Information Leakage of Feistel Ciphers

Howard M. Heys

Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, Newfoundland, Canada A1B 3X5
email: howard@engr.mun.ca

Abstract: In this paper, we examine the information leakage between sets of plaintext and ciphertext bits in symmetric-key block ciphers. The paper demonstrates the effectiveness of information leakage as a measure of cipher security by relating information leakage to linear cryptanalysis and by determining a lower bound on the amount of data required in an attack from an upper bound on information leakage. As well, a model is developed which is used to estimate the upper bound on the information leakage of a general Feistel block cipher. For a cipher that fits the model well, the results of the analysis can be used as a measure in determining the number of rounds required for security against attacks based on information leakage. It is conjectured that the CAST-128 cipher fits the model well and using the model it is predicted that information leaked from 20 or fewer plaintext bits is small enough to make an attack on CAST-128 infeasible.

Key Words: cryptography, Feistel cipher, linear cryptanalysis, symmetric-key block ciphers

I. Introduction

In his landmark paper, Shannon [1] introduced concepts relating cryptography and information theory. As well, the paper laid the foundation for the structure of modern symmetric-key block ciphers, proposing an iterated architecture composed of a number of rounds of simple cryptographic mixing operations capable of producing "confusion" and "diffusion". The best known architecture to which these concepts are practically applied is referred to as a Feistel cipher [2] and this is the structure used for the popular Data Encryption Standard (DES) [3]. Another Feistel cipher that we shall later refer to in our paper is called CAST-128 [4]. CAST-128 has been implemented in several commercial network security products [5]. Many other modern Feistel ciphers have been proposed, including some of the candidates for the Advanced Encryption Standard (AES) [6].

To date, the major focus in block cipher analysis has been the cryptanalysis of specific algorithms such as DES. This focus on cryptanalysis has lead to the discovery of several powerful cryptanalysis techniques that are applicable to a broad category of ciphers. Most notably, the two general attacks of differential cryptanalysis [7] and linear cryptanalysis [8] have had considerable influence on the analysis and subsequent design of ciphers. Differential attacks are based on predicting likely differences occurring in cipher data in response to specific differences in pairs of plaintexts; linear attacks are based on predicting highly likely linear expressions of plaintext, ciphertext, and key bits.

There have been several attempts to establish provable security in block ciphers. Most notably, Luby and Rackoff [9] provided proof that a Feistel cipher of 3 rounds based on a pseudo-random function generator as the round function is a pseudo-random permutation generator and, hence, is provably secure against chosen plaintext attacks. It was also shown that 4 rounds provide provable security against an adaptive chosen plaintext/ciphertext attack. The result is theoretical in that it relies on the round functions to be provably pseudo-random function generators. However, several practical proposals have used the Luby-Rackoff result as the rationale for ciphers using round functions constructed from function generators such as hash functions [10][11][12].

Other approaches to provable security have attempted to prove the immunity of ciphers to notable attacks such as differential and linear cryptanalysis [13][14][15] by carefully selecting the structure and components of the cipher. Typically, the round functions of a Feistel cipher are significantly simpler than a hash function, implying more rounds are required than the optimistic 4 rounds of the Luby-Rackoff result. Often, the security of ciphers relies on heuristic arguments of the difficulty of finding useful differential characteristics and linear approximations [16]. For example, the security of CAST-128 has been argued on this basis [17]. Despite many attempts at provable security in block cipher design, the fact remains that all practical cipher designs still rely on security proofs premised on reasonable assumptions of the behaviour of the cipher such as the independence of the data involved in different rounds.

Several papers have attempted to relate the information leakage of a proposed cipher structure and its components directly to the design and analysis of the security of the cipher. Some work in this regard has focussed on ciphers that employ substitution boxes or S-boxes to achieve the cipher nonlinearity (i.e., the confusion). Forré [18] considered information theoretic measures in the construction of S-boxes with the objective of constructing S-boxes to minimize the mutual information between sets of input and output bits. In [19], Dawson and Tavares extended this work to include both static and dynamic

measures of mutual information or information leakage. The work by Sivabalan, Tavares, and Peppard [20] examined the information leakage for a simple substitution-permutation network with a study of different S-box constructions. However, since their information leakage measures were derived experimentally, the results apply to a very small system and must be extrapolated to larger ciphers.

Zhang, Tavares, and Campbell [21] focussed on characterizing the cryptographic properties of boolean functions within the framework of information theory. However, their results were not applied to any actual ciphers. In [22], Youssef and Tavares examined the information leakage of randomly selected functions of multiple output bits. Exact expressions and simple upper bounds on the expected information leakage of a randomly selected function were derived and discussed in the context of S-box generation. However, since the results give average information leakage, rather than upper bounds on information leakage, they cannot be used to derive estimates of the security of ciphers constructed from the randomly selected functions.

In this paper, we relate the information leakage of a cipher to cryptanalysis, and, in particular, linear cryptanalysis. As well, we develop a model of the information leakage of a general Feistel cipher. As a result, the methods that we employ in modelling the cipher are able to estimate upper bounds on the information leakage in the cipher and from this it is possible to place a lower bound on the amount of data (i.e., the number of ciphertexts) required in an attack.

II. Background

In this section, we outline some of the basic background and notation required for the paper. In particular, we review the notion of a Feistel cipher and some required fundamentals from information theory.

A. Feistel Ciphers

We consider first the concept of the Feistel structure for symmetric-key ciphers. Notationally, we shall consider that the cipher transforms an N-bit block of plaintext P to the ciphertext C by passing the data through a sequence of r rounds. As aligns with common practice, we shall assume that the number of rounds, r, is even. In a Feistel cipher, the plaintext is divided into the left half-block P_L and right half-block P_R and passed into the first round. In each round, the round function f operates on the right half of the cipher data block and is parameterized with the subkey associated with the round. Each subkey is derived from the master key K.

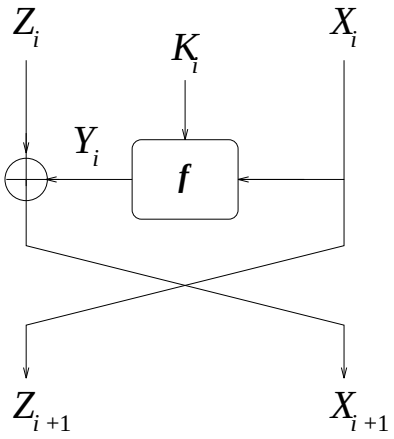

Figure 1: Round i in a Feistel Cipher

The general structure of a round in a Feistel cipher is illustrated in Figure 1. In round i of a Feistel cipher, the right half of the data block, X_i, is used as the input to the round function (parameterized with subkey K_i) with the resulting output, Y_i, being bit-wise XORed with the left half, Z_i. Then the unmodified right half and the modified left half are swapped and the data is passed to the next round as X_{i+1} and Z_{i+1}. This process is repeated for the desired number of rounds. Letting C_L and C_R represent the left and right half of the ciphertext, respectively, the general flow of data though an r-round Feistel cipher is given by:

$$\begin{array}{l} X_1 \leftarrow P_R \\ Z_1 \leftarrow P_L \\ \text{for } i = 1 \text{ to } r \text{ do} \\ \quad X_{i+1} = Z_i \oplus Y_i \\ \quad Z_{i+1} = X_i \\ C_L \leftarrow X_{r+1} \\ C_R \leftarrow Z_{r+1} \end{array} \tag{1}$$

where $Y_i = f_{K_i}(X_i)$ and "$\oplus$" represents bit-wise XOR on two binary vectors. Note that X_i, Y_i, $Z_i \in \{0,1\}^{N/2}$; the size of K_i depends on the method used to mix the subkey bits into the data in each round.

The round function is essentially responsible for the cryptographic security of the cipher (i.e., it typically achieves most of the confusion and diffusion). Generally, the more rounds applied, the more secure the cipher. The structure and nature of the round function varies significantly from cipher to cipher. One method for implementing the parameterization of the round function with the subkey is to mix a subkey of $N/2$ bits with the data bits using a bit-wise XOR prior to the nonlinear mapping of the data. That is, $Y_i = f(X_i \oplus K_i)$ where f is an unkeyed $N/2 \times N/2$ nonlinear mapping. This is similar to the keying methodology used in DES.

A notable characteristic of a Feistel cipher is that the decryption procedure is the same as encryption except the subkeys must be applied in reverse order. In general, the determination of the subkeys, $K_1, \ldots, K_r$, from the master key K (which is usually significantly shorter than the sum of all subkey lengths) is accomplished by a key scheduling algorithm. Together the key scheduling and the round function structure essentially define the cipher. It is important to note that the Feistel structure ensures that the cipher is a bijective mapping for all keys regardless of the nature of the round function and the method used to apply key bits.

B. Information Theory Notation and Definitions

In this section, we introduce the notation that we shall use when dealing with some of the basic definitions in the field of information theory. Consider a system with a random m-bit input vector X and a random n-bit output vector Y. The entropy of Y is defined to be

$$H(Y) = -\sum_{y \in \{0,1\}^n} Pr(Y = y) \log_2 Pr(Y = y) \tag{2}$$

where $Pr(Y = y)$ represents the probability that Y takes on a value y. The conditional entropy is given by

$$H(Y|X) = -\sum_{x \in \{0,1\}^m} \sum_{y \in \{0,1\}^n} Pr(Y = y, X = x) \cdot \log_2 Pr(Y = y|X = x) \tag{3}$$

and the average mutual information of the system is given by

$$I(X;Y) = H(Y) - H(Y|X). \tag{4}$$

Note that in the remainder of the paper, for convenience, we shall frequently use a more concise form for representing probabilities where $Pr(y)$ is equivalent to $Pr(Y = y)$ and $Pr(y|x)$ is equivalent to $Pr(Y = y|X = x)$.

In this paper, we use the term *information leakage* to refer to the average mutual information between subsets of plaintext bits and ciphertext bits. For example, the information leakage between a set of m bits of plaintext and a set of n bits of ciphertext for a particular key value K is given by

$$I_K(P^{(m)}; C^{(n)}) = n - H(C^{(n)}|P^{(m)}). \tag{5}$$

In general, in the paper, we shall use the superscript notation "(n)" to label a vector formed by taking n bits of the base vector. The right side of (5) follows since, assuming random plaintexts are applied to the cipher, the unconditional entropy of any n bits of

ciphertext is n due to the bijective nature of a block cipher. We will sometimes drop the subscript K when referring to an information leakage which applies to all keys.

The smaller the value of information leakage, the less information is available on the output based on knowledge of the input (and vice versa). In the case of ciphers, a secure cipher would have negligible information leakage for subsets of plaintexts bits and ciphertext bits for all keys. This implies that no information about the ciphertext bits can be determined given knowledge of the plaintext bit values, making it impossible to use information leakage between the plaintext and ciphertext to find the key.[1] A practical cipher should have an information leakage (for all subsets of plaintext and ciphertext bits) that is small enough to imply that an impractically large number of ciphertexts are required to derive the key.

III. Relationship of Information Leakage to Cryptanalysis

Before modelling the information leakage of a cipher, we first examine the effectiveness of information leakage as a measure of cipher security. Specifically, we will consider the relationship between linear cryptanalysis and information leakage and then investigate the direct relationship between information leakage and the amount of data required in an attack.

A. Information Leakage and Linear Cryptanalysis

Consider the linear expression formed from the XOR of m bits of plaintext P, n bits of ciphertext C, and l bits of key K. This is represented as

$$\begin{aligned} P[i_1] \oplus P[i_2] \oplus \ldots \oplus P[i_m] \oplus C[j_1] \oplus C[j_2] \oplus \ldots \oplus C[j_n] \\ = K[h_1] \oplus K[h_2] \oplus \ldots \oplus K[h_l] \end{aligned} \tag{6}$$

where, for example, $P[i]$ represents the i-th bit of plaintext P.

Basic linear cryptanalysis [8] is a known plaintext attack and attacks ciphers by exploiting an expression of the form of (6) which has a high probability of being satisfied (or not satisfied) given a random, uniformly distributed selection of plaintexts as input. Ideally for a secure cipher, any linear expression of the form above should have a probability of exactly $1/2$ of being satisfied.

Let ρ represent the probability that the linear expression holds for a random selection

[1] In practice, attacks such as differential and linear cryptanalysis typically attempt to use information of the penultimate round and the knowledge of the ciphertext to derive information about the subkey applied to the last round. However, for the purposes of the analysis in this paper, we shall focus on information leakage at the output of the cipher. Although we could trivially consider the analysis for one less round, the results would not be significantly different.

of plaintext P. The *probability bias* for the linear expression can be represented as $\epsilon = |\rho - 1/2|$. It can be shown that the number of known plaintexts required in a linear attack based on an expression with a probability bias of ϵ is proportional to ϵ^{-2} [8]. Hence, for a cipher to be immune to linear cryptanalysis, we require $\epsilon \to 0$ for all possible linear expressions. That is, for all subsets of plaintext, ciphertext, and key bits the expression of the form of (6) has a probability of holding that is very close to $\rho = 1/2$. (In practice, for security against a linear attack, it is sufficient that $\epsilon < 2^{-N/2}$ for a block cipher with a block size of N, so that the amount of data required in the attack is greater than the number of plaintexts available for the cipher.)

Linear cryptanalysis considers the random application of plaintexts and the resulting pseudo-random generation of ciphertexts for a particular fixed key, i.e., the key under attack. Hence, the probability ρ is actually the probability that a particular subset of plaintext bits and a particular subset of ciphertext bits sum to a fixed value which is determined from the appropriate subset of key bits of the unknown fixed key. In a cipher such as DES, the fixed value is derived by the XOR sum of key bits as in (6) because the subkeys are mixed into the data at each round using the XOR operation, allowing the elimination of variables representing data within the cipher. The result is an expression involving only the XOR of plaintext, ciphertext, and key bits that can be arranged to solve for the equivalent of one key bit given by the right side of equation (6). If the plaintexts utilized in the attack are assumed to be randomly selected, it follows that the values of the m plaintexts bits are randomly generated from a uniform distribution. Assuming $m < N$, the n ciphertext bits are random with respect to the m plaintext bits since each ciphertext bit is dependent on all plaintext bits (under the reasonable expectation that the cipher is non-degenerate) and the $N - m$ remaining plaintext bits are random and independent of the m plaintext bits in equation (6).

In the following theorem, we relate the property of information leakage and the immunity of a cipher to linear cryptanalysis.

Definition: A cipher is defined to be (m, n)-*immune* to linear cryptanalysis if, for all subsets of u plaintext bits, $1 \leq u \leq m$, and v ciphertext bits, $1 \leq v \leq n$,

$$P[i_1] \oplus P[i_2] \oplus \ldots \oplus P[i_u] \oplus C[j_1] \oplus C[j_2] \oplus \ldots \oplus C[j_v] = 0 \tag{7}$$

holds with a probability of $1/2$ for all keys, assuming a random, uniformly distributed plaintext input P.

We require the following lemma from [23] in the proof of the theorem. A compact proof of the lemma is given in [24].

Lemma 1 [23]

Consider a discrete random variable Z and a random binary vector of m bits, $X = [X[1], X[2], \ldots, X[m]]$. Then $I(X;Z) = 0$ if, and only if, $I(X[i_1] \oplus X[i_2] \oplus \ldots \oplus X[i_u]; Z) = 0$ for all subsets of u bits of X for all u, $1 \leq u \leq m$.

Theorem

For all K and all vectors formed from m plaintext bits $P^{(m)}$ and n ciphertext bits $C^{(n)}$, $I_K(P^{(m)}; C^{(n)}) = 0$ if, and only if, the cipher is (m,n)-immune to linear cryptanalysis.

Proof:

From the lemma it follows that

$$I_K(P^{(m)}; C^{(n)}) = 0 \Leftrightarrow I_K(P[i_1] \oplus \ldots \oplus P[i_u]; C[j_1] \oplus \ldots \oplus C[j_v]) = 0 \qquad (8)$$

for all subsets of u bits of the m plaintext bits and v bits of the n ciphertext bits where $1 \leq u \leq m$ and $1 \leq v \leq n$. Since the XOR sums of plaintext and ciphertext bits are balanced, this is equivalent to the cipher being (m,n)-immune. □

It should be noted that the definition of immunity to linear cryptanalysis is equivalent to the concept of correlation immunity [25]. In fact, a cipher that is (m,n)-immune to linear cryptanalysis has all boolean functions formed by the linear combination of n or fewer ciphertext bits being m-th order correlation immune. Further, since any linear combination of ciphertext bits forms a balanced boolean function, such functions are m-resilient [26].

Motivated by the strong relationship between linear cryptanalysis and information leakage, we shall use tools similar to those applied when considering linear cryptanalysis to develop a model for the security of a Feistel cipher on the basis of an information leakage paradigm. However, it is important to note that limitations of the usefulness of the relationship between linear cryptanalysis and information leakage do exist. First, as we shall see, the techniques employed in this paper are only applicable for information leakages involving modest numbers of plaintext and ciphertext bits. Theoretically, full immunity to linear cryptanalysis requires (m,n)-immunity to linear cryptanalysis with $m = N$ and $n = N$. Secondly, the concept of (m,n)-immunity to linear cryptanalysis is of theoretical interest only. In practice, ciphers can be resistant to linear cryptanalysis without satisfying (m,n)-immunity with $m = N$ and $n = N$. As long as all linear expressions have a probability bias sufficiently close to 0 (not necessarily exactly 0), linear cryptanalysis cannot be employed successfully. This explains why, for all ciphers with an N-bit plaintext P and ciphertext C, although clearly $I_K(P;C) = N \neq 0$ when the key is fixed, the

cipher can still be resistant to linear cryptanalysis: this does not contradict the theorem because, although the bias on any linear expression from the cipher is small enough to make practical attacks impossible, the cipher is not theoretically (N, N)-immune to linear cryptanalysis.

B. Complexity of Attacks Based on Information Leakage

In this section, we consider the complexity of directly attacking a cipher with non-zero information leakage based on distinguishing between a uniform distribution and the distribution of a set of ciphertext bits conditioned on a set of plaintext bits. This allows us to relate the amount of information leakage in a cipher to the security of the cipher as given by the amount of data (i.e., the number of ciphertexts) required to attack the cipher.

Consider exploiting n ciphertext bits, $C^{(n)}$. Since a Feistel cipher is a bijective mapping, $C^{(n)}$ is uniformly distributed and $H(C^{(n)}) = n$. However, for a vector $P^{(m)}$ representing m plaintext bits, it is certainly possible that $H(C^{(n)}|P^{(m)}) \neq n$. This implies for at least one point, $P^{(m)} = p^{(m)}$ and $C^{(n)} = c^{(n)}$, $Pr(c^{(n)}|p^{(m)}) \neq 1/2^n$.

Assume that the largest probability bias for $Pr(c^{(n)}|p^{(m)})$ is given by

$$\epsilon_{max} = \max_{p^{(m)}, c^{(n)}} |Pr(c^{(n)}|p^{(m)}) - 1/2^n|. \tag{9}$$

(Note that we now use a definition of probability bias which refers to the deviation from a probability of $1/2^n$, as opposed to a probability of $1/2$ as in linear cryptanalysis.) In order to distinguish information that might be used in determining the key or a subset of the key bits, the cryptanalyst would be interested in distinguishing the conditional probability distribution from the uniform distribution. We shall refer to the uniform distribution as distribution 1 and the distribution of $C^{(n)}$ conditioned on a value of $P^{(m)} = p^{(m)}$ as distribution 2. Let $\rho_1 = 1/2^n$ and $\rho_2 = Pr(c^{(n)}|p^{(m)})$. Hence, $1/2^n - \epsilon_{max} \leq \rho_2 \leq 1/2^n + \epsilon_{max}$. Assuming N_S data samples are taken, consider the random variable representing the number of times a sample has a particular value $c^{(n)}$. The resulting distributions are binomial in nature for both cases with mean $\mu_i = N_S\rho_i$ and variance $\sigma_i^2 = N_S\rho_i(1 - \rho_i)$ where the index $i \in \{1, 2\}$ is used to indicate the distribution. The cryptanalyst, in order to exploit information leakage, would like the distance between the two means to be large in comparison to the standard deviation, making it easy to distinguish the conditional distribution from the uniform distribution. However, the cipher designer will, at least implicitly, try to ensure that the conditional distribution is close enough to uniform to be indistinguishable for all but an infeasible amount of data N_S.

Assume that in order for the distributions to be distinguishable, the two distributions must have means such that

$$|\mu_2 - \mu_1| \geq \max\{\sigma_1, \sigma_2\} \tag{10}$$

where

$$\max_{c^{(n)}} |\mu_2 - \mu_1| = N_S \epsilon_{max}. \tag{11}$$

For $\epsilon_{max} \ll 1/2^n$, $\sigma_2 \approx \sigma_1$, and

$$\sigma_1 = \left[N_S \frac{1}{2^n} \left(1 - \frac{1}{2^n} \right) \right]^{1/2}. \tag{12}$$

This results in

$$N_S \geq \frac{2^n - 1}{2^{2n} \epsilon_{max}^2} \tag{13}$$

and, hence,

$$N_S \geq \frac{1}{2^{n+1} \epsilon_{max}^2}. \tag{14}$$

As expected, from the work of Matsui [8], the amount of data required, N_S, varies inversely with the square of the probability bias. It must be noted however that in the case of linear cryptanalysis, the probability of interest deviates from $1/2$, whereas the information leakage approach exploits a bias from a probability of $1/2^n$. The lower bound on N_S also appears to decrease exponentially in n, implying for a fixed ϵ_{max} that the lower bound on the amount of data required decreases significantly with more bits involved in the attack. The perspective at this point, however, is incomplete in that we must consider that ϵ_{max} is also a function of n. We deal with this issue by modelling a general Feistel cipher in Section V to examine the overall effect on the lower bound on N_S as a function of n.

In order for the cipher to be not susceptible to attack, $N_S > 2^{N-m}$, where N is the block size and m is the number of fixed plaintext bits. This ensures that the amount of data required to attack the cipher is greater than the available plaintexts which have a fixed value for m particular bits.

Although in this section, we have related the amount of data required in the attack to the probability bias, it is also possible to relate N_S directly to the information leakage of the cipher. In Section IV-C, we shall give a lower bound on the value of N_S given an upper bound on the information leakage I_{max} where $I(P^{(m)}; C^{(n)}) \leq I_{max}$.

IV. Foundational Theory for Information Leakage Models

Before discussing a model of the information leakage of a general Feistel cipher, it is necessary to consider some basic mathematical tools that will be required in the analysis.

A. Generalized Piling-Up Lemma

The computation of the information leakage of the cipher will be done in an iterative fashion, from round to round. It will be based on using principles similar to those used in the calculation of the probability used in linear cryptanalysis [8]. Fundamental to that analysis is the "piling-up lemma" used to determine the probability that the XOR sum of independent binary random variables is equal to 0 or 1. In this section, we generalize the lemma to apply to vectors consisting of multiple bits combined by bit-wise XOR.

Consider

$$Y = X_1 \oplus X_2 \oplus \ldots \oplus X_k \tag{15}$$

where X_i is a randomly generated n-bit vector, "$\oplus$" represents bit-wise XOR, and all X_i are independent. Let ϵ represent the maximum bias from $1/2^n$ for the probability of the occurrence of a specific value of X_i over all i, $1 \leq i \leq k$. That is,

$$\epsilon = \max_{1 \leq i \leq k,\ x \in \{0,1\}^n} |Pr(X_i = x) - 1/2^n|. \tag{16}$$

Define ϵ_{max} to represent the maximum bias from $1/2^n$ for the probability that Y takes on a particular value y. That is,

$$\epsilon_{max} = \max_{y \in \{0,1\}^n} |Pr(Y = y) - 1/2^n|, \tag{17}$$

We now state the lemma that we shall refer to as the *generalized piling-up lemma.*

Lemma 2 (Generalized Piling-up Lemma)

Consider the bit-wise XOR of $k \geq 2$ random, independent, n-bit vectors as given in (15) and the definitions of ϵ and ϵ_{max} given in (16) and (17). Then

$$\epsilon_{max} \leq 2^{(k-1)n} \epsilon^k. \tag{18}$$

Moreover, if $\epsilon < 1/2^n$, then $\epsilon_{max} < \epsilon$.

Proof:

We shall prove (18) by using induction on k.

Base Case: Let $k = 2$. We are interested in determining an upper bound on $Pr(X_1 \oplus X_2 = y)$ where $y \in \{0,1\}^n$. Without loss of generality, we shall consider y to be the all zeros vector.

Let $Pr(X_1 = j) = 1/2^n + \delta'_j$ and $Pr(X_2 = j) = 1/2^n + \delta''_j$ where j is the integer representation of a vector. The probability that the XOR of two vectors is all zeros is equivalent to the probability that the two vectors are the same:

$$Pr(X_1 \oplus X_2 = 0) = \sum_{j=0}^{2^n-1} \left(\frac{1}{2^n} + \delta'_j\right)\left(\frac{1}{2^n} + \delta''_j\right) \tag{19}$$

where we have used the independence of X_1 and X_2 to determine each term of the summation. Now

$$Pr(X_1 \oplus X_2 = 0) = \frac{1}{2^n} + \frac{1}{2^n}\sum_{j=0}^{2^n-1}(\delta'_j + \delta''_j) + \sum_{j=0}^{2^n-1}\delta'_j\delta''_j. \tag{20}$$

Since $\sum_j Pr(X_1 = j) = 1$, we have $\sum_j \delta'_j = 0$. Similarly, $\sum_j \delta''_j = 0$. Hence, the middle term in (20) is 0 and since, by definition, $\epsilon = \max_j\{|\delta'_j|, |\delta''_j|\}$ then

$$\left|Pr(X_1 \oplus X_2 = 0) - \frac{1}{2^n}\right| \leq \epsilon_{max} \leq 2^n\epsilon^2 \tag{21}$$

as expected from the lemma for $k = 2$.

Induction Step: For our induction hypothesis, we shall assume that the lemma holds for the XOR of $k-1$ random n-bit vectors. Hence, letting ϵ' represent the upper bound on the bias for the sum of $k-1$ vectors, we have

$$\epsilon' \leq 2^{(k-2)n}\epsilon^{k-1} \tag{22}$$

and $\epsilon' < \epsilon$. Now we can use the same reasoning as in the base case if we consider X_1 to be the vector representing the sum of $k-1$ vectors and X_2 represents the k-th vector. Since now $|\delta'_j| \leq \epsilon'$ and $|\delta''_j| \leq \epsilon$ for all j, (18) follows straightforwardly from (20).

The proof of the lemma is completed by noting that

$$\epsilon_{max} \leq 2^{(k-1)n}\epsilon^k = (2^n\epsilon)^{k-1}\epsilon < \epsilon \tag{23}$$

if $\epsilon < 1/2^n$. □

Note that if $\epsilon \geq 1/2^n$, the right side of (18) evaluates to greater than or equal to $1/2^n$ and, hence, the expression of (18) is not useful in determining an upper bound on the bias that converges to zero as the number of vectors k increases. Since in our analysis the number of vectors will increase proportionally to the number of rounds in the cipher, it is desirable to consider scenarios in which $\epsilon < 1/2^n$ so that we can derive an upper bound on the bias, and consequently the information leakage, which is converging to zero.

B. A Lower Bound on Entropy

Consider an n-bit random vector Y, which has probabilities $Pr(Y = i) = 1/2^n + \alpha_i$ where i is an integer representation of Y with $0 \leq i \leq 2^n - 1$. The entropy of Y is given by

$$H(Y) = -\sum_{i=0}^{2^n-1} \left(\frac{1}{2^n} + \alpha_i\right) \log_2 \left(\frac{1}{2^n} + \alpha_i\right). \tag{24}$$

Consider now the expansion of $\log_2(1/2^n + \alpha_i)$ using a Taylor series:

$$\log_2 \left(\frac{1}{2^n} + \alpha_i\right) = -n + \lambda \left[\sum_{j=1}^{\infty} (-1)^{j+1} \cdot \frac{2^{jn}}{j} {\alpha_i}^j\right] \tag{25}$$

where $\lambda = \log_2 e$. Substituting (25) into (24) and using $\sum_i \alpha_i = 0$ gives

$$H(Y) = n - \lambda \sum_{j=2}^{\infty} \left[(-1)^j \frac{2^{(j-1)n}}{j(j-1)} \sum_{i=0}^{2^n-1} {\alpha_i}^j\right]. \tag{26}$$

Now assuming $|\alpha_i| \ll 1/2^n$, the $j = 2$ term dominates the summation so that

$$H(Y) \approx n - \lambda \cdot 2^{n-1} \sum_{i=0}^{2^n-1} {\alpha_i}^2. \tag{27}$$

Letting $\epsilon_{max} = \max_i |\alpha_i|$, the entropy of Y is bounded approximately by the following:

$$H(Y) \geq n - \lambda \cdot 2^{2n-1} \epsilon_{max}^2. \tag{28}$$

A similar expression can be derived for the entropy of Y conditioned on a specific value x of another random variable X. In this case, a bound on $H(Y|x)$ can be computed using (28) where ϵ_{max} is the maximum bias such that $|Pr(Y = i|x) - 1/2^n| \leq \epsilon_{max}$.

C. An Upper Bound on Information Leakage

An expression can also be developed for an upper bound on the average mutual information between an n-bit random vector Y and an m-bit random vector X. If Y and X are uniformly distributed, $H(Y) = n$ and $H(X) = m$. Using the lower bound derived for entropy in the previous section, we know

$$\min_{x \in \{0,1\}^n} H(Y|x) \geq n - \lambda \cdot 2^{2n-1} \epsilon_{max}^2 \tag{29}$$

where $\epsilon_{max} = \max_{i,x} |Pr(Y = i|x) - 1/2^n|$. Also, we note that $H(Y|X) \geq \min_x H(Y|x)$ and, hence, the average mutual information is upper bounded by I_{max} as below:

$$I(X;Y) \leq n - \min_{x \in \{0,1\}^n} H(Y|x) \leq I_{max} \tag{30}$$

where

$$I_{max} = \lambda \cdot 2^{2n-1} \epsilon_{max}^2 \tag{31}$$

with ϵ_{max} being the largest bias of the conditional probability over all values for Y and X.

If we consider X to represent m bits of plaintext and Y to represent n bits of ciphertext, we can now directly relate the upper bound on information leakage to a lower bound on the amount of data, N_S, required in an attack. Substituting (31) into (14) gives

$$N_S \geq \frac{\lambda \cdot 2^n}{4 \cdot I_{max}}. \tag{32}$$

As expected, an increase in the amount of information leakage (actually the upper bound on information leakage) lowers the amount of data required (actually the lower bound on N_S). However, to this point in our development we still do not have a clear picture of the relationship between the number of ciphertext bits involved in the leakage, n, and the upper bound on leakage I_{max}. We address this issue in the next section by modelling a bound on the probability bias ϵ_{max}, and subsequently information leakage, of a generalized Feistel cipher.

V. Modelling Information Leakage of a Feistel Cipher

In this section, we focus specifically on developing a model for determining a measure of the information leakage of a Feistel cipher. For reasons of clarity, we shall initially consider a model for the round function, followed by an analysis of the simplest case, referred to as the "restricted case", and eventually an analysis of the more difficult "general case".

The determination of the information leakage of a Feistel cipher involves the iterative computation of bounds on probabilities. The development of such bounds requires the assumption that the output bits of the round function in round i are independent of the output bits of the round function in round $i \pm 2j$ where j is an integer. This will allow us to use the generalized piling-up lemma.[2] Intuitively, for our analysis, the independence assumption should be a reasonable approximation because we will be restricting our analysis to modest sized subsets of plaintext and ciphertext bits and the large number of

[2]The applicability of the piling-up lemma is also appropriate in circumstances where we model the cipher to have random, independent, uniformly distributed subkeys and the subkey bits are mixed with the data using an XOR operation. This random subkey model is frequently assumed in the consideration of a cipher's resistance to linear cryptanalysis as an appropriate approximation to the actual situation of a fixed key. This topic is thoroughly discussed in [27].

remaining plaintext bits are assumed to be random and independent. The appropriateness of the independence assumption is further supported by experimental observations as detailed in Section VII-A.

A. Modelling the Round Function

In determining bounds on the information leakage of a cipher, we will inevitably require a bound on the probability of a value taken by a vector formed from a subset of ciphertext bits conditioned on a vector formed from a subset of plaintext bits. This will require determining a bound on the conditional probabilities of the round function of the form $Pr(Y^{(n)} = y^{(n)}|X^{(m)} = x^{(m)}) \equiv Pr(y^{(n)}|x^{(m)})$ where $Y^{(n)}$ represents a vector formed from n output bits and $X^{(m)}$ represents a vector formed from m input bits.

In our model of the round function, we do not consider a specific method of keying the round function, but rather simply assume that each function generated under a key has some fundamental random behaviour. Consider modelling the behaviour of a round function in a Feistel cipher for the scenario that a set of m input bits are fixed and a set of n output bits are targetted to be of interest. Let $L = N/2 - m$ represent the number of non-fixed input bits. In the generation of the round function for a particular subkey, we assume that 2^L values of the n output bits are randomly selected from a uniform distribution. These values correspond to the 2^L inputs derived from the selections for the L non-fixed input bits combined with the m fixed bits. We do not necessarily assume that the round function is a pseudo-random function generator as defined in [9]. Although a pseudo-random function generator satisfies the model, a round function can satisfy the model without being a perfect pseudo-random function generator. Our model would apply, for example, if we randomly generated an $N/2 \times N/2$ mapping for the round function f and keyed it by XORing $N/2$ subkey bits with the data at the input to the function.

The model for the round function can be viewed as equivalent in nature to an occupancy problem of randomly throwing a number of balls into a collection of bins. The number of balls is given by 2^L and the number of bins is given by 2^n. The number of balls in a particular bin is given by the binomial distribution:

$$Pr(k \text{ balls in a bin}) = \left(\begin{array}{c} 2^L \\ k \end{array} \right) \left(\frac{1}{2^n}\right)^k \left(1 - \frac{1}{2^n}\right)^{2^L - k}. \tag{33}$$

The mean number of balls in a bin is $\mu = 2^{L-n}$ and the variance of the number of balls in a bin is given by

$$\sigma^2 = 2^{L-n}(1 - 1/2^n) < 2^{L-n} = \mu \tag{34}$$

where the inequality approaches an equality for $n \gg 1$. Noting that the binomial distribution is approximately Gaussian under the conditions that mean $\mu \gg 0$ and the standard deviation $\sigma \ll \mu$, we can assume with high probability that the number of balls in a bin, k, will be within 8σ of the mean[3] if the total number of balls, 2^L, is large. That is, the condition

$$\mu - 8\sigma < k < \mu + 8\sigma \tag{35}$$

is highly probable if $\mu \gg 0$ and $\sigma \ll \mu$.

Now, the random variable value k is related to the conditional probability of getting a particular n-bit value $y^{(n)}$ for the output $Y^{(n)}$ given the m-bit fixed input value of $x^{(m)}$. In particular,

$$Pr(y^{(n)}|x^{(m)}) = k/2^L \tag{36}$$

implying

$$\frac{1}{2^n} - \frac{8\sigma}{2^L} < Pr(y^{(n)}|x^{(m)}) < \frac{1}{2^n} + \frac{8\sigma}{2^L} \tag{37}$$

with high probability. This results in the bias ϵ_f being bounded, with high probability, as in

$$\epsilon_f = \max_{x^{(m)},y^{(n)}} |Pr(y^{(n)}|x^{(m)}) - 1/2^n| \leq \frac{8\sigma}{2^L} < 2^{3-(L+n)/2}. \tag{38}$$

For $N = 64$, $L = N/2 - m = 32 - m$ and (38) becomes

$$\epsilon_f < 2^{-13+(m-n)/2} \tag{39}$$

for a round function that fits the model well.

In the derivation of the information leakage of the entire cipher, we will be using the generalized piling-up lemma based on the bound for the probability biases for the round function as given in (38) above. However, we must consider restrictions on the values of m and n in deriving ϵ_f for the round function so that ϵ_f does not exceed $1/2^n$ since the piling-up lemma requires $\epsilon_f < 1/2^n$. Therefore, we will be able to analyze a cipher as long as $(L+n)/2 - 3 > n$. Hence, $L - n > 6$ or

$$m + n < N/2 - 6. \tag{40}$$

For $N = 64$, we therefore require $m + n < 26$ in our model of the round function in order to be able to compute a useful upper bound on the information leakage of the cipher using the techniques in the following sections.

[3]The factor of 8 was chosen as an arbitrarily large enough value to ensure negligible probabilities in the tails of the distribution.

As an example cipher for which our analysis is applicable, we conjecture that our model of a round function fits well with the CAST-128 round function because of its non-degenerate nature: all input bits influence all output bits. Supporting experimental evidence is presented in Section VII-B. It should be noted though that the model would not work well for all ciphers. For example, DES has S-boxes with small output sizes and therefore does not have a non-degenerate round function. As a result, the input bits to the round function have large, localized influences on the output bits, implying a poor fit to the model.

It should also be noted that it is not necessary to use the model to apply the methodology used in the analysis of the following sections. All we really require for our approach to be useful is that the probability biases in a round function be bounded by a value of $\epsilon_f < 1/2^n$ for the values of m and n of interest. In a real cipher, it may be possible, using the characteristics of the round function, to determine a bound on the value of ϵ_f.

B. Modelling Multiple Rounds: Restricted Case

We now extend our analysis to the full Feistel cipher. We consider in the first instance the information leakage involving only n bits in the left half of the ciphertext, C_L, and n bits in the right half of the plaintext, P_R, such that the n bits of C_L are in the same relative bit positions as the n bits in P_R. (The arguments can also be applied when considering bits from only the right half of the ciphertext and the left half of the plaintext.) We refer to this as the *restricted case*.

Let a vector formed from a set of n bits of C_L be represented as the random vector $C_L^{(n)}$ and a particular value of $C_L^{(n)}$ be represented by $c_L^{(n)}$. In the restricted case, we are interested in the probability of the occurrence of the value $c_L^{(n)}$ given that the corresponding n bits of P_R are fixed at a value $p_R^{(n)}$. Hence, we represent this probability as $Pr(c_L^{(n)}|p_R^{(n)})$. The remaining $N-n$ bits of plaintext are assumed to be independent, uniformly distributed random variables.

Let the input to the round function of round i be X_i and the output of the round function be Y_i. From the definition of a general Feistel cipher given in Section II-A, a vector representing n bits of C_L can be derived from

$$C_L^{(n)} = P_R^{(n)} \oplus \sum_{i=1}^{r/2} Y_{2i}^{(n)} \tag{41}$$

where the large summation represents a bit-wise XOR on the vectors and it is assumed that the number of rounds r is even. In the expression, all vectors are formed from the same positions for the n bits. Note that the relationship between $C_L^{(n)}$ and the input

$P_R^{(n)}$ is determined by the sum of the $Y_{2i}^{(n)}$ vectors (i.e., the output of every second round function). Ideally for no information leakage, $\epsilon_{max} \rightarrow 0$, where

$$\epsilon_{max} = \max_{p_R^{(n)}, c_L^{(n)}} |Pr(c_L^{(n)}|p_R^{(n)}) - 1/2^n|. \tag{42}$$

We shall determine ϵ_{max} by using the generalized piling-up lemma, making use of the assumption that the bits of interest of all Y_{2i} and P_R are independent. In the use of expression (18), we must consider the maximum value for the probability bias $|Pr(y_{2i}^{(n)}|p_R^{(n)}) - 1/2^n|$. However, we assume that all variables on the right side of equation (41) are independent and, therefore,

$$Pr(y_{2i}^{(n)}|p_R^{(n)}) = Pr(y_{2i}^{(n)}) \tag{43}$$

where $Pr(y_{2i}^{(n)})$ is the probability of a particular value for n bits of the output of the round function in round $2i$ assuming a uniform distribution across all inputs to the round function. It can, in fact, be shown that $Y_2^{(n)}$ is independent of $P_R^{(n)}$; however we must rely on the appropriateness of our assumption of independence between all $Y_{2i}^{(n)}$ and $P_R^{(n)}$.

Using our model of a round function, we let

$$|Pr(y_{2i}^{(n)}) - 1/2^n| \leq \epsilon_f \tag{44}$$

where ϵ_f is determined from the round function model of the previous section with the number of fixed input bits to the round function given by $m = 0$.

The generalized piling-up lemma can now be used to determine a bound on the probability of $Pr(c_L^{(n)}|p_R^{(n)})$ resulting in

$$\epsilon_{max} \leq 2^{(r/2-1)n} {\epsilon_f}^{r/2} \tag{45}$$

for a cipher of r rounds.

For a cipher with a block size of N, from the development of the previous section, the probability bias ϵ_f can be bounded for all values of $n < N/2 - 6$ (since $m = 0$). Hence, an upper bound, I_{max}, on the information leakage $I(P_R^{(n)}; C_L^{(n)})$ can be determined for a large number of cases for n using (31). As well, this can be related to the amount of data that is required to distinguish the information leaked in the cipher, given by N_S in equation (14) or equivalently (32). For a cipher of block size $N = 64$, the results for several values of n have been computed and are presented in Table 1. The value of min N_S given in the table represents the lower bound on N_S computed from (14) or (32).

n	#rounds (r)	ϵ_{max}	I_{max}	$\min N_S$
1	2	8.63×10^{-5}	2.15×10^{-8}	2^{25}
	4	1.49×10^{-8}	6.41×10^{-16}	2^{50}
	6	2.57×10^{-12}	1.91×10^{-23}	2^{75}
	8	4.44×10^{-16}	5.69×10^{-31}	2^{100}
2	2	6.10×10^{-5}	4.30×10^{-8}	2^{25}
	4	1.49×10^{-8}	2.56×10^{-15}	2^{49}
	6	3.64×10^{-12}	1.53×10^{-22}	2^{73}
	8	8.88×10^{-16}	9.10×10^{-30}	2^{97}
4	2	3.05×10^{-5}	1.72×10^{-7}	2^{25}
	4	1.49×10^{-8}	4.10×10^{-14}	2^{47}
	6	7.28×10^{-12}	9.78×10^{-21}	2^{69}
	8	3.55×10^{-15}	2.33×10^{-27}	2^{91}
8	2	7.63×10^{-6}	2.75×10^{-6}	2^{25}
	4	1.49×10^{-8}	1.05×10^{-11}	2^{43}
	6	2.91×10^{-11}	4.00×10^{-17}	2^{61}
	8	5.68×10^{-14}	1.53×10^{-22}	2^{79}
12	2	1.91×10^{-6}	4.40×10^{-5}	2^{25}
	4	1.49×10^{-8}	2.69×10^{-9}	2^{39}
	6	1.16×10^{-10}	1.64×10^{-13}	2^{53}
	8	9.09×10^{-13}	1.00×10^{-17}	2^{67}
16	2	4.77×10^{-7}	7.04×10^{-4}	2^{25}
	4	1.49×10^{-8}	6.88×10^{-7}	2^{35}
	6	4.66×10^{-10}	6.72×10^{-10}	2^{45}
	8	1.46×10^{-11}	6.56×10^{-13}	2^{55}
24	2	2.98×10^{-8}	1.80×10^{-1}	2^{25}
	4	1.49×10^{-8}	4.51×10^{-2}	2^{27}
	6	7.45×10^{-9}	1.13×10^{-2}	2^{29}
	8	3.73×10^{-9}	2.82×10^{-3}	2^{31}

Table 1: Information Leakage for Restricted Case ($N = 64$)

As can be seen from the table, as expected, the upper bound on information leakage decreases and the lower bound on the amount of data required increases as the number of rounds increases. As the number of bits involved in the information leakage, n, increases, the lower bound on the amount of data required decreases. Considering ciphers of 8 rounds, for $n = 24$ the lower bound is small enough ($< 2^{64-n}$) that the cipher cannot be claimed to be secure against attacks based on information leakage. (This is not to say that the cipher is insecure; it just cannot be claimed that the cipher is secure.)

We emphasize that this is a very specific case, where (1) the number of plaintext bits equals the number of ciphertext bits, (2) the plaintext bits are in one half and the ciphertext bits are in the corresponding half, and (3) the n ciphertext bits correspond to

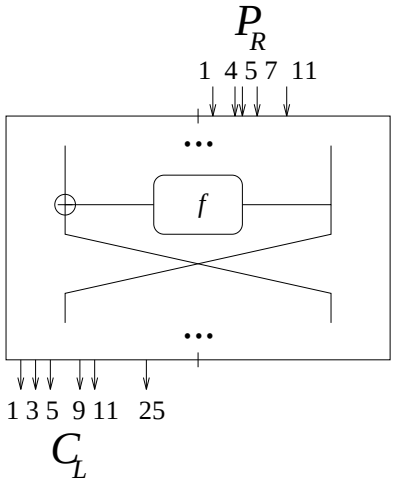

Figure 2: Example Scenario for Isolated-Half Case

exactly the same bit positions as the n plaintext bits.

C. Modelling Multiple Rounds: Isolated-Half Case

Next we consider the broader case of m plaintext bits and n ciphertext bits with $n \neq m$ but such that all bits are still in C_L for the ciphertext and P_R for the plaintext. (Similarly, we could consider C_R and P_L.) We refer to this as the *isolated-half case* since we still consider the bits of interest are confined to just one half of the ciphertext and the corresponding half of the plaintext. Let $P_R^{(n_0+n_1)} = [P_R^{(n_0)}, P_R^{(n_1)}]$ represent the vector derived from the set of plaintext bits and $C_L^{(n_1+n_2)} = [C_L^{(n_1)}, C_L^{(n_2)}]$ represent the vector derived from the set of ciphertext bits where n_0 represents the number of plaintext bits that are not aligned with the ciphertext bits, n_1 represents the number of bit positions common to both the plaintext and ciphertext vectors, and n_2 represents the number of ciphertext bits not aligned with the plaintext bits. When considering the information leakage, we consider the distribution of the values of the $n_1 + n_2$ ciphertext bits when the $n_0 + n_1$ plaintext bits are fixed and the remaining plaintext bits are randomly and independently selected.

As an example, consider the information leakage from bits $\{1, 4, 5, 7, 11\}$ in the plaintext right half to bits $\{1, 3, 5, 9, 11, 25\}$ in the ciphertext left half as shown in Figure 2. In this case, $n_0 = 2$, $n_1 = 3$, and $n_2 = 3$ with $P_R^{(n_0)} = [P_R[4], P_R[7]]$, $P_R^{(n_1)} = [P_R[1], P_R[5], P_R[11]]$, $C_L^{(n_1)} = [C_L[1], C_L[5], C_L[11]]$, and $C_L^{(n_2)} = [C_L[3], C_L[9], C_L[25]]$. As well, the n_2 bits of C_L may be considered aligned with n_2 bits of plaintext, represented as $P_R^{(n_2)}$, although these plaintext bits are not part of the information leakage expression.

We can modify equation (41) to highlight the component subsets of bits:

$$\begin{array}{l} C_L^{(n_1)} = P_R^{(n_1)} \oplus \sum_{i=1}^{r/2} Y_{2i}^{(n_1)} \\ C_L^{(n_2)} = P_R^{(n_2)} \oplus \sum_{i=1}^{r/2} Y_{2i}^{(n_2)}. \end{array} \tag{46}$$

As in the previous section, we assume that the bits of interest of all Y_{2i} and P_R are independent and, hence, it is evident from (46) that the $n_1 + n_2$ ciphertext bits are not directly influenced by the set of n_0 plaintext bits. (That is, the $n_1 + n_2$ ciphertext bits are not derived by XORing the n_0 plaintext bits with the outputs of round functions.) Hence, our independence assumption implies that there is no leakage of the n_0 bits of plaintext to the $n_1 + n_2$ ciphertext bits.[4] Also the n_2 ciphertext bits are influenced by the random, uniformly distributed bits $P_R^{(n_2)}$ which are independent of the set of $n_0 + n_1$ plaintext bits of interest. Hence, the entropy of the $n_1 + n_2$ ciphertext bits *conditioned* on the $n_0 + n_1$ plaintext bits is given by the entropy of the n_1 ciphertext bits conditioned on the n_1 plaintext bits plus the unconditional entropy of the n_2 ciphertext bits. That is,

$$H(C_L^{(n_1+n_2)}|P_R^{(n_0+n_1)}) = H(C_L^{(n_1)}|P_R^{(n_1)}) + n_2. \tag{47}$$

Hence,

$$I(P_R^{(n_0+n_1)}; C_L^{(n_1+n_2)}) = I(P_R^{(n_1)}; C_L^{(n_1)}). \tag{48}$$

An upper bound on the information leakage can therefore be computed as in the restricted case of the previous section.

D. Modelling Multiple Rounds: General Case

Finally, we must consider the completely general case of m plaintext bits and n ($\neq m$) ciphertext bits and any combination of bit positions, i.e., the bits are not constrained to only one half of the plaintext and ciphertext. We use the following notation to refer to bits involved in the information leakage expression:

n_{L0} - the number of bits in P_L not aligned with bits in C_R
n_{L1} - the number of aligned bits in P_L and C_R
n_{L2} - the number of bits in C_R not aligned with bits in P_L
n_{R0} - the number of bits in P_R not aligned with bits in C_L
n_{R1} - the number of aligned bits in P_R and C_L
n_{R2} - the number of bits in C_L not aligned with bits in P_R

[4] Although the n_0 plaintext bits must influence the values of the $n_1 + n_2$ ciphertext bits, the independence assumption is premised on the assumption that negligible information leakage would occur between the n_0 plaintext bits and the $n_1 + n_2$ ciphertext bits. For example, in the trivial case of $r = 2$, the influence of the n_0 plaintext bits is diffused through the two round functions and it is reasonable to expect the information leakage would be negligible in comparison to the $n_1 + n_2$ plaintext bits directly influencing the corresponding ciphertext bits.

Hence, in determining the information leakage we consider a total of $n_{L0}+n_{L1}+n_{R0}+n_{R1}$ plaintext bits to be fixed and are interested in the information leaked to $n_{R1}+n_{R2}+n_{L1}+n_{L2}$ bits of ciphertext, assuming that the non-fixed plaintext bits are random and independent. Our objective then is the consideration of

$$I(P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}; C_L^{(n_{R1}+n_{R2})}, C_R^{(n_{L1}+n_{L2})}). \tag{49}$$

We know

$$H(C_L^{(n_{R1}+n_{R2})}, C_R^{(n_{L1}+n_{L2})}) = n_{R1}+n_{R2}+n_{L1}+n_{L2} \tag{50}$$

and so need to consider the conditional entropy

$$H(C_L^{(n_{R1}+n_{R2})}, C_R^{(n_{L1}+n_{L2})} | P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}). \tag{51}$$

Similar to the reasoning in the previous section, we rationalize that the ciphertext bits counted in n_{R2} and n_{L2} are randomized by bits that are uniformly distributed and independent of the plaintext bits of interest. Hence,

$$\begin{aligned} &H(C_L^{(n_{R1}+n_{R2})}, C_R^{(n_{L1}+n_{L2})} | P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}) \\ &\quad = H(C_L^{(n_{R1})}, C_R^{(n_{L1})} | P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}) + H(C_L^{(n_{R2})}, C_R^{(n_{L2})}) \\ &\quad = H(C_L^{(n_{R1})}, C_R^{(n_{L1})} | P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}) + n_{R2} + n_{L2} \end{aligned} \tag{52}$$

and therefore

$$\begin{aligned} &I(P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}; C_L^{(n_{R1}+n_{R2})}, C_R^{(n_{L1}+n_{L2})}) \\ &\quad = I(P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}; C_L^{(n_{R1})}, C_R^{(n_{L1})}). \end{aligned} \tag{53}$$

However, we can no longer consider the influence of fixed plaintext bits in one half in isolation and we must consider the influence of some number of fixed bits of the plaintext right half when determining the information leakage from the plaintext left half and vice versa.

Consider the output of the first round function Y_1. According to our round function model, we assume that the randomness of Y_1 is generated by the $L_o = N/2 - (n_{R0}+n_{R1})$ non-fixed bits of the input to the round function. The bits of Y_1 are combined with the bits of P_L by the XOR operation and we are specifically interested in the bits $P_L^{(n_{L1})}$ (since $P_L^{(n_{L0})}$ does not directly affect the ciphertext bits of interest and as a result of our assumption of independence between output bits of every 2nd round function and plaintext bits). The round function model is based on 2^{L_o} balls being tossed into the $2^{n_{L1}}$ bins of interest. In the context of the notation used in Section V-A, $m = n_{R0}+n_{R1}$ and $n = n_{L1}$. From this reasoning we can compute a bound on the probability bias ϵ_f at the output of the round function and, consequently, determine the probability bias for the corresponding n_{L1} bits at the output of the XOR.

At the output of the round function of round 2, we are interested in the bias of the bits of Y_2 corresponding to the n_{R1} bits from P_R. To model the 2nd round function, since we are interested in an upper bound on the probability bias, we shall assume the worst case scenario, where the fixed $n_{L0}+n_{L1}$ bits of P_L constrain the input to the right half of round 2 to take on only 2^{L_e} values where $L_e = N/2 - (n_{L0} + n_{L1})$. Clearly, this is pessimistic since the $n_{L0} + n_{L1}$ bits of P_L that are fixed, are randomized when XORed with the corresponding bits of Y_1, resulting in X_2 capable of taking on $2^{N/2}$ values. However, X_2 is not necessarily uniformly distributed. So the distribution for the vector of output bits of interest from the 2nd round function Y_2 is modelled by considering tossing 2^{L_e} balls into $2^{n_{R1}}$ bins and consequently bounding the probability bias.

Extrapolating the analysis from round 2 to find an upper bound on the information leakage, we use the pessimistic approach and model all odd rounds of the cipher by considering 2^{L_o} balls tossed into $2^{n_{L1}}$ bins and all even rounds by considering 2^{L_e} balls tossed into $2^{n_{R1}}$ bins. Using this model, the upper bound on the biases for the outputs of all the round functions can be determined. We can then apply the piling-up lemma to derive an upper bound on the biases of the overall cipher for the left and right halves of ciphertext. That is, the following can be computed:

$$\epsilon_{Lmax} = \max |Pr(c_L^{(n_{R1})} | p_L^{(n_{L0}+n_{L1})}, p_R^{(n_{R0}+n_{R1})}) - 1/2^{n_{R1}}| \tag{54}$$

and

$$\epsilon_{Rmax} = \max |Pr(c_R^{(n_{L1})} | p_L^{(n_{L0}+n_{L1})}, p_R^{(n_{R0}+n_{R1})}) - 1/2^{n_{L1}}|. \tag{55}$$

In order to compute a bound on the information leakage of the combined left and right halves of the ciphertext, we use the approximation[5]

$$\begin{aligned} & I(P_L^{(n_{L0}+n_{L1})}, P_R^{(n_{R0}+n_{R1})}; C_L^{(n_{R1})}, C_R^{(n_{L1})}) \\ & \quad \approx I(P_R^{(n_{R0}+n_{R1})}, P_L^{(n_{L0}+n_{L1})}; C_L^{(n_{R1})}) \\ & \qquad + I(P_R^{(n_{R0}+n_{R1})}, P_L^{(n_{L0}+n_{L1})}; C_R^{(n_{L1})}). \end{aligned} \tag{56}$$

Hence, we may employ (54) and (55) to determine an upper bound on the information leakage of each ciphertext half separately and then combine them as per (56) to determine the upper bound on the information leakage. Subsequently, a lower bound on the amount of data required in an attack based on information leakage can be computed. A number of sample values of plaintext and ciphertext bit sets have been considered and are presented in Table 2 for a cipher of block size $N = 64$ and $r = 8$ rounds.

[5]Expression (56) is true with equality for ciphers which XOR subkeys to data in each round under the assumption of random, independent, uniformly distributed subkeys.

$n_{L0} - n_{L1} - n_{R0} - n_{R1}$	ϵ_{Lmax}, ϵ_{Rmax}	I_{max}	$\min N_S$
$0-1-0-1$	1.78×10^{-15}	1.82×10^{-29}	2^{96}
$1-1-1-1$	7.11×10^{-15}	2.91×10^{-28}	2^{92}
$3-1-3-1$	1.14×10^{-13}	7.46×10^{-26}	2^{84}
$0-2-0-2$	1.42×10^{-14}	4.66×10^{-27}	2^{90}
$2-2-2-2$	2.27×10^{-13}	1.19×10^{-24}	2^{82}
$1-3-1-3$	4.55×10^{-15}	1.91×10^{-23}	2^{80}
$0-4-0-4$	9.09×10^{-13}	3.06×10^{-22}	2^{78}
$4-4-4-4$	2.33×10^{-10}	2.00×10^{-17}	2^{62}
$8-4-8-4$	5.96×10^{-8}	1.31×10^{-12}	2^{46}
$4-8-4-8$	9.54×10^{-7}	8.60×10^{-8}	2^{38}
$8-8-8-8$	2.44×10^{-4}	5.64×10^{-3}	2^{22}

Table 2: Information Leakage for General Case ($N = 64$, $r = 8$)

Note that in the analysis of the general case, as per (40), in order for the bound on the probability bias of our modelled round function to be small enough for convergence as the number of rounds increases, for $N = 64$, $n_{L0} + n_{L1} + n_{R1} < 26$ and $n_{L1} + n_{R0} + n_{R1} < 26$. In Table 2, we have assumed that $n_{L2} = n_{R2} = 0$: since no information is leaked through the n_{L2} and n_{R2} bits, there is no value in including them in an attack. In fact, including them only increases the number of plaintexts required in the attack.

From the table, we observe the general trend of a decrease in the lower bound on the amount of data required for an attack, N_S, as the number of bits involved in the plaintext increases. For the scenarios with a large number of involved bits we find the lower bound on N_S is too small to be able to claim security against exploiting information leakage. However, due to the nature of the analysis, it seems very likely that the small value of the lower bound is the result of the looseness of the bound, rather than a sincere security risk.

VI. Application of the Model to CAST-128

In this section, we consider the results of the model applied to a cipher with the parameters of CAST-128 [4]. We conjecture that the model of the round function applies well to the CAST-128 round function and, hence, the results in this section accurately represent security bounds for CAST-128. This is discussed further below and is supported with experimental evidence in Section VII-B.

CAST-128 is a Feistel cipher of block size $N = 64$ and 16 rounds. It has a 32×32 round

function generated by dividing the data block into 4 bytes and using the values of the bytes to look-up the 32-bit output values of four 8×32 S-boxes. The outputs of the S-boxes in the CAST-128 round function are combined to produce a 32-bit output using 3 operations: addition modulo-2^{32}, subtraction modulo-2^{32}, and 32-bit XOR. There are 37 subkey bits mixed into each round: 32 bits are combined with the incoming data using either addition, subtraction, or XOR and this is followed by a key-dependent rotation determined from the remaining 5 bits of subkey.

There are actually 3 defined round functions in CAST-128 with the round number determining which function is applied. The functions differ in which operations are used for combining the outputs of the S-boxes. As an example, round function f_1 with a 32-bit input X and a 32-bit output Y is defined to be:

$$\begin{aligned} D &= ((K_m + X) \leftarrow K_r \\ Y &= ((S_1[D_1] \oplus S_2[D_2]) + S_3[D_3]) - S_4[D_4] \end{aligned} \tag{57}$$

where S_i represents S-box i, D_i represents the 8-bit input to S-box i, and K_m and K_r represent the combining and rotation subkeys, respectively. The operators "$\oplus$", "$+$", and "$-$" represent XOR, addition, and subtraction, respectively, and the notation "$V \leftarrow U$" represents a left rotation of vector V by U bits. Note that D and K_m represent 32-bit vectors and K_r is 5 bits in length. Round functions f_2 and f_3 are similar in nature.

A notable property of the CAST-128 round function is that all input bits influence all output bits. This occurs because the 8-bit input of each S-box influences a 32-bit output which is combined with the other 32-bit S-box outputs to produce the 32-bit output of the round function. This is a strength of the CAST cipher which is not present in DES, for example, and is an important consideration when modelling the round function. In the CAST-128 round function, as in our round function model, we expect there can be some small amount of information leakage from any set of input bits to any set of output bits. (Contrast this to the DES round function, where it is trivally true that input bits have no influence on most output bits, implying the information leakage is 0 for some sets of inputs and outputs but for other sets the information leakage is high.)

In order to examine the security bounds for the CAST-128 cipher, we computed the information leakage with $N = 64$ and $r = 16$ for all valid combinations of n_{L0}, n_{L1}, n_{R0}, and n_{R1} such that in the round function model, (40) is not violated. We observed that the smallest lower bound for any number of plaintext and ciphertext bits occurs for scenarios where the number of ciphertext bits is 1 and the number of plaintext bits is maximized in the plaintext half not corresponding to the half of the ciphertext bit. That is, $n_{L1} = 1$ and $n_{R1} = 0$ with n_{R0} maximized is a scenario which gives the smallest lower

n_P	Smallest min N_S (n_{L0}-n_{L1}-n_{R0}-n_{R1})		Restricted min N_S (n_{L0}-n_{L1}-n_{R0}-n_{R1})	
2	2^{192}	(0-1-1-0)	2^{193}	(0-2-0-0)
4	2^{176}	(0-1-3-0)	2^{179}	(0-4-0-0)
6	2^{160}	(0-1-5-0)	2^{165}	(0-6-0-0)
8	2^{144}	(0-1-7-0)	2^{151}	(0-8-0-0)
10	2^{128}	(0-1-9-0)	2^{137}	(0-10-0-0)
12	2^{112}	(0-1-11-0)	2^{123}	(0-12-0-0)
14	2^{96}	(0-1-13-0)	2^{109}	(0-14-0-0)
16	2^{80}	(0-1-15-0)	2^{95}	(0-16-0-0)
18	2^{64}	(0-1-17-0)	2^{81}	(0-18-0-0)
20	2^{48}	(0-1-19-0)	2^{67}	(0-20-0-0)
22	2^{32}	(0-1-21-0)	2^{53}	(0-22-0-0)

Table 3: Smallest Lower Bound on N_S ($N = 64$, $r = 16$)

bound on N_S for a given number of plaintext bits n_P. (The number of bits of n_{L0} does not influence the leakage and resulting bound on N_S.) So, in consideration of (40), if $n_P < 26$, then $n_{L0} = 0$, $n_{L1} = 1$, $n_{R0} = n_P - 1$, and $n_{R1} = 0$ minimizes the bound on N_S and, if $n_P \geq 26$, then $n_{L0} = n_P - 25$, $n_{L1} = 1$, $n_{R0} = 24$, and $n_{R1} = 0$ minimizes the bound on N_S. (We could also reverse the role of the left and right halves and the results would be the same.) It is not surprising that these scenarios minimize the bound on N_S because such scenarios maximize the bias of the round function ϵ_f.

In Table 3, we present, as a function of the number of plaintext bits involved in the information leakage expression n_P, the minimum computed bounds on N_S and the corresponding scenario for a cipher of $N = 64$ and $r = 16$ as in CAST-128. For comparison, we also give the value of the bound on N_S for the restricted case of the same number of bits. As the number of bits increases, the smallest lower bound and the restricted case lower bound diverge somewhat. In an attack based on information leakage, n_P plaintext bits are fixed, allowing a maximum of 2^{N-n_P} plaintexts available to mount the attack. For $n_P \leq 20$, the amount of data required exceeds 2^{64-n_P}, implying that the cipher cannot be attacked using information leakages involving 20 or fewer bits.

VII. Experimental Results

In this section, we present a limited set of experimental results supporting the applicability of the model and the conjecture that CAST-128 is well represented by the model.

n	#rounds	Theoretical ϵ_{max}	Average Experimental ϵ_{max}	Largest Experimental ϵ_{max}
1	2	3.906×10^{-1}	3.906×10^{-1}	3.946×10^{-1}
	4	3.052×10^{-1}	3.052×10^{-1}	3.089×10^{-1}
	6	2.384×10^{-1}	2.385×10^{-1}	2.428×10^{-1}
	8	1.863×10^{-1}	1.865×10^{-1}	1.915×10^{-1}
2	2	1.953×10^{-1}	1.953×10^{-1}	2.005×10^{-1}
	4	1.526×10^{-1}	1.526×10^{-1}	1.573×10^{-1}
	6	1.192×10^{-1}	1.193×10^{-1}	1.246×10^{-1}
	8	9.313×10^{-2}	9.327×10^{-2}	9.871×10^{-2}
3	2	9.766×10^{-2}	9.767×10^{-2}	10.315×10^{-2}
	4	7.629×10^{-2}	7.692×10^{-2}	8.090×10^{-2}
	6	5.960×10^{-2}	5.964×10^{-2}	6.407×10^{-2}
	8	4.657×10^{-2}	4.663×10^{-2}	5.160×10^{-2}

Table 4: Experimental Results for Restricted Case ($N = 24$)

A. Verification of Analysis on a Small Cipher

In this section, we consider a small scale Feistel cipher. Specifically, we examine a cipher with a block size of $N = 24$ and with a round function generated as a randomly selected 12×12 mapping. The small block size makes it possible to consider probability biases that are large enough to be verified using several thousand test encryptions.

Our specific objective in the experiments was to verify the correctness of the approach to computing the probability biases used in bounding the information leakage of the cipher, essentially verifying the applicability of the piling-up lemma. The first set of experiments involved verification of the restricted case discussed in Section V-B. Following these experiments, several scenarios from the general case discussed in Section V-D are examined.

1) Restricted Case

To verify the restricted case scenarios, our experimental approach generated a number of ciphers by randomly selecting a round function and then modifying it in specific bits to create probability biases that were large enough to be easily measured at the cipher output. The round function in each cipher was not keyed but the results would be identical to the scenario where we XOR a subkey to the data at the input to the round function. For a specific set of 3 input bits, the randomly generated round function was modified so that the probability bias from 1/8 was exactly $\epsilon_f = 25/256$ for all values. Similarly, for a 2 bit set, the probability bias from 1/4 of $\epsilon_f = 25/128$ was established

for all values; for a particular bit, the probability bias from 1/2 was $\epsilon_f = 25/64$.

With each generated cipher, many random test encryptions were executed for the different cases with a set of either $n = 1$, 2, or 3 bits held fixed in the plaintext right half while the remaining plaintext bits were randomly selected. The fixed bits corresponded to the bits in the 2nd round at the output of the round function that were generated to give the probability biases ϵ_f above.

The experimental probabilities associated with the corresponding ciphertext bits were determined for 1000 different ciphers as a function of the number of rounds. The average biases and the largest biases across all ciphers were determined for the different scenarios and these are tabulated in Table 4 and compared to the values computed according to equation (45), with ϵ_f as appropriate from above. Note that the experimental results compare very favourably with the theoretical results derived for the restricted case of Section V-B. The experimental average bias is very close to the expected value in all cases; the largest bias is marginally greater than the expected value which is not surprising given that these values are derived by picking the largest from 1000 experimental sets of data. This gives us encouragement that the assumption of independence used in the analysis (that allowed the application of the piling-up lemma) appears to give a reasonable approximation of the behaviour of the cipher. The analysis for the isolated-half case of Section V-C follows straightforwardly from the restricted case and, hence, we do not present any experimental verification of such scenarios here.

2) General Case

The general case of Section V-D relies exclusively on the "balls-in-bins" approach to modelling the behaviour of the round function and so cannot be verified by a specially modified round function as we did for the restricted case. However, we have observed experimental results consistent with the upper bound derived in the analysis of Section V-D. Specifically, we have randomly generated 1000 ciphers, each based on a randomly generated 12×12 round function (without modified bits as before) and, using a number of test plaintexts with some fixed bits (n_{L1} bits in P_L and n_{R1} in P_R), have examined the probability bias associated with the corresponding ciphertext bits. In Table 5, the resulting experimental probability biases are compared against their theoretical upper bound values derived using the methodology of Section V-D. Note that $n_{L1} + n_{R1} < N/2 - 6 = 6$ as per the discussion of Section V-A. As expected, the experimental results fall below the theoretical bounds. The theoretical upper bounds on the bias do not appear to be very tight.

$n_{L1} - n_{R1}$	#rounds	Theor. ϵ_{Lmax}	Exp. ϵ_{Lmax}	Theor. ϵ_{Rmax}	Exp. ϵ_{Rmax}
$1-1$	2	1.250×10^{-1}	3.286×10^{-2}	1.250×10^{-1}	4.003×10^{-2}
	4	3.125×10^{-2}	2.193×10^{-3}	3.125×10^{-2}	2.579×10^{-3}
	6	7.813×10^{-3}	2.021×10^{-3}	7.813×10^{-3}	2.034×10^{-3}
	8	1.953×10^{-3}	1.300×10^{-3}	1.953×10^{-3}	1.531×10^{-3}
$2-2$	2	1.250×10^{-1}	2.309×10^{-2}	1.250×10^{-1}	4.781×10^{-2}
	4	6.250×10^{-2}	2.593×10^{-3}	6.250×10^{-2}	2.993×10^{-3}
	6	3.125×10^{-2}	2.434×10^{-3}	3.125×10^{-2}	2.331×10^{-3}
	8	1.563×10^{-2}	2.622×10^{-3}	1.563×10^{-2}	2.455×10^{-3}
$2-3$	2	8.839×10^{-2}	2.476×10^{-2}	1.768×10^{-1}	6.748×10^{-2}
	4	6.250×10^{-2}	2.722×10^{-3}	1.250×10^{-1}	3.597×10^{-3}
	6	4.419×10^{-2}	2.408×10^{-3}	8.839×10^{-2}	3.287×10^{-3}
	8	3.125×10^{-2}	2.288×10^{-3}	6.250×10^{-2}	2.736×10^{-3}
$3-1$	2	2.500×10^{-1}	3.253×10^{-2}	6.250×10^{-2}	2.869×10^{-2}
	4	1.250×10^{-1}	2.211×10^{-3}	3.125×10^{-2}	2.332×10^{-3}
	6	6.250×10^{-2}	2.298×10^{-3}	1.563×10^{-2}	2.181×10^{-3}
	8	3.125×10^{-2}	2.692×10^{-3}	7.813×10^{-3}	2.524×10^{-3}

Table 5: Experimental Results for General Case ($N = 24$)

B. Applicability of Round Function Model to CAST-128

It is not conceivable to verify the applicability of the model to the CAST-128 round function for all combinations of appropriately sized input and output sets (i.e., so that $m + n < 26$). It is particularly difficult since the 32×32 size of the CAST-128 round function makes the probability biases very small, with typically several million tests of round function values to determine experimentally. So, we focussed our tests on a small number of cases for m and n and used 10 randomly selected subsets of m plaintext bits and n ciphertext bits for each case. We used function f_1 in our experiments with K_m and K_r both being all zeroes. Table 6 illustrates the largest bias values that were found in experiments using 10^9 test values for the different scenarios and compares these against the corresponding bound used in the model. As conjectured, no experimental values exceeded the bound suggested by the model.

VIII. Conclusions

In this paper, we have examined information leakage between bits of plaintext and ciphertext as a measure of cipher security. Information leakage has been related to cryptanalysis such that a lower bound on the amount of data required to attack a cipher by exploiting information leakage can be determined from an upper bound on the information leakage

$m-n$	Theoretical Maximum ϵ_f	Largest Experimental ϵ_f
$1-1$	1.221×10^{-4}	9.151×10^{-5}
$2-2$	1.221×10^{-4}	6.775×10^{-5}
$8-4$	4.883×10^{-4}	2.807×10^{-4}
$8-8$	1.221×10^{-4}	6.517×10^{-5}
$10-6$	4.883×10^{-4}	2.662×10^{-4}
$11-1$	3.906×10^{-3}	5.709×10^{-4}
$12-4$	1.953×10^{-3}	6.988×10^{-4}
$12-12$	1.221×10^{-4}	6.599×10^{-5}

Table 6: Experimental Probability Biases for CAST-128

of the cipher. The theoretical equivalence of zero information leakage and immunity to linear cryptanalysis has also been established.

The paper has also presented a model of a general Feistel cipher that can be applied to get a general sense of the amount of information leakage (in the form of an upper bound) as a function of the number of rounds, block size, and selection of bits involved in the leakage. The usefulness of the model is supported with experimental evidence and the applicability of the model to the CAST-128 cipher has been discussed. As an example use of the model, for the CAST-128 cipher, we predict that the information leakage from 20 or fewer bits of plaintext is such that an attack would require more than the available plaintexts - an impossible requirement. For scenarios of information leakage involving more than 20 bits, it is not possible to conclude that the CAST-128 cipher is immune to attacks based on information leakage. However, it is important to note that this is a limitation of the analysis in this paper which gives lower bounds on the amount of data required, rather than a negative reflection on the security of CAST-128.

The model is most useful for information leakage involving small numbers of plaintext and ciphertext bits. It is likely that, due to the conservative nature of the round function model and the slow convergence of the upper bound on the probability bias for the cipher when round function bias is close to $1/2^n$, the analysis provides a loose upper bound on information leakage, particularly when the number of involved bits becomes large. Therefore, the amount of data required for an attack involving a large number of bits is a correspondingly loose lower bound. Future work could involve finding better methods for bounding the information leakage involving larger numbers of bits.

Acknowledgements

The author wishes to thank the anonymous referee that indicated the relationship between the definition of (m, n)-immunity to linear cryptanalysis and the concept of correlation immunity.

References

[1] C.E. Shannon, "Communication Theory of Secrecy Systems", *Bell System Technical Journal*, vol. 28, pp. 656-715, 1949.

[2] 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.

[3] National Bureau of Standards, "Data Encryption Standard (DES)", *Federal Information Processing Standard 46*, 1977.

[4] C.M. Adams, "Constructing Symmetric Ciphers Using the CAST Design Procedure", *Designs, Codes, and Cryptography*, vol. 12, no. 3, pp. 283-316, 1997.

[5] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 2nd edition, Prentice-Hall, 1999.

[6] E. Roback and M. Dworkin, "Conference Report: First Advanced Encryption Standard (AES) Candidate Conference, Ventura, CA, August 20-22, 1998", *Journal of Research of National Institute of Standards and Technology*, vol. 104, no. 1, pp. 97-105, 1999.

[7] E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like cryptosystems", *Journal of Cryptology*, vol. 4, no. 1, pp. 3-72, 1991.

[8] M. Matsui, "Linear Cryptanalysis Method for DES Cipher", *Advances in Cryptology - EUROCRYPT '93*, Lecture Notes in Computer Science 765, Springer-Verlag, pp. 386-397, 1994.

[9] M. Luby and C. Rackoff, "How to Construct Pseudorandom Permutations from Pseudorandom Functions", *SIAM Journal of Computing*, vol. 17, no. 2, pp. 373-386, 1988.

[10] R. Anderson and E. Biham, "Two Practical and Provably Secure Block Ciphers: BEAR and LION", *Fast Software Encryption*, Lecture Notes in Computer Science 1039, Springer-Verlag, pp. 113-120, 1996.

[11] S. Lucks, "Faster Luby-Rackoff Ciphers", *Fast Software Encryption*, Lecture Notes in Computer Science 1039, Springer-Verlag, pp. 189-203, 1996.

[12] C. Adams, "The SHADE Cipher: An Efficient Hash-Function Based Feistel Network", *presented at IEEE 1997 Canadian Conference on Electrical and Computer Engineering*, St. John's, Canada, 1997.

[13] L. Knudsen and K. Nyberg, "Provable Security Against a Differential Attack", *Journal of Cryptology*, vol. 8, no. 1, pp. 27-37, 1995.

[14] M. Matsui, "New Structure of Block Ciphers with Provable Security Against Differential and Linear Cryptanalysis", *Fast Software Encryption*, Lecture Notes in Computer Science 1039, Springer-Verlag, pp. 205-218, 1996.

[15] S. Vaudenay, "Provable Security for Block Ciphers by Decorrelation", *Symposium on Theoretical Aspects of Computer Science - STACS 98*, Paris, France, Lecture Notes in Computer Science 1373, Springer-Verlag, pp. 249-275, 1998.

[16] H.M. Heys and S.E. Tavares, "Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis", *Journal of Cryptology*, vol. 9, no. 1, pp. 1-19, 1996.

[17] J. Lee, H.M. Heys, and S.E. Tavares, "Resistance of a CAST-Like Encryption Algorithm to Linear and Differential Cryptanalysis", *Designs, Codes, and Cryptography*, vol. 12, no. 3, pp. 267-282, 1997.

[18] R. Forré, "Methods and Instruments for Designing S-boxes", *Journal of Cryptology*, vol. 2, no. 3, pp. 115-130, 1990.

[19] 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 - EUROCRYPT '91*, Lecture Notes in Computer Science 547, Springer-Verlag, pp. 352-367, 1991.

[20] 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 - CRYPTO '92*, Lecture Notes in Computer Science 740, Springer-Verlag, pp. 260-279, 1993.

[21] M. Zhang, S.E. Tavares, and L.L. Campbell, "Information Leakage of Boolean Functions and its Relationship to Other Cryptographic Criteria", *Proceedings of 2nd ACM Conference on Computer and Communications Security*, Fairfax, Virginia, pp. 156-165, 1994.

[22] A.M. Youssef and S.E. Tavares, "Information Leakage of a Randomly Selected Boolean Function", *Proceedings of Information Theory and its Applications II*, Lecture Notes in Computer Science 1133, Springer-Verlag, pp. 41-52, 1996.

[23] G.Z. Xiao and J.L. Massey, "A Spectral Characterization of Correlation-Immune Combining Functions", *IEEE Transactions on Information Theory*, vol. IT-34, no. 3, pp. 569-571, 1988.

[24] L. Brynielsson, "A Short Proof of the Xiao-Massey Lemma", *IEEE Transactions on Information Theory*, vol. IT-35, no. 6, pp. 1344, 1989.

[25] T. Siegenthaler, "Correlation-Immunity of Nonlinear Combining Functions for Cryptographic Applications", *IEEE Transactions on Information Theory*, vol. IT-30, no. 5, pp. 776-780, 1984.

[26] B. Chor, O. Goldreich, J. Hastad, J. Freidmann, S. Rudich, and R. Smolensky "The Bit Extraction Problem or t-Resilient Functions", *Proceedings of 26th IEEE Symposium on Foundations of Computer Science*, pp. 396-407, 1985.

[27] K. Nyberg, "Linear Approximation of Block Ciphers", *Advances in Cryptology - EUROCRYPT '94*, Lecture Notes in Computer Science 950, Springer-Verlag, pp. 439-444, 1995.

List of Figures

List of Tables