# Structure in the $S$-Boxes of the DES
(extended abstract)

by

E. F. Brickell
Bell Communications Research
Morristown, NJ 07960

J. H. Moore*
Sandia National Laboratories
Albuquerque, NM 87185

M. R. Purtill**
Massachusetts Institute of Technology
Cambridge, MA 02139

## ABSTRACT

The $S$-boxes used in the DES are the major cryptographic component of the system. Any structure which they possess can have far reaching implications for the security of the algorithm. Structure may exist as a result of design principles intended to strengthen security. Structure could also exist as a "trapdoor" for breaking the system. This paper examines some properties which the $S$-boxes satisfy and attempts to determine a reason for such structure to exist.

## INTRODUCTION

The DES (Data Encryption Standard) was certified by the NBS in 1975 [NBS1]. A complete description of the DES can also be found in either [D] or [K]. The major nonlinear component of the DES is a function $f$ that involves the $S$-boxes. $f$ is a function that takes as input 32 bits of partially enciphered message and 48 bits of key and produces 32 bits of partially enciphered message as output. $f$ uses eight $S$-boxes. Each $S$-box is a function from 6 bits into 4 bits. To be more precise, let $a = (a_1 \ldots a_{32})$ be 32 bits of partially enciphered message and let $k = (k_1 \ldots k_{48})$ be 48 bits of key. Then to form $f(a,k)$, $a$ is expanded to a 48 bit $b$ by duplicating the bits that have an index that is 0 or 1 mod 4 in the following manner. Let

$$b = (b_1 \ldots b_{48}) = (a_{32}a_1a_2a_3a_4a_5a_4a_5a_6a_7a_8a_9a_8a_9 \ldots a_{28}a_{29}a_{28}a_{29}a_{30}a_{31}a_{32}a_1).$$

Let $c_i = b_i + k_i$ for $1 < i < 48$. (+ will refer to addition mod 2 throughout this paper.) Let $c = (c_1 \ldots c_{48})$. $c_{6(i-1)+1} \cdots c_{6(i-1)+6}$ will be the 6 input bits into the $i$'th $S$-box. Let $d_{4(i-1)+1} \cdots d_{4(i-1)+4}$ be the output bits from the $i$'th $S$-box. Let $d = (d_1 \ldots d_{32})$. Then $f(a,k) = d$.

*This work performed at Sandia National Laboratories supported by the U.S. Department of Energy under contract number DE-AC04-76DP00789.

**This work performed while the author was visiting Bell Communications Research.

A.M. Odlyzko (Ed.): Advances in Cryptology - CRYPTO '86, LNCS 263, pp. 3-8, 1987.

The eight $S$-boxes used in the DES are listed in Table 1. Each $S$-box, $S_i$, is described by a matrix, $M_i$, with 4 rows and 16 columns. Then $S_i(c_1 \ldots c_6) = M_i(c_1c_6, c_2c_3c_4c_5)$. It is clear from examining the $S$-boxes in Table 1 that each row of an $S$-box is a permutation of the integers 0 to 15. We will list this as property P0 of the $S$-boxes.

P0. Each row of an $S$-box is a permutation of the integers 0 to 15.

There have also been other properties found [K],[L],[S] that the $S$-boxes satisfy that would not likely be satisfied if the boxes were chosen randomly. We want to study these properties to see if some of them are related. The purpose of this paper is to attempt to find a minimal set of properties satisfied by the $S$-boxes. That is, a set of properties such that if we generate random boxes designed to satisfy these properties, then these boxes will satisfy all of the unusual properties of the $S$-boxes that we have been able to find.

## S-BOX DESIGN PRINCIPLES

We would like to know what properties the $S$-boxes were designed to satisfy. This information has never been published and if 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 [L]. These were included in the report of the second workshop on the DES held by the NBS in 1976 [BGK]. In their comments, the NSA labelled the following as "design criteria" for the $S$-boxes.

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 $S(x + 001100)$ must differ in at least 2 bits.

The following were labelled by the NSA as "caused by design criteria."

P4. $S(x) \neq S(x + 11EF00)$ for any choice of $e$ and $f$.

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.

In this paper, an attempt is made to link any structure found in the $S$-boxes to these design principles. A basic tool used in this study was the generation of new boxes designed to meet a subset of these properties, but chosen randomly subject to the constraints imposed by the properties. These new boxes were compared with the DES $S$-boxes in order to identify further structure.

## RELATION BETWEEN PROPERTIES

When we generated boxes satisfying P0, P2, and P3, we found that P5 also held. P5 can be stated in a more obvious manner. For an $S$-box $S$, let $p_iS$ be the projection of $S$ onto the $i$'th output bit. i.e. if $S(x) = (d_1d_2d_3d_4)$, then $p_iS(x) = d_i$. Since each row of the $S$-box is a permutation, the list of $p_iS(x)$ over all 6 bit inputs, $x$, will contain exactly 32 1's and 32 0's. Consider the same list with one of the input bits fixed. For example, consider the list over all 6 bit inputs $x = x_1 \ldots x_6$ such that $x_i = 0$. If $i = 1$ or 6, then the list will contain exactly 16 1's

and 16 0's. If $i = 2,3,4$, or 5, the list will not necessarily contain the same number of 1's and 0's. P5 states that the $S$-boxes were chosen to minimize this difference between the number of 1's and 0's. In Table 2, we tabulated the number of 1's in each of the lists $p_j S_i(x)$ where $x$ ranges over all 6 bit inputs satisfying $x_k = 0$ for $2<k<5$, $1<i<8$, $0<j<3$. If the $S$-boxes satisfied only the row permutation property, then we would expect the distribution of the number of 1's to look like the distribution in Table 3. Table 4 contains the distribution of 100 boxes generated to satisfy only P0, P2, and P3. The similarity with Table 2 is apparent.

Given an $S$-box $S$, we can define $R_i$ to be the permutation defined by the $i$'th row of $S$. Each row of an $S$-box is a Boolean vector function from 4 input bits to 4 output bits. Each such output bit can be viewed as a function of the 4 input bits and each input bit can be viewed as a function of the 4 output bits. These 256 functions belong to the set $\boldsymbol{F}$ of all Boolean functions from 4 bits to 1 bit having the property that exactly 8 inputs are mapped to 0 and 8 are mapped to 1. Define an equivalence relation on $\boldsymbol{F}$ by $f \sim g$ if there is a function $\alpha$ which is a permutation on 4 elements and also allows for complementation of those elements, so that $f\ \alpha = g$ or $f\ \alpha = \bar{g}$. The elements of $\boldsymbol{F}$ fall into 58 equivalence classes under this relation. However, the subset $\boldsymbol{F}_s$ of $\boldsymbol{F}$ used in the rows of the $S$-boxes, fall into only 22 of these classes.

In the boxes we generated satisfying P0, P2, P3, and P4, there were two classes of functions that frequently appeared that were not in $\boldsymbol{F}_s$. One class contained the function $f(w,x,y,z) = w+x+y+z$. The other class contained the function $g(w,x,y,z) = w+x+y$. It is possible that these two classes were prohibited in the $S$-boxes because of property 1. We then generated new boxes which did not contain these two classes of functions. The distribution of classes of functions in $\boldsymbol{F}$ used in these new boxes was found to be similar to that of the DES $S$-boxes. Thus we define property P1'.

P1'. None of the four bit to one bit functions used in a row of an $S$-box is equivalent to the sum of four bits or to the sum of three bits.

We now return to the function $f$ defined in the introduction. If a vector $K$ of 48 bits is fixed, then $f|_K$ is a function from 32 bits to 32 bits. ($f|_K(a) = f(a,K)$.) However, $f|_K$ is not one-to-one or onto, so an investigation of the image of $f|_K$ was initiated. A set $X$ of one thousand 32-bit vectors was randomly generated and for several different key vectors $K$, the integers $|\{y : f|_K(y) = x\}|$ were calculated for each $x \epsilon X$. For the $S$-boxes, for all keys tested, about 1/2 of the elements of $X$ had exactly one pre-image and about 1/3 of them were not in the image set. However, these two statistics were reversed in boxes that satisfied only P0, P1', P2, and P3. Since it seems desirable to make the image set as large as possible, it appears that this shift in the distribution may be due to some design principle.

This characteristic of the image set appears to be caused by property P4. Random boxes were generated which satisfied properties P0, P1', P2, P3, and also P4. The tabulation of $|\{y : f|_K(y) = x\}|$ was then repeated for these boxes. The distribution obtained from these new boxes was not significantly different from that of the DES $S$-boxes.

The relationship between property 4 and the image set becomes clearer with a deeper look at the implications of property 4. To be precise, some notation is required. For a 32-bit vector X, let $S_i(X)$ denote the 6-bits of $X$ used in the input to the $i$-th $S$-box. Also, $S_i$ and $S_j$ will be called *consecutive* if $|i-j| = 1$ or if $\{i,j\} = \{1,8\}$. Using property 4 and the expansion operator, the following result can be proven.

*Theorem* 1

If $X$ and $Y$ are 32-bit vectors for which $f|_K(X) = f|_K(Y)$, for some key $K$, then $S_i(X) \neq S_i(Y)$ for at least 3 consecutive $S_i's$ and the Hamming distance between $X$ and $Y$ is at least 4.

Without property 4, the smallest distance between such an $X$ and $Y$ would be 2 and would involve differences in only 2 consecutive $S_i's$. Thus property 4 tends to make the possibility that two different inputs would be mapped onto the same output less likely. This both links the property to the shift in the distribution and appears to be a desirable cryptographic result.

At Crypto'85, Shamir [S] presented a paper in which he pointed out some unusual patterns in the $S$-boxes and questioned why such patterns exist. In particular, if the 6-bit input to an $S$-box is labelled as $ABCDEF$, then these patterns demonstrated an extremely high correlation between even hamming weight outputs and either $B = 0$ or $B = 1$. The boxes that we generated that satisfied P0 - P4 also tended to exhibit the same patterns, although the correlations were not as strong. This does indicate that the probabilities given in Shamir's talk do not fairly evaluate the chances of these patterns occurring, since the acknowledged design criteria seem to make them likely to exist.

## CONCLUSION

All of the structure of the $S$-boxes that we have described appears to be the result of design principles. The question that remains is whether this is a complete list of the design principles used in creating the $S$-boxes. This question could be answered in the negative if further structure was discovered in the $S$-boxes that did not occur in the boxes created using these design principles.

*References*

[BGK].
Dennis K. Branstead, Jason Gait, and Stuart Katzke, "Report of the Workshop on Cryptography in Support of Computer Security," National Bureau of Standads, September 21-22, 1976, NBSIR 77-1291, September 1977.

[D]. Dorothy E.R. Denning, "Cryptography and Data Security," Addison-Wesley, Menlo Park, California, 1983.

[K]. Alan G. Konheim, "Cryptography, A Primer," John Wiley, New York, 1981.

[L]. Lexar Corporation, "An Evaluation of the NBS Data Encryption Standard," unpublished report, Lexar Corporation, 11611 San Vicente Blvd., Los Angeles, 1976.

[NBS1].
National Bureau of Standards, "Encryption Algorithm for Computer Data Protection," Federal Register, 40, March 17, 1975, pp. 12134-12139.

[S]. Adi Shamir, "On the Security of DES," Advances in Cryptology, Proceedings of Crypto 85, pp.280-281.

| column / row | 0000 | 0001 | 0010 | 0011 | 0100 | 0101 | 0110 | 0111 | 1000 | 1001 | 1010 | 1011 | 1100 | 1101 | 1110 | 1111 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S-box 1** | | | | | | | | | | | | | | | | |
| 00 | 1110 | 0100 | 1101 | 0001 | 0010 | 1111 | 1011 | 1000 | 0011 | 1010 | 0110 | 1100 | 0101 | 1001 | 0000 | 0111 |
| 01 | 0000 | 1111 | 0111 | 0100 | 1110 | 0010 | 1101 | 0001 | 1010 | 0110 | 1100 | 1011 | 1001 | 0101 | 0011 | 1000 |
| 10 | 0100 | 0001 | 1110 | 1000 | 1101 | 0110 | 0010 | 1011 | 1111 | 1100 | 1001 | 0111 | 0011 | 1010 | 0101 | 0000 |
| 11 | 1111 | 1100 | 1000 | 0010 | 0100 | 1001 | 0001 | 0111 | 0101 | 1011 | 0011 | 1110 | 1010 | 0000 | 0110 | 1101 |
| **S-box 2** | | | | | | | | | | | | | | | | |
| 00 | 1111 | 0001 | 1000 | 1110 | 0110 | 1011 | 0011 | 0100 | 1001 | 0111 | 0010 | 1101 | 1100 | 0000 | 0101 | 1010 |
| 01 | 0011 | 1101 | 0100 | 0111 | 1111 | 0010 | 1000 | 1110 | 1100 | 0000 | 0001 | 1010 | 0110 | 1001 | 1011 | 0101 |
| 10 | 0000 | 1110 | 0111 | 1011 | 1010 | 0100 | 1101 | 0001 | 0101 | 1000 | 1100 | 0110 | 1001 | 0011 | 0010 | 1111 |
| 11 | 1101 | 1000 | 1010 | 0001 | 0011 | 1111 | 0100 | 0010 | 1011 | 0110 | 0111 | 1100 | 0000 | 0101 | 1110 | 1001 |
| **S-box 3** | | | | | | | | | | | | | | | | |
| 00 | 1010 | 0000 | 1001 | 1110 | 0110 | 0011 | 1111 | 0101 | 0001 | 1101 | 1100 | 0111 | 1011 | 0100 | 0010 | 1000 |
| 01 | 1101 | 0111 | 0000 | 1001 | 0011 | 0100 | 0110 | 1010 | 0010 | 1000 | 0101 | 1110 | 1100 | 1011 | 1111 | 0001 |
| 10 | 1101 | 0110 | 0100 | 1001 | 1000 | 1111 | 0011 | 0000 | 1011 | 0001 | 0010 | 1100 | 0101 | 1010 | 1110 | 0111 |
| 11 | 0001 | 1010 | 1101 | 0000 | 0110 | 1001 | 1000 | 0111 | 0100 | 1111 | 1110 | 0011 | 1011 | 0101 | 0010 | 1100 |
| **S-box 4** | | | | | | | | | | | | | | | | |
| 00 | 0111 | 1101 | 1110 | 0011 | 0000 | 0110 | 1001 | 1010 | 0001 | 0010 | 1000 | 0101 | 1011 | 1100 | 0100 | 1111 |
| 01 | 1101 | 1000 | 1011 | 0101 | 0110 | 1111 | 0000 | 0011 | 0100 | 0111 | 0010 | 1100 | 0001 | 1010 | 1110 | 1001 |
| 10 | 1010 | 0110 | 1001 | 0000 | 1100 | 1011 | 0111 | 1101 | 1111 | 0001 | 0011 | 1110 | 0101 | 0010 | 1000 | 0100 |
| 11 | 0011 | 1111 | 0000 | 0110 | 1010 | 0001 | 1101 | 1000 | 1001 | 0100 | 0101 | 1011 | 1100 | 0111 | 0010 | 1110 |
| **S-box 5** | | | | | | | | | | | | | | | | |
| 00 | 0010 | 1100 | 0100 | 0001 | 0111 | 1010 | 1011 | 0110 | 1000 | 0101 | 0011 | 1111 | 1101 | 0000 | 1110 | 1001 |
| 01 | 1110 | 1011 | 0010 | 1100 | 0100 | 0111 | 1101 | 0001 | 0101 | 0000 | 1111 | 1010 | 0011 | 1001 | 1000 | 0110 |
| 10 | 0100 | 0010 | 0001 | 1011 | 1010 | 1101 | 0111 | 1000 | 1111 | 1001 | 1100 | 0101 | 0110 | 0011 | 0000 | 1110 |
| 11 | 1011 | 1000 | 1100 | 0111 | 0001 | 1110 | 0010 | 1101 | 0110 | 1111 | 0000 | 1001 | 1010 | 0100 | 0101 | 0011 |
| **S-box 6** | | | | | | | | | | | | | | | | |
| 00 | 1100 | 0001 | 1010 | 1111 | 1001 | 0010 | 0110 | 1000 | 0000 | 1101 | 0011 | 0100 | 1110 | 0111 | 0101 | 1011 |
| 01 | 1010 | 1111 | 0100 | 0010 | 0111 | 1100 | 1001 | 0101 | 0110 | 0001 | 1101 | 1110 | 0000 | 1011 | 0011 | 1000 |
| 10 | 1001 | 1110 | 1111 | 0101 | 0010 | 1000 | 1100 | 0011 | 0111 | 0000 | 0100 | 1010 | 0001 | 1101 | 1011 | 0110 |
| 11 | 0100 | 0011 | 0010 | 1100 | 1001 | 0101 | 1111 | 1010 | 1011 | 1110 | 0001 | 0111 | 0110 | 0000 | 1000 | 1101 |
| **S-box 7** | | | | | | | | | | | | | | | | |
| 00 | 0100 | 1011 | 0010 | 1110 | 1111 | 0000 | 1000 | 1101 | 0011 | 1100 | 1001 | 0111 | 0101 | 1010 | 0110 | 0001 |
| 01 | 1101 | 0000 | 1011 | 0111 | 0100 | 1001 | 0001 | 1010 | 1110 | 0011 | 0101 | 1100 | 0010 | 1111 | 1000 | 0110 |
| 10 | 0001 | 0100 | 1011 | 1101 | 1100 | 0011 | 0111 | 1110 | 1010 | 1111 | 0110 | 1000 | 0000 | 0101 | 1001 | 0010 |
| 11 | 0110 | 1011 | 1101 | 1000 | 0001 | 0100 | 1010 | 0111 | 1001 | 0101 | 0000 | 1111 | 1110 | 0010 | 0011 | 1100 |
| **S-box 8** | | | | | | | | | | | | | | | | |
| 00 | 1101 | 0010 | 1000 | 0100 | 0110 | 1111 | 1011 | 0001 | 1010 | 1001 | 0011 | 1110 | 0101 | 0000 | 1100 | 0111 |
| 01 | 0001 | 1111 | 1101 | 1000 | 1010 | 0011 | 0111 | 0100 | 1100 | 0101 | 0110 | 1011 | 0000 | 1110 | 1001 | 0010 |
| 10 | 0111 | 1011 | 0100 | 0001 | 1001 | 1100 | 1110 | 0010 | 0000 | 0110 | 1010 | 1101 | 1111 | 0011 | 0101 | 1000 |
| 11 | 0010 | 0001 | 1110 | 0111 | 0100 | 1010 | 1000 | 1101 | 1111 | 1100 | 1001 | 0000 | 0011 | 0101 | 0110 | 1011 |

Table 1: DES S-boxes

| | | |
|---|---|---|
| | .1 | |
| | .3 | |
| | 1.0 | |
| | 3.0 | .3 |
| .8 | 6.8 | 1.1 |
| 6.3 | 12.1 | 5.7 |
| 25.8 | 17.1 | 20.5 |
| 35.2 | 19.2 | 40.9 |
| 25.0 | 17.1 | 22.5 |
| 6.3 | 12.1 | 7.2 |
| .8 | 6.8 | 1.6 |
| | 3.0 | .2 |
| | 1.0 | |
| | .3 | |
| | .1 | |

| Table 2a: S boxes | Table 2b: Expected value | Table 2c: boxes satisfying P0,P2,P3 |
|---|---|---|

Table 2: Property P5

distribution of number of 1's in output with 1 input bit fixed