DB inside a Cloud (DBAAS), the functions (Func (i) and Func_Mc (i)) will be included among the outsourced DB objects.

Otherwise, these solutions can be compromised by a specific attack, which remains the retro-engineering attack. An attacker can reveal the source code from the executables of the solution files (Algo-client1 and Algo-client2) by using special tools. The attacker can get the return of the functions and can determine how these functions provide the new keys. In this case, we suggest putting all the code of the user's queries in stored procedures, which will also be placed on the DB server.

IV. CONCLUSION

Databases are the favorite targets of attackers due to the values of the information they contain and their volume. They are vulnerable to several typologies of attacks, especially the internal attacks. When implementing an encryption strategy at the application level to secure DB, the internal attack threats become more dangerous for the sensitive DB, and the probability of attacking the encryption keys is considerable. The attacker can use them to decrypt all sensitive data without leaving any traces.

In this work, we have proposed and implemented two solutions that protect encryption keys inside applications. The principle of these solutions is to transform the encryption keys defined at the user's query to other keys considered the real keys for encryption and decryption. This transformation is performed using functions stored within DB server. Our solution is considered efficient and simple when implementation, and it is more adapted to a trusted DB server. In forthcoming works, we aim to adapt our solutions to cover more complicated attacks made by the internal attackers on sensitive DB.

REFERENCES


