Ensuring Data Security in Databases Using Format Preserving Encryption

Conference Paper · January 2018

DOI: 10.1109/CONFLUENCE.2018.8442626

Shikha Gupta
Netaji Subhas Institute of Technology
8 PUBLICATIONS **20** CITATIONS
SEE PROFILE

Mohit Agarwal
University
16 PUBLICATIONS **205** CITATIONS
SEE PROFILE

Ensuring Data Security in Databases Using Format Preserving Encryption

Shikha Gupta[1], Satbir Jain[3]
Computer Engineering
NSIT
New Delhi, India
shikha.gpt1@gmail.com, jain_satbir@yahoo.com

Mohit Agarwal[2]
Department of Physics and Computer Science
DayalBagh Educational Institute
Agra, India
rs.mohitag@gmail.com

***Abstract*— In the current scenario data security has become an important issue with the growth of digital media. Many users and the applications are accessing the data both from inside and outside the database. Hence, the database as well as data within these databases has become the key target for most of the attackers. Many cryptographic schemes have been designed to solve this problem. Encryption plays an important role in providing the data confidentiality to data stored within the databases. But, the problem in adopting the standard encryption methods is that they may cause a damage to the existing schema as well as to the underlying applications or database as the output length is different from the input length and it also changes the format of data. This paper proposes a Format Preserving Encryption method by accumulating with Advance encryption standard(AES), eXclusive OR operation and a translation method for 16 digit numeric data. Format preserving encryption technique is used to minimizes the databases changes by preserving the format as well as the length of the input data.**

Keywords— Data security, Format preserving encryption, Advanced encryption standard(AES), Data Length, Database, numeric data.

I. Introduction

Data is a major asset for every organization whether it is sensitive or non-sensitive in nature. In this contemporary world, the organizations biggest challenge is security of data i.e. sensitive in nature. Sensitive data means the data that should not be made public such as credit card numbers, pan numbers etc. Many organizations have been targeted by an increasing number of attacks that focuses on stealing the personal information. Thus, it has created awareness among the users that have motivated many organization to find suitable methods for securing the data to minimize the consequence of losing data. To protect confidentiality and personal information cryptography is widely used in everyday applications such as to do online transfer through Internet, VPN technologies, encryption of files or complete hard disks. Many security techniques are being implemented for protecting the data. Cryptography[6] is the most effective way to protect the data. The word cryptography was mostly used as an synonym for encryption, but now a days it deals with a much wider range of security techniques. The four major goal in cryptography to protect information could be defined as follows: [5]

1. Message confidentiality (or privacy)
2. Message integrity
3. Sender authentication
4. Sender non-repudiation

Encryption is the best method to protect the data stored within the databases, while maintaining high database performance.

The standard encryption schemes can only maintain the data security. The valuable information stored in database could be accessed by the unauthorized users or attackers for malicious purpose. Therefore, it is necessary to apply effective and secure encryption/decryption schemes to enhance the security of data. Various encryption schemes like AES,DES, Blowfish are used to encrypt the sensitive data stored in databases, but at the same time it also degrades the performance that can leads to various key costs:

- It requires large processing time for encrypting the sensitive data.
- Extra storage space is required for storing encrypted data.
- Overhead of query response time and allocated resources for decrypting data to process those queries.

A. Need for Format Preserving Encryption

To overcome these problems as stated above a new symmetric encryption scheme is gaining attention named as Format preserving encryption. This technique is a bit different from standard encryption schemes named as AES, DES[1][2]. It is a rapidly growing cryptography tool for providing security in database systems that covers the goal of confidentiality in cryptography. Applying an FPE scheme leads to various advantages stated as follows:

- It increases the security of database systems by maintaining the format of data as well as maintains transparency within the database.
- It is also well suited for masking of data[21]. Data masking will hide the original form of data and replaces it with some random data.

Format preserving encryption as the name suggest is a technology that aims to perform encryption of data without disrupting the format. It means encrypting the data in such a way so that the output has same format as input. This feature makes it to have some advantages over standard encryption methods. A typical example would be the data of numeric format, such as a credit card number. The motivation for using FPE comes from the problems associated with integrating encryption into underlying application or schema, with well defined data architectures. Since FPE can preserve the output in the same format as the input, it is appropriate to encrypt the format-sensitive data that is numeric in nature.

An efficient format preserving encryption scheme for numeric data is proposed in this paper involves the accumulation of Advance encryption standard(AES), eXclusive OR operation and Translation method.

It can improve performance as compare to existing schemes and is provably secure.

978-1-5386-1719-9/18/$31.00 ©2018 IEEE

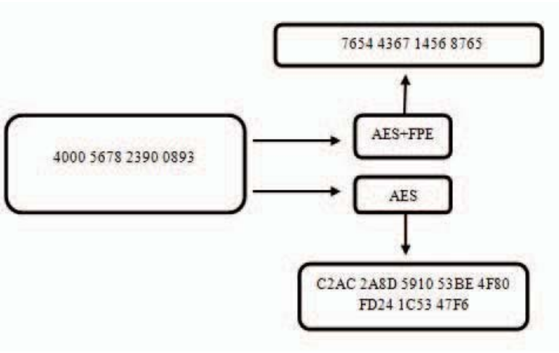

Fig 1. Encryption of a credit card number using AES and AES+FPE

The rest of the paper comprises of following sections. Section 2 describes the related work of FPE. In Section 3, we review the definition of FPE as well of AES. Section 4, provides the description of our proposed work. Section 5, describes the result obtained. Finally the conclusion is drawn in Section 6 followed by the references.

II. Related Work

In the last few years, researches on applied cryptography have developed various practical encryption methods [7,8,10] a paradigm is the Format-Preserving Encryption (FPE). FPE was first proposed in 1981[12], in which a DES based approach was defined to encrypt the strings of a fixed alphabet E which can be defined as E={0,1....9 and a,b.....z}. In this scheme a pad is created with DES algorithm which has the length of the input data and is from the alphabet E. Later the pad is then added letter by letter to input data modulo the biggest character of the alphabet. But this scheme turns out to be defective and easily attacked by intruders. In 1997, Brightwell and Smith[13] proposed a technique named Datatype preserving encryption. Their aim was to encrypt the DB entries without making changes to the data type. The scheme fails because the architecture is complex and cryptographically naive as it doesn't provide relevant security.

In 2002, Black and Rogaway [14] proposed 3 FPE methods: prefix cipher, cycle-walking and generalized-Feistel and suggested that these ciphers can be used to construct FPE schemes on any arbitrary finite domain.

In 2009, Bellare [15] introduces the concept of rank then-encipher approach (or RtE). The scheme defines that it is possible to construct any FPE scheme based on integer FPEs by building a bijection between the target domain and an integer domain. This idea is proves to be strong and useful as it reduces all FPE problems to the integer FPE problem, and has been used flexibly as a basic construction method.

In 2010, Philip Rogaway [16] surveyed over various sizes of domains which can be used for tiny space, small space and large space FPE encryption. In tiny space it uses three methods: Prefix Cipher, Knuth Shuffle and Permutation numbering. For small space encryption it uses FFX mode i.e. "Format preserving" feistel based encryption which supports larger message space but they have a limitation of block size and block cipher used. In 2013, [17] a scheme was proposed defining the enhancement of prefix cipher.

This technique is simple to implement and require less space and time because instead of storing 32 digit hex number in a table, it uses only numeric digit from the table and discard the remaining digits. The other scheme defines the overhead of FPE (FIPS 74-8). The author examines in this scheme that instead of using DES, the use of AES and blowfish will give better results. This scheme have several pros and cons as well. No authentication and randomization of data is defined but it require less storage space and more secure. In 2014, Richard Agbeyibor [18] compares various NIST standards of FPE mechanisms such as FF1(FFX)[16],FF2(VAES3),FF3(BPS) based on input dataset, entropy measurement and implementation, performance and hardware design. They concludes that FF3 is as secure as FF1 and FF2 and requires least hardware resources.

This paper proposes a data security technique based on format preserving encryption or data type preservation. this encryption scheme helps to meet various security challenges posed by protecting diverse types of information. Format preservation provides several distinct benefits that build on solid strong-encryption practices. The main aim of FPE is to encrypt the data without modifying all of the systems that uses that data; such as database field, queries and all the application program.

III. Methodology

A. Format Preserving Encryption

In general, FPE is defined as a symmetric key (K) cipher that encrypts a input message (A) into a output message (B) that has the same format as of(A). The recent research[9,10,11] states two classical definitions of FPE:

1. Basic FPE: It defines the problem that FPE solves, i.e., it makes sure the output falls in the same domain of the input. FPE can be described as a function shown in Eq.1

$$E: X\times K=K \tag{1}$$

where,
E: a reversible function that performs permutation.
X: is called the key space
K: specifies the domain of the input message respectively the output message.

2. *The generalized FPE*: It emphasizes the complexity of the FPE lies in the complexity of message space.

B. Advance Encryption Standard

AES[1] is define as a secret or private key encryption standard developed by Vincent Rijmen and Joan Daemen in 1999, to overcome the disadvantages of DES[2] algorithm. AES is symmetric block cipher algorithm in which the same key is used for encryption and the reverse transformation, decryption[3]. The algorithm must determine the block and key sizes before applying it to the input data. AES allows key sizes of 128, 192, and 256 bits [4]. In standard encryption algorithm (AES), the length of the input block, the output block and the State is 128 bits. AES encryption process performs ten rounds. The first nine rounds will repeat the following four transformations- Sub Byte, Shift Rows, Mix Columns and Add Round Key. In tenth round of the process, only three transformations are performed namely Sub Byte, Shift Rows and Add Round Key.

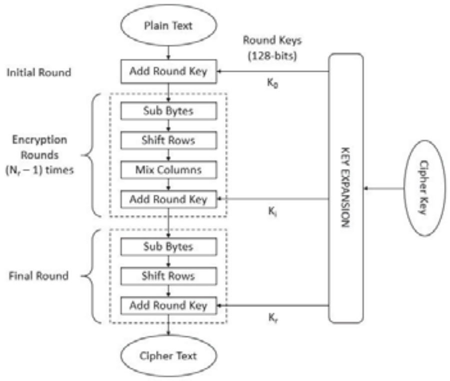

Fig 2. Flowchart of AES Encryption Technique

1) Sub Byte transformation: It uses S-box substitution table.
2) It Shifts the rows of State array by different offsets.
3) Mix column operation is performed by mixing the data within each column of the State array.
4) Add round key function is defined by adding a round key to the State.

IV. Proposed Format Preserving Encryption Technique

FPE was designed in accordance with block cipher by using AES-128 bit encryption algorithm as the base for encrypting the data. This paper defines an efficient format preservation encryption technique by using AES to overcome the disadvantages of DES[2] algorithm used to preserve the format of input data. The proposed technique is designed on the basis of two steps which are defined in order to preserve the format of the 16 digit plaintext data as shown in Figure.2 and it will also preserves the referential integrity of data. After the completion of the AES algorithm, the resultant output will serve as input to the eXclusive OR operation to retain the original format and data type of the input data. The eXclusive OR operation will divide the resultant cipher text is into groups. The operation is performed on each group. After the completion of this step a block of desired size is obtained. Now a translation method is applied to convert the output of eXclusive OR operation into a desired format.

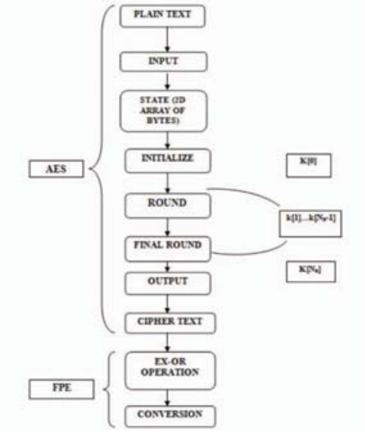

Fig 3. Flow of Proposed Encryption Technique

The use of this technique is suitable for all types of data format, but in this paper the technique is analyzed on the numerical data such as credit card numbers. This technique is analyzed on several credit card numbers using AES-128 bit

V. Results and discussions

A. Steps used topreserve the format of input data

1) *eXclusive OR Operation:* The eXclusive OR operation will divide the resultant 128 bit cipher text is into 8 bit groups. The operation is performed on each group.The last higher order bytes are eXclusive OR with the lower order bytes shown in Table.2. After the completion of this step we will get a 64-bit block. At the end of last round the 16 digit number is encrypted as 128 bit data output.

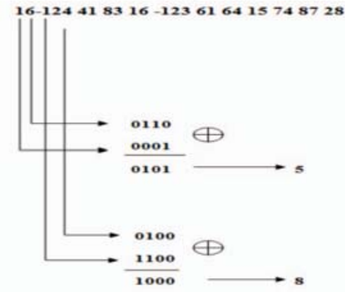

Fig 4. eXclusive OR operation performed on AES output

Similarly, applying the eXclusive OR operation to the remaining groups we will get the hexadecimal digits as the outcome shown in Table 1.

Table 1 eXclusive OR Operation between the two digits of AES output

Output obtained from AES (Decimal Form)		eXclusive OR Operation		Output obtained in Binary form	Hexadecimal Value
1	6	0001	0110	0111	7
12	4	1100	0100	1000	8
4	1	0100	0001	0101	5
8	3	1000	0011	1011	B
1	6	0001	0110	0111	7
12	3	1100	0011	1111	F
6	1	0110	0110	0111	7
6	4	0110	0100	0010	2
1	5	0001	0101	0100	4
7	4	0111	0100	0011	3
8	7	1000	0111	1111	F
2	8	0010	1000	1010	A
6	3	0110	0011	0101	5
3	0	0011	0000	0011	3
6	4	0110	0100	0010	2
7	8	0111	1000	1111	F

2) *Translation Method:* In this step, a translation method is applied by using 5211 coding to the hex digits to get the precise 16 decimal digits; these decimal digits are within the valid range from 0 to 9. The letters from A to F in hexadecimal represents 2 two digit numbers. Instead of applying ordinary conversion we are applying 5211 decimal conversion to get the valid decimal value as shown in Table 3. At the end of this step the input data and output data are same in format and type.

Table 2 Conversion Of Hexadecimal Into Decimal Number Using 5211 Coding

HEX	Binary	5211 Coding	Decimal
7	0111	0+2+1+1	4
8	1000	5+0+0+0	5
5	0101	0+2+0+1	3
B	1011	5+0+1+1	7
7	0111	0+2+1+1	4
F	1111	5+2+1+1	9
7	0111	0+2+1+1	4
2	0010	0+0+1+0	1
4	0100	0+2+0+0	2
3	0011	0+0+1+1	2
F	1111	5+2+1+1	9
A	1010	5+0+1+0	6
5	0101	0+2+0+1	3
3	0011	0+0+1+1	2
2	0010	0+0+1+0	1
F	1111	5+2+1+1	9

The 16 digit output obtained is given by 4537494122963219. The basic idea is to use a strong block cipher such as AES and then combining AES with format preservation to increase the attacker's burden.

Progress in the field of cryptology is based on the practice of making the algorithms public and inviting interested parties to find the flaws. It is in this spirit that our method is presented. The technique proposed in the paper, results in the preservation of data length as well format of the input data. This implies that by preserving the format of the data by combining it with a strong encryption algorithm is as secure as, an AES algorithm.

VI. CONCLUSION

It was examined that Format preserving encryption is an interesting and rapidly growing technology. In this paper a new and efficient FPE scheme is proposed for encrypting integer data of 16 digit by using AES, exclusive OR operation and a translation method is used to overcome the shortcomings of existing schemes like Prefix method, cycle walking[14]and Length preserving encryption scheme[22].

These techniques have various pros and cons which are overcome by the Format preserving encryption scheme used in this paper by using a secure underlying scheme named Advance encryption standard(AES). The proposed scheme is more flexible than techniques shown in Table 3. In future the work might be extended to encrypt numeric data of size 0-19 digits and will also apply data masking techniques to provide more security to sensitive data.

Table 3 Comparison of Proposed Technique with [14] and [22]

	Method	Storage requirements	Execution
Our Work Done	AES, XOR operation and a translation method is applied	No additional storage is required.	It preserves the length as well as format of input data.
[14]	Prefix cipher, Cycle walking	Extra storage for random keys and to hold tables is required	Applicable for small length of input data. Repetitive encryption with small size data sets.
[22]	Length preserving Encryption	No additional storage is required.	It preserves only the length of input data.

REFERENCES

[1] AES, "Advanced Encryption Standard", National Inst. of Standards and Technology (NIST), FIPS-197, 2001.

[2] DES, "Data Encryption Standard", National Institute of Standards and Technology (NIST), Federal Information Processing Standards (FIPS) , Pub 46, 1977.

[3] Daemen, J., Rijmen, V, "The block cipher Rijndael, Smart Card research and Applications", LNCS 1820, Springer , pp. 288-296,1998.

[4] Kaufman, C., Perlman, R.., Speciner. M, Network Security, Private Communication in a Public World. 2nd ed. Prentice Hall PTR, 2002.

[5] Wikipedia article on Cryptography, http://en.wikibooks.org/wiki/Cryptography/Introduction

[6] Wikipedia article on symmetric encryption, http://en.wikipedia.org/wiki/Symmetric-key_algorithm.

[7] Bellare, M., Ristenpart, T., Rogaway, P., Stegers, T, "Format-preserving encryption", Lecture Notes in Computer Science, vol.45, no.5, pp.295–312. 2009.

[8] Luby, M., Rackoff, C, "How to construct pseudorandom permutations from pseudo-random functions", Siam Journal on Computing, vol.17, no.2, pp.373–386,1988.

[9] Li, J.W., Jia, C.F., Liu, Z.L., Li, M ,"FPE scheme based on k-splits feistel network" Journal on Communications, vol.33, no.4, pp. 62–68, 2012.

[10] Liu, Z.L., Jia, C.F., Li, J.W, "Research on the format-preserving encryption techniques", Journal of Software, vol.23, no.1, pp. 152–170, 2012.

[11] Liu, Z. L., Jia, C. F., Jing-Wei, L. I, "Research on the format-preserving encryption modes", Journal on Communications, vol.32, no.6, 184–190, 2011.

[12] Guidelines for Implementing and Using the NBS Data Encryption Standard: "First DES-based", FPE approach. https://www.thc.org/root/docs/cryptography/fips74.html, April 1.

[13] Smith, H.E, Brightwell, M, "Using Datatype-Preserving Encryption to Enhance Data Warehouse Security", 20th National Information Systems Security Conference, NIST, pp.141, 1997.

[14] Black, J., Rogaway, P, "Ciphers with arbitrary finite domains", Cryptographers Track at the RSA Conference, pp. 114-130, Springer, 2002.

[15] Bellare, M., Ristenpart, T., Rogaway, P., Stegers, T, "Format preserving encryption", Springer, 2009.

[16] Bellare, M., Rogaway, P., Spies, T, "The FFX mode of operation for format-preserving encryption",NIST submission, 2010.

[17] Mallaiah, K., Ramachandram, S., Gorantala, S,"Performance Analysis of Format Preserving Encryption" , (FIPS PUBS 74-8) over block ciphers for Numeric data, IEEE, 2013.

[18] Agbeyibor, R., Butts, J., Grimaila, M., Mills, R.: Evaluation of format preserving encryption algorithms for critical infrastructure protection, Springer, 2014.

[19] Chandrashekar, P., Dara, S., Muralidhara, V.N, "Efficient format preserving encrypted databases", International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 1-4. IEEE, July 2015.

[20] Wang, P., Luo, H., Liu, J, "Format-preserving encryption for Excel", International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp.1-2. IEEE, May 2016.

[21] Cui, B.J., Zhang, B.H., Wang, K.Y, "A Data Masking Scheme for Sensitive Big Data based on Format-Preserving Encryption", International Conference on Computational Science and Engineering (CSE) and International Conference on Embedded and Ubiquitous Computing (EUC), pp.519-524. IEEE, 2017.

[22] Gupta, S., Jain, S., Govil, A," An Innovative Length Preserving Encryption Scheme for Sensitive Data Security", 4th International conference on Computing for Sustainable Global Development(INDIACOM). IEEE, 2017.