BULETINUL Universităţii Petrol – Gaze din Ploieşti	Vol. LXI No. 2/2009	39 - 48	Seria Matematică - Informatică - Fizică

A Survey on the Usage of Substitution Tables in DES and AES Algorithms

Gabriela Moise

Petroleum –Gas University of Ploieşti, Informatics Department, Bd. Bucureşti, 39, Ploieşti
e-mail: gmoise@upg-ploiesti.ro

Abstract

In this article, there are presented the substitution tables used in cryptosystems (DES and AES) and their role in increasing the security of the cryptographic algorithms. The substituting tables are non-linear permutation functions and they are mathematically formalized using the Boolean functions. The easiest way to obtain new cryptographic algorithms is to use random S-boxes. The research directions are concentrated on the constructions of Boolean functions that have good cryptographic properties.

Key words: *cryptosystem, S-box, DES algorithm, AES algorithm*

Introduction

The cryptographic techniques enable its users to transmit information in a secure way over insecure communication media. The objectives of the cryptography regarding the security of the information are: confidentiality, authentication, integrity and non-repudiation. [2]

Confidentiality assures the secret of the information. Even if unauthorized users intercept it, the information is encrypted, so that these users cannot understand the message.

Authentication implies entities' identification. The authentication procedures make sure that both the transmitter and the receiver are the right entities.

Integrity assures that the message has not suffered any unwanted modification (altered or corrupted).

Non-repudiation means that a part of the system cannot deny previously occurred events that triggered the message. Without this property, a receiver might declare that the message was not received.

A cryptosystem is a system consisting of two entities: the former contains an encryption algorithm and the latter - a decryption algorithm. The encryption algorithm processes the plain text using a key and generates the cipher text. The decryption algorithm processes the cipher text using a key and generates an intelligible text.

There are two types of cryptosystems: symmetric and asymmetric cryptosystems (or cryptosystem with public key).

In a symmetric cryptosystems, the encryption and decryption algorithms use the same key and in an asymmetric cryptosystems, the encryption and decryption used two different keys: en encryption key and a decryption key.

In an asymmetric cryptosystem, the encryption key is public and decryption key is a private key.

A cryptosystem can be formalized as a five-tuple:

$$(P, C, K, E, D),$$

where:

P is a finite set of plaintexts,

C is a finite set of ciphertexts,

K is a finite set of keys.

E is a set of encryption algorithms.

D is a set of decryption algorithms. [2]

For each key k, there are an encryption algorithm and a decryption algorithm. The encryption and decryption algorithms are functions $e_k : P \to C$ and $d_k : C \to P$ with the following property: $d_k(e_k(p)) = p$ for every plaintext p.

The Shannon's model of a cryptosystem is presented in figure no.1.

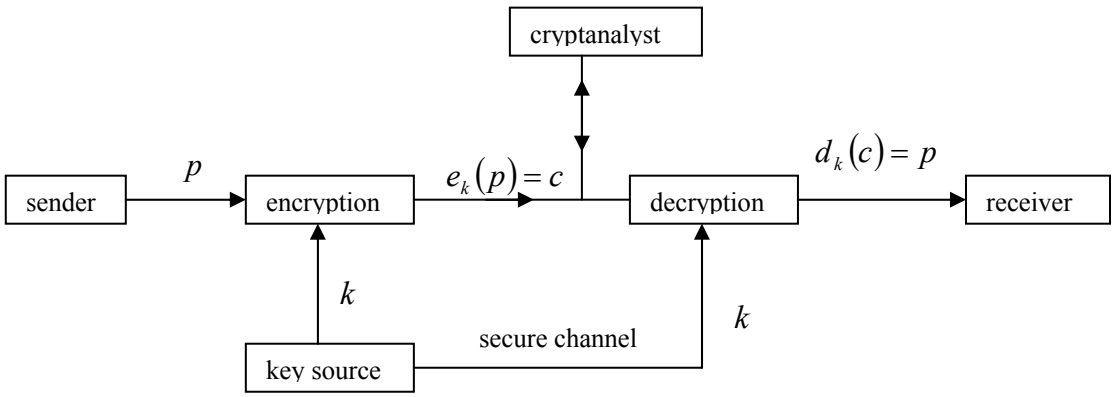

Fig. 1. Shannon's model of a cryptosystem

Shannon defines a perfect secrecy in [9] as follows (equation 1):

$$\Pr[p \mid c] = \Pr[p], \tag{1}$$

where p is a plaintext and c is a ciphertext.

The interpretation of this property is: the probability to discover the plaintext corresponding to an observed ciphertext is the same as that to discover the plaintext without knowing the cipher text.

Shannon proved in [9] that the one-time-pad (OTP) system provides the perfect secrecy. One-time–pad is a crypto algorithm developed by Vernam in 1917. The algorithm is very simple, but, unfortunately, it is not practical, as the key's length has to be the same as the message's

length and each key can be used only once. Although Shannon's demonstration was later infirmed by Wang, the OTP system still has a good cryptographic property. [11]

In this paper, there are presented two symmetric cryptographic algorithms (DES algorithm with its derivation Triple DES and AES algorithm) and the importance of the substitution boxes (S-boxes) in the cryptographic algorithms and the criteria designed to provide good S-boxes.

Standards of the Cryptography: DES and AES Algorithms

Data Encryption Standard (DES) Algorithm

The National Bureau of Standards (NBS) adopted DES algorithm as a new standard in 1977. DES algorithm is the first modern encryption algorithm. It is implemented in the encryption and decryption functions from Shannon's model. The algorithm is based on Lucifer algorithm developed by IBM. DES is an encryption algorithm that operates with block ciphers. DES encrypts a 64-bit block into a new 64-bit block using an encryption key with a 64-bit length. Although the length of the key is 64 bits, there are used only 56 bits because the most suitable bit in each byte is a part byte. DES algorithm was adopted as a new standard in 1977. The algorithm is based on 16 iterations that use a function depending on 16 keys generated from the primary key. [14] [4]

DES algorithm consists of the following steps:

1. the plaintext is divided into blocks of 64 bits and permuted based on a permutation table:
   $p_0 = IP(p)$, where p is a block of 64 bits and IP is the initial permutation.

   $p_0 = L_0 R_0$

   L_0 is the first half of 32 bits from p_0 and R_0 is the last half of 32 bits from p_0.
2. there are generated 16 keys starting from the primary key. This key is permuted using a permutation table. The result is divided in two blocks. At a j step, each pair of blocks is formed using the operation left shift from the previous blocks.
3. there are realized 16 processing rounds over the messages using the 16 keys. At each round i, there are calculated $L_i R_i$ using the following formulas (2):

$$L_i = R_{i-1}, \tag{2}$$

$$R_i = L_{i-1} \oplus f\left(R_{i-1}, K_i\right),$$

   where $K_1, K_2, \ldots, K_{16}$ are 48 bits length and obtained from the primary key K. The process of keys' generation is called "key schedule".

   The block of input data (with 64 bits) is split into two halves, L and R

   L_i is the left-most 32 bits and R_i is the right-most 32 bits of the block at the round i.

   The left half of the input block is the right half of the output block from the previous round.

   The function f operates on the right most half of the block from the previous round and the key K_i, generated at round i. The result is a block of 32 bits.

   The right half is the result of the XOR addition between left half of the output block from the previous round and the calculation f.

4. after the 16th iteration, there are concatenated two halves of the 32 bits, therefore resulting a 64-bit block.
5. the last block is permuted using the inverse function from step 1.

Regarding the security of DES, the algorithm has "a single main critique: the DES has a relatively short key length. This is regarded as the only most serious weakness of the DES. Attacks related to this weakness involve exhaustively testing keys, using a known pair of plaintext and ciphertext messages, until the correct key is found." [13]

The encryption function used at step no. 3 uses substitution boxes known as S-boxes. S-boxes are represented as tables with 4 rows and 16 columns.

Considering a 48-bit binary number, this is split in 8 blocks of 6 bits:

$$B[1], B[2], \ldots B[8].$$

$S[n][row][column]$ has the following signification: the first bit of $B[k]$ is used as the row index and the middle four bits are used as column indices. *S* maps a binary number to a binary number. In the DES algorithm, there are used 8 S-boxes given in detail in [16]. In table 1, it is presented the S-box used at round 1.

Table 1. S-BOX 1-8 of DES algorithm

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
1	0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

The key schedule consists of the following steps:

1. the parity bits are eliminated from the primary key and the rest of bits are permuted using the permutation $PC1$. The permutation $PC1$ is presented in table 2.

Table 2. PC1 Permutation

57	49	41	33	25	17	9
1	58	50	42	34	26	18
10	2	59	51	43	35	27
19	11	3	60	52	44	36
63	55	47	39	31	23	15
7	62	54	46	38	30	22
14	6	61	53	45	37	29
21	13	5	28	20	12	4

2. for each round i from 1 to 16, there are calculated (formula 3):

$$\begin{aligned} C_i &= LS_i\left(C_{i-1}\right) \\ D_i &= LS_i\left(D_{i-1}\right) \\ K_i &= PC2\left(C_i D_i\right) \end{aligned} \tag{3}$$

where LS_i is one or two left shifts and $PC2$ is a permutation presented in table 3. LS_i is one left shift for $i = 1,2,9,16$ and two left shift for the other values of i.

The schema of DES algorithm is presented in figure no 2.

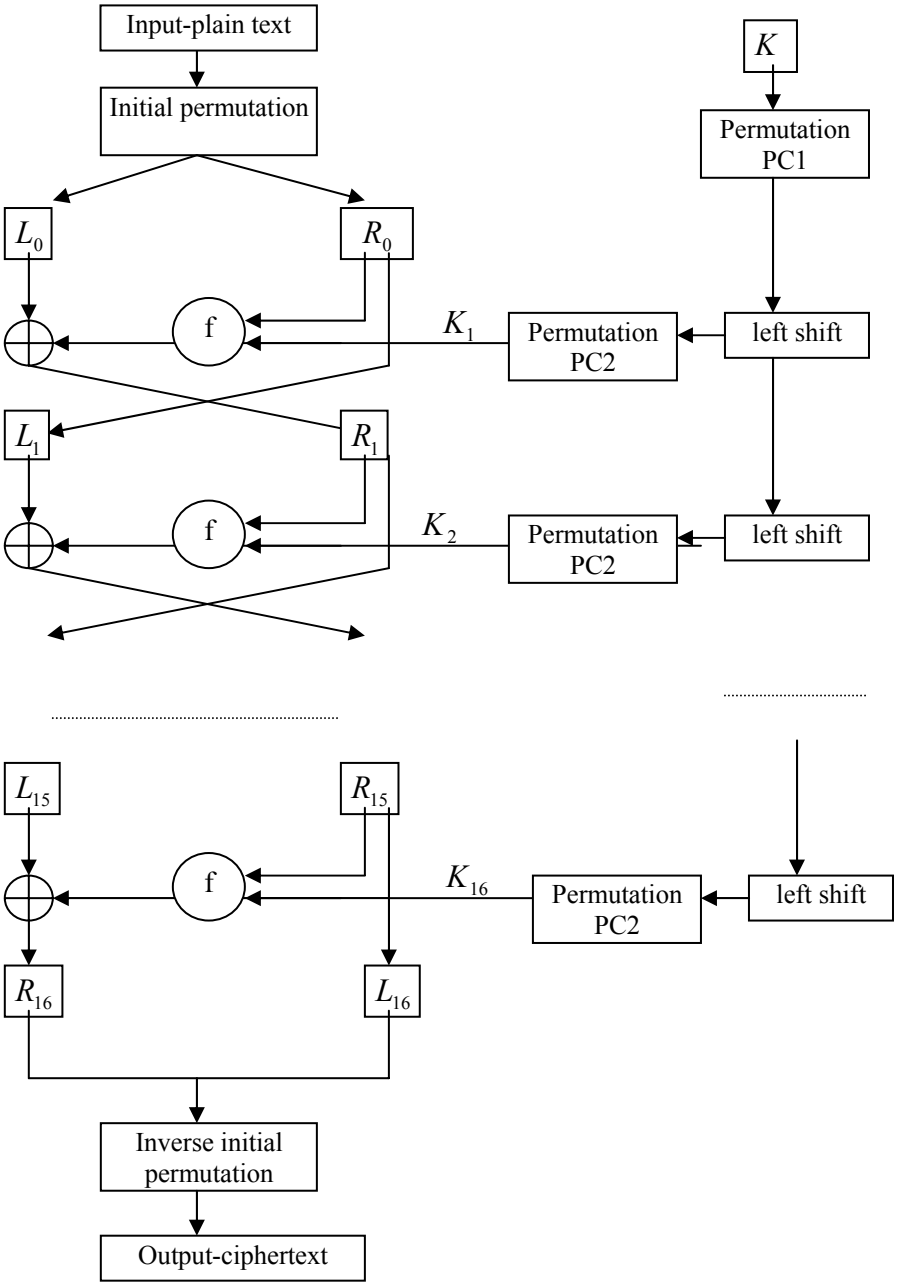

Fig. 2. DES algorithm's schema

Table 3. PC2 Permutation

14	17	11	24	1	5
3	28	15	6	21	10
23	19	12	4	26	8
16	7	27	20	13	2
41	52	31	37	47	55
30	40	51	45	33	48
44	49	39	56	34	53
46	42	50	36	29	32

The 3DES Algorithm

The triple DES (3DES) is an extension of the DES algorithm, consisting in running DES for several times using different keys. The triple DES version proposed by Tuchman follows an encryption-decryption-encryption scheme (EDE). [10]

The formalism of encryption and decryption 3DES is described in formula 4:

$$c = DES_{k3}\left(DES_{k2}^{-1}\left(DES_{k1}(p)\right)\right) \tag{4}$$

3DES runs three times DES algorithm, K_1, K_2, K_3 are three different keys. There are variants of 3DES algorithm as: encryption, decryption, encryption or encryption, encryption, encryption. In the case of $K_1 = K_3$, 3DES is called 3DES with two keys (formula 5).

$$\begin{aligned} c &\leftarrow e_{k1}\left(d_{k2}\left(e_{k1}(p)\right)\right) \\ p &\leftarrow d_{k1}\left(e_{k2}\left(d_{k1}(p)\right)\right) \end{aligned} \tag{5}$$

3DES algorithm has the advantage of the resistance at the “meet-in-middle” attacks, but it is too slow. So, there was proposed a 128-bit block cipher, called Advanced Encryption Standard.

Advanced Encryption Standard (AES) Algorithm

The AES is a formal encryption method adopted in 2000 by the National Institute of Standards and Technology of the US Government. Joan Daeman and Vincent Rijmen proposed the method, initially called the Rijndael algorithm. [1]

The algorithm is a symmetric algorithm and uses only one key of the three encryption keys possible: 128 bits (16 bytes), 192 bits (24 bytes) or 256 bits (32 bytes).

The AES algorithm converts a block of 128 bits to 128 bits of ciphertext. The plaintext is converted in a 4x4 matrix, called “state”. The initial encryption key is expanded into a table of 32-bit values. After that, the table is subdivided into groups of 32-bit values. The number of keys depends on the initial size of the key. A “round” is the base unit of transformation. The number of rounds depends on the key size: 10 for a key with a 128- bit length, 12 for a key with

a 192-bit length and 14 for a key with a 256-bit length, respectively. The algorithm of each round consists of four steps:

1. subbytes(state) – consists in a non-linear substitution of each byte within the state matrix with a new value,
2. shiftrows(state) – operates on each row of the "state" and it is a transposition cipher that permutes the positions of the row's elements,
3. mixcolumns(state) – operates on each column of the "state" and consists in four iterations,
4. addroundkey(state, key) – a Round Key is added using an exclusive OR operation.

The final round is slightly modified, as the "mixcolumns" function is removed.

For further information on AES, see [15].

At step "subbyte" there is used a nonlinear and invertible substitution table (Rijndael S-box) which operates on bytes. The substitution is obtained by means of composing two transformations [15]:

1. each byte is replaced with its inverse multiplicative in the finite field $GF(2^8)$. The 00 byte is mapped to itself.
2. the result is processed using an affine transformation over GF(2).

Each byte of the state is replaced with the element from S-box corresponding to the column given by the least significant nibble, and the row given by the most significant nibble.

The Rijndael S-box is described in the table 4. The key schedule procedure uses the same S-box. The complexity and the resistance of the AES algorithm are given by the transformations on the blocks of bits using the S-boxes. The Rijndael S-box is resistant to linear and differential cryptanalysis.

Table 4. Rijndael S-BOX

63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16

In the matrix presented in table 5, there is presented a comparison between the performances of the triple DES and AES algorithms.

Table 5. 3DES vs. AES

	Triple DES	***AES***
Key size	168 bits	129/192/256 bits
Block size	64 bits	128 bits
Speed	Low	Higher
Speed depends on key size	No	Yes
Strength	Moderate	Greater

The security of the cryptosystem depends on the substitution boxes. Shannon defined two design principles related to a secure cryptographic system:

- the principle of confusion:

  "the cipher text statistics should depend on the plaintext statistics in a manner too complicated to be exploited by the cryptanalyst."

- the principle of diffusion:

  "each digit of the plaintext and each digit of the secret key should influence many digits of the cipher text" [9]

A good cryptosystem has to respect the avalanche effect that is stated as "small change at the input resulting in a large change at the output is referred". [6]

Design of the Substitution Tables (S-boxes) in Cryptosystems

Generalizing, a S-box can be defined as a Boolean mapping:

$$S-box: B_2^m \to B_2^n, \text{ where } B_2 = \{0,1\}.$$

$\{x_1, x_2, \ldots, x_m\} \to \{y_1, y_2, \ldots, y_n\}$ and $y_i = f_i(x_1, x_2, \ldots x_m)$, where f_i is a Boolean function. A Boolean function can be described using a truth table.

A Boolean function is balanced if in the output column of the truth table the number of 0's equals the number of 1's. The Hamming weight of a binary vector is the number of 1's. The Hamming distance between two vectors is the numbers of items through which they differ.

To understand the role of S-boxes it is necessary to define the concept of cryptanalysis. The cryptanalysis is the process of finding the plain text without knowing the cryptographic key. The cryptosystems can be attacked using linear and differential cryptanalysis. The linear cryptanalysis makes linear approximation of the component functions of the cryptographic algorithm. Matsui proposed the linear cryptanalysis method presented in [7]. The differential cryptanalysis consists in analyzing the changes occurred as a result of changing the inputs. [3]

S-boxes have to resist to the linear cryptanalysis and differential cryptanalysis. In order to achieve these desiderata, Lineham and Gilliver [6] stated the necessary characteristics of S-boxes:

"1. High level of compliance with the Strict Avalanche Criteria;

2. Non-linear;

3. High degree resistance to Differential Cryptanalysis."

Strict Avalanche Criteria (SAC), introduced by Webster and Tavares in [12], refers to changes produced in the outputs due to the changes in inputs. SAC is a combination of the concepts of completeness and avalanche effect: "If a cryptographic function is to satisfy the strict avalanche criterion, then each output bit should change with a probability of one half whenever a single input bit is complemented." [12]

A Boolean function satisfies the SAC criteria if the following condition is satisfied (formula 6):

$$f(x) \oplus f(x \oplus \alpha) \text{ is balanced for any } \alpha \text{ such that } wt(\alpha) = 1 \tag{6}$$

where $wt(\alpha)$ is the hamming weight of α.

The property of non-linearity refers to the fact that the component functions should not be expressed as linear functions of the inputs. Nyberg demonstrated in [8] that the perfect non-linear transformations exist and can be implemented using two construction methods. Also, it is provided a corollary: " For a perfect nonlinear binary S-box the dimension of the input space is at least twice the dimension of the output space".

The Rijndael S-box satisfies the condition of resistance to Differential Cryptanalysis.

To analyze the differential cryptanalysis, there are compared the XOR of two inputs (inputs difference) and the XOR of their outputs (outputs difference) of the S-boxes.

According to the definition of the Boolean functions, the above-mentioned criteria may be expressed as: non-linear component functions of the algorithm, balanced component functions and balanced, highly non-linear, high algebraic degree combinations of functions.

Conclusions

The substitution boxes are used in the cryptographic systems in order to increase the complexity and security of the algorithms implemented in the cryptographic systems. Goldreich defined two steps to design good cryptographic systems: a definitional step and a constructive step. [5]

In the constructive step, there are designed good S-boxes by means of composing non-linear Boolean functions. The paper entitled A Survey on the Usage of Substitution Table in the Cryptosystems presents the role of S-boxes, also known as substitution table or look-up table, in the cryptosystems. Changing S-boxes can generate new cryptographic algorithms. The main role of S-boxes is to provide a non-linear and random distribution of the plaintext messages over the space of the cipher- text messages.

References

1. Allman, S. - *Encryption and Security: the Advanced Encryption Standard*, http://www.edn.com /contents/images/253789.pdf, accessed on 10 October, 2009
2. Atanasiu, A. - *Criptografie*, http://www.galaxyng.com/adrian_atanasiu/cript.htm/, accessed on 1st December, 2009
3. Biham, E., Shamir, A. - Diferential Cryptanalysis of DES-like Cryptosystems, *Advances in Cryptology* - CRYPTO'90, *Lecture Notes in Computer Science*, pp. 2-21, Springer-Verlag, 1991
4. Cangea, O. - *Transmisia si criptarea datelor*, Ed. Matrix Rom, Bucuresti, 2008
5. Goldreich, O. - Cryptography and Cryptographic Protocols, *Distributed Computing*, Vol. 16, Issue 2-3, pp. 177–199, 2003
6. Lineham, A., Gulliver, T.A. - Heuristic S-box Design, *Contemporary Engineering Sciences*, Vol. 1, no. 4, pp. 147 – 168, 2008

7. Matsui, M. - Linear Cryptanalysis Method for DES cipher, *Advances in Cryptology* - EUROCRYPT'93, *Lecture Notes in Computer Science*, pp. 386-397, Springer-Verlag, 1994
8. Nyberg, K. - *Perfect Non-linear S-boxes*, Springer Verlag, 1998
9. Shannon, C.E. - Communication Theory of Secrecy Systems, *Bell System Technical Journal*, 28, pp. 656–715, 1949
10. Tuchman, W. - Hellman Presents No Shortcut Solutions to DES, *IEEE Spectrum*, July 1979
11. Wang, Y. - *Confirmation of Shannon's Mistake about Perfect Secrecy of One-time-pad*, e-print arXiv:0709.4420, 2007
12. Webster A.F., Tavares, S.E. - On the Design of S-boxes, *Advances in Cryptology*, Crypto '85, *Lecture Notes in Computer Science*, vol. 219, pp. 523-534, 1985
13. Wenbo, M. - *Modern Cryptography Theory and Practice*, Prentice Hall PTR, 2003
14. *** - Data Encryption Standard, *Federal Information Processing Standard (FIPS) Publication* 46, National Bureau of Standards, U.S. Department of Commerce, Washington D.C., 1977
15. *** - *Specifications for the Advanced Encryption Standard (AES)*, 2001, http://csrc.nist.gov /publications/fips/fips197/fips-197.pdf, accessed on 1st December, 2009

O trecere în revistă a utilizării tabelelor de substituţie în algoritmii DES şi AES

Rezumat

În acest articol sunt prezentate tabelele de substituţie utilizate în criptosisteme (DES şi AES) şi rolul lor în creşterea securităţii algoritmilor criptografici. Tabelele de substituţie sunt funcţii de permutare neliniare şi sunt formalizate matematic prin funcţii booleene. O modalitate pentru a obţine noi algoritmi criptografici o constituie utilizarea de tabele de substituţie randomizate. Direcţiile de cercetare sunt concentrate pe construirea de funcţii booleene cu proprietăţi criptografice bune.