A Semi-Linear Relation between Inputs and Outputs of DES S-Boxes

Mohammad Etemad
Koç University
Rumelifeneri yolu, 34450, Sariyer, Istanbul, Turkey

Saeedeh Anvari
Koç University
Rumelifeneri yolu, 34450, Sariyer, Istanbul, Turkey

ABSTRACT

The Data Encryption Standard (DES) is the most widely used cryptosystem developed by a team of cryptographers working at IBM. DES has been cryptanalyzed intensively by researchers, but no efficient attack has been found on DES so far. This is mainly due to the lack of an obvious algebraic relation in the structure of S-boxes, which makes it impossible to use known methods to attack DES. S-boxes are the nonlinear part of DES with strong properties. This paper presents a semi-linear relation between input and output of S-boxes that could be used to cryptanalyze DES. This is based on Differential Cryptanalysis method proposed by Biham and Shamir.

Keywords

DES, S-Box, Differential Cryptanalysis.

1. INTRODUCTION

The Data Encryption Standard (DES) is a US national standard and it has been the most widely used cryptosystem worldwide for most of the last 30 years. It is a block cipher which encrypts blocks of length of 64 bits to produce ciphertext blocks of the same size under the control of a 56-bit key (for more information, see [2, 7]). DES is composed of 16 rounds which all perform the same operation. In every round, a different 48-bit subkey is used which is derived from the main 56-bit key. (We leave out the initial and final permutation (IP, IP^{-1}) because they have no effect on the security of the system, as well as on our analysis.) Each round takes 32-bit inputs L_{i-1} and R_{i-1} from the previous round and produces 32-bit outputs L_i and R_i for $1 \leq i \leq 16$, as follows [3]:

$$\begin{cases} L_i = R_i \\ R_i = L_{i-1} \oplus F(R_{i-1}, K_i) \end{cases}$$
$$where\ F(R_{i-1}, K_i) = P(S(E(R_{i-1}) \oplus K_i))$$

The structure of a round of DES is depicted in Fig. 1. All parts of a round of DES, except S-boxes, are linear, and hence reversible. S-boxes bring nonlinearity to block ciphers and strengthen their cryptographic security [12], which makes crypt-analysis of DES more complicated. This means that S-boxes play the most important role in DES-like cryptosystems. Once we found a relationship between their inputs and output, we can cryptanalyze DES much easier.

In this paper, a semi-linear relation between inputs and outputs of S-boxes is presented whose aim is to ease the cryptanalysis of DES. The relation is based on Differential Cryptanalysis introduced by Biham and Shamir [4, 5]. This way, we can find the inputs by XORing outputs of the S-Box.

This paper is organized as follows: Section 2 reviews previous works on DES. In section 3, the proposed method is introduced and conclusions will follow in section 4.

2. Related Work

Since the introduction of DES, there have been a lot of studies about the properties of S-boxes. Thanks to these studies, much of these properties are revealed, also the primary properties were never published [8].

There were some criteria in the design of DES S-boxes, as Brickell et al. wrote in [9]: “We would like to know what properties the S-boxes were designed to satisfy. This information was never published and in fact, the only source for specific “design principles” appears to be responses from the NSA to a study of the DES made by the Lexar Corporation. There were included in the report of the second workshop on the DES held by the NBS in 1976.”

The NSA mentioned the following “design criteria” in design of S-boxes [8, 9]:

- P1. No S-box is a linear or affine function of the input.
- P2. Changing 1 input bit to an S-box results in changing at least 2 output bits.
- P3. $S(x)$ and $S(x \oplus 001100)$ must differ in at least 2 bits.

The followings were labeled by the NSA as “caused by design criteria” [8]:

- P4. $S(x) \neq S(X \oplus 11ab00)$ for any choice of a and b.
- P5. The S-boxes were chosen to minimize the difference between the number of 1’s and 0’s in any S-box output when any single input bit is held constant.

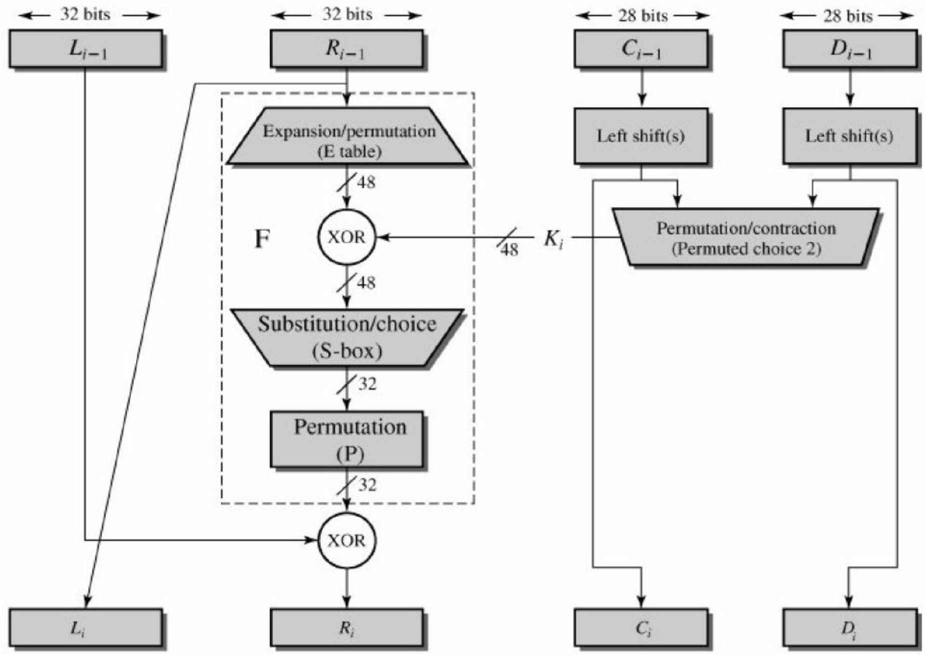

Fig. 1- The structure of a round of DES [2]

D. Coppersmith listed and discussed the criteria for S-boxes and permutations of DES [10]. He stated 8 design criteria for S-boxes and 3 for permutation. In 1991, Dawson and Tavares introduced an expanded set of design criteria for creating good S-boxes based on information theoretic concepts [11]. They also showed that an S-box that meets those criteria is immune to differential cryptanalysis. "We could not find S-boxes with substantially better information theoretic properties than the S-boxes of DES and which also meet the acknowledged DES design criteria", also they mentioned.

In 1990, Biham and Shamir discovered a powerful attack method, named *Differential Cryptanalysis*, which is applicable to any block cipher [4]. The basic method uses pairs of plaintext related by a constant difference that can be defined in several ways through XOR operation. Differences of the corresponding ciphertexts are then computed, hoping to detect statistical patterns in their distribution.

Beside these properties, applying XOR operation to triplets and quadruples of outputs of S-boxes yields recovery of their inputs that will be discussed in the following section.

3. The New Behavior of S-Boxes

The research done on DES and its S-boxes, especially after 1990, has identified many of their properties. But, no direct relation between inputs and outputs of S-boxes has been found yet. Here, such a relation is going to be described. We are about to find the inputs of S-boxes by XORing their outputs with high certainty.

Suppose that a, b are inputs of ith S-box, and $S_i(a), S_i(b)$ are corresponding outputs. We do XOR all possible distinct pairs of inputs (i.e. all possible pairs consisting of a, b). There are 2016 of such pairs ($\binom{64}{2} = 2016$). A little part of computation for S1 is shown in Table 1. A detailed analysis of this table reveals that only 12.2% (=246/2016) of rows consisting of $a \oplus b$, $S1(a) \oplus S1(b)$ are unique, i.e. in 12.2% of cases a, b could be recovered with certainty. On the other hand, in 87.8% of cases it is impossible to recover a, b with certainty due to the repeating behavior of rows. The repeating behavior of S1, S2, S3, and S4 is shown in Table 2. First row shows the number of repetition, the second row shows the number of rows repeated that many times, and the third row shoes the percentages for S-box S1. The row after them, presents the same information for the next S-boxes.

We go further, and perform similar operations with 3 ciphertexts. We do XOR all possible distinct triplets of inputs of S-boxes (i.e. all possible triplets consisting of a, b, c) in pairs. There are 41664 of such triplets ($\binom{64}{3} = 41664$). A little part of computation for S1 is shown in Table 3. A detailed analysis of this table reveals that 91.45% of rows consisting of $a \oplus b$, $a \oplus c$, $b \oplus c$, $S1(a) \oplus S1(b)$, $S1(a) \oplus S1(c), S1(b) \oplus S1(c)$ are unique, i.e. in 91.45% of cases a, b, c could be recovered with certainty. In other cases (just 8.55% of cases), it's impossible to recover a, b, c with certainty due to the repeating behavior of rows. The repeating behavior of S1, S2, S3, and S4 is shown in Table 4.

Table 1- Part of results of XORing pairs of S1 inputs.

a	b	$a \oplus b$	$S1(a) \oplus S1(b)$
0	1	1	14
0	2	2	10
0	3	3	1
0	4	4	3
…	…	…	…
2	18	16	14
2	19	17	2
2	20	22	2
2	21	23	8
…	…	…	…

Continuing on this way, we perform similar operations with 4 ciphertexts. We do XOR all possible distinct quadruples of inputs of S-boxes (i.e. all possible quadruples consisting of a, b, c, d) in pairs. There are 635376 of such quadruples ($\binom{64}{4} = 635376$). The results are interesting. A little part of computation for S1 is shown in Table 5. A detailed analysis of this table reveals that 99.6% of rows consisting of $a \oplus b$, $a \oplus c$, a$\oplus$ d, $b \oplus c$, $b \oplus d$, $c \oplus d$, $S1(a) \oplus S1(b)$, $S1(a) \oplus S1(c), S1(a) \oplus S1(d)$, $S1(b) \oplus S1(c)$, $S1(b) \oplus S1(d), S1(c) \oplus S1(d)$ are unique, i.e. in 99.6% of cases a, b, c, d could be recovered with high certainty. In other cases (just 0.4% of cases), however, it's impossible to recover a, b, c, d with high certainty due to the repeating behavior of rows. The repeating behavior of S1, S2, S3, and S4 is shown in Table 6.

Table 2- The repeating behavior of rows.

Repetition times		1	2	3	4	5	6	7	8	9
S1	no. of rows	246	232	168	84	46	24	12	1	-
	percentage	12.20%	23.02%	25.0%	16.67%	11.41%	7.14%	4.17%	0.40%	-
S2	no. of rows	248	216	165	94	50	15	12	4	1
	percentage	12.30%	21.43%	24.55%	18.65%	12.40%	4.46%	4.17	1.59%	0.45%
S3	no. of rows	256	234	141	99	53	20	8	4	-
	percentage	12.70%	23.21%	20.98%	19.64%	13.14%	5.95%	2.78%	1.59%	-
S4	no. of rows	96	268	128	156	-	24	-	29	-
	percentage	4.76%	26.59%	19.05%	30.95%	-	7.14%	-	11.51%	-

Table 3- Part of results of XORing possible pairs of S1 inputs.

a	b	c	$a \oplus b$	$a \oplus c$	$b \oplus c$	$S1(a) \oplus S1(b)$	$S1(a) \oplus S1(c)$	$S1(b) \oplus S1(c)$
0	1	2	1	2	3	14	10	4
0	1	3	1	3	2	14	1	15
0	1	4	1	4	5	14	3	13
0	1	5	1	5	4	14	9	7
0	1	6	1	6	7	14	15	1
…	…		…	…	…	…	…	…
6	9	61	15	59	52	15	7	8
6	9	62	15	56	55	15	1	14
6	9	63	15	57	54	15	12	3
6	10	11	12	13	1	14	3	13
6	10	12	12	10	6	14	10	4
…	…		…	…	…	…	…	…

Table 4- The repeating behavior of rows.

Repetition times		1	2	3	4
S1	no. of rows	38100	1687	62	1
	percentage	91.45%	8.10%	0.45%	0.01%
S2	no. of rows	37909	1780	65	-
	percentage	90.99%	8.54%	0.47%	-
S3	no. of rows	37802	1810	74	5
	percentage	90.73%	8.69%	0.53%	0.05%
S4	no. of rows	36280	2588	48	16
	percentage	87.08%	12.42%	0.35%	0.15%

Table 5- Part of results of XORing possible pairs of S1 inputs.

a	b	c	d	$a \oplus b$	$a \oplus c$	$a \oplus d$	$b \oplus c$	$b \oplus d$	$c \oplus d$	$S1(a) \oplus S1(b)$	$S1(a) \oplus S1(c)$	$S1(a) \oplus S1(d)$	$S1(b) \oplus S1(c)$	$S1(b) \oplus S1(d)$	$S1(c) \oplus S1(d)$
0	1	2	3	1	2	3	3	2	1	14	10	1	4	15	11
0	1	2	4	1	2	4	3	5	6	14	10	3	4	13	9
0	1	2	5	1	2	5	3	4	7	14	10	9	4	7	3
0	1	2	6	1	2	6	3	7	4	14	10	15	4	1	5
0	1	2	7	1	2	7	3	6	5	14	10	10	4	4	0
…	…			…	…	…	…	…	…	…	…	…	…	…	…
1	1	3	5	31	40	62	55	33	22	13	5	7	8	10	2
1	1	3	5	31	40	63	55	32	23	13	5	0	8	13	5
1	1	3	5	31	40	56	55	39	16	13	5	2	8	15	7
1	1	3	5	31	40	57	55	38	17	13	5	8	8	5	13
1	1	3	5	31	40	58	55	37	18	13	5	12	8	1	9
…	…			…	…	…	…	…	…	…	…	…	…	…	…

Table 6- The repeating behavior of rows.

Repetition times		1	2	3	4
S1	no. of rows	632891	1241	1	-
	percentage	99.61%	0.39%	0.00%	-
S2	no. of rows	632270	1553	-	-
	percentage	99.51%	0.49%	-	-
S3	no. of rows	632270	1550	2	-
	percentage	99.51%	0.49%	0.00%	-
S4	no. of rows	628296	3524	8	2
	percentage	98.89%	1.11%	0.00%	0.00%

4. Conclusion

In this paper, a semi-linear relation between inputs and output of S-boxes is presented which is based on multiple XORing, and can be used to cryptanalyze DES. Since S-boxes are the only nonlinear part of DES, simulating them with a linear relation will be of great importance. This relation, along with other cryptanalysis methods like Linear Cryptanalysis, will produce more efficient ways of attacking DES.

5. Acknowledgement

We would like to extend our gratitude to Dr. Rasool Jalili, for his endless support.

6. REFERENCES

[1] C. Paar, J. Pelzl, 2010. Understanding Cryptography, Springer-Verlag.

[2] W. Stallings, 2005. Cryptography and Network Security, 4th Edition, Prentice Hall.

[3] A. Menezes, P. Van Oorschot, S. Vanstone, 1996. Handbook of Applied Cryptography, CRC Press.

[4] E. Biham, A. Shamir, 1991. Differential Cryptanalysis of DES-like Cryptosystems, Journal of cryptology, Vol. 4, No. 1.

[5] E. Biham, A. Shamir, 1992. Differential Cryptanalysis of the full 16-round DES, Advances in Cryptology - Crypto ’92, Springer LNCS, Vol. 740, pp. 487–496. .

[6] M. Matsui, 1993. Linear Cryptanalysis Method for DES Cipher, in Advances in Cryptology, EUROCRYPT’93, Springer LNCS, Vol. 765, pp. 386–397.

[7] National Bureau of Standards, 1977. Data Encryption Standard, U.S. Department of Commerce.

[8] N.T. Courtois, G. Castagnos, L. Goubin, What do DES S-boxes Say to Each Other?, Cryptology ePrint Archive, http://eprint.iacr.org/

[9] E.F. Brickell, J.H. Moore, M.R. Purtill, 1986. Structure in the S-Boxes of DES**,** Crypto’86, Springer LNCS, Vol. 1440, pp. 3-7.

[10] Don Coppersmith, 1994. The Data Encryption Standard (DES) and its strength against attacks, IBM Journal of Research and Development, Vol. 38, No. 3, pp. 243-250.

[11] M. Dawson, S. Tavares. 1991. An Expanded Set of S-box Design Criteria Based on Information Theory and its Relation to Differential-Like Attacks**,** Advances in Cryptology - EUROCRYPT '91.

[12] P. P. Mar, K. M. Latt, 2008. New analysis methods on strict avalanche criterion of S-boxes, World Academy of Science, Engineering and Technology 48, pp. 150-154.