J. Cryptology (1996) 9: 1–19

Journal of
CRYPTOLOGY

Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis*

Howard M. Heys and Stafford E. Tavares
Department of Electrical and Computer Engineering,
Queen's University, Kingston, Ontario, Canada K7L 3N6

Communicated by Thomas A. Berson

Received 11 November 1993 and revised 22 September 1994

Abstract. In this paper we examine a class of product ciphers referred to as substitution-permutation networks. We investigate the resistance of these cryptographic networks to two important attacks: differential cryptanalysis and linear cryptanalysis. In particular, we develop upper bounds on the differential characteristic probability and on the probability of a linear approximation as a function of the number of rounds of substitutions. Further, it is shown that using large S-boxes with good diffusion characteristics and replacing the permutation between rounds by an appropriate linear transformation is effective in improving the cipher security in relation to these two attacks.

Keywords. Product cipher, Substitution-permutation network, S-box, Differential cryptanalysis, Linear cryptanalysis.

1. Introduction

The class of product ciphers considered in this paper is based on principles introduced by Shannon [28]. Shannon suggested that secure, practical product ciphers may be constructed using a "mixing transformation" consisting of a number of layers or rounds of "confusion" and "diffusion". The confusion component is a nonlinear substitution on a small subblock and the diffusion component is a linear mixing of the subblock connections in order to diffuse the statistics of the system.

Feistel [13] and Feistel *et al.* [14] were the first to introduce a practical architecture based on Shannon's concepts with a network structure consisting of a sequence of rounds of small substitutions (referred to as S-boxes), easily implemented by table lookup and connected by bit position permutations or transpositions. Such ciphers are generally

* This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Telecommunications Research Institute of Ontario, and was presented at the rump session of CRYPTO '93. Howard Heys is now with Electrical Engineering, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X5.

referred to as substitution-permutation networks or SPNs. The fundamental principles of an SPN form the foundation for many modern product ciphers, including DES [20], FEAL [29], and LOKI [10].

Recent cryptanalysis techniques have had a notable effect on the perceived security of many product ciphers. For example, DES has been found to be theoretically cryptanalyzable by differential cryptanalysis using a chosen plaintext approach [5] and by linear cryptanalysis using a known plaintext approach [18]. In this paper we examine the security of SPNs with respect to these two powerful cryptanalysis techniques and suggest structures that aid in resisting the attacks. In particular, we develop upper bounds on the probability of a differential characteristic and on the deviation of the probability of a linear approximation from the ideal value of $\frac{1}{2}$. The objective of such an analysis is to determine a flexible architecture that can be efficiently implemented in as few rounds as possible to provide suitably small probabilities for differential characteristics and linear approximations.

2. Background

We consider a general N-bit SPN as consisting of R rounds of $n \times n$ S-boxes. The number of S-boxes used in each round is represented by M where $M = N/n$. The plaintext and ciphertext are N-bit vectors denoted as $\mathbf{P} = [P_1\ P_2\ P_N]$ and $\mathbf{C} = C_1\ C_2 \cdots C_N]$, respectively. An S-box in the network is defined as an n-bit bijective mapping S: $\mathbf{X} \to \mathbf{Y}$ where $\mathbf{X} = [X_1\ X_2\ \cdots\ X_n]$ and $\mathbf{Y} = [Y_1\ Y_2\ \cdots\ Y_n]$. A simple example of an SPN is illustrated in Fig. 1 with $N = 16$, $R = 4$, and $n = 4$.

In general S-boxes may be keyed using one or both of the following methods:

1. Selection keying: key bits are used to select which mapping from a set of mappings is to be used for a particular S-box.
2. XOR mask keying: key bits are XORed with the network bits prior to entering an S-box.

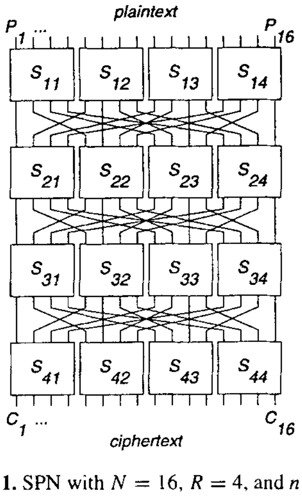

Fig. 1. SPN with $N = 16$, $R = 4$, and $n = 4$.

Note that method 2 may actually be considered as a special case of method 1. Method 2, however, ensures that all mappings in the set of possible mappings for an S-box are from the same cryptographic equivalence class [30]. We assume in our discussion that the network is keyed using XOR mask keying by XORing N bits of key (as determined by the key-scheduling algorithm) before the first substitution, after the last substitution, and between all substitutions. Decryption is performed by applying the key-scheduling algorithm in reverse and using the inverse S-boxes.

Rather than strictly confining ourselves to the basic form of substitutions connected by permutations, in this paper we consider the more general model of substitutions connected by invertible linear transformations. However, for consistency, we still refer to the more general architecture as an SPN.

Many papers have examined the cryptographically desirable properties of SPNs and their components. Acknowledged design criteria for the network S-boxes include nonlinearity [26], [19], [3], [21] and information-theoretic properties [15], [12]. Preferred permutation structures promote the influence of input bits [16], [4], [11].

Of particular importance to our discussion is the notion of nonlinearity and we use the following nonlinearity measures when referring to a boolean function or an S-box. The nonlinearity of an n-input boolean function, f: $\{0, 1\}^n \rightarrow \{0, 1\}$, is defined as the Hamming distance to the nearest affine function:

$$NL(f) = \min_{U_1,\ldots,U_n,V\in\{0,1\}} \#\left\{\mathbf{X} \mid f(\mathbf{X}) \neq \bigoplus_{i=1}^{n} U_i X_i \oplus V\right\}. \tag{1}$$

Consequently, the nonlinearity of an $n \times n$ bijective mapping or S-box S is defined as the minimum nonlinearity of all nonzero linear combinations of output functions:

$$NL(S) = \min_{W_1,\ldots,W_n\in\{0,1\},\ \text{all}\ W_i\neq 0} NL\left(\bigoplus_{i=1}^{n} W_i f_i\right), \tag{2}$$

where f_i represents the n-input function of the ith output of the S-box. Letting S^{-1} represent the inverse of S-box S, it can be shown that $NL(S^{-1}) = NL(S)$ [22].

3. Two Important Classes of Cryptanalysis

In this section we discuss two important classes of cryptanalysis which have had significant success against product ciphers.

(a) *Differential Cryptanalysis*

In a series of papers [5]–[8] Biham and Shamir successfully demonstrate the susceptibility of several product ciphers to differential cryptanalysis. Notably, differential cryptanalysis has been successful in breaking weakened versions of DES and can theoretically compromise the security of the full 16-round DES algorithm using 2^{47} chosen plaintexts. As well, differential cryptanalysis has been successfully applied to the FEAL cipher for up to 31 rounds of substitutions.

Differential cryptanalysis is an attack which examines changes in the output of the cipher in response to controlled changes in the input. In general, we are interested in bit changes or XOR differences within the network when two plaintexts, $\mathbf{P}'$ and $\mathbf{P}''$ are selected as inputs. We represent the XOR difference of the two plaintexts by $\Delta\mathbf{P} = \mathbf{P}' \oplus \mathbf{P}''$. Let the input and output difference to a particular round i be represented by $\Delta\mathbf{U}_i$ and $\Delta\mathbf{V}_i$, respectively. Differential cryptanalysis relies on the existence of highly probable "characteristics" where an r-round characteristic, Ω_r, is defined as a sequence of difference pairs: $\Omega_r = \{(\Delta\mathbf{U}_1, \Delta\mathbf{V}_1), \ldots, (\Delta\mathbf{U}_r, \Delta\mathbf{V}_r)\}$. The algorithm tries an appropriate number of chosen plaintexts with $\Delta\mathbf{P} = \Delta\mathbf{U}_1$ and counts the number of times that a subkey consisting of a subset of the key bits is consistent with the ciphertext difference, $\Delta\mathbf{C}$, assuming that the characteristic has occurred. If the characteristic occurs with probability p_{Ω_r}, the correct subkey bits are consistent with a probability of at least p_{Ω_r}. After an appropriate number of trials (typically several times more than $1/p_{\Omega_r}$ chosen plaintext pairs) the correct subkey will be counted significantly more times than incorrect subkeys.

In this paper we assume that a characteristic probability is determined by the product of the probabilities of the occurrence of a one-round difference pair. Letting $P(\Delta\mathbf{U}_i, \Delta\mathbf{V}_i)$ represent the probability of occurrence of the ith-round difference pair, then

$$p_{\Omega_r} = \prod_{i=1}^{r} P(\Delta\mathbf{U}_i, \Delta\mathbf{V}_i). \qquad (3)$$

Equation (3) gives exactly the characteristic probability taken over the independent distributions of plaintext and key. Hence, it strictly applies only when the plaintext and the keys applied at each round are independent and uniformly randomly selected for the encryption of each plaintext pair. In practice, (3) has been found to provide a reasonable estimate of the characteristic probability in ciphers with mutually dependent round keys.

Differential cryptanalysis of a basic SPN can be applied similarly to the attack on DES-like ciphers. For a DES-like cipher, differential cryptanalysis determines key bits associated with the input to the last round function by using knowledge (directly available from the right half of the ciphertext) of the two input values (and their difference) to the last round function combined with probabilistic knowledge of the output difference of the last round function. Similarly, differential cryptanalysis of a basic SPN can be used to determine the key bits XORed to the output of the last round of S-boxes by using knowledge of the two ciphertext values (and their difference) and the probabilistic knowledge of the input difference to the last round of S-boxes.

Hence, a differential attack of an SPN may be successful if the cryptanalyst is aware of a highly probable characteristic for the first $R-1$ rounds, Ω_{R-1}. The attack targets the round R S-boxes that are affected by the output changes of the characteristic, $\Delta\mathbf{V}_{R-1}$. The targeted subkey contains the key bits which are XORed with the output of the targeted S-boxes. Consequently, trying all subkey values, the cryptanalyst can use the known ciphertext values to decrypt the portion of round R associated with the target S-boxes. (Ciphertext pairs which have bit changes in the output of nontargeted S-boxes may be discarded since they cannot be generated by characteristic Ω_{R-1}). If the XOR difference of the target S-box inputs determined by the partial decryption corresponds to $\Delta\mathbf{V}_{R-1}$, then the corresponding subkey count is incremented. The actual subkey may be deduced as the key which is consistent most frequently over a number of trials.

Similarly to the analysis of the differential cryptanalysis of DES by Biham and Shamir [5], it can be assumed that, in circumstances where a highly likely $(R-1)$-round characteristic of probability $p_{\Omega_{R-1}}$ is known, the number of chosen plaintexts required to determine the subkey may be approximated by N_D where

$$N_D = \frac{1}{p_{\Omega_{R-1}}}. \tag{4}$$

In practice, the number of chosen plaintexts required will be greater than N_D since we have neglected the factor of 2 (which arises from the fact that the chosen plaintexts are encrypted in pairs) and since many incorrect subkeys, as well as the correct subkey, are counted at least once.

Let $\Delta\mathbf{X}$ and $\Delta\mathbf{Y}$ represent the input and ouput XOR differences, respectively, to an S-box when a plaintext difference $\Delta\mathbf{P}$ is applied to the cipher. The existence of highly probable characteristics depends on two factors: the distribution of S-box XOR difference pairs, $(\Delta\mathbf{X}, \Delta\mathbf{Y})$, and the diffusion of bit changes within the network. We define the probability of an S-box XOR pair $(\Delta\mathbf{X}, \Delta\mathbf{Y})$ to be the probability that $\Delta\mathbf{Y}$ occurs given that one of the input values for $\mathbf{X}$ is randomly selected and the other is related by the difference $\Delta\mathbf{X}$. Let the probability of the most likely S-box XOR pair (other than $(\Delta\mathbf{X} = \mathbf{0}, \Delta\mathbf{Y} = \mathbf{0})$) be p_δ.

Characteristics derived from S-box XOR pairs with high probabilities will typically occur with high probability. Several authors [12], [21], [2] have related the information-theoretic and nonlinear (bentness) properties of S-boxes to minimizing p_δ and suggest that S-boxes based on these principles provide resistance to differential cryptanalysis. In [25] O'Connor shows that, for large n, the S-box XOR pair probability is expected to be at most $n/2^{n-1}$. Hence, the expected maximum XOR pair probability decreases as the size of the S-box is increased. For 8 × 8 S-boxes, the expected maximum XOR pair probability satisfies $p_\delta \le 2^{-4}$.

High probability characteristics will also occur when poor diffusion of bit changes results in a characteristic involving a small number of S-boxes [17], [25]. Consider, for example, a four-round characteristic for an SPN with 4 × 4 S-boxes that have a maximum XOR pair probability of $p_\delta = \frac{1}{4}$. It is possible that a characteristic might exist with only one S-box affected in each round, i.e., an input change of one bit leads to an output change of one bit in all rounds. This is illustrated by the highlighted lines in Fig. 2(a). Since such a characteristic involves the fewest number of S-boxes possible, it is clear that the probability of a four-round characteristic is bounded by

$$p_{\Omega_4} \le \left(\frac{1}{4}\right)^4 = 2^{-8}. \tag{5}$$

Assume now, instead, that all S-boxes are such that a one-bit input change must cause at least two output bits to change and that the permutation used in the network is such that no two outputs of an S-box are connected to one S-box in the next round. The

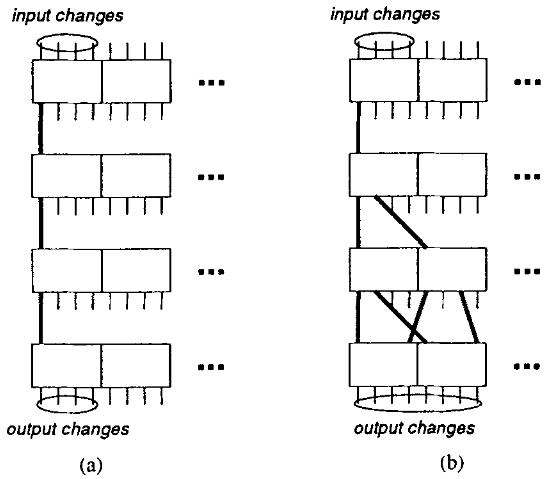

Fig. 2. High probability characteristics.

four-round characteristic which affects the fewest number of S-boxes is similar to that shown in Fig. 2(b). Assuming $p_\delta = \frac{1}{4}$, we now find that the characteristic probability is bounded by

$$p_{\Omega_4} \leq \left(\frac{1}{4}\right)^6 = 2^{-12}, \qquad (6)$$

which is a significantly smaller characteristic probability than the previous case.

(b) *Linear Cryptanalysis*

In [18] Matsui presents an effective linear cryptanalysis method for DES. The attack uses a known plaintext technique to extract key information by finding a linear equation consisting of plaintext, ciphertext, and key terms which is statistically likely to be satisfied. The full 16-round DES algorithm is susceptible to the attack with 2^{47} known plaintexts and it is shown that the attack can even be modified to be successful on an eight-round version of DES with 2^{29} encrypted ASCII-coded English blocks using a ciphertext-only attack. In order to attack an SPN using the linear cryptanalysis technique, the cryptanalyst is interested in the best R-round linear approximation of the form

$$P_{i_1} \oplus \cdots \oplus P_{i_\gamma} \oplus C_{j_1} \oplus \cdots \oplus C_{j_\zeta} = K_{k_1} \oplus \cdots \oplus K_{k_\eta}. \qquad (7)$$

If we let p_L represent the probability that (7) is satisfied, in order for the linear approximation to be valid $p_L \neq \frac{1}{2}$ and the best expression is the equation for which $|p_L - \frac{1}{2}|$ is maximized. If the magnitude $|p_L - \frac{1}{2}|$ is large enough and sufficient plaintext–ciphertext pairs are available, the equivalent of one key bit, expressed by the XOR sum of the key bits on the right-hand side of (7), may be guessed as the value that most often satisfies the linear approximation.

A basic linear attack, presented as Algorithm 1 in [18], may be executed using an algorithm based on a maximum likelihood approach. If $p_L > \frac{1}{2}$, then the sum of the key bits is assumed to be 0 if the left-hand side of (7) equals 0 for more than half the known plaintext–ciphertext pairs tested, or the sum of the key bits is assumed to be 1 if the left-hand side equals 1 for more than half the pairs. If $p_L < \frac{1}{2}$, then the sum of the key bits is assumed to be 1 if the left-hand side of (7) equals 0 for more than half the known plaintext–ciphertext pairs tested, or the sum of the key bits is assumed to be 0 if the left-hand side equals 1 for more than half the pairs.

An appropriate linear expression is derived by combining a number of linear expressions for different rounds such that any intermediate terms (i.e., terms that are not plaintext, ciphertext, or key terms) are cancelled. Let the best linear approximation of an S-box, in the form $a_1X_1 \oplus \cdots \oplus a_nX_n = b_1Y_1 \oplus \cdots \oplus b_nY_n$, be satisfied with probability p_ε assuming input $\mathbf{X}$ is randomly selected. In this paper we consider the probability that a system linear expression is satisfied to be taken over the independent distributions of plaintext and key. Hence, since the key bits XORed to the network bits prior to entering the S-boxes are independent and uniformly random, the inputs to the S-boxes involved in the linear approximation are independent and uniformly random. Under this assumption, it then follows from Lemma 3 in [18] that

$$\left|p_L - \tfrac{1}{2}\right| \le 2^{\alpha-1} \left|p_\varepsilon - \tfrac{1}{2}\right|^\alpha , \qquad (8)$$

where α is the number of S-box linear approximations combined to give the overall linear approximation.

In Lemma 2 of [18] Matsui develops an expression for the number of plaintexts required by the basic linear attack (Algorithm 1 in [18]). From this it is shown that the number of known plaintexts required to give a 97.7% confidence in the correct key bit may be approximated by N_L where

$$N_L = \left|p_L - \tfrac{1}{2}\right|^{-2} . \qquad (9)$$

It is obvious that N_L can be increased by decreasing $|p_L - \frac{1}{2}|$. Hence, selecting S-boxes for which $p_\varepsilon \to \frac{1}{2}$ will clearly aid in thwarting the attack. As well, the larger the number of S-boxes, α, involved in the system equation, the smaller $|p_L - \frac{1}{2}|$ and the more known plaintexts required for the cryptanalysis.

4. S-box Design Criteria

In this section we consider S-box design criteria that are relevant to the two attacks and examine the procedures that may be followed to generate S-boxes that satisfy such design constraints.

(a) *Diffusion*

As suggested in the previous section, S-boxes that effectively diffuse bit changes increase resistance to differential cryptanalysis. The diffusion properties of an S-box can be considered by examining the relationship between input and output XORs. Let $wt(\cdot)$

represent the Hamming weight of the specified argument and consider the following definition.

Definition 1. An S-box satisfies a *diffusion order of* λ, $\lambda \geq 0$, if, for $wt(\Delta\mathbf{X}) > 0$,

$$wt(\Delta\mathbf{Y}) > \begin{cases} \lambda + 1 - wt(\Delta\mathbf{X}), & wt(\Delta\mathbf{X}) < \lambda + 1, \\ 0, & \text{otherwise}. \end{cases} \tag{10}$$

Note that all bijective S-boxes satisfy $\lambda = 0$ and that DES S-boxes satisfy $\lambda = 1$ [9]. As well, the diffusion order is bidirectional, i.e., the inverse S-box S^{-1} satisfies the same diffusion order as S-box S.

Other properties related to the diffusiveness of an S-box are the strict avalanche criterion (SAC) [31] and the propagation criterion [27] (also referred to as higher-order SAC [1]). An S-box satisfies SAC if, given that a single input bit is complemented, the probability that each output bit changes is exactly $\frac{1}{2}$. Similarly, an S-box satisfies the propagation criterion order k if each output bit changes with a probability of $\frac{1}{2}$ when k or less input bits are complemented. The SAC and propagation criterion properties of an S-box imply that the expected number of output changes will not be small (i.e., on average half the output bits will change) even if the number of input changes is small. However, unlike the diffusion order of an S-box, SAC and the propagation criterion cannot be used to guarantee a lower bound on the number of output changes given a small number of input changes. As will be seen in Theorem 1, it is this guaranteed lower bound on the number of output changes defined by the diffusion order which is useful in ensuring low probability differential characteristics.

Let Π represent the set of permutations for which no two outputs of an S-box are connected to one S-box in the next round. Note that the set Π will only be nonempty if $M \geq n$.

Lemma 1. *Let ψ_{r-1} and ψ_{r+1} represent the number of S-boxes included in a characteristic from round $r-1$ and round $r+1$, respectively. For an SPN with $M \geq n$ S-boxes in each round, using a permutation $\pi \in \Pi$ and S-boxes with a diffusion order of λ,*

$$\psi_{r-1} + \psi_{r+1} \geq \lambda + 2. \tag{11}$$

Proof. Let w_X and w_Y represent the number of input and output bit changes for a particular S-box in round r selected such that $w_X \neq 0$. From the constraint placed on the permutations of Π and considering that $M \geq n$ and $w_X, w_Y \leq n$, we see that $\psi_{r-1} \geq w_X$ and $\psi_{r+1} \geq w_Y$. Hence,

$$\psi_{r-1} + \psi_{r+1} \geq w_X + w_Y. \tag{12}$$

From the definition of diffusion order, $w_X + w_Y \geq \lambda + 2$ and the inequality of (11) follows. □

Theorem 1. *Consider an SPN of R rounds of M S-boxes such that R is a multiple of 4 and $M \geq n$. Using a permutation $\pi \in \Pi$, the probability of an $(R-1)$-round characteristic satisfies*

$$p_{\Omega_{R-1}} \leq (p_\delta)^{((\lambda+2)/2)R-(\lambda+1)}, \tag{13}$$

where all S-boxes satisfy diffusion order λ and p_δ represents the maximum S-box XOR pair probability.

Proof. An upper bound on the most probable $(R-1)$-round characteristic can be derived by considering the concatenation of the most probable $(R-4)$-round characteristic and the most probable three-round characteristic. Further, a bound on the most likely $(R-4)$-round characteristic can be determined as $(R-4)/4$ iterations of the most probable four-round characteristic, and hence, the $(R-1)$-round characteristic probability satisfies

$$p_{\Omega_{R-1}} \leq (p_{\Omega_4}^{\max})^{(R-4)/4}(p_{\Omega_3}^{\max}), \tag{14}$$

where $p_{\Omega_3}^{\max}$ and $p_{\Omega_4}^{\max}$ are upper bounds on the probability of three- and four-round characteristics, respectively.

In general, an upper bound on a characteristic probability can be derived by determining the characteristic which involves the fewest number of S-boxes. From Lemma 1, the minimum number of S-boxes used by a characteristic in any four consecutive rounds is $2(\lambda+2)$ and therefore

$$p_{\Omega_4}^{\max} = (p_\delta)^{2(\lambda+2)}. \tag{15}$$

As well, by considering that the constraint of Lemma 1 applies to the first and third rounds of a three-round characteristic and that the second round has only one S-box, the minimum number of S-boxes used by a characteristic in any three consecutive rounds is $\lambda+3$. Therefore,

$$p_{\Omega_3}^{\max} = (p_\delta)^{\lambda+3}. \tag{16}$$

Combining (14), (15), and (16) results in (13) and the theorem is proven. □

From Theorem 1 we see that S-boxes satisfying a high diffusion order can be used to decrease the upper bound on characteristic probabilities and thereby strengthen a network against differential cryptanalysis. One obvious approach to generate such S-boxes would be to select randomly an $n \times n$ bijective mapping and discard those which do not satisfy the appropriate property. Unfortunately, we have found experimentally that S-boxes which satisfy diffusion orders of $\lambda \geq 1$ are extremely rare and cannot generally be found by random search. The following lemma is useful in determining the likelihood of finding such S-boxes.

Lemma 2. *Assume that the event that an S-box XOR pair $(\Delta\mathbf{X}, \Delta\mathbf{Y})$ violates diffusion order $\lambda = 1$ is independent of other XOR pairs violating $\lambda = 1$. Then the probability that a randomly selected $n \times n$ bijective S-box satisfies diffusion order $\lambda = 1$ is given by*

$$P(\lambda=1) = \left[\frac{2^n-1-n}{2^n-1}\right]^{n \cdot 2^{n-1}}. \tag{17}$$

Proof. Since the assignment of any two output values and their corresponding $\Delta\mathbf{Y}$ is random, the probability that the XOR pair $(\Delta\mathbf{X}, \Delta\mathbf{Y})$ satisfies $\lambda = 1$ given $wt(\Delta\mathbf{X}) = 1$ is simply

$$\begin{aligned} P(wt(\Delta\mathbf{Y}) > 1 \mid wt(\Delta\mathbf{X}) = 1) &= \frac{\#\{\Delta\mathbf{Y} \mid wt(\Delta\mathbf{Y}) > 1\}}{\#\{\Delta\mathbf{Y} \mid wt(\Delta\mathbf{Y}) \neq 0\}} \\ &= \frac{2^n - 1 - n}{2^n - 1}. \end{aligned} \tag{18}$$

Equation (17) follows by utilizing the independence assumption with the exponent determined by considering the number of unique input pairs for which $wt(\Delta\mathbf{X}) = 1$. Letting $\mathbf{X}'$ and $\mathbf{X}''$ represent the S-box inputs such that $\Delta\mathbf{X} = \mathbf{X}' \oplus \mathbf{X}''$, it may be seen that

$$\#\{(\mathbf{X}', \mathbf{X}'') \mid wt(\Delta\mathbf{X} = \mathbf{X}' \oplus \mathbf{X}'') = 1\} = n \cdot 2^{n-1} \tag{19}$$

and the lemma is proven. □

We have found experimentally that (17) is a good approximation of the probability that an S-box satisfies $\lambda = 1$. Table 1 lists the estimated probability from (17) that an S-box satisfies $\lambda = 1$ for a number of values of n, as well as the experimental value determined as described below. It is clear that, as n increases, S-boxes which satisfy $\lambda = 1$ become increasingly impractical to find by random search.

Consider S-boxes for which $\lambda > 1$. An S-box which satisfies diffusion order λ must, by definition, satisfy diffusion order $\lambda - 1$. Hence, the probability that a randomly selected S-box satisfies $\lambda > 1$ is less than or equal to the probability that the S-box satisfies $\lambda = 1$ and we conclude that as n increases such S-boxes are also impractical to find by random search.

In Fig. 3 we present an algorithm to select the S-box output values using a depth-first-search approach as an efficient method of generating S-boxes that satisfy a particular diffusion order. In the algorithm of Fig. 3 we use the variables i and $S(i)$ to represent, in decimal form, the S-box input and corresponding output, respectively. As well, $rand(\cdot)$ represents the random selection of an element from the specified set.

Considering the algorithm of Fig. 3 and letting $P(I_i \mid I_1 I_2 \cdots I_{i-1})$ represent the probability of iteration i being successful given iterations 1 to $i - 1$ are successful, the

Table 1. Probability of randomly selecting an S-box with $\lambda = 1$.

n	Estimated $P(\lambda = 1)$	Experimental $P(\lambda = 1)$
3	1.2×10^{-3}	3.6×10^{-3}*
4	4.9×10^{-5}	3.8×10^{-5}
5	7.7×10^{-7}	5.2×10^{-7}
6	4.5×10^{-9}	2.5×10^{-9}
7	9.3×10^{-12}	9.2×10^{-12}
8	6.7×10^{-15}	4.9×10^{-15}

*Actual value is $144/8! \approx 3.6 \times 10^{-3}$.

```
$\Gamma = \{0, 1, 2, \ldots, 2^n - 1\}$
$\Lambda_0 = \Gamma$
$i = 0$
do
  if $(\Lambda_i \neq \{\emptyset\})$ then
    $S(i) = rand(\Lambda_i)$
    $\Lambda_i = \Lambda_i - \{S(i)\}$
    if $((i, S(i))$ satisfy $\lambda)$ then
      $\Gamma = \Gamma - \{S(i)\}$
      $i = i + 1$
      $\Lambda_i = \Gamma$
    endif
  else
    $i = i - 1$
    $\Gamma = \Gamma + \{S(i)\}$
  endif
while $(i \leq 2^n - 1)$
output: $(i, S(i))$ for $0 \leq i \leq 2^n - 1$
end
```

Fig. 3. Algorithm to find S-boxes satisfying diffusion order λ.

probability of a randomly selected S-box satisfying $\lambda = 1$ can be determined using the chain rule:

$$P(\lambda = 1) = \prod_{i=1}^{2^n-1} P(I_i \mid I_1 I_2 \cdots I_{i-1}). \tag{20}$$

Utilizing experimental values of $P(I_i \mid I_1 I_2 \cdots I_{i-1})$ determined from executions of the algorithm, it is possible therefore to derive an experimental estimate using (20). The resulting experimental probabilities for different n are listed in Table 1.

There are limitations to the applicability of the depth-first-search algorithm. For example, while the algorithm successfully found many 8 × 8 S-boxes which satisfied diffusion orders of $\lambda = 1$ and $\lambda = 2$, it could not successfully find S-boxes with $\lambda \geq 3$. In the next section we show that, although the algorithm is designed to find S-boxes that satisfy a particular diffusion order, it is also valuable in generating S-boxes which are cryptographically strong in other respects.

(b) *Nonlinearity*

An important cryptographic property for product ciphers is nonlinearity. Since the S-boxes are the only nonlinear components of an SPN, it is crucial to consider the amount of nonlinearity required in S-boxes to provide adequate overall SPN security. The linear cryptanalysis method of Matsui [18] is one basis for determining the amount of nonlinearity required in an S-box.

Consider an SPN in which the lowest nonlinearity of an S-box is $NL_{\min}$, i.e., $NL(S) \geq NL_{\min}$ for all S-boxes. Then the best linear approximation of an S-box occurs with

Table 2. Nonlinearities of 8 × 8 S-boxes.

λ	Min NL	Max NL	$\%NL = 94$	$\%NL = 96$	$\%NL = 98$
0	86	98*	38.5	23.5	5.5
1	86	96	48	26	0
2	36	96	34	2	0

*S-boxes with $NL(S) = 100$ have been found using a more thorough search.

probability p_ε where

$$\left|p_\varepsilon - \tfrac{1}{2}\right| = \frac{2^{n-1} - NL_{\min}}{2^n}. \tag{21}$$

Since there must be at least one S-box approximation included in the linear expression of (7) for each round, the best possible linear approximation has $\alpha = R$ and satisfies

$$\left|p_L - \tfrac{1}{2}\right| \leq 2^{R-1}\left|p_\varepsilon - \tfrac{1}{2}\right|^R \leq 2^{R-1}\left[\frac{2^{n-1} - NL_{\min}}{2}\right]^R. \tag{22}$$

It is known that there are $n \times n$ bijective mappings for which $NL(S) \geq 2^{n-1} - 2^{n/2}$ [23]. Assuming that S-boxes are used that have $NL(S) = 2^{n-1} - 2^{n/2}$, combining (9) and (22) we see that the number of known plaintexts required to determine one bit of key is at least $2^{nR-2(R-1)}$. For example, if an eight-round SPN was constructed using 8 × 8 S-boxes with $NL(S) = 112$, it would take about 2^{50} known plaintexts to determine one key bit.

In [24] O'Connor shows that, as n gets larger, the expected distance of a randomly selected n-bit function (not necessarily balanced) from the nearest affine function increases and p_ε approaches the ideal value of $\frac{1}{2}$. In view of this, we expect that, as n gets large, S-boxes with high nonlinearities will be plentiful and easy to find by random search.

In order to confirm this intuition, 200 8 × 8 bijective S-boxes (i.e., $\lambda = 0$) were randomly generated and their nonlinearities examined. As well, 50 S-boxes were constructed using the depth-first-search algorithm for the diffusion orders of $\lambda = 1$ and $\lambda = 2$. The results are given in Table 2. We surmise that, as the diffusion characteristics become more constraining, the S-box nonlinearities are adversely affected. However, for $\lambda = 0$, 1, or 2, it is still reasonable to expect to find S-boxes with high nonlinearities of 94 or 96.

5. Linear Transformations Between Rounds

The permutations of an SPN belong to a specialized class of the set of linear transformations that may be used to achieve Shannon's diffusion effect. In this section we consider another class of invertible linear transformations that may be used between rounds of S-boxes to increase the resistance to differential and linear cryptanalysis.

Let N be even and consider the class of invertible linear transformations defined by

$$\mathbf{V} = \pi(\mathcal{L}(\mathbf{U})), \tag{23}$$

where $\mathbf{V} = [V_1 \ V_2 \ \cdots \ V_N]$ is the vector of input bits to a round of S-boxes, $\mathbf{U} = [U_1 \ U_2 \ \cdots \ U_N]$ is the vector of bits from the previous round output, $\pi \in \Pi$, and $\mathcal{L}(\mathbf{U}) = [L_1(\mathbf{U}) \ \cdots \ L_N(\mathbf{U})]$. The set Π is defined to be the set of permutations for which no two outputs of an S-box are connected to one S-box in the next round and

$$L_i(\mathbf{U}) = U_1 \oplus \cdots \oplus U_{i-1} \oplus U_{i+1} \oplus \cdots \oplus U_N. \tag{24}$$

The linear transformation may be efficiently implemented by noting that each $L_i(\mathbf{U})$ can be simply determined by XORing U_i with the XOR sum of all U_j, $1 \leq j \leq N$, i.e.,

$$L_i(\mathbf{U}) = U_i \oplus Q, \tag{25}$$

where

$$Q = \bigoplus_{j=1}^{N} U_j. \tag{26}$$

The following lemma illustrates the effect of the linear transformation on the diffusion of bit changes within the network.

Lemma 3. *Let $W = \mathcal{L}(\mathbf{U})$ where $\mathcal{L}(\cdot)$ is defined above and $\mathbf{W} = [W_1 \ W_2 \ \cdots \ W_N]$. Let $\Delta\mathbf{U} = [\Delta U_1 \ \cdots \ \Delta U_N]$ be the XOR difference between two arbitrary values of $\mathbf{U}$, and $\Delta\mathbf{W} = [\Delta W_1 \ \Delta W_2 \ \cdots \ W_N]$ is the resulting XOR difference for $\mathbf{W}$. Then*

$$\Delta\mathbf{W} = \begin{cases} \Delta\mathbf{U}, & wt(\Delta\mathbf{U}) \text{ even}, \\ \overline{(\Delta\mathbf{U})}, & wt(\Delta\mathbf{U}) \text{ odd}, \end{cases} \tag{27}$$

where $\overline{(\Delta\mathbf{U})}$ is the complement of $\Delta\mathbf{U}$.

Proof. Let $\Delta\mathbf{U} = \mathbf{U}' \oplus \mathbf{U}''$ and $\Delta U_i = U_i' \oplus U_i''$. Therefore,

$$\begin{aligned} \Delta W_i &= L_i(\mathbf{U}') \oplus L_i(\mathbf{U}'') \\ &= [U_i' \oplus Q'] \oplus [U_i'' \oplus Q''] \\ &= U_i' \oplus U_i'' \oplus Q' \oplus Q'' \\ &= \Delta U_i \oplus \Delta Q, \end{aligned} \tag{28}$$

where

$$\Delta Q = \bigoplus_{j=1}^{N} \Delta U_j. \tag{29}$$

If $wt(\Delta\mathbf{U})$ is even, then $\Delta Q = 0$ and

$$\Delta W_i = \begin{cases} 1, & \Delta U_i = 1, \\ 0, & \Delta U_i = 0. \end{cases} \tag{30}$$

If $wt(\Delta\mathbf{U})$ is odd, then $\Delta Q = 1$ and

$$\Delta W_i = \begin{cases} 1, & \Delta U_i = 0, \\ 0, & \Delta U_i = 1. \end{cases} \tag{31}$$

Equation (27) follows and the lemma is proven. □

Lemma 3 is useful in developing the following result.

Theorem 2. *Consider an SPN of R rounds of M S-boxes such that R is a multiple of* 4 *and $M \geq n$. Let $n \geq 3$ and let each S-box satisfy diffusion order λ such that $\lambda \leq (n-1)/2$. Using the linear transformation of* (23), *the probability of an $(R-1)$-round characteristic satisfies*

$$p_{\Omega_{R-1}} \leq (p_\delta)^{((\lambda+2)/2)R-(\lambda+1)}, \tag{32}$$

where p_δ represents the maximum S-box XOR pair probability. Further, for $\lambda = 0$, the characteristic probability can be more tightly bounded by

$$p_{\Omega_{R-1}} \leq (p_\delta)^{(3/2)R-2}. \tag{33}$$

Proof. Consider separately the case for general λ and the case for $\lambda = 0$.

(i) (General λ) As in the proof of Theorem 1, consider determining the upper bound on the most probable $(R-1)$-round characteristic from the concatenation of $(R-4)/4$ iterations of the most probable four-round characteristic with the most probable three-round characteristic. Hence, a characteristic probability satisfies (14).

Consider first the determination of the most likely four-round characteristic in order to determine $p_{\Omega_4}^{\max}$ of (14). Let ψ_i, $1 \leq \psi_i \leq M$, represent the number of S-boxes from round i involved in the characteristic and let η_i represent the number of bit changes after the substitutions of round i and before the linear transformation $\pi(\mathcal{L}(\cdot))$. Consider two cases for the values of η_i for four consecutive rounds r to $r+3$.

In the first case assume that at least one of η_r, η_{r+1}, or η_{r+2} are odd. Without loss of generality assume that η_r is odd. If we let $\psi_r < M$, this implies that $\psi_{r+1} \geq n$, since, from Lemma 3, the n bits from an S-box in round r with no output changes must result in n bit changes after $\mathcal{L}(\cdot)$, which, due to the nature of the permutation π, must then affect n different S-boxes in round $r+1$. Further, since, in general, $\psi_i \geq 1$, then

$$\psi_r + \psi_{r+1} + \psi_{r+2} + \psi_{r+3} \geq n + 3. \tag{34}$$

Since $M \geq n$, (34) also holds if $\psi_r = M$.

Now consider the second case where all of η_r, η_{r+1}, and η_{r+2} are even. From Lemma 3 and the definition of the permutation π, it may be seen that Lemma 1 may be applied as in the proof of Theorem 1 and, therefore,

$$\psi_r + \psi_{r+1} + \psi_{r+2} + \psi_{r+3} \geq 2(\lambda + 2). \tag{35}$$

Since $\lambda \leq (n-1)/2$, (35) always holds and

$$p_{\Omega_4}^{\max} = (p_\delta)^{2(\lambda+2)}. \tag{36}$$

A bound on the probability of a three-round characteristic may be determined similarly to the four-round characteristic. In this case if at least one of η_r or η_{r+1} is odd, then

$$\psi_r + \psi_{r+1} + \psi_{r+2} \geq n + 2. \tag{37}$$

If both η_r and η_{r+1} are even, then

$$\psi_r + \psi_{r+1} + \psi_{r+2} \geq \lambda + 3. \tag{38}$$

Hence, since $\lambda \leq (n-1)/2$, (38) always holds and

$$p_{\Omega_3}^{\max} = (p_\delta)^{\lambda+3}. \tag{39}$$

From (14) in the proof of Theorem 1, we can now see that, for general λ, (32) holds.

(ii) ($\lambda = 0$) From (35), we have

$$\psi_r + \psi_{r+1} + \psi_{r+2} + \psi_{r+3} \geq 4. \tag{40}$$

However, for the case where η_r, η_{r+1}, and η_{r+2} are all even, $\psi_i \neq 1$ for any two consecutive rounds since the permutation π spreads the effect of more than one output change to more than one S-box. However, if $\psi_r = 1$ and $\eta_r = 2$, then $\psi_{r+1} = 2$. Hence, $\psi_r + \psi_{r+1} \geq 3$ and, consequently,

$$\psi_r + \psi_{r+1} + \psi_{r+2} + \psi_{r+3} \geq 6. \tag{41}$$

From (34) we can see that (41) also holds for the case where one or more of η_r, η_{r+1}, and η_{r+2} is odd, as long as $n \geq 3$.

Similarly, it may be shown that

$$\psi_r + \psi_{r+1} + \psi_{r+2} \geq 4 \tag{42}$$

and, applying (14), we have now proven the case for $\lambda = 0$. □

Note that for $\lambda = 0$ the linear transformation has decreased the upper bound on the characteristic probability and for $\lambda > 0$ the bound on the characteristic probability has remained unchanged.

Consider now the effects of the linear transformation on the applicability of linear cryptanalysis. Using the linear transformation ensures that there are a large number of S-box approximations included in the system linear approximation, thereby increasing the number of required plaintexts.

Theorem 3. *Consider an SPN of R rounds of M S-boxes such that R is even and $M \geq n$. Using the linear transformation of* (23), *the best possible R-round linear approximation requires $\alpha = 3R/2$ S-box approximations and the probability of the linear approximation satisfies*

$$\left|p_L - \tfrac{1}{2}\right| \leq 2^{(3/2)R-1}\left|p_\varepsilon - \tfrac{1}{2}\right|^{(3/2)R}, \tag{43}$$

where p_ε represents the probability of the best S-box linear approximation.

Proof. Using the linear transformation of (23), it is impossible to involve only one S-box per round in the linear approximation. Let the number of S-boxes from round i involved in the overall system linear approximation be represented by ψ_i. Consider round r to contribute only one S-box to the linear approximation, i.e., $\psi_r = 1$. The linear approximation of this S-box involves a linear combination of the input bits, $a_1X_1 \oplus a_2X_2 \oplus \cdots \oplus a_nX_n$, where $\mathbf{a} = [a_1 \ \cdots \ a_n]$, $a_i \in \{0, 1\}$, and a linear combination of the output bits, $b_1Y_1 \oplus b_2Y_2 \oplus \cdots \oplus b_nY_n$, where $\mathbf{b} = [b_1 \ \cdots \ b_n]$, $b_i \in \{0, 1\}$, so that the probability of

$$\bigoplus_{i=1}^{n} a_i X_i = \bigoplus_{i=1}^{n} b_i Y_i \tag{44}$$

does not equal $\frac{1}{2}$. (Note that the trivial case of $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = \mathbf{0}$ is of no use in linear cryptanalysis and is ignored.)

Without loss of generality, assume that the S-box included in the system linear approximation from round r is the first S-box so that

$$\mathbf{X} = [X_1 \quad X_2 \quad \cdots \quad X_n] = [V_1 \quad V_2 \quad \cdots \quad V_n]$$

where V_i is the ith input bit to round r. The input to round r is determined by the permutation π so that $V_i = L_{j_i}(\mathbf{U})$ where $\mathbf{U}$ is the vector of output bits from the S-boxes of round $r - 1$. Subsequently, we have $X_i = U_{j_i} \oplus Q$ where Q is defined in (26) and each U_{j_i}, $1 \le i \le n$, comes from a different S-box (as a result of the definition of the permutation π). We now have

$$\begin{aligned} \bigoplus_{i=1}^{n} a_i X_i &= \bigoplus_{i=1}^{n} a_i \cdot (U_{j_i} \oplus Q) \\ &= \bigoplus_{i=1}^{n} a_i U_{j_i} \oplus \bigoplus_{i=1}^{n} a_i Q \\ &= \begin{cases} \displaystyle\bigoplus_{i=1}^{n} a_i U_{j_i} \oplus Q, & wt(\mathbf{a})\text{ odd}, \\ \displaystyle\bigoplus_{i=1}^{n} a_i U_{j_i}, & wt(\mathbf{a})\text{ even}. \end{cases} \end{aligned} \tag{45}$$

Hence, if $wt(\mathbf{a})$ is odd, then the sum used for the input of the round r S-box is determined by $N - wt(\mathbf{a})$ outputs of round $r - 1$ since a term is removed from Q when $a_i = 1$. If $wt(\mathbf{a})$ is even, then the sum used for the input of the round r S-box is determined by $wt(\mathbf{a})$ outputs of round $r - 1$ since a term is only included in the summation when $a_i = 1$. If, for example, $wt(\mathbf{a}) = 1$, then the corresponding S-box input bit used in the linear approximation is a function of $N - 1$ output bits from round $r - 1$ and, hence, $\psi_{r-1} = M$. If, however, $wt(\mathbf{a}) = 2$, then $\psi_{r-1} = 2$. Hence, considering other values for $wt(\mathbf{a})$, $1 \le wt(\mathbf{a}) \le n$, we may now conclude that, given $\psi_r = 1$, $\psi_{r-1} \ge 2$.

A similar analysis may be used to determine a lower bound on the number of S-boxes included in the linear approximation from round $r + 1$, ψ_{r+1}, given $\psi_r = 1$. This is possible due to the following easily verifiable observations: $\mathcal{L}^{-1} \equiv \mathcal{L}$, $\pi^{-1} \in \Pi$, and

$\mathcal{L}(\pi(\cdot)) \equiv \pi(\mathcal{L}(\cdot))$. Hence, we have

$$\mathbf{U}^* = \pi^{-1}(\mathcal{L}(\mathbf{V}^*)), \tag{46}$$

where $\mathbf{U}^*$ is the vector of output bits of the round r substitutions and $\mathbf{V}^*$ is the vector of input bits to the round $r+1$ substitutions. Since (46) is of a similar form to (23), we may determine the bound for ψ_{r+1} analogously to the bound for ψ_{r-1}. Hence, it follows that $\psi_{r+1} \geq 2$ given $\psi_r = 1$.

We conclude, therefore, that the number of S-boxes involved in the linear approximation from any two consecutive rounds must be at least three and for an R-round SPN, assuming R is even, $\alpha \geq 3R/2$. □

Note that results similar to Lemma 3, Theorem 2, and Theorem 3 can be derived for $\mathcal{L}(\mathbf{U})$ defined as other invertible linear transformations where each $L_i(\mathbf{U})$ may contain fewer than the $N-1$ terms of (24).

6. Summary of Results

In Table 3, for SPNs of eight rounds, we have summarized lower bounds on the values of N_D and N_L (defined in (4) and (9), respectively). The networks are assumed to be composed of 8×8 S-boxes where the maximum S-box XOR pair probability is $p_\delta = 2^{-4}$ and the minimum S-box nonlinearity is $NL_{\text{min}} = 96$. Results are presented for networks using permutations from the set Π and for networks using a linear transformation of the form of (23). Note that the analysis of Table 3 is equally applicable to the decryption as well as the encryption network. (This is important since the decryption network may also be attacked using either cryptanalysis method.)

Consider a 64-bit eight-round SPN that uses a linear transformation of the form of (23) and 8×8 S-boxes with $\lambda = 2$, $p_\delta = 2^{-4}$, and $NL_{\text{min}} = 96$. Assume that the network is keyed using a 64-bit key with XOR mask keying. Application of the key bits at each round is determined by a key-scheduling algorithm. Such a network has high values of $N_D^{\text{min}} = 2^{52}$ and $N_L^{\text{min}} = 2^{50}$, is comparable in size with DES (64-bit blocks, 56-bit key), but is implemented in half the number of rounds.

Table 3. Resistance to cryptanalysis for networks with $R = 8$ using 8×8 S-boxes with $p_\delta = 2^{-4}$ and $NL_{\text{min}} = 96$.

Type	λ	N_D^{min}	N_L^{min}
Permutation $\pi(\cdot)$	0	2^{28}	2^{34}
	1	2^{40}	
	2	2^{52}	
Linear transform $\pi(\mathcal{L}(\cdot))$	0	2^{40}	2^{50}
	1	2^{40}	
	2	2^{52}	

7. Conclusion

In this paper we have developed bounds on the probabilities of a differential characteristic and a linear approximation for substitution-permutation networks. It is important to note that the bounds are of interest, not because they give a provable lower bound on the complexity of the cryptanalysis, but because they suggest the level of difficulty required in implementing the attacks. For example, in a differential attack, the cryptanalyst typically identifies a high probability input difference to the last round by searching for high probability differential characteristics. Similarly, for linear cryptanalysis, a good linear approximation can be practically used by a cryptanalyst to determine which subsets of plaintext and ciphertext bits to examine in the attack.

The analysis presented in this paper suggests the following general design principles for substitution-permutation networks:

- Large, randomly selected S-boxes are very likely to have high nonlinearity.
- S-boxes which have good diffusion properties increase the resistance to differential cryptanalysis.
- The use of an appropriate linear transformation between rounds increases the resistance to linear cryptanalysis.

Consequently, with an appropriate selection of S-boxes and linear transformations between rounds of substitutions, security in relation to differential and linear cryptanalysis can be improved, resulting in an efficient implementation with fewer rounds required to provide adequate security.

Acknowledgments

The authors appreciatively acknowledge the assistance of Stephen Morton who provided much of the software required in the researching of this paper. The authors are also grateful to the referees, whose insightful comments significantly improved the presentation of the paper.

References

[1] C. M. Adams. A Formal and Practical Design Procedure for Substitution-Permutation Network Cryptosystems. Ph.D. thesis, Queen's University, Kingston, Ontario, 1990.

[2] C. M. Adams. On immunity against Biham and Shamir's differential cryptanalysis. *Information Processing Letters*, 41(2):77–80, 1992.

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

[4] F. Ayoub. The design of complete encryption networks using cryptographically equivalent permutations. *Computers and Security*, 2:261–267, 1982.

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

[6] E. Biham and A. Shamir. Differential cryptanalysis of FEAL and N-Hash. *Advances in Cryptology: Proceedings of EUROCRYPT '91*, Springer-Verlag, Berlin, pages 1–16, 1991.

[7] E. Biham and A. Shamir. Differential cryptanalysis of Snefru, Khafre, REDOC-II, LOKI, and Lucifer. *Advances in Cryptology: Proceedings of CRYPTO '91*, Springer-Verlag, Berlin, pages 156–171, 1992.

[8] E. Biham and A. Shamir. Differential cryptanalysis of the full 16-round DES. *Advances in Cryptology: Proceedings of CRYPTO '92*, Springer-Verlag, Berlin, pages 487–496, 1993.

[9] E. F. Brickell, J. H. Moore, and M. R. Purtill. Structures in the S-boxes of DES. *Advances in Cryptology: Proceedings of CRYPTO '86*, Springer-Verlag, Berlin, pages 3–8, 1987.

[10] L. Brown, J. Pieprzyk, and J. Seberry. LOKI—a cryptographic primitive for authentication and secrecy applications. *Advances in Cryptology: Proceedings of AUSCRYPT '90*, Springer-Verlag, Berlin, pages 229–236, 1990.

[11] L. Brown and J. R. Seberry. On the design of permutation P in DES type cryptosystems. *Advances in Cryptology: Proceedings of EUROCRYPT '89*, Springer-Verlag, Berlin, pages 696–705, 1989.

[12] M. H. Dawson and S. E. Tavares. An expanded set of S-box design criteria based on information theory and its relation to differential-like attacks. *Advances in Cryptology: Proceedings of EUROCRYPT '91*, Springer-Verlag, Berlin, pages 352–367, 1991.

[13] H. Feistel. Cryptography and computer privacy. *Scientific American*, 228(5):15–23, 1973.

[14] H. Feistel, W. A. Notz, and J. L. Smith. Some cryptographic techniques for machine-to-machine data communications. *Proceedings of the IEEE*, 63(11):1545–1554, 1975.

[15] R. Forré. Methods and instruments for designing S-boxes. *Journal of Cryptology*, 2(3):115–130, 1990.

[16] J. B. Kam and G. I. Davida. A structured design of substitution-permutation encryption networks. *IEEE Transactions on Computers*, 28(10):747–753, 1979.

[17] L. R. Knudsen. Iterative characteristics of DES and s^2-DES. *Advances in Cryptology: Proceedings of CRYPTO '92*, Springer-Verlag, Berlin, pages 497–511, 1993.

[18] M. Matsui. Linear cryptanalysis method for DES cipher. *Advances in Cryptology: Proceedings of EUROCRYPT '93*, Springer-Verlag, Berlin, pages 386–397, 1994.

[19] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. *Advances in Cryptology: Proceedings of EUROCRYPT '89*, Springer-Verlag, Berlin, pages 549–562, 1990.

[20] National Bureau of Standards. *Data Encryption Standard (DES)*. Federal Information Processing Standard Publication 46, U.S. Department of Commerce, January 1977.

[21] K. Nyberg. Perfect nonlinear S-boxes. *Advances in Cryptology: Proceedings of EUROCRYPT '91*, Springer-Verlag, Berlin, pages 378–386, 1991.

[22] K. Nyberg. On the construction of highly nonlinear permutations. *Advances in Cryptology: Proceedings of EUROCRYPT '92*, Springer-Verlag, Berlin, pages 92–98, 1992.

[23] K. Nyberg. Differentially uniform mappings for cryptography. *Advances in Cryptology: Proceedings of EUROCRYPT '93*, Springer-Verlag, Berlin, pages 55–64, 1994.

[24] L. O'Conñor. An Analysis of Product Ciphers Based on the Properties of Boolean Functions. Ph.D. thesis, University of Waterloo, Ontario, 1992.

[25] L. J. O'Connor. On the distribution of characteristics in bijective mappings. *Advances in Cryptology: Proceedings of EUROCRYPT '93*, Springer-Verlag, Berlin, pages 360–370, 1994.

[26] J. Pieprzyk and G. Finkelstein. Towards effective nonlinear cryptosystem design. *IEE Proceedings, Part E*, 135(6):325–335, 1988.

[27] B. Preneel, W. Van Leekwijck, L. Van Linden, R. Goevarts, and J. Vanderwalle. Propagation characteristics of boolean functions. *Advances in Cryptology: Proceedings of EUROCRYPT '90*, Springer-Verlag, Berlin, pages 161–173, 1991.

[28] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.

[29] A. Shimizu and S. Miyaguchi. Fast data encipherment algorithm: FEAL. *Advances in Cryptology: Proceedings of EUROCRYPT '87*, Springer-Verlag, Berlin, pages 267–278, 1988.

[30] M. Sivabalan, S. E. Tavares, and L. E. Peppard. On the design of SP networks from an information-theoretic point of view. *Advances in Cryptology: Proceedings of CRYPTO '92*, Springer-Verlag, Berlin, pages 260–279, 1993.

[31] A. F. Webster and S. E. Tavares. On the design of S-boxes. *Advances in Cryptology: Proceedings of CRYPTO '85*, Springer-Verlag, Berlin, pages 523–534, 1986.