DATA ENCRYPTION STANDARD



Indian Institute of Technology (ISM) Dhanbad

Akshay Dobariya

Guidede By:

16JE001934

Dr. Arup Kr. Pal

4th semester

B.tech Computer Science and Engineering

INTRODUCTION

The DES (Data Encryption Standard) algorithm is the most widely used encryption algorithm in the world. For many years, and among many people, "secret code making" and DES have been synonymous. And despite the recent coup by the Electronic Frontier Foundation in creating a \$220,000 machine to crack DES-encrypted messages, DES will live on in government and banking for years to come through a life- extending version called "triple-DES."

How does DES work? This article explains the various steps involved in DES-encryption, illustrating each step by means of a simple example. Since the creation of DES, many other algorithms (recipes for changing data) have emerged which are based on design principles similar to DES. Once you understand the basic transformations that take place in DES, you will find it easy to follow the steps involved in these more recent algorithms.

But first a bit of history of how DES came about is appropriate, as well as a look toward the future.

The National Bureau of Standards

On May 15, 1973, during the reign of Richard Nixon, the National Bureau of Standards (NBS) published a notice in the Federal Register soliciting proposals for cryptographic algorithms to protect data during transmission and storage. The notice explained why encryption was an important issue.

Over the last decade, there has been an accelerating increase in the accumulations and communication of digital data by government, industry and by other organizations in the private sector. The contents of these communicated and stored data often have very significant value and/or sensitivity. It is now common to find data transmissions which constitute funds transfers of several million dollars, purchase or sale of securities, warrants for arrests or arrest and conviction records being communicated between law enforcement agencies, airline reservations and ticketing representing investment and value both to the airline and passengers, and health and patient care records transmitted among physicians and treatment centers.

The increasing volume, value and confidentiality of these records regularly transmitted and stored by commercial and government agencies has led to heightened recognition and concern over their exposures to unauthorized access and use. This misuse can be in the form of theft or defalcations of data records representing money, malicious modification of business inventories or the interception and misuse of confidential information about people. The need for protection is then apparent and urgent.

It is recognized that encryption (otherwise known as scrambling, enciphering or privacy transformation) represents the only means of protecting such data during

transmission and a useful means of protecting the content of data stored on various media, providing encryption of adequate strength can be devised and validated and is inherently integrable into system architecture. The National Bureau of Standards solicits proposed techniques and algorithms for computer data encryption. The Bureau also solicits recommended techniques for implementing the cryptographic function: for generating, evaluating, and protecting cryptographic keys; for maintaining files encoded under expiring keys; for making partial updates to encrypted files; and mixed clear and encrypted data to permit labelling, polling, routing, etc. The Bureau in its role for establishing standards and aiding government and industry in assessing technology, will arrange for the evaluation of protection methods in order to prepare guidelines.

NBS waited for the responses to come in. It received none until August 6, 1974, three days before Nixon's resignation, when IBM submitted a candidate that it had developed internally under the name LUCIFER. After evaluating the algorithm with the help of the National Security Agency (NSA), the NBS adopted a modification of the LUCIFER algorithm as the new Data Encryption Standard (DES) on July 15, 1977.

DES was quickly adopted for non-digital media, such as voice-grade public telephone lines. Within a couple of years, for example, International Flavors and Fragrances was using DES to protect its valuable formulas transmitted over the phone ("With Data Encryption, Scents Are Safe at IFF," *Computerworld* 14, No. 21, 95 (1980).)

Meanwhile, the banking industry, which is the largest user of encryption outside government, adopted DES as a wholesale banking standard. Standards for the wholesale banking industry are set by the American National Standards Institute (ANSI). ANSI X3.92, adopted in 1980, specified the use of the DES algorithm.

How DES Works in Detail

DES is a *block cipher*--meaning it operates on plaintext blocks of a given size (64-bits) and returns ciphertext blocks of the same size. Thus DES results in a *permutation* among the 2^64 (read this as: "2 to the 64th power") possible arrangements of 64 bits, each of which may be either 0 or 1. Each block of 64 bits is divided into two blocks of 32 bits each, a left half block **L** and a right half **R**. (This division is only used in certain operations.)

Example: Let \mathbf{M} be the plain text message $\mathbf{M} = 0123456789 \text{ABCDEF}$, where \mathbf{M} is in hexadecimal (base 16) format. Rewriting \mathbf{M} in binary format, we get the 64-bit block of text:

 $\mathbf{M} = 0000\ 0001\ 0010\ 0011\ 0100\ 0101\ 0110\ 0111\ 1000\ 1001\ 1010\ 1011\ 1100\ 1101$

 $L = 0000\ 0001\ 0010\ 0011\ 0100\ 0101\ 0110\ 0111$ $R = 1000\ 1001\ 1010\ 1011\ 1100\ 1101\ 1110$

The first bit of **M** is "0". The last bit is "1". We read from left to right.

DES operates on the 64-bit blocks using *key* sizes of 56- bits. The keys are actually stored as being 64 bits long, but every 8th bit in the key is not used (i.e. bits numbered 8, 16, 24, 32, 40, 48, 56, and 64). However, we will nevertheless number the bits from 1 to 64, going left to right, in the following calculations. But, as you will see, the eight bits just mentioned get eliminated when we create subkeys.

Example: Let **K** be the hexadecimal key $\mathbf{K} = 133457799BBCDFF1$. This gives us as the binary key (setting 1 = 0001, 3 = 0011, etc., and grouping together every eight bits, of which the last one in each group will be unused):

 $\mathbf{K} = 00010011\ 00110100\ 01010111\ 01111001\ 10011011\ 10111100\ 110111111$

The DES algorithm uses the following steps:

Step 1: Create 16 sub keys, each of which is 48-bits long.

The 64-bit key is permuted according to the following table, **PC-1**. Since the first entry in the table is "57", this means that the 57th bit of the original key **K** becomes the first bit of the permuted key **K**+. The 49th bit of the original key becomes the second bit of the permuted key. The 4th bit of the original key is the last bit of the permuted key. Note only 56 bits of the original key appear in the permuted key.

PC-1									
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			

Example: From the original 64-bit key

 $\mathbf{K} = 00010011\ 00110100\ 01010111\ 01111001\ 10011011\ 10111100\ 110111111$

we get the 56-bit permutation

 \mathbf{K} + = 1111000 0110011 0010101 0101111 0101010 1011001 1001111 0001111

Next, split this key into left and right halves, C_0 and D_0 , where each half has 28 bits.

Example: From the permuted key K+, we get

 $C_{\theta} = 1111000 \ 0110011 \ 0010101 \ 0101111$ $D_{\theta} = 0101010 \ 1011001 \ 1001111 \ 0001111$

With C_0 and D_0 defined, we now create sixteen blocks C_n and D_n , 1 <= n <= 16. Each pair of blocks C_n and D_n is formed from the previous pair C_{n-1} and D_{n-1} , respectively, for n = 1, 2, ..., 16, using the following schedule of "left shifts" of the previous block. To do a left shift, move each bit one place to the left, except for the first bit, which is cycled to the end of the block.

Iteration Number	Number of Left Shifts
1	1
2	1
3	2
4	2
5	2
6	2
7	2
8	2
9	1
10	2
11	2
12	2
13	2
14	2
15	2
16	1

This means, for example, C_3 and D_3 are obtained from C_2 and D_2 , respectively, by two left shifts, and C_{16} and D_{16} are obtained from C_{15} and D_{15} , respectively, by one left shift. In all cases, by a single left shift is meant a rotation of the bits one place to the left, so that after one left shift the bits in the 28 positions are the bits that were previously in positions 2, 3,..., 28, 1.

Example: From original pair pair C_{θ} and D_{θ} we obtain:

```
C_{\theta} = 11110000110011001010101011111

D_{\theta} = 010101010101100111110001111
```

```
C_I = 11100001100110010101010111111

D_I = 1010101011011100111100011110
```

```
C_2 = 11000011001100101010101111111

D_2 = 0101010110011001111000111101
```

```
C_3 = 00001100110010101010111111111
D_3 = 01010110011001111100011110101
```

```
C_4 = 00110011001010101111111111100

D_4 = 010110011001111100011110101
```

```
C_5 = 110011001010101011111111110000

D_5 = 011001100111110001111010101
```

```
C_6 = 001100101010101111111111000011

D_6 = 100110011111000111101010101
```

```
C_7 = 110010101010111111111100001100

D_7 = 01100111110001111010101010110
```

```
C_8 = 001010101011111111110000110011

D_8 = 10011110001111101010101011001
```

```
C_9 = 010101010111111111100001100110

D_9 = 00111100011111010101010110011
```

```
C_{10} = 010101011111111110000110011001

D_{10} = 1111000111101010101011001100
```

$$C_{II} = 010101111111111000011001100101$$

 $D_{II} = 1100011110101010101100110011$

$$C_{12} = 010111111111100001100110010101$$

 $D_{12} = 0001111010101010110011001111$

$$C_{13} = 01111111110000110011001010101$$

 $D_{13} = 0111101010101011001100111100$

$$C_{14} = 1111111000011001100101010101$$

 $D_{14} = 1110101010101100110011110001$

$$C_{15} = 11111000011001100101010101111$$

 $D_{15} = 10101010101100110011111000111$

$$C_{16} = 11110000110011001010101011111$$

 $D_{16} = 0101010101111001100111110001111$

We now form the keys K_n , for 1 <= n <= 16, by applying the following permutation table to each of the concatenated pairs C_nD_n . Each pair has 56 bits, but **PC-2** only uses 48 of these.

PC-2									
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				

Therefore, the first bit of K_n is the 14th bit of C_nD_n , the second bit the 17th, and so on, ending with the 48th bit of K_n being the 32th bit of C_nD_n .

Example: For the first key we have $C_1D_1 = 1110000 \ 1100110 \ 0101010 \ 1011111 \ 1010101 \ 0110011 \ 0011110 \ 0011110$

which, after we apply the permutation PC-2, becomes

For the other keys we have

So much for the subkeys. Now we look at the message itself.

Step 2: Encode each 64-bit block of data.

There is an *initial permutation* **IP** of the 64 bits of the message data **M**. This rearranges the bits according to the following table, where the entries in the table show the new arrangement of the bits from their initial order. The 58th bit of **M** becomes the first bit of **IP**. The 50th bit of **M** becomes the second bit of **IP**. The 7th bit of **M** is the last bit of **IP**.

			IP				
- 0	- 0		0.4	0.6	4.0	4.0	
58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

Example: Applying the initial permutation to the block of text **M**, given previously, we get

 $\mathbf{M} = 0000\ 0001\ 0010\ 0011\ 0100\ 0101\ 0110\ 0111\ 1000\ 1001\ 1010\ 1011\ 1100\ 1101$

 $\mathbf{IP} = 1100\ 1100\ 0000\ 0000\ 1100\ 1100\ 1111\ 1111\ 1111\ 0000\ 1010\ 1010\ 1111\ 0000$

Here the 58th bit of **M** is "1", which becomes the first bit of **IP**. The 50th bit of **M** is "1", which becomes the second bit of **IP**. The 7th bit of **M** is "0", which becomes the last bit of **IP**.

Next divide the permuted block **IP** into a left half L_{θ} of 32 bits, and a right half R_{θ} of 32 bits.

Example: From **IP**, we get L_0 and R_0

 $L_{\theta} = 1100 \ 1100 \ 0000 \ 0000 \ 1100 \ 1100 \ 1111 \ 1111$ $R_{\theta} = 1111 \ 0000 \ 1010 \ 1010 \ 1111 \ 0000 \ 1010$

We now proceed through 16 iterations, for 1 <= n <= 16, using a function f which operates on two blocks--a data block of 32 bits and a key K_n of 48 bits--to produce a block of 32 bits. Let + denote XOR addition, (bit-by-bit addition modulo 2). Then for \mathbf{n} going from 1 to 16 we calculate

$$L_n = R_{n-1}$$

 $R_n = L_{n-1} + f(R_{n-1}, K_n)$

This results in a final block, for n = 16, of $L_{16}R_{16}$. That is, in each iteration, we take the right 32 bits of the previous result and make them the left 32 bits of the current

step. For the right 32 bits in the current step, we XOR the left 32 bits of the previous step with the calculation f.

Example: For n = 1, we have

 $K_I = 000110 \ 110000 \ 001011 \ 101111 \ 111111 \ 000111 \ 000001 \ 110010$ $L_I = R_0 = 1111 \ 0000 \ 1010 \ 1010 \ 1111 \ 0000 \ 1010 \ 1010$ $R_I = L_0 + f(R_0, K_I)$

It remains to explain how the function f works. To calculate f, we first expand each block R_{n-1} from 32 bits to 48 bits. This is done by using a selection table that repeats some of the bits in R_{n-1} . We'll call the use of this selection table the function E. Thus $E(R_{n-1})$ has a 32 bit input block, and a 48 bit output block.

Let **E** be such that the 48 bits of its output, written as 8 blocks of 6 bits each, are obtained by selecting the bits in its inputs in order according to the following table:

E BIT-SELECTION TABLE

32	1	2	3	4	5
52	Τ.	2	5	-	J
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

Thus the first three bits of $\mathbf{E}(\mathbf{R}_{n-1})$ are the bits in positions 32, 1 and 2 of \mathbf{R}_{n-1} while the last 2 bits of $\mathbf{E}(\mathbf{R}_{n-1})$ are the bits in positions 32 and 1.

Example: We calculate $\mathbf{E}(\mathbf{R}_{\theta})$ from \mathbf{R}_{θ} as follows:

 $\mathbf{R}_{\theta} = 1111\ 0000\ 1010\ 1010\ 1111\ 0000\ 1010\ 1010$ $\mathbf{E}(\mathbf{R}_{\theta}) = 011110\ 100001\ 010101\ 010101\ 011110\ 100001\ 010101\ 010101$

(Note that each block of 4 original bits has been expanded to a block of 6 output bits.)

Next in the f calculation, we XOR the output $\mathbf{E}(\mathbf{R}_{n-1})$ with the key \mathbf{K}_n :

$$\mathbf{K}_n + \mathbf{E}(\mathbf{R}_{n-1}).$$

Example: For K_I , $\mathbf{E}(R_{\theta})$, we have

 $K_I = 000110\ 110000\ 001011\ 101111\ 1111111\ 000111\ 000001\ 110010$ $\mathbf{E}(\mathbf{R}_{\theta}) = 011110\ 100001\ 010101\ 010101\ 011110\ 100001\ 010101\ 010101$ $K_I + \mathbf{E}(\mathbf{R}_{\theta}) = 011000\ 010001\ 011110\ 111010\ 100001\ 100110\ 010100\ 100111.$

We have not yet finished calculating the function f. To this point we have expanded R_{n-1} from 32 bits to 48 bits, using the selection table, and XORed the result with the key K_n . We now have 48 bits, or eight groups of six bits. We now do something

strange with each group of six bits: we use them as addresses in tables called "S boxes". Each group of six bits will give us an address in a different S box. Located at that address will be a 4 bit number. This 4 bit number will replace the original 6 bits. The net result is that the eight groups of 6 bits are transformed into eight groups of 4 bits (the 4-bit outputs from the S boxes) for 32 bits total.

Write the previous result, which is 48 bits, in the form:

$$K_n + \mathbf{E}(R_{n-1}) = B_1B_2B_3B_4B_5B_6B_7B_8,$$

where each B_i is a group of six bits. We now calculate

$$S_1(B_1)S_2(B_2)S_3(B_3)S_4(B_4)S_5(B_5)S_6(B_6)S_7(B_7)S_8(B_8)$$

where $S_i(B_i)$ referres to the output of the *i*-th **S** box.

To repeat, each of the functions S1, S2,..., S8, takes a 6-bit block as input and yields a 4-bit block as output. The table to determine S_1 is shown and explained below:

s1

Col	umn	Number	•

No.	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

If S_I is the function defined in this table and B is a block of 6 bits, then $S_I(B)$ is determined as follows: The first and last bits of B represent in base 2 a number in the decimal range 0 to 3 (or binary 00 to 11). Let that number be i. The middle 4 bits of B represent in base 2 a number in the decimal range 0 to 15 (binary 0000 to 1111). Let that number be j. Look up in the table the number in the i-th row and j-th column. It is a number in the range 0 to 15 and is uniquely represented by a 4 bit block. That block is the output $S_I(B)$ of S_I for the input B. For example, for input block B = 011011 the first bit is "0" and the last bit "1" giving 01 as the row. This is row 1. The middle four bits are "1101". This is the binary equivalent of decimal 13, so the column is column number 13. In row 1, column 13 appears 5. This determines the output; 5 is binary 0101, so that the output is 0101. Hence $S_I(011011) = 0101$.

The tables defining the functions $S_1,...,S_8$ are the following:

			S1				
14 4 0 15 4 1 15 12	13 1 7 4 14 8 8 2	2 15 11 14 2 13 13 6 2 4 9 1	8 3 1 10 11 15 7 5	10 6 6 12 12 9 11 3	12 5 11 9 7 3 14 10	5 3 10 5	8
			S2				
15 1 3 13 0 14 13 8	8 14 4 7 7 11 10 1	15 2 8 10 4 13	4 9 14 12 1 5 2 11	0 1 8 12	13 12 10 6 6 9 12 0	9 11 3 2	5
			s 3				
10 0 13 7 13 6 1 10	9 14 0 9 4 9 13 0	6 3 15 3 4 6 8 15 3 6 9 8	10 2 0 11	13 12 8 5 1 2 15 14	7 11 14 12 12 5 3 11	11 15 10 14	1 7
			S4				
7 13 13 8 10 6 3 15	14 3 11 5 9 0 0 6	0 6 9 6 15 0 12 11 7 10 1 13	3 4 13 15	2 8 7 2 1 3 4 5	5 11 12 1 14 5 11 12	10 14 2 8	9
			S 5				
2 12 14 11 4 2 11 8	4 1 2 12 1 11 12 7	7 10 11 4 7 13 10 13 7 1 14 2	6 8 1 5 8 15 13 6	5 3 0 15 9 12 15 0	15 13 10 3 5 6 9 10	9 8 3 0	6 14
			s 6				
12 1 10 15 9 14 4 3	10 15 4 2 15 5 2 12		8 0 5 6 3 7 10 11		10 1	11 3 13 11	8
			S7				
4 11 13 0 1 4 6 11	2 14 11 7 11 13 13 8	4 9 1 12 3 7		3 5 15 6	12 2 8 0	10 6 15 8 5 9 2 3	6 2
			S8				
13 2 1 15 7 11 2 1	8 4 13 8 4 1 14 7	10 3 7 9 12 14		5 6 6 10	14 5 11 0 13 15 0 3	14 9 3 5	2 8

Example: For the first round, we obtain as the output of the eight ${\bf S}$ boxes:

 $K_I + \mathbf{E}(R_0) = 011000\ 010001\ 011110\ 111010\ 100001\ 100110\ 010100\ 100111.$

 $S_1(B_1)S_2(B_2)S_3(B_3)S_4(B_4)S_5(B_5)S_6(B_6)S_7(B_7)S_8(B_8) = 0101\ 1100\ 1000\ 0010\ 1011\ 0101\ 1001\ 0111$

The final stage in the calculation of f is to do a permutation P of the S-box output to obtain the final value of f:

$$f = P(S_1(B_1)S_2(B_2)...S_8(B_8))$$

The permutation **P** is defined in the following table. **P** yields a 32-bit output from a 32-bit input by permuting the bits of the input block.

		Р	
16	7	20	21
29	12	28	17
1	15	23	26
5	18	31	10
2	8	24	14
32	27	3	9
19	13	30	6
22	11	4	25

Example: From the output of the eight **S** boxes:

$$S_1(B_1)S_2(B_2)S_3(B_3)S_4(B_4)S_5(B_5)S_6(B_6)S_7(B_7)S_8(B_8) = 0101\ 1100\ 1000\ 0010\ 1011$$

we get

$$f = 0010\ 0011\ 0100\ 1010\ 1010\ 1001\ 1011\ 1011$$

$$\mathbf{R}_1 = \mathbf{L}_0 + \mathbf{f}(\mathbf{R}_0, \mathbf{K}_1)$$

- = 1100 1100 0000 0000 1100 1100 1111 1111
- + 0010 0011 0100 1010 1010 1001 1011 1011
- = 1110 1111 0100 1010 0110 0101 0100 0100

In the next round, we will have $L_2 = R_I$, which is the block we just calculated, and then we must calculate $R_2 = L_I + f(R_I, K_2)$, and so on for 16 rounds. At the end of the sixteenth round we have the blocks L_{16} and R_{16} . We then *reverse* the order of the two blocks into the 64-bit block

$$R_{16}L_{16}$$

and apply a final permutation **IP-1** as defined by the following table:

IP-1								
40	8	48	16	56	24	64	32	
39	7	47	15	55	23	63	31	
38	6	46	14	54	22	62	30	
37	5	45	13	53	21	61	29	
36	4	44	12	52	20	60	28	
35	3	43	11	51	19	59	27	
34	2	42	10	50	18	58	26	
33	1	41	9	49	17	57	25	

That is, the output of the algorithm has bit 40 of the preoutput block as its first bit, bit 8 as its second bit, and so on, until bit 25 of the preoutput block is the last bit of the output.

Example: If we process all 16 blocks using the method defined previously, we get, on the 16th round,

 $L_{16} = 0100\ 0011\ 0100\ 0010\ 0011\ 0010\ 0011\ 0100$ $\mathbf{R}_{16} = 0000\ 1010\ 0100\ 1100\ 1101\ 1001\ 1001\ 0101$

We reverse the order of these two blocks and apply the final permutation to

 $R_{16}L_{16} = 00001010\ 01001100\ 11011001\ 10010101\ 01000011\ 01000010\ 00110010$ 00110100

 $IP^{-1} = 10000101\ 11101000\ 00010011\ 01010100\ 00001111\ 00001010\ 10110100$ 00000101

which in hexadecimal format is

85E813540F0AB405.

This is the encrypted form of M = 0123456789ABCDEF: namely, C =85E813540F0AB405.

Decryption is simply the inverse of encryption, follwing the same steps as above, but reversing the order in which the subkeys are applied.