Linear dependencies in product ciphers

H. M. Gustafson A. N. Pettitt
School of Mathematics

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

Queensland University of Technology
GPO Box 2434, Brisbane QLD 4001

Abstract

In this paper we will survey three forms of statistical dependency found in block ciphers. Each dependency is based on some notion of linearity, forming the basis of several attacks which are particularly applicable to product ciphers. We consider linear relationships between the plaintext and ciphertext bits, using elementary arguments from linear algebra, and then using linear relationships under real number addition based on canonical correlation analysis. Linear structures [4] are also examined, which are a form of linearity that leads to degeneracy in the key, meaning that certain bits do not affect the ciphertext. We show that most functions are not expected to have a linear structure, though even partial linearity in this respect leads to a powerful attack known as differential cryptanalysis. Lastly, we consider linear approximation as a cryptanalytic tool, and present the recent linear cryptanalysis due to Matsui on the Data Encryption Standard (DES).

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

1 Introduction

A block cipher E is a family of encryption functions that acts on n characters of data (usually bits), with typical values of n being in the range of 64 to 2048. Examples of block ciphers include LUCIFER, DES and RSA [6]. The two major properties to be considered in the design of a block cipher are (a) to minimize the statistical relationship between the plaintext and ciphertext, and (b) to strongly suggest that the key cannot be recovered in time that is significantly less than the expected cost of exhaustive key search. First, we observe that (a) does not imply (b), in that a strong pseudorandom function is not necessarily resistant to cryptographic attacks. For example, it is known that sequences produced by linear feedback registers can be selected to satisfy the randomness postulates of Golumb [8], but the initial register contents and tapping information can be recovered by inspecting a small amount of ciphertext [6]. Second, (b) is not a proof of security since this would essentially be a solution to a major open problem in computational complexity theory [7]. Often (b) will be an accreditation given to the cipher after a thorough, and most likely protracted, examination of its properties by cryptanalysts; even so, it is only a conjecture that the cipher is in fact secure. On this point, the history of DES is informative. When released in the mid seventies, IBM stated that 17 years of research had been consumed in the design and analysis of the algorithm. To this day, all reported 'weaknesses' of DES are either unlikely to occur (for example, selecting a so-called weak key [20]), or require such substantial computational resources to take advantage of (for example, differential cryptanalysis [3]) At present, and probably always, DES is considered to be a very strong cipher with an 'unfortunately small' key (56 bits).

It is possible to apply a large number of statistical tests to a block cipher E. Some of these tests are adapted directly from the theory of random sequences [11, 14], such as tests related to runs and frequency distributions, and sequence complexity such as the Lempel-Ziv and linear complexities [10]. Specific tests for block ciphers E are designed to measure the sensitivity of E to changes in the plaintext or key, say by complementing one or two bit positions in each parameter. Such tests are said to examine the 'avalanche' properties of the cipher [28] where a small change in one parameter causes a correspondingly large change in the ciphertext. One limitation of statistical tests is that, in general, while they identify anomalies, the tests do not suggest methods to remedy an anomaly. Similarly, statistical tests do not necessarily indicate how an anomaly can be developed into a weakness. It is this essentially 'nonconstructive' nature of statistical tests which limits their use in the design and analysis of block ciphers.

In this paper we will survey three forms of statistical dependency found in block ciphers each based on some notion of linearity. These attacks will apply particularly to product ciphers [6] which are block ciphers built from smaller components such as look-up tables (S-boxes S) and permutations (P-boxes P). A basic product cipher is shown in Figure 1. We begin in §2 by considering linear relationships between the plaintext and ciphertext bits, using elementary arguments from linear algebra. We also investigate the application of canonical correlation analysis to cryptanalysis, which examines linear relationships under real number addition. In

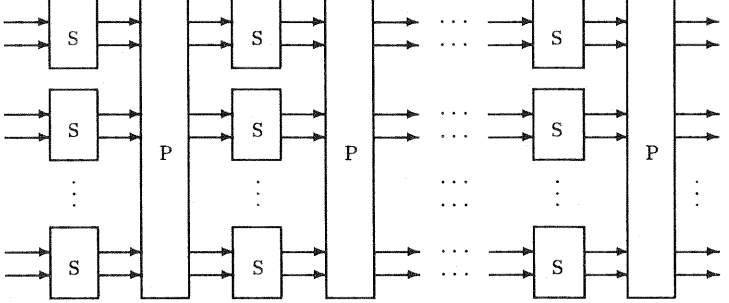

Figure 1: A general product cipher showing substitutions and transpositions.

§3 we consider linear structures [4], a form of linearity that leads to degeneracy in the key (here degeneracy means that when the influence of the key is modelled as a boolean function f, certain keys bits do not affect the function). We show that most functions are not expected to have linear structures, though even partial linearity in this respect leads to a powerful attack known as differential cryptanalysis. Lastly, we consider a linear approximation as a cryptanalytic tool, and present the recent attack of Matsui [18] on the Data Encryption Standard (DES) [22].

In this paper we assume that the block ciphers investigated combine n input bits $P = p_1, p_2, \ldots, p_n \in Z_2^n$, known as the plaintext, under the action of an m-bit key $K = k_1, k_2, \ldots, k_m \in Z_2^m$ to give n output bits $C = c_1, c_2, \ldots, c_n \in Z_2^n$, known as the ciphertext. We will say that for each fixed key K, a block cipher E is an n-bit function if $E : Z_2^n \rightarrow Z_2^n$.

2 Plaintext/Ciphertext Linearity

Dependencies that exist between subsets of plaintext, ciphertext and key bits could decrease the cost of searching the keyspace. In one of the the worst cases the cipher is a linear mapping, allowing the cipher to be totally determined after inspecting a relatively small amount of ciphertext. Several such dependencies, including the linearity just mentioned, can be detected through statistical methods [10, 15], such as the χ^2-test. In the next few sections we examine several form of linearity in block ciphers.

2.1 The Standard Linear Relationship

Suppose that we have a block cipher with block length of n and an encryption function denoted by E. This block cipher is *affine* if for each key there exists an $n \times n$ matrix A and $n \times 1$ vector b both over Z_2, such that $C = AP \oplus b$ for all plaintext-ciphertext pairs $(P, C = E(P))$. Testing for this relationship involves encrypting four plaintext blocks: the null block, P_0, two random blocks P_1 and P_2, and their modulo

two addition, $P_3 = P_1 \oplus P_2$. If the modulo-two addition of the ciphertexts obtained on encryption of P_1 and P_2 equals the modulo two addition of the ciphertexts obtained on encryption of P_0 and P_3 then the cipher may satisfy the affine property and, if so, an equation of the form $C = AP \oplus b$ is true for this cipher. If a block cipher satisfies the affine property then the cipher would not be recommended for use in encryption as knowledge of $n+1$ plaintext-ciphertext pairs would be sufficient to obtain A and b and hence the information on any further plaintext blocks would easily be obtained from the intercepted ciphertext only.

The set of all nonsingular $k \times k$ matrices over the field F_q is called the general linear group and is denoted as $GL(k, F_q)$. As it is known that the order of $GL(k, F_q)$ is $q^{(k^2-k)/2} \cdot \prod_{i=1}^{k}(q^i - 1)$ [17], then the number of linear mappings from $E : Z_2^n \rightarrow Z_2^n$ is

$$2^{(n^2-n)/2} \cdot \prod_{i=1}^{n}(2^i - 1) \;\leq\; 2^{2n^2}. \tag{1}$$

Since as n tends to infinity, $2^{2n^2}/2^n! = 0$, it follows that most n-bit to n-bit mappings are not linear. We may further enquire as to the probability that any particular ciphertext bit c_i is a linear function of some subset of the plaintext. That is, does there exist some i, $1 \leq i \leq n$ such that $c_i = f_i(p_1, p_2, \ldots, p_n) = a_0 + \sum_{j=1}^{n} a_j p_j$, $a_j \in Z_2$. Let $\mathcal{NL}^{n,n}$ be the set of invertible n-bit to n-bit mappings for which no bit of the mapping is described by an affine boolean function.

Theorem 2.1 (Gordon and Retkin [9]) $|\mathcal{NL}^{n,n}|$ is given as

$$|\mathcal{NL}^{n,n}| \;=\; 2^n! + \sum_{k=1}^{n}(-1)^k \cdot 2^k \cdot \binom{n}{k} \cdot (2^{n-k}!)^{2^k} \cdot \prod_{i=0}^{k-1}(2^n - 2^i). \tag{2}$$

Corollary 2.1 $2^n! \sim |\mathcal{NL}^{n,n}|$. □

Then with high probability if E is selected at random then E will be nonlinear in all ciphertext bits. Another form of linearity is *partial linearity*. A function f is said to be *partially linear*, or simply *p-linear*, if there exists a subset $Y = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, $1 \leq k \leq n$, of the variables such that

$$f(X) \;=\; g(x'_1, x'_2, \ldots, x'_{n-k}) \;+\; \sum_{1 \leq j \leq k} m_j x_{i_j} \tag{3}$$

where $\{x'_1, x'_2, \ldots, x'_{n-k}\} = \{x_1, x_2, \ldots, x_n\} - Y$, $m_j \in Z_2$, $1 \leq j \leq k$. These functions were previously studied by Beale and Monaghan [2] where they were called *linear-in* functions, and are discussed further in §3.

2.2 Gaussian Elimination

The previous test detected if each ciphertext bit may be written as a linear combination of plaintext bits. Further, we may enquire if some *subset* of the ciphertext bits can be written as a linear combination of plaintext bits. For $X = x_1 x_2 \cdots x_n \in Z_2^n$

let $X[i_1, i_2, \ldots, i_a]$ denote the XOR sum $x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_a}$ where $1 \leq i_1 < i_2 < \cdots < i_a \leq n$ and $1 \leq a \leq n$. Consider an equation of the form

$$P[i_1, i_2, \ldots, i_a] \oplus C[j_1, j_2, \ldots, j_b] \oplus a_0 \;=\; 0 \tag{4}$$

where $a_0 \in Z_2$, which indicates that the sum of a subset of plaintext bits with a subset of ciphertext bits is constant (corrected to 0 by the a_0 term). Dependencies of the form in (4) can be tested as follows. Select $(2n+1)$ plaintext/ciphertext pairs (P_i, C_i), $1 \leq i \leq 2n+1$, where $P_i = p_{i1}, p_{i2}, \ldots, p_{in}$ and $C_i = c_{i1}, c_{i2}, \ldots, c_{in}$. Then construct a $(2n+1) \times (2n+1)$ matrix A matrix where the first row is all ones, and column i contains the bits of plaintext P_i followed by the bits of ciphertext C_i. If when performing row reductions on the matrix A a row of all zeros is encountered, then a dependency of the form in (4) must exist.

If E has a dependency of the form in (4), then the test will report it (true dependency); on the other hand, even if E has no dependency of the form in (4), the test may report a dependency for the given sample of plaintext/ciphertext pairs (P_i, C_i) (false dependency). We could sample N matrices $A_1, A_2, \ldots, A_N$ and true dependencies would be found in each A_i if they existed. However, we would like to know how large N should be before any false dependencies induced by the plaintext/ciphertext sample would be unlikely to occur in all N sample matrices. To answer this question, observe that a matrix of full rank has no dependencies. We will make the assumption that a cipher E which has no true dependencies when sampled produces matrices A_i that are random over Z_2 (except for the first row). The probability that a random $k \times k$ matrix B has full rank is

$$q \;=\; \prod_{i=0}^{k-1} \left(1 - 2^{i-k}\right). \tag{5}$$

Let A' be defined as the matrix obtained from A by adding the first row of A to all other rows that have a 1 in the first column, and then deleting the first row and column from the resulting matrix. Clearly, if A has full rank then the $(2n \times 2n)$ matrix A' will also have full rank. ¿From (5), the probability that A' has full rank is $q = 0.2887$ when $n = 64$. Then the probability that at least one matrix in a random sample of N such matrices will have full rank is $1 - (1-q)^N$. Thus by solving $1 - (1-q)^N = p$ we are confident that in a sample of N matrices, the probability of producing at least one matrix of full rank is p. For example, when $n = 64$, a sample of 21 matrices has a probability of 99.9% to yield a matrix with full rank.

2.3 Linear Relationship Under Real Number Addition

Another method for examining linear relationships in block ciphers is to apply *Canonical Correlation analysis* as was first suggested by Carlisle Adams in his PhD thesis [1, p. 91]. This method investigates linear relationships under real number addition between two variables X and Y that are expressed as linear combinations of experimental observations (given below). Canonical analysis may be employed to determine the *best linear relationship* that exists between the X and Y variables.

This technique was originally developed by Hotelling [12, p.321], and in our case will involve calculating coefficients $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ plus an associated *canonical correlation*, λ_j, which measures the extent of the linear correlation between X and Y. More general information on canonical correlation analysis can be found in the book of Cooley and Lohnes [5, p.168].

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

$$Y_i = \alpha_{j1}p_{i1} + \alpha_{j2}p_{i2} + \cdots + \alpha_{jn}p_{in} \quad (6)$$

$$X_i = \beta_{j1}c_{i1} + \beta_{j2}c_{i2} + \cdots + \beta_{jn}c_{in} \quad (7)$$

The coefficient vectors $\boldsymbol{\alpha}_j = \alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jn}$ and $\boldsymbol{\beta}_j = \beta_{j1}, \beta_{j2}, \ldots, \beta_{jn}$ are calculated so that the corresponding correlation between the variables X and Y is maximized. The $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ vectors contain the weights of each bit position in the resulting canonical variables. The analysis requires the calculation of three $n \times n$ covariance matrices: R_{CC} measuring correlation between ciphertext bits, R_{PP} measuring correlation between plaintext bits, and R_{CP} measuring correlation between ciphertext and plaintext bits. If $R_{CC}[i,j]$ is the entry for the ith row and jth column, $1 \leq i, j \leq n$, then

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

$$\bar{C}_i = \frac{1}{N} \cdot \sum_{k=1}^{N} c_{ki}$$

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

$$R_{PP}[i,j] = \frac{1}{N-1} \cdot \sum_{k=1}^{N} (p_{ki} - \bar{P}_i)(p_{kj} - \bar{P}_j)$$

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

$$\bar{P}_i = \frac{1}{N} \cdot \sum_{k=1}^{N} p_{ki}.$$

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

$$(R_{PP}^{-1} \cdot R_{PC} \cdot R_{CC}^{-1} \cdot R_{CP} - \lambda_j I)\boldsymbol{\alpha}_j = 0 \quad (8)$$

where λ_j is the eigenvalue corresponding to the vector of weights $\boldsymbol{\alpha}_j$ subject to the condition that $\boldsymbol{\alpha}_j{}^T \cdot R_{PP} \cdot \boldsymbol{\alpha}_j = 1$. The corresponding vector of coefficients $\boldsymbol{\beta}_j$ is determined from the equation $\boldsymbol{\beta}_j = R_{CC}^{-1} \cdot R_{CP} \cdot \boldsymbol{\alpha}_j \cdot \sqrt{\lambda_j}$. Each value of λ_j determines a canonical correlation which measures the strength of a linear relationship $Y = aX + b$ corresponding to the set of points (X_i, Y_i), determined by substituting $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ in (6) for the sample of N plaintext-ciphertext pairs chosen.

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

Our analysis of this method shows that its application to block ciphers can be used to determine the existence of equal Hamming weights between subsets of plaintext and ciphertext positions in a cipher, with 100% correlation. This occurs, for example, in a transposition of plaintext to ciphertext bit positions. Canonical Correlation analysis aims to find a relationship between plaintext and ciphertext so that some part of the plaintext may be determined from an intercepted ciphertext. Unless the cipher exhibits the properties to yield 100% correlation, different samples of plaintext-ciphertext pairs will yield different canonical correlations λ_j, different coefficient vectors $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ and different linear equations $Y = aX + b$, for the same key.

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

3 Linear structures

A divide-and-conquer attack on the keyspace of a cipher is a method for partitioning the key bits into $d > 1$ distinct sets $w_1, w_2, \ldots, w_d$ such that each set w_i can be searched independently. If such a partition can be found then the cost of testing all possible keys becomes $O(2^{|w^*|})$ steps where $w^* = \max_{1 \leq i \leq d} w_i$, rather than $O(2^{|w_1|+|w_2|+\cdots+|w_d|})$ steps by obvious methods. Such a partition will exist if, for example, a known subset of the ciphertext depends on only k out of m key bits, will permit the key to be recovered in approximately $2^k + 2^{m-k}$ steps. We see that if $k \approx m/2$ then the key can be recovered in time which is approximately the square root of the time to perform exhaustive search. We will examine a class of boolean

functions, known as functions with linear structures, that admit divide-and-conquer attacks of this type. These functions have been used by Chaum and Evertse [4] to perform an attack on DES that is faster than exhaustive search when DES is reduced to less than 8 rounds. In what follows, we will represent an n-bit boolean function f as a polynomial $f(X) \in Z_2[x_1, x_2, \cdots, x_n]$, called the Algebraic Normal Form (ANF) of f.

Recall that p-linear functions were defined as

$$f(X) = g(x'_1, x'_2, \ldots, x'_{n-k}) + \sum_{1 \le j \le k} m_j x_{i_j}. \tag{9}$$

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

Lemma 3.1 (Lai [16]) Let $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$ be a set of linearly independent linear structures for the n-bit function f, where $1 \le k \le n$. Then there exists an $n \times n$ matrix M with coefficients over Z_2 such that if $g(X) = g(x_1, x_2, \ldots, x_n) = f((x_1, x_2, \ldots, x_n)M)$ then the ANF of $g(x_1, x_2, \ldots, x_n)$ is given as

$$g(X) = x_1 m_1 + x_2 m_2 + \cdots + x_k m_k + g(x_{k+1}, x_{k+2}, \ldots, x_n) \tag{10}$$

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

□

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

Proof. By Lemma 3.1 let $\mathbf{b_i} = \mathbf{e_i}$, $1 \le i \le k$, without loss of generality. However it follows from (10) that there are 2^k ways to choose the m_i, and $2^{2^{n-k}}$ ways to choose the $(n-k)$-bit function g. □

Thus if f is a function that has linear structures $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$, an appropriate basis change for Z_2^n transforms f into a p-linear function. The cryptanalyst can take advantage of the linear structures in f if some of the m_i in (10) are zero, which will eliminate the influence of some variables (possibly key bits) on the ciphertext.

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

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

Define M as the matrix

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{11}$$

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

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

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

Let $\mathcal{LS}^n$ be the set of n-bit boolean functions that have a linear structure $b \neq 0$. O'Connor and Klapper [21] have shown that most functions do not have linear structure, and in particular, that

$$\lim_{n\to\infty} |\mathcal{LS}^n|/((2^n - 1) \cdot 2^{2^{n-1}+1}) = 1. \tag{12}$$

3.1 Differential Cryptanalysis

Let $\oplus$ denote the addition of two elements from Z_2 and let $+$ denote the addition of two elements from Z_2^n. Observe that if $\mathbf{b}$ is a linear structure in f then either $f(X) \oplus f(X + \mathbf{b}) = 0$ with probability 1, or $f(X) \oplus f(X + \mathbf{b}) = 1$ with probability 1. We then say that inputs of difference $\mathbf{b}$ in f lead to output difference $\mathbf{d} \in \{0, 1\}$ with probability 1. However consider finding vectors $\mathbf{b}$ such that $f(X) \oplus f(X + \mathbf{b}) = c$ with probability p. In the same way, for $S : Z_2^r \to Z_2^s$, consider finding vectors $\mathbf{b} \in Z_2^r$ and $\mathbf{d} \in Z_2^s$ such that $S(X) + S(X + \mathbf{b}) = \mathbf{d}$ with some probability p. Unlike linear structures, pairs of input/output differences $\mathbf{b}, \mathbf{d}$ can always be found for a mapping S, though the probability p of the difference equation being true may be small. Attacks based on (highly) probable input/output difference pairs are collectively referred to as *differential cryptanalysis*. This attack has been popularized by Biham and Shamir [3] who have applied it to a wide range of cryptosystems, especially those that combine the key with the ciphertext using exclusive-or.

4 Affine approximation

When the cryptanalyst cannot find any direct linear relationships in a cipher, it may be possible to derive information about the key by considering linear approximation. To proceed further we will require several other definitions. The distance between two n-bit functions f and g is defined as

$$d(f, g) = |\{\ f(a) \neq g(a)\ :\ a \in Z_2^n\ \}| \tag{13}$$

which is the number of function values they disagree in. Further information about the properties of a boolean function can be gained by considering its Walsh transform [13]. For an n-bit boolean function f, the Walsh transform $F(\omega)$ of f at the point $\omega \in Z_2^n$ is defined as

$$F(\omega) = \sum_{X \in Z_2^n} f(X) \cdot (-1)^{X \cdot \omega} \tag{14}$$

where $X \cdot \omega$ is the dot product of X and ω. The Walsh transform can be computed in $O(n2^n)$ steps [13].

Let $\mathcal{A}^n = \{ f \mid f = a_0 + \sum a_i x_n,\ a_i \in Z_2 \}$ be the set of n-bit affine functions, observing that $|\mathcal{A}^n| = 2^{n+1}$. The functional nonlinearity of the n-bit function f is the minimum distance of f to $\mathcal{A}^n$, or the smallest number of function values that must be changed to make f affine. Formally, for an n-bit function f we define the functional nonlinearity $\delta(f, \mathcal{A}^n)$ of f to be

$$\delta(f, \mathcal{A}^n) \quad = \quad \min_{g \in \mathcal{A}^n} d(f, g). \tag{15}$$

It follows that $\delta(f, \mathcal{A}^n)$ is an integer bound as $0 \leq \delta(f, \mathcal{A}^n) \leq 2^{n-1}$; the lower bound follows from the possibility that f is affine, and the upper bound from the observation that if $\delta(f, g) > 2^{n-1}$ then $\delta(f, g \oplus 1) < 2^{n-1}$. Let g be called a best affine approximator (BAA) to f if $g \in \mathcal{A}^n$ and $d(f, g) = \delta(f, \mathcal{A}^n)$. We note that a best affine approximator is not unique in general.

Rueppel [26] characterized the distance of an n-bit function f to the set $\mathcal{A}^n$ as

$$\delta(f, \mathcal{A}^n) \quad = \quad 2^{n-1} - \max_{\omega \in Z_2^n} \frac{|\hat{F}(\omega)|}{2}. \tag{16}$$

Each invocation of the Walsh transform $\hat{F}(\omega)$ requires $O(n2^n)$ operations [13], which implies that determining $\delta(f, \mathcal{A}^n)$ via (16) will cost $O(n2^{2n})$ operations. Using (16) Meier and Staffelbach [19] were able to prove that for even n, the bent functions [25] attained the maximum possible distance from the set of linear functions, and also the maximum possible distance from the set of linear structures. A boolean function f is called *bent* if $|\hat{F}(\omega)| = 2^{n/2}$ for all $\omega \neq \mathbf{0} \in Z_2^n$ [25]. It then follows from (16) that for bent functions $\delta(f, \mathcal{A}^n) = 2^{n-1} - 2^{\frac{n}{2}-1}$. Let $\mathcal{B}^n$ be the set of n-bit bent functions, n even. At present there is no known closed form or tight bound for $|\mathcal{B}^n|$. According to Preneel *et al.* [24] the best known bound for the number of bent functions is

$$(2^{2^{\frac{n}{2}}})2^{\frac{n}{2}!} \ < \ |\mathcal{B}^n| \ < \ 2^{2^{n-1}+\frac{1}{2}\cdot\binom{n}{n/2}}$$

from which it follows that an 8-bit function is bent with probability less than 2^{-90}. If the nonlinear order of a function is restricted to be 2, then the number of such functions that are bent is $\prod_{i=0}^{n/2-1}(2^{2i-1} - 1) \cdot 2^{2i}$ [27].

4.1 Affine approximation of general mappings

We now address the problem of evaluating the nonlinearity of a set of functions $F = [f_1, f_2, \ldots, f_m]$, particularly when the functions are grouped as the outputs of an S-box. S-boxes are lookup tables which map m_1 input bits to m_2 output bits. Let an (m_1, m_2)-bit S-box $S : Z_2^{m_1} \rightarrow Z_2^{m_2}$ be realized by the m_2-tuple of m_1-bit functions $F_S = [f_1, f_2, \ldots, f_{m_2}]$. Pieprzyk and Finkelstein [23] have proposed expressions for measuring the nonlinearity of mappings S based on maximizing the individual nonlinearities of the f_i for those cases where S is a permutation. However, there is no accepted general metric for measuring the nonlinearity of S, but intuitively, any

such measure would be based on the nonlinearities of the f_i. Consider the following two measures of the nonlinearity $\delta(S)$ of S [23]:

$$\delta(S) = \min_i \delta(f_i, \mathcal{A}^m), \delta(f_i^{-1}, \mathcal{A}^m) \tag{17}$$

$$\delta(S) = \frac{\sum_{i=1}^{m} \delta(f_i, \mathcal{A}^m) + \delta(f_i^{-1}, \mathcal{A}^m)}{2m} \tag{18}$$

where S^{-1} is realized by $F_S^{-1} = [f_1^{-1}, f_2^{-1}, \ldots, f_m^{-1}]$. Pieprzyk and Finkelstein [23] have observed that under the definition in (17), all the permutations in the S-boxes of DES attain the same nonlinearity but vary under the definition given in (18).

Example 4.1 Consider the S-box $S3$ from DES which is listed below with its BAA $L(S3)$, which is a mapping where $m_1 = 6$ and $m_2 = 4$. The balance property is preserved in that $|L(S3)^{-1}(X)| = 4$, for all $X \in Z_2^4$, but observe that each row is no longer a permutation of the integers $\{0, 1, \ldots, 15\}$.

$$S3 = \begin{bmatrix} 15 & 1 & 8 & 14 & 6 & 11 & 3 & 4 & 9 & 7 & 2 & 13 & 12 & 0 & 5 & 10 \\ 3 & 13 & 4 & 7 & 15 & 2 & 8 & 14 & 12 & 0 & 1 & 10 & 6 & 9 & 11 & 5 \\ 0 & 14 & 7 & 11 & 10 & 4 & 13 & 1 & 5 & 8 & 12 & 6 & 9 & 3 & 2 & 15 \\ 13 & 8 & 10 & 1 & 3 & 15 & 4 & 2 & 11 & 6 & 7 & 12 & 0 & 5 & 14 & 9 \end{bmatrix}$$

$$L(S3) = \begin{bmatrix} 14 & 0 & 1 & 15 & 4 & 10 & 11 & 5 & 5 & 11 & 10 & 4 & 15 & 1 & 0 & 14 \\ 7 & 9 & 8 & 6 & 13 & 3 & 2 & 12 & 12 & 2 & 3 & 13 & 6 & 8 & 9 & 7 \\ 8 & 6 & 7 & 9 & 2 & 12 & 13 & 3 & 3 & 13 & 12 & 2 & 9 & 7 & 6 & 8 \\ 1 & 15 & 14 & 0 & 11 & 5 & 4 & 10 & 10 & 4 & 5 & 11 & 0 & 14 & 15 & 1 \end{bmatrix}.$$

The mapping $L(S3)$ is realized by the 4 affine functions

$$\begin{aligned} g_1 &= x_2 + x_4 + x_6 \\ g_2 &= x_1 + x_2 + x_3 + x_4 + x_5 + 1 \\ g_3 &= x_1 + x_4 + x_5 + 1 \\ g_4 &= x_2 + x_3 + x_4 + x_5 + x_6 + 1. \end{aligned}$$

The Hamming distances between the individual table entries is given as

$$w(S3 \oplus L(S3)) = \begin{bmatrix} 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 & 2 & 1 & 0 & 1 & 1 & 3 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 2 & 2 & 0 & 1 & 0 & 1 & 1 & 3 \\ 2 & 3 & 1 & 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 & 3 & 0 & 3 & 1 & 1 \end{bmatrix}.$$

Over 70% of the table entries have an error of 1 bit or less. □

5 Linear Cryptanalysis

We will now give a short exposition on a new method for cryptanalyzing DES based on linear approximation due to Matsui [18]. The basis of the attack is finding approximate linear relationship between certain bits in the plaintext, ciphertext and

key. Recall that, for $X = x_1 x_2 \cdots x_n \in Z_2^n$, $X[i_1, i_2, \ldots, i_a]$ denotes the XOR sum $x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_a}$ where $1 \leq i_1 < i_2 < \cdots < i_a \leq n$ and $1 \leq a \leq n$. Let P be a plaintext, C its ciphertext, and K the key used to encrypt P. Consider an equation of the form

$$P[i_1, i_2, \ldots, i_a] \oplus C[j_1, j_2, \ldots, j_b] \quad = \quad K[k_1, k_2, \ldots, k_c] \tag{19}$$

where the LHS is equal to the RHS with some probability q^*. This means that if we fix the key K and consider all possible plaintexts, then XORing certain subsets of the plaintext and ciphertext bits equals a certain XORed subset of the key bits with probability p. Intuition would suggest that if the bit subsets are selected randomly then the probability of (19) being true should be close to $1/2$. Matsui has shown that $K[k_1, k_2, \ldots, k_c]$ can be determined accurately using the maximum likelihood method if a sufficient amount N_L of known ciphertext is available. In particular, if the approximation in (19) is correct with probability q^*, the attack is expected to be successful 98% of the time when $N_L \approx |q^* - 1/2|^{-2}$. The result of the attack is the knowledge of one bit of information concerning the key, namely the value of $K[k_1, k_2, \ldots, k_c]$.

We will now consider how the approximation in (19) is found. Assume that the round function F of a product block cipher being attacked is fixed, and that a subkey K_i is XORed with the current ciphertext during the operation of F at round i. An approximation (19) for r rounds is found by determining several linear approximations (or linearizations) $\tau_1, \tau_2, \cdots, \tau_r$ to the F function, with the property that $\sum_i \tau_i \pmod 2$ only involves plaintext, ciphertext and key bits as unknowns. That is, all terms involving input or output to internal rounds of the cipher that cannot be represented as plaintext or ciphertext cancel. For DES Matsui has shown that such τ_i can be found by considering linearizations of the S-boxes. That is, by finding linear dependencies between the input and outputs to an S-box, a dependency of the form in (19) can be determined.

6 Conclusion

In this paper we have examined several forms of linearity as applied to cryptanalysing block ciphers. We began with considering how to detect if the ciphertext was a linear transformation of the plaintext, or if a similar relationship holds between proper subsets of the plaintext and ciphertext. Direct enumeration of $GL(k, F_q)$ shows that most invertible mappings E are not linear; further, if E maps n-bits to n-bits, then Theorem 2.1 showed that it is also unlikely that any output bit of E is described by a linear function. The notion of linear dependency is extended by using canonical correlation in §2, but this approach appears only to be useful in detecting permutation mappings. Any attack attempting to exploit correlation due to linearity merely by observing a large number of plaintext-ciphertext pairs is unlikely succeed. That is, designing statistical tests to detect linearity without taking into account the internal structure of the cipher are unlikely to detect any correlation.

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

References

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

[2] M. Beale and M. F. Monaghan. Encryption using random boolean functions. In H. J. Beker and F. C. Piper, editors, *Cryptography and Coding*, pages 219–230. Clarendon Press, 1989.

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

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

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

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

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

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

[9] J. Gordon and H. Retkin. Are big S-boxes best? In T. Beth, editor, *Cryptography, proceedings, Burg Feuerstein*, pages 257–262, 1982.

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

[11] H. Gustafson, E. P. Dawson, and W. Caelli. Comparison of block ciphers. *Advances in Cryptology, AUSTCRYPT 90, Lecture Notes in Computer Science, vol. 453, J. Seberry and J. Piepryzk eds., Springer-Verlag*, pages 208–220, 1990.

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

[13] M. G. Karpovsky. *Finite Orthogonal series in the design of digital devices.* John Wiley and Sons, 1976.

[14] D. E. Knuth. *The Art of Computer Programming : Volume 2, Seminumerical Algorithms.* Addsion Wesley, 1981.

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

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

[17] R. Lidl and H. Neiderreiter. *Introduction to finite fields and their applications.* Cambridge University Press, 1986.

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

[19] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. *Advances in Cryptology, EUROCRYPT 89, Lecture Notes in Computer Science, vol. 434, J.-J. Quisquater, J. Vandewalle eds., Springer-Verlag*, pages 549–562, 1990.

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

[21] L. J. O'Connor and A. Klapper. Algebraic nonlinearity and its applications to cryptography. *accepted to the Journal of Cryptology*, August, 1993.

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

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

[24] B. Preneel, W. Van Leekwijck, L. Van Linden, R. Govaerts, and J. Vandewalle. Propagation characteristics of boolean functions. *Advances in Cryptology, EUROCRYPT 90, Lecture Notes in Computer Science, vol. 473, I. B. Damgård ed., Springer-Verlag*, pages 161–173, 1991.

[25] O. S. Rothaus. On bent functions. *Journal of Combinatorial Theory (A)*, 20:300–305, 1976.

[26] R. A. Rueppel. *Design and Analysis of Stream Ciphers.* Springer–Verlag, 1986.

[27] N. J. A. Sloane and F. J. MacWilliams. *The Theory of Error Correcting Codes.* North–Holland Publishing Company, 1977.

[28] A. F. Webster and S. E. Tavares. On the design of S-boxes. *Advances in Cryptology, CRYPTO 85, H. C. Williams ed., Lecture Notes in Computer Science, vol. 218, Springer-Verlag*, pages 523–534, 1986.

(Received 14/12/93)