Provable Security Against a Differential Attack*

Kaisa Nyberg ** and Lars Ramkilde Knudsen

Aarhus University, DK-8000 Aarhus C.

Abstract. The purpose of this paper is to show that there exist DES-like iterated ciphers, which are provably resistant against differential attacks. The main result on the security of a DES-like cipher with independent round keys is Theorem 1, which gives an upper bound to the probability of s-round differentials, as defined in [4] and this upper bound depends only on the round function of the iterated cipher. Moreover, it is shown that there exist functions such that the probabilities of differentials are less than or equal to 2^{3-n}, where n is the length of the plaintext block. We also show a prototype of an iterated block cipher, which is compatible with DES and has proven security against differential attacks.

Key words. DES-like ciphers, Differential cryptanalysis, Almost perfect nonlinear permutations, Markov Ciphers.

1 Introduction

A DES-like cipher is a block cipher based on iterating a function, called F, several times. Each iteration is called a round. The input to each round is divided into two halves. The right half is fed into F together with a round key derived from a key schedule algorithm. The output of F is added (modulo 2) to the left half of the input and the two halves are swapped except for the last round. The plaintext is the input to the first round and the ciphertext is the output of the last round.

In [1] E. Biham and A. Shamir introduced differential cryptanalysis of DES-like ciphers. In their attacks they make use of characteristics, which describe the behaviour of input and output differences for some number of consecutive rounds. The probability of a one-round characteristic is the conditional probability that given a certain difference in the inputs to the round we get a certain difference in the outputs of the round. Lai and Massey [4] observed that for the success of differential cryptanalysis it may not be necessary to fix the values of input and output differences for the intermediate rounds in a characteristic. They introduced the notion of *differentials.* The probability of an s-round differential is the conditional probability that given an input difference at the first round, the output difference at the s'th round will be some fixed value. Note that

* A preliminary version of this paper was presented in the rump session at Crypto '92 and will appear in the proceedings.

** The work of the author on this project is supported by MATINE Board, Finland.

the probability of an s-round differential with input difference A and output difference B is the sum of the probabilities of all s-round characteristics with input difference A and output difference B. For $s \leq 2$ the probabilities for a differential and for the corresponding characteristic are equal, but in general the probabilities for differentials will be higher.
In order to make a successful attack on a DES-like iterated cipher by differential cryptanalysis the existence of good characteristics is sufficient. On the other hand to prove security against differential attacks for DES-like iterated ciphers we must ensure that there is no differential with a probability high enough to enable successful attacks.

2 Resistance against differential attacks

A DES-like iterated cipher with block size $2n$ and with r rounds is defined as follows. Let

$$\begin{aligned} \mathbf{f} &: GF(2)^m \rightarrow GF(2)^n, \quad m \geq n \\ E &: GF(2)^n \rightarrow GF(2)^m, \text{ an affine expansion mapping} \end{aligned}$$

and let $K = (K_1, K_2, ..., K_r)$, where $K_i \in GF(2)^m$, be the r round keys. The round function (in the i'th round)

$$F : GF(2)^n \times GF(2)^m \rightarrow GF(2)^n$$

is then defined $F(X, K_i) = \mathbf{f}(E(X) + K_i)$, where '+' is the bitwise addition modulo 2.

Given a plaintext $X = (X_L, X_R)$ and a key $K = (K_1, K_2, ..., K_r)$ the ciphertext $Y = (Y_L, Y_R)$ is computed in r rounds. Set $X_L(0) = X_L$ and $X_R(0) = X_R$ and compute for $i = 1, 2, ..., r$

$$\begin{aligned} X_L(i) &= X_R(i-1) \\ X_R(i) &= F((X_R(i-1), K_i) + X_L(i-1) \\ X(i) &= (X_L(i), X_R(i)) \end{aligned}$$

Set $Y_L = X_R(r)$ and $Y_R = X_L(r)$.
The difference between two n-bit blocks is defined as

$$\Delta X = X + X^*$$

An s-round characteristic is an (s+1)-tuple $(\beta(0), \beta(1), ..., \beta(s))$ considered as the possible values of $(\Delta X(0), \Delta X(1), ..., \Delta X(s))$, whereas an s-round differential is a pair $(\beta(0), \beta(s))$ considered as the possible values of $(\Delta X(0), \Delta X(s))$ [1, 4]. To prove resistance against differential cryptanalysis we need to find the best differentials, so for the remainder of this paper we consider only differentials.

Differential attacks use s-round differentials to push forward the information of a fixed input difference at the first round to the s'th round independently of the used key. In this paper we will show that it is possible to choose the

round function so that no single differential is useful. Given a plaintext pair X, X^*, chosen by the cryptanalyst and r independent uniformly random round keys $K_1, K_2, ..., K_r$, unknown to the cryptanalyst, the differential may or may not hold. It is natural to measure the rate of success for the cryptanalyst by the probability of the differential taken over the distributions of X and K. The probability of a one-round differential $(\Delta X(0) = \alpha, \Delta X(1) = \beta)$ is

$$P(\Delta X(1) = \beta \mid \Delta X(0) = \alpha)$$

which by the property of Markov ciphers, as defined in [4], is equal to

$$P(\Delta X(1) = \beta \mid \Delta X(0) = \alpha,\ X = \gamma)$$

for all values γ of X, if the round key K is uniformly distributed. Hence the probability of a one round differential is independent of the distribution of X and is taken over the distribution of K. Assuming that the round keys $K_1, K_2, ..., K_r$ are mutually independent it follows that the probability of an s-round characteristic is the product of the probabilities of the individual rounds. Then the probability of an s-round differential equals (see also [4])

$$P(\Delta X(s) = \beta(s) \mid \Delta X(0) = \beta(0)) = \sum_{\beta(1)} \sum_{\beta(2)} .... \sum_{\beta(s-1)} \prod_{i=1}^{s} P(\Delta X(i) = \beta(i) \mid \Delta X(i-1) = \beta(i-1))$$

We denote by p_{max} the highest probability for a non-trivial one-round differential achievable by the cryptanalyst, i.e.

$$p_{max} = max_{\beta} max_{\alpha_R \neq 0} P(\Delta X(1) = \beta \mid \Delta X(0) = \alpha)$$

where α_R is the right half of α. We shall show in Sect. 3 that the round function of a DES-like cipher can be chosen in such a way that p_{max} is small.

Theorem 1 *It is assumed that in a DES-like cipher with* $\mathbf{f} : GF(2)^m \rightarrow GF(2)^n$ *the round keys are independent and uniformly random. Then the probability of an s-round differential,* $s \geq 4$, *is less than or equal to* $2p_{max}^2$.

Proof: We shall first give the proof for $s = 4$, i.e.

$$P(\Delta X(4) = \beta \mid \Delta X(0) = \alpha) \leq 2p_{max}^2$$

for any $\beta, \alpha\ (\neq 0)$. Let α_L, α_R and β_L, β_R be the left and right halves of α and β. We denote by $\Delta X_R(i)$ the right input differences at the i'th round, see Figure 1. Let $\delta \rightarrow \epsilon$ denote that, in order for the s-round differential (α, β) to occur, it is *necessary* that inputs to F with difference δ lead to outputs with difference ϵ. We split the proof into cases where $\beta_L = 0$ and $\beta_L \neq 0$.
Note that when $\beta_L = 0$ then $\beta_R \neq 0$, otherwise $\alpha_L = \alpha_R = \beta_L = \beta_R = 0$, which is of no use in differential cryptanalysis. Similarly if $\alpha_L = 0$ then $\alpha_R \neq 0$.

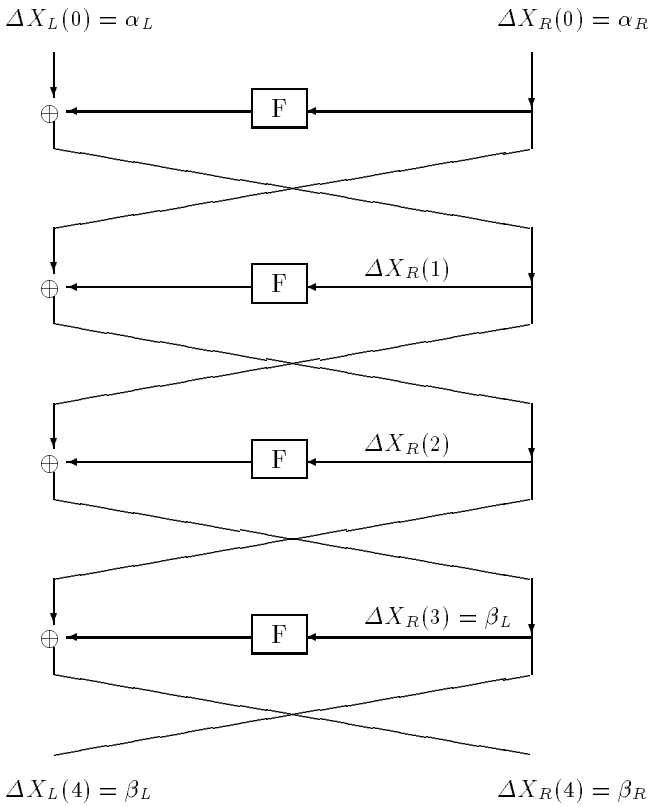

Fig. 1. The four round differential

1. $\beta_L = 0$. Then clearly $\Delta X_R(2) = \beta_R \neq 0$. If $\Delta X_R(1) = 0$ then $\Delta X_R(2) = \alpha_R = \beta_R \neq 0$. It then follows that $\alpha_R = \beta_R \rightarrow \alpha_L$ in the first round and $\Delta X_R(2) = \beta_R \rightarrow 0$ in the third round, both combinations with probability at most p_{max}. If $\Delta X_R(1) \neq 0$ then it follows that for any given $\Delta X_R(1)$ the second round must be $\Delta X_R(1) \rightarrow \alpha_R + \beta_R$ and the third round must be $\Delta X_R(2) = \beta_R \rightarrow \Delta X_R(1)$, both combinations with probability at most p_{max}. We obtain

$$
\begin{aligned}
&P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha) \\
&= \sum_{\Delta X_R(1)} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha)\, P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \\
&= P(\Delta X_R(1) = 0 \,|\, \Delta X(0) = \alpha)\, P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1) = 0) \\
&+ \sum_{\Delta X_R(1) \neq 0} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha)\, P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \\
&\leq p_{max}^2 + \sum_{\Delta X_R(1) \neq 0} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha) \cdot p_{max}^2
\end{aligned}
$$

$\leq 2p_{max}^2$

since $\sum_{\Delta X_R(1)\neq 0} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha) \leq 1$.

2. $\beta_L \neq 0$. We consider first the 3-round differential obtained by fixing $\Delta X_R(1)$. We obtain

$$\begin{aligned}
&P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \\
&\quad = \sum_{\Delta X_R(2)} P(\Delta X_R(2) \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \cdot \\
&\qquad\qquad P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1), \Delta X_R(2)) \\
&\quad = P(\Delta X_R(2) = 0 \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \cdot \\
&\qquad P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1), \Delta X_R(2) = 0) \\
&\qquad + \sum_{\Delta X_R(2)\neq 0} P(\Delta X_R(2) \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \cdot \\
&\qquad\qquad P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1), \Delta X_R(2)) \\
&\qquad \leq p_{max} \cdot p_{max} + \sum_{\Delta X_R(2)\neq 0} P(\Delta X_R(2) \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \cdot p_{max}^2 \\
&\qquad \leq 2p_{max}^2
\end{aligned}$$

The above shows that Theorem 1 holds for s-round differentials for $s \geq 3$ if $\beta_L \neq 0$. In the first inequality we used that if $\Delta X_R(2) = 0$ then $\Delta X_R(1) \neq 0$, since otherwise $\alpha_L = \alpha_R = 0$. Now

$$\begin{aligned}
&P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha) \\
&\quad = \sum_{\Delta X_R(1)} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha)\, P(\Delta X(4) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \\
&\quad \leq \sum_{\Delta X_R(1)} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha) \cdot 2p_{max}^2 \\
&\quad \leq 2p_{max}^2
\end{aligned}$$

Let now $s > 4$. Then

$$\begin{aligned}
&P(\Delta X(s) = \beta \,|\, \Delta X(0) = \alpha) \\
&= \sum_{\Delta X(s-4)} P(\Delta X(s-4) \,|\, \Delta X(0) = \alpha)\, P(\Delta X(s) = \beta \,|\, \Delta X(0) = \alpha, \Delta X(s-4))
\end{aligned}$$

Since we assumed that the round keys are independent and uniformly random it follows from the proof for $s = 4$ that

$$P(\Delta X(s) = \beta \,|\, \Delta X(0) = \alpha, \Delta X(s-4)) = P(\Delta X(s) = \beta \,|\, \Delta X(s-4)) \leq 2p_{max}^2$$

Thus $P(\Delta X(s) = \beta \,|\, \Delta X(0) = \alpha) \leq 2p_{max}^2$. □

If $\mathbf{f}$ is a permutation, Theorem 1 can be proved for $s \geq 3$. It comes from the fact that to have equal outputs of one round we must have equal inputs.

Theorem 2 *It is assumed that the function* $\mathbf{f}$ *in a DES-like cipher is a permutation and that the round keys are independent and uniformly random. Then the probability of an s-round differential for* $s \geq 3$ *is less than or equal to* $2p_{max}^2$.

Proof: We give the proof for $s = 3$. The general case can then be proved like in the preceding theorem. Again we separate between two cases and use the same notation as before.
1. $\beta_L = 0$. Then $\Delta X_R(0) = \alpha_R \neq 0$, otherwise different inputs would have to yield equal outputs in the second round, but that is not possible, since $\mathbf{f}$ is a permutation. The difference in the inputs at the first round is $\alpha_R \neq 0$ and the difference in the inputs at the second round is $\beta_R \neq 0$, thus
$P(\Delta X(3) = \beta \,|\, \Delta X(0) = \alpha) \leq p_{max}^2$.
2. $\beta_L \neq 0$. Like in the proof of Theorem 1 we split into cases where $\Delta X_R(1)$ is zero or not. Note that if $\Delta X_R(1) = 0 \Rightarrow \alpha_L \neq 0$ otherwise $\alpha_R \rightarrow \alpha_L = 0 \Rightarrow \alpha_R = 0$. We obtain

$$
\begin{aligned}
&P(\Delta X(3) = \beta \,|\, \Delta X(0) = \alpha) \\
&\quad = \sum_{\Delta X_R(1)} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha)\, P(\Delta X(3) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \\
&\quad = P(\Delta X_R(1) = 0 \,|\, \Delta X(0) = \alpha)\, P(\Delta X(3) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1) = 0) \\
&\quad + \sum_{\Delta X_R(1) \neq 0} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha)\, P(\Delta X(3) = \beta \,|\, \Delta X(0) = \alpha, \Delta X_R(1)) \\
&\quad \leq p_{max}^2 + \sum_{\Delta X_R(1) \neq 0} P(\Delta X_R(1) \,|\, \Delta X(0) = \alpha) \cdot p_{max}^2 \\
&\quad \leq 2p_{max}^2
\end{aligned}
$$

□

3 Almost perfect nonlinear permutations

First we show that the maximum probability p_{max} of a one-round differential has an upperbound that can be expressed in terms of the function $\mathbf{f}$. Let

$$p_{\mathbf{f}} = max_b max_{a \neq 0} P(\mathbf{f}(Y + a) + \mathbf{f}(Y) = b).$$

Then

$$
\begin{aligned}
p_{max} &= max_\beta max_{\alpha_R \neq 0} P(\Delta X(1) = \beta \,|\, \Delta X(0) = \alpha) \\
&= max_\beta max_{\alpha_R \neq 0} P(\alpha_L + \mathbf{f}(E(X + \alpha_R) + K) + \mathbf{f}(E(X) + K) = \beta_R) \\
&\leq p_{\mathbf{f}},
\end{aligned}
$$

where we have assumed that E is affine and denoted $K + E(X)$ by Y. If K is uniformly distributed then so is Y.
For a mapping $\mathbf{f} : GF(2)^m \rightarrow GF(2)^n$ the lower bound for $p_{\mathbf{f}}$ is 2^{-n}. Mappings attaining this lower bound were investigated in [7], where they are called

perfect nonlinear generalizing the definition of perfect nonlinearity given for Boolean functions in [6]. It was shown in [7] that perfect nonlinear mappings from $GF(2)^m \to GF(2)^n$ only exist for m even and $m \geq 2n$. Hence they can be adapted for use in DES-like ciphers only with expansion mappings that double the block length.
If the round function of a DES-like cipher does not involve any expansion, i.e. in the case when $\mathbf{f} : GF(2)^m \to GF(2)^n$ is a permutation, the trivial lower bound for $p_{\mathbf{f}}$ is 2^{1-n}, since then the difference

$$\mathbf{f}(\mathbf{x}+\mathbf{w}) + \mathbf{f}(\mathbf{x})$$

obtains half of the values in $GF(2)^n$ twice and never the other half of the values. We shall call the permutations with $p_{\mathbf{f}} = 2^{1-n}$ *almost perfect nonlinear.* The purpose of this section is to show that such permutations exist. For unexplained terminology we refer to [5].
Let $m = nd$, where n is odd. In [8] permutations $\mathbf{f}$ of $GF(2^m) = GF(2^d)^n$ were constructed to satisfy the following property:

(P) Every nonzero linear combination of the components of $\mathbf{f}$ is a nondegenerate quadratic form $\mathbf{x}^t\mathbf{C}\mathbf{x}$ in n indeterminates over $GF(2^d)$ with $rank(\mathbf{C}+\mathbf{C}^t) = n-1$.

It follows immediately from the definition that the coordinate functions of a permutation with (P) are complete, that is, depend on all input variables.
The main result of this section is the following theorem.

Theorem 3 *Let $\mathbf{f} : GF(2^d)^n \to GF(2^d)^n$, n odd, be a permutation satisfying (P). Then $p_{\mathbf{f}} = 2^{d(1-n)}$.*

Our proof of the theorem is based on the following three lemmata concerning properties of linear structures of quadratic forms. Recall that a linear structure $\mathbf{w}$ of $f : \mathbf{F}^n \to \mathbf{F}$, $\mathbf{F}$ a field, is a vector in $\mathbf{F}^n$ such that $f(\mathbf{x}+\mathbf{w}) + f(\mathbf{x})$ is constant as $\mathbf{x}$ varies. The linear structures of a quadratic form $f(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x}$ in n indeterminates over $GF(2^d)$ with $rank(\mathbf{A}+\mathbf{A}^t) = n-1$ form a one-dimensional linear subspace of $GF(2^d)^n$ (see [8], Prop. 3).
For a quadratic form f and every fixed $\mathbf{w}$ the function $f(\mathbf{x}+\mathbf{w}) + f(\mathbf{x})$ of $\mathbf{x}$ is affine or constant. From this we get the first lemma.

Lemma 1 *Let $\mathbf{w} \in GF(2^d)^n$ be not a linear structure of $f : GF(2^d)^n \to GF(2^d)$, $f(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x}$. Then the function $f(\mathbf{x}+\mathbf{w}) + f(\mathbf{x})$ of $\mathbf{x}$ is balanced, i.e. obtains each value in $GF(2^d)$ equally many times.*

Lemma 2 *Let $f(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x}$ be a quadratic form in n indeterminates over $GF(2)$ such that $rank(\mathbf{A}+\mathbf{A}^t) = n-1$. Then f is nondegenerate if and only if $f(\mathbf{w}) \neq 0$ for the nonzero linear structures $\mathbf{w}$ of f* (see also Lemma 4.1. in [3]).

Proof: Let

$$\varphi(x_1, \ldots, x_n) = x_1x_2 + \ldots + x_{n-2}x_{n-1} + \delta x_n^2$$

$\delta = 0$ or 1, be the quadratic forms to which all quadratic forms $f(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x}$ with $rank(\mathbf{A} + \mathbf{A}^t) = n - 1$ are equivalent (see [5], Ch.6.2). It means that there is a linear transformation $\mathbf{T}$ of coordinates such that $f(\mathbf{x}) = \varphi(\mathbf{T}\mathbf{x})$. Then $\mathbf{w}$ is a linear structure of f if and only if $\mathbf{T}\mathbf{w} = (0, 0, \ldots, 0, a)$, where $a \in GF(2^d)$. Then f is nondegenerate if and only if φ is nondegenerate which is true if and only if $\delta = 1$. But $\delta = 1$ if and only if

$$f(\mathbf{w}) = \varphi(\mathbf{T}\mathbf{w}) = \varphi(0, \ldots, 0, a) = \delta a^2 \neq 0$$

for $a \neq 0$. □

Lemma 3 *Let $\mathbf{f} : GF(2^d)^n \to GF(2^d)^n$ be a permutation with property (P). Then every nonzero $\mathbf{w} \in GF(2^d)^n$ is a linear structure of a nonzero linear combination of the components of $\mathbf{f}$.*

Proof: Let $\mathbf{u}$ be a nonzero vector in $GF(2^d)^n$ and let $\mathbf{w} \in GF(2^d)^n$ be a nonzero linear structure of $\mathbf{u} \cdot \mathbf{f}$. Then $\lambda\mathbf{w}$, $\lambda \in GF(2^d)$ are the linear structures of $c\mathbf{u} \cdot \mathbf{f}$, $c \in GF(2^d)$. Hence it suffices to show that if $\mathbf{u}_1 \cdot \mathbf{f}$ and $\mathbf{u}_2 \cdot \mathbf{f}$ share a nonzero linear structure then there is $c \in GF(2^d)$ such that $\mathbf{u}_1 = c\mathbf{u}_2$.
Let $\mathbf{w}$ be a nonzero linear structure of $\mathbf{u}_1 \cdot \mathbf{f}$ and $\mathbf{u}_2 \cdot \mathbf{f}$. Then $\mathbf{w}$ is also the linear structure of $(c_1\mathbf{u}_1 + c_2\mathbf{u}_2) \cdot \mathbf{f}$, for all $c_1, c_2 \in GF(2^d)$. Since $\mathbf{u}_1 \cdot \mathbf{f}$ and $\mathbf{u}_2 \cdot \mathbf{f}$ are nondegenerate it follows from Lemma 2 that

$$\mathbf{u}_1 \cdot \mathbf{f}(\mathbf{w}) \neq 0 \,\text{and}\, \mathbf{u}_2 \cdot \mathbf{f}(\mathbf{w}) \neq 0.$$

Hence there exists $c \neq 0$ such that $c\mathbf{u}_1 \cdot \mathbf{f}(\mathbf{w}) = \mathbf{u}_2 \cdot \mathbf{f}(\mathbf{w})$ or, what is the same,

$$(c\mathbf{u}_1 + \mathbf{u}_2) \cdot \mathbf{f}(\mathbf{w}) = 0.$$

If $c\mathbf{u}_1 \neq \mathbf{u}_2$, then $(c\mathbf{u}_1 + \mathbf{u}_2) \cdot \mathbf{f}$ is nondegenerate which cannot be true by Lemma 2. Consequently $c\mathbf{u}_1 = \mathbf{u}_2$. □
Now Theorem 3 is a consequence of the following

Theorem 4 *Let $\mathbf{f} = (f_1, f_2, \ldots, f_n)$: $GF(2^d)^n \to GF(2^d)^n$ be a permutation that satisfies (P). Then for every fixed nonzero difference $\mathbf{w} \in GF(2^d)^n$ of the inputs to $\mathbf{f}$, the differences of the outputs lie in an affine hyperplane of $GF(2^d)^n$ and are uniformly distributed there.*

Proof: Let $\mathbf{w}$ be a nonzero input difference for $\mathbf{f}$. Then by Lemma 3 there is $\mathbf{v} \in GF(2^d)^n$, $\mathbf{v} \neq 0$, such that $\mathbf{w}$ is the linear structure of $\mathbf{v} \cdot \mathbf{f}$ and by Lemma 2

$$\mathbf{v} \cdot \mathbf{f}(\mathbf{x} + \mathbf{w}) + \mathbf{v} \cdot \mathbf{f}(\mathbf{x}) = \mathbf{v} \cdot \mathbf{f}(\mathbf{w}) \neq 0$$

for all $\mathbf{x} \in GF(2^d)^n$. We denote $b_0 = \mathbf{v} \cdot \mathbf{f}(\mathbf{w})$.
Let $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ be linearly independent vectors in $GF(2^d)^n$ such that

$$\mathbf{v} \notin \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}.$$

Then by Lemma 1 for every $\mathbf{u} \in \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$ the function

$$\mathbf{x} \mapsto \mathbf{u} \cdot \mathbf{f}(\mathbf{x} + \mathbf{w}) + \mathbf{u} \cdot \mathbf{f}(\mathbf{x})$$

obtains each value in $GF(2^d)$ equally many times. Consequently (see [5]), for every $(b_1, \ldots, b_{n-1}) \in GF(2^d)^{n-1}$, the system of equations

$$\mathbf{u}_i \cdot \mathbf{f}(\mathbf{x}+\mathbf{w}) + \mathbf{u}_i \cdot \mathbf{f}(\mathbf{x}) = b_i,\ i = 1, \ldots, n-1,$$

has 2^d solutions $\mathbf{x} \in GF(2^d)^n$. Hence the system of n equations:

$$\begin{aligned} (2) \quad \mathbf{u}_i \cdot \mathbf{f}(\mathbf{x}+\mathbf{w}) + \mathbf{u}_i \cdot \mathbf{f}(\mathbf{x}) &= b_i, i = 1, \ldots, n-1, \\ \mathbf{v} \cdot \mathbf{f}(\mathbf{x}+\mathbf{w}) + \mathbf{v} \cdot \mathbf{f}(\mathbf{x}) &= b \end{aligned}$$

has 2^d solutions if $b = b_0$ and no solutions if $b \neq b_0$. Every system of n equations

$$f_i(\mathbf{x}+\mathbf{w}) + f_i(\mathbf{x}) = a_i,\ i = 1, 2, \ldots, n$$

is a linear transformation of (2), from which the claim follows. □

By a similar argumentation one can prove the following generalization of Theorem 3.

Theorem 5 *Let* $\mathbf{f}$ *be a permutation in* $GF(2^d)^n$, n *odd, with property* (P) *and let* $f_1, \ldots, f_n$ *be the components of* $\mathbf{f}$ *with respect to some arbitrary fixed basis over* $GF(2^d)$. *Let* $l \leq n$ *and set* $\mathbf{h} = (f_1, f_2, \ldots, f_l)$. *Then* $p_{\mathbf{h}} = 2^{d(1-l)}$.

From the results of Section 2 we now obtain

Theorem 6 *Assume that in a DES-like cipher the function* $\mathbf{f}$ *is a mapping from* $GF(2)^{nd}$ *to* $GF(2)^{ld}$, $n \geq l$, *obtained from a permutation in* $GF(2^d)^n$ *with* (P) *by discarding* $n-l$ *output coordinates. Then* $p_{\mathbf{f}} = 2^{d(1-l)}$. *Moreover, if* $n > l$, *then the probability of every* r*-round differential,* $r \geq 4$, *is less than or equal to* $2^{2d(1-l)+1}$, *assuming that the round keys are uniformly random and independent. If* $n = l$, *the probability of every* r*-round differential,* $r \geq 3$, *is less than or equal to* $2^{2d(1-l)+1}$.

4 Class of permutations with property (P)

In this section we show that the permutations $\mathbf{f}(\mathbf{x}) = \mathbf{x}^{2^k+1}$ in $GF(2^{nd})$ with $k = 0 \bmod d$, $gcd(k, n) = 1$, and n odd, have property (P), when considered as permutations of $GF(2^d)^n$.

Let $\alpha_1, \ldots, \alpha_n$ be a basis in $GF(2^{nd})$ over $GF(2^d)$ and $\beta_1, \ldots, \beta_n$ be its dual basis. Let $\mathbf{x} = \sum_{i=1}^n x_i \alpha_i$, $x_i \in GF(2^d)$. Then the i'th component $f_i(\mathbf{x})$ of $\mathbf{f}(\mathbf{x})$ with respect to the basis $\alpha_1, \ldots, \alpha_n$ is

$$\begin{aligned} f_i(\mathbf{x}) &= Tr(\beta_i \mathbf{x}^{2^k+1}) \\ &= Tr(\beta_i (\sum_{j=1}^n x_j \alpha_j)(\sum_{l=1}^n x_l \alpha_l)^{2^k}) \\ &= \sum_{j=1}^n \sum_{l=1}^n Tr(\beta_i \alpha_j \alpha_l^{2^k}) x_j x_l \\ &= \sum_{j=1}^n \sum_{l=1}^n Tr(\gamma_i \alpha_j (\gamma_i \alpha_l)^{2^k}) x_j x_l \end{aligned}$$

where $\gamma_i \in GF(2^{nd})$ is such that $\gamma_i^{2^k+1} = \beta_i$, $i = 1, 2, ..., n$.
Now it is straightforward to check that $Tr(\gamma_i \alpha_j (\gamma_i \alpha_l)^{2^k}) \in GF(2^d)$ is the entry on the j'th row and l'th column in the matrix $\mathbf{A}_i = \mathbf{B}_i^t \mathbf{R}^k \mathbf{B}_i$ where

$$\mathbf{B}_i = \begin{pmatrix} \gamma_i\alpha_1 & \gamma_i\alpha_2 & \cdots & \gamma_i\alpha_n \\ (\gamma_i\alpha_1)^2 & (\gamma_i\alpha_2)^2 & \cdots & (\gamma_i\alpha_n)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (\gamma_i\alpha_1)^{2^{n-1}} & (\gamma_i\alpha_2)^{2^{n-1}} & \cdots & (\gamma_i\alpha_n)^{2^{n-1}} \end{pmatrix}$$

is a $n \times n$ regular matrix over $GF(2^{nd})$ and

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

is the cyclic shift for which $rank(\mathbf{R}^k + (\mathbf{R}^k)^t) = n - 1$ if $gcd(k, n) = 1$. Then also $\mathbf{x}^t \mathbf{R} \mathbf{x}$ is nondegenerate. Consequently,

$$f_i(\mathbf{x}) = \mathbf{x}^t \mathbf{A}_i \mathbf{x}$$

and

$$rank(\mathbf{A}_i + \mathbf{A}_i^t) = rank(\mathbf{B}_i^t(\mathbf{R}^k + (\mathbf{R}^k)^t)\mathbf{B}_i) = rank(\mathbf{R}^k + (\mathbf{R}^k)^t) = n - 1$$

over $GF(2^{nd})$. Thus $rank(\mathbf{A}_i + \mathbf{A}_i^t) = n - 1$ also over $GF(2^d)$, since the rank does not decrease when going to a subfield and it cannot be n. By the linearity of the trace function the same holds for every nonzero linear combination of the components f_i of $\mathbf{f}$. This completes the proof of property (P) for $\mathbf{f}$.

5 A prototype of a DES-like cipher for encryption

Let $\mathbf{g}(\mathbf{x}) = \mathbf{x}^3$ in $GF(2^{33})$. There are several efficient ways of implementing this power polynomial and each of them suggest a choice of a basis in $GF(2^{33})$. Let us fix a basis and discard one output coordinate. Then we have a function $\mathbf{f} : GF(2)^{33} \to GF(2)^{32}$. The 64-bit plaintext block is divided into two 32-bit halves L and R. The plaintext expansion is an affine mapping $E : GF(2)^{32} \to GF(2)^{33}$. Each round take a 32 bit input and a 33 bit key. The round function is $L \| R \mapsto R \| L + \mathbf{f}(E(R) + K)$.
In [2] E. Biham and A. Shamir introduced an improved differential attack on 16-round DES. This means, that in general for an r-round DES-like cipher the existence of an $(r-2)$-round differential with a sufficiently high probability may enable a successful differential attack. From Theorem 6 we have that every four and five round differential of this block cipher has probability less than or equal to 2^{-61}. Therefore we suggest at least six rounds for the block cipher. All round

keys should be independent, therefore we need at least 198 key bits.
More examples of permutations $\mathbf{f}$ for which p_{max} is low can be found in [9]. The examples include the inverses of $x \mapsto x^{2^k+1}$ and the mappings $x \mapsto x^{-1}$, whose coordinate functions are of higher nonlinear order than quadratic.

6 Acknowledgements

We would like to thank D. Coppersmith and an anonymous referee for comments that improved the paper.

References

1. E. Biham, A. Shamir. *Differential Cryptanalysis of DES-like Cryptosystems.* Journal of Cryptology, Vol. 4 No. 1 1991.
2. E. Biham, A. Shamir. *Differential Cryptanalysis of the full 16-round DES.* Technical Report # 708, Technion - Israel Institute of Technology.
3. P. Camion, C. Carlet, P. Charpin, N. Sendrier. *On Correlation-immune functions.* Advances in Cryptology - Crypto '91. Lecture Notes in Computer Science 576, Springer-Verlag, 1992, pp. 86-100.
4. X. Lai, J. L. Massey, S. Murphy. *Markov Ciphers and Differential Cryptanalysis.* Advances in Cryptology - Eurocrypt '91. Lecture Notes in Computer Science 547, Springer-Verlag, 1992, pp. 17-38.
5. R. Lidl, H. Niederreiter. *Finite Fields.* Encyclopedia of Mathematics and its applications, Vol. 20. Addison-Wesley, Reading, Massachusetts, 1983.
6. W. Meier, O. Staffelbach. *Nonlinearity criteria for cryptographic functions.* Advances in Cryptology - Eurocrypt '89. Lecture Notes in Computer Science, 434, Springer-Verlag, 1990, pp. 549-562.
7. K. Nyberg. *Perfect nonlinear S-boxes.* Advances in Cryptology - Proceedings of Eurocrypt '91. Lecture Notes in Computer Science 547, Springer Verlag, 1991, pp. 378-386.
8. K. Nyberg. *On the construction of highly nonlinear permutations.* Advances in Cryptology - Eurocrypt '92. Lecture Notes in Computer Science, 658, Springer-Verlag, 1993, pp. 92-98.
9. K. Nyberg. *Differentially uniform mappings for cryptography.* Proceedings of Eurocrypt '93 (to appear).