Secure Linear Transformation Based Cryptosystem using Dynamic Byte Substitution

Adi Narayana Reddy K *aadi.iitkgp@gmail.com*
Dept. of CSE & HITAM
Gowdavelli, Medchal, Hyderabad,India

Vishnuvardhan B
Dept. of IT & JNTUH College of Engineering
Nachupally, Kondagattu, Karimnagar, India

Abstract

Many classical cryptosystems are developed based on simple substitution. Hybrid cryptosystem using byte substitution and variable length sub key groups is a simple nonlinear algorithm but the cryptanalyst can find the length of each sub key group because the byte substitution is static and if the modulo prime number is small then byte substitution is limited to first few rows of S-box. In this paper an attempt is made to introduce the nonlinearity to the linear transformation based cryptosystem using dynamic byte substitution over GF (2^8). The secret value is added to the index value to shift the substitution to a new secret location dynamically. This adds extra security in addition to non-linearity, confusion and diffusion. The performance evaluation of the method is also studied and presented.

Keywords: Dynamic Byte Substitution; Hill Cipher; Pseudo-random Numbers; Sub Key Groups.

1. INTRODUCTION

The information must be secure while transmission. Cryptography is one of the methods to attain security. The encryption and decryption process converts intelligible information into unintelligible form and vice versa. There are various encryption algorithms to provide security for the information. Traditional symmetric key ciphers use substitution, in which each character is replaced by other character. The mathematician Lester S. Hill invented the Hill cipher in 1929. Hill cipher is a classical substitution technique that has been developed based on linear transformation. Many modifications improved the security of the Hill cipher but all modified algorithms are linear. The nonlinearity is one of the measures to make the modified algorithms more secure. The byte substitution using S-box provides nonlinearity. Consider the table of S-box of size 16 x16, contains the permutation of all possible values of 8 bits with the leftmost 4 bits as row value and the rightmost 4 bits as a column value. The row and column values serve as indices into the S-box to select a unique 8-bit output. The dynamic byte substitution shifts the substitution to a new secret location. The efficiency of the algorithm is one of the most important aspects to be studied.

Hill cipher was developed based on simple linear transformation. It is easy to implement, the processing speed is high and high throughput. But it is vulnerable to known plaintext attack and the inverse of every shared key matrix may not exist all the time. It is a simple traditional symmetric key cipher algorithm. In Hill cipher, the plaintext to be transmitted through the insecure communication channel is partitioned into ‘m’ groups, each of size ‘n’ and these partitioned groups are called blocks. Assume that both ‘n’ and ‘m’ are positive integers and Mi is the ith partitioned block. Transform each of the block Mi, one at a time using secret key matrix. Map each character with unique numeric value like A = 0, B = 1 ... to produce the ‘n’ characters in each of the partitioned block. The ith cipher text block Ci can be obtained by encrypting the ith plaintext block Mi using Equation (1) as:

$$C_i = M_i K \bmod m \qquad (1)$$

In which K is an n×n key matrix. The plain text can be obtained from the decrypted cipher text using Equation (2) as:

$$M_i = C_i K^{-1} \bmod m \qquad (2)$$

In which K-1 is the key inverse and it exist only if the GCD (det K (mod m), m) = 1. According to Overbey [12] the key space of the Hill cipher is precisely GL (n, Zm) - the group of n x n matrices that are invertible over Zm for a predetermined modulus m and the key space of a prime modulus is larger than composite modulus.

Many researchers improved the security of linear transformation based cryptosystem. Yeh, Wu et al. [21] presented an algorithm which thwarts the known-plaintext attack, but it is not efficient for dealing bulk data, because too many mathematical calculations. Saeednia et al [16] improved the original Hill cipher, which prevents the known-plaintext attack on encrypted data but it is vulnerable to known-plaintext attack on permutated vector because the permutated vector is encrypted with the original key matrix. Ismail [8] tried a new scheme HillMRIV (Hill Multiplying Rows by Initial Vector). Rangel-Romeror et al. [13] proved it is vulnerable to known-plaintext attack and also proved that if IV is not chosen carefully, some of the new keys to be generated may not be invertible over Zm, this make encryption/decryption process useless. Lin C.H. et al. [11] improved the security of Hill cipher by using several random numbers. It thwarts the known-plaintext attack but the algorithm is not efficient and is vulnerable to the chosen-cipher text attack and it was proved by Mohsen Toorani et al. [19, 20] and he improved the security with one-way hash function but Keliher, L. et al [10] proved that it is still vulnerable to attacks . Ahmed Y Mahmoud et al [3, 4, 5] improved the algorithm by using eigen values but it is not efficient because the time complexity is more and too many seeds are exchanged. Reddy, K.A. et al. [14, 15] improved the security of the cryptosystem by using circulant matrices but the time complexity is more and is linear. Dunkelman, O. et al. [6] presented that byte substitution step in AES is a perfectly nonlinear function. Kaipa, A.N.R. et al. [9] improved the algorithm by adding nonlinearity using byte substitution over GF (28) and substitution using variable length sub key groups. The cryptanalyst can gain the length of each sub key group because the byte substitution is static. In the study, an attempt is made to introduce dynamic byte substitution over GF (28) and variable length sub key groups using pseudo random number sequence.

2. PROPOSED METHOD

In this paper the nonlinearity is introduced using byte substitution over GF (2^8). The byte substitution is static; to make it dynamic a secret value is added to the static index to shift the substitution to a secret location. To add extra security variable length sub keys are generated using sequence of pseudo random numbers.

2.1 Algorithm

Let M be the message to be transmitted securely. The message is divided into m blocks each of size n where is positive integer (greater than 1) and pad the last block if necessary. Let C_i be the ciphertext of the i^{th} block corresponding to i^{th} block of plaintext M_i. Choose a prime number p < (2^8) and base value b. The algorithm is explained in the following steps.

1. Select a vector of 'n' relatively prime numbers (k_1, k_2, … k_n). Rotate each row vector relatively right to the preceding row vector to generate a shared key matrix K_{nxn}.
2. Let r = $\sum_{i=1}^{n} k_i$ mod p. Generate a sequence of 'p' pseudo random numbers S_i with initial seed value as r. Now generate the sub-key group as i from pseudo-random number sequence as, $j = (i + S_i) \bmod b$, for all $i \in S_{Gj}$ and $b < \lfloor p/2 \rfloor$. (i.e. the sub-key groups are formed with the pseudorandom number sequence.)

3. Encryption: The encryption process follows the following steps

   I. Initially apply the transformation on the plaintext block as Y = KM mod p.

   II. The index of very element of the vector Y is calculated as, Index = Y, where Index = (Index1, Index2, …, Indexn)

   III. Share a secret value to shift the substitution to a new secret location. The following equations explains the dynamic byte substitution

   NewIndex1 = (Index 1+ secret value) mod 256

   NewIndexi = (NewIndexi-1 + Indexi) mod 256, for i = 2, 3, …, n (3)

   The static Index has been shifted to new secret location. Now substitute each element of the vector NewIndex by an element from the S-box table and the resultant vector as Z

   IV. Select 'n' elements in queue from the corresponding sub key group SG using z1 mod b as index of the sub key group and add to Z over mod 28 to generate cipher text block C.

4. Transfer the pair (C, index), where index = z_1 mod b to other end

5. Decryption: The decryption process follows the following step. After receiving the pair, select 'n' elements in queue from the received indexed sub key group and subtract these elements from C and the resultant vector becomes Z. Substitute each element of vector Z by an element from the inverse S-box table. The inverse S-box table produces the NewIndex vector. Now apply the following inverse process on the NewIndex vector to get Index vector as:

   $Index_1$= ($NewIndex_1$ – secret value) mod 256

   $Index_i$ = ($NewIndex_i$ – $NewIndex_{i-1}$) mod 256, for i = 2, 3, …, n (4)

   The vector Index becomes Y. The multiplication of the resultant vector Y with the inverse key matrix produces the original plaintext as

   $$M = K^{-1}Y \bmod p \quad (5)$$

The following example explains the algorithm

2.2 Example

Consider a prime number 'p' as 29 and the set of relatively prime numbers (5, 27, 13). Generate shared key matrix K_{3x3}. Assume the plaintext block M= [8, 13, 5]. The Table 1 and Table 2 shows the S-box and inverse S-box using GF (2^8) field with irreducible polynomial $x^8 + x^4 + x^3 + x + 1$ and are presented in the Appendix. Generate a sequence of 'p' pseudo-random number with initial seed value as 45 and secret value as 45. Assume b = 3 and generate the three sub-key groups (S_G) from the random number sequence.

SG [0] = {1, 3, 4, 6, 7, 8, 11, 14, 16, 19, 23, 24, 26, 27}

SG [1] = {0, 2, 5, 9, 13, 13}

SG [2] = {10, 15, 17, 18, 20, 21, 23, 25, 28}

The process of encryption outputs the ciphertext C = [160, 172, 69]. After communication of the pair (C, 1) and the decryption process outputs the plaintext M = [8, 13, 5].

3. PERFORMANCE ANALYSIS OF THE PROPOSED METHOD

The performance of the proposed method is evaluated by considering running time (i.e. time complexity), result analysis and security analysis.

3.1 Time Complexity

The time complexity measures the running time of the algorithm. The time complexity of the proposed algorithm to encrypt the 'm' plaintext blocks are O (mn^2) and to decrypt 'm' ciphertext blocks also O (mn^2) which is shown in the equation (6),where 'n' is size of each block, which is same as that of original Hill cipher. In this process T_{Enc} and T_{Dec} denote the running time for encryption and decryption of 'm' block of plaintext respectively. From step 3.1 the number of multiplications is n^2 and the number of additions is n^2-n because K is an n x n non-singular key matrix and the size of plaintext/ciphertext is 'n'.

$$\begin{aligned} T_{Enc}(m) &\cong m(n^2)T_{Mul} + m(n^2-n)T_{Add} + mnT_S + mnT_{SV} \\ T_{Dec}(m) &\cong m(n^2)T_{Mul} + m(n^2-n)T_{Add} + mnT_{IS} + mnT_{SV} \end{aligned} \quad (6)$$

Where T_{Add} , T_{Mul} , T_S , T_{IS} and T_{SV} are the time complexities for scalar modular addition, multiplication, byte substitution, inverse byte substitution and secret vector addition respectively.

$$\begin{aligned} T_{Enc}(m) &\cong m(n^2)c_1 + m(n^2-n)c_2 + mnc_3 + mnc_4 \\ T_{Dec}(m) &\cong m(n^2)c_1 + m(n^2-n)c_2 + mnc_5 + mnc_4 \end{aligned} \quad (7)$$

In which c_1, c_2, c_3, c_4 and c_5 are the time for scalar multiplication, scalar addition, byte substitution, inverse byte substitution and secret vector addition respectively. These times taken for encryption and decryption are same as original Hill cipher and less than other modified Hill ciphers.

3.2 Result Analysis

The result analysis of the method is carried with by considering a large encrypted data. The runs test and correlation coefficients are carried on the data.

3.2.1 Runs Test

The runs test is carried on to check the randomness of the encrypted data. The data falls into two separate categories such as above and below to a median. The test result shows that the encrypted data is random since runs test gives -0.525 as a result. This value is better than the hybrid cryptosystem proposed by Kaipa, A.N.R. et al [8].

3.2.2 Correlation Coefficient

Correlation coefficient is a number between -1 and 1 which quantifies the relation between two variables. The correlation coefficient is 1 in case an increasing linear relation, -1 in case of decreasing linear relation and some value in between in all other cases indicating the linear dependence between the variables. The figure 2 shows that the correlation distribution of plaintext and ciphertext for the proposed algorithm. The correlation coefficient has calculated for the dynamic byte substitution and hybrid cryptosystem based on static byte substitution proposed by Kaipa, A.N.R. et al [8] and the values are (-0.0545) and (-0.06005) respectively. The dynamic byte substitution is better than static byte substitution. The results are presented in scatter plot in figure 1.

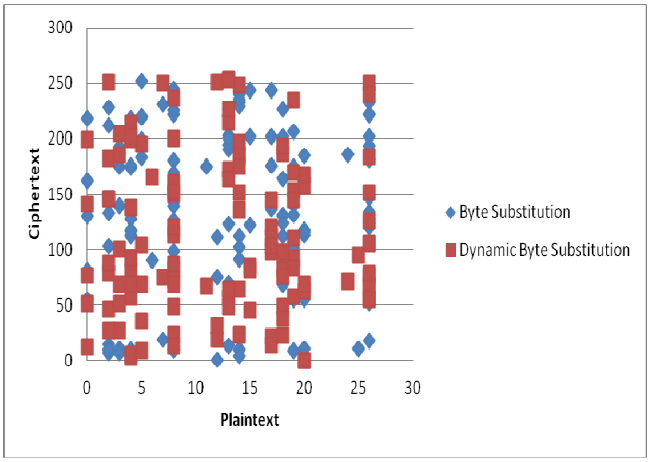

FIGURE 1: Relation between Plaintext and Ciphertext.

3.3 Comparative Study

3.3.1 Time Complexity

The Fig 2 shows that the running time of encryption process in sec over various block sizes. The 5 million characters as input. The results show that our method takes same time as Hill cipher and Toorani's improved algorithm.

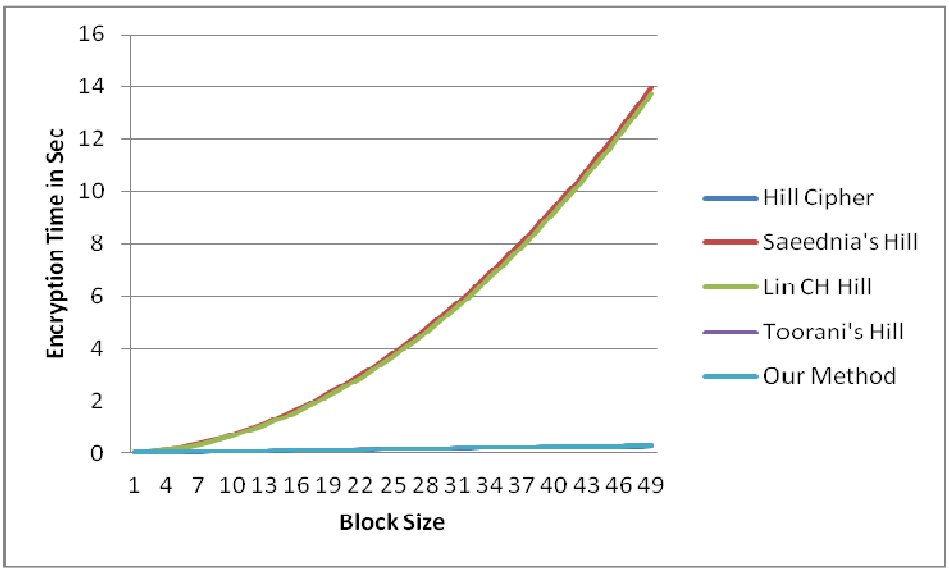

FIGURE 2: Encryption Time in Sec for various sizes of blocks.

3.3.2 Avalanche Effect

The avalanche effect is a characteristic of an encryption algorithm in which a small change in plaintext or key should produce large change in the corresponding ciphertext. The avalanche effect of our proposed method is same as AES algorithm. The avalanche effect has been calculated on 5 million characters by considering random keys the results shows that it was at least 71% because our proposed method uses dynamic byte substitution.

3.4 Security Analysis

The proposed method overcomes all the security drawbacks of the linear transformation based cryptosystem. The byte substitution introduces non-linearity and dynamic byte substitution shifts

the static location to the new secret location. The dynamic byte substitution provides security against linear, differential and algebraic attacks and Dunkelman, O. et al. [5] presented that the byte substitution is perfect nonlinear function. The variable length sub key groups add another degree of security to the method. It thwarts the ciphertext-only attack if the modulo a prime number is large and dimension of key matrix is at least 6 x 6. The proposed method thwarts known-plaintext attack because 'n' plaintext, ciphertext pairs cannot be used to solve 'n' unknown elements of key matrix K and n^2 unknowns of secret shift values and n^2 unknowns of sub key groups. It also thwarts chosen-plaintext and chosen –ciphertext attacks because the dynamic byte substitution using S-box over GF (2^8) adds nonlinearity and confusion to the cryptosystem and also variable length sub key groups adds extra security to diffuse the relation between key and ciphertext. From the analysis the adversary needs

4. CONCLUSION

The proposed cryptosystem is similar to linear transformation based Hill cipher but the nonlinear is included by using byte substitution over GF (2^8). The dynamic byte substitution adds extra security by shifting the substitution to a secret location. This overcomes the problems of hybrid cryptosystem based on static byte substitution. This method encrypts the same plaintext blocks to different ciphertext blocks and the cryptanalyst cannot know the length of sub key groups. It is free from linear and differential cryptanalysis attack along with known-plaintext and ciphertext only attack. It provides more security than the existing algorithms with less key size because set of 'n' elements is enough to form the n x n key matrix. It required memory is less compared to other improvements and the processing speed is high so we can use this algorithm in mobile environment.

5. REFERENCES

[1] Advanced Encryption Standard (AES) 2001. Federal Information Processing Standards (FIPS), publication November 26.

[2] Abidin, A.F.A. and Chuan, O.Y. and Kamel Ariffin, Muhammad Rezal 2011. A novel enhancement technique of the hill cipher for effective cryptographic purposes. Journal of Computer Science, 7 (5). pp. 785-789. ISSN 1549-3636

[3] Ahmed, Y.M. and A.G. Chefranov, 2009. Hill cipher modification based on eigenvalues hcm-EE. Proceedings of the 2th International Conference on Security of Information and Networks, Oct. 6-10, ACM Press, New York, USA., pp: 164-167. DOI: 10.1145/1626195.1626237

[4] Ahmed, Y.M. and Alexander Chefranov, 2011. Hill cipher modification based on pseudo-random eigen values HCM-PRE. Applied Mathematics and Information Sciences (SCI-E) 8(2), pp. 505-516.

[5] Ahmed, Y.M. and Alexander Chefranov. Hill cipher modification based generalized permutation matrix SHC-GPM, Information Science letter, 1, pp. 91-102

[6] Dunkelman, O. and Keller, N. 2007. A New Criterion for Nonlinearity of Block Ciphers, IEEE Transactions on Information Theory, Vol. 53, No. 11, 3944-3957. DOI: 10.1109/TIT.2007.907341

[7] Hill, L.S., 1929. Cryptography in an Algebraic Alphabet. Am. Math. Monthly, 36: 306-312. http://www.jstor.org/discover/10.2307/2298294?uid=3738832&uid=2129&uid=2&uid=70&uid=4&sid=21102878411191

[8] Ismail, I.A., M. Amin and H. Diab, 2006. How to repair the hill cipher. J. Zhej. Univ. Sci. A., 7: 2022-2030. DOI: 10.1631/jzus.2006.A2022

[9] Kaipa, A.N.R., V.V. Bulusu, R.R. Koduru and D.P. Kavati, 2014. A Hybrid Cryptosystem using Variable Length Sub Key Groups and Byte Substitution. J. Comput. Sci., 10:251-254

[10] Keliher, L. and A.Z. Delaney, 2013. Cryptanalysis of the toorani-falahati hill ciphers. Mount Allison University. http://eprint.iacr.org/2013/592.pdf

[11] Lin, C.H., C.Y. Lee and C.Y. Lee, 2004. Comments on Saeednia's improved scheme for the hill cipher. J. Chin. Instit. Eng., 27: 743-746. DOI: 10.1080/02533839.2004.9670922

[12] Overbey, J., W. Traves and J. Wojdylo, 2005. On the keyspace of the hill cipher. Cryptologia, 29: 59-72. DOI: 10.1080/0161-110591893771

[13] Rangel-Romeror, Y., R. Vega-Garcia, A. Menchaca-Mendez, D. Acoltzi-Cervantes and L. Martinez-Ramos *et al*., 2008. Comments on "How to repair the Hill cipher". J. Zhej. Univ. Sci. A., 9: 211-214. DOI: 10.1631/jzus.A072143

[14] Reddy, K.A., B. Vishnuvardhan, Madhuviswanath and A.V.N. Krishna, 2012. A modified hill cipher based on circulant matrices. Proceedings of the 2nd International Conference on Computer, Communication, Control and Information Technology, Feb. 25-26, Elsevier Ltd., pp: 114-118. DOI: 10.1016/j.protcy.2012.05.016

[15] Reddy, K. A., B. Vishnuvardhan, Durgaprasad, 2012. Generalized Affine Transformation Based on Circulant Matrices. Internatonal Journal of Distributed and Parallel Systems, Vol. 3, No. 5, pp. 159-166

[16] Saeednia, S., 2000. How to make the hill cipher secure. Cryptologia, 24: 353-360. DOI: 10.1080/01611190008984253

[17] Stallings, William. Cryptography and Network Security, Third edition, PHI/Pearson.

[18] Forouzan, B.A. and Mukhopadhyay, D. Cryptography and Network Security, Second edition, TMH.

[19] Toorani, M. and A. Falahati, 2009. A secure variant of the hill cipher. Proceedings of the IEEE Symposium on Computers and Communications, Jul. 5-8, IEEE Xplore Press, Sousse, pp: 313-316. DOI: 10.1109/ISCC.2009.5202241

[20] Toorani, M. and A. Falahati, 2011. A secure cryptosystem based on affine transformation. Sec. Commun. Netw., 4: 207-215. DOI: 10.1002/sec.137

[21] Yeh, Y.S., T.C. Wu, C.C. Chang and. W.C. Yang, 1991. A new cryptosystem using matrix transformation. Proceedings of the 25th IEEE International Carnahan Conference on Security Technology, Oct. 1-3, IEEE Xplore Press, Taipei, pp: 131-138. DOI: 10.1109/CCST.1991.202204

Appendix

Table 1 S-box generated from the polynomial $x^8 + x^4 + x^3 + x + 1$

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
1	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
2	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
3	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
4	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
5	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF
6	D0	EF	AA	FB	43	4D	33	85	45	F9	02	7F	50	3C	9F	A8
7	51	A3	40	8F	92	9D	38	F5	BC	B6	DA	21	10	FF	F3	D2
8	CD	0C	13	EC	5F	97	44	17	C4	A7	7E	3D	64	5D	19	73
9	60	81	4F	DC	22	2A	90	88	46	EE	B8	14	DE	5E	0B	DB
A	E0	32	3A	0A	49	06	24	5C	C2	D3	AC	62	91	95	E4	79
B	E7	CB	37	6D	8D	D5	4E	A9	6C	56	F4	EA	65	7A	AE	08
C	BA	78	25	2E	1C	A6	B4	C6	E8	DD	74	1F	4B	BD	8B	8A
D	70	3E	B5	66	48	03	F6	0E	61	35	57	B9	86	C1	1D	9E
E	E1	F8	98	11	69	D9	8E	94	9B	1E	87	E9	CE	55	28	DF
F	8C	A1	89	0D	BF	E6	42	68	41	99	2D	0F	B0	54	BB	16

Table 2 Inverse S-box generated from the polynomial $x^8 + x^4 + x^3 + x + 1$

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	52	09	6A	D5	30	36	A5	38	BF	40	A3	9E	81	F3	D7	FB

1	7C	E3	39	82	9B	2F	FF	87	34	8E	43	44	C4	DE	E9	CB
2	54	7B	94	32	A6	C2	23	3D	EE	4C	95	0B	42	FA	C3	4E
3	08	2E	A1	66	28	D9	24	B2	76	5B	A2	49	6D	8B	D1	25
4	72	F8	E6	64	86	68	98	16	D4	A4	5C	CC	5D	65	B6	92
5	6C	70	48	50	FD	ED	B9	DA	5E	15	46	57	A7	8D	9D	84
6	90	D8	AB	00	8C	BC	D3	0A	F7	E4	58	05	B8	B4	45	06
7	D0	2C	1E	8F	CA	3F	0F	02	C1	AF	BD	03	01	13	8A	6B
8	3A	91	11	41	4F	67	DC	EA	97	F2	CF	CE	F0	B4	E6	73
9	96	AC	74	22	E7	AD	35	85	E2	F9	37	E8	1C	75	DF	6E
A	47	F1	1A	71	1D	29	C5	89	6F	B7	62	0E	AA	18	BE	1B
B	FC	56	3E	4B	C6	D2	79	20	9A	DB	C0	FE	7C	CD	5A	F4
C	1F	DD	A8	33	88	07	C7	31	B1	12	10	59	27	80	EC	5F
D	60	51	7F	A9	19	B5	4A	0D	2D	E5	7A	9F	93	C9	9C	EF
E	A0	E0	3B	4D	AE	2A	F5	B0	C8	EB	BB	3C	83	53	99	61
F	17	2B	04	7E	BA	77	D6	26	E1	69	14	63	55	21	0C	7D