Avalanche Characteristics of Substitution-Permutation Encryption Networks

Howard M. Heys and Stafford E. Tavares, member IEEE

Abstract — This paper develops analytical models for the avalanche characteristics of a class of block ciphers usually referred to as substitution-permutation encryption networks or SPNs. An SPN is considered to display good avalanche characteristics if a one bit change in the plaintext input is expected to result in close to half the ciphertext output bits changing. Good avalanche characteristics are important to ensure that a cipher is not susceptible to statistical attacks and the strength of an SPN's avalanche characteristics may be considered as a measure of the randomness of the ciphertext. The results presented in this paper demonstrate that the avalanche behaviour of encryption networks can be improved by using larger S-boxes. As well, it is shown that increasing the diffusion properties of the S-boxes or replacing the permutations by diffusive linear transformations is effective in improving the network avalanche characteristics.

H. Heys is with Electrical Engineering, Memorial University of Newfoundland, St. John's, Newfoundland, Canada, A1B 3X5.

S. Tavares is with the Department of Electrical Engineering, Queen's University, Kingston, Ontario, Canada, K7L 3N6.

This work was done at Queen's University and supported by the Natural Sciences and Engineering Research Council of Canada and the Telecommunications Research Institute of Ontario.

List of Index Terms

(1) avalanche

(2) block ciphers

(3) cryptography

(4) S-boxes

(5) substitution-permutation encryption networks

List of Figure Captions

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

Figure 2. SPN utilizing Permutation Type II with $N = 27$, $R = 4$, and $n = 3$

Figure 3. Theoretical Avalanche of SPN with $N = 64$

Figure 4. Improving Avalanche for SPNs with $N = 64$ and $n = 4$

Figure 5. Improving Avalanche for SPNs with $N = 64$ and $n = 8$

I. Introduction

Since its introduction in 1977, the Data Encryption Standard (DES) [1] has become the most widely applied private key block cipher. Initial analysis of the DES algorithm suggested that the 56–bit key size was too small and that DES would eventually succumb to an exhaustive key search using specialized hardware [2]. Recently, a hardware design to effectively break DES using exhaustive key search was outlined by Wiener [3]. Unfortunately, it is not generally known how to efficiently modify the DES algorithm to allow for different block or key sizes. This suggests that there is a need to replace DES with an efficient, secure, flexible block cipher whose design is well understood. The results presented in this paper are a contribution to the achievement of this objective.

Avalanche is an important cryptographic property of private key block ciphers. We say that a cipher satisfies the avalanche criterion if changing a single plaintext bit is expected to result in one half of the ciphertext bits changing. Satisfaction of this criterion is necessary for the ciphertext to be random and, depending on the keying methodology, can assist in making a cipher resistant to certain statistical attacks such as key clustering attacks [4][5]. In this paper we develop an analytical model of the avalanche characteristics of an important class of block ciphers, referred to as substitution-permutation encryption networks (SPNs). We examine the tendency of a network to display good avalanche characteristics as the number of rounds is increased and investigate the effect of modifying various design constraints such as S-box size. The objective of the analysis is to determine architectures that will allow an efficient implementation with the fewest number of rounds necessary to achieve a suitable level of security.

II. Background

Feistel [6] was the first to suggest that an SPN architecture consisting of rounds of nonlinear substitutions (S-boxes) connected by bit position permutations was a simple, effective implementation of Shannon's principle of a "mixing transformation" based on the concepts of "confusion" and "diffusion" [7]. (Note that this basic SPN architecture differs from a DES-like architecture which also uses substitutions and permutations for a mixing transformation which operates on only half the block at a time.) It has been shown that this basic SPN structure can be used to construct ciphers which possess good cryptographic properties such as completeness [8] and, as shown in [9], resistance to differential cryptanalysis [10] and linear cryptanalysis [11]. Many modern block ciphers, including DES [1], FEAL [12], LOKI [13][14], and IDEA [15][16], while deviating from the basic SPN model, are based on Shannon's fundamental concepts. In this paper we consider the basic SPN model because, compared to other ciphers, it is a simple but elegant structure for which it is generally possible to prove security properties.

In general, we shall consider an N-bit SPN to be composed of R rounds of $n \times n$ S-boxes with each round consisting of $M = N/n$ S-boxes. We denote the plaintext input as $\mathbf{P} = [P_1\ P_2\ ...\ P_N]$, $P_i \in \{0,1\}$, and the ciphertext output as $\mathbf{C} = [C_1\ C_2\ ...\ C_N]$, $C_i \in \{0,1\}$. S-boxes in the network are defined as a bijective mapping $S\ :\ \mathbf{X} \rightarrow \mathbf{Y}$ where $\mathbf{X} = [X_1\ ...\ X_n]$, $X_i \in \{0,1\}$, and $\mathbf{Y} = [Y_1\ ...\ Y_n]$, $Y_i \in \{0,1\}$. The interconnection between consecutive rounds of S-boxes is typically achieved by a permutation of the bit positions from the output of a round to the input of the next round such that no two output bits of an S-box are connected to one S-box in the next round.A simple example of an SPN is illustrated in Figure 1 with $N = 16$, $R = 4$, and $n = 4$.

In general, the SPN cipher may be keyed by applying key bits to the S-boxes 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, and

(2) XOR keying: key bits are XORed with the network bits prior to entering an S-box.

Since, by definition, the avalanche characteristic of a cipher assumes that the cipher key is fixed, the method of keying the S-boxes does not affect the analysis in this paper. Hence, in this paper we do not address the issue of keying the cipher. However, the relationship between key avalanche and a key clustering attack is thoroughly discussed in [5].

When considering the avalanche characteristics of a block cipher, we are interested in the bit changes or XOR differences within the network when two plaintexts, $\mathbf{P}'$ and $\mathbf{P}''$, are selected as inputs such that $wt(\Delta\mathbf{P} = \mathbf{P}' \oplus \mathbf{P}'') = 1$ where $\Delta\mathbf{P} = \mathbf{P}' \oplus \mathbf{P}''$ represents the bit-wise XOR of $\mathbf{P}'$ and $\mathbf{P}''$ and $wt(\cdot)$ represents the Hamming weight of the specified vector. The ciphertext change resulting from $\mathbf{P}'$ and $\mathbf{P}''$ is represented by $\Delta\mathrm{C}$. We refer to an S-box input change as $\Delta\mathbf{X}$ and an S-box output change as $\Delta\mathbf{Y}$. The number of input and output bit changes of an S-box is given by $wt(\Delta\mathbf{X})$ and $wt(\Delta\mathbf{Y})$, respectively; the number of ciphertext bit changes is represented by $wt(\Delta\mathrm{C})$.

The concept of avalanche in SPNs was informally introduced by Feistel [6] and Feistel, Notz, and Smith [17] as the property of a small number of bit changes in plaintext leading to an “avalanche” of changes in subsequent rounds resulting in a large number of ciphertext bit changes. We shall define avalanche formally as follows:

Definition 1: A cipher is said to satisfy the *avalanche criterion* if, for each key, on average half of the ciphertext bits change when one plaintext bit is changed. That is, $\boldsymbol{E}\left(wt(\Delta\mathrm{C}) \mid wt(\Delta\mathbf{P}) = 1\right) = N/2$.

An extension to this definition was proposed by Webster and Tavares [18] and is referred to as the strict avalanche criterion (SAC). Boolean functions satisfying SAC and their relationships to cryptographic systems have been examined by several authors [19][20][21][22][23].

Definition 2: A cipher is said to satisfy the *strict avalanche criterion (SAC)* if, for each key, each ciphertext bit changes with a probability of 1/2 when a single plaintext bit is changed. That is, $P(\Delta C_t = 1 \mid \Delta \mathbf{P} = \mathbf{e}_i) = 1/2$ for $1 \leq t \leq N$ and $1 \leq i \leq N$ where $\mathbf{e}_i = [e_1 \; e_2 \; ... \; e_N]$ with $e_i = 1$ and $e_j = 0, \; j \neq i$.

It is apparent that the avalanche criterion and SAC are very similar. SAC imposes the extra conditions that a particular plaintext bit is changed and all ciphertext bits are equally likely to change given the one bit plaintext change. Therefore, although a network satisfying SAC must satisfy the avalanche criterion, satisfaction of the avalanche criterion does not necessarily imply satisfaction of SAC.

The avalanche criterion and SAC are of interest in the design of ciphers since the satisfaction of these criteria is a necessary condition for the randomness of the ciphertext. However, it should be noted that satisfaction of these criteria is not sufficient to ensure the security of the cipher. For example, the powerful cryptanalysis techniques of differential and linear cryptanalysis have been very effective on ciphers which have been shown to reasonably satisfy SAC. In particular, although DES has been found experimentally to reasonably satisfy SAC after a modest number of rounds [18], differential and linear cryptanalysis have been applied effectively on the full 16 round algorithm [24][11].

III. Modelling Avalanche

In this section, we develop models for the avalanche characteristics of SPNs. We shall consider two general network models, distinguished by the nature of their permutation component. The models are:

Model A — Stochastic Permutation

- a network where the permutation between any two rounds is modelled as a random variable whose values are equally likely

Model B — Deterministic Permutation

- a network which has a specified fixed permutation between rounds

In either case, the identity permutation is used before and after the last round of S-boxes.

For both network models, we assume any set of one or more input bit changes to an S-box results in the number of output bit changes represented by the random variable D, i.e., $D = wt(\Delta\mathbf{Y})$. We assume the likelihood of a particular non-zero value for D is given by considering that all possible values of $\Delta\mathbf{Y}$ belonging to the set of $2^n - 1$ non-zero changes are equally likely. Hence, the probability distribution of D is given by

$$P_D(D=0) = \begin{cases} 1 & , wt(\Delta\mathbf{X}) = 0 \\ 0 & , wt(\Delta\mathbf{X}) \geq 1 \end{cases} \tag{1}$$

and

$$P_D(D=d) = \begin{cases} 0 & , wt(\Delta\mathbf{X}) = 0 \\ \frac{\binom{n}{d}}{2^n-1} & , wt(\Delta\mathbf{X}) \geq 1 \end{cases} \tag{2}$$

for $1 \leq d \leq n$. Note that the S-box model essentially represents an average over all randomly selected S-boxes and is not intended to characterize the behaviour of an actual physically realizable S-box. However, as experimental evidence suggests, modelling the number of output changes of each S-box as a random variable appears to be a suitable approximation when considering an SPN constructed using randomly selected fixed S-boxes.

For ease of notation, throughout the paper we will represent the probability of a specific value z of random variable Z, $P(Z = z)$, as simply $P(z)$. Further, for random variable D representing the number of output bit changes of an S-box, we maintain the subscript D on the probability operator P in order ensure clarity. Hence, $P_D(D = d)$ is represented by $P_D(d)$.

(a) Model A — Stochastic Permutation

Model A is implemented by recursively calculating the probability distribution of the number of bit changes after each round of substitutions given a one bit plaintext change. The recursive probability is taken over the uniform distribution of permutations. Let W_r represent the random variable corresponding to the number of bit changes after round r given a one bit plaintext change, i.e.,

$$W_r = \sum_{s=1}^{M} wt(\Delta \mathbf{Y}_{rs}) \tag{3}$$

where we have included the subscripts r and s for $\Delta \mathbf{Y}$ to represent the output change of an S-box numbered s, $1 \leq s \leq M$ in round r, $1 \leq r \leq R$. We wish to determine the probability distribution $P(W_r = w_r)$ given $P(W_{r-1} = w_{r-1})$.

Let l_r be a value of the random variable L_r representing the number of S-boxes in round r affected by a change. Under the assumptions of the model, the number of output bit changes in round r is strictly dependent on the number of S-boxes in round r affected by an input bit change and, therefore, from the total probability theorem, we have

$$P(w_r) = \sum_{l_r=1}^{M} P(w_r \mid l_r) \cdot P(l_r). \tag{4}$$

Hence, the probability distribution of interest is given by

$$P(w_r) = \sum_{l_r=1}^{M} P(w_r \mid l_r) \sum_{w_{r-1}=1}^{N} P(l_r \mid w_{r-1}) \cdot P(w_{r-1}). \tag{5}$$

In our methodology, it is necessary therefore to determine an expression for the two conditional probabilities: $P(w_r \mid l_r)$ and $P(l_r \mid w_{r-1})$.

Consider first $P(l_r \mid w_{r-1})$, the probability that l_r S-boxes in round r are affected by changes given that there are w_{r-1} output changes of the round $r-1$ substitutions. This can be determined by considering the number of selections of w_{r-1} bit changes, $N_W(w_{r-1})$, and the number of selections of w_{r-1} bit changes that affect exactly l_r S-boxes, $N_{LW}(l_r, w_{r-1})$, and is, subsequently, given by

$$P(l_r \mid w_{r-1}) = N_{LW}(l_r, w_{r-1})/N_W(w_{r-1}) \tag{6}$$

where

$$N_W(w) = \binom{N}{w} \tag{7}$$

and $N_{LW}(l, w)$ is determined as in Lemma 2. Equation (6) incorporates the stochastic nature of the permutation by assuming that any selection of w_{r-1} bits from the output of the round $r-1$ substitutions results in a random selection of the w_{r-1} corresponding input bits to round r where all selections of the w_{r-1} input bits are equally likely.

Lemma 1 (Generalization of Inclusion-Exclusion Principle [25, p.106]): Consider a set of objects, each of which may or may not possess each property from a total set of ϕ properties. If the properties are symmetric, the number of objects which possess exactly t properties, $\Gamma(t)$, $0 \le t \le \phi$, is given by

$$\Gamma(t) = \sum_{i=t}^{\phi} (-1)^{i-t} \binom{i}{t} \binom{\phi}{i} \Gamma^*(i) \tag{8}$$

where $\Gamma^*(i)$ represents the number of objects which have at least i particular properties.

Lemma 2: Assuming that each round consists of M $n \times n$ S-boxes, the number of selections of w input bit changes to a round that affect exactly l S-boxes is given by

$$N_{LW}(l, w) = A(M-l, w, M) \tag{9}$$

where

$$A(M-l, w, M) = \sum_{i=M-l}^{M} (-1)^{i-(M-l)} \binom{i}{M-l} \binom{M}{i} A^*(i, w, M) \tag{10}$$

with

$$A^*(i, w, M) = \binom{(M-i)n}{w}. \tag{11}$$

Proof: Let $A(M-l, w, M)$ represent the number of selections of w bit changes so that exactly $M-l$ S-boxes do not have input bit changes. Therefore, $N_{LW}(l, w) = A(M-l, w, M)$. Subsequently, letting $A^*(i, w, M)$ represent the number of selections such that at least i particular S-boxes do not have input changes and applying Lemma 1 leads directly to equation (10). The quantity $A^*(i, w, M)$ is given in equation (11) as the number of selections of w bit changes for the remaining $M-i$ S-boxes which may or may not have changes. ☐

Consider now the determination of $P(w_r \mid l_r)$, the probability distribution of output changes of the round r S-boxes given that the inputs to l_r S-boxes are affected by changes from round $r-1$. Let $\mathbf{d} = [d_1\ d_2\ ...\ d_{l_r}]$ where $d_i \in \{1, ..., n\}$ is the number of output changes, $wt(\Delta\mathbf{Y})$, in the i-th S-box that has had an input change. Define

$$\Lambda = \left\{ \mathbf{d} \mid \sum_{i=1}^{l_r} d_i = w_r \right\} \tag{12}$$

to represent the values of $\mathbf{d}$ for which there are a total of w_r output bit changes. It may be seen that

$$P(w_r \mid l_r) = \sum_{\mathbf{d}\in\Lambda} P(\mathbf{d}) \tag{13}$$

where $P(\mathbf{d})$ represents the probability of a particular $\mathbf{d}$ occurring. Since each S-box operates independently, we have

$$P(\mathbf{d}) = \prod_{i=1}^{l_r} P_D(d_i). \tag{14}$$

The probability of $P(w_r \mid l_r)$ could be computed by examining all $\mathbf{d} \in \{1, ..., n\}^{l_r}$ to find all $\mathbf{d} \in \Lambda$. However, this is very computationally intensive since there are n^{l_r} possible values of $\mathbf{d}$ for a particular value of l_r. In order to determine the complete distribution for $P(w_r \mid l_r)$ we must consider all values of l_r and, therefore, there are $N(\mathbf{d})$ different values of $\mathbf{d}$ to be considered, where

$$N(\mathbf{d}) = \sum_{l_r=1}^{M} n^{l_r} = \frac{n(n^M - 1)}{n - 1}. \tag{15}$$

For example, a 64-bit network using 4×4 S-boxes, would require consideration of approximately 2^{32} vector values.

Instead, it is more efficient to determine $P(w_r \mid l_r)$ by summing over only unique unordered arrangements of the elements of $\mathbf{d}$. We introduce a vector of l_r elements, $\tilde{\mathbf{d}}$, derived by sorting the elements of $\mathbf{d}$ from smallest to largest. That is, $\tilde{\mathbf{d}} = \left[d_{i_1}\ d_{i_2}\ ...\ d_{i_{l_r}}\right]$ where $d_{i_1} \leq d_{i_2} \leq ... \leq d_{i_{l_r}}$. Now in order to determine the complete distribution of $P(w_r \mid l_r)$ the number of vector values to consider is given by[1]

$$N\left(\tilde{\mathbf{d}}\right) = \binom{M + n}{M} - 1. \tag{16}$$

Therefore, for a 64-bit network using 4×4 S-boxes there are only about 2^{12} different vector values to consider.

---

[1] The value of $N\left(\tilde{\mathbf{d}}\right)$ is determined in the following manner [26, p.38]. Consider sorted arrangements of the M S-boxes according to the values of $wt(\Delta \mathbf{Y}_{rs})$, $1 \leq s \leq M$. There are $n + 1$ classes of S-boxes corresponding to $0 \leq wt(\Delta \mathbf{Y}_{rs}) \leq n$ and a sorted arrangement is identified by placing n imaginary separators between S-boxes when the class changes. Hence, the number of arrangements is given by considering the selection of the placements of the n separators among $M + n$ different elements. From this, one is subtracted to account for the fact that the placement of all separators after M S-boxes is equivalent to $l_r = 0$ and is invalid.

Define Φ to be the set of possible $\widetilde{\mathbf{d}}$ for a particular value of l_r. For each sorted vector $\widetilde{\mathbf{d}}$ there are $l_r!/(\beta_1!\beta_2!...\beta_n!)$ corresponding unsorted $\mathbf{d}$ where $\beta_\alpha = \#\{i \mid 1 \leq i \leq l_r,\ d_i = \alpha\}$. Therefore,

$$P(w_r \mid l_r) = \sum_{\mathbf{d} \in \Phi \cap \Lambda} \frac{l_r!}{\beta_1!\beta_2!...\beta_n!} P(\mathbf{d}) \tag{17}$$

where $P(\mathbf{d})$ is given as before and $\Phi \cap \Lambda$ represents the set of all $\widetilde{\mathbf{d}}$ for which there are a total of w_r output bit changes.

It can also be shown that $P(w_r \mid l_r)$ can be efficiently calculated by

$$P(w_r \mid l_r) = \frac{1}{(2^n - 1)^{l_r}} \sum_{i=0}^{l_r} (-1)^i \binom{l_r}{i} \binom{(l_r - i)n}{w_r}. \tag{18}$$

However, in order to establish the foundation for the remainder of the analysis in the paper, we have explicitly examined the calculation of $P(w_r \mid l_r)$ by the methodology of summing over the unordered arrangements of the elements of $\mathbf{d}$. It should be noted that, for many networks of practical size, the methodology of computing over all unordered arrangements of the elements of $\mathbf{d}$ as in equation (17) is reasonable. However, in cases where very large networks are of interest, equation (18) may be used as a more efficient means of computation.

Using equations (5), (6), and (17) we can now recursively determine the probability distribution of bit changes and thereby determine the expected number of bit changes after round r, $\boldsymbol{E}(W_r)$, given an initial distribution of

$$P(W_0 = w_0) = \begin{cases} 1 & , w_0 = 1 \\ 0 & , w_0 \neq 1 \end{cases} \tag{19}$$

where W_0 represents the number of plaintext bit changes. Subsequently, we may determine the minimum number of rounds required to reasonably satisfy the avalanche criterion.

Note that, due to the stochastic nature of the permutation, for $r \geq 2$ all output bits are equally likely to change regardless of the input change distribution and, hence, $\boldsymbol{E}(W_r)/N$ can be used as a measure of the network's adherence to SAC after r rounds.

(b) Model B — Deterministic Permutation

In general, modelling the avalanche characteristics of an SPN with a given fixed permutation is a more difficult problem than the stochastic permutation case. The approach to the problem depends on the permutation. One particularly interesting class of permutations was identified by Ayoub [27]. This class is similar to the permutation structure introduced by Kam and Davida [8] as a methodology for providing completeness for networks which satisfy $N = n^R$. (A network is said to be *complete* if each output bit is a nondegenerate function of all input bits.) Ayoub extended this structure to a class of permutations, referred to as Cryptographically Equivalent Permutations (CEP), to be used in a network with an arbitrary number of rounds. Permutations belonging to the class of CEP have the characteristics that they are optimal (in the sense that they achieve completeness in the fewest number of consecutive rounds), allow the use of the same permutation for each round, and are applicable to networks which satisfy the constraint that $log\ N/log\ n \leq 3$. In this section we consider two useful permutation structures belonging to the class of CEP.

In order to simply describe the permutations we introduce the concept of bit partitions. The N bits of an SPN round may be divided into partitions of n^2 contiguous bits (or n contiguous S-boxes) with the first partition beginning at the first bit. A bit coordinate position in round r may be identified by $\left(\widehat{p}_r, \widehat{s}_r, \widehat{b}_r\right)$ where $\widehat{p}_r$, $1 \leq \widehat{p}_r \leq N/n^2$, represents the partition to which the bit belongs, $\widehat{s}_r$, $1 \leq \widehat{s}_r \leq n$, represents the S-box number within the partition, and $\widehat{b}_r$, $1 \leq \widehat{b}_r \leq n$, represents the bit number within the S-box. Assuming that bit i of round $r-1$ is connected to bit j of round r, the two permutations of interest satisfy the following constraints:

Permutation Type I:

$$\begin{aligned} &N = n^2 \\ &i \equiv \left(\widehat{p}_{r-1} = 1, \widehat{s}_{r-1}, \widehat{b}_{r-1}\right) \\ &j \equiv \left(\widehat{p}_r = 1, \widehat{s}_r = \widehat{b}_{r-1}, \widehat{b}_r = \widehat{s}_{r-1}\right) \end{aligned} \tag{20}$$

Permutation Type II:

$$\begin{aligned} &N = n^3 \\ &i \equiv \left(\widehat{p}_{r-1}, \widehat{s}_{r-1}, \widehat{b}_{r-1}\right) \\ &j \equiv \left(\widehat{p}_r = \widehat{b}_{r-1}, \widehat{s}_r = \widehat{p}_{r-1}, \widehat{b}_r = \widehat{s}_{r-1}\right). \end{aligned} \tag{21}$$

The network of Figure 1 uses a Type I permutation. Figure 2 illustrates an SPN utilizing a Type II permutation for which $N = 27$, $R = 4$, and $n = 3$.

Permutation Type I:

The recursive model for permutation Type I is based on finding the probability distribution of the number of affected S-boxes in a round given the probability distribution of the number of affected S-boxes in the previous round. Therefore, we are interested in determining $P(L_{r+1} = l_{r+1} \mid L_r = l_r) \equiv P(l_{r+1} \mid l_r)$.

Let $P^*(\delta, \mathbf{d})$ represent the probability that a specific $\mathbf{d}$ occurs and at least δ particular S-boxes are not affected in the next round. Applying Lemma 1, it may be seen that

$$P(l_{r+1} = n - t \mid l_r) = \sum_{\mathbf{d} \in \Phi} \frac{l_r!}{\beta_1! \beta_2! \ldots \beta_n!} \sum_{\delta=t}^{n} (-1)^{\delta - t} \binom{\delta}{t} \binom{n}{\delta} P^*(\delta, \mathbf{d}) \qquad (22)$$

where $\mathbf{d}$, Φ, and β_α are defined as before. The factor $l_r!/(\beta_1!\beta_2!\ldots\beta_n!)$ represents the number of different values of $\mathbf{d}$ corresponding to a sorted vector of Φ.

The probability that at least δ particular S-boxes are not affected in the next round from an S-box with d_i output changes, $P^*(\delta \mid d_i)$, is simply determined by

$$P^*(\delta \mid d_i) = \binom{n-\delta}{d_i} / \binom{n}{d_i} \qquad (23)$$

where the numerator represents the number of ways of selecting d_i bits from the $n - \delta$ bits connected to the S-boxes that may be affected in the next round and the denominator is the total number of ways of selecting d_i bits from n bits. Subsequently, $P^*(\delta, \mathbf{d})$ is given by

$$P^*(\delta, \mathbf{d}) = \prod_{i=1}^{l_r} P^*(\delta, d_i) = \prod_{i=1}^{l_r} P^*(\delta \mid d_i) \cdot P_D(d_i). \qquad (24)$$

Assuming an initial probability distribution of

$$P(L_1 = l_1) = \begin{cases} 1 & , l_1 = 1 \\ 0 & , l_1 \neq 1 \end{cases} \qquad (25)$$

it is simple to recursively calculate $P(l_{r+1})$ using

$$P(l_{r+1}) = \sum_{l_r=1}^{M} P(l_{r+1} \mid l_r) \cdot P(l_r). \tag{26}$$

The probability distribution of bit changes for a particular round can be determined by equation (4) where $P(w_r \mid l_r)$ is determined as in equation (17). As before, the tendency of a network to satisfy the avalanche criterion as the number of rounds is increased can be examined by calculating the expected bit changes after each round, $\boldsymbol{E}(W_r)$.

Note that, since we assume that the S-box properties are symmetric and the permutation is symmetric, any variation between the probabilities of different output bits changing are eliminated after the second round and, hence, the SAC nature of a network using permutation Type I can be considered to be given by $\boldsymbol{E}(W_r)/N$ for $r \geq 2$.

Permutation Type II:

The model for this scenario determines the probability distribution of bit changes per round by recursively considering the probability distribution of an unordered list of the number of S-boxes with input changes per partition.

Let $\mathrm{h}_r = \left[\widehat{\mathrm{h}}_1^{(r)} \ldots \widehat{\mathrm{h}}_n^{(r)}\right]$ where $\widehat{\mathrm{h}}_j^{(r)} = \left[\widehat{h}_1^{(rj)} \ldots \widehat{h}_n^{(rj)}\right]$ with $\widehat{h}_i^{(rj)} = 1$ if the i-th S-box of the j-th partition in round r has an input change and $\widehat{h}_i^{(rj)} = 0$ if it does not. Then, $wt\left(\widehat{\mathrm{h}}_j^{(r)}\right)$ represents the number of S-boxes in partition j of round r with input changes. Now define $\mathrm{g}_r = \left[\, g_1^{(r)} \ldots g_n^{(r)}\right]$ to represent the elements of $\left[wt\left(\widehat{\mathrm{h}}_1^{(r)}\right) \ldots wt\left(\widehat{\mathrm{h}}_n^{(r)}\right)\right]$ conveniently sorted from the smallest element to the largest element. That is, $\mathrm{g}_r = \left[\, g_1^{(r)} \ldots g_n^{(r)}\right] = \left[wt\left(\widehat{\mathrm{h}}_{j_1}^{(r)}\right) \ldots wt\left(\widehat{\mathrm{h}}_{j_n}^{(r)}\right)\right]$ where $wt\left(\widehat{\mathrm{h}}_{j_1}^{(r)}\right) \leq \ldots \leq wt\left(\widehat{\mathrm{h}}_{j_n}^{(r)}\right)$. We are interested in determining $P(\mathrm{g}_r \mid \mathrm{g}_{r-1})$.

Lemma 3: For an SPN using permutation Type II, let $S_A^{(i)}$ represent a set of n S-boxes in round $r-1$ where each S-box belongs to partition i and $S_B^{(i)}$ represent a set of n S-boxes in round r

where each S-box is the i-th S-box of a partition. Then the connections between $S_A^{(i)}$ and $S_B^{(i)}$ are equivalent to permutation Type I.

Without loss of generality, let the number of S-boxes of $S_A^{(i)}$ with input changes be $g_i^{(r-1)}$ and the number of S-boxes of $S_B^{(i)}$ with input changes be determined by

$$l_r^{(i)} = \sum_{j=1}^{n} \widehat{h}_i^{(rj)}. \tag{27}$$

Hence, the probability distribution of interest is given by

$$P(\mathrm{g}_r \mid \mathrm{g}_{r-1}) = \sum_{\mathbf{h}_r \in \mathrm{Z}} \prod_{i=1}^{n} P\Big(l_r^{(i)} \mid g_i^{(r-1)}\Big) / \binom{n}{l_r^{(i)}} \tag{28}$$

where Z is the set of $\mathbf{h}_r$ which, when the weights of their elements are sorted, correspond to g_r. As implied by Lemma 3, we can determine $P\Big(l_r^{(i)} \mid g_i^{(r-1)}\Big)$ using equation (22). Note that the denominator in the product term is necessary since we are only interested in one particular selection of $l_r^{(i)}$ S-boxes for each $\mathbf{h}_r$ in the summation.

Let T represent the set of all possible g_r. The probability of g_r can be calculated using

$$P(\mathrm{g}_r) = \sum_{\mathrm{g}_{r-1} \in \mathrm{T}} P(\mathrm{g}_r \mid \mathrm{g}_{r-1}) \cdot P(\mathrm{g}_{r-1}) \tag{29}$$

and letting

$$P(\mathrm{g}_1) = \begin{cases} 1 & , \mathrm{g}_1 = [0 \ldots 0\ 1] \\ 0 & , otherwise. \end{cases} \tag{30}$$

The probability of bit changes for round r is simply calculated from

$$P(w_r) = \sum_{\mathrm{g}_r \in \mathrm{T}} P(w_r \mid \mathrm{g}_r) \cdot P(\mathrm{g}_r) \tag{31}$$

where

$$P(w_r \mid \mathrm{g}_r) = \sum_{\mathbf{d} \in \Phi \cap \Lambda} \frac{l_r!}{\beta_1! \beta_2! \ldots \beta_n!} P(\mathbf{d}) \tag{32}$$

with

$$l_r = \sum_{j=1}^{n} g_j^{(r)} \tag{33}$$

and $\mathbf{d}$, Φ, Λ, and β_α are defined as before and $P(\mathbf{d})$ is determined as in equation (14). Again, the avalanche behaviour of the network may be determined by $\boldsymbol{E}(W_r)$. It should be noted that the methodology of equations (27) to (33) is computationally intensive and, hence, is limited to networks of modest size. For example, networks which use permutation Type II with $n = 8$ and $N = 512$ cannot be modeled since equation (28) cannot be evaluated in a practical amount of time.

Due to the symmetries in the S-boxes and the permutation, for $r \geq 3$, a network using permutation Type II will have all output bits changing equiprobably and, hence, its SAC behaviour revealed by examination of $\boldsymbol{E}(W_r)/N$.

(c) Results

The resulting graphs of expected bit changes as a function of the number of rounds in the network are presented in Figure 3 for both model types applied to a network with $N = 64$ based on 4×4 and 8×8 S-boxes. Note that in all cases the expected number of bit changes approaches the ideal number of $N/2 = 32$ as the number of rounds in the network is increased. As well, there is only a small difference in the performance of Model A and Model B, implying that little advantage is gained in using an optimal set of permutations.

It is apparent that the network composed of 8×8 S-boxes satisfies the avalanche criterion in fewer rounds than the network composed of 4×4 S-boxes. Define ϵ to be the error of an R round network from perfectly satisfying SAC such that $\epsilon = \mid 1 - \boldsymbol{E}\,(W_r)/(N/2) \mid$. The results of Model A and B indicate that a 64–bit SPN reasonably satisfies SAC, with $\epsilon \lesssim 10^{-5}$, for $R \geq 16$ when $n = 4$ and for $R \geq 7$ when $n = 8$.

In order to assess the validity of the analytical methods used, Figure 3 also includes experimental graphs of the avalanche behaviour for the networks using the deterministic permutations. For the network with 4×4 S-boxes, a set of 16 S-boxes was randomly selected and used in all rounds with the rounds connected by permutation Type II; for the network with 8×8 S-boxes, a set of 8 S-boxes was randomly selected and used in all rounds with the rounds connected by permutation Type I. The results presented are the expected ciphertext bit changes for a single input bit change based on 10^5 randomly selected plaintexts. From the graphs, it is apparent that the experimental results agree very closely with the analytical results of Model B. For the network based on 4×4 S-boxes, the relative errors of the experimental values were less than 2% of the values determined from the analytical model; for the network using 8×8 S-boxes, the relative errors were less than 1%. For both network types, the relative error decreased significantly as the number of rounds increased: for a large number of rounds ($R > 6$), the relative error in both networks was about 0.05%.

IV. Improving Avalanche Characteristics of Networks

In this section we consider two methodologies for improving the avalanche characteristics of SPNs: (1) improving the diffusion characteristics of the S-boxes and (2) using a diffusive linear transformation between rounds of S-boxes. In both cases, we assume that the permutation is stochastic so that the analysis is tractable and since Figure 3 suggests that a stochastic permutation is a good approximation of a deterministic permutation model.

(a) Diffusion Characteristics of S-boxes

The avalanche characteristics of an SPN may be improved by selecting S-boxes for the network which have strong diffusion characteristics, i.e., the property that a small input change leads to a large output change. For example, one property which improves the diffusion characteristics of an S-box is the property that we refer to as guaranteed avalanche (GA).

Definition 3: An S-box satisfies the property of *guaranteed avalanche of order* γ if, for a one bit input change, at least γ output bits change, i.e., $wt(\Delta\mathbf{X}) = 1 \Rightarrow wt(\Delta\mathbf{Y}) \geq \gamma$.

Note that a guaranteed avalanche of order $\gamma = n$ is not possible in a bijective S-box since there are n S-box input changes such that $wt(\Delta\mathbf{X}) = 1$ and only one S-box output change satisfying $wt(\Delta\mathbf{Y}) = n$. Guaranteed avalanche order 2 is an acknowledged DES S-box design criterion [28].

Consider the S-box model of equations (1) and (2). In order to modify the model to create a general model for an S-box which satisfies guaranteed avalanche of order γ, we shall apply equation (1) and replace equation (2) with probabilities conditioned on the number of input bit changes. Letting $P'_D(d) \equiv P_D(d \mid wt(\Delta\mathbf{X}) = 1)$ and $P''_D(d) \equiv P_D(d \mid wt(\Delta\mathbf{X}) > 1)$, equation (2) is replaced with

$$P'_D(d) = \begin{cases} \dfrac{\binom{n}{d}}{\sum\limits_{j=\gamma}^{n} \binom{n}{j}} & , d \geq \gamma \\ 0 & , d < \gamma \end{cases} \tag{34}$$

and

$$P''_D(d) = \frac{\binom{n}{d} - n \cdot P'_D(d)}{2^n - 1 - n} \tag{35}$$

for $1 \leq d \leq n$. Note that it is assumed that the value of γ is realizable in bijective S-boxes and that equations (34) and (35) represent valid probabilities. For example, $\gamma \neq n$ since $\gamma = n$ implies $P'_D(n) = 1$ and $P''_D(n) < 0$.

The motivation for the model is similar to the previous S-box model. The probabilities $P'_D(d)$ and $P''_D(d)$ are not intended to reflect probabilities that are necessarily found in any physically realizable S-box. Rather, the S-box model characteristics are determined from an average of all S-boxes which satisfy the guaranteed avalanche order. The case for $wt(\Delta\mathbf{X}) = 1$ arises from assuming that selection of $\Delta\mathbf{Y}$ is uniformly distributed over the set of values such that $D = wt(\Delta\mathbf{Y}) \geq \gamma$. Considering the case of $wt(\Delta\mathbf{X}) > 1$, we see that the denominator of equation (35) represents the number of values of $\Delta\mathbf{X}$ for which $wt(\Delta\mathbf{X}) > 1$ and the numerator represents the number of values of $\Delta\mathbf{Y}$ for which $wt(\Delta\mathbf{Y}) = d$, adjusted to remove the expected number of $\Delta\mathbf{Y}$ values used for the n values of $\Delta\mathbf{X}$ for which $wt(\Delta\mathbf{X}) = 1$.

Considering this new S-box model within the context of the Model A network with stochastic permutations, the development follows similarly to equations (4) through (19), except that we must now consider separately the case of an S-box with a one bit input change and the case of an S-box with more than a one bit input change.

Let l'_r represent the number of S-boxes in round r for which $wt(\Delta\mathbf{X}) = 1$ and l_r represent the number of S-boxes in round r for which $wt(\Delta\mathbf{X}) \geq 1$. Now equation (5) becomes

$$P(w_r) = \sum_{l_r=1}^{M} \sum_{l'_r=0}^{l_r} P\big(w_r \mid l'_r, l_r\big) \sum_{w_{r-1}=1}^{N} P\big(l'_r, l_r \mid w_{r-1}\big) \cdot P(w_{r-1}). \tag{36}$$

Consider first the determination of

$$P\big(l'_r, l_r \mid w_{r-1}\big) = N_{L'LW}\big(l'_r, l_r, w_{r-1}\big)/N_W(w_{r-1}) \tag{37}$$

where $N_W(w)$ is given by equation (7) and $N_{L'LW}(l', l, w)$ is determined as in the following lemma.

Lemma 4: Assuming that each round consists of M $n \times n$ S-boxes, the number of selections of w input bit changes to a round that affect exactly l S-boxes such that l' S-boxes have changes in one input bit only is given by

$$N_{L'LW}\big(l', l, w\big) = \binom{M}{l} \cdot B\big(l', l, w\big) \tag{38}$$

where

$$B(l',l,w) = \sum_{i=l'}^{l} (-1)^{i-l'} \binom{i}{l'} \binom{l}{i} B^*(i,l,w) \tag{39}$$

and

$$B^*(i,l,w) = n^i \sum_{j=0}^{l-i} (-1)^j \binom{l-i}{j} A^*(j,w-i,l-i) \tag{40}$$

with $A^*(j,w-i,l-i)$ calculated from equation (11).

Proof: Let $B(l',l,w)$ represent the number of selections of w bit changes for l S-boxes such that each S-box has at least one input bit change and exactly l' S-boxes have only a one bit input change. Allowing for the number of ways of selecting l S-boxes from the total set of M S-boxes, $N_{L'LW}(l',l,w)$ is given by equation (38).

Using the generalization of the inclusion-exclusion principle in Lemma 1, $B(l',l,w)$ is straightforwardly determined as in equation (39), where $B^*(i,l,w)$ represents the number of selections such that at least i particular S-boxes have a one bit input change and all l S-boxes have at least one input bit change.

The quantity $B^*(i,l,w)$ is given in equation (40) as the number of selections of single bit changes within the i S-boxes, n^i, multiplied by the number of selections such that the remaining $w-i$ bit changes are placed so that each of the remaining $l-i$ S-boxes has at least one input bit changing (or equivalently that no S-boxes have no input bit changes). The number of selections of the remaining $w-i$ bit changes are determined using Lemma 1 where we have utilized $A^*(j,w-i,l-i)$ from equation (11) as the number of selections of the $w-i$ bit changes so that at least j of the $l-i$ S-boxes do not have a bit change. □

We may now determine $P(w_r \mid l'_r, l_r)$ similarly to equation (13) except that we must now consider the two different cases for the number of input bits changed. Define the vector $\mathbf{d}' = \left[d'_1 \; d'_2 \; ... \; d'_{l'_r}\right]$ such that $d'_i \in \{\gamma, ..., n\}$ represents the number of output changes, $wt(\Delta\mathbf{Y})$, of the i-th S-box for which $wt(\Delta\mathbf{X}) = 1$. Similarly, define the vector $\mathbf{d}'' = \left[d''_1 \; d''_2 \; ... \; d''_{l_r - l'_r}\right]$ such that $d''_i \in \{1, ..., n\}$ represents the number of output changes, $wt(\Delta\mathbf{Y})$, of the i-th S-box for which $wt(\Delta\mathbf{X}) > 1$. We can determine the probability using

$$P\big(w_r \mid l'_r, l_r\big) = \sum_{(\mathbf{d}',\mathbf{d}'')\in\Lambda^*} P\big(\mathbf{d}', \mathbf{d}''\big) \tag{41}$$

where

$$\Lambda^* = \left\{ \big(\mathbf{d}', \mathbf{d}''\big) \mid \sum_{i=1}^{l'_r} d'_i + \sum_{i=1}^{l_r - l'_r} d''_i = w_r \right\}. \tag{42}$$

The probability $P(\mathbf{d}', \mathbf{d}'')$ is given by

$$P\big(\mathbf{d}', \mathbf{d}''\big) = \left[\prod_{i=1}^{l'_r} P'_D\big(d'_i\big)\right] \left[\prod_{i=1}^{l_r - l'_r} P''_D\big(d''_i\big)\right]. \tag{43}$$

To improve the efficiency of the algorithm calculating the conditional probability, we can conveniently consider unordered arrangements of the elements of the vector pair $(\mathbf{d}', \mathbf{d}'')$. Define $\widetilde{\mathbf{d}}'$ and $\widetilde{\mathbf{d}}''$ to be vectors derived by sorting $\mathbf{d}'$ and $\mathbf{d}''$, respectively. Let Φ^* represent the set of possible values for $\left(\widetilde{\mathbf{d}}', \widetilde{\mathbf{d}}''\right)$ for a particular value of l_r and l'_r. For each sorted vector pair $\left(\widetilde{\mathbf{d}}', \widetilde{\mathbf{d}}''\right)$ there are

$$N_{\Phi^*} = \frac{l'_r!}{\beta'_1! ... \beta'_n!} \cdot \frac{(l_r - l'_r)!}{\beta''_1! ... \beta''_n!} \tag{44}$$

corresponding vector pair values for $(\mathbf{d}', \mathbf{d}'')$ where $\beta'_\alpha = \#\{i \mid 1 \le i \le l'_r,\ d'_i = \alpha\}$ and $\beta''_\alpha = \#\{i \mid 1 \le i \le l_r - l'_r,\ d''_i = \alpha\}$. The resulting equation for the conditional probability can be written as

$$P\big(w_r \mid l'_r, l_r\big) = \sum_{(\mathbf{d}',\mathbf{d}'')\in\Phi^*\cap\Lambda^*} N_{\Phi^*} \cdot P\big(\mathbf{d}', \mathbf{d}''\big). \tag{45}$$

The probability $P(w_r)$ can now be determined using equation (36) and, subsequently, $\boldsymbol{E}(W_r)$ can be calculated.

The results for 64–bit SPNs using 4×4 and 8×8 S-boxes with different GA orders are illustrated in Figure 4 and Figure 5 respectively. The graphs of Figures 4 and 5 suggest that satisfaction of the avalanche criterion occurs in fewer rounds if the S-boxes satisfy higher orders of guaranteed avalanche. However, it is important to stress that it is unclear whether it is possible to find S-boxes that satisfy high orders of guaranteed avalanche and that still reasonably satisfy other known design criteria such as nonlinearity [29].

(b) Linear Transformations Between Rounds

It is also possible to improve the avalanche characteristics of a block cipher based on Shannon's principles of confusion and diffusion by using a suitable invertible linear transformation in place of a permutation between rounds of S-boxes. Let N be even and consider the class of invertible linear transformations defined by

$$\mathbf{V} = \pi(\mathcal{L}(\mathbf{U})) \tag{46}$$

where $\mathbf{V} = [V_1 \ V_2 \ ... \ V_N]$ is the vector of input bits to a round of S-boxes, $\mathbf{U} = [U_1 \ U_2 \ ... \ U_N]$ is the vector of bits from the previous round output, $\mathcal{L}(\mathbf{U}) = [L_1(\mathbf{U}) \ ... \ L_N(\mathbf{U})]$, and π is a stochastic permutation, uniformly selected from the set of all permutations (as in Model A). The i-th element of $\mathcal{L}(\mathbf{U})$ is defined by

$$L_i(\mathbf{U}) = U_1 \oplus ... \oplus U_{i-1} \oplus U_{i+1} \oplus ... \oplus U_N. \tag{47}$$

Using such a transformation between rounds is useful in rapidly diffusing bit changes within the network as is seen in the following lemma.

Lemma 5: Let $\mathbf{Z} = \mathcal{L}(\mathbf{U})$ where $\mathbf{Z} = [Z_1 \ Z_2 \ ... \ Z_N]$. Let $\Delta\mathbf{U} = [\Delta U_1 \ ... \ \Delta U_N]$ be the XOR difference between two arbitrary values of $\mathbf{U}$, and $\Delta\mathbf{Z} = [\Delta Z_1 \ \Delta Z_2 \ ... \ \Delta Z_N]$ is the resulting

XOR difference for $\mathbf{Z}$. Then

$$wt(\Delta\mathbf{Z}) = \begin{cases} wt(\Delta\mathbf{U}) & , wt(\Delta\mathbf{U})\ even \\ N - wt(\Delta\mathbf{U}) & , wt(\Delta\mathbf{U})\ odd. \end{cases} \tag{48}$$

Hence, using the linear transformation of equation (46) between rounds of substitutions is helpful in promoting avalanche because changes of small, odd weight are translated into large changes. For example, if $N = 64$, a one bit change from the output of round r becomes a 63 bit change to the input of round $r + 1$. Applying the linear transformation to the output of each round of S-boxes resulted in the dramatic improvement of the avalanche behaviour of the network as illustrated in Figures 4 and 5. It should be noted that the linear transformation of equation (46) does not effectively diffuse bit changes when the number of bit changes is even: changes of small, even weight are translated into changes of the same small weight. For example, a two bit change from the output of round r becomes a two bit change to the input of round $r + 1$.

V. Conclusion

We have presented analytical methods for modelling the avalanche characteristics of substitution-permutation encryption networks. The results clearly indicate that networks composed of large 8×8 S-boxes satisfy the avalanche criterion and SAC in fewer rounds than those based on smaller 4×4 S-boxes. Further it is shown that strengthening the diffusion properties of the S-boxes or using a diffusive linear transformation between rounds can improve the avalanche characteristics of the network, facilitating the construction of efficient ciphers with fewer rounds required for adequate security.

Acknowledgments

The authors are grateful to the anonymous referees, whose insightful comments significantly improved the presentation of the paper.

References

[1] National_Bureau_of_Standards, “Data Encryption Standard (DES),” *Federal Information Processing Standard Publication 46*, 1977.

[2] W. Diffie and M. Hellman, “Exhaustive cryptanalysis of the NBS data encryption standard,” *Computer*, vol. 10, pp. 74–84, 1977.

[3] M. J. Wiener, “Efficient DES key search,” tech. rep., School of Computer Science, Carleton University, Ottawa, Canada, May 1994. Presented at the Rump Session of CRYPTO ’93.

[4] W. Diffie and M. E. Hellman, “Privacy and authentication: An introduction to cryptography,” *Proceedings of the IEEE*, vol. 67, no. 3, pp. 397–427, 1979.

[5] H. M. Heys and S. E. Tavares, “Key clustering in substitution-permutation network cryptosystems,” *presented at Workshop on Selected Areas in Cryptography (SAC ’94)*, Queen’s University, Kingston, Canada, May 1994.

[6] H. Feistel, “Cryptography and computer privacy,” *Scientific American*, vol. 228, no. 5, pp. 15–23, 1973.

[7] C. E. Shannon, “Communication theory of secrecy systems,” *Bell System Technical Journal*, vol. 28, pp. 656–715, 1949.

[8] J. B. Kam and G. I. Davida, “A structured design of substitution-permutation encryption networks,” *IEEE Transactions on Computers*, vol. 28, no. 10, pp. 747–753, 1979.

[9] H. M. Heys and S. E. Tavares, “Substitution-permutation networks resistant to differential and linear cryptanalysis,” accepted for publication in *Journal of Cryptology*, Sept. 1994.

[10]E. Biham and A. Shamir, “Differential cryptanalysis of DES-like cryptosystems,” *Journal of Cryptology*, vol. 4, no. 1, pp. 3–72, 1991.

[11]M. Matsui, “Linear cryptanalysis method for DES cipher,” *Advances in Cryptology: Proceedings of EUROCRYPT ’93*, Springer-Verlag, Berlin, pp. 386–397, 1994.

[12]A. Shimizu and S. Miyaguchi, “Fast data encipherment algorithm: FEAL,” *Advances in Cryptology: Proceedings of EUROCRYPT ’87*, Springer-Verlag, Berlin, pp. 267–278, 1988.

[13]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, pp. 229–236, 1990.

[14]L. Brown, M. Kwan, J. Pieprzyk, and J. Seberry, “Improving resistance to differential cryptanalysis and the redesign of LOKI,” *Advances in Cryptology: Proceedings of ASIACRYPT ’91*, Springer-Verlag, Berlin, pp. 36–50, 1993.

[15]X. Lai and J. Massey, “A proposal for a new block encryption standard,” *Advances in Cryptology: Proceedings of EUROCRYPT ’90*, Springer-Verlag, Berlin, pp. 389–404, 1991.

[16]X. Lai, J. Massey, and S. Murphy, “Markov ciphers and differential cryptanalysis,” *Advances in Cryptology: Proceedings of EUROCRYPT ’91*, Springer-Verlag, Berlin, pp. 17–38, 1991.

[17]H. Feistel, W. A. Notz, and J. L. Smith, “Some cryptographic techniques for machine-to-machine data communications,” *Proceedings of the IEEE*, vol. 63, no. 11, pp. 1545–1554, 1975.

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

[19]R. Forré, “The strict avalanche criterion: Spectral properties of boolean functions and an extended definition,” *Advances in Cryptology: Proceedings of CRYPTO ’88*, Springer-Verlag, Berlin, pp. 450–468, 1990.

[20]C. M. Adams and S. E. Tavares, “The structured design of cryptographically good S-boxes,” *Journal of Cryptology*, vol. 3, no. 1, pp. 27–41, 1990.

[21]S. Lloyd, “Counting functions satisfying a higher order strict avalanche criterion,” *Advances in Cryptology: Proceedings of EUROCRYPT ’89*, Springer-Verlag, Berlin, pp. 63–74, 1990.

[22]B. Preneel, W. Van_Leekwijck, L. Van_Linden, R. Govaerts, and J. Vandewalle, “Propagation characteristics of boolean functions,” *Advances in Cryptology: Proceedings of EUROCRYPT ’90*, Springer-Verlag, Berlin, pp. 161–173, 1991.

[23]K. Kim, T. Matsumoto, and H. Imai, “A recursive construction method of S-boxes satisfying strict avalanche criterion,” *Advances in Cryptology: Proceedings of CRYPTO ’90*, Springer-Verlag, Berlin, pp. 545–553, 1991.

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

[25]W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons, 3rd ed., 1968.

[26]F. S. Roberts, *Applied Combinatorics*. Englewood Cliffs, N.J.: Prentice-Hall, 1984.

[27]F. Ayoub, “The design of complete encryption networks using cryptographically equivalent permutations,” *Computers and Security*, vol. 2, pp. 261–267, 1982.

[28]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, pp. 3–8, 1987.

[29]W. Meier and O. Staffelbach, “Nonlinearity criteria for cryptographic functions,” *Advances in Cryptology: Proceedings of EUROCRYPT ’89*, Springer-Verlag, Berlin, pp. 549–562, 1990.