Migration Model for unsecure Database driven Software System to Secure System using Cryptography

Mr. Mahendra Kumar Shrivas  
Head Of Department Information Technology  
Academic City College, Kumasi, Ghana  
mahendra.shrivas.gh@ieee.org

Dr. Augustine Amoako  
HOD of Accountancy and Accounting Information Systems, Kumasi Polytechnic, Kumasi, Ghana  
wutispe@yahoo.co.uk

Mr. Samuel Odame Boateng  
Head Of Department, ICT  
St. Joseph Seminary Senior High School, Mampong - Ashanti, Ghana  
odameph@gmail.com

Dr. Thomas Yeboah  
Head Of Department, ICT  
Christen Service University College, Kumasi, Ghana  
thomyebs24@gmail.com

ABSTRACT: Most of the software systems are having role based operation model where each user, based on their job role having some functionality to perform. Role based access privileges are basic security implementation in most of the database driven software system. Users or Operators of the system enters the records into the software system using graphical user interface (GUI), which is stored into the database after data validation.

Generally administrators are having all privileges and can perform all system and functional operations. Administrator can be divided into following categories:-
1. Software System Administrators
2. Network Administrators
3. Server Administrators
4. Database Administrators

Network, Server & Database administrators are more powerful than software system administrator as they are having full privileges and can able to do changes in the systems which is almost untraceable unless complete system audit is performed to trace the mismatch of manual record and software system record.

Data in the databases are unsecure as confidential organization record is stored on servers in unencrypted form, which is not secure from insider and outsider attack.

This research work shows how to protect data confidentiality even when attackers get access to all the data stored on servers. Also authors are proposing a migration model which can be used to secure existing unsecure database driven software system.

As a case study authors have taken an existing software system and applying AES based encryption and decryption with the key initialization vector (IV) and Code Block Chaining (CBC) mode with PKCS 5 padding because it has a very high security performance.

KEYWORDS: Algorithms, Authentication, Cipher, Cryptography, CBC, PKCS5, Encryption, Decryption, Database Security, AES

I. INTRODUCTION

Nowadays there are exclusive headline stories in the newspapers almost every day millions of data being leaked or computer systems are being hacked by cyber thieves. The cyber thieves are unknown mysterious people who perform their cyber criminal activities from remote locations without being tracked. They are having expert knowledge of software systems, databases, computer networks and operating system's tools. They find any loophole into the victims computer network and system and then initiate their anonymous activities.

Department of Defense (DOD) was also hacked once. Hacker were able to download some of personal and financial information of their employees[2]. It is not just about DOD almost all major tech giants had been hacked including Apple, Sony, eBay, Amazon, Google, Microsoft, etc.

It does not matter how well you have implemented all security policies, there are equally chances of being hacked by cyber criminals. It is just a matter of discovering or invention of a technique to break down the security systems.

Leakage of confidential data plagues many computing systems today. For example, last year marks a peak in data breaches: about 740 million records were exposed, the largest number so far.[7] Some of the biggest data breaches of 2014 [8][4] and 2015 [1][10][5] are listed below:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Organization Name</th>
<th>People Affected (Millions)</th>
<th>Stolen Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>eBay</td>
<td>145.0</td>
<td>Encrypted Passwords, Unencrypted Customer details</td>
</tr>
<tr>
<td>2</td>
<td>Target</td>
<td>110.0</td>
<td>Credit and debit card details along with Customer details</td>
</tr>
<tr>
<td>3</td>
<td>JPMorgan Chase</td>
<td>76.0</td>
<td>Names, Addresses, Phone Numbers and Emails</td>
</tr>
<tr>
<td>Reported Month Year</td>
<td>Organization Name</td>
<td>People Affected (Millions)</td>
<td>Stolen Information</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>January 2015</td>
<td>Morgan Stanley</td>
<td>3.5</td>
<td>Personal Information, Debit/Credit Card and Account Information</td>
</tr>
<tr>
<td>January 2015</td>
<td>MyTF1</td>
<td>2.0</td>
<td>Personal Information, Debit/Credit Card and Account Information</td>
</tr>
<tr>
<td>January 2015</td>
<td>Topface Russian dating</td>
<td>20.0</td>
<td>Personal Information</td>
</tr>
<tr>
<td>February 2015</td>
<td>Anthem, Inc</td>
<td>78.8</td>
<td>Personal Information, Medical Information</td>
</tr>
<tr>
<td>February 2015</td>
<td>Uber</td>
<td>50.0</td>
<td>Driver Information</td>
</tr>
<tr>
<td>March 2015</td>
<td>Premera BlueCross BlueShield</td>
<td>11.2</td>
<td>Personal Information, Health</td>
</tr>
</tbody>
</table>

Table 1: Biggest Data Breaches of 2014

<table>
<thead>
<tr>
<th>Reported Month Year</th>
<th>Organization Name</th>
<th>People Affected (Millions)</th>
<th>Stolen Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 2015</td>
<td>Office Of Personnel Management</td>
<td>4.2</td>
<td>Federal Worker’s Details, Social Security Numbers</td>
</tr>
<tr>
<td>May 2015</td>
<td>CareFirst BlueCross BlueShield</td>
<td>1.1</td>
<td>Personal Information, Health</td>
</tr>
<tr>
<td>May 2015</td>
<td>Office Of Personnel Management</td>
<td>21.1</td>
<td>Federal Worker’s Details, Social Security Numbers</td>
</tr>
<tr>
<td>May 2015</td>
<td>Gaana.com</td>
<td>10.0</td>
<td>Personal Information, User Login Information</td>
</tr>
<tr>
<td>May 2015</td>
<td>mSpy</td>
<td>4.0</td>
<td>Personal Information</td>
</tr>
<tr>
<td>May 2015</td>
<td>Internal Revenue Service (IRS)</td>
<td>1.0</td>
<td>Personal Information, Account Information</td>
</tr>
<tr>
<td>June 2015</td>
<td>Army National Guard</td>
<td>8.5</td>
<td>Personal Information, Social Security Numbers</td>
</tr>
<tr>
<td>June 2015</td>
<td>Japan’s National Pension System</td>
<td>1.3</td>
<td>Personal Information, Account Information</td>
</tr>
<tr>
<td>June 2015</td>
<td>US Office of Personnel Information OPM.Gov</td>
<td>32.0</td>
<td>Personal Information, Account Information</td>
</tr>
<tr>
<td>June 2015</td>
<td>health insurance company, Premera Blue Cross</td>
<td>11.0</td>
<td>Personal Information, Medical Information</td>
</tr>
<tr>
<td>July 2015</td>
<td>Hacking Team</td>
<td>1.0</td>
<td>Security tool, Email Details of Client and Client’s Target</td>
</tr>
<tr>
<td>July 2015</td>
<td>St. Francis Health / Medical Informatics Engineering</td>
<td>1.5</td>
<td>Personal Information, Medical Information</td>
</tr>
<tr>
<td>July 2015</td>
<td>UCLA Health System</td>
<td>4.5</td>
<td>Personal Information, Medical Information</td>
</tr>
<tr>
<td>August 2015</td>
<td>Ashley Madison</td>
<td>37.0</td>
<td>Personal Information, Email Log</td>
</tr>
<tr>
<td>August 2015</td>
<td>Carphone Warehouse</td>
<td>3.4</td>
<td>Personal Information, Payment Card Details</td>
</tr>
</tbody>
</table>

Table 2: Biggest Data Breaches of 2015
II. PROBLEM IDENTIFICATION

Most applications store sensitive data on servers, so preventing data leakage from servers is a crucial task towards protecting data confidentiality. In fact, if computer systems are not connected to the Internet then also these systems are not secure. Organization’s computers which operate in LAN environment frontage the threat of hacking from inside attacker. Employees or former employees who have access of the systems are also treat to the security of an organization's computers and networks.

Currently most of the existing computerized application are web based and operates in LAN or Internet environment. These applications helps in effective and controlled running of the operations such as storing, mining, and centralizing of the activities of an entire organization.

Authors have observed that most of the systems are role based. As per their roles access permission is given into the software system. Generally in the software system which operates in LAN or Internet environment can have following administrator level users apart from software system administrator user :

1. Network Administrators
2. Server Administrators
3. Database Administrators

Unfortunately, the existing system has some security issues ranging from an intruder, Insider and Administrators which all of these persons gain access to a Computer system and try to extract valuable information. Even though the Administrator has the privileges to administer a computer system, yet uses his Administration rights in order to extract valuable information does not auger well. For example there has been an instance where some people with administrative rights have tampered with and made changes to sensitive data unlawfully through the back end of the System.

Authors have observed the pattern used in biggest data breaches of 2014 and 2015 is usages of malwares and malicious software, which is used to sniff the data packets between the communication of applications and databases, sometime the tools sniff data directly from input sources.

In most of the cases attack start with a phishing email, once attackers received any reply from victims, they fetch network and system information from email headers which lead to series of more sophisticated attacks. Once the hackers have control over target computer system, they create new user accounts or use existing accounts and get administrator privileges. After getting administrator rights they start stealing information they need. In authors point of view administrators are the biggest threat for the Software System security. (TechTarget, 2015)

III. LITERATURE REVIEW

The techniques needed to protect data belong to the field of cryptography. Cryptography is a study of secret (crypto-) and writing (-graphy). It is the science or art of encompassing the principles and methods transforming message into some coded form and then transforming that coded message back to its original form. As the field of cryptography has advanced; cryptography today is assumed as the study of techniques and applications of securing the integrity and authenticity of transfer of information under difficult circumstances.

Actually, the subject has three terms cryptography, cryptology, and cryptanalysis, which are often used interchangeably. Technically, cryptology is the all-inclusive term for the study of communication over non secure channels.

The process of designing systems to do this is called cryptography. Again, Cryptanalysis is the procedures, processes, and methods used to translate or interpret secret writings or communication as codes and ciphers for which the key is unknown.

A. Cryptographic Algorithms

The classes include Symmetric Cryptography and Asymmetric Cryptography.

B. Symmetric Cryptography

Symmetric Cryptography is the most traditional form of cryptography. The scheme uses the similar key to encode and decode of information. Thus, symmetric cryptography involved parties that share a common secret (password, pass phrase, or key). Data is encrypted and decrypted using the same key. These algorithms tend to be comparatively fast, but they cannot be used unless the involved parties have already exchanged keys.

C. Common Symmetric Algorithms

Common symmetric algorithms are listed below along with their ratings[11] :-

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description</th>
<th>Key Length</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blowfish</td>
<td>Block cipher developed by Schneier</td>
<td>1-448 bits</td>
<td>L</td>
</tr>
<tr>
<td>DES</td>
<td>DES adopted as a U.S. government standard in 1977</td>
<td>56 bits</td>
<td>§</td>
</tr>
<tr>
<td>IDEA</td>
<td>Block cipher developed by Massey and Xuejia</td>
<td>128 bits</td>
<td>L</td>
</tr>
<tr>
<td>MARS</td>
<td>AES finalist developed by IBM</td>
<td>128-256</td>
<td>Ø</td>
</tr>
</tbody>
</table>
Key Cryptography uses two keys in the encryption system. Some of asymmetrical encryption algorithms are listed below:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Developed by</th>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC2</td>
<td>Block cipher developed by Rivest</td>
<td>1976</td>
<td>Generate a shared secret key to secure information exchange</td>
</tr>
<tr>
<td>RC4</td>
<td>Stream cipher developed by Rivest</td>
<td>1976</td>
<td>Data encrypting and signing of data</td>
</tr>
<tr>
<td>RC5</td>
<td>Block cipher developed by Rivest</td>
<td>1994</td>
<td>Similar to RSA but for small device</td>
</tr>
<tr>
<td>RC6</td>
<td>AES finalist developed by RSA Labs</td>
<td>1994</td>
<td>Transmission of digital signatures and key</td>
</tr>
<tr>
<td>Rijndael OR AES</td>
<td>NIST selection for AES, developed by Daemen and Rijmen</td>
<td>1994</td>
<td></td>
</tr>
<tr>
<td>Serpent</td>
<td>AES finalist developed by Anderson, Biham, and Knudsen</td>
<td>1994</td>
<td></td>
</tr>
<tr>
<td>Triple-DES</td>
<td>A three-fold application of the DES algorithm</td>
<td>1994</td>
<td></td>
</tr>
<tr>
<td>Twofish</td>
<td>AES candidate developed by Schneier</td>
<td>1994</td>
<td></td>
</tr>
</tbody>
</table>

Key to ratings:

W) Excellent algorithm. This algorithm is widely used and is believed to be secure, provided that keys of sufficient length are used.

L) Algorithm appears strong but is being phased out for other algorithms that are faster or thought to be more secure.

§) Algorithm appears to be strong but will not be widely deployed because it was not chosen as the AES standard.

Table 3: Symmetric Algorithms

In encryption and decryption process the security key is very important. Encrypted data is stored in the server which can be access by the attacker but without security key they cannot do anything with encrypted data unless they invest huge amount of computer resources and time. Thus the security key should be stored offline [7] [9].

E. Existing Software System Specification

There is an existence School Management System developed by authors which is deployed at St. Joseph Seminary Senior High School, Kumasi. It is an Enterprise Resource Planning (ERP) System which is build using Java Swing and used MySql 5.0 as database server. The SMS consists of more than ten modules which include: Registration module, Account module, Transaction module, Exams module, Attendance module, Transcript module, HR modules and so on.

The SMS Software System was developed using MVC design architecture along with DAO layer for housing all the modules. For modular application development Model View Controller (MVC) is the first choice as it divides application into three independent reusable parts called; the model, the view and the controller.

The SMS is based on client and server technology and having more than 15 client nodes connected with MySql database server via LAN.

It is an undisputable fact that the existing system poses less security threats to attackers or hackers. If someone hacks into the database of the system, various vital areas and sensitive parts will be exposed because in database tables, data is being stored as plain text.

F. Algorithms Consideration

As described in table 3 AES is excellent algorithms as compare to other symmetrical algorithms.

The AES256 is a symmetrical encryption algorithm used in this process. In fact this algorithm was selected due to the high standard of the security. Because the length of the key which is 256 bits and the number of hashes [3], it takes a murderously long time for a malware hacker to perform a dictionary attack. So the AES256 is very difficult to crack and it's ciphertext is almost invulnerable to attack when properly implemented.

Also with the PKCS5Padding, it is a method where extra bits are added in the plaintext which makes desire...
size of data block for encryption. This helps to increase the brute force strength which ensures good security[14].

Therefore to arrest this situation authors are applying AES 256 bit based encryption and decryption with the key Initialization Vector and Code Block Chaining mode with PKCS5 padding because it has a high security performance.

Organizations and System developers can use custom build, well tested and more stronger encryption algorithms to ensure higher data security.

IV. MIGRATION MODEL TO SECURE DATABASE DRIVEN SOFTWARE SYSTEMS

After observing various database driven software systems authors have developed migration model which can be used to secure any existing unsecure database driven software systems which have developed using modular approach. Authors have proposed following steps in the migration model:-

1. Affected Table Identification and Justification
2. Field consideration for encryption
3. Column data type modification
4. Padding Factor and Initialization Vector
5. Encryption Key Consideration
6. Encryption Process
7. Decryption Process
8. Re-Assessment

1. Affected Table Identification and Justification
This section describes the affected tables which will ensure the security and purpose of the Implementation of the encryption. Proper analysis should be carried out in identification of table for encryption. In case of SMS, Course table cannot be chosen for encryption as information about course offering by schools is already in public domain.

In view of that, Student, Employee, Fees, Account, Transcript, Transaction, Exams Attendance, Payroll and Login tables have been identified for the encryption.

2. Field Consideration for Encryption
After identification of the tables, fields need to be identified for encryption. All the information of the table should not be encrypted because of encryption total length of the data may increase because AES operates on 8 bytes plain text data block, so if given data for the column is less than 8 bytes system should add some padded information to make it multiple of 8 bytes. Also primary key, foreign key and index key field should not be chosen for encryption as these fields are vital for proper functioning of database systems. Encryption of these fields may cause information mismatch which leads to violation of system integrity.

3. Column data type modification
Encrypted data stored in the table’s column, that is why there is a need of alteration of the data type of the column. The cipher text resulting from encryption is in binary form. To store this text column data type should be any one of following :-

- BLOB
- VARBINARY
- BINARY
- CHAR BIT DATA
- VARCHAR BIT DATA

4. Padding Factor and Initialization Vector
In AES algorithm, there are different modes of operations which include; ECB, CBC, OFB, CFB, CTR, and XTS. Here Authors used CBC mode with PKCS5 padding. Which have information service such as secrecy and validity. If plaintext is larger than required block size there is a need to apply cipher's single block operation constantly. This mode adds extra bytes to plaintext so that it can be converted into the required block size which is multiple of 8 bytes[13]. The procedure of applying this operation is described by mode of operation algorithm. The PKCS5 padding follows the following rules[13]:

- The number of bytes to be padded = 8 – (number Of Bytes (clear Text) mod 8).
- Depending on the length of the clear text data 1 to 8 bytes will be added to the clear text data.
- All padded bytes have the same value - the number of bytes padded.

However, for each encryption operation most of the modes require a unique binary sequence which is called initialization vector (IV).

The Initialization vector (IV) or starting variable (SV) is a block of bits. Several modes use it to randomize the encryption. By this randomization it produces distinct cipher texts even if the same plaintext is encrypted multiple times, without the need for a slower re-keying process. To make it reality in the PKCS5 Padding, a block cipher works on fixed block size units, so the length of coming data or messages are variable. And the CBC mode used concludes that the block must be padded before encryption. In padding method, extra bits are added in text message to make require size of block for encryption. Several padding schemes exist. In simplest padding scheme Authors added null bytes to the plaintext to bring its length up to a multiple of the block size.

5. Encryption Key Consideration
Encryption keys are the backbone of security of the System. AES-256 bit encryption key is proposed by the authors as it is relative strong. Hacker should have to invest huge amount of time and computer resources to
crack the password using various techniques like brute-force attack, dictionary attacks.
While choosing the encryption key following points need to be consider:

1. Key should follow the rules of strong password
2. Key length must be as per algorithms used
3. Key should not have any dictionary words
4. Key information should be stored in secured location. System which is not connected with the Internet or LAN is preferable.

6. Encryption Process
The Figure 1 below shows the Encryption Processes.

![Figure 1: Encryption Process](image1.png)

**Explanation**

**Step 1**
User supply message into the available fields.

**Step 2**
User click Button to activate Event from view.

**Step 2a**
Controller Request some data from model or populate form data into model.

**Step 2b**
Data from Controller is communicated and handled by DAO.

**Step 3a**
Data from DAO is processed and encrypted with AES - 256 coupled with PKCS5Padding.

**Step 3b**
After encryption, the encrypted data is then send to the Database for Storage and future retrieval.

**Method for Encryption**
Procedure can be step wised explained as follows:-

**Step 1.** Read the input from model objects.

**Step 2.** Input the AES key

**Step 3.** Perform the AES logical operations

**Step 4.** Generate ciphertext.

**Step 5.** Finally, return the encrypted data.

7. Decryption Process
The Figure 2 below shows the Decryption Processes.

![Figure 2: Decryption Process](image2.png)

**Explanation**
When the user query the system or click on the Search button for retrieval the followings activities take place.

**Step 1**
Data is fetched from the database.

**Step 2**
Data pass through DAO to AES-256 for decryption where the PKCS5Padding is remove.

**Step 3**
The decrypted data is displayed in model section.

**Step 4**
Data is then sent to the controller.

**Step 5**
Finally, Data is shown at the view section.

**Step 6**
The output is shown to the user.

**Method for Decryption**
Procedure can be step wised explained as follows:

**Step 1.** Read the cipher text from the database

**Step 2.** Input AES Key.

**Step 3.** Remove AES-CBC-PKCS5 padding from the cipher text.

**Step 4.** Perform Decryption AES logical operations

**Step 5.** Return the generated plaintext

8. Re-Assessment
Re-Assessment is the last but not the least step of the proposed migration model. In this phase one should
verify that whether proper process had been followed in
the migration of database to apply encryption.

This phase also covers the verification of tables and
column which has been chosen for encryption. Logical
reasoning, details discussion and consultation from
security specialist should be done before finalization of
information that needs to be encrypted. Unless and until
the information is very important implementers are not
advised to encrypt the field because this may increase
disc requirement and also speed of the system is little
bit slow due to encryption and decryption process.

V. SYSTEM IMPLEMENTATION

The School Management System consists of four main
parts:

1. Java Swing based frontend to input and
display the data.
2. Java Class which uses Java Cryptography
Architecture (JCA) APIs for encryption and
decryption of sensitive data[6].
3. Controller, Model and DAO layer which uses
various methods for inter communication
4. MySql database tables to store the encrypted
data

When Add button is clicked while application is in
execution, after front end data validation it takes values
from various text fields and a message will display to
signify that the data has been transferred from the view
(StudentForm) to database via Controller and DAO
class after encryption. Controller populate data from
view (Figure 3) to model (Student) after population
controller creates and object of DAO (StudentDAO)
and passes model object to DAO class’s function where
AES encrypt() is triggered to encrypt the data through to
Model (Student) before it is finally store into the
Database.

Now, to make it a reality and affirmation, the Figure 4
below shows the detail of the recent student data stored
with the Roll Number JOS1 from the backend of the
project which is MySql Database.

Moreover, for decryption, Search on the Student Form
will be used. So, if the Search button is clicked while
application is in execution, it takes the command or the
parameter based on the values from the search input
fields which is transferred from the view (Student
Form) to Controller (StudentDAO) where AES is
initiated to decrypt the data which is coming from the
database. Decrypted data is stored into the model
(Student) and return to the Controller from DAO layer.
The data is finally sent back and displayed at the
various text fields on the View (Student Form). See
that in the Figure 5 snapshot.

Finally this is the completion of the encryption and
decryption in Java with the AES algorithm using Model
View Controller Architecture.

However system can be developed using any
programming languages, for any Operating System and
for any application which uses database.

VI. RECOMMENDATIONS AND
CONCLUSION

Data breaches are the biggest technical hits of 2014-
2015. One such incidence happens and whole
organization business collapse within an hour. The impact of such incidence can be seen in company's stock values. Latest data breach of Ashley Madison took some precious human lives which cannot be recovered by compensation or by any mean. One can rebuild business but trust of millions of valuable customers cannot be rebuild. The trust which took lots of continues efforts and years of effective business services cannot be regain easily by just saying we are sorry we were unable to protect your sensitive information.

Organizations have to take responsibilities and have to be more accountable to the customers so that valuable organization's data and customer information can be secure to avoid any such disaster.

Authors believe that this migration model could help those organizations, who are using unsecure database driven software system. This migration model can be adopted for various software systems which had implemented using different programming languages, databases and operates on any operating systems.

VII. REFERENCES


