BRICS

**Basic Research in Computer Science**

# Partial and Higher Order Differentials and Applications to the DES

**Lars R. Knudsen**

**BRICS Report Series** **RS-95-9**

**ISSN 0909-0878** **February 1995**

See back inner page for a list of recent publications in the BRICS Report Series. Copies may be obtained by contacting:

BRICS
Department of Computer Science
University of Aarhus
Ny Munkegade, building 540
DK - 8000 Aarhus C
Denmark

Telephone: +45 8942 3360
Telefax: +45 8942 3255
Internet: BRICS@brics.dk

BRICS publications are in general accessible through WWW and anonymous FTP:

`http://www.brics.dk/`
`ftp ftp.brics.dk (cd pub/BRICS)`

# Partial and Higher Order Differentials and Applications to the DES

Lars R. Knudsen
BRICS[*], Aarhus University
email:ramkilde@daimi.aau.dk

January 13, 1995

**Abstract**

In 1994 Lai considered higher order derivatives of discrete functions and introduced the concept of higher order differentials. We introduce the concept of partial differentials and present attacks on ciphers presumably secure against differential attacks, but vulnerable to attacks using higher order and partial differentials. Also we examine the DES for partial and higher order differentials and give a differential attack using partial differentials on DES reduced to 6 rounds using only 46 chosen plaintexts with an expected running time of about the time of 3,500 encryptions. Finally it is shown how to find a minimum nonlinear order of a block cipher using higher order differentials.

# 1 Introduction

Differential cryptanalysis [1] was introduced by Biham and Shamir. Lai considered higher order derivatives of discrete functions [6] and the concept of higher order differentials was introduced. As a special case binary functions were considered, which is relevant for cryptanalysis of block ciphers. The cryptographic significance of higher order differentials was discussed, but no applications given. Knudsen and Nyberg [8] showed that block ciphers exist secure against a differential attack using first order differentials, as proposed by Biham and Shamir.

[*] Basic Research in Computer Science, Centre of the Danish National Research Foundation

In this paper we introduce the concept of **partial** differentials, i.e. differentials where only a part of the difference in the ciphertexts (after a number of rounds) can be predicted. We show examples of Feistel block ciphers secure against a differential attack using first order differentials, but vulnerable to a differential attack using partial differentials and higher order differentials, thus illustrating that one should be careful when claiming for resistance against differential attacks. Finally, we give a method of how to find a minimum nonlinear order of a block cipher using higher order differentials.

## 2 Differential Attacks

In this paper we consider Feistel ciphers. A **Feistel cipher** with block size $2n$ and with $r$ rounds is defined as follows. The round function $g$ is

$$\begin{aligned} g : GF(2)^n \times GF(2)^n \times GF(2)^m &\rightarrow GF(2)^n \times GF(2)^n \\ g(X, Y, Z) &= (Y,\ f(Y, Z) + X) \end{aligned}$$

where $f$ can be any function taking two arguments of $n$ bits and $m$ bits respectively and producing $n$ bits. '+' is a commutative group operation on the set of $n$ bit blocks.

Given a plaintext $P = (P^L, P^R)$ and $r$ round keys $K_1, K_2, ..., K_r$ the ciphertext $C = (C^L, C^R)$ is computed in $r$ rounds. Set $C_0^L = P^L$ and $C_0^R = P^R$ and compute for $i = 1, 2, ..., r$

$$(C_i^L, C_i^R) \;=\; (C_{i-1}^R, f(C_{i-1}^R, K_i) + C_{i-1}^L)$$

Set $C_i = (C_i^L, C_i^R)$ and $C^L = C_r^R$ and $C^R = C_r^L$.

Traditionally, the round keys $(K_1, K_2, ..., K_r)$, where $K_i \in GF(2)^m$, are computed by a key schedule algorithm on input a master key $K$.

The differential attacks exploit that pairs of plaintexts with certain differences yield other certain differences in the corresponding ciphertexts with a non-uniform probability distribution. For a pair of plaintexts, which are not discarded by a filtering process, see [1, 2], one tries for all values of the round key in the last round, if the expected difference in the ciphertexts occur. This is repeated several times and the most suggested value is taken to be the value of the secret key of the last round. Now all ciphertexts can be decrypted one round and a weaker cipher attacked in the same way but with a smaller complexity.

The signal to noise ratio, $S/N$ [1, 2], is the number of times the right key is counted over the number of times a random key is counted.

$$S/N = \frac{|K| \times p}{\gamma \times \lambda}$$

where $p$ is the probability of the differential used in the attack, $|K|$ is the number of possible values of the key, we are looking for, $\gamma$ is the number of keys suggested by each pair of plaintexts and $\lambda$ is the ratio of non-discarded pairs to all pairs, see [1, 2] for further details. For our attacks in this paper $\lambda = 1$. If $S/N \leq 1$ then a differential attack will not succeed.

Sometimes one also calls the function $f$, the round function. We adopt this convention for convenience, since it should cause no confusion.

For the remainder of this paper we will assume that the round keys are independent and uniformly random and of size $n$, i.e. half the block size. The difference of two quantities is always taken to be the operation for which the difference is independent on the value of the inserted key. Therefore when considering differences for the round function $f$ we will write $f(x)$ instead of $f(x, k)$. We will assume that the difference of two quantities chosen in an attack is the exclusive-or operation, if not stated explicitly otherwise. The complexity of the attacks is measured as the number of encryptions of the full cipher that an attacker has to perform for success.

# 3 Partial Differentials

In a conventional differential attack on a $2n$ bit Feistel cipher, a differential is a tool to predict an $n$ bit value of the ciphertext after a certain number of rounds. But as we will show now it is not always necessary to predict the full $n$ bit value. Even a 1 bit value suffices in some cases. A differential that predicts only parts of an $n$ bit value is called a **partial differential**.

In [7] it is shown that the functions $f(x) = x^{-1}$ in $GF(2^n)$, where $f(x) = 0$ for $x = 0$, are differentially 2-uniform for odd $n$ and differentially 4-uniform for even $n$, i.e. the highest probability of a non-trivial one round differential is $2/2^n$ and $4/2^n$ respectively. In both cases the nonlinear order of the outputs is $n - 1$ [7]. As an example consider a 5 round cipher using as round function

$$f(x, k) = (x \oplus k)^{-1}$$

in $GF(2^n)$ for $n$ odd. From the results of [8] this cipher is highly resistant against differential attacks using full differentials, since any 3 round differential has a probability of at most $2^{3-2n}$ according to Th. 2 of [8], that is, using differentials, where full $n$ bit differences are used. In an attack counting on the round key of the last round the signal to noise ratio is

$$S/N < \frac{2^n \times 2^{3-2n}}{1 \times 1} < 1$$

for $n > 3$ and the attack will not succeed. In an attack counting on the round keys of the last two rounds only a 2 round differential is needed. And since the concepts of characteristics and differentials coincide for 2 rounds in a Feistel cipher it is easy to see that there exists a differential with a probability of $2/2^n$ and that this differential obtains a maximum probability. The signal to noise ratio is

$$S/N = \frac{2^{2n} \times 2^{1-n}}{1 \times 1} = 2^{n+1}$$

and the attack will succeed with complexity $2^n$ chosen plaintexts and running time of about $2^{3n}$.

However, for every non-trivial input difference to one round there are only $2^{n-1}$ possible differences in the outputs, each one with a probability of $2/2^n$, since the round function is differentially 2-uniform and the exclusive-or operation is commutative. That is, for a non-trivial input difference we get one bit of information about the output differences. From this fact we can construct a 2 round differential of probability one, where only one bit of the differences after 2 rounds of encryption is predicted. In a differential attack counting on the round keys of the last two rounds for every pair of plaintexts only half the possible values of the keys will be suggested. We obtain

$$S/N = \frac{2^{2n} \times 1}{2^{(2n-1)} \times 1} = 2$$

and the attack will succeed with sufficiently many pairs of chosen plaintexts. We implemented the attack on a 5 round 18 bit cipher with a key of 45 bits using as round function $f(x) = x^{-1}$ in $GF(2^9)$. Using 18 pairs of chosen plaintexts in 100 tests only one pair of keys was found, the right keys in the fourth and fifth rounds.

The attack can be generalised and the following result holds.

**Theorem 3.1** *Let $f(x,k) : GF(2^n) \times GF(2^n) \rightarrow GF(2^n)$ be the round function in a 5 round Feistel cipher with block size $2n$ bits using 5 round keys, each of size $n$ bits. Let $\alpha$ $(\neq 0)$ be an input difference for which only a fraction $W$ of all output differences are possible. Then a differential attack using partial differentials has a complexity of $2L$ chosen plaintexts and a running time of about $L \times 2^{2n}$, where $L$ is the smallest integer s.t. $(W)^L < 2^{-2n}$. The value of $L$ is at most $2n+1$.*

Proof: Consider the following attack.

1. Let $\alpha$ be the non-trivial difference of two inputs to $f$, for which only a fraction $W$ of the output differences can occur.
2. Compute a table $T$ (initialised to zero in all entries), s.t. for $i = 0,..,2^n - 1$, $T[f(i) \oplus f(i \oplus \alpha))] = 1$.
3. Choose plaintext $P_1$ at random and set $P_2 = P_1 \oplus (\alpha \parallel 0)$.
4. Get the encryptions $C_1$ and $C_2$ of $P_1$ and $P_2$
5. For every value $k_5$ of the round key $RK_5$ do
   (a) Decrypt the ciphertexts $C_1, C_2$ one round using $k_5$. Denote these ciphertexts $D_1, D_2$.
   (b) For every value $k_4$ of the round key $RK_4$ do
      i. Calculate $t_i = f(D_i^R \oplus k_4)$ for $i = 1, 2$.
      ii. If $T[t_1 \oplus t_2 \oplus D_1^L \oplus D_2^L] > 0$ then output $k_5$ and $k_4$.

Since the nonlinear order of $f(x)$ can be as high as $n-1$, the information about the output differences we get from a given input difference is not necessarily easily determined. Therefore we may have to compute a table $T$, s.t. for a given input difference $\alpha$, if $T[\beta] > 0$ then an output difference $\beta$ is possible. The inputs to the first round are equal and the inputs to the second round has difference $\alpha$. That is, we can compute a fraction $W$ of all possible values of the output difference of the fourth round from the right halves of the ciphertexts and from the values in table T. Upon termination about $W \times 2^{2n}$ of the possible values of $(RK_4, RK_5)$ have been suggested, one of which is the right pair of keys. By repeating the attack sufficiently many times only one unique pair of keys, the right pair of keys, will be left suggested. Any other keys will be suggested with probability $W$ for each run of the above attack. Therefore after trying $L$ pairs of plaintexts any key but the right key, is suggested $L$ times with a probability of $(W)^L$ and if $(W)^L < 2^{-2n}$ with a high probability the right keys are uniquely determined. Finally, note that since $W \leq 1/2$, $\min_L : (1/2)^L < 2^{-2n} = 2n + 1$. □

The attack can be extended to work on ciphers with any number of rounds by counting on all but the first three round keys.

# 4 Higher Order Differentials

In [6] the definition of derivatives of cryptographic functions was given.

**Definition 4.1 ((Lai [6]))** *Let $(S, +)$ and $(T, +)$ be Abelian groups. For a function $f : S \mapsto T$, the derivative of $f$ at the point $a \in S$ is defined as*

$$\Delta_a f(x) = f(x+a) - f(x)$$

*The i'th derivative of $f$ at the point $a_1, ..., a_i$ is defined as*

$$\Delta^{(i)}_{a_1,\ldots,a_i} f(x) = \Delta_{a_i}(\Delta^{(i-1)}_{a_1,\ldots,a_{i-1}} f(x))$$

Note that the characteristics and differentials used by Biham and Shamir in their attacks correspond to the first order derivative described by Lai. Therefore it seems natural to extend the notion of differential into **higher order differentials**.

**Definition 4.2** *A one round differential of order $i$ is an $(i+1)$-tuple $(\alpha_1, ..., \alpha_i, \beta)$, s.t. $\Delta^{(i)}_{\alpha_1,...,\alpha_i} f(x) = \beta$*

When considering functions over $GF(2)$ the points $a_1, ..., a_i$ must be linearly independent for the $i$'th derivative not to be trivial zero.

**Proposition 4.1 ((Lai [6]))** *Let $L[a_1, a_2, ..., a_i]$ be the list of all $2^i$ possible linear combinations of $a_1, a_2, ..., a_i$. Then*

$$\Delta^{(i)}_{a_1,...,a_i} f(x) = \sum_{\gamma \in L(\alpha_1,...,\alpha_i)} f(P \oplus \gamma)$$

*If $a_i$ is linearly dependent of $a_1, ..., a_{i-1}$, then*

$$\Delta^{(i)}_{a_1,...,a_i} f(x) = 0$$

**Proposition 4.2 ((Lai [6]))** *Let $ord(f)$ denote the nonlinear order[1] of a multi-variable polynomial function $f(x)$. Then*

$$ord(\Delta_a f(x)) \leq ord(f(x)) - 1$$

This leads to the following Proposition.

**Proposition 4.3** *If $\Delta_{a_1,...,a_i} f(x)$ is not a constant, then the nonlinear order of $f$ is greater than $i$.*

Proof: From Prop. 4.2 it follows that

$$ord(f) \geq ord(\Delta_{a_1} f(x)) + 1 \geq .............. \geq ord(\Delta_{a_1,...,a_i} f(x)) + i$$

□

## 4.1 Attacks using higher order differentials

In the previous section we showed how to exploit partial information of differentials. One may ask the following question: does round functions exist, which does not leak any partial information for any non-trivial difference? The answer is positive and in the following we give an

[1] In [6] called the nonlinear degree.

example of a 5 round Feistel cipher, for which the round function is differentially 1-uniform i.e. for every non trivial input difference all output differences occur exactly once. We show that differential attacks on this cipher using higher order differentials are much more efficient than conventional differential attacks. We generalise the result to any 5 round Feistel cipher.

**Theorem 4.1** *Let $f(x,k) = (x+k)^2 \bmod p$, p prime, be the round function in a Feistel cipher of block size $log_2 p^2$, where '+' is addition modulo p and the difference of two quantities, x and y, is $x-y \bmod p$. f is differentially 1-uniform, a non-trivial one round differential has a probability of $1/p$. Secondly, the second order derivative of f is constant.*

Proof: To prove the first statement, consider a fixed $a \neq 0 \bmod p$. Then

$$\begin{aligned} f(x) - f(x+a) &=_p f(y) - f(y+a) \Leftrightarrow \\ x^2 - (x^2 + a^2 + 2ax) &=_p y^2 - (y^2 + a^2 + 2ay) \Leftrightarrow \\ 2ax &=_p 2ay \Leftrightarrow 2a(x-y) =_p 0 \Leftrightarrow x =_p y \end{aligned}$$

since $p$ is prime. To prove the second statement, let $a_1, a_2$ be constants, then

$$\begin{aligned} \Delta_{a_1,a_2} f(x) &= f(x+a_1+a_2) - f(x+a_1) - f(x+a_2) + f(x) \\ &= x^2 + (a_1+a_2)^2 + 2(a_1+a_2)x - (x^2 + a_1^2 + 2a_1 x) \\ &\quad -(x^2 + a_2^2 + 2a_2 x) + x^2 \\ &= (a_1+a_2)^2 - a_1^2 - a_2^2 \\ &= 2a_1 a_2 \end{aligned}$$

□

**Theorem 4.2** *Let $f(x,k) = (x+k)^2 \bmod p$, p prime, be the round function in a 5 round Feistel cipher of block size $\log_2 p^2$ with independent round keys, i.e. a key size of $5 \times \log_2 p$. A differential attack using first order differentials needs about $2p$ chosen plaintexts and has a running time of about $p^3$.*

Proof: When doing a differential attack counting on the round key in the fifth round of the above cipher we need a 3 (or 4) round differential. It is easy to see that there exists a 3 round differential with a probability of $1/p$ and that this differential obtains a maximum probability. We obtain

$$S/N = \frac{p \times 1/p}{1 \times 1} = 1$$

This attack is not possible, since the right key cannot be distinguished from other random keys. When doing a differential attack counting on the round keys in both the fourth and fifth rounds we need only a 2 round differential. There exists a 2 round differential with a probability of $1/p$, which is a maximum probability for the above cipher. In this case we obtain

$$S/N = \frac{p^2 \times 1/p}{1 \times 1} = p$$

This attack is possible. We need about $2p$ chosen plaintexts and for every pair of plaintexts we do two rounds of encryption for every $p^2$ possible keys of the fourth and fifth rounds. Therefore we obtain a complexity of about $p^3$. □

**Theorem 4.3** *Let* $f(x,k) = (x+k)^2 \bmod p$*,* $p$ *prime, be the round function in a 5 round Feistel cipher of block size* $\log_2 p^2$ *with independent round keys, i.e. a key size of* $5 \times \log_2 p$*. A differential attack using second order differentials needs about* 8 *chosen plaintexts with a running time of about* $p^2$*.*

Proof: Consider $\Delta_{\alpha,\beta} f(x)$ where $\alpha = a \parallel 0$ and $\beta = b \parallel 0$ for some fixed $a, b$, i.e the left halves of $\alpha$ and $\beta$ are zero. See Fig. 1, where $(0,0)$ denotes the trivial second order derivative of $f$ and where in the second round the second order derivative is $(a, b, 2 \times a \times b)$. Consider the following attack

1. Choose plaintext $P_1$ at random.

2. Set $P_2 = P_1 + \alpha$, $P_3 = P_1 + \beta$ and $P_4 = P_1 + \alpha + \beta$.

3. Get the encryptions $C_1, ..., C_4$ of $P_1, ..., P_4$

4. For every value $k_5$ of the round key $RK_5$ do

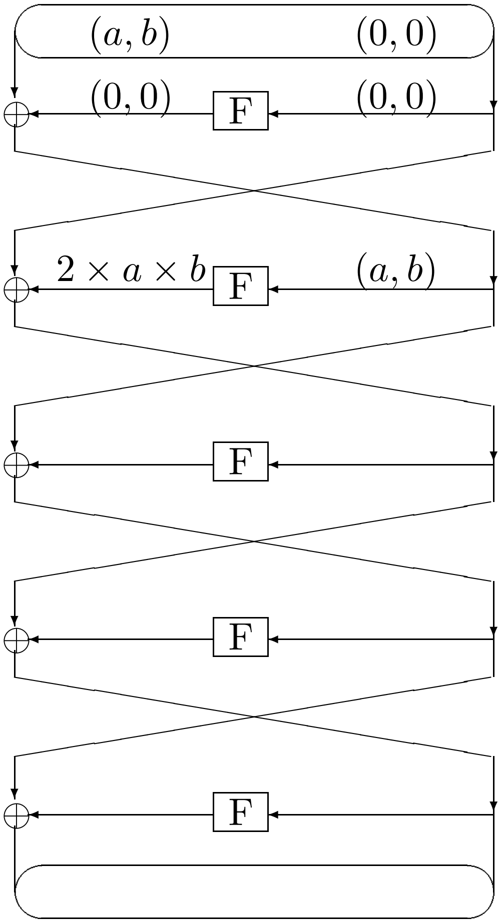

Figure 1: A second order differential of a five round Feistel cipher

(a) Decrypt all ciphertexts $C_1, ..., C_4$ one round using $k_5$. Denote these 4 ciphertexts $D_1, ..., D_4$.

(b) For every value $k_4$ of the round key $RK_4$ do

   i. Calculate $t_i = f(D_i^R + k_4)$ for $i = 1, .., 4$.

   ii. If $(t_1 + t_4 - (t_2 + t_3)) - (D_1^L + D_4^L - (D_2^L + D_3^L)) = 2 \times a \times b$ then output $k_5$ and $k_4$.

Here $X^L$ and $X^R$ denote the left and right halves of $X$ respectively. In the first round all inputs to the $f$-function are equal. In the second round the inputs form a second order differential with $(a, b, 2 \times a \times b)$. Since this differential has probability 1 according to Th. 4.1, the difference in the four inputs to the third round is $\Gamma = 2 \times a \times b$. Therefore the difference in the outputs of the fourth round can be computed as the exclusive-or sum of $\Gamma$ and of the right halves of the ciphertexts. Upon termination a few keys will have been suggested, among which the right keys appear, since the two round second order differential has probability 1. Therefore by repeating this attack a few times only one value of $(RK_4, RK_5)$ is suggested every time. This value is guaranteed to be the secret fourth and fifth round key. The signal to noise ratio of the attack is

$$S/N = \frac{p^2 \times 1}{1 \times 1} = p^2$$

where we have assumed that one key in average is suggested by each pair of plaintexts. Now it is trivial to find the remaining three round keys by similar attacks on cryptosystems with less than five rounds. As in [1, 2] we can pack the chosen plaintexts in economical structures, thus as an example obtain four second order differentials from 8 chosen plaintexts.□

If the prime $p$ above is of cardinality, say about $2^{25}$, according to Th. 4.2 a differential attack using first order differential has a complexity of about $2^{75}$ using about $2^{26}$ chosen plaintexts, i.e. not at all a practical attack. According to Th. 4.3 a differential attack using second order differentials has a complexity of about $2^{50}$ using only about 8 chosen plaintexts, a practical attack or at least not far from being one.

The attack in the proof of Th. 4.3 can be applied to any 5 round Feistel cipher, where the round function contains no expansion and where the output coordinates are quadratic, i.e. the nonlinear order of $f$ is 2. Furthermore the attack can be converted into an attack on any 5 round Feistel cipher. For convenience let us now consider functions over $GF(2)$. We state explicitly the definition of higher order differentials for this important case.

**Definition 4.3** *A one round differential of order $i$ is an $(i+1)$-tuple $(\alpha_1, ..., \alpha_i, \beta)$, s.t. all $\alpha_j$'s are linearly independent and*

$$\sum_{\gamma \in L(\alpha_1,\ldots,\alpha_i)} F(P \oplus \gamma) = \beta$$

It is seen there are $2^i$ plaintexts in an $i$-order differential.

**Theorem 4.4** *Let $f(x, k)$ be the round function in a 5 round Feistel cipher of block size $2n$ with independent round keys, i.e. a key size of $5 \times n$ bits. Assume that the nonlinear order of $f$ is $r$. Then a differential attack using $r$-order differentials needs about $2^{r+1}$ chosen plaintexts with a running time of about $2^{2n+r}$.*

Proof: According to Prop. 4.3 the $r$-order derivative of a function of nonlinear order $r$ is a constant. Therefore we can obtain a 2 round $r$-order differential with probability 1 and do a similar attack as in the proof of Th. 4.3. □

To illustrate the above attack, we consider now the differentially uniform mappings $f(x) = x^{2^k+1}$ in $GF(2^n)$ described in [8].

**Lemma 4.1** *Consider the permutation $f(x) = x^{2^k+1}$ in $GF(2^n)$ for $n$ odd and $gcd(k, n) = 1$. $f$ is differentially 2-uniform and the second order derivative of $f$, $\Delta_{\alpha,\beta} f(x)$ is a constant with the value $\Gamma = \alpha \times \beta \times (\alpha^{2^k-1} \oplus \beta^{2^k-1})$, where $'\times'$ is multiplication in $GF(2^n)$.*

Proof: The first statement is proved in [8] and that the second derivative is a constant follows from Prop. 4.2. The actual constant can be computed in a straightforward way and is omitted here (see [5]). □

We implemented the attack of Th. 4.4 counting on both the fourth and fifth round key using second order differentials in a five round Feistel cipher with $f(x)$ of Lemma 4.1 as round function and with $n = 9$ and $k = 1$, i.e. a 18 bit cipher with a 45 bit key. In 100 tests using 12 chosen plaintexts only one pair of keys was suggested and every time this pair was the right values of the fourth and fifth secret round keys. By using quartets as defined in [1, 2] the number of chosen plaintexts can be reduced to about 8. Note that for this cipher the probability of any 3 round differential of first order is at most $2^{3-2n}$ [8], where $2n$ is the block size. Also note that the example cipher of [8] has 6 rounds, and is therefore not vulnerable to the above attacks.

| The outputs of S-box | Does not affect S-boxes |
|---|---|
| 1 | 1, 7 |
| 2 | 2, 6 |
| 3 | 3, 1 |
| 4 | 4, 2 |
| 5 | 5, 8 |
| 6 | 6, 4 |
| 7 | 7, 5 |
| 8 | 8, 3 |

Table 1: Flow of the S-box output bits.

# 5 Applications to the DES

## 5.1 Partial differentials of the DES

For the DES [9] there are partial differentials with probability one. When two inputs to the $F$-function are equal in the inputs to an S-box, the outputs from that S-box are always equal, independent of the values of the inputs to other S-boxes. These partial differentials are used to a wide extent in Biham and Shamirs attacks on the DES [1, 2].

The output of an S-box affects the inputs of at most six S-boxes in the following round, because of the P-permutation, see Table 1. This fact can be used to construct a four round partial differential for the DES with probability one, which gives knowledge about the difference of eight bits in the ciphertext after four rounds. Consider a pair of plaintexts where the right halves are equal and the left halves differ, such that the inputs to only one S-box, say S-box 1, are different after the E-expansion. The first round in the differential holds always, and in the second round the outputs of all S-boxes except S-box 1 are equal. In the inputs to the third round the inputs of two S-boxes, S-boxes 1 and 7, are always equal, since S-box 1 does not affect these S-boxes according to Table 1. Therefore the outputs of these S-boxes are equal, and the xor of eight bits in the right halves of the ciphertexts after three rounds are known, since the xor in the inputs in the second round is known. The right halves after three rounds equal the left halves after four rounds, therefore the xor of eight bits after four rounds of

| S-box | Input xor (hex) | Bit $i$ ($y_i$) of output xor | Probability $(p-1/2)\times 64$ |
|---|---|---|---|
| 1 | 24 | 3 | -20 |
| 2 | 2 | 2 | 20 |
|  | c | 4 | 20 |
|  | 20 | 2 | 28 |
|  | 22 | 2 | -20 |
|  | 2d | 1 | 20 |
| 3 | 2 | 1 | 20 |
|  | 4 | 1 | 20 |
|  | 10 | 4 | 24 |
|  | 20 | 2 | 24 |
| 5 | 1 | 2 | 20 |
| 6 | 4 | 3 | 20 |
|  | c | 4 | 20 |
|  | 24 | 1 | -24 |
| 7 | 2 | 2 | 24 |
|  | c | 2 | 20 |
|  | e | 2 | -20 |
|  | 20 | 4 | 24 |
| 8 | 1 | 2 | 20 |
|  | 10 | 3 | 20 |
|  | 20 | 4 | 20 |

Table 2: The partial 1-bit output differentials with $|p-1/2| \geq 20/64$ for the 8 S-boxes of DES.

encryption are known with probability one. This differential can be used to attack the DES with 6 rounds in a differential attack using only a few chosen plaintexts as we will show in the next section.

There are other interesting partial differentials for the DES. Consider a six bit input difference (xor) to one S-box, $x_1, x_2, x_3, x_4, x_5, x_6$ and the corresponding difference in the outputs $y_1, y_2, y_3, y_4$. Instead of considering all 4 output bits as in traditional differentials, we consider only one of the $y_i$'s. The probabilities of these partial *1-bit output* differentials in the ideal case will be 1/2. In Table 2 for all 8 S-boxes the partial differentials for which $|p-1/2| \geq 20/64$ are listed, where $p$ is the probabilities of the differentials. Note that if $p$ is the probability that an xor bit is one, $1-p$ is the probability that the bit is zero. As an example consider S-box 2, where an input xor of $20_x$ leads to an output xor, for which the xor of the second most significant bits of the outputs is one in 60 out of all 64 possible pairs of inputs. It is also interesting to note that for the S-box 4, the probabilities of partial 1-bit output differentials are all between 20/64 and 44/64, i.e. $|p - 1/2| \leq 12/64$ for S-box 4. S-box 4 has been the subject of much debate since the publication of the DES and it has been conjectured the weakest S-box. It is 75% redundant, see [3] and it has a strange difference distribution table (see [1, 2]). To our knowledge the above properties show for the first time a case, where S-box 4 is the strongest of the 8 S-boxes.

Another interesting property of the S-boxes is revealed by considering pairs of input where the only the two middle input bits differ, i.e. xors $04_x$, $08_x$ and $0c_x$. These xors are of particular interest in differential cryptanalysis, since this allows neighbouring S-boxes to have equal inputs, i.e. xors $00_x$. For these input xors, the probability that one particular bit in the output xor is zero is at most 36/64 for S-boxes no. 2, 3 and 7. For the S-boxes 1, 5, 6 and 8 the probability is at most 32/64 and for S-box 4 at most 24/64. We can use the above partial differentials to construct a four round differential, which gives knowledge about the difference of more than eight bits in the ciphertext after four rounds.

As an example, for S-box 7, an input xor of $04_x$ will yield an output xor, such that the xor of the second output bits ($y_2$) is zero with probability 36/64. Consider a pair of plaintexts with difference $00000020_x \mid 00000000_x$, that is where the right halves are equal and the left halves differ in only one bit. After one round of encryption the difference will always be $00000000_x \mid 00000020_x$. After two rounds of encryption the difference will be $00000020_x \mid Y_x$ with probability 36/64, where $E(Y)$ is different in the inputs to only S-boxes 1, 2, 6, and 8. Therefore the outputs of S-boxes 3, 4, 5, and 7 will be equal after three rounds of encryption. In other words one gets knowledge of the xor of 16 bits in the right halves of the ciphertexts after three rounds and therefore in the left halves of the ciphertexts after four rounds of encryption with probability 36/64.

In a similar way, one can use two of the combinations of Table 2, namely the input xor $04_x$ for S-box 6 and the input xor $04_x$ for S-box 3, both with probability 52/64 to obtain a four round partial differential with probability $(52/64)^2 \simeq 42/64$ where the xor of nine ciphertext bits are known.

Note that although the above differentials can be used to deduce key bits of the DES with 6 or fewer rounds in a partial differential attack, it is also clear that when considering the DES with more than 6 rounds the method will only work locally in the first few and last few rounds of the cipher.

Finally we note that in [10] Preneel et al. considered, what they call *reduced exors*, in differential attacks on the DES in CFB mode. The reduced exors have some resemblance with partial differentials.

### 5.1.1 Attack on 6 round DES.

In this section we consider the DES [9] reduced to 6 rounds. We take the first 6 rounds of the standard and omit the initial and final permutation, since they are of no importance for our attack.

**Theorem 5.1** *There exists a differential attack on DES with 6 rounds, which finds the secret key using 46 chosen plaintexts in expected time the time of about 3,500 encryptions, which can be done in a few seconds on a PC.*

Proof: We consider a differential chosen plaintext attack using the differential in Fig. 2 and a similar differential where all the quantities $20000000_x$ are replaced by $40000000_x$. Assume first that the outputs of the first round have difference $\alpha$. The inputs to the third round differ in only two bits both affecting only S-box 1. According to the above discussion, the inputs with difference $X$ to the fourth round are equal in the inputs to the S-boxes 1 and 7. Therefore eight bits of the difference $Y$ are zero. Since the difference of the inputs to the third round is known, the attacker knows eight bits of the difference of the outputs of the F-function in the sixth round, since he knows the difference in the ciphertexts. These eight bits are the output bits of S-boxes 1 and 7. The attacker now tries for all 64 possible values of the key whether the inputs to S-box 1 yield the computed expected output difference, and does the same for S-box 7. For every pair of ciphertexts used in the analysis for both S-boxes the attacker will get an average of 4 suggested key values, among which the right key values

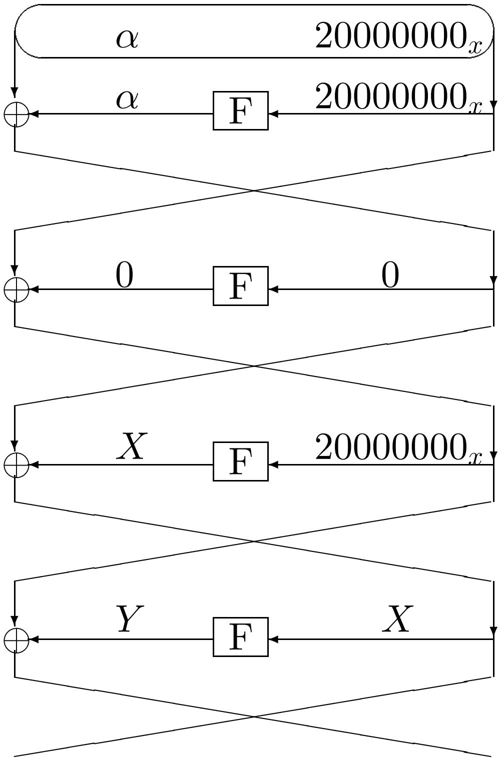

Figure 2: A 4 round differential of DES.

appear, since the used differential has probability one. By trying a few pairs, e.g. four pairs with a high probability only one key value, the right key value, will be left suggested by all pairs.

In the following, let $K_{i,j}$ denote the six bit key in S-box no. $j$ in the $i$'th round and let $P$ be the 32 bit linear permutation in the DES round function, see [9]. '|' and '||' denotes concatenation of 4 bit and 32 bit strings respectively.

We assumed above that the difference of the outputs of the first round is $\alpha$, which it will not always be. First we note that since the inputs to the first round differ in the inputs to only one S-box, there are only 16 possible values of $\alpha$. Choose a set of 4 plaintexts

$$P_i = A_i \parallel P_R$$

for $i = 0, ..., 3$, where $A_i = P(a_i \mid r_0 \mid r_1 \mid ... | r_5 \mid r_6)$, where $a_i = i$, each of 4 bits, the $r_k$'s are randomly chosen 4 bit numbers and $P_R$ is a randomly chosen 32 bit string. Next choose a set of 4 plaintexts

$$P_{1,j} = B_j \parallel P_R \oplus \Phi_{1,1}$$

for $j = 0, ..., 3$, where $B_j = P(b_j \mid r_0 \mid r_1 \mid ... \mid r_5 \mid r_6)$, $\Phi_{1,1} = 20000000_x$ and $b_0 = 0_x$, $b_1 = 4_x$, $b_2 = 8_x$, $b_3 = c_x$.

By combining each of the four plaintexts $P_i$ with each of the four plaintexts $P_{1,j}$ one obtains one pair of plaintexts with difference

$$P(h_x \mid 0 \mid 0 \mid 0 \mid 0 \mid 0 \mid 0 \mid 0) \parallel \Phi_{1,1} \tag{1}$$

for all values of $h = 0, ..., f_x$, that is, from these eight plaintexts one pair of plaintexts is a right pair with respect to the characteristic in Fig. 2.

To get more right pairs choose a set of 4 plaintexts

$$P_{2,j} = B_j \parallel P_R \oplus \Phi_{1,2}$$

for $j = 0, ..., 3$, where $\Phi_{1,2} = 40000000_x$, and a set of 4 plaintexts

$$P_{3,j} = A_i \parallel P_R \oplus \Phi_{1,1} \oplus \Phi_{1,2}$$

for $i = 0, ..., 3$.

By combining the set $P_{2,j}$ with the set $P_{3,j}$ one obtains another pair of plaintexts with difference (1) for all values of $h = 0, ..., f_x$.

By combining the set $P_{1,j}$ with the set $P_{2,j}$ and combining the set $P_i$ with the set $P_{3,j}$ one obtains 2 pairs of plaintexts with difference

$$P(h_x \mid 0 \mid 0 \mid 0 \mid 0 \mid 0 \mid 0 \mid 0) \parallel \Phi_{1,2}$$

for all values of $h = 0, ..., f_x$. Note that the characteristics just defined both affect the same S-box in the first round. Get the encryptions of the 16 plaintexts $P_i, P_{1,j}, P_{2,j}$ and $P_{3,j}$.

The attack proceeds as follows.

1. For every value $k_{1,1}$ of the key $K_{1,1}$ in S-box 1 in the first round do
   (a) Let $k_{1,*}$ be the 48 bit key obtained from the concatenation of the value of $k_{1,1}$ and 42 randomly chosen bits.
   Compute $c_0 = F(k_{1,*}, P_R)$ and $c_1 = F(k_{1,*}, P_R \oplus \Phi_{1,1})$. Now

$c_0 \oplus c_1 = P(y \mid 0 \mid 0 \mid ... \mid 0)$ for some hex value $y$. Find the plaintext $P_i$ and $P_{1,j}$, such that $c_0 \oplus c_1 = A_i \oplus B_j$. The pair of plaintexts $P_i$ and $P_{1,j}$ is a right pair with respect to the characteristic in Fig. 2. Next compute $c_2 = F(k_{1,*}, P_R \oplus \Phi_{1,2})$ and $c_3 = F(k_{1,*}, P_R \oplus \Phi_{1,1} \oplus \Phi_{1,2})$. Find the plaintext $P_{2,j}$ and $P_{3,j}$, such that $c_2 \oplus c_3 = B_j \oplus A_i$. The pair of plaintexts $P_{2,j}$ and $P_{3,j}$ is a right pair with respect to the characteristic in Fig. 2.

Repeat this procedure finding 2 right pairs $P_i$ and $P_{2,j}$, $P_{1,j}$ and $P_{3,j}$ for the second characteristic.

(b) Use the four right pairs in the differential attack described above. First do the attack on S-box 1 in the last round. If one key value $k_{6,1}$ of $K_{6,1}$ is suggested by all four pairs, perform the differential attack on S-box 7 in the last round. If one key value $k_{6,7}$ of $K_{6,7}$ is suggested by all four pairs, take $k_{6,1}$ and $k_{6,7}$ as the key values of $K_{6,1}$ and $K_{6,7}$ and take $k_{1,1}$ as the value of $K_{1,1}$.

The above attack finds 18 key bits with a high probability. In step 1(a) above we need not do a complete evaluation of the F-function, only the computation of the one S-box involved is needed. For every value of $K_{1,1}$ we do 4 S-box evaluations. Then for every value of $K_{6,1}$ we do 8 S-box evaluations, one for each of the 8 ciphertexts in the 4 pairs. The search for $K_{6,7}$ is done only when one key value of $K_{6,1}$ is suggested all four times. Totally the time used is about the time of $2^{15}$ S-box evaluations, about the time of 500 encryptions of six round DES. Note that the differential used in the attack has probability one. More key bits can be found in similar attacks by plaintexts yielding other characteristics.

With an additional 2 sets of each 16 plaintexts involving other S-boxes in the first round one finds 54 key bits. By a careful choice of each of the 2 sets one of the plaintext $P_i$ in the above described attack can be reused. Since the DES has dependent round keys some of the key bits tried in the first and in the sixth round are identical. Using the S-boxes 1, 2 and 5 in the first round is an optimal choice and the attack finds 45 bits of the 56 bit secret key. The remaining 11 bits can be found by exhaustive search. The attack needs a total of 46

| No. of chosen plaintexts | No. of keybits found |
|---|---|
| 7 | 8 |
| 16 | 18 |
| 31 | 33 |
| 46 | 45 |

Table 3: Complexities of our attacks on DES with 6 rounds.

plaintexts and runs in time about 3,500 encryptions of six round DES, which can be done in a few seconds on a PC. □

There are possible variations of the above attack, which are listed in Table 3. It should be noted that the linear attack combined with differential 'techniques' by Hellman and Langford [4] exploits the same phenomenon as in our attack, but the two attacks are different. Finally we note that in [10] Preneel et al. considered, what they call *reduced exors*, in differential attacks on the DES in CFB mode. The reduced exors have some resemblance with partial differentials.

## 5.2 Higher order differentials of the DES

In this section we consider higher order differentials for the 8 S-boxes of the DES. Table 4 lists the probabilities of the most likely $n$'th order differentials for the 8 S-boxes of the DES, for $n = 1, ..., 4$. Note that the probability of any fifth order differential is one, since the output coordinates of the DES S-boxes have order 5 (see [11]) and the fifth derivative is a constant according to Prop. 4.2. The numbers for S-box 4 in Table 4 are substantially different from

the numbers of the other S-boxes and there exist 3. order differentials with probability one. For example with $\alpha_1 = 25_x$, $\alpha_2 = 24_x$ and $\alpha_3 = 30_x$ the third order differential of S-box 4, $\Delta_{\alpha_1,\alpha_2,\alpha_3}(S4) = f_x$ with probability one. Note that $\Delta_{\alpha_1,\alpha_2,\alpha_3}(S4)$ is the exclusive-or of eight six bit inputs. We have found no way of exploiting higher order differentials for the DES, other than by attacking a four round version of the DES. However since the DES with four rounds is trivially broken using first order differentials, this application is not of much use.

| S-box no. | 1. order | 2. order | 3. order | 4. order |
|---|---|---|---|---|
| 1 | 16 | 24 | 48 | 64 |
| 2 | 16 | 28 | 48 | 64 |
| 3 | 16 | 28 | 40 | 64 |
| 4 | 16 | 48 | 64 | 64 |
| 5 | 16 | 28 | 40 | 64 |
| 6 | 16 | 24 | 40 | 64 |
| 7 | 16 | 28 | 40 | 64 |
| 8 | 16 | 28 | 40 | 64 |

Table 4: The probabilities (× 64) of the best higher order differential s for the 8 S-boxes of DES.

# 6 Computing the Nonlinear Order

In [11] it was considered to cryptanalyse the DES by the method of formal coding. The conclusion was that this is hardly possible. It was shown also that the nonlinear order of any of the 8 S-boxes in the DES is 5. An open question is, what is the order of the outputs for the full 16 round DES. In general, a cipher will be vulnerable to attacks like the method of formal coding if the nonlinear order of the outputs is too low. Higher order differentials can be used to determine a lower bound of the nonlinear order of a block cipher.

**Test for nonlinear order**

Input: $E_K(\cdot)$, a block cipher, a key $K$, plaintexts $x_1 \neq x_2$ and $r$, an integer.

Output: $i \leq r$, a minimum nonlinear order of $E_K$.

Let $a_1, a_2, ..., a_i$ be linearly independent.

```
1. Set $i = 1$
2. Compute $y_1 = \Delta_{a_1,...,a_i} E_K(x_1)$ and $y_2 = \Delta_{a_1,...,a_i} E_K(x_2)$
3. If $y_1 = y_2$ output $i$ and stop
4. If $i \geq r$ output $i$ and stop
5. Set $i = i + 1$ and go to step (2)
```

If in step (3), $y_1 \neq y_2$ then the nonlinear order is greater than $i$ according to Prop. 4.3. If $y_1 = y_2$ then the nonlinear order may be greater than $i$, because it is possible for other values of $x'_1$ and $x'_2$ that $y'_1 \neq y'_2$. However the above test must stop, since if the $i$'th derivative of $f$ is constant, then the $i + r$'th derivative of $f$ is zero for all $r > 0$. Also, note that computing an $i$'th order derivative of $f$, is equivalent to computing two times an $i - 1$'st order derivative of $f$. Therefore the values of $y_1, y_2$ can be stored and re-used in following steps.

To test a block cipher $E$, pick a random key $K$ and two random plaintexts and run the test for nonlinear order. If the output of the test is $d$ then the nonlinear order of $E_K$ is at least $d$. Repeat this procedure for as many keys and plaintexts as desired. The input $r$ and the test in step (4) is necessary for block ciphers like the DES and $r$ should be chosen not much greater than 32, since it takes about $2^r$ encryptions to check a nonlinear order of $r$.

# 7 Concluding Remarks

We have shown applications for partial and higher order differentials. We presented ciphers secure against conventional differential attacks, but vulnerable to attacks using either partial or higher order differentials. We showed interesting partial and higher order differentials for the DES and presented a differential attack on DES with 6 rounds using partial differentials with complexity of about 46 chosen plaintexts and a running time of about the time of 3,500 encryptions. Finally we presented a method to test the nonlinear order of a block cipher using higher order differentials.

In the above attacks we have exploited the small number of rounds in the Feistel ciphers we have analysed. It is an open problem, whether differential attacks based on higher order differentials are applicable to ciphers with more than 5 rounds. This seems to require a method of iterating higher order differentials to more than two rounds in the same way as with first order differentials. Partial differentials can be combined with conventional differentials to refine attacks using the latter. It is an open problem whether partial differentials can improve the attacks on DES [1, 2] for more than 6 rounds.

# 8 Acknowledgements

The author wish to thank Luke O'Connor and Bart Preneel for helpful comments on an earlier version of this paper.

# References

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

[2] E. Biham and A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard.* Springer Verlag, 1993.

[3] M.E. Hellman, R. Merkle, R. Schroppel, L. Washington, W. Diffie, S. Pohlig, and P. Schweitzer. Results of an initial attempt to cryptanalyze the NBS Data Encryption Standard. Technical report, Stanford University, U.S.A., September 1976.

[4] M. E. Hellman and S. K. Langford. Differential–linear cryptanalysis. In Y. G. Desmedt, editor, *Advances in Cryptology - Proc. Crypto'94, LNCS 839*, pages 26–39. Springer Verlag, 1994.

[5] L.R. Knudsen. *Block Ciphers - Analysis, Design and Applications.* PhD thesis, Aarhus University, Denmark, 1994, DAIMI PB – 485.

[6] X. Lai. Higher order derivatives and differential cryptanalysis. In *Proc. "Symposium on Communication, Coding and Cryptography", in honor of James L. Massey on the occasion of his 60'th birthday, Feb. 10-13, 1994, Monte-Verita, Ascona, Switzerland*, 1994. To appear.

[7] K. Nyberg. Differentially uniform mappings for cryptography. In T. Helleseth, editor, *Advances in Cryptology - Proc. Eurocrypt'93, LNCS 765*, pages 55–64. Springer Verlag, 1993.

[8] K. Nyberg and L.R. Knudsen. Provable security against differential cryptanalysis. In E.F. Brickell, editor, *Advances in Cryptology - Proc. Crypto'92, LNCS 740*, pages 566–574. Springer Verlag, 1993.

[9] National Bureau of Standards. Data encryption standard. Federal Information Processing Standard (FIPS), Publication 46, National Bureau of Standards, U.S. Department of Commerce, Washington D.C., January 1977.

[10] B. Preneel, M. Nuttin, V. Rijmen, and J. Buelens. Differential cryptanalysis of the CFB mode. In D.R. Stinson, editor, *Advances in Cryptology - Proc. Crypto'93, LNCS 773*, pages 212–223. Springer Verlag, 1993.

[11] I. Schaumüller-Bichl. The method of formal coding. In *Cryptography - Proc., Burg Feuerstein, 1992, LNCS 149*, pages 235–255. Springer Verlag, 1982.

## Recent Publications in the BRICS Report Series

**RS-95-9** **Lars R. Knudsen.** ***Partial and Higher Order Differentials and Applications to the DES.*** **February 1995. 24 pp.**

**RS-95-8** **Ole I. Hougaard, Michael I. Schwartzbach, and Hosein Askari.** ***Type Inference of Turbo Pascal.*** **February 1995. 19 pp.**

**RS-95-7** **David A. Basin and Nils Klarlund.** ***Hardware Verification using Monadic Second-Order Logic.*** **January 1995. 13 pp.**

**RS-95-6** **Igor Walukiewicz.** ***A Complete Deductive System for the $\mu$-Calculus.*** **January 1995. 39 pp.**

**RS-95-5** **Luca Aceto and Anna Ingólfsdóttir.** ***A Complete Equational Axiomatization for Prefix Iteration with Silent Steps.*** **January 1995. 27 pp.**

**RS-95-4** **Mogens Nielsen and Glynn Winskel.** ***Petri Nets and Bisimulations.*** **January 1995. 36 pp. To appear in TCS.**

**RS-95-3** **Anna Ingólfsdóttir.** ***A Semantic Theory for Value–Passing Processes, Late Approach, Part I: A Denotational Model and Its Complete Axiomatization.*** **January 1995. 37 pp.**

**RS-95-2** **François Laroussinie, Kim G. Larsen, and Carsten Weise.** ***From Timed Automata to Logic - and Back.*** **January 1995. 21 pp.**

**RS-95-1** **Gudmund Skovbjerg Frandsen, Thore Husfeldt, Peter Bro Miltersen, Theis Rauhe, and Søren Skyum.** ***Dynamic Algorithms for the Dyck Languages.*** **January 1995. 21 pp.**

**RS-94-48** **Jens Chr. Godskesen and Kim G. Larsen.** ***Synthesizing Distinguishing Formulae for Real Time Systems.*** **December 1994. 21 pp.**

**RS-94-47** **Kim G. Larsen, Bernhard Steffen, and Carsten Weise.** ***A Constraint Oriented Proof Methodology based on Modal Transition Systems.*** **December 1994. 13 pp.**

**RS-94-46** **Amos Beimel, Anna Gál, and Mike Paterson.** ***Lower Bounds for Monotone Span Programs.*** **December 1994. 14 pp.**