Provable Security Against Differential Cryptanalysis

Kaisa Nyberg *[1] and Lars Ramkilde Knudsen[2]

[1] Finnish Defence Forces, University of Helsinki, (on leave)***
[2] Aarhus University, DK-8000 Aarhus C.

1 Introduction

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 r-round differentials, as defined in [3] 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^{2-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.

2 Differential Cryptanalysis of DES-like iterated ciphers

A DES-like cipher is a block cipher based on iterating a function, called $\mathbf{F}$, several times. Each iteration is called a round. The input to each round is divided into two halves. The right half is fed into $\mathbf{F}$ together with a round key derived from a keyschedule algorithm. The output of $\mathbf{F}$ 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.

Notation: Let the block size of the cipher be $2n$ and the size of the round key be m, $m \geq n$. Let $\mathbf{f} : GF(2)^m \rightarrow GF(2)^n$ and $E : GF(2)^n \rightarrow GF(2)^m$, an affine expansion mapping. Let L_i, R_i be the left and right halves of the input to the i'th round. Then $L_{i+1} = R_i$ and $R_{i+1} = \mathbf{f}(E(R_i) \oplus K_i) \oplus L_i$ and $\mathbf{F}(R_i, K_i) = \mathbf{f}(E(R_i) \oplus K_i)$.

In [1] Biham and 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

* The work of the author on this project is supported by MATINE Board, Finland.
*** Current address: Prinz Eugen-Straße 18/6, A-1040 Vienna.

E.F. Brickell (Ed.): Advances in Cryptology - CRYPTO '92, LNCS 740, pp. 566-574, 1993.

in the outputs of that round. Assume that in every round the inputs $E(R) \oplus K$ to $\mathbf{f}$ are independent and random. This assumption is satisfied if the round keys are uniformly random and independent. Then the probability of an r-round characteristic is obtained by multiplying the probabilities of the r one-round characteristics.

Lai and Massey [3] observed that for the success of differential cryptanalysis it is not 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 r-round differential is the conditional probability that given an input difference at the first round, the output difference at the r'th round will be some fixed value. Note that the probability of an r-round differential with input difference A and output difference B is the sum of the probabilities of all r-round characteristics with input difference A and output difference B. For $r \leq 2$ 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.

The difference of two inputs $E(R) \oplus K$ and $E(R^*) \oplus K$ to $\mathbf{f}$ is $E(R) \oplus E(R^*)$. Since we assume E to be affine, the difference of two inputs depends only on the difference $R \oplus R^*$. Hence for DES-like ciphers the round probabilities of characteristics only depend on the intrinsic properties of $\mathbf{f}$. Given $\mathbf{f} : GF(2)^m \rightarrow GF(2)^n$ denote

$$p_{max} = 2^{-m} max_{\beta} max_{\alpha \neq 0} \#\{X \in GF(2)^m | \mathbf{f}(X \oplus \alpha) \oplus \mathbf{f}(X) = \beta\}$$

That is, p_{max} is the highest probability for a non trivial one-round characteristic or differential.

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

Proof: We shall first give the proof for $r = 4$. Let α_L and α_R be the left and right halves of the input difference at the first round and β_L and β_R be corresponding halves of the output difference at the last round. Either $\beta_L \neq 0$ or $\beta_R \neq 0$ or both. We shall give the proof in the case $\beta_L = 0$, $\beta_R \neq 0$, the other two cases are similar. We denote by $\Delta R(i)$ the right input differences to the i'th round, $i = 2, 3, 4$. Let $\beta_L = 0$ and $\beta_R \neq 0$. Then $\Delta R(4) = \beta_L = 0$ and $\Delta R(3) = \beta_R$. We separate between two cases: $\alpha_R \neq \beta_R$ and $\alpha_R = \beta_R$.

1. $\alpha_R \neq \beta_R$. Then $\Delta R(2) \neq 0$. For any given $\Delta R(2) \neq 0$ there is exactly one way of getting β_L, β_R from the input differences α_R and $\Delta R(2)$ at the second round, and the probability is less than or equal to p_{max}^2. Hence the probability of the four round differential is less than or equal to p_{max}^2.

2. $\alpha_R = \beta_R$. If $\Delta R(2) = 0$ it follows that the output difference from $\mathbf{F}$ at the

third round $\Delta\mathbf{F}(R(3)) = 0$, which happens with probability less than or equal to p_{max}, because $\Delta R(3) = \beta_R \neq 0$. Since $\alpha_R \neq 0$ we have

$$Prob(\Delta R(2) = 0 \,|\, \alpha_L, \alpha_R) \leq p_{max}$$

If $\Delta R(2) \neq 0$ the probability that $\Delta\mathbf{F}(R(3)) = \Delta R(2)$ is less than or equal to p_{max}. We also need to have $\Delta\mathbf{F}(R(2)) = 0$, which is true with probability less than or equal to p_{max}. So we obtain

$$\begin{aligned}
&Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R) \\
&\quad = \sum_{\Delta R(2)} Prob(\Delta R(2) \,|\, \alpha_L, \alpha_R)\, Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R, \Delta R(2)) \\
&\quad = Prob(\Delta R(2) = 0 \,|\, \alpha_L, \alpha_R)\, Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R, \Delta R(2) = 0) \\
&\quad + \sum_{\Delta R(2) \neq 0} Prob(\Delta R(2) \,|\, \alpha_L, \alpha_R)\, Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R, \Delta R(2)) \\
&\quad \leq p_{max}^2 + \sum_{\Delta R(2) \neq 0} Prob(\Delta R(2) \,|\, \alpha_L, \alpha_R) \cdot p_{max}^2 \\
&\quad \leq 2p_{max}^2
\end{aligned}$$

Let now $r > 4$. Then

$$\begin{aligned}
&Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R) \\
&\qquad = \sum_{\Delta L(r-3), \Delta R(r-3)} [Prob(\Delta L(r-3), \Delta R(r-3) \,|\, \alpha_L, \alpha_R) \cdot \\
&\qquad\qquad Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R, \Delta L(r-3), \Delta R(r-3))]
\end{aligned}$$

Since we assumed that the inputs to $\mathbf{f}$ are independent and uniformly random it follows from the proof for $r = 4$ that

$$\begin{aligned}
&Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R, \Delta L(r-3), \Delta R(r-3)) = \\
&\qquad\qquad Prob(\beta_L, \beta_R \,|\, \Delta L(r-3), \Delta R(r-3)) \leq 2p_{max}^2
\end{aligned}$$

Thus $Prob(\beta_L, \beta_R \,|\, \alpha_L, \alpha_R) \leq 2p_{max}^2$. □

If $\mathbf{f}$ is a permutation, then in every characteristic between two zero rounds there has to be at least two nonzero rounds and the following result can be proved.

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

Proof: We give the proof for $r = 3$. The general case can then be proved like in the preceding theorem. Again we separate between three cases and use the same notation as before.

1. $\beta_L = 0,\ \beta_R \neq 0$. In this case the third round of each characteristic is a zero-round. At the second round the input difference $\Delta R(2) = \beta_R \neq 0$ results in

an output difference $\alpha \neq 0$ with probability less than or equal to p_{max}. At the first round we get the output difference $\alpha_L \oplus \beta_R \neq 0$ with probability less than or equal to p_{max} from the input difference $\Delta R(1) = \alpha_R \neq 0$. Hence $Prob(\beta_L, \beta_R \mid \alpha_L, \alpha_R) \leq p_{max}^2$.

2. $\beta_L \neq 0$, $\beta_R = 0$. Now the output difference at the third round equals $\Delta R(2)$ and it is different from zero. Given $\Delta R(2) \neq 0$ the probability of the third round is less than or equal to p_{max} and the same holds for the second round. Consequently

$$\begin{aligned} &Prob(\beta_L, \beta_R \mid \alpha_L, \alpha_R) \\ &= \sum_{\Delta R(2) \neq 0} Prob(\Delta R(2) \mid \alpha_L, \alpha_R)\, Prob(\beta_L, \beta_R \mid \alpha_L, \alpha_R, \Delta R(2)) \\ &\leq \sum_{\Delta R(2)} Prob(\Delta R(2) \mid \alpha_L, \alpha_R) \cdot p_{max}^2 \;\leq\; p_{max}^2 \end{aligned}$$

3. $\beta_L \neq 0$, $\beta_R \neq 0$. Assume first that $\Delta R(2) = 0$. Then for every characteristic the probability of the third round is less than or equal to p_{max}, the probability of the second round is one and the probability of the first round is less than or equal to p_{max}. Secondly, given $\Delta R(2) \neq 0$, the probability of the third round is less than or equal to p_{max} and the same is true for the second round. Hence $Prob(\beta_L, \beta_R \mid \alpha_L, \alpha_R) \leq p_{max}^2$ also in this case. □

3 Almost perfect nonlinear permutations

For a mapping $\mathbf{f} : GF(2)^m \rightarrow GF(2)^n$ the lower bound for p_{max} 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 $GF(2)^m \rightarrow GF(2)^n$ 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 $\mathbf{f} : GF(2)^m \rightarrow GF(2)^n$ is a permutation, the trivial lower bound for p_{max} 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 $p_{max} = 2^{1-n}$ *almost perfect nonlinear*. The purpose of this section is to show that such permutations exist.

Assume that $m = nd$, where m, n, d are all odd integers. In [8] permutations $\mathbf{f}$ in $GF(2^m) = GF(2^d)^n$ were constructed to satisfy the following property:

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

Indeed the following theorem holds.

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

For the sake of simplicity we shall give the proof in the case where $d = 1$ and $m = n$.

Lemma 1 *A quadratic form* $f(\mathbf{x}) = \mathbf{x^t A x}$ *in* n *indeterminates over* $GF(2)$ *is balanced if and only if* $f(\mathbf{w}) \neq 0$ *for the linear structure* $\mathbf{w}$ *of* f.

Recall that a linear structure $\mathbf{w}$ of $f : \mathbf{F}^n \to \mathbf{F}$ is a nonzero vector in $\mathbf{F}^n$ such that $f(\mathbf{x} + \mathbf{w}) + f(\mathbf{x})$ is constant. It was also shown in [8] that a quadratic form $f(\mathbf{x}) = \mathbf{x^t A x}$ in n indeterminates over $GF(2)$ with $rank(\mathbf{A} + \mathbf{A^t}) = n - 1$ has exactly one linear structure.
Proof of Lemma 1: Let

$$\varphi(x_1, ..., x_n) = x_1x_2 + ... + x_{n-2}x_{n-1} + \delta x_n = \mathbf{x^t C x}$$

$\delta = 0$ or 1, be the quadratic forms to which all quadratic forms $f(\mathbf{x}) = \mathbf{x^t A x}$ with $rank(\mathbf{A} + \mathbf{A^t}) = n - 1$ are equivalent (see [4]). It means that there is a linear transformation $\mathbf{T}$ of coordinates such that $f(\mathbf{x}) = \varphi(\mathbf{Tx})$. Then $\mathbf{w}$ is a linear structure of f if and only if $\mathbf{Tw} = (0, 0, ..., 0, 1)$. Then f is balanced if and only if φ is balanced which is true if and only if $\delta = 1$. But $\delta = 1$ if and only if

$$f(\mathbf{w}) = \varphi(\mathbf{Tw}) = \varphi(0, ..., 0, 1) = 1.$$

□

Lemma 2 *Let* $\mathbf{w} \in GF(2)^n$ *be not the linear structure of* $f : GF(2)^n \to GF(2)$, $f(\mathbf{x}) = \mathbf{x^t A x}$ *with* $rank(\mathbf{A} + \mathbf{A^t}) = n - 1$. *Then*

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

is balanced.

Proof: It sufficies to show that

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

is balanced for every $\mathbf{w} \neq (0, ..., 0, 1)$. But this is true since

$$\begin{aligned}&\varphi(\mathbf{x} + \mathbf{w}) + \varphi(\mathbf{x}) = \\ &(x_1 + w_1)(x_2 + w_2) + ... + (x_{n-2} + w_{n-2})(x_{n-1} + w_{n-1}) + x_n + w_n + \\ &x_1x_2 + ... + x_{n-2}x_{n-1} + x_n.\end{aligned}$$

is a non-constant affine or linear function for every $\mathbf{w} \neq (0, ..., 0, 1)$. □

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

Proof: It sufficies to show that two different linear combinations of the components of $\mathbf{f}$ have different linear structures. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be nonzero vectors in $GF(2)^n$ and let $\mathbf{w}_1$ and $\mathbf{w}_2$ be the linear structures of $\mathbf{u}_1 \cdot \mathbf{f}$ and $\mathbf{u}_2 \cdot \mathbf{f}$, respectively. If $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$ it follows that $\mathbf{w}$ is also the linear structure of $(\mathbf{u}_1 + \mathbf{u}_2) \cdot \mathbf{f}$. Since $\mathbf{u}_1 \cdot \mathbf{f}$ and $\mathbf{u}_2 \cdot \mathbf{f}$ are balanced it follows from Lemma 1 that

$$\mathbf{u}_1 \cdot \mathbf{f}(\mathbf{w}) = \mathbf{u}_2 \cdot \mathbf{f}(\mathbf{w}) = 1$$

and consequently

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

If $\mathbf{u}_1 \neq \mathbf{u}_2$, then $(\mathbf{u}_1 + \mathbf{u}_2) \cdot \mathbf{f}$ is balanced. Thus by Lemma 1, $\mathbf{u}_1 = \mathbf{u}_2$. □
Now Theorem 3 for $d = 1$ is a consequence of the following

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

Proof: Let $\mathbf{w}$ be a nonzero input difference for $\mathbf{f}$. Then by Lemma 3 there is $\mathbf{v} \in GF(2)^n$, $\mathbf{v} \neq 0$, such that $\mathbf{w}$ is the linear structure of $\mathbf{v} \cdot \mathbf{f}$ and by Lemma 1

$$\mathbf{v} \cdot \mathbf{f}(\mathbf{x} + \mathbf{w}) + \mathbf{v} \cdot \mathbf{f}(\mathbf{x}) = 1$$

for all $\mathbf{x} \in GF(2)^n$.
Let $\mathbf{u}_1, ..., \mathbf{u}_{n-1}$ be linearly independent vectors in $GF(2)^n$ such that

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

Then by Lemma 2 for every $\mathbf{u} \in span\{\mathbf{u}_1, ..., \mathbf{u}_{n-1}\}$ the function

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

is balanced, which means (see [4]) that for every $(b_1, ..., b_{n-1}) \in GF(2)^{n-1}$ the system of equations

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

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

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

has 2 solutions if $b = 1$ and no solutions if $b = 0$. Every system of n equations

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

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

From the results in Section 2 we now obtain

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

4 Examples of permutations with property (P)

Pieprzyk [9] observed that the permutations $\mathbf{f}(\mathbf{x}) = \mathbf{x}^{2^k+1}$ in $GF(2^n)$ with $gcd(k, n) = 1$, $1 \leq k < n$ and n odd are at a large distance from the linear mappings. We shall show that these permutations have property (P).
Let $\alpha_1, ..., \alpha_n$ be a basis in $GF(2^n)$ over $GF(2)$ and $\beta_1, ..., \beta_n$ be its dual basis. Let $\mathbf{x} = \sum_{i=1}^{n} x_i\alpha_i$, $x_i \in GF(2)$. Then the i'th component $f_i(\mathbf{x})$ of $\mathbf{f}(\mathbf{x})$ with respect to the basis $\alpha_1, ..., \alpha_n$ 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 $\gamma_i \in GF(2^n)$ is such that $\gamma_i^{2^k+1} = \beta_i$, $i = 1, 2, ..., n$.
Now it is straightforward to check that $Tr(\gamma_i\alpha_j(\gamma_i\alpha_l)^{2^k})$ is the jl'th entry in the matrix $\mathbf{A}_i = \mathbf{B}_i^t \mathbf{R}^k \mathbf{B}_i$ where

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

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

$$\mathbf{R} = \begin{pmatrix} 0\,1\,0\,\ldots\,0 \\ 0\,0\,1\,\ldots\,0 \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ 0\,0\,0\,\ldots\,1 \\ 1\,0\,0\,\ldots\,0 \end{pmatrix}$$

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

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

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 $GF(2^n)$. Thus $rank(\mathbf{A}_i + \mathbf{A}_i^t) = n - 1$ also over $GF(2)$, 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 $\mathbf{f}$. Moreover, since $\mathbf{f}$ is a permutation, they are all balanced, which completes the proof of property (P) for $\mathbf{f}$.
Matsumoto and Imai proposed in [5] a public key cryptosystem C^*, which is based on power polynomials $\mathbf{x}^{2^k+1}$. If the round function of an iterated DES-like cipher of block size 64 makes use of the mapping $\mathbf{x}^{2^k+1}$ as proposed below in Section 5, the description of the round function for efficient implementation would be less than the minimum size of the public key for C^* cryptosystem.

5 A prototype of a DES-like cipher for encryption

Let $\mathbf{g}(\mathbf{x}) = \mathbf{x}^3$ in $GF(2^{37})$. There are several efficient ways of implementing this power polynomial and each of them suggest a choice of a basis in $GF(2^{37})$. Let us fix a basis and discard five output coordinates. Then we have a function $\mathbf{f} : GF(2)^{37} \rightarrow GF(2)^{32}$. The 64-bit plaintext block is divided into two 32-bit halves L and R. The plaintext expansion is an affine mapping $E : GF(2)^{32} \rightarrow GF(2)^{37}$. Each round take a 32 bit input and a 37 bit key. The round function is $L||R \mapsto R||\, L \oplus \mathbf{f}(E(R) \oplus K)$.
In [2] Biham and 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 too 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 222 key bits. This is equivalent to four DES keys, where all parity bits plus two other bits are discarded.

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. 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.
4. R. Lidl, H. Niederreiter. *Finite Fields.* Encyclopedia of Mathematics and its applications, Vol. 20. Addison-Wesley, Reading, Massachusetts, 1983.
5. T. Matsumoto, H. Imai. *Public quadratic polynomial-tuples for efficient signature-verification and message-encryption.* Advances in Cryptology - Eurocrypt '88. Lecture Notes in Computer Science, Springer Verlag, 1989.
6. W. Meier, O. Staffelbach. *Nonlinearity criteria for cryptographic functions.* Proceedings of Eurocrypt '89, Springer Verlag 1990, 549-562.
7. K. Nyberg. *Perfect nonlinear S-boxes.* Advances in Cryptology - Proceedings of Eurocrypt '91. Lecture Notes in Computer Science 547, Springer Verlag.
8. K. Nyberg. *On the construction of highly nonlinear permutations.* Advances in Cryptology - Proceedings of Eurocrypt '92 (to appear).
9. J. Pieprzyk. *On bent permutations.* Technical Report CS91/11; The University of New South Wales.