ELSEVIER

Information Processing Letters 56 (1995) 249–252

Information Processing Letters

Resistance of balanced s-boxes to linear and differential cryptanalysis

A.M. Youssef, S.E. Tavares *

Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

Received 1 April 1995; revised 1 August 1995
Communicated by S.G. Akl

Abstract

We study the marginal density of the XOR distribution table, and the linear approximation table entries of regular substitution boxes (s-boxes). Based on this, we show that the fraction of good s-boxes (with regard to immunity against linear and differential cryptanalysis) increases dramatically with the number of input variables.

Keywords: Cryptography; Private-key cryptosystems; Combinatorics; Statistical cryptanalysis

1. Introduction

Differential cryptanalysis [1], and linear cryptanalysis [3] are currently the most powerful cryptanalytic attacks on private-key block ciphers. The complexity of differential cryptanalysis depends on the size of the largest entry in the XOR table, the total number of zeroes in the XOR table, and the number of nonzero entries in the first column in that table [1,8]. The complexity of linear cryptanalysis depends on the size of the largest entry in the linear approximation table (LAT).

One requirement in s-box design is to have a balanced s-box (also known as a regular s-box). This means that each output symbol should appear an equal number of times when the input is varied over all possible values.

Gordon and Retkin [2] calculated the probability that one or more of the output coordinates of a random, reversible s-box is an affine function. By showing that this probability deceases dramatically with the number of input variables, they conjectured that larger s-boxes are better. In this letter, we provide further evidence for their conjecture by showing that the fraction of good s-boxes, with regard to immunity against linear and differential cryptanalysis, increases dramatically with the number of input variables.

2. Definitions

For a given s-box constructed from a mapping $f(X): \mathrm{Z}_2^n \to \mathrm{Z}_2^m$, the linear approximation table entry $LAT(\alpha, \beta)$ is defined as [3]:

$$LAT(\alpha, \beta) = \#\{X \in \mathrm{Z}_2^n \mid \alpha \cdot X = \beta \cdot f(X)\} - 2^{n-1},$$

* Corresponding author. Email: tavares@ee.queensu.ca.

SSDI 0020-0190(95)00156-5

where $\alpha \in Z_2^n$, $\beta \in \{Z_2^m\}/\underline{0}$, and $\alpha \cdot X$ denotes the inner product of the vectors α and X evaluated over Z_2.

The XOR table entry $N_{\Delta x \Delta y}$ is defined as [1]:

$$N_{\Delta x \Delta y} = \#\{X \in Z_2^n \mid f(X \oplus \Delta x) \ominus f(X) = \Delta y\},$$

where $\Delta x \in Z_2^n$, $\Delta y \in Z_2^m$.

It is straightforward to notice that $LAT(\alpha, \beta) = 2^{n-1} - d(\alpha \cdot X, \beta \cdot f(X))$, where

$$d(\alpha \cdot X, \beta \cdot f(X)) = \#\{X \in Z_2^n \mid \alpha \cdot X \oplus \beta \cdot f(X) = 1\}.$$

From [4], the total number of balanced s-boxes with n input bits, and m output bits is given by

$$B(n, m) = \frac{2^n!}{(2^{n-m}!)^{2^m}}, \quad n \geq m.$$

3. Linear approximation table of balanced s-boxes

Let $wt(\alpha)$ be the hamming weight of the binary vector α; which is the number of 1's in the vector α. For $wt(\alpha) = 0$, we have $LAT(\alpha, \beta) = 0$ as $\beta \cdot f(X)$ is always a balanced function for a balanced s-box.

For $\alpha \neq \underline{0}$, $\beta \neq \underline{0}$ we have

Lemma 1.

$$P\{d(\alpha \cdot X, \beta \cdot f(X)) = 2l\} = \frac{\binom{2^{n-1}}{l}^2 (2^{n-1}!)^2}{2^n!}.$$

Proof. For $wt(\beta) = k > 0$, we have $(2^{n-1}!/(2^{n-k}!)^{2^{k-1}})^2$ ways of arranging the bits of $f(X)$ corresponding to nonzero bits of β such that $d(\alpha \cdot X, \beta \cdot f(X)) = 0$. This result follows directly by noting that the number of arrangements of N different objects, for each of which there are C copies, is $(NC)!/(C!)^N$. In our case, we have 2^{k-1} k-bit symbols with the corresponding XOR equal 0, and 2^{k-1} k-bit symbols with the corresponding XOR equal 1. Each of these symbols occurs 2^{n-k} times. Thus we have $(2^{n-1}!/(2^{n-k}!)^{2^{k-1}})$ possible arrangements for the symbols corresponding to the 0's of $\alpha \cdot X$ and a similar number of possible arrangements for the symbols corresponding to the 1's of $\alpha \cdot X$.

Since we can partition the input X into 2^k distinct sets, all X's in a given set are assigned the same common value of the k bits of $f(X)$ corresponding to the nonzero bits of β. We still need to assign the remaining $n-k$ output bits for each X. Each X within a given set must be assigned a distinct $n-k$ tuple of the remaining output bits for 2^{n-m} times, each set can be assigned in $2^{n-k}!/(2^{n-m}!)^{2^{m-k}}$ ways, so the remaining bits can be assigned in $(2^{n-k}!/(2^{n-m}!)^{2^{m-k}})^{2^k}$ ways.

Thus we have

$$\left(\frac{2^{n-1}!}{(2^{n-k}!)^{2^{k-1}}}\right)^2 \left(\frac{2^{n-k}!}{(2^{n-m}!)^{2^{m-k}}}\right)^{2^k} = \frac{(2^{n-1}!)^2}{(2^{n-m}!)^{2^m}}$$

distinct balanced functions with $d(\alpha \cdot X, \beta \cdot f(X)) = 0$. Using the same argument, and by noting that we have $(2_l^{n-1})^2$ ways to generate balanced functions that have $d(\alpha \cdot X, \beta \cdot f(X)) = 2l$ from $\alpha \cdot X$, we have

$$\binom{2^{n-1}}{l}^2 \frac{(2^{n-1}!)^2}{(2^{n-m}!)^{2^m}}$$

ways of generating distinct balanced s-boxes with $d(\alpha \cdot X, \beta \cdot f(X)) = 2l$. By dividing by the total number of balanced s-boxes, $B(n, m)$, we get Lemma 1. □

Theorem 2. *For any integer value M_{LAT}, $0 \leq M_{LAT} \leq 2^{n-2}$, let N_{LAT}^* denote the number of LATs with any entry having absolute value $\geq 2M_{LAT}$; then N_{LAT}^* is upper bounded by*

$$\frac{N_{LAT}^*}{B(n, m)} < \frac{2(2^{n-1}!)^2(2^n - 1)(2^m - 1)}{2^n!} \times \sum_{l=M_{LAT}}^{2^{n-2}} \binom{2^{n-1}}{2^{n-2} + l}^2.$$

Proof. For $l \neq 0$ we have

$$P\{LAT(\alpha, \beta) = \pm 2l\} = 2P\{d(\alpha \cdot X, \beta \cdot f(X)) = 2^{n-1} + 2l\}.$$

Using Lemma 1, and by noting that the number of LATs with entries having absolute value $\geqslant 2M_{LAT}$ is upper bounded by the number of these entries we get Theorem 2. □

Numerical substitution in the formula above shows that the fraction of balanced s-boxes with undesirable LATs decreases dramatically as the number of inputs increases. To give a numerical example, consider the case where $M_{LAT} = 2^{n-5}$, for $n = 12$, $m = 6$ we have $N^*_{LAT}/B(12, 6) < 3.86 * 10^{-10}$.

4. XOR distribution table of balanced s-boxes

Throughout the rest of this letter, the entry $N_{\underline{0}\underline{0}} = 2^n$, is not taken into consideration as it does not have any cryptographic significance.

Lemma 3. *For* $\Delta x \neq \underline{0}$, $\Delta y \neq 0$, *the number of balanced functions with* $N_{\Delta x \Delta y} \geqslant k$ *is upper bounded by* $\Psi_{n,m}(k)$, *where*

$$\Psi_{n,m}(k) = \binom{2^{n-1}}{k} 2^k \sum_{\Sigma k_i = k} G(k_1, k_2, \ldots, k_{2^{m-1}}),$$

$$G(k_1, k_2 \ldots k_{2^{m-1}}) = C(k; k_1, k_2, \ldots, k_{2^{m-1}}) \times C(2^n - 2k; l_1, l_1, l_2, l_2, \ldots, l_{2^{m-1}}, l_{2^{m-1}}),$$

$$C(k; k_1, k_2 \ldots k_{2^{m-1}}) = \frac{k!}{k_1! k_2! \cdots k_{2^{m-1}}!},$$

$$l_i = 2^{n-m} - k_i, \quad l_i \geqslant 0, \quad k_i \geqslant 0.$$

Proof. By noting that we have 2^m distinct output symbols, and each of them is repeated 2^{n-m} times, it is easy to see that for a given $\Delta y \neq \underline{0}$, $\Delta x \neq \underline{0}$ we have 2^{m-1} distinct XOR pairs, each of them is repeated 2^{n-m} times.

There is only one way to choose k_i pairs from the set i, $i = 1, 2, \ldots, 2^{m-1}$ (as the pairs within a given set are indistinguishable). These k pairs can be permuted in $C(k; k_1, k_2, \ldots, k_{2^{m-1}})$ ways. The remaining $2^n - 2k$ output symbols can be permuted into $C(2^n - 2k; l_1, l_1, l_2, l_2, \ldots, l_{2^{m-1}}, l_{2^{m-1}})$ ways, where $l_i = 2^{n-m} - k_i$. Note that we have 2 possible orders for each pair, giving 2^k total possible orders, and $\binom{2^{n-1}}{k}$ possible choices for the X positions of these k pairs.

The construction approach described above does not guarantee that these balanced functions are all distinct, and so $\Psi_{n,m}(k)$ is an upper bound. □

Lemma 4. *For* $\Delta x \neq \underline{0}$, $\Delta y = \underline{0}$, *the number of balanced functions with* $N_{\Delta x \underline{0}} \geqslant k$ *is upper bounded by* $\Phi_{n,m}(k)$, *where*

$$\Phi_{n,m}(k) = \binom{2^{n-1}}{k} \sum_{\Sigma k_i = k} D(k_1, k_2, \ldots, k_{2^m}),$$

$$D(k_1, k_2, \ldots, k_{2^{m-1}}) = C(k; k_1, k_2, \ldots, k_{2^m}) \times C(2^n - 2k; l_1, l_2, \ldots, l_{2^m}),$$

$$l_i = 2^{n-m} - 2k_i, \quad l_i \geqslant 0, \quad k_i \geqslant 0.$$

Proof. Similar to the proof of Lemma 3. □

Lemma 5. *The exact number of balanced functions with* $N_{\Delta x \Delta y} = k$, $\Delta x \neq \underline{0}$, $\Delta y \neq \underline{0}$ *(denoted by* $\Lambda_{n,m,\Delta y}(k)$*) is given by*

$$\Lambda_{n,m,\Delta y}(k) = \sum_{i=k}^{2^{n-1}} (-1)^{i-k} \binom{i}{k} \Psi_{n,m}(i).$$

Similarly, if we denote the exact number of balanced functions with $N_{\Delta x \Delta y} = k$, $\Delta x \neq \underline{0}$, $\Delta y = \underline{0}$ *by* $\Lambda_{n,m,\underline{0}}(k)$, *then we have*

$$\Lambda_{n,m,\underline{0}}(k) = \sum_{i=k}^{2^{n-1}} (-1)^{i-k} \binom{i}{k} \Phi_{n,m}(i).$$

Proof. Follows directly by using the inclusion-exclusion principle [7]. □

Using the above results, and by noting that the number of the XOR distribution tables with entries having absolute value $\geqslant 2M_{XOR}$ is upper bounded by the number of these entries we get

Theorem 6. *The fraction of balanced functions with maximum XOR table entry* $\geqslant 2M_{XOR}$, $0 \leqslant M_{XOR} \leqslant 2^{n-1}$, *is upper bounded by*

$$\frac{N^*_{XOR}}{B(n, m)} < \frac{(2^n - 1)(2^m - 1)}{B(n, m)} \sum_{k=M_{XOR}}^{2^{n-1}} \Lambda_{n,m,\Delta y}(k) + \frac{(2^n - 1)}{B(n, m)} \sum_{k=M_{XOR}}^{2^{n-1}} \Lambda_{n,m,\underline{0}}(k).$$

Numerical substitution in the formula above shows that the fraction of balanced s-boxes with undesirable XOR distribution tables decreases dramatically as the number of inputs increases.

One special case of interest is for permutations, i.e., when $n = m$. For this case (see [6] for more details on this special case) we have

$$\frac{N^*_{LAT}}{B(n,n)} < \frac{2(2^{n-1}!)^2(2^n-1)^2}{2^n!} \sum_{l=M_{LAT}}^{2^{n-2}} \left(\begin{matrix} 2^{n-1} \\ 2^{n-2}+l \end{matrix} \right)^2.$$

It is also straightforward to see that

$$\psi_{n,n}(k) = \left(\begin{matrix} 2^{n-1} \\ k \end{matrix} \right)^2 2^k (2^n - 2k)!k!$$

and

$$\phi_{n,n}(k) = \begin{cases} \dfrac{2^n!}{(2^{n-m}!)^{2^m}} & k = 0, \\ 0 & k > 0 \end{cases}$$

and hence for $0 < M_{XOR} \leqslant 2^{n-1}$, we have

$$\frac{N^*_{XOR}}{B(n,n)} < \frac{(2^n-1)^2}{2^n!} \sum_{k=M_{XOR}}^{2^{n-1}} \Lambda_{n,n,\Delta y}(k).$$

To give a numerical example, consider the case where $M_{XOR} = n$, for $n = 12$ we have $N^*_{XOR}/B(n, n) < 8.9 * 10^{-6}$.

5. Conclusion

We have derived an upper bound on the fraction of balanced functions (s-boxes) having a specified lower bound on the maximum entry in the XOR distribution table or the LAT. For reasonably small values of these maximum entries, this fraction decreases dramatically with the number of input variables. Simulation results show that this bound, in its non trivial range, approaches the actual distribution.

It is worth noting that our statistical results might have some practical limitations; as describing a large randomly chosen s-box requires a large amount of memory which might be impractical in some applications. This suggests that other methods such as algebraic constructions, which trade-off memory requirements and computational speed (see [5] for an example of such methods), offer alternative approaches.

References

[1] E. Biham and A. Shamir, Differential cryptanalysis of DES-like cryptosystems, in: *Advances in Cryptology: Proc. Crypto '90* (Springer, Berlin, 1991) 1–21.

[2] J. Gordon and H. Retkin, Are big S-boxes best?, in: *Proc. Workshop on Cryptography*, Lecture Notes in Computer Science **149** (Springer, Berlin, 1982) 257–262.

[3] M. Matsui, Linear cryptanalysis method for DES cipher, in: *Advances in Cryptology: Proc. Eurocrypt '93* (Springer, Berlin, 1994) 366–397.

[4] C. Mitchell, Enumerating boolean functions of cryptographic significance, *J. Cryptology* (1990) 155–170.

[5] K. Nyberg, On the construction of highly nonlinear permutations, in: *Advances in Cryptology: Proc. Eurocrypt '92* (Springer, Berlin, 1992) 92–98.

[6] L.J. O'Connor, On the distribution of charachteristics in bijective mappings, in: *Advances in Cryptology: Proc. of Eurocrypt '93* (Springer, Berlin, 1994) 360–370.

[7] F.S. Roberts, *Applied Combinatorics* (Prentice-Hall, Englewood Cliffs, NJ, 1984).

[8] J. Seberry, X. Zhang and Y. Zheng, Systematic generation of cryptographically robust s-boxes, in: *Proc. 1st ACM Conf. on Computer and Communications Security*, Fairfax, VA (1993) 172–182.