SECRET-KEY CRYPTOSYSTEMS II

Linearity in block ciphers

E. P. Dawson L. J. O'Connor
CRC for Distributed Systems Technology *
and
Information Security Research Centre

H. M. Gustafson
School of Mathematics

Queensland University of Technology
GPO Box 2434, Brisbane QLD 4001

Abstract

In this paper we will survey the principle of linearity in block ciphers. We consider linear relationships between the plaintext and ciphertext bits, using elementary arguments from linear algebra, and then using linear relationships under real number addition based on canonical correlation analysis. Linear structures [3] are also examined, which are a form of linearity that leads to degeneracy in the key, meaning that certain bits do not affect the ciphertext. We show that most functions are not expected to have a linear structure, though even partial linearity in this respect leads to a powerful attack known as differential cryptanalysis. Lastly, we consider linear approximation as a cryptanalytic tool, and present the recent linear cryptanalysis due to Matsui on the Data Encryption Standard (DES).

*The work reported in this paper has been funded in part by the Cooperative Research Centres program through the Department of the Prime Minister and Cabinet of Australia.

1 Introduction

A block cipher E is a family of encryption functions that acts on n characters of data (usually bits), with typical values of n being in the range of 64 to 2048. The two major properties to be considered in the design of a block cipher are (a) to minimize the statistical relationship between the plaintext and ciphertext, and (b) to strongly suggest that the key cannot be recovered in time that is significantly less than the expected cost of exhaustive key search. First, we observe that (a) does not imply (b), in that a strong pseudorandom function is not necessarily resistant to cryptographic attacks. For example, it is known that sequences produced by linear feedback registers can be selected to satisfy the randomness postulates of Golumb [7], but the initial register contents and tapping information can be recovered by inspecting a small amount of ciphertext [5]. Second, (b) is not a proof of security since this would essentially be a solution to a major open problem in computational complexity theory [6]. Often (b) will be an accreditation given to the cipher after a thorough, and most likely protracted, examination of its properties by cryptanalysts; even so, it is only a conjecture that the cipher is in fact secure. On this point, the history of DES is informative. When released in the mid seventies, IBM stated that 17 years of research had been consumed in the design and analysis of the algorithm. To this day, all reported 'weaknesses' of DES are either unlikely to occur (for example, selecting a so-called weak key [13]), or require such substantial computational resources to take advantage of (for example, differential cryptanalysis [2]) At present, and probably always, DES is considered to be a very strong cipher with an 'unfortunately small' key (56 bits).

In this paper we will survey three forms of statistical dependency found in block ciphers each based on some notion of linearity. These attacks will apply particularly to product ciphers [5] which are block ciphers built from smaller components such as look-up tables (S-boxes S) and permutations (P-boxes P). We begin in §2 by considering linear relationships between the plaintext and ciphertext bits, using elementary arguments from linear algebra. We also investigate the application of canonical correlation analysis to cryptanalysis, which examines linear relationships under real number addition. In §3 we consider linear structures [3], a form of linearity that leads to degeneracy in the key (here degeneracy means that when the influence of the key is modelled as a boolean function f, certain keys bits do not affect the function). We show that most functions are not expected to have linear structures, though even partial linearity in this respect leads to a powerful attack known as differential cryptanalysis. Lastly, we consider a linear approximation as a cryptanalytic tool, and present the recent attack of Matsui [12] on the Data Encryption Standard (DES) [15].

2 Plaintext/Ciphertext Linearity

Dependencies that exist between subsets of plaintext, ciphertext and key bits could decrease the cost of searching the keyspace. In one of the the worst cases the cipher is a linear mapping, allowing the cipher to be totally determined after inspecting a relatively small amount of ciphertext. Several such dependencies, including the linearity just mentioned, can be detected through statistical methods [8, 10], such as the χ^2-test. In the next few sections we examine several form of linearity in block ciphers.

2.1 Gaussian Elimination

In this test we enquire if some *subset* of the ciphertext bits can be written as a linear combination of plaintext bits. For $X = x_1 x_2 \cdots x_n \in Z_2^n$ let $X[i_1, i_2, \ldots, i_a]$ denote the XOR sum $x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_a}$ where $1 \leq i_1 < i_2 < \cdots < i_a \leq n$ and $1 \leq a \leq n$. Consider an equation of the form

$$P[i_1, i_2, \ldots, i_a] \oplus C[j_1, j_2, \ldots, j_b] \oplus a_0 \;=\; 0 \tag{1}$$

where $a_0 \in Z_2$, which indicates that the sum of a subset of plaintext bits with a subset of ciphertext bits is constant (corrected to 0 by the a_0 term). Dependencies of the form in (1) can be tested as follows. Select $(2n+1)$ plaintext/ciphertext pairs (P_i, C_i), $1 \leq i \leq 2n+1$, where $P_i = p_{i1}, p_{i2}, \ldots, p_{in}$ and $C_i = c_{i1}, c_{i2}, \ldots, c_{in}$. Then construct a $(2n+1) \times (2n+1)$ matrix A matrix where the first row is all ones, and column i contains the bits of plaintext P_i followed by the bits of ciphertext C_i. If when performing row reductions on the matrix A a row of all zeros is encountered, then a dependency of the form in (1) must exist.

If E has a dependency of the form in (1), then the test will report it (true dependency); on the other hand, even if E has no dependency of the form in (1), the test may report a dependency for the given sample of plaintext/ciphertext pairs (P_i, C_i) (false dependency). We could sample N matrices $A_1, A_2, \ldots, A_N$ and true dependencies would be found in each A_i if they existed. However, we would like to know how large N should be before any false dependencies induced by the plaintext/ciphertext sample would be unlikely to occur in all N sample matrices. To answer this question, observe that a matrix of full rank has no dependencies. We will make the assumption that a cipher E which has no true dependencies when sampled produces matrices A_i that are random over Z_2 (except for the first row). The probability that a random $k \times k$ matrix B has full rank is

$$q \;=\; \prod_{i=0}^{k-1} \left(1 - 2^{i-k}\right). \tag{2}$$

Let A' be defined as the matrix obtained from A by adding the first row of A to all other rows that have a 1 in the first column, and then deleting the first row and column from

the resulting matrix. Clearly, if A has full rank then the $(2n \times 2n)$ matrix A' will also have full rank. From eq. (2), the probability that A' has full rank is $q = 0.2887$ when $n = 64$. Then the probability that at least one matrix in a random sample of N such matrices will have full rank is $1 - (1 - q)^N$. Thus by solving $1 - (1 - q)^N = p$ we are confident that in a sample of N matrices, the probability of producing at least one matrix of full rank is p. For example, when $n = 64$, a sample of 21 matrices has a probability of 99.9% to yield a matrix with full rank.

It should be noted that each independent equation in the form of (1) reduces by one bit the entropy between the intercepted ciphertext and the unknown plaintext. If there are n independent such equations then the cipher is affine.

2.2 Linear Relationship Under Real Number Addition

Another method for examining linear relationships in block ciphers is to apply *Canonical Correlation analysis* as was first suggested by Carlisle Adams in his PhD thesis [1, p. 91]. This method investigates linear relationships under real number addition between two variables X and Y that are expressed as linear combinations of experimental observations (given below). Canonical analysis may be employed to determine the *best linear relationship* that exists between the X and Y variables. This technique was originally developed by Hotelling [9, p.321], and in our case will involve calculating coefficients $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ plus an associated *canonical correlation*, λ_j, which measures the extent of the linear correlation between X and Y. More general information on canonical correlation analysis can be found in the book of Cooley and Lohnes [4, p.168].

An analysis on a sample of N plaintext-ciphertext pairs, (P_i, C_i), $1 \leq i \leq N$, where $P_i = p_{i1}, p_{i2}, \ldots, p_{in}$ and $C_i = c_{i1}, c_{i2}, \ldots, c_{in}$ is performed as follows: using several covariance matrices (defined below), an equation with n solutions is established, where each solution yields a measure of canonical correlation, λ_j and resulting weight vectors $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ corresponding to the plaintext and ciphertext bit vectors. Each canonical correlation measures the strength of a line of the form $Y = aX + b$ to fit the set of points (X_i, Y_i), where X and Y are the corresponding canonical variables. The points (X_i, Y_i) are expressed as linear functions of the plaintext and ciphertext bits:

$$Y_i = \alpha_{j1} p_{i1} + \alpha_{j2} p_{i2} + \cdots + \alpha_{jn} p_{in} \quad (3)$$

$$X_i = \beta_{j1} c_{i1} + \beta_{j2} c_{i2} + \cdots + \beta_{jn} c_{in} \quad (4)$$

The coefficient vectors $\boldsymbol{\alpha}_j = \alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jn}$ and $\boldsymbol{\beta}_j = \beta_{j1}, \beta_{j2}, \ldots, \beta_{jn}$ are calculated so that the corresponding correlation between the variables X and Y is maximized. The $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ vectors contain the weights of each bit position in the resulting canonical variables.

The analysis requires the calculation of three $n \times n$ covariance matrices: R_{CC} measuring correlation between ciphertext bits, R_{PP} measuring correlation between plaintext bits, and R_{CP} measuring correlation between ciphertext and plaintext bits. If $R_{CC}[i,j]$ is the entry for the ith row and jth column, $1 \leq i,j \leq n$, then

$$\begin{aligned} R_{CC}[i,j] &= \frac{1}{N-1} \cdot \sum_{k=1}^{N} (c_{ki} - \bar{C}_i)(c_{kj} - \bar{C}_j) \\ \bar{C}_i &= \frac{1}{N} \cdot \sum_{k=1}^{N} c_{ki} \end{aligned}$$

where $\bar{C}_i$ is the sample mean for ciphertext bit c_i. The matrices R_{PP} and R_{CP} are similarly defined:

$$\begin{aligned} R_{PP}[i,j] &= \frac{1}{N-1} \cdot \sum_{k=1}^{N} (p_{ki} - \bar{P}_i)(p_{kj} - \bar{P}_j) \\ R_{CP}[i,j] &= \frac{1}{N-1} \cdot \sum_{k=1}^{N} (c_{ki} - \bar{C}_i)(p_{kj} - \bar{P}_j) \\ \bar{P}_i &= \frac{1}{N} \cdot \sum_{k=1}^{N} p_{ki}. \end{aligned}$$

Also, let $R_{PC} = R_{CP}^T$ be the transpose of R_{CP}. The analysis involves the calculation of n eigenvalues and eigenvectors of the equation

$$(R_{PP}^{-1} \cdot R_{PC} \cdot R_{CC}^{-1} \cdot R_{CP} - \lambda_j I)\boldsymbol{\alpha}_j = 0 \tag{5}$$

where λ_j is the eigenvalue corresponding to the vector of weights $\boldsymbol{\alpha}_j$ subject to the condition that $\boldsymbol{\alpha}_j{}^T \cdot R_{PP} \cdot \boldsymbol{\alpha}_j = 1$. The corresponding vector of coefficients $\boldsymbol{\beta}_j$ is determined from the equation $\boldsymbol{\beta}_j = R_{CC}^{-1} \cdot R_{CP} \cdot \boldsymbol{\alpha}_j \cdot \sqrt{\lambda_j}$. Each value of λ_j determines a canonical correlation which measures the strength of a linear relationship $Y = aX + b$ corresponding to the set of points (X_i, Y_i), determined by substituting $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ in eq. (3) for the sample of N plaintext-ciphertext pairs chosen.

Observe that $0 \leq \lambda_j \leq 1$ with $\lambda_j = 1$ indicating 100% correlation, for which all calculated points (X_i, Y_i) lie on a straight line. For each value of λ_j the resultant line $Y = aX + b$ is obtained using statistical regression analysis and determines the *line-of-best-fit* relating to the sample of plaintext-ciphertext pairs chosen.

Our analysis of this method shows that its application to block ciphers can be used to determine the existence of equal Hamming weights between subsets of plaintext and ciphertext positions in a cipher, with 100% correlation. This occurs, for example, in a transposition of plaintext to ciphertext bit positions. Canonical Correlation analysis

aims to find a relationship between plaintext and ciphertext so that some part of the plaintext may be determined from an intercepted ciphertext. Unless the cipher exhibits the properties to yield 100% correlation, different samples of plaintext-ciphertext pairs will yield different canonical correlations λ_j, different coefficient vectors $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ and different linear equations $Y = aX + b$, for the same key.

Each equation determined from this analysis, as in the method of Gaussian Elimination, will yield one bit of information between plaintext and ciphertext bits. A number of equations would be desired to give sufficient information to effectively determine sufficient plaintext bits from any intercepted ciphertext. A similar analysis could be carried out by combining the plaintext and ciphertext vectors to represent the X variable and the key vector as the Y variable. The number of solutions is limited by the length of the smaller variable, Y. Linear relationships relating plaintext and ciphertext bits to key bits would be more useful in determining information about the key. This method could be applied to the S-boxes of newly developed ciphers emulating DES or the internal functions of symmetric block ciphers, to determine the existence of linear equations under real number addition. As expected, the S-boxes of DES yielded such equations with very low canonical correlation measures.

3 Linear structures

A divide-and-conquer attack on the keyspace of a cipher is a method for partitioning the key bits into $d > 1$ distinct sets $w_1, w_2, \ldots, w_d$ such that each set w_i can be searched independently. If such a partition can be found then the cost of testing all possible keys becomes $O(2^{|w^*|})$ steps where $w^* = \max_{1 \le i \le d} w_i$, rather than $O(2^{|w_1|+|w_2|+\cdots+|w_d|})$ steps by obvious methods. Such a partition will exist if, for example, a known subset of the ciphertext depends on only k out of m key bits, will permit the key to be recovered in approximately $2^k + 2^{m-k}$ steps. We see that if $k \approx m/2$ then the key can be recovered in time which is approximately the square root of the time to perform exhaustive search. We will examine a class of boolean functions, known as functions with linear structures, that admit divide-and-conquer attacks of this type. These functions have been used by Chaum and Evertse [3] to perform an attack on DES that is faster than exhaustive search when DES is reduced to less than 8 rounds. In what follows, we will represent an n-bit boolean function f as a polynomial $f(X) \in Z_2[x_1, x_2, \cdots, x_n]$, called the Algebraic Normal Form (ANF) of f.

Recall that p-linear functions were defined as

$$f(X) = g(x'_1, x'_2, \ldots, x'_{n-k}) + \sum_{1 \le j \le k} m_j x_{i_j}. \tag{6}$$

Equivalently, if $\mathbf{e_i}$ is the ith unit vector, a function f is p-linear in k variables if there exists a set $B = \{\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}\} \subseteq \{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_k}\}$ such that for all $\mathbf{b_i} \in B$, $f(X) \oplus f(X + \mathbf{b_i})$ is invariant for all $X \in Z_2^n$. Here $e_i \in Z_2^n$ is the ith unit vector. Linear structures are an extension of p-linearity in that B is an arbitrary subset of Z_2^n. The relation between p-linearity and linear structures is given in the next lemma.

Lemma 3.1 (Lai [11]) Let $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$ be a set of linearly independent linear structures for the n-bit function f, where $1 \leq k \leq n$. Then there exists an $n \times n$ matrix M with coefficients over Z_2 such that if $g(X) = g(x_1, x_2, \ldots, x_n) = f((x_1, x_2, \ldots, x_n)M)$ then the ANF of $g(x_1, x_2, \ldots, x_n)$ is given as

$$g(X) = x_1 m_1 + x_2 m_2 + \cdots + x_k m_k + g(x_{k+1}, x_{k+2}, \ldots, x_n) \tag{7}$$

where $m_i = f(\mathbf{b_i}) \oplus f(\mathbf{0}) \in Z_2$ for $1 \leq i \leq k$.

□

Corollary 3.1 Let $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$ be a set of linearly independent vectors. There are $2^{2^{n-k}+k}$ n-bit functions for which $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$ are linear structures.

Proof. By Lemma 3.1 let $\mathbf{b_i} = \mathbf{e_i}$, $1 \leq i \leq k$, without loss of generality. However it follows from eq. (7) that there are 2^k ways to choose the m_i, and $2^{2^{n-k}}$ ways to choose the $(n-k)$-bit function g. □

Thus if f is a function that has linear structures $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$, an appropriate basis change for Z_2^n transforms f into a p-linear function. The cryptanalyst can take advantage of the linear structures in f if some of the m_i in eq. (7) are zero, which will eliminate the influence of some variables (possibly key bits) on the ciphertext.

Example 3.1 The 4-bit function f has $\mathbf{b} = 1110$ as its only linear structure where

$$f(X) = x_2 + x_1x_2 + x_1x_3 + x_2x_3 + x_3x_4 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4.$$

Define M as the matrix

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{8}$$

If $g(x_1, x_2, \ldots, x_n) = f((x_1, x_2, \ldots, x_n)M)$ then

$$g(X) = x_3 + x_2x_4 + x_3x_4 + x_2x_3x_4.$$

As the first column of M is b, then $\mathbf{e_1}$ is a linear structure in g, and g is degenerate in x_1 as $f(\mathbf{b}) = f(\mathbf{0}) = 0$. □

Let $\mathcal{LS}^n$ be the set of n-bit boolean functions that have a linear structure $b \neq 0$. O'Connor [14] has shown that most functions do not have linear structure, and in particular, that

$$\lim_{n\to\infty} |\mathcal{LS}^n|/((2^n - 1) \cdot 2^{2^{n-1}+1}) = 1. \tag{9}$$

4 Linear Cryptanalysis

We will now give a short exposition on a new method for cryptanalyzing DES based on linear approximation due to Matsui [12]. The basis of the attack is finding approximate linear relationships between certain bits in the plaintext, ciphertext and key. Recall that, for $X = x_1 x_2 \cdots x_n \in Z_2^n$, $X[i_1, i_2, \ldots, i_a]$ denotes the XOR sum $x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_a}$ where $1 \leq i_1 < i_2 < \cdots < i_a \leq n$ and $1 \leq a \leq n$. Let P be a plaintext, C its ciphertext, and K the key used to encrypt P. Consider an equation of the form

$$P[i_1, i_2, \ldots, i_a] \oplus C[j_1, j_2, \ldots, j_b] \;=\; K[k_1, k_2, \ldots, k_c] \tag{10}$$

where the LHS is equal to the RHS with some probability p. This means that if we fix the key K and consider all possible plaintexts, then XORing certain subsets of the plaintext and ciphertext bits equals a certain XORed subset of the key bits with probability p. Intuition would suggest that if the bit subsets are selected randomly then the probability of eq. (10) being true should be close to $1/2$.

Let the cryptanalyst have a sample of N plaintext/ciphertext pairs P_i, C_i, all encrypted under the same key K. Let there be $0 \leq T \leq N$ pairs for which the LHS of eq. (10) is equal to 0. Then consider the following procedure to determine $K[k_1, k_2, \ldots, k_c]$, which Matsui has called the 'maximum likelihood method':

If $T > N/2$ then
    guess $K[k_1, k_2, \ldots, k_c] = 0$ when $p > 1/2$ or
    guess $K[k_1, k_2, \ldots, k_c] = 1$ when $p < 1/2$
else
    guess $K[k_1, k_2, \ldots, k_c] = 1$ when $p > 1/2$ or
    guess $K[k_1, k_2, \ldots, k_c] = 0$ when $p < 1/2$.

The reasoning behind the method is quite straightforward: if a majority of the plaintext/ciphertext pairs in the sample of size N give the LHS of eq. (10) to be zero and the probability of the LHS equaling the RHS is less than one half, then guess the RHS to be one (the complement of the RHS). Similar reasoning prevails in the case where a

minority of the sample gives the LHS of eq. (10) to be zero. It should be clear that this method is more likely to succeed as the value of p moves away from $1/2$.

When the maximum likelihood method makes a correct prediction we obtain 1 bit of information about the key (namely the value of the XOR of c bits of the key). So to gain a significant amount of information about the key we would then require several relations of the form in eq. (10). By approximating S-box S5 we find that

$$R_i[15] \oplus K_i[22] \quad = \quad F(R_i, K_i)[7, 18, 24, 29] \tag{11}$$

with probability $12/64 = 0.19$. For 3-round DES, using this approximation in the first and third rounds we have that

$$P_L[7, 18, 24, 29] \oplus P_R[15] \oplus C_L[7, 18, 24, 29] \oplus C_R[15] \quad = \quad K_1[22] \oplus K_3[22]. \tag{12}$$

is true probability $p = (12/64)^2 + (1 - 12/64)^2 = 0.70$, and is of the form desired for eq. (10). Similarly for 5-round DES, Matsui has found that the approximation

$$\begin{aligned} P_L[15] \oplus P_R[7, 18, 24, 27, 28, 29, 30, 31] \oplus C_L[15] \oplus C_R[7, 18, 24, 27, 28, 29, 30, 31] \\ = K_1[42, 43, 45, 46] \oplus K_2[22] \oplus K_4[22] \oplus K_5[42, 43, 45, 46]. \end{aligned} \tag{13}$$

holds with probability 0.519.

We still need to determine when the maximum likelihood method is expected to correctly predict the sum of the key bits. Recall that the maximum likelihood method operates on a sample of N plaintext/ciphertext pairs. The success (correct prediction) of the method will increase the larger N becomes since our sample probabilities are more likely to be close to the true probabilities. It is natural to ask how large N needs to be before we expect the maximum likelihood method to make the correct prediction say 90% of the time. Using the normal distribution Matsui gives such estimates and they are listed in Table 1. Using these results, if given $(|0.519 - \frac{1}{2}|)^{-2} = 2800$ plaintext/ciphertext pairs, the maximum likelihood method can predict the key bits in eq. (13) with 97.7% success. The attack can be modified to yield more than one bit of information about the key, and the reader is referred to Matsui's paper for details.

N	$\frac{1}{4} \cdot (\lvert p - \frac{1}{2} \rvert)^{-2}$	$\frac{1}{2} \cdot (\lvert p - \frac{1}{2} \rvert)^{-2}$	$(\lvert p - \frac{1}{2} \rvert)^{-2}$	$2 \cdot (\lvert p - \frac{1}{2} \rvert)^{-2}$
Success rate	84.1%	92.1%	97.7%	99.8%

Table 1: The success rate of the maximum likelihood method.

5 Conclusion

In this paer we have examined several forms of linearity as applied to cryptanalysing block ciphers. We began with considering how to detect if the ciphertext was a linear transformation of the plaintext, or if a similar relationship holds between proper subsets of the plaintext and ciphertext. The notion of linear dependency is extended by using canonical correlation in §2, but this approach appears only to be useful in detecting permutation mappings. Any attack attempting to exploit correlation due to linearity merely by observing a large number of plaintext-ciphertext pairs is unlikely to succeed. That is, designing statistical tests to detect linearity without taking into account the internal structure of the cipher are unlikely to detect any correlation.

On the other hand, the internal mappings used in a product cipher, the S-boxes, are much smaller in size than the block size. It is quite possible to select S-boxes that exhibit linear dependencies and not contradict the enumeration results above since they are asymptotic. Differential and linear cryptanalysis have shown that poorly chosen S-boxes can lead to attacks on product ciphers even when the ciphertext itself may be highly nonlinear. The cryptanalyst need only attack the cipher round by round, establishing and extending dependencies from one round to the next, hopefully inducing some correlation in the ciphertext. This suggests that the designer not only construct a highly nonlinear cipher, but must select highly nonlinear S-boxes to achieve this. For example, each S-box should have a linear correlation as close to one half as possible.

References

[1] C. M. Adams. *A formal and practical design procedure for Substitution-Permutation network cryptosystem.* PhD thesis, Department of Electrical Engineering, Queen's Univeristy at Kingston, 1990.

[2] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[3] D. Chaum and J.-H. Evertse. Cryptanalysis of DES with a reduced number of rounds. *Advances in Cryptology, CRYPTO 85, H. C. Williams ed., Lecture Notes in Computer Science, vol. 218, Springer-Verlag*, pages 192–211, 1986.

[4] W. W. Cooley and P. R. Lohnes. *Multivariate Data Analysis.* Wiley, New York, 1971.

[5] D. E. Denning. *Cryptography and Data Security.* Addison–Wesley Publishing Company, 1982.

[6] M. R. Garey and D. S. Johnson. *Computers and Intractability, A Guide to the Theory of NP-completeness.* W. H. Freeman and Co., San Francisco, 1979.

[7] S. Golumb. *Shift Register Sequences.* Aegean Park Press, 1982.

[8] H. Gustafson, E. Dawson, L. Nielsen, and W. Caelli. Measuring the strength of ciphers. In G. Gable and W. Caelli, editors, *IFIP Transactions, IT Security: The Need for International Cooperation, Elsevier Science Publishers B.V., North-Holland*, pages 235–247, 1992.

[9] H. Hotelling. Canonical analysis. *Biometrika*, 28:321–377, 1936.

[10] A. Konheim. *Cryptography: a primer.* Wiley, 1981.

[11] X. Lai. Linear structures of functions over prime fields. unpublished manuscript, 1990.

[12] M. Matsui. Linear cryptanalysis method for DES cipher. *abstracts of papers, EUROCRYPT 93, Norway, May.*

[13] C. H. Meyer and S. M. Matyas. *Cryptography: A new dimension in computer security.* Wiley, 1982.

[14] L. J. O'Connor. *An analysis of product ciphers using boolean functions.* PhD thesis, Department of Computer Science, University of Waterloo, 1992.

[15] National Bureau of Standards. Data Encryption Standard. FIPS PUB 46, Washington, D. C. (January 1977).

DES can be Immune to Linear Cryptanalysis

Kwang-jo Kim Sang-jin Lee Sang-jun Park Dai-ki Lee

ETRI, KOREA

Abstract In this paper, we propose the necessary conditions how to strengthen DES S-boxes against linear cryptanalysis. Combined with our design criteria of DES S-boxes against differential cryptanalysis presented in JW-ISC'93 [5], we show that the total security of DES against both linear cryptanalysis and differential cryptanalysis can be improved.

1 Introduction

Two ways of cryptanalysis have been published in the open literature, which can break DES [1] more efficiently than key-exhaustive search. They are the differential cryptanalysis[2],[3] by Biham and Shamir and linear cryptanalysis[8] by Matsui in 1990 and 1993, respectively. Differential cryptanalysis is more efficient than key-exhaustive search when a set of input XORed values are probably correlated with the output XORed values. Thus, differential cryptanalysis belongs to a chosen plaintext attack in a sense that the attacker should choose a particular set of input values.

On the other hand, linear cryptanalysis tries to find the partial key information so that a linearly approximated expression holds. Linear cryptanalysis can be said to be a known plaintext attack. The complexity to break DES by linear cryptanalysis is about 2^{47}. A ciphertext-only attack is also possible when the plaintext consists of 7-bit ASCII codes only, where the complexity to find 7 key bits of the 16-round DES is about 1.82×2^{53}. Descriptions on the linear cryptanalysis in detail can be found in [8]. In SCIS'94 [9], Matsui improved the linear cryptanalysis method of the 16-round DES and showed that the reduced complexity is about 2^{43}. The common point of two attacking methods is to use the cryptanalytic properties of DES S-boxes.

In JW-ISC'93 [5], we have suggested an additional design criterion of DES-like S-boxes so that DES can be resistant to differential cryptanalysis, *i.e.*, $S(x) \neq S(x \oplus 11ef10)$ for any DES S-box S. Our criteria are shown [6], [7] to lead to a simple and robust strengthening method of DES against differential cryptanalysis. Biham [4] suggested that the security of our DES, denoted as "s^3DES", can be improved from linear cryptanalysis if the order of S1-box and S2-box are reversed. The original order has almost the same strength as of DES against linear cryptanalysis. When reversed, it becomes 2^{15} times secure. Some order, for example, 32145678, makes s^3DES weaker. This requires that the order of S-boxes in DES F-function should be considered carefully to improve for the immunity of DES to linear cryptanalysis.

Thus, to withstand DES against linear cryptanalysis, we focus ourselves on how to redesign DES S-boxes including their ordering requirements.

Like differential cryptanalysis, an attack by linear cryptanalysis is to be successful if any n-round linear iterative approximation holds with high probability. If this probability is low enough compared to the current DES, the linear cryptanalysis is no more efficient than the key-exhaustive search.

In this paper, we discuss the uniformity of a linear distribution table in a DES-like S-box and the necessary conditions to strengthen DES S-boxes against linear cryptanalysis. We also suggest that how to locate S-boxes in a specific position of DES F-function.

2 Preliminaries

The following notations are adopted throughout this paper and the rightmost bit is referred to as the zero-*th* bit.

- I_i : The input value of *i-th* round in DES F-function.
- O_i : The output value of *i-th* round in DES F-function.
- K_i : The key value of *i-th* round in DES F-function.
- $X[Z] = \dotplus_{k \in Z} X[k]$, where $Z \subset \{0, 1, \ldots, 47\}$ and $X[k]$ is the k-*th* bit of X which is one of I_i, O_i and K_i.
- a_x : The hexadecimal value of a.
- $W(\alpha)$: The Hamming weight of α.
- For $x, y \in GF(2)^n$, $x \bullet y$ denotes the dot product of x and y.

Definition 1 (Linear distribution table) *For a given DES S-box S, we define $NS(\alpha, \beta)$ as the number of times minus 32 out of 64 input patterns of S, such that an XORed value of the input bits masked by α coincides with an XORed value of the output bits masked by β, that is to say,*

$$NS(\alpha, \beta) \;=\; \#\{x \in GF(2)^6 | x \bullet \alpha = S(x) \bullet \beta\} - 32$$

where $\alpha \in GF(2)^6$ and $\beta \in GF(2)^4$. We refer the complete table for every α and β to be the linear distribution table as shown in Figure 1. For a specific S-box, S_i $(i = 1, \ldots, 8)$, we denote its linear distribution table as $NS_i(\alpha, \beta)$.

Definition 2 (Linear approximation) *For a given expression $I[Z_1] \dotplus O[Z_2] = K[Z_3]$ with probability $p + 1/2$, this linear approximation is denoted as*

$$A\colon I[Z_1], K[Z_3] \;\longrightarrow\; O[Z_2] \quad \textit{with } p.$$

We denote this expression as $A, B, C, \cdots$. Also $\delta(A)$ denotes the set of S-boxes necessary to express A and $\#\delta(A) = |\delta(A)|$.

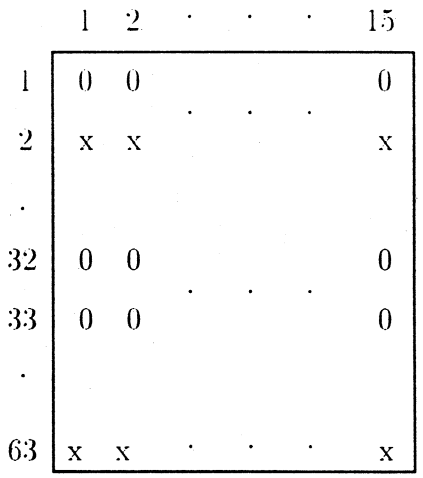

Figure 1: Linear distribution table of DES S-boxes

Example 1 In DES, the following linear approximation from $NS_5(10_x, f_x) = -20/64$ holds

$$A \ : \ I[15], K[22] \longrightarrow O[7, 18, 24, 29] \ \text{with} \ -20/64.$$

Then, $\delta(A) = \{S5\text{-}box\}$ and $\#\delta(A) = 1$. □

To attack n-round DES by linear cryptanalysis in general, we need to find the useful linear approximation of $(n-1)$-round DES. When the linear approximation of $(n-1)$-round DES holds with probability $q = p + \frac{1}{2}$, the number of plaintexts which the attacker needs are about $|p|^{-2}$ by Lemma 2 in [8].

Thus, a linear approximation of 15-round DES is necessary to break the full 16-round DES. When this approximation holds with probability p_{15}, the necessary condition that linear cryptanalysis is no more efficient than key-exhaustive search is $p_{15}^{-2} \geq 2^{56}$, *i.e.*, $|p_{15}| \leq 2^{-28}$.

In order that DES can be resistant to linear cryptanalysis, it is necessary to find S-boxes which make the probability of any linear approximation small. We may rearrange other components like P-permutation or E-expansion in DES F-function so that DES can be resistant to linear cryptanalysis. This, however, cannot be a solution against differential cryptanalysis because they are linear (*i.e.*, they do not affect the complexity of differential cryptanalysis).

We have checked a set of linear distribution tables of random DES-like S-boxes through computer experiments. It was possible to obtain occasionally such an S-box that the maximal absolute value in its linear distribution table is smaller than 16. Note that the maximal entry of absolute values in the linear distribution table of S5-box and S6-box in DES are 20 and 14, respectively. This characteristic of DES S5-box is a fundamental tool to break DES by linear cryptanalysis.

Anyway it is reasonable to fix the allowable maximal absolute value of linear distribution table to be 16. As a rule of thumb, we set up the first condition as below :

Condition 1 *The allowable maximal absolute value in a linear distribution table of any S-box must be less than 16.*

3 Uniformity of a Linear Distribution Table

The next question is how the values in the linear distribution table of a DES-like S-box to be distributed. In this section, we discuss the uniformity of the linear distribution table in a DES-like S-box.

Definition 3 (Walsh transform) *For a given Boolean function $f(x): GF(2)^n \rightarrow GF(2)$, the Walsh transform of f, denoted by F, is given by;*

$$F(w) = \sum_{x \in GF(2)^n} (-1)^{f(x)} \cdot (-1)^{x \bullet w}$$

where $w \in GF(2)^n$.

Definition 4 (Nonlinearity) *Let $f(x): GF(2)^n \rightarrow GF(2)$ denote any Boolean function. Define*

$$\begin{aligned} d(f, L_w) &= \#\{x \in GF(2)^n : f(x) \neq L_w(x)\} \quad \text{and} \\ \epsilon(f, L_w) &= \#\{x \in GF(2)^n : f(x) = L_w(x)\}, \end{aligned}$$

where $L_w(x) = w \bullet x$ is an arbitrary linear function.

Then,

$$\begin{aligned} F(w) &= 2^n - 2d(f, L_w) \\ &= 2\epsilon(f, L_w) - 2^n. \end{aligned}$$

$d(f, L_w)$ is also called as a nonlinearity of any function f which means a distance from a set of affine functions and L_w can be considered to be some form of linear approximation of a function f.

Theorem 1 (Parseval's theorem [11]) *For any Boolean function, $f : GF(2)^n \rightarrow GF(2)$,*

$$\sum_{w \in GF(2)^n} \left(\frac{F(w)}{2^n}\right)^2 = 1.$$

We can obtain the following corollary due to Parseval's theorem.

Corollary 1

$$\sum_{w \in GF(2)^n} \left(\frac{\epsilon(f, L_w) - 2^{n-1}}{2^n}\right)^2 = \frac{1}{4}.$$

For a given function, $f : GF(2)^n \rightarrow GF(2)^m$, f can be expressed in terms of m Boolean functions, $f_1, f_2, \ldots, f_m$ as $f = (f_1, f_2, \ldots, f_m)$. For $\beta = (\beta_1, \ldots, \beta_n) \in GF(2)^m$, $L_\beta(f) : GF(2)^m \rightarrow GF(2)$ can be defined as below :

$$L_\beta(f) = \sum_{i=1}^{m} f_i \cdot \beta_i$$

By **Theorem 1**, for any $\beta \in GF(2)^m$,

$$\sum_{w \in GF(2)^n} \left(\frac{e(L_w, L_\beta(f)) - 2^{n-1}}{2^n}\right)^2 = \frac{1}{4}.$$

Thus for all $\beta \in GF(2)^m$,

$$\sum_{\beta \in GF(2)^m} \sum_{w \in GF(2)^n} \left(\frac{e(L_w, L_\beta(f)) - 2^{n-1}}{2^n}\right)^2 = \frac{2^m}{4}.$$

This can be interpreted as any Boolean function should have a linear distribution table with nonzero entry. The linear distribution table of DES S-boxes which can be considered as 15 Boolean functions has always nonzero entry.

Next, we compute the average value and variation of the linear distribution table for any Boolean function.

Theorem 2 *For any Boolean function, $f(x) : GF(2)^n \to GF(2)$, the average value, m, of $S(w) = e(f, L_w) - 2^{n-1}$ is equal to $1/2$ or $-1/2$ and its variation, σ^2, is equal to $2^{n-2} - 1/4$.*

Proof:

$$\begin{aligned}
\sum_w S(w) &= \sum_w \frac{F(w)}{2} \\
&= \frac{1}{2} \sum_w \sum_x (-1)^{f(x)} (-1)^{L_w(x)} \\
&= \frac{1}{2} \sum_x \left(\sum_w (-1)^{L_w(x)}\right) (-1)^{f(x)} \\
&= \frac{1}{2} (-1)^{f(0)} 2^n.
\end{aligned}$$

Thus,

$$m = \begin{cases} 1/2 & \text{if } f(0) = 0 \\ -1/2 & \text{if } f(0) = 1. \end{cases}$$

Since $\sum_w S(w)^2/2^n = 2^{n-2}$, $\sigma^2 = 2^{n-2} - 1/4$. □

The following result is based on a series of computer experiments.

Conjecture 1 *For any permutation $: GF(2)^n \to GF(2)^n$, the entry values in a form of its linear distribution table are bounded within $\pm 4 \cdot 2^{n/2-1}$ (i.e., 4σ) and the maximal absolute value varies around 4σ.*

This conjecture indirectly tells us that **Condition 1** is a reasonable constraint.

4 Iterative Linear Approximation

As the number of round in DES increases, the probability of linear approximation rapidly decreases. It will be impratical to find all linear approximations. Any iterative characteristic is a very useful tool to do a linear cryptanalysis successfully as in differential cryptanalysis. Using an iterative linear approximation with high probability, we can find a partial key information by linear cryptanalysis efficiently than by key-exhaustive search.

Definition 5 (nR iterative linear approximation) *The n-round (simply, nR) iterative linear approximation is defined as*

$$I_1[Z_1] \oplus I_n[Z_n] = K_2[Z_2] \oplus \cdots \oplus K_{n-1}[Z_{n-1}].$$

For the consecutive n-rounds, the XORed values of n-2 keys in an (n-2)-round can be expressed by its input and output XORed values. When this expression holds with probability $q = p + 1/2$, the probability of this linear approximation is to be p. Also, we denote nR iterative linear approximation as $-A_1 \cdots A_{n-2}-$ and its concatenated expression as $-A_1 \cdots A_{n-2} - A_{n-2} \cdots A_1 -$.

Example 2 Let two linear approximations A and B be

$$A: I[Z_1].K[Z_3] \longrightarrow O[Z_2]. \quad (1)$$
$$B: I[Z_2].K[Z_4] \longrightarrow O[Z_1]. \quad (2)$$

From Eqns. (1) and (2), we get

$$I_2[Z_1] \div O_2[Z_2] = K_2[Z_3]. \quad (3)$$
$$I_3[Z_2] \div O_3[Z_1] = K_3[Z_4]. \quad (4)$$

Since

$$I_1[Z_2] \div I_3[Z_2] = O_2[Z_2] \quad (5)$$
$$I_2[Z_1] \div I_4[Z_1] = O_3[Z_1]. \quad (6)$$

XORing Eqns. (3), (4), (5) and (6) term by term, we end up with a 4R linear approximation after cancelling common terms:

$$I_1[Z_2] \div I_4[Z_1] = K_2[Z_3] \div K_3[Z_4]. \quad (7)$$

By concatenating Eq.(7), we can get a 7R linear approximation as follows:

$$I_1[Z_2] \div I_7[Z_2] = K_2[Z_3] \div K_3[Z_4] \div K_5[Z_4] \div K_6[Z_3].$$

□

By Lemma 3 in [8], if we find an nR iterative linear approximation with probability p, then we can also obtain $(k \cdot (n-1)+1)$R linear approximation with probability $2^{k-1}p^k$ when applying nR iterative linear approximation k times.

In this section, we discuss the probability of 3R, 4R, 5R, and nR ($n > 6$) iterative linear approximations to prevent DES from being broken by a successful linear cryptanalysis.

4.1 3R Iterative Linear Approximation

The 3R iterative linear approximation has a form of $I_1[Z_1] \oplus I_3[Z_1] = K_1[Z_2]$, *i.e.*, there exists a linear correlation between key and output subblocks without input subblock as $O_i[Z_1] = K_i[Z_2]$. This case always occurs when a repeated input of two outer bits to a DES S-box are given to two neighbouring S-boxes. Thus, we can build the 3R iterative linear approximation from this case.

Theorem 3 *There exists a 3R iterative linear approximation if and only if the input of Si-box and the input of S(i+1)-box are* 3_x *and* 30_x *respectively.*

If $NS_i(3_x, \beta_1)$ and $NS_{i+1}(30_x, \beta_2)$ are not equal to zero, we can build some 3R iterative linear approximations. From the 3R iterative linear approximation $-A-$ with probability p, we can build the 15-round linear approximation as :

$$-A-A-A-A-A-A-A-$$

and the total probability for this approximation to hold is $2^6 p^7$.

Example 3 In DES, $NS_7(3_x, f_x) = 8$, $NS_8(30_x, d_x) = -12$, where the probability p of 3R iterative linear approximation is $2 \cdot \frac{8}{64} \cdot \frac{-12}{64} = \frac{-3}{64}$. Using this probability, the total complexity, $Comp_1$, to break the full-round DES by linear cryptanalysis is

$$Comp_1 = (2^6|p|^7)^{-2} = (\frac{3^7}{2^{36}})^{-2} = 2^{49.8}.$$

□

Thus, the necessary condition that this attack is no more efficient than key-exhaustive search is $2^6|p|^7 \leq 2^{-28}$, *i.e.*, $|p| \leq 2^{-4.9}$. In other words,

$$|2 \cdot \frac{NS_i(3_x, \beta_1)}{64} \cdot \frac{NS_{i+1}(30_x, \beta_2)}{64}| \leq 2^{-4.9}. \quad (8)$$

In [5], we suggested an additional design criterion for constructing DES-like S-boxes against differential cryptanalysis, *i.e.*, $S(x) \neq S(x \oplus 11efg0)$ for any efg. If any DES-like S-box satisfies this criterion, the values of $NS(30_x, \beta)$ and $NS(31_x, \beta)$ are to be zero for any β so that the LHS of Eq. (8) is always equal to zero.

Condition 2 $S(x) \neq S(x \oplus 11efg0)$ *for any* efg.

4.2 4R Iterative Linear Approximation

We discuss cases when a 4R iterative linear approximation occurs from two given linear approximations such as,

$$A : I_2[Z_1], K_2[Z_3] \longrightarrow O_2[Z_2] \quad (9)$$

$$B : I_3[Y_1], K_3[Y_3] \longrightarrow O_3[Y_2] \quad (10)$$

If we linearly approximate the 2nd round and 3rd round function of DES to A and B, respectively, $I_2[Z_1]$ should must be equal to the XORed value between the 3rd round output and 4th round input, and $I_3[Y_1]$ should be equal to the XORed value between the 1st round input and 2nd round output in order to get an useful 4R iterative linear approximation (refer to **Example 2**).

Theorem 4 *By concatenating two linear approximations Eqns. (9) and (10) with probability p_1 and p_2, respectively, the condition for building a 4R iterative linear approximation is*

$$Z_1 = Y_2, Z_2 = Y_1.$$

Then, the 4R iterative linear approximation is of the form

$$I_1[Z_2] \oplus I_4[Z_1] = K_2[Z_3] \oplus K_3[Y_3] \tag{11}$$

with probability $2p_1p_2$.

If the 4R iterative linear approximation $-AB-$ with probability p is given, we can build the 15-round linear approximation as

$$-AB - BA - AB - BA - AB.$$

The necessary condition that this attack is no more efficient than key-exhaustive search is $|2^4p^5| \leq 2^{-28}$, *i.e.*, $|p| \leq 2^{-6.4}$.

We can obtain the best 4R iterative linear approximation when $\#\delta(A) = \#\delta(B) = 1$. This means that the output of Si-box equals the input of Sj-box at the next round and the output of Sj-box equals the input of Si-box at the previous round. If the corresponding probabilities are p_1 and p_2, respectively, the total probability of the 4R linear approximation is $2p_1p_2$. Thus, values of p_1 and p_2 should be less than $2^{-3.7}$. However, values of p_1 and p_2 must be zero when we consider a 5R iteration linear approximation, which will be discussed in the next section.

Condition 3 *The followings (18 cases in total) are necessary so that the 4R iterative linear approximation will not occur.*

- *S1-box* : $NS_1(4,4) = NS_1(2,2) = 0$
- *S2-box* : $NS_2(4,4) = NS_2(2,1) = 0$
- *S3-box* : $NS_3(8,4) = NS_3(4,8) = 0$
- *S4-box* : $NS_4(8,4) = NS_4(2,2) = 0$
- *S5-box* : $NS_5(16,1) = NS_5(8,8) = NS_5(2,4) = 0$
- *S6-box* : $NS_6(16,4) = NS_6(4,8) = NS_6(2,2) = 0$
- *S7-box* : $NS_7(4,8) = NS_7(2,1) = 0$
- *S8-box* : $NS_8(16,1) = NS_8(2,4) = 0$

If we choose DES-like S-boxes satisfying **Condition 3**, the possible 4R linear approximations will be as in Table 1.

Let each probability of 4 cases be p_1, p_2, p_3, and p_4. All $p_i (i = 1, \ldots, 4)$ must be less than $2^{-2.35}$ since $|2^3p_1p_2p_3p_4| \leq 2^{-6.4}$. These probabilities should be less than 2^{-3} in considering 5R iterative linear approximation, which will be discussed in the next section. Moreover, since the output (or input) of an S-box is equal to an input (or output) of two different S-boxes, the Hamming weight between input and output of an S-box must be less than or equal to 2.

Table 1: Possible cases of 4R linear approximation

Case	Source S-boxes	Destination S-boxes
1	Si	Sk and/or Sl
2	Sj	Sk and/or Sl
3	Sk	Si and/or Sj
4	Sl	Si and/or Sj

Condition 4 *For* $\alpha \in GF(2)^6$ *and* $\beta \in GF(2)^4$,

$$W(\alpha), W(\beta) \leq 2 \implies |NS(\alpha, \beta)| \leq 8$$

If we build linear approximation with more than five S-boxes, the holding probability $|2^4 p_1 p_2 p_3 p_4 p_5|$ must be less than $2^{-6.4}$ and we guess that there exists at least one S-box such that $|NS(\alpha, \beta)| \leq 12$ among five S-boxes.

Thus, if we can find DES-like S-boxes satisfying both **Condition 3** and **Condition 4** simultaneously, the linear cryptanalysis with the 4R iterative approximation would be less efficient than key-exhaustive search.

Example 4 In DES, the following 4R iterative linear approximation holds :

$$I_1[0] \longrightarrow O_1[5] \ : \ NS_8(2, 8) = -2,$$
$$I_2[5] \longrightarrow O_1[0] \ : \ NS_7(4, 8) = 4.$$

Based on these approximations, the total complexity, $Comp_2$, to break the full-round DES by linear cryptanalysis is to be

$$Comp_2 = (2^4(|\frac{4}{64}\frac{-2}{64}|)^5)^{-2} = 2^{81}.$$

□

4.3 5R Iterative Linear Approximation

In [8], Matsui showed that DES can be broken within the complexity of 2^{47} using 5R iterative linear approximation.

Theorem 5 *When three linear approximations are given by*

$$A \ : \ I_2[Z_1], K_2[Z_3] \longrightarrow O_2[Z_2]$$
$$B \ : \ I_3[Y_1], K_3[Y_3] \longrightarrow O_3[Y_2]$$
$$C \ : \ I_4[X_1], K_4[X_3] \longrightarrow O_4[X_2],$$

we can obtain a 5R iterative linear approximation only if $Z_1 = Y_2 = X_1, Y_1 = Z_2 \cup X_2 - Z_2 \cap X_2$, *and the 5R iterative linear expression* $-ABC-$ *is of the form*

$$I_1[Z_2] \doteq I_5[X_2] \ = \ K_2[Z_3] \doteq K_3[Y_3] \doteq K_4[X_3].$$

Example 5 In DES, we know 5R iterative linear approximation $-ABC-$ as

$$\begin{array}{lll} A & : & I[15], K[22] \longrightarrow O[7, 18, 24] \quad \text{with } 10/64 \\ B & : & I[29], K[44] \longrightarrow O[15] \quad \text{with } -2/64 \\ C & : & I[15], K[22] \longrightarrow O[7, 18, 24, 29] \quad \text{with } -20/64. \end{array}$$

The total probability is equal to $2^{-7.3}$. □

Using the 5R iterative linear approximation $-ABC-$, we can build 15-round linear approximation as

$$-ABC - CBA - ABC - DE.$$

The probabilities of D and E will be less than 2^{-2} by **Condition 1**. If the probability of $-ABC-$ is p, $|p|^3$ must be less than 2^{-28}, *i.e.*, $|p| \leq 2^{-9.4}$.

If three linear approximation (ABC) in 5R iterative linear approximation consists of only three S-boxes, each probability must be less than $2^{-3.8}$. In order to satisfy this,

$$|NS(\alpha, \beta)| \leq 4 \quad \text{with} \quad W(\alpha) = 1 \tag{12}$$

where $\alpha \in GF(2)^6$ and $\beta \in GF(2)^4$.

When $\alpha = 10_x$, we could not find any DES-like S-boxes satisfying Eq. (12) through computer experiments. Thus, other design condition except Eq. (12) should be considered. However, if the 5R iterative linear approximation consists of linear approximation from three S-boxes, then $|Y_2| = 1$ since $Z_1 = Y_2 = X_1$ and $\#\delta(A) = \#\delta(B) = 1$ from P-permutation in DES F-function. By the same reason, $|Y_1| = 1$ because $\#\delta(B) = 1$ and the input of B is a subset of A or C. Also, we cannot build a 5R iterative linear approximation with nonzero probability which consists of three S-boxes satisfying **Condition 3** because $\delta(A) = \delta(C) = \{\text{Si-box}\}$ and $\delta(B) = \{\text{Sj-box}\}$. Moreover, if DES-like S-boxes satisfy **Condition 3**, we cannot build a 5R iterative linear approximation with nonzero probability which consists of four S-boxes, *i.e.*, $\delta(B) = \{\text{Sj-box}, \text{Sk-box}\}$. Thus, we can build a 5R iterative linear approximation consisted of more than five S-boxes.

Let $\delta(A) = \delta(C) = \{\text{Si-box}, \text{Sj-box}\}$ and $\delta(B) = \{\text{Sk-box}\}$, the probability of linear approximation from Si-box and Sj-box at the 2nd (4th) round be p_1 and p_2 (p_4 and p_5), respectively, and the probability of linear approximation at the 3rd round from Sk-box be p_3. Then, the input of the Sk-box coming from the output of Si-box and/or Sj-box will have a Hamming weight less than 2. By **Condition 4**, p_3 will be less than 2^3, thus there will be one bit difference in the output of Si-box (Sj-box) at the output of A (C).

Condition 5 *For* $\alpha \in GF(2)^6$, β_1 *and* $\beta_2 \in GF(2)^4$.
$W(\alpha) = 1$ *and* $W(\beta_1 + \beta_2) = 1 \Longrightarrow |NS(\alpha, \beta_1) \cdot NS(\alpha, \beta_2)| \leq 48$.

When the above condition is given, $p_1 p_4$ (and/or $p_2 p_5$) $\leq 2^{-6.4}$. The probability of this 5R iterative linear approximation is

$$2^4 |p_1 p_4 p_3 p_2 p_5| \leq 2^4 \cdot 2^{-6.4} \cdot 2^{-3} \cdot 2^{-2} \cdot 2^{-2} = 2^{-9.4}.$$

Thus **Condition 5** is necessary in order that linear cryptanalysis by 5R iterative linear approximation is less efficient than key-exhaustive search. For a given 5R iterative linear approximation $(-A_1 A_2 A_3-)$ consisted of five S-boxes, if the probability of linear approximation coming from one of S-boxes is less than $2^{-2.4}$, then the probability of $-A_1 A_2 A_3-$ will be $2^{-9.4}$ by **Condition 1**.

4.4 nR Iterative Linear Approximation

We can generalize from 5R iterative linear approximation into nR iterative linear approximation. When $n-2$ linear approximations are given below :

$$\begin{aligned} A_2 &: I_2[X_2], K_2[Y_2] \longrightarrow O_2[Z_2] \text{ with } p_2 \\ A_3 &: I_3[X_3], K_3[Y_3] \longrightarrow O_3[Z_3] \text{ with } p_3 \\ &\cdot \\ A_{n-1} &: I_{n-1}[X_{n-1}], K_{n-1}[Y_{n-1}] \longrightarrow O_{n-1}[Z_{n-1}] \text{ with } p_{n-1}, \end{aligned}$$

then we can obtain an nR iterative linear approximation in the form of

$$I_1[Z_2] \oplus I_n[Z_{n-1}] = K_2[Y_2] \oplus \cdots \oplus K[n-1][Y_{n-1}]$$

only if $X_k = Z_{k-1} \triangle Z_{k+1} = Z_{k-1} \cup Z_{k+1} - Z_{k-1} \cap Z_{k+1}$ for $k = 3, 4, \ldots, n-2$, $X_2 = Z_3$, and $X_{n-1} = Z_{n-2}$. The total probability is $2^{n-3} \prod p_2 \cdots p_{n-1}$. For the 16-round DES, n=6,7 and 8 are possible cases. In these cases, we can concatenate this nR iterative linear approximation in order to do a successful linear attack. If we can lower the total probability, then linear attack for DES will be difficult.

In [10], Matsui suggested a way of linear approximation for DES called as "Type-I approximation"which means that any nR linear approximation can be derived when at most one S-box is approximated in a single round. Due to **Condition 3**, however, nR iterative linear approximation with nonzero probability like Type-I approximation cannot be obtained.

5 Concluding Remarks

In this paper, we proposed five necessary conditions in order to immunize DES S-boxes from linear cryptanalysis. If we find DES-like S-boxes satisfying these additional five conditions, we can conclude that DES with new S-boxes is resistant to linear cryptanalysis and differential cryptanalysis as well. Even if new S-boxes are substituted into the current DES S-boxes, the key-exhaustive search machine suggested by Wiener[12] can be valid for a successful attack of new DES. Against this attack, we suggest that the key scheduling part of DES can be redesigned to increase the current DES key size.

In the full paper, we will suggest new DES-like S-boxes satisfying our conditions and discuss the complexity of new DES against both differential cryptanalysis and linear cryptanalysis.

References

[1] "Data Encryption Standard", FIPS-Pub. 46, National Bureau of Standards (former NIST), 1977.

[2] E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystems", *Crypto'90 Extended Abstracts*, UCSB, 1990.

[3] E. Biham and A. Shamir, "Differential Cryptanalysis of the full 16-round DES", *Crypto'92 Extended Abstracts*, 1992.

[4] E. Biham, "A Comment on s^3DES S-boxes", Private Communication, Jan.4, 1994.

[5] K. Kim, S. Park and S. Lee, "Reconstruction of s^2DES S-boxes and their Immunity to Differential Cryptanalysis", Proceedings of 1993 Korea-Japan Joint Workshop on Information Security and Cryptology (JW-ISC'93), Oct.24-26, Seoul, 1993.

[6] K. Kim, S.Lee and S. Park, "How to Strengthen DES against Differential Attack", Submitted to Eurocrypt'94.

[7] Lars R. Knudsen, "An Analysis of Kim, Park and Lee's DES-like S-boxes", Private Communication, Jun.18, 1993.

[8] M. Matsui, "Linear Cryptanalysis Method for DES Cipher", *Eurocrypt'93 Extended Abstracts*, 1993.

[9] M. Matsui, "Linear Cryptanalysis Method for DES Cipher(III) (in Japanese)", Proc. of SCIS'94, Biwako, Japan, pp.4A.1-11, Jan., 1994.

[10] M. Matsui, "On Correlation between the Order of S-boxes and the Strength of DES", *preprint*.

[11] W. Meier and O. Staffelbach, "Nonlinearity Criteria for Cryptographic Functions", *Advances in Cryptology-Eurocrypt'89*, pp.549-562, Springer-Verlag, 1990.

[12] M.J. Wiener, "Efficient DES Key Search", *Crypto'93 Extended Abstracts*, 1993.

Linear Approximation Versus Nonlinearity

Josef Pieprzyk
Chris Charnes
Jennifer Seberry*

Center for Computer Security Research
Department of Computer Science
University of Wollongong
Wollongong, NSW 2500, AUSTRALIA

e-mail: josef@cs.uow.edu.au
charnes@cs.uow.edu.au
jennie@cs.uow.edu.au

Abstract

Recently Matsui [2] announced an attack on the DES algorithm. The attack relies on approximation of S-boxes by linear functions. To find out the best linear approximation, Matsui defines the linear approximation tables (LAT) for S-boxes. In this work we examine the relation between Matsui's linear approximation tables and nonlinearities of corresponding S-boxes.

1 Introduction

The recent cryptographic attack introduced by Matsui [2] relies on the approximation of S-boxes by linear functions. For a given S-box, every output or linear combination of outputs can be approximated by linear functions. Matsui showed how to find the best linear approximation for the S-boxes used in the DES algorithm and how to use this to break the algorithm. However there appears to be some misunderstanding of the relationship between this attack

*Support for this project was provided in part by the Australian Research Council under the reference number A491131885

and the significance of the linear approximation tables (LAT), introduced by Matsui. In this paper we will explain the relation between nonlinearity and the linear approximation tables. We prove that linear approximation tables give the nonlinearities of every linear combination of output functions. Hence the design criteria for S-boxes now has to include an extended measure of nonlinearity of S-boxes; in particular this measure must be consistent with the measure introduced by Nyberg in [4]. Some preliminary comments about the influence of linear cryptanalysis on the design of S-boxes can be found in [9].

2 Background

We denote by $x \in \{0,1\}^n = X^n$ a binary string of length n. A Boolean function f is defined as a mapping

$$f : X^n \longrightarrow X.$$

The set of all n-variable linear Boolean functions is

$$L_n = \{f \mid f : X^n \to X; f = a_1x_1 \oplus \ldots \oplus a_nx_n\},$$

where $a_i \in \{0,1\}$ and $x_i \in X$ for $i = 1,\ldots,n$, and $\oplus$ is the Exclusive-OR operation. The set of n-variable affine Boolean functions is

$$A_n = \{f \mid \ell \in L_n; f = \ell \oplus a_0\}$$

where $a_0 \in X$.

Any Boolean function $f(x)$ can be represented in the form of a truth table. The table is described by the following vector

$$f(x) = (f_0, f_1, \ldots, f_{2^n-1}),$$

where f_i is the value of the function $f(\alpha_i)$ and α_i is the binary representation of the integer i; i.e., $i = \sum_{j=1}^{n} 2^{j-1}\alpha_i[j]$ and $\alpha_i = (\alpha_i[1],\ldots,\alpha_i[n])$.

Definition 2.1 *The Hamming distance between two Boolean functions $f,g : X^n \to X$, is defined as*

$$d(f,g) = wt(f_0 \oplus g_0, f_1 \oplus g_1, \ldots, f_{2^n-1} \oplus g_{2^n-1}),$$

where $wt(\alpha)$ is the weight–the number of ones of the binary string $\alpha \in X^n$.

The nonlinearity of Boolean functions is defined as, (see also [1],[3],[4],[6]).

Definition 2.2 *The nonlinearity $\mathcal{N}(f)$ of a Boolean function $f : X^n \to X$ is*

$$\mathcal{N}(f) = \min_{\ell \in A_n} d(\ell, f),$$

i.e., is the minimal distance between the function f and the set of affine functions.

Nonlinearity can also be expressed in terms of the Walsh transform $\hat{F}(u)$ of f:

$$\mathcal{N}(f) = 2^{n-1} - \max_{u \in X^n} \frac{|\hat{F}(u)|}{2},$$

see [8].

Definition 2.3 *A $(n \times m)$ S-box is a collection of m functions $F_i(x)$, $i = 1, \ldots, m$, in n Boolean variables $x = (x_1, \ldots, x_n)$ for which*

$$S(x) = (F_1(x), \ldots, F_m(x)).$$

The next definition is taken from Matsui's paper [2].

Definition 2.4 *A linear approximation table LAT for a S-box $S(x)$ is*

$$LAT_S(\alpha, \beta) = \#\{x \mid 0 \le x \le 2^n - 1; (\bigoplus_{i=1}^{n}(x[i] \bullet \alpha[i])) = (\bigoplus_{j=1}^{m}(S(x)[j] \bullet \beta[j]))\},$$

where $\bullet$ is the bitwise AND operation, $x = \sum_{i=1}^{n} 2^{i-1} x[i]$ and $\alpha = \sum_{i=1}^{n} 2^{i-1} \alpha[i]$ $(x[i], \alpha[i] \in X)$.

The nonlinearity of a $(n \times m)$ S-box $S(x)$ is (see Nyberg [4])

$$\mathcal{N}(S(x)) = \min_{w=(w_1,\ldots,w_m) \in X^m; v \in X} \mathcal{N}(w_1 F_1 \oplus \ldots \oplus w_m F_m \oplus v). \quad (1)$$

3 Properties of Linear Approximation Tables

A binary string $\alpha = (\alpha[1], \ldots, \alpha[n])$ generates a linear function $\ell_\alpha(x) \in L_n$ defined as

$$\ell_\alpha(x) = x_1 \alpha[1] \oplus \ldots \oplus x_n \alpha[n].$$

Also a binary string $\beta = (\beta[1], \ldots, \beta[n])$ gives a Boolean function $S_\beta(x)$ defined as the linear combination of output functions of a S-box:

$$S_\beta(x) = F_1(x)\beta[1] \oplus \ldots \oplus F_n(x)\beta[n].$$

Lemma 3.1

$$LAT_S(\alpha,\beta) = 2^n - d(\ell_\alpha(x), S_\beta(x)). \tag{2}$$

Proof : The (α,β)-entry of the LAT_S table indicates the number of arguments x for which the values of $\ell_\alpha(x)$ and $S_\beta(x)$ coincide. On the other hand, the distance $d(\ell_\alpha(x), S_\beta(x))$ gives the number of arguments for which the two functions differ. Thus equation (2) holds. □

Lemma 3.2 *For a fixed vector $\beta \in X^n$, the following inequality holds*

$$\forall_{\alpha\in X^n}\mathcal{N}(S_\beta(x)) \leq LAT_S(\alpha,\beta) \leq 2^n - \mathcal{N}(S_\beta(x)). \tag{3}$$

Proof : ¿From equation (2) we have

$$\min_{\alpha\in X^n} LAT_S(\alpha,\beta) \leq 2^n - d(\ell_\alpha(x), S_\beta(x)).$$

Note that

$$\min_{\alpha\in X^n} d(\ell_\alpha(x), S_\beta(x)) = \min_{\ell\in L_n} d(\ell(x), S_\beta(x))$$

but this is

$$\begin{cases} \mathcal{N}(S_\beta(x)) & \text{if } \ell_{min} \in L_n; \\ 2^n - \mathcal{N}(S_\beta(x)) & \text{otherwise,} \end{cases}$$

where ℓ_{min} is a linear function for which

$$\min_{\ell\in L_n} d(\ell(x), S_\beta(x)) = d(\ell_{min}, S_\beta(x)).$$

□

The (α,β)-entry specifies the closest linear function which approximates $S_\beta(x)$. Since the linear functions $\ell_\alpha(x)$ comprise the whole of L_n, it follows that there must be an α^* which gives best approximation to the function $S_\beta(x)$, moreover

$$LAT_S(\alpha^*,\beta) = \begin{cases} \mathcal{N}(S_\beta(x)) \text{ or;} \\ 2^n - \mathcal{N}(S_\beta(x)). \end{cases} \tag{4}$$

So we have the following theorem.

Theorem 3.1 *For every function $S_\beta(x)$, the best linear approximation is given by the function $\ell_{\alpha^*}(x)$, moreover expression (4) holds.*

This theorem has the following corollaries.

Corollary 3.1 *Let* $S(x) = (F_1(x), \ldots, F_n(x))$ *and* β *be a combination of functions* $F_i(x)$ *such that* $S_\beta(x)$ *is affine, then there is a single* α^* *such that* $LAT_S(\alpha^*, \beta) = 0$ *or* 2^n. *The other entries of the table are:* $LAT_S(\alpha, \beta) = 2^{n-1}$, *where* $\alpha \neq \alpha^*$.

Example: Consider $GF(2^4)$ and the field automorphism $\sigma : x \rightarrow x^2$. Since σ is a linear operation, the linear approximation table has the form:

8	8	16	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	16	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	16
8	16	8	8	8	8	8	8	8	8	8	8	8	8	8
16	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	16	8
8	8	8	8	8	8	8	8	8	8	8	8	16	8	8
8	8	8	8	8	8	8	16	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	16	8	8	8	8
8	8	8	16	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	16	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	16	8	8	8	8	8
8	8	8	8	8	8	8	8	16	8	8	8	8	8	8
8	8	8	8	8	16	8	8	8	8	8	8	8	8	8
8	8	8	8	16	8	8	8	8	8	8	8	8	8	8

It is easy to see that every function $S_\beta(x)$ can be represented by a linear function (where the corresponding entry is $2^4 = 16$).

Corollary 3.2 *¿From a given* $LAT_S(\alpha, \beta)$, *it is possible to recover all the nonlinearities of* $S_\beta(x)$ *by selecting the minimal and the maximal values from the column* $LAT_S(\alpha, \beta)$; $\alpha \in X^n$. *Denote these two values by* $LAT_S(\alpha_{min}, \beta)$ *and* $LAT_S(\alpha_{max}, \beta)$. *Then the nonlinearity of* $S_\beta(x)$ *is* $\min(LAT_S(\alpha_{min}, \beta), 2^n - LAT_S(\alpha_{max}, \beta))$.

Corollary 3.3 *The nonlinearity of* $S(x)$ *is*

$$\mathcal{N}(S(x)) = \min_{\beta \in X^m} \mathcal{N}(S_\beta(X)). \tag{5}$$

4 Linear Approximation Tables of Permutations

In this section we assume that $m = n$ and $S(x) = (F_1(x), \ldots, F_n(x))$ is a permutation. Therefore $S(x)$ has an inverse: $S^{-1}(x) = (F_1^{-1}(x), \ldots, F_n^{-1}(x))$. The following theorem describes the relation between the linear approximation tables of $S(x)$ and $S^{-1}(x)$.

Theorem 4.1

$$LAT_S(\alpha, \beta) = LAT_{S^{-1}}(\beta, \alpha).$$

Proof : The definition of $LAT_S(\alpha,\beta)$ states that

$$LAT_S(\alpha,\beta) = \#\{x \mid 0 \le x \le 2^n - 1; (\bigoplus_{i=1}^{n}(x[i] \bullet \alpha[i])) = (\bigoplus_{j=1}^{n}(S(x)[j] \bullet \beta[j]))\}.$$

Since our S-box is a permutation we can count the number of output values $y = S(x)$ instead of $x = S^{-1}(y)$, this does not change the entries $LAT_S(\alpha,\beta)$. Therefore

$$\begin{aligned} LAT_S(\alpha,\beta) &= \#\{y \mid 0 \le y \le 2^n - 1; \\ &\quad (\bigoplus_{i=1}^{n}(S^{-1}(y)[i] \bullet \alpha[i])) = (\bigoplus_{j=1}^{m}(y[j] \bullet \beta[j]))\} \\ &= LAT_{S^{-1}}(\beta,\alpha). \end{aligned}$$

□

Corollary 4.1 *If $\ell_\alpha(x)$ is the best linear approximation of $S_\beta(x)$ then $\ell_\beta(x)$ is the best linear approximation of $S_\alpha^{-1}(x)$.*

There have been several definitions proposed for the nonlinearity of permutations. In earlier work, see [7], the nonlinearity of a permutation was defined as the minimum value of nonlinearities of the components; so

$$\mathcal{N}_{(1)}(S(x)) = \min_{i=1,\ldots,n} \mathcal{N}(F_i(x)),$$

where $S(x) = (F_1(x),\ldots,F_n(x))$.

But there are permutations whose every component is highly nonlinear, and yet some components of the inverse permutation have low nonlinearity. In view of this it was concluded that the appropriate measure of nonlinearity of permutations should be

$$\mathcal{N}_{(2)}(S(x)) = \min_{i=1,\ldots,n} (\mathcal{N}(F_i(x)), \mathcal{N}(F_i^{-1}(x)),$$

where $S^{-1}(x) = (F_1^{-1}(x),\ldots,F_n^{-1}(x))$ is the inverse of $S(x)$.

Regarding the linear approximation attack of Matsui [2], it is obvious that the nonlinearity of a permutation should be defined using its linear approximation table, or equivalently by definition (1) (see Nyberg [4]). Assume that

$$\gamma = \max_{\alpha,\beta=1,\ldots,2^n-1} \mid LAT_S(\alpha,\beta) - 2^{n-1} \mid$$

then the nonlinearity of a permutation $S(x)$ is

$$\mathcal{N}(S(x)) = 2^{n-1} - \gamma. \tag{6}$$

It is obvious that

$$\mathcal{N}_{(1)}(S(x)) \geq \mathcal{N}_{(2)}(S(x)) \geq \mathcal{N}(S(x)).$$

The nonlinearity $\mathcal{N}(S(x))$ can be obtained from the LAT_S table by selecting the column (indexed by β) which has the the smallest nonlinearity

$$\mathcal{N}(S(x)) = \min_{\beta=1,\ldots,2^n-1} \mathcal{N}(S_\beta(x)). \tag{7}$$

The same value can also be obtained by selecting the row (indexed by α) with the smallest nonlinearity, so

$$\mathcal{N}(S^{-1}(x)) = \min_{\alpha=1,\ldots,2^n-1} \mathcal{N}(S_\alpha^{-1}(x)). \tag{8}$$

Therefore $\mathcal{N}(S(x)) = \mathcal{N}(S^{-1}(x))$. (Compare this with Theorem 1 in Nyberg [4].)

Example: Consider the cubing permutation in $GF(2^4)$. The linear approximation table for this permutation has the following form.

10	10	10	10	10	10	12	8	8	6	8	6	8	6	6
10	6	6	10	10	6	8	8	8	10	12	6	8	10	10
10	10	10	6	6	6	8	8	8	6	8	10	12	10	10
6	6	10	6	10	10	8	8	12	10	8	10	8	6	10
10	6	6	10	10	6	8	8	8	10	12	6	8	10	10
6	10	6	10	6	10	8	12	8	10	8	10	8	10	6
10	6	6	10	10	6	8	8	8	10	12	6	8	10	10
6	10	6	10	6	10	8	12	8	10	8	10	8	10	6
10	10	10	6	6	6	8	8	8	6	8	10	12	10	10
10	10	10	6	6	6	8	8	8	6	8	10	12	10	10
6	6	10	6	10	10	8	8	12	10	8	10	8	6	10
10	10	10	10	10	10	12	8	8	6	8	6	8	6	6
10	10	10	10	10	10	12	8	8	6	8	6	8	6	6
6	10	6	10	6	10	8	12	8	10	8	10	8	10	6
6	6	10	6	10	10	8	8	12	10	8	10	8	6	10

The smallest entry is 6 and the largest is 12, so the nonlinearity is $\min(6, 16-12) = 4$. All pairs (α, β) with entries 12 give the most effective approximation of $S_\beta(x)$ by the complement of the linear function $\ell_\alpha(x)$.

5 Conclusions

The core of Matsui's [2] linear cryptanalysis is the linear approximation table. We showed that these tables not only give nonlinearity profiles of the output functions, but also characterize the nonlinearities of their linear combinations. In view of Matsui's attack designers of S-boxes now have to include an additional requirement related to the nonlinearity of S-boxes. The nonlinearity of a S-box is the the smallest nonlinearity of the linear combinations of

output functions – this definition of nonlinearity was introduced by Nyberg [4]. To make an encryption algorithm resistant to linear cryptanalysis, it is necessary to use S-boxes of the highest possible nonlinearity.

Note that linear cryptanalysis fails if all the entries are 2^{n-1}. Thus we should design S-boxes so that their linear approximation tables contain entries close to 2^{n-1}. There are two independent ways of achieving this. The first way is to design S-boxes with the highest possible nonlinearity (getting the best design for a fixed size n of the S-box). The second way is to design S-boxes for a large parameter n, as nonlinearities grow asymptotically to 2^{n-1}. It turns out that even a random selection of S-boxes, for a large enough parameter n, can generate a highly nonlinear box ([5]).

References

[1] C. Adams and S. Tavares. The structured design of cryptographically good S-boxes. *Journal of Cryptology*, 3:27–41, 1990.

[2] M. Matsui. Linear cryptanalysis method for DES cipher. Abstracts of EUROCRYPT'93, May 1993.

[3] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. *Proceedings of EUROCRYPT'89, Lecture Notes in Computer Science, Advances in Cryptology*, 434:549–562, 1989.

[4] K. Nyberg. On the construction of highly nonlinear permutations. In *Extended Abstracts - Eurocrypt'92*, pages 89–94, May 1992.

[5] L.J. O'Connor. An analysis of product ciphers based on the properties of Boolean functions. PhD thesis, the University of Waterloo, 1992. Waterloo, Ontario, Canada.

[6] J. Pieprzyk and G. Finkelstein. Towards effective nonlinear cryptosystem design. *IEE Proceedings-E, Computers and Digital Techniques*, 135(6):325–335, November 1988.

[7] J.P. Pieprzyk. On bent permutations. In *Proceedings of the International Conference on Finite Fields, Coding Theory, and Advances in Communications and Computing, Las Vegas*, August 1991.

[8] R.A. Rueppel. *Stream ciphers, in G. Simmons (ed.), Contemporary Cryptology - The Science of Information Integrity.* IEEE Press, New York, 1992.

[9] J. Seberry, X.M. Zhang, and Y. Zheng. Systematic generation of cryptographically robust S-boxes. Proceedings of the 1st ACM Conference on C0mputer and Communication Security, November 1993.