IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Secure Unique Identification using Encrypted Storage in NoSQL Database

Anand Shende[1], Omkar Gurav[2], Swapnil Shirode[3], Piyush Govekar [4], S.N.Zaware [5]

Student, Computer, AISSMS IOIT, Pune, Maharashtra, India[1,2,3,4]

Asst. Professor, Computer, AISSMS IOIT, Pune, Maharashtra, India[5]

Abstract: Database helps us to collect, retrieve, organise and manage the data. The database used in the project is used to associate user's personal documents together and keep it available for the user anytime he requires. The requirement of the project were to have a unique identity. To satisfy the requirement Adhar Card was used. This is used because there is only one card per person. The database is going to record all sensitive information of the user, so there is a need to provide security to it. This is done using encryption. Various algorithms were studied in the paper, the one suitable out of them was ETSFS (Extended Transpose Substitute Folding Shifting). NoSQL is chosen because it satisfies the need of storing images as well as text efficiently, compared to SQL database. There has been a study as to where all an encryption process should take place. The encryption/decryption here is provided on the Client Application end. This project should provide a unique identity for the database which is secured using encryption. This will help the user to keep the personal information_online and not carry it around everywhere physically.

Keywords: NoSQL database, ETSFS algorithm, Application side Encryption.

I. INTRODUCTION

This project is an implementation of the NoSQL database. The NoSQL database is used because it has to store all sensitive and personal information in the database, which would not only include textual data but also images. This won't be easy with structured language.

The NoSQL database used is MongoDB. MongoDB has many features which supported the system to be implemented. One of them being the GridFs for storing files and other would be the scalability. Survey shows the attacks on MongoDB database. (Source: informationage.com, thehackernews.com).

MongoDB currently provides security in form of authorization, authentication, encryption (TLS and SSL), auditing. To provide better security from above mentioned attacks the implementation of ETSFS was necessary. As the encryption and decryption was on the client side a lightweight encryption algorithm was required to be implemented hence ETSFS.

Whenever a question arises about providing security to databases. The usual answer is to use Encryption and Decryption. There are many Encryption techniques, the standard ones are DES (Data Encryption Scheme), AES (Advanced Encryption Scheme) and their variants. But they have been cracked in the past, there is a time for change, so ETSFS (Extended Transpose Substitution Folding Shifting) was used.

The project provides security at two levels- first is data-in-motion and the other is data-at-rest. The data is going to be sent from the Client to Server, before sending this data, it will be encrypted, and this is Data-in-motion protection. The data is stored as it is in the database, without decryption, and this is known as Data-at-rest protection. If every user was given a unique identity then it would be easy to maintain data. This job was done by the Indian Government by giving the Adhar Card. Using this Adhar Card, its UID (Unique Identifier) can be used by the user to save all the personal information. For example, the medical records of each person can be associated with this UID. Also that he can have more than one medical record for the same UID.

Whenever we go out for shopping, we have to carry many cards (i.e.-credit and debit) along with us. Implementing this project will taper off the need to carry a number of cards. All you would need is just an image of the Adhar's QR Code.

II. BACKGROUND

The three main properties of database security are –
1. Confidentiality - this means that we need to protect personal data so that nobody else can interfere with one's life (this is an exception in the case of doctors).
2. Integrity this means that nobody without authority can change or modify the information in the database.
3. Availability this means that the information should be available always.

There are four types of flows according to [1], which are to be taken care of to provide database security –
1. Access control this give access to only some of the people that have the rights to get access to the database

DOI 10.17148/IJARCCE.2016.54124

and the information within. If the access control rights are given in the wrong hands, the effects can be hazardous.
2. Information flow control If the flow of information is not proper then this gives the intruder a chance to latch up on the information and make wrong use of it.
3. Cryptographic flow control this flow is the flow of encrypted data in the network.
4. Inference control in this type of flow it is taken care that the information reaches only the required person and with given rights.

There are two aspects to the Encryption process which have to be discussed in detail-
1. Algorithm.
2. Location where algorithm is implemented.

A few Algorithms that were studied in paper [2] are -
1. DES:
DES is an acronym Data Encryption Standard, operating on 64 bits using a secret key which is of 56-bit long. A key is chosen randomly. The same key is used to encrypt the message and the same is used to decrypt it. Six bits are mapped to groups of four bits. It is done in 16 rounds.

Process of DES encryption:-
- The input key is used to obtain sixteen 48-bit keys. These are used as sub keys. Each sub key is used in each round.
- It is expanded from 32 bits to 48 bits using another fixed table.
- The result is combined with the sub key for that round.
- The 48 resulting bits are then transformed again to 32 bits. In the next round, combination is used as the left part.

`

2. TRIPLE DES:
Triple DES is the advancement of DES and covers the problems of DES. It uses three 56-bit DES keys, which creates a key with the length of 168 bits. The key is then split into three same length keys.

The process is-
- First key is for encryption.
- Second key is for decryption.
- Third key is for another encryption.

3. RSA
Stands after the name of developers: Ron Rivest, Adi Shamir, and Leonard Adleman. RSA is asymmetric, which means that the same key for encryption cannot be used for decryption. It uses a combination of public and private keys.

The process of creating keys in RSA is as follows:
- Two prime numbers are taken, *p* and *q*. 'n' is given by p multiplied by q. Considering '*e*' should be greater than 1. *Gcd (e, (p-1), (q-1)) = 1.*
- Then find the multiplicative inverse 'd' of *e* modulo *(p-1) (q-1)*. '(*n*, *e*)' is the public key. 'd' is the private key.

4. AES – Advanced Encryption Standard
AES varies between 128, 192 and 256 bits. AES is symmetric. The AES encryption process: (1) Key Schedule and (2) Encryption. The key schedule generates several sub keys.

Steps taken to encrypt are:
- ***SubBytes.***
- ***ShiftRows***.
- ***MixColumns***.
- ***AddRoundKey.***

The standard algorithms such as DES and AES. They have been used for a number of years now. They are easy to implement. But the disadvantage is that they have been cracked from time to time. Their variants were also used to get some change but then even they were cracked. So using them any further can increase the threats.

The paper [3] studies that the AES algorithm takes much more time to encrypt and decrypt than it should. So they proposed an algorithm that could make this process faster. This was successfully shown by them. The advantages of this algorithm were that it took less time for encryption and decryption. But this made the algorithm complex.

The paper [4] says that we can use the ASCII value of each character and then encrypt it. This algorithm employed a key, which was used to encrypt another key called as a secret key. The encrypted secret key was used to encrypt the text. This gave a cipher text. This algorithm successfully implemented the encryption using ASCII values for characters. But the disadvantages of this were that the key had to be same as the length of the text and this key had to be entered by the user. So this made the algorithm inflexible.

The paper [5] transforms the previous algorithm and makes it automatic and flexible by making the key generate randomly. This gives us the flexibility to keep the text of any size.

The most suitable algorithm was found in paper [6], ETSFS – Extended Transposition-Substitution-Folding-Shifting. This covers the disadvantages of all the above algorithms. The steps are –
1. Transpose.
2. Substitution.
3. Folding.
4. Shifting.

The paper [1] presented a survey on papers based on database encryption. The paper [7] discusses a different strategy in which the database is encrypted end to end. Usually in database encryption the data residing in the databases is encrypted, but here the database scheme is itself encrypted. This is termed as Mixed Cryptography by the authors. It is seen that the security level has increased while the complexity and time required has also increased. This is the downfall of the strategy. The algorithm used for

DOI 10.17148/IJARCCE.2016.54124

this type of encryption is any symmetric algorithm. This project considers the server to be a multi-party server.

There are three places where encryption is implemented-
1. Server
2. Trusted Party
3. Client

This technique provides three tier encryption and is hence very secure. This also encrypts the queries at the client end which makes transmission secure as well. But this makes the performance of the queries low. Also here access control methods are not defined.

The paper [8] studies the encryption technique used in Microsoft SQL Server 2008. It provides security to tables, table space and columns. Before this we couldn't provide security to the databases that resided on removable media. This technique has the capability to provide security not only to the databases that reside in the internal devices but also on the external devices. This meant that databases on floppy, CD or Hard Disks can be protected in the same way. It uses a Master Key that was used to encrypt the key used by the client, this encryption gave rise to a certificate. A certificate is a digital mark or a signature that is used by the receiver to recognise who the sender. So the sender can make his identity be known. This will help the server recognise who the sender is and can make the decision of whether the databases should be shown or no. The Transparent Data Encryption consists of:-
1. Authentication- Every user is known and registered to the Server. Only recognised people can use this functionality.
2. Validation- The server must validate the request of each client. This makes the process synchronised.
3. Data Protection- The data of each client is protected.

This technique eliminates the problem of illegal access. Nobody can access and lay hands on your data without your permission. The cost of user management is efficiently handled. This reduces its cost. The privacy management is also maintained. This does not provide encryption across communication channels as the previous paper provided. This can highly effect affect the security across the network as the data is completely vulnerable. Not optimal for sensitive data. The database can never be opened without the certificate from the client. This increases the delay and not only that it makes the cost high because we need to send not only information we need (i.e. the request) but also the certificate. Extra information has to be sent. And also that the certificate can be easy modified and altered. This is very difficult to maintain then.

The paper [9] explains that we can use encryption no only on the database or storage but also on the application level. The encryption in this paper is applied on-
1. Storage level
2. Database level
3. Application level.

The advantage is that the security is maintained and it cannot be tampered with. The encryption keys can also be kept hidden and never be exposed to the outside world. The only disadvantage is that the algorithm is very complex.

The goal of the paper [10] is to design a database system that can be encrypted as well as its performance is high. The advantages were
1. Fast indexing operation
2. Low decryption overhead

The disadvantages were
1. Complex algorithm
2. Costly

The method used is Fast Comparison Encryption. The encryption is performed at the Data Ware House level.

III. PROPOSED SYSTEM

The main objective of the project is to make use of the unique identifier provided by the Indian Government. The objective is also to make this identifier applicable in day to day life applications such as in the field of medical health or to help make transactions. This identifier clubs multiple medical reports and can be accessed from anywhere and at any time. We can also make transactions using this UID.

A. *Architectural diagram*

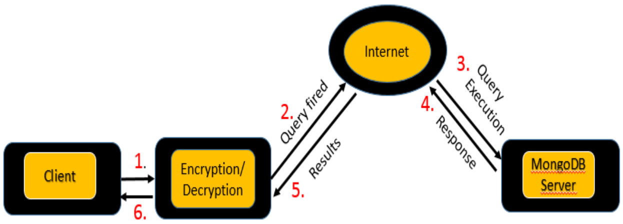


Fig-1 Architecture diagram

The given fig-1 is architectural diagram of our proposed system. As shown in given figure our system is divided into two parts
1. Server Side
2. Client Side

from the above figure, it is cogent that only the Client can encrypt and decrypt the data. We use SSL protocol to transfer data. The server uses JSP and is connected to the database via a JDBC (Java Database Connectivity). Whenever we take the data it is encrypted and is represented by '%'. The unencrypted or plain text is represented by '*'. The S stands for the query.

Across the SSL we have the same combination of query and data as S and %. But in the SSL, the query and data are % and %. So data remains consistent at the start and at the end of SSL. The data is stored in % form in database. Whenever we retrieve from the database, then only it is decrypted at the Client end.

B. System Description

The user first has to fill the registration using his Adhar Card. We'll first check whether an account has been created for this Adhar Card. If yes then an account is created and then personal information can be coupled with it. There are two types of users-

1. User
2. Administrator

The user has the rights to –

1. Insert
2. Update
3. Retrieve

The admin have tasks like-

1. Upgrade
2. Maintain

IV.METHODOLOGY

1) Client:

The client is an end from where he can interact with the Server. For doing so, he will have to first register with the Server. After creating an account he can couple documents with the account. After creating the account he can update, retrieve, insert and delete the documents coupled with the account.

Server:

The server side uses a Servlet. It takes the data from the client and stores it as it is in the database without decrypting. The Server understands the query from the client and works accordingly to reply back with the answer to the query or to store the data on insert option. It can also delete upon the delete query initiated by the client.

2) SSL (Secure Socket Layer Protocol) with GRID Networks.

SSL Protocol provides the high degree of security over the Web Network. Basically it is used for secure communication over an internet. SSL protocol builds a secure tunnel based communication interface between the client and the server. Here the SSL is used in the grid network to resolve the load over the network after the authentication of both sender and receiver.

The stage of SSL protocol are as follows:

i. To establish the key of safety communication.
ii. Server authentication.
iii. Client authentication.
iv. End stage

3) Encryption/Decryption

The encryption is done at the client end and sent to the server end. The decryption is also provided by the client end only.

4) Query and Flow

The information when entered, it will first encrypt the data and then sent to the server. The server then takes the data and stores it as it is without decrypting. This saves the time required for decryption.

When the client tries to retrieve data, it sends the query. The query is then encrypted by SSL Client and is decrypted by SSL Server. The decrypted query is sent to the Server. The Server then returns the documents requested by the Client.

The update query also works on the same terms. The data is updated by the server on the client's request.

5) Transactions

The project enables us to make transactions with a QR Code image. The project will take the information to the Payment Gateway and the rest will be handled by the Gateway itself. This helps to also include transactions in the project.

V.ALGORITHM

The algorithm which the paper [6] proposes is an ETSFS algorithm. The algorithm in this paper has been a modification of it –

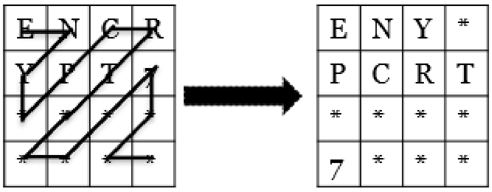
Algorithm for Encryption –

1. The string entered is converted into an array of integers. This is a 2-dimensional array of 4 rows and 4 columns.

E	N	C	R
Y	P	T	7
*	*	*	*
*	*	*	*

The above is the plotting of the string 'ENcrypt7' in a 2D array form.

2. Transpose.



E	N	C	R
Y	P	T	7
*	*	*	*
*	*	*	*

E	N	Y	*
P	C	R	T
*	*	*	*
7	*	*	*

We just move around the elements of the array in the way shown by the arrows.

3. Substitution.

In the substitution phase, the values of the characters are substituted by other values. This adds to the security. Substitution is a three step process-

1. Characterizing
2. Keys formation
3. Operation.

During the first phase, the character is categorized into small caps (a-z), capital alphabets (A-Z), digits (0-9) and symbols (! @#$%*).

DOI 10.17148/IJARCCE.2016.54124

Characterizing gives the value of 'M'. For alphabets (i.e. small and capital) M is 26. While for digits it is 10 and for symbols it is 7. Three user given keys are convert into 12 keys.

Next the character's values are taken in its numeric form. For example, A is replaced by 1, C by 3, Y by 25. Now we add the keys and take modulus with M.

Consider the character as 'E', its numeric is 5. The key element is suppose 23. Then adding 23 and 5 will give 2. Which is the character 'B'.

E	N	Y	*
P	C	R	T
*	*	*	*
7	*	*	*

→

B	L	X	+
Q	E	U	X
	!	*	
8	+	&	@

4. Folding.
The matrix is mirrored about the left diagonal first and then the right diagonal. All the elements along the diagonal are only interchanged. Now for the rest of the elements the matrix is mirrored about the horizontal centre and then the vertical centre. Now the rest of the elements are interchanged.

B	L	X	+
Q	E	U	X
	!	*	
8	+	&	@

→

@	+	&	8
X	*	!	Q
	U	E	
+	L	X	B

5. Shifting.
This shifts the rows upwards by 1. The upper row is added down at the end.

@	+	&	8
X	*	!	Q
	U	E	
+	L	X	B

→

X	*	!	Q
	U	E	
+	L	X	B
@	+	&	8

VI.CONCLUSION

The project started with the team finding a suitable database first. The team came across the need to add documents dynamically which can be done using MongoDB. The team also realized that MongoDB does not provide security measures. This was a task that the team worked upon and decided to use encryption for these purposes. The ETSFS algorithm was chosen after studying a number of algorithms. The project was visualized from end to end and then requirements and input-outputs were decided. The project plan was then decided in a team meeting.

ACKNOWLEDGMENT

We would like to thank all those who somehow have given a contribution to this project and taking this project on the path of success.

REFERENCES

[1] Iqra Basharat, Farooque Azam Database Security and Encryption: A Survey Study. International Journal of Computer Applications (0975 888) Volume 47 No.12, June 2012

[2] Ali Makhmali, Hajar Mat Jani; Comparative Study on Encryption Algorithms and Proposing a Data Management Structure. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 6, JUNE 2013 ISSN 2277-8616.

[3] Obaida Mohammad Awad Al-Hazaimeh, A NEW APPROACH FOR COMPLEX ENCRYPTING AND DECRYPTING DATA. International Journal of Computer Networks & Communications (IJCNC) Vol.5, No.2, March 2013.

[4] A. Mathur, An ASCII value based data encryption algorithm and its comparison with other symmetric data encryption algorithms. International Journal on Computer Science and Engineering (IJCSE), Vol. 4, pp. 1650-1657, Sep 2012 ISSN: 09753397.

[5] Satyajeet R. Shinde, Rahul Patil ,An Encryption Algorithm Based on ASCII Value of Data, International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7232-7234.

[6] Hanan A. Al-Souly, Abeer S. Al-Sheddi, Heba A. Kurdi;Fast, Lightweight Symmetric Encryption Algorithm for Secure Database. (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Extended Papers from Science and Information Conference 2013.

[7] Kadhem, H.; Amagasa, A Novel Framework for Database Security based on Mixed Cryptography. Internet and Web Applications and Services, 2009. ICIW '09. Fourth International Conference on; Publication Year: 2009,Page(s):163170.

[8] Dr. Anwar Pasha Abdul Gafoor Deshmukh; Transparent Data Encryption-Solution for Security of Database Contents. International Journal of Computer Science and Applications, Vol. 2, March 2011.

[9] Luc Bouganim; Yanli GUO, Database Encryption; Encyclopedia of Cryptography and Security. S. Jajodia and H. van Tilborg (Ed.) 2009.

[10] Tingjian Ge, Stan Zdonik Fast, Secure Encryption for Indexing in a Column Oriented DBMS. 2007 IEEE 23rd International Conference on Data Engineering (2007) Publisher: IEEE, Page(s): 676-685.

[11] https://www.digicert.com/ssl.htm

[12] https://en.wikipedia.org/wiki/Payment_gateway

[13] https://www.digicert.com/ssl.htm

[14] https://en.wikipedia.org/wiki/Transport_Layer_Security

DOI 10.17148/IJARCCE.2016.54124