J. Cryptology (1995) 8: 67–86

Journal of
CRYPTOLOGY

On the Distribution of Characteristics in Bijective Mappings

Luke O'Connor*
Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Communicated by Adi Shamir

Received 9 December 1992 and revised 11 October 1993

Abstract. Differential cryptanalysis is a method of attacking iterated mappings based on differences known as characteristics. The probability of a given characteristic is derived from the XOR tables associated with the iterated mapping. If π is a mapping $\pi: Z_2^m \to Z_2^m$, then for each ΔX, $\Delta Y \in Z_2^m$ the XOR table for π gives the number of input pairs of difference $\Delta X = X + X'$ for which $\pi(X) + \pi(X') = \Delta Y$.

The complexity of a differential attack depends upon two properties of the XOR tables: the density of zero entries in the table, and the size of the largest entry in the table. In this paper we present the first results on the expected values of these properties for a general class of mappings π. We prove that if $\pi: Z_2^m \to Z_2^m$ is a bijective mapping, then the expected size of the largest entry in the XOR table for π is bounded by $2m$, while the fraction of the XOR table that is zero approaches $e^{-1/2} = 0.60653$. We are then able to demonstrate that there are easily constructed classes of iterated mappings for which the probability of a differential-like attack succeeding is very small.

Key words. Differential cryptanalysis, Iterated mapping, Product cipher.

1. Introduction

Differential cryptanalysis is a statistical attack popularized by Biham and Shamir [3], [5] that has been applied to a wide range of cryptosystems including LUCIFER, DES, FEAL, REDOC, and Kahfre [7], [8], [10], [17], [18], [25]. The attack is universal in that it can be used against any cryptographic mapping which is constructed from iterating a fixed round function (compare this with the universality of the birthday paradox against hash functions). For this reason the

*The author is presently employed by the Distributed System Technology Center, Brisbane, Australia. Correspondence should be sent to ISRC, QUT Gardens Point, 2 George Street, GPO Box 2434, Brisbane Queensland 4001, Australia; email: oconnor@sleet.fit.qut.edu.au.

differential attack must be considered one of the most general cryptanalytic attacks known to date. The main shortcoming of differential cryptanalysis is that large amounts of chosen-ciphertext may be required to determine the key, which will not be possible in most practical circumstances. Nevertheless, differential cryptanalysis has caused the revision and redesign of several iterated mappings [1], [5], [6], [19] and together with linear cryptanalysis are the only known attacks which can theoretically recover DES keys in time less than the expected cost of exhaustive search [4]. Importantly, the method has shown that the security of DES is not significantly increased if independent subkeys are used.

We give a brief description of differential cryptanalysis with reference to product ciphers, though any iterated mapping could be used. For a product cipher E that consists of R rounds, let $E_r(X, K)$ be the ciphertext of the plaintext X under the key K for r rounds, $1 \le r \le R$, where $E_R(X, K) = E(X, K) = C$ is the mapping for X. Let $\Delta C(r) = E_r(X, K) + E_r(X', K)$ be the difference between the ciphertexts of two plaintexts X, X' after r rounds where $1 \le r \le R$. An r-round *characteristic* is defined as an $(r + 1)$-tuple $\Omega_r(\Delta X, \Delta Y_1, \Delta Y_2, \ldots, \Delta Y_r)$ where ΔX is a plantext difference, and the ΔY_i are ciphertext differences. A plaintext pair X, X' of difference ΔX is called a *right pair* with respect to a key K and a characteristic $\Omega_r(\Delta X, \Delta Y_1, \Delta Y_2, \ldots, \Delta Y_r)$ if, when the pair X, X' is encrypted, $\Delta C(i) = \Delta Y_i$ for $1 \le i \le r$. That is, the characteristic correctly predicts the ciphertext differences at each round. The characteristic Ω_r has probability p^{Ω_r} if a fraction p^{Ω_r} of the plaintext pairs of difference ΔX are right pairs. On the other hand, if X, X' is not a right pair, then it is said to be a *wrong pair* (with respect to the characteristic and the key). A table which records the number of pairs of difference ΔX that give the output difference ΔY for a mapping π is called the *XOR table distribution* of π. A characteristic $\Delta X, \Delta Y$ is said to be *impossible* for π if its corresponding XOR table entry is zero. Also a characteristic is said to be nonzero if $w(\Delta X) > 0$, where $w(\cdot)$ is the Hamming weight function.

Assume that we wish to determine the subkey K_R that is being used in round R. The method of differential cryptanalysis proceeds as follows:

Step 1. Find a highly probably r-round characteristic $\Omega_r(\Delta X, \Delta Y_1, \Delta Y_2, \ldots, \Delta Y_r)$ which gives (partial) information about the input and output differences of the round mapping F at round R.

Step 2. Uniformly select a ciphertext pair X, X' with difference ΔX and encrypt this pair, assuming that X, X' is a right pair. Determine candidate subkeys $K'_1, K'_2, \ldots, K'_d$ such that each K'_i could have caused the observed output difference. Increment a counter for each candidate subkey K'_i.

Step 3. Repeat Step 2 until one subkey K'_R is distinguished as being counted significantly more often than other subkeys. Take K'_R to be the actual subkey.

If X, X' is a right pair, then one of the candidate subkeys $K'_1, K'_2, \ldots, K'_d$ is the actual subkey K_R, and K_R will be counted for each right pair. On the other hand, if X, X' is a wrong pair, then we assume that the candidate keys are distributed uniformly over the set of possible subkeys for the round, and K_R will

be counted with small probability. Similarly, we assume that any key other than the actual subkey will also be counted infrequently. It is then natural to define the *complexity* of a differential cryptanalysis to be the number of encrypted plaintext pairs of a specified difference required to determine the key or subkey. From experiments on restricted versions of DES, Biham and Shamir [3] found that the complexity of the attack was approximately c/p^{Ω^*}, where p^{Ω^*} is the probability of the characteristic being used, and c is a constant bound as $2 < c < 8$.

A variant of the attack is to perform Step 2 using only a subset of the subkey bits which could be counted for that round. For example, in DES each subkey is 48 bits in length which could possibly require 2^{48} counters to record the individual frequencies of the candidate subkeys. It is then more practical to count on fewer key bits, and for DES it is natural to count on $6k$ key bits, representing the subkey bits entering k S-boxes. Observe that those S-boxes S_j whose subkeys are *not being counted* may still be used to discard, before counting, those ciphertext pairs X, X' which yield an impossible characteristic for S_j when X, X' is assumed to be a right pair. That is, if for S-box S_j the observed input/output difference is $\Delta X_j, \Delta Y_j$, and ΔX_j, ΔY_j is an impossible characteristic in the XOR table for S_j, then the pair encrypted to produce the differences ΔX_j, ΔY_j cannot be a right pair. By filtering plaintext pairs in this way, the ratio of right to wrong pairs that will be counted is enlarged, and the actual subkey will be distinguished more directly. Thus the density of impossible characteristics in an S-box is important to determine the effectiveness of this filtering process.

1.1. *Results*

Observe that an r-round characteristic is simply the concatenation of r 1-round, or single round, characteristics defined on the round mapping F. It then follows that the probability of the r-round characteristic Ω_r can be bound as $p^{\Omega_r} \leq (p^{\Omega})^r$ where p^{Ω} is the *probability of the most likely nonzero single round characteristic*. At present there are no general bounds known for p^{Ω}. The main result of this paper is to bound p^{Ω} when the round mapping F is derived from a set of bijective mappings. These results will lead to bounds on the probability of characteristics in a large class of product ciphers.

Let $\pi\colon Z_2^m \to Z_2^m$ be an bijective mapping, which is referred to as an m-bit permutation. The set of all m-bit permutations is known as the symmetric group on 2^m objects and is denoted as S_{2^m}. Let $\Lambda_\pi(\Delta X, \Delta Y)$ be the value of the XOR table entry of the pair $\Delta X, \Delta Y \in Z_2^m$ for the permutation $\pi \in S_{2^m}$. We consider $\Lambda_\pi(\Delta X, \Delta Y)$ as a random variable $\Lambda_\pi(\Delta X, \Delta Y)\colon S_{2^m} \to \{0, 2, \ldots, 2^m\}$, assuming the uniform distribution on the set S_{2^m}. We prove (Theorem 2.1) that

$$\Pr(\Lambda_\pi(\Delta X, \Delta Y) = 0) = \frac{1}{2^m!} \cdot \sum_{k=0}^{2^{m-1}} (-1)^k \cdot \binom{2^{m-1}}{k}^2 \cdot 2^k \cdot k! \cdot (2^m - 2k)! \quad (1)$$

from which we are able to determine the exact distribution of $\Lambda_\pi(\Delta X, \Delta Y)$ (see Corollary 2.1). Note that (1) gives the distribution of a single entry in the XOR

table, but we are interested in global properties of the table such as the largest entry and the fraction of the table that is zero. Fortunately, we are able to manipulate the distribution of (1) to yield bounds on such global properties of the XOR table. In Section 3 we prove, for large m, the expected probability of the most likely nonzero characteristic for an m-bit permutation is at most $m/2^{m-1}$ when the uniform distribution on S_{2^m} is assumed. Equivalently, the expected maximum entry in the XOR table for nonzero characteristics is at most $2m$ for large m. Experiments indicate that this bound is significant for $m \geq 8$ (see Table 1 in Section 3). Theorem 3.1 indicates that the individual entries of an XOR table are expected to be distributed in the interval $[0, 2, \ldots, 2m]$. At this time we are not able to determine the exact distribution of the entries within this interval, but we are able to show that most XOR table entries are in fact zero. We prove (Corollary 2.2) that the expected fraction of the XOR table for nonzero characteristics that is zero approaches $e^{-1/2} = 0.60653$. In another way, approximately 60% of the entries for nonzero characteristics will be zero for a permutation selected uniformly. It follows that impossible characteristics can then be used to discard a high percentage of wrong pairs X, X' which give no probabilistic information about the actual key.

The sections of this paper are arranged as follows. In Section 1.2 some relevant notation is defined. In Section 2 we present the Pairing Theorem which is the counting result that is used to prove the major results concerning the distribution of characteristics in XOR tables. Later in Section 2 we determine the expected fraction of the XOR table that is zero, and use these calculations in Section 3 to prove our results concerning the largest entry of the XOR table. In Sections 3.1 and 3.2 we use our previous results to bound the probability of characteristics in two common product ciphers.

1.2. *Notation*

Throughout the paper we let $[\cdot]$ denote a boolean predicate which evaluates to zero or one. For example, the sum $\sum_{i=1}^{n}[n \text{ is prime}]$ computes $\pi(n)$, the number of primes less than or equal to n, while $\varphi(n) = \sum_{i=1}^{n}[\gcd(i, n) = 1]$. This notation should not be confused with $\mathbf{E}[\alpha]$ which is the expected value of the random variable α.

We now formalize some of the notation given in the introduction. For a given $\pi \in S_{2^m}$, define $\Lambda_\pi(\Delta X, \Delta Y)$ as

$$\Lambda_\pi(\Delta X, \Delta Y) = \sum_{\substack{X, X' \in Z_2^m \\ \Delta X = X + X'}} [\pi(X) + \pi(X') = \Delta Y]. \tag{2}$$

Then $2^{-m} \cdot \Lambda_\pi(\Delta X, \Delta Y)$ is a random variable giving the probability that the difference in the output of the mapping π is ΔY when the difference of the input pair X, X' is ΔX. For all $\pi \in S_{2^m}$, observe that when $\Delta X = 0$ or $\Delta Y = 0$ it follows that $\Lambda_\pi(\Delta X, \Delta Y) = 0$, unless $\Delta X = \Delta Y = 0$ whereupon $\Lambda_\pi(\Delta X, \Delta Y) = 2^m$. The distribution of $\Lambda_\pi(\Delta X, \Delta Y)$ taken over all possible $\Delta X, \Delta Y \in Z_2^m$ is known as the *pairs XOR distribution table for* π, or simply the XOR table for π. A characteristic is a sequence of differences. Unless otherwise

stated, when we speak of a characteristic $\Delta X, \Delta Y$ for a product cipher we refer to a nonzero single-round characteristic.

Example 1.1. For an m-bit permutation π, let XOR_π be the $2^m \times 2^m$ matrix where $\mathrm{XOR}_\pi(i, j) = \Lambda_\pi(i, j)$, $0 \le i, j \le 2^m - 1$, where i, j are treated as 3-bit binary vectors. Observe that $\mathrm{XOR}_\pi(0, 0) = 8$, and all other entries in the first row or column of $\mathrm{XOR}(\pi)$ are zero. For $\pi = (7, 2, 4, 1, 5, 6, 3, 0)$, where $\pi(0) = 7$, $\pi(1) = 2$, and so on, the corresponding XOR table is given by

$$\mathrm{XOR}_\pi = \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 4 & 0 \\ 0 & 2 & 2 & 0 & 2 & 0 & 0 & 2 \\ 0 & 2 & 2 & 0 & 2 & 0 & 0 & 2 \\ 0 & 2 & 2 & 0 & 2 & 0 & 0 & 2 \\ 0 & 2 & 2 & 0 & 2 & 0 & 0 & 2 \end{bmatrix} \tag{3}$$

Notice that if each entry in the XOR table is divided by 2^m, then the resulting matrix will be doubly stochastic.

The XOR table for an m-bit permutation π has the following general form:

$$\mathrm{XOR}_\pi = \begin{bmatrix} 2^m & 0 & 0 & \cdots & 0 \\ 0 & a_{1,1} & a_{1,2} & \cdots & a_{1,2^m-1} \\ 0 & a_{2,1} & a_{2,2} & \cdots & a_{2,2^m-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & a_{2^m-1,1} & a_{2^m-1,2} & \cdots & a_{2^m-1,2^m-1} \end{bmatrix} = \begin{bmatrix} 2^m & \mathbf{0} \\ \mathbf{0} & A_\pi \end{bmatrix}. \tag{4}$$

We are interested in the properties of the $(2^m - 1) \times (2^m - 1)$ submatrix $A_\pi = [a_{i,j}]$, $1 \le i, j \le 2^m - 1$, which corresponds to that portion of the XOR table entries attributed to nonzero characteristics. In this paper we show that, for large m, approximately 60% of the entries in A_π are zero and the largest entry in A_π is expected to be bounded by $2m$.

It is evident that all the entries of an XOR table are even, since summation in (2) is taken over all unordered pairs. However, for clarity in the counting to follow we consider the pairs to be ordered. To this end, define $\Lambda'_\pi(\Delta X, \Delta Y)$ as

$$\Lambda'_\pi(\Delta X, \Delta Y) = \frac{\Lambda_\pi(\Delta X, \Delta Y)}{2}. \tag{5}$$

2. The Pairing Theorem

Observe that a characteristic $\Delta X, \Delta Y$ corresponds to a pairing of the inputs and outputs of a permutation π (namely, the pairs X, X' and Y, Y' where $\Delta X = X + X'$ and $\Delta Y = Y + Y'$). For φ: $A \to B$, let Π_A and Π_B be pairings on the

sets A and B, respectively. Theorem 2.1 determines the number of functions φ which take no pair of Π_A to a pair in Π_B, and is referred to as the Pairing Theorem. All our results concerning the distribution of characteristics in bijective mappings are derived from the Pairing Theorem.

Theorem 2.1 (The Pairing Theorem). *Let* $A = \{a_1, a_2, \ldots, a_{2d}\}$ *and* $B = \{b_1, b_2, \ldots, b_{2d}\}$ *be sets of distinct elements. Let* $\Pi_A \subseteq A \times A$ *and* $\Pi_B \subseteq B \times B$ *be unordered pairs, such that* $a_i(b_i)$ *occurs in one pair of* $\Pi_A(\Pi_B)$ *for* $1 \leq i \leq 2d$. *Then the number* $\Phi(d)$ *of bijective functions* φ: $A \to B$ *such that, for all* $(a_i, a_j) \in \Pi_A, (\varphi(a_i), \varphi(a_j)) \notin \Pi_B$ *is*

$$\Phi(d) = \sum_{k=0}^{d} (-1)^k \cdot \binom{d}{k}^2 \cdot 2^k \cdot k! \cdot (2d - 2k)!. \tag{6}$$

Proof. Order the elements of Π_B as (b'_i, b'_{d+i}), $1 \leq i \leq d$. For $1 \leq i \leq d$ define $P(i)$ as

$$P(i) = \{\varphi | (\varphi(a), \varphi(a')) = (b'_i, b'_{d+i}), (a, a') \in \Pi_A\}$$

which is the number of functions φ that map some pair of Π_A to the pair $(b'_i, b'_{d+i}) \in \Pi_B$. It follows that

$$\Phi(d) = (2d)! - \left| \bigcup_{1 \leq j \leq d} P(j) \right| = (2d)! + \sum_{\substack{S \subseteq \{1,2,\ldots,d\} \\ S \neq \{\emptyset\}}} (-1)^{|S|} \cdot \left| \bigcap_{j \in S} P(j) \right| \tag{7}$$

using the inclusion–exclusion principle [13]. For $1 \leq k \leq d$ define the integers

$$P(i'_1, i'_2, \ldots, i'_k) = \left| \bigcap_{1 \leq j \leq k} P(i'_j) \right| \tag{8}$$

and by symmetry $P(1, 2, \ldots, k) = P(i'_1, i'_2, \ldots, i'_k) \overset{\text{def}}{=} \mathrm{P(d, k)}$. From (7) it then follows that

$$\Phi(d) = (2d)! + \sum_{k=1}^{k} (-1)^k \cdot \binom{d}{k} \cdot P(d, k). \tag{9}$$

It remains to determine $P(d, k)$ for $1 \leq k \leq d$. To this end, order the pairs within Π_A as (a'_i, a'_{d+i}) for $1 \leq k \leq d$. Then $P(d, k)$ is the number of functions φ for which there are k pairs (a''_i, a''_{d+i}) such that $\{\varphi(a''_i), \varphi(a''_{d+i})\} = \{(b'_i, b'_{d+i})\}$, $1 \leq i \leq k$. There are $\binom{d}{k}$ ways to choose the k pairs (a''_i, a''_{d+i}) from Π_A, $k!$ ways of assigning the (a''_i, a''_{d+i}) to the $b'_i, b'_{d+i})$, and 2^k ways of assigning

(a''_i, a''_{d+i}) to a particular pair in Π_B. It then follows that

$$P(d,k) = \binom{d}{k} \cdot 2^k \cdot k! \cdot (2d - 2d)!, \tag{10}$$

where $(2d - 2k)!$ is the number of ways to assign the elements in $A - \{a''_i, a''_{d+i} | 1 \le i \le k\}$. We then have that

$$\Phi(d) = (2d)! + \sum_{k=1}^{d} (-1)^k \cdot \binom{d}{k} \cdot P(d,k)$$

$$= \sum_{k=0}^{d} (-1)^k \cdot \binom{d}{k}^2 \cdot 2^k \cdot k! \cdot (2d - 2k)!,$$

which completes the proof of the theorem. □

Observe that $\Phi(2^{m-1})$ will give the number of permutations π for which the entry in the XOR table for ΔX, ΔY is zero; also, since φ is bijective we are able to determine the probability that exactly k pairs of difference ΔX lead to difference ΔY. We have defined differences using the "+" operator but we observe that the Pairing Theorem will apply to any notion of difference that pairs input and output differences uniquely, i.e., given ΔX and X there is only one X' such that $\Delta X = X + X'$. Using the Pairing Theorem we now derive the distribution of the random variable $\Lambda'_\pi(\Delta X, \Delta Y)$ as defined in (5).

Corollary 2.1. *For any fixed nonzero* ΔX, $\Delta Y \in Z_2^m$, *assuming* π *is chosen uniformly from the set* S_{2^m} *and* $0 \le k \le 2^{m-1}$,

$$\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)] = \sum_{k=0}^{2^{m-1}} \binom{2^{m-1}}{k}^2 \cdot \frac{k \cdot k! \cdot 2^k \cdot \Phi(2^{m-1} - k)}{2^m!}. \tag{11}$$

Proof. From the Pairing Theorem the number of $\pi \in S_{2^m}$ for which $\Lambda_\pi(\Delta X, \Delta Y) = 0$ is $\varphi(2^{m-1})$. By definition $P(d,k)$ is the number of mappings which take any k pairs from Π_A to the fixed k pairs $(b'_i, b'_{i+d}) \in \Pi_B$ where $1 \le i \le k$. Then from (10) it follows that

$$\frac{P(d,k) \cdot \Phi(d-k)}{(2d - 2k)!} \tag{12}$$

is the number of mappings which take exactly k pairs from Π_A to (b'_i, b'_{i+k}). Then the number of m-bit permutations for which k pairs of difference ΔX can be mapped into k fixed pairs of difference ΔY is

$$\sum_{\substack{S \subseteq [2^{m-1}] \\ |S| = k}} \frac{P(2^{m-1}, k) \cdot \Phi(2^{m-1} - k)}{(2^m - 2k)!} = \binom{2^{m-1}}{k} \cdot \frac{P(2^{m-1}, k) \cdot \Phi(2^{m-1} - k)}{(2^m - 2k)!}.$$

It follows from (10) that

$$|\{\pi : \Lambda'_\pi(\Delta X, \Delta Y) = k\}| = \binom{2^{m-1}}{k} \cdot \frac{P(2^{m-1}, k) \cdot \Phi(2^{m-1} - k)}{(2^m - 2k)!}$$

$$= \binom{2^{m-1}}{k}^2 \cdot k! \cdot 2^k \cdot \Phi(2^{m-1} - k)$$

$$= 2^m! \cdot \Pr(\Lambda'_\pi(\Delta X, \Delta Y) = k). \tag{13}$$

The theorem follows from the definition of expectation. □

Now consider any row of the XOR table corresponding to an input difference ΔX. Since there are 2^m pairs of difference ΔX, each of the 2^m column entries in the row is expected to be 1. We prove this formally in the next theorem by determining that $\mathbf{E}[\Lambda'_\pi(\Delta X), \Delta Y]$ tends to the constant $\frac{1}{2}$. The proof also contains the information required to show that approximately 60% of the XOR table is expected to be zero.

Remark. The proof of the next theorem is based on approximating a summation $\Sigma_k T_k$ by its first term as $\Sigma_k T_1 \cdot \Pi_{1 \le j < k}(T_{j+1}/T_j)$. If the T_k are exponentially decreasing then this method is very accurate, which is the case for $\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)]$.

Theorem 2.2. *For any fixed nonzero $\Delta X, \Delta Y \in Z_2^m$ and assuming π is chosen uniformly from the set S_{2^m},*

$$\lim_{m \to \infty} \mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)] = \tfrac{1}{2}. \tag{14}$$

Proof. Consider estimating $\Phi(d)$ from (6) when $d = 2^{m-1}$. The expression for $\Phi(2^{m-1})$ is an alternating summation with $2^{m-1} + 1$ terms $T_\varphi(m, k)$ for $0 \le k \le 2^{m-1}$ where

$$T_\varphi(m, k) = (-1)^k \cdot \binom{2^{m-1}}{k}^2 \cdot 2^k \cdot k! \cdot (2^m - 2k)!. \tag{15}$$

Then for k, $0 \le k < 2^{m-1}$, the ratio of successive terms is

$$\frac{T_\varphi(m, k+1)}{T_\varphi(m, k)} = -\frac{(2^{m-1} - k)^2}{(k+1)^2} \cdot \frac{2^{k+1} \cdot (k+1)!}{2^k \cdot k!} \cdot \frac{(2^m - 2k - 2)!}{(2^m - 2k)!}$$

$$= \frac{-2(2^{m-1} - k)^2}{(k+1)(2^m - 2k)(2^m - 2k - 1)}$$

$$= \frac{-2(2^{m-1} - k)^2}{4(k+1)(2^{m-1} - k)^2 - (k+1)(2^m - 2k)}$$

$$= -\left[2(k+1) \cdot \left(1 - \frac{1}{2^m - 2k}\right)\right]^{-1}. \tag{16}$$

It then follows that, for $1 \le k \le 2^{m-1}$,

$$\frac{T_\varphi(m,k)}{T_\varphi(m,0)}(-1)^k \cdot \prod_{j=0}^{k-1}\left[2(j+1)\cdot\left(1-\frac{1}{2^m-2j}\right)\right]^{-1}$$

$$= \frac{(-1)^k}{2^k\cdot k!}\cdot\prod_{j=0}^{k-1}\left[1-\frac{1}{2^m-2j}\right]^{-1}$$

$$= \frac{(-1)^k}{2^k\cdot k!}\cdot\exp\left[-\sum_{j=0}^{k-1}\ln\left(1-\frac{1}{2^m-2j}\right)\right]$$

$$\stackrel{\text{def}}{=} \frac{(-1)^k}{2^k\cdot k!}\cdot\varepsilon_k. \tag{17}$$

Observe that the expression in (17) is equal to the kth coefficient of the Taylor series expansion for

$$e^{-1/2} = \sum_{k\ge 0}\frac{(-1)^k}{2^k\cdot k!} \qquad \text{if} \quad \varepsilon_k = 1.$$

For this reason, the ε_k may be considered as error terms representing the deviation from the kth coefficient for $e^{-1/2}$. We now consider obtaining asymptotic estimates of these error terms. We have that

$$\varepsilon_k = \exp\left[-\sum_{j=0}^{k-1}\ln\left(1-\frac{1}{2^m-2j}\right)\right]$$

$$= \exp\left[-\sum_{j=0}^{k-1}\left[-\frac{1}{2^m-2j}+O\left(\frac{1}{(2^m-2j)^2}\right)\right]\right]$$

$$= \exp\left[\sum_{j=0}^{k-1}\frac{1}{2^m-2j}+O\left(\frac{1}{(2^{m-1}-j)^2}\right)\right]$$

from which it follows that

$$\ln(\varepsilon_k) < \sum_{j=0}^{k-1}\frac{1}{2^m-2k}+O\left(\frac{1}{(2^{m-1}-k)^2}\right) = \frac{k}{2^m-2k}+O\left(\frac{k}{(2^{m-1}-k)^2}\right).$$

Then $\varepsilon_k \to 1$ as $m \to \infty$ when $k = o(2^{m-1})$. On the other hand, when $k = 2^{m-1}$ we have that

$$\ln(\varepsilon_{2^{m-1}}) = \tfrac{1}{2}\cdot\sum_{j=0}^{2^{m-1}}\frac{1}{j}+O\left(\frac{1}{j^2}\right) = \frac{H_{2^{m-1}}}{2}+O\left(2-\frac{1}{2^{m-1}}\right)$$

since $\sum_{j=1}^{t} j^{-2} < 1 + \int_1^t j^{-2} = 2 - 1/t$. Here H_n is the nth harmonic number and it is known that $H_n = \ln n + O(1)$ [12]. Thus $\ln(\varepsilon_{2^{m-1}}) = O(m) + O(1)$, and

$\varepsilon_{2^{m-1}} = O(2^m)$. If we observe that $\varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_{2^{m-1}}$, then $\varepsilon_k = O(2^m)$ as $m \to \infty$ when $k \notin o(2^{m-1})$.

If $\varepsilon_0 \overset{\text{def}}{=} 1$ we may then write $\Phi(2^{m-1})$ as

$$\Phi(2^{m-1}) = T_\varphi(m,0) \cdot \left[\sum_{k=0}^{m} \frac{(-1)^k \cdot \varepsilon_k}{2^k \cdot k!} + \sum_{k=m+1}^{2^{m-1}} \frac{O(2^m)}{2^k \cdot k!} \right]$$

$$= T_\varphi(m,0) \cdot \left[\sum_{k=0}^{m} \frac{(-1)^k \cdot \varepsilon_k}{2^k \cdot k!} + O\left(\frac{(2^{m-1} - m - 1) \cdot 2^m}{2^m \cdot (m+1)!} \right) \right]$$

$$= T_\varphi(m,0) \cdot \left[\sum_{k=0}^{m} \frac{(-1)^k \cdot \varepsilon_k}{2^k \cdot k!} + O\left(\frac{(2^m - m)}{(m+1)!} \right) \right].$$

Then for large m we have that

$$\frac{\Phi(2^{m-1})}{T_\varphi(m,O)} - e^{-1/2} = \sum_{k=0}^{m} \frac{(-1)^k \cdot (\varepsilon_k - 1)}{2^k \cdot k!} + O\left(\frac{(2^m - m)}{(m+1)!} - \frac{1}{2^m \cdot m!} \right)$$

$$< \sum_{k=1}^{m} \frac{(-1)^k \cdot (e^{O(k/(2^m - 2k))} - 1)}{2^k \cdot k!} + O\left(\frac{2^m}{(m+1)!} \right)$$

$$= \sum_{k=1}^{m} \frac{O(k/(2^m - 2k))}{2^k \cdot k!} + o(1)$$

$$< \sum_{k=1}^{m} \frac{O(1)}{2^k \cdot (k-1)! \cdot (2^m - 2k)} + o(1)$$

$$< O\left(\frac{1}{2^m} \right) + o(1) < \varepsilon$$

for any $\varepsilon > 0$ and sufficiently large m. Then by definition

$$\lim_{m \to \infty} \frac{\Phi(2^{m-1})}{T_\varphi(m,o)} = e^{-1/2}$$

and

$$\Phi(2^{m-1}) \sim e^{-1/2} \cdot T_\varphi(m,o) \sim e^{-1/2} \cdot 2^m!. \tag{18}$$

A similar method can be used to determine an asymptotic estimate of $\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)]$. To this end, define 2^{m-1} terms $T_\Lambda(m,k)$ for $1 \le k \le 2^{m-1}$ as

$$T_\Lambda(m,k) = \binom{2^{m-1}}{k}^2 \cdot k \cdot k! \cdot 2^k \cdot \Phi(2^{m-1} - k) \tag{19}$$

and $\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)] = \Sigma_{1 \le k \le 2^{m-1}} T_\Lambda(m,k)/2^m!$. Then for large m and $1 \le k \le 2^{m-1} - 1$ the ratio of successive terms is

$$\frac{T_\Lambda(m,k+1)}{T_\Lambda(m,k)} = \left[2k \cdot \left(1 + \frac{1}{2^m - 2k}\right)\right]^{-1}. \tag{20}$$

As in the first part of the proof, it can be shown that $T_\Lambda(m,1)$ is the dominant term amongst the $T_\Lambda(m,k)$. For large m it then follows that

$$\frac{\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)]}{2^m! \cdot T_\Lambda(m,1)} - e^{1/2} < O\left(\frac{1}{2^m}\right) + o(1) < \varepsilon$$

for any $\varepsilon > 0$ and sufficiently large m. Then by definition

$$\lim_{m \to \infty} \frac{\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)]}{2^m! \cdot T_\Lambda(m,1)} = e^{1/2}$$

and

$$\mathbf{E}[\Lambda'_\pi(\Delta X, \Delta Y)] \sim \frac{e^{1/2} \cdot T_\Lambda(m,1)}{2^m!}.$$

The theorem now follows since

$$\begin{aligned} T_\Lambda(m,1) &= 2 \cdot (2^{m-1})^2 \cdot \Phi(2^{m-1}) \\ &\sim 2 \cdot 2^{2m-2} \cdot e^{-1/2} \cdot T_\varphi(m,0) \\ &\sim 2^{2m-1} \cdot e^{-1/2} \cdot (2^m - 2)! \end{aligned}$$

and

$$\lim_{m \to \infty} \frac{e^{1/2} \cdot T_\Lambda(m,1)}{2^m!} = \lim_{m \to \infty} \frac{2^{2m-1} \cdot (2^m - 2)!}{2^m!} = \lim_{m \to \infty} \frac{2^{2m-1}}{2^{2m} - 2^m} = \tfrac{1}{2}. \quad \square$$

As noted in the introduction, the presence of impossible characteristics assists in discarding certain plaintext pairs which cannot give any probabilistic information concerning the actual key. It has been observed that 20%–30% of the characteristics in the S-boxes of DES are impossible. Let $\Lambda_{m,0}$ be the expected number of nonzero characteristics $\Delta X, \Delta Y$ which have zero entries in the XOR table of a uniformly selected m-bit permutation. We are able to compute $\Lambda_{m,0}$ as a direct application of the Pairing Theorem and the previous theorem.

Corollary 2.2. *For any fixed nonzero $\Delta X, \Delta Y \in Z_2^m$ and assuming π is chosen uniformly from the set S_{2^m},*

$$\lim_{m \to \infty} \Lambda_{m,0} = \frac{(2^m - 1)^2}{e^{1/2}}. \tag{21}$$

Proof. By the definition of $\Lambda_{m,0}$ and (18) it follows that

$$\Lambda_{m,0} = \frac{1}{2^m!} \cdot \sum_{\substack{\Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \sum_{\pi \in S_{2^m}} [\Lambda'_\pi(\Delta X, \Delta Y) = 0)]$$

$$= \frac{1}{2^m!} \cdot \sum_{\substack{\Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \Phi(2^{m-1})$$

$$\sim \frac{(2^m - 1)^2}{2^m!} \cdot \Phi(2^{m-1})$$

$$\sim \frac{(2^m - 1)^2}{e^{1/2}}. \tag{22}$$

This completes the proof of the theorem. □

It now follows that approximately 60% of the entries of the A_π submatrix defined in (4) are zero since $e^{-1/2} = 0.6065$. Then, from (4), the total fraction of an XOR table that is expected to be zero is

$$\frac{(2^m - 1)^2}{e^{1/2} \cdot 2^{2m}} + \frac{2^{m+1} - 2}{2^{2m}}.$$

3. The Largest Entry in the XOR Table

For a random m-bit permutation π let Λ^*_m be defined as follows

$$\Lambda^*_m \stackrel{\text{def}}{=} \max_{\substack{\Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \Lambda'_\pi(\Delta X, \Delta Y). \tag{23}$$

Thus twice Λ^*_m is the size of the largest XOR entry for the mapping π, and will bound the probability of the most likely difference passing through π. Using the Pairing Theorem we are able to bound the expected value of Λ^*_m.

Theorem 3.1. *Assuming the uniform distribution on the set* S_{2^m},

$$\lim_{m \to \infty} \frac{\mathbf{E}[\Lambda^*_m]}{m} \leq 1. \tag{24}$$

Proof. For $1 \leq k \leq 2^{m-1}$, let $\Lambda'_{m,k}$ be the expected number of nonzero characteristics $\Delta X, \Delta Y$ for which $\Lambda'_\pi(\Delta X, \Delta Y) = k$. Further, let $\Pr(\Lambda'_\pi = k)$ be the

probability that an m-bit permutation has a nonzero characteristic $\Delta X, \Delta Y$ for which $\Lambda'_\pi(\Delta X, \Delta Y) = k$. The proof rests on the following inequalities:

$$\Pr(\Lambda^*_m = k) < \Pr(\Lambda'_\pi = k) \leq \Lambda'_{m,k}.$$

The left inequality follows from definitions and the right inequality follows from expanding the expectation:

$$\Lambda'_{m,k} = \sum_{i=0}^{2^{m-1}} i \cdot \Pr(i \text{ characteristics of size } k)$$

$$> \Pr(1 \text{ characteristic of size } k).$$

We prove that $\Sigma_{k>m} k \cdot \Lambda'_{m,k} = o(1)$ from which the theorem follows. For $0 \leq k \leq 2^{m-1}$ we have by definition that

$$\Lambda'_{m,k} = \frac{1}{2^m!} \cdot \sum_{\pi \in S_{2^m}} \sum_{\substack{\Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} [\Lambda'_\pi(\Delta X, \Delta Y) = k)].$$

Then using the Pairing Theorem it follows that

$$\begin{aligned}
\Lambda'_{m,k} &= \frac{1}{2^m!} \cdot \sum_{\substack{\delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \sum_{\pi \in S_{2^m}} [\Lambda'_\pi(\Delta X, \Delta Y) = k)] \\
&= \frac{1}{2^m!} \cdot \sum_{\substack{\Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \binom{2^{m-1}}{k} \cdot \frac{P(2^{m-1}, k) \cdot \Phi(2^{m-1} - k)}{(2^m - 2k)!} \\
&= \frac{1}{2^m!} \sum_{\substack{\Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \binom{2^{m-1}}{k}^2 \cdot 2^k \cdot k! \cdot \Phi(2^{m-1} - k)! \\
&= \frac{(2^m - 1)^2}{2^m!} \cdot \binom{2^{m-1}}{k}^2 \cdot 2^k \cdot k! \cdot \Phi(2^{m-1} - k)!. \qquad (25)
\end{aligned}$$

If the sequence in (25) is plotted for increasing k, then when $k > m$ the terms become negligible in size. Consider obtaining asymptotic estimates of (25) for $k = m + 1$, which is achieved in two steps. Using (18) from the proof of Theorem 2.2 and asymptotic estimates of the factorial function [14, p. 221], the three largest terms of (25) can be estimated as

$$\begin{aligned}
\frac{\binom{2^{m-1}}{m+1}^2 \cdot \varphi(2^{m-1} - (m+1))}{2^m!} &\sim \frac{\binom{2^{m-1}}{m+1}^2 \cdot (2^m - 2m - 2)!}{e^{1/2} \cdot 2^m!} \\
&= O(1) \cdot \frac{(2^m - 2m - 2)^{1/2} \cdot 2^{m - m/2} \cdot e^{2m+3/2}}{(2^{m-1} - m - 1) \cdot (m+1)^{2m-3}} \\
&= O(1) \cdot \frac{2^{m - m/2} \cdot e^{2m+3/2}}{(2^{m-1} - m - 1)^{1/2} \cdot (m+1)^{2m+3}}.
\end{aligned}$$

Then estimating the remaining terms of (25) in a similar way, which are 2^{m+1}, $(2^m - 1)^2$, and $(m + 1)!$, we have that

$$\Lambda'_{m,m+1} = O(1) \cdot \frac{2^{4m-m/2-1} \cdot e^{m+1/2}}{(2^{m-1} - m - 1)^{1/2} \cdot (m + 1)^{m+2}} \tag{26}$$

from which it follows that $\lim_{m\to\infty}(m + 1) \cdot \Lambda'_{m,m+1} = 0$. Observe that, for large m, $\Lambda'_{m,k}$ defined in (25) and $T_\varphi(m, k)$ as defined in (15) only differ by the multiplicative factor

$$\frac{(-1)^k \cdot (2^m - 1)^2}{2^m! \cdot e^{1/2}}.$$

It then follows from (16) that, for large m, and $1 \le k \le 2^{m-1} - 1$,

$$\frac{\Lambda'_{m,k+1}}{\Lambda'_{m,k}} = \left[2(k + 1) \cdot \left(1 - \frac{1}{2^m - 2k}\right)\right]^{-1}. \tag{27}$$

We then have

$$\begin{aligned}
\lim_{m\to\infty} \sum_{k=m+1}^{2^{m-1}} k \cdot \Lambda'_{m,k} &\le \lim_{m\to\infty} \sum_{k=m+1}^{2^{m-1}} k \cdot \Lambda'_{m,m+1} \\
&= \lim_{m\to\infty} \Lambda'_{m,m+1} \cdot \sum_{k=m+1}^{2^{m-1}} k \\
&= \lim_{m\to\infty} \Lambda'_{m,m+1} \cdot O(2^{2m}) \\
&= \lim_{m\to\infty} O(1) \cdot \frac{2^{6m-m/2-1} \cdot e^{m+1/2}}{(2^{m-1} - m - 1)^{1/2} \cdot (m + 1)^{m+2}} \\
&= 0
\end{aligned}$$

and thus $\Sigma_{k=m+1}^{2^{m-1}} k \cdot \Lambda'_{m,k} = o(1)$. Finally observe that, for large m,

$$\begin{aligned}
\mathbf{E}[\Lambda^*_m] &= \sum_{k=1}^{m} k \cdot \Pr(\Lambda^*_m = k) + \sum_{k=m+1}^{2^{m-1}} k \cdot \Pr(\Lambda^*_m = k) \\
&= \sum_{k=1}^{m} k \cdot \Pr(\Lambda^*_m = k) + \sum_{k=m+1}^{2^{m-1}} O(k \cdot \lambda'_{m,k}) \\
&\le m \cdot \left[1 - \sum_{k>m} O(\Lambda'_{m,k})\right] + o(1) \\
&= m \cdot (1 + o(1)) + o(1).
\end{aligned}$$

This completes the proof of the theorem. □

Table 1. The distribution of characteristics.

m	$(m+1)\cdot \Lambda'_{m,m+1}$	$\Sigma_{k=m+1}^{2^{m-1}} k\cdot \Lambda'_{m,k}$	$\overline{\mathbf{E}[\Lambda^*_m]}$	Min	Max
4	0.76863	0.87258	3.114	2	6
5	0.25793	0.28436	3.839	3	6
6	0.80244×10^{-1}	0.86489×10^{-1}	4.495	3	7
7	0.22027×10^{-1}	0.23498×10^{-1}	5.126	4	8
8	0.53856×10^{-2}	0.57019×10^{-2}	5.606	5	8
9	0.11818×10^{-2}	0.12438×10^{-2}	6.190	6	8
10	0.23470×10^{-3}	0.24584×10^{-3}	6.700	6	9

Let $\overline{\mathbf{E}[\Lambda^*_m]}$ be an empirical estimate of $\mathbf{E}[\Lambda^*_m]$ based on a sample of 10,000 random permutations. Further, let min (max) be the smallest (largest) maximum XOR entry found across the 10,000 permutations. Table 1 lists these quantities for $m = 4, 5, \ldots, 10$. We see that, for $m \geq 6$, $(m+1)\cdot \Lambda_{m,m+1}$ is a good approximation to the tail of the summation for $\mathbf{E}[\Lambda^*_m]$ beginning $k = m+1$. The reader is reminded that the results of Table 1 have been derived using $\Lambda'_\pi(\Delta X, \Delta Y)$ where $\Lambda'_\pi(\Delta X, \Delta Y) = \Lambda_\pi(\Delta X, \Delta Y)/2$.

Recall that p^Ω was defined in the introduction as the probability of the most likely single-round characteristic for an iterated mapping. The main use of Theorem 3.1 is its application in constructing classes of product ciphers for which p^Ω is bounded. If a product cipher uses s S-boxes, and each S-box has $\Lambda^*_m \leq m$, then $p^\Omega \leq \Lambda^*_m/2^{m-1} = m/2^{m-1}$. The probability of this being the case is less than $(1 - \Sigma_{k>m}\Lambda'_{m,K})^s$, which from Theorem 3.1 will approach one when m is large and $s = o(m^m)$. More practically, in Table 1 it was easy to find 10,000 8-bit permutations for which $\Lambda^*_8 \leq 8$, which could be used to construct a suitable product cipher with a known bound on p^Ω. In the ext two sections we describe two such product ciphers, assuming that 8-bit permutations with $\Lambda^*_8 \leq 8$ can be found directly by random selection.

3.1. *Characteristics in SP-Networks*

In this section we derive bounds on the expected value of p^Ω assuming that the round mapping F is based on random m-bit permutations selected uniformly from S_{2^m}. Consider the network shown in Fig. 1. Let F consist of s S-boxes implementing m-bit permutations $\pi_1, \pi_2, \ldots, \pi_s$ such that $F: Z_2^{m\cdot s} \to Z_2^{m\cdot s}$

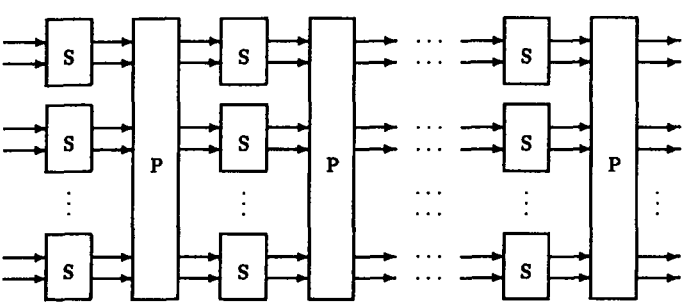

Fig. 1. The general SP-network product cipher.

where π_1 operates on the first block of s bits, π_2 operates on the second block of s bits, and so on. Then if we redefine Λ_m^* as

$$2^{m-1} \cdot p^{\Omega} \leq \Lambda_m^* \stackrel{\text{def}}{=} \max_{\substack{\pi \in \{\pi_1, \pi_2, \ldots, \pi_S\} \\ \Delta X, \Delta Y \in Z_2^m \\ w(\Delta X), w(\Delta Y) > 0}} \Lambda'_{\pi}(\Delta X, \Delta Y)$$

it follows that $\Lambda_m^*/2^{m-1}$ is the probability of the most likely characteristic across all s permutations in F. Then for any r-round characteristic Ω_r containing no zero differences it also follows that

$$p^{\Omega_r} \leq \left(\frac{\Lambda^*}{2^{m-1}}\right)^r. \tag{28}$$

In general the bound in (28) is not expected to be equality when $r > 1$. This discrepancy is accounted for by observing that the most likely characteristics have a zero input difference to a subset of the S-boxes, which means that these differences cause the expected output difference with probability 1. However, the avalanche effect diffuses the nonzero output difference of an S-box at round i to the inputs of several S-boxes at round $i + 1$, making it likely that more S-boxes at round $i + 1$ will have nonzero input difference than at round i, thus decreasing the probability of the characteristic predicting the difference from one round to the next. Thus characteristics are chosen not only because they are probable, but also because they may limit the influence of the avalanche effect on causing nonzero input differences to S-boxes.

Consider a 16-round 64-bit product cipher E for which the round mapping consists of 8×8-bit permutations followed by a transposition of the 64 ciphertext bits, which is an example of an SP-network. Then to predict the input difference to round 16 requires a 15-round characteristic Ω_{15} where the input difference to each of the first 15 rounds is nonzero. Let us assume that the permutations are selected uniformly from S_{2^8} and that at each round there is only one S-box which has a nonzero input difference. It then follows that

$$p^{\Omega_{15}} \leq (p^{\Omega})^{15} \leq \left(\frac{8}{2^7}\right)^{15} = 0.86736 \times 10^{-18}. \tag{29}$$

On the other hand, if Ω_{15} has nonzero input differences to two S-boxes at 7 out of the 15 rounds, then

$$p^{\Omega_{15}} \leq \left(\frac{8}{2^7}\right)^{2 \cdot 7} \cdot \left(\frac{8}{2^7}\right)^{8} = 0.32311 \times 10^{-26}. \tag{30}$$

3.2. *Characteristics in DES-Like Networks*

DES-like networks are symmetric ciphers whose round function is of the form shown in Fig. 2. The ciphertext is divided into halves, the left half L_r and the right half R_r. The round function F acts on R_r under the action of K_r, the

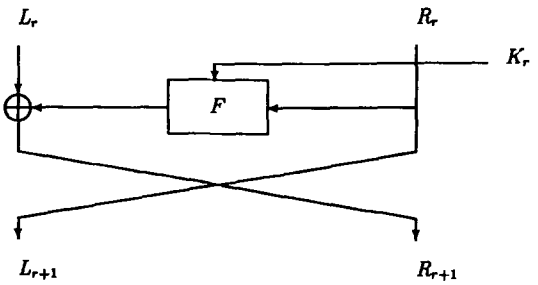

Fig. 2. The round function of a DES-like cipher.

subkey round r. For the SP-networks displayed in Fig. 1, if at any point a characteristic Ω_r predicts a zero difference, then all subsequent differences will be zero since the round mapping is bijective. However, in the case of a DES-like mapping, the round function F need not be bijective and nonzero input differences to F can be used to produce zero output differences.

For $a = a_1a_2 \in Z^{2m}$ where $a_i \in Z_2^m$, let $r(a) = a_2a_1$. An r-round characteristic for a DES-like mapping $\Omega_r = (\Delta X, \Delta Y_1, \ldots, \Delta Y_r)$ is said to be *iterative* if $\Delta X = r(\Delta Y_r)$. Taking into account the swapping operation at each round, an iterative characteristic essentially maps plaintexts of difference ΔX to ciphertexts of difference ΔX in r rounds. We observe that k r-round iterative characteristics can be concatenated to form a (kr)-round characteristic, $k > 0$. The best-known characteristic that has been used against DES is a 2-round iterative characteristic found by Biham and Shamir [5] that is concatenated $6\frac{1}{2}$ times to break 16-round DES.

Let F be defined as in the previous section to consist of s S-boxes implementing m-bit permutations $\pi_1, \pi_2, \ldots, \pi_s$ such that $F: Z_2^{m \cdot s} \to Z_2^{m \cdot s}$ where π_1 operates on the first block of s bits, π_2 operates on the second block of s bits, and so on. Also let the output of these substitutions be acted on by an (ms)-bit permutation P. With respect to DES, consider removing the E expansion and creating four new S-boxes that are 8-bit permutations; the P permutation is retained. A new key schedule yielding 32-bits for K_r would need to be devised.

Let α_r be the left-half difference at round r, β_r the right-half difference at round r, and γ_r the output difference of F at round r. Then these differences are listed in Table 2 for rounds 1–4. It is straightforward to argue that no 2-round iterative characteristics will exist when the F function is bijective, unless $\alpha_1 = \beta_1 = 0$. On the other hand, we prove that 3-round characteristics

Table 2. Round differences.

α_r	β_r	Round r
α_1	β_1	1
β_1	$\alpha_1 + \gamma_1$	2
$\alpha_1 + \gamma_1$	$\beta_1 + \gamma_2$	3
$\beta_1 + \gamma_2$	$\alpha_1 + \gamma_1 + \gamma_3$	4

are possible. We call α a fixed point [15] of F if inputs of difference α to F can lead to an output difference of α in F (that is, $\Lambda_F(\alpha, \alpha) > 0$).

Lemma 3.1. *Let α be a fixed point of F. Then Ω_3 is a 3-round nonzero iterative characteristic if*

$$\Delta X \in \{\alpha 0, 0\alpha, \alpha\alpha\}. \tag{31}$$

Proof. If Ω_3 is a 3-round iterative characteristic, then from Table 2 we must have that $\alpha_1 = B_1 + \gamma_2$ and $\beta_1 = \alpha_1 + \gamma_1 + \gamma_3$ which implies that $\gamma_1 + \gamma_2 + \gamma_3 = 0$. There are three cases to consider corresponding to the three possible values of ΔX in (31). We prove the case where $\Delta X = \alpha 0$ explicitly, and the other cases are similar. If $\Delta X = \alpha 0$, then $\gamma_1 = 0$, and $\gamma_2 = \alpha = \alpha_1$ since α is a fixed point of F. However, then $\beta_1 + \gamma_2 = \alpha$ and $\gamma_3 = \alpha$ from which we have that $\alpha_4 = \alpha_1 = \alpha$ and $\beta_4 = \alpha_1 + 0 + \alpha_1 = 0$. □

For each difference α such that $\Lambda_F(\alpha, \alpha) > 0$, three 3-round iterative characteristics of the form in (31) exist. Each of these three characteristics has a round difference of the form $\alpha 0$ which will go to the difference 0α with probability 1. Then once in every three rounds differences are predicted with certainty implying that

$$p^{\Omega_r} \leq (p^{\Omega})^{2\cdot\lfloor r/3\rfloor + \lfloor (r \bmod 3)/2\rfloor}. \tag{32}$$

Consider a Feistel-cipher similar to DES obtained by removing the E expansion and replacing the S-boxes by four 8-bit bijective mappings followed by a 32-bit permutation. Then the probability of a 15-round characteristic Ω_{15} is bound as

$$p^{\Omega_{15}} \leq (p^{\Omega})^{10} = \left(\frac{8}{2^7}\right)^{10} = 0.90949 \times 10^{-12}. \tag{33}$$

The bound is lowered further if we assume that more than one S-box has a nonzero input difference at a round which has a nonzero input difference.

4. Conclusion and Remarks

The method of differential cryptanalysis is based on the distribution of multi-round characteristics $\Omega_r(\Delta X, \Delta Y_1, \Delta Y_2, \ldots, \Delta Y_r)$. The probability of Ω_r correctly predicting ciphertext differences at each round in turn depends on the distribution of single-round characteristics $\Delta Y_i, \Delta Y_{i+1}$ for the round mapping F. When F consists of S-boxes implementing m-bit permutations, we have shown that the probability of the most likely single-round characteristic is expected to be less than $m/2^{m-1}$. It may well be the case that Theorem 3.1 can be improved to show that $\lim_{m\to\infty} \mathbf{E}[\Lambda_m^*]/m = 0$. Further research may attempt to prove that $\mathbf{E}[\Lambda_m^*] = O(\log m)$ for example.

Our results then show that a relatively simple design can produce product ciphers for which all characteristics Ω_r are expected to (correctly) predict

differences with low probability. We further note that random m-bit permutations can be generated efficiently [24], and that the fraction of permutations that are linear [11] or degenerate [22] in any output bit is tending to zero rapidly as a function of m. On the other hand, Biham and Shamir [5] found that replacing the S-boxes of DES by random 4-bit permutations yielded systems that were far weaker than the original DES. The weakness of these S-boxes appears to be due to the dimension of the permutation rather than the use of permutations *per se*. The XOR properties of S-boxes that are constructed from several permutations, as in the case of DES, is considered by O'Connor [23].

An apparent defense against differential cryptanalysis would be to design a round mapping F for which the corresponding XOR table contains uniform or nearly uniform entries. It has been shown by Nyberg [19], and independently by Adams [1], that it is possible to construct mappings $\pi: Z_2^{m_1} \rightarrow Z_2^{m_2}$ for which each entry of XOR_π is $2^{m_1 - m_2}$ when the input difference is nonzero. For the construction to be possible it must be the case that $m_1 \geq 2m_2$ which implies that the mapping cannot be bijective. Detombe and Tavares [9] have shown that for bijective mappings $\pi: Z_2^m \rightarrow Z_2^m$ the most balanced XOR tables are those for which each row has 2^{m-1} entries that are two, with the remaining XOR entries being zero. In both cases the mappings are constructed from boolean functions that are either bent or almost bent. More recently, several other such constructions have been found [2], [20].

We have concentrated on characteristics Ω_r but more important to the system designer are *differentials*. A differential is similar to a characteristic except that only an input difference ΔX and output difference $\Delta Y_r = \Delta Y$ are specified while the intermediate differences $\Delta Y_1, \Delta Y_2, \ldots, \Delta Y_{r-1}$ are unspecified and may assume any values which lead to ΔY at the rth round. The notion of a differential follows from modeling differences using Markov chains, as suggested by Lai [16]. It may be the case that all characteristics are unlikely but high probability differentials exist. A deeper analysis using Markov chains will be required to bound the probability of the most likely differential in a cipher. Notwithstanding, Nyberg and Knudsen [21] have shown that the probability of any differential is bounded from above by $2 \cdot (p^\Omega)^2$, regardless of the number of rounds.

Acknowledgments

I would like to thank Prabahkar Ragde for his assistance in developing the results in this thesis. I would also like to thank the referees for their cogent comments and for the correction of several errors in the original manuscript.

References

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

[2] T. Beth and C. Ding. On almost perfect nonlinear permutations. *Advances in Cryptology—EUROCRYPT 93*, Lecture Notes in Computer Science, vol. 765, T. Helleseth, ed., Springer-Verlag, Berlin, pages 65–76, 1994.

[3] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.
[4] E. Biham and A. Shamir. Differential cryptanalysis of the full 16-round DES. Technical Report 708, Technion, Israel Institute of Technology, Haifa, 1991.
[5] E. Biham and A. Shamir. Differential cryptanalysis of Snefru, Khafre, REDOCII, LOKI, and LUCIFER. *Advances in Cryptology, CRYPTO 91,* Lecture Notes in Computer Science, vol. 576, J. Feigenbaum, ed., Springer-Berlag, Berlin, pages 156–171, 1992.
[6] L. P. Brown, M. Kwan, J. Pieprzyk, and J. Seberry. Improving resistance to differential cryptanalysis and the redesign of LOKI. *Advances in Cryptology, ASIACRYPT 91,* Lecture Notes in Computer Science, vol. 739, H. Imai *et al.*, eds., Springer-Verlag, Berlin, pages 36–50, 1993.
[7] L. P. Brown, J. Pieprzyk, and J. Seberry. LOKI—a cryptographic primitive for authentication and secrecy applications. *Advances in Cryptology, AUSCRYPT 90,* Lecture Notes in Computer Science, vol. 453, J. Seberry and J. Pieprzyk, eds., Springer-Verlag, Berlin, pages 229–236, 1990.
[8] T. Cusick and M. Wood. The REDOC-II cryptosystem. *Advances in Cryptology, CRYPTO90,* Lecture Notes in Computer Science, vol. 537, A. J. Menezes and S. A. Vanstone, ed., Springer-Verlag, Berlin, pages 545–563, 1991.
[9] J. Detombe and S. Tavares. Constructing large cryptographically strong S-boxes. *Advances in Cryptology, AUSCRYPT 92,* Lecture Notes in Computer Science, vol. 718, J. Seberry and Y. Zheng, eds., Springer-Verlag, Berlin, pages 165–181, 1993.
[10] H. Feistel. Cryptography and computer privacy. *Scientific American,* 228(5):15–23, 1973.
[11] J. Gordon and H. Retkin. Are big S-boxes best? *Cryptography, Proceedings, Burg Feuerstein,* T. Beth, ed., Springer-Verlag, Berlin, pages 257–262, 1982.
[12] R. L. Graham, D. E. Knuth, and O. Patshnik. *Concrete Mathematics, A Foundation for Computer Science.* Addison-Wesley, Reading, MA, 1989.
[13] M. Hall. *Combinatorial Theory.* Blaisdell, Waltham, MA, 1967.
[14] M. Hofri. *Probabilistic Analysis of Algorithms.* Springer-Verlag, New York, 1987.
[15] L. R. Knudsen. Cryptanalysis of LOKI. *Advances in Cryptology, ASIACRYPT 91,* Lecture Notes in Computer Science, vol. 739, H. Imai *et al.*, eds., Springer-Verlag, Berlin, pages 237–246, 1993.
[16] X. Lai. *On the Design and Security and Block Ciphers.* ETH Series in Information Processing, J. Massey, ed. Hartung-Gorre Verlag, Konstanz, 1992.
[17] X. Lai and J. L. Massey. A proposal for a new block encryption standard. *Advances in Cryptology, EUROCRYPT 90,* Lecture Notes in Computer Science, vol. 473, I. B. Damgård, ed., Springer-Verlag, Berlin, pages 389–404, 1991.
[18] R. Merkle. Fast software encryption functions. *Advances in Cryptology, CRYPTO 90,* Lecture Notes in Computer Science, vol. 537, A. J. Menezes and S. A. Vanstone, ed., Springer-Verlag, Berlin, pages 476–501, 1991.
[19] K. Nyberg. Perfect nonlinear S-boxes. *Advances in Cryptology, EUROCRYPT 91,* Lecture Notes in Computer Science, vol. 547, D. W. Davies, ed., Springer-Verlag, Berlin, pages 378–386, 1991.
[20] K. Nyberg. Differentially uniform mappings for cryptography. *Advances in Cryptology—EUROCRYPT 93,* Lecture Notes in Computer Science, vol. 765, T. Helleseth, ed., Springer-Verlag, Berlin, pages 55–64, 1994.
[21] K. Nyberg and L. R. Knudsen. Provable security against differential cryptanalysis. Talk given at the Rump Session of CRYPTO 92, August, 1992.
[22] L. J. O'Connor. Enumerating nondegenerate permutations. *Advances in Cryptology, EUROCRYPT 91,* Lecture Notes in Computer Science, vol. 547, D. W. Davies, ed., Springer-Verlag, Berlin, pages 368–377, 1991.
[23] L. J. O'Connor. On the distribution of characteristics in composite permutations. *Advances in Cryptology—CRYPTO 93,* Lecture Notes in Computer Science, vol. 773, D. R. Stinson, ed., Springer-Verlag, Berlin, pages 403–412, 1994.
[24] E. M. Reingold, J. Nievergeld, and N. Deo. *Combinatorial Algorithms: Theory and Practice.* Prentice-Hall, Englewood Cliffs, NJ, 1976.
[25] A. Shimizu and S. Miyaguchi. Fast data encipherment algorithm FEAL. *Advances in Cryptology, EUROCRYPT 87,* Lecture Notes in Computer Science, vol. 304, D. Chaum and W. L. Price, eds., Springer-Verlag, Berlin, pages 267–278, 1988.