Feistel Schemes and Bi-Linear Cryptanalysis

(extended abstract)

Nicolas T. Courtois

Axalto Smart Cards Crypto Research,
36-38 rue de la Princesse, BP 45, F-78430 Louveciennes Cedex, France,
courtois@minrank.org

Abstract. In this paper we introduce the method of bi-linear cryptanalysis (BLC), designed specifically to attack Feistel ciphers. It allows to construct periodic biased characteristics that combine for an arbitrary number of rounds. In particular, we present a practical attack on DES based on a 1-round invariant, the fastest known based on such invariant, and about as fast as the best Matsui's attack. For ciphers similar to DES, based on small S-boxes, we claim that BLC is very closely related to LC, and we do not expect to find a bi-linear attack much faster than by LC. Nevertheless we have found bi-linear characteristics that are strictly better than the best Matsui's result for 3, 7, 11 and more rounds.
For more general Feistel schemes there is no reason whatsoever for BLC to remain only a small improvement over LC. We present a construction of a family of practical ciphers based on a big Rijndael-type S-box that are strongly resistant against linear cryptanalysis (LC) but can be easily broken by BLC, even with 16 or more rounds.

Key Words: Block ciphers, Feistel schemes, S-box design, inverse-based S-box, DES, linear cryptanalysis, generalised linear cryptanalysis, I/O sums, correlation attacks on block ciphers, multivariate quadratic equations.

1 Introduction

In spite of growing importance of AES, Feistel schemes and DES remain widely used in practice, especially in financial/banking sector. The linear cryptanalysis (LC), due to Gilbert and Matsui is the best known plaintext attack on DES, see [4, 25, 27, 16, 21]. (For chosen plaintext attacks, see [21, 2]).

A straightforward way of extending linear attacks is to consider nonlinear multivariate equations. Exact multivariate equations can give a tiny improvement to the last round of a linear attack, as shown at Crypto'98 [18]. A more powerful idea is to use probabilistic multivariate equations, for every round, and replace Matsui's biased linear I/O sums by nonlinear I/O sums as proposed by Harpes, Kramer, and Massey at Eurocrypt'95 [9]. This is known as Generalized Linear Cryptanalysis (GLC). In [10, 11] Harpes introduces partitioning cryptanalysis (PC) and shows that it generalizes both LC and GLC. The correlation cryptanalysis (CC) introduced in Jakobsen's master thesis [13] is claimed even more general. Moreover, in [12] it is shown that all these attacks, including also Differential Cryptanalysis are closely related and can be studied in terms of the Fast Fourier Transform for the cipher round function. Unfortunately, computing this transform is in general infeasible for a real-life cipher and up till now, nonlinear multivariate I/O sums played a marginal role in attacking real ciphers. Accordingly, these attacks may be excessively general and there is probably no substitute to finding and studying in details interesting special cases.

At Eurocrypt'96 Knudsen and Robshaw consider applying GLC to Feistel schemes [20], and affirm that in this case non-linear characteristics cannot be joined together. We will demonstrate that GLC can be applied to Feistel ciphers, which is made possible with our "Bi-Linear Cryptanalysis" (BLC) attack.

2 Feistel Schemes and Bi-Linear Functions

Differential [2] and linear attacks on DES [25, 1] have periodic patterns with invariant equations for some 1, 3 or 8 rounds. In this paper we will present several new practical attacks with periodic structure for DES, including new 1-round invariants.

2.1 The Principle of the Bi-Linear Attack on Feistel Schemes

In one round of a Feistel scheme, one half is unchanged, and one half is linearly combined with the output of the component connected to the other half. This will allow bi-linear I/O expressions on the round function to be combined together. First we will give an example with one product, and extend it to arbitrary bi-linear expressions. Then in Section 3 we explain the full method in details (with linear parts present too) for an arbitrary Feistel schemes. Later we will apply it to get concrete working attacks for DES and other ciphers.

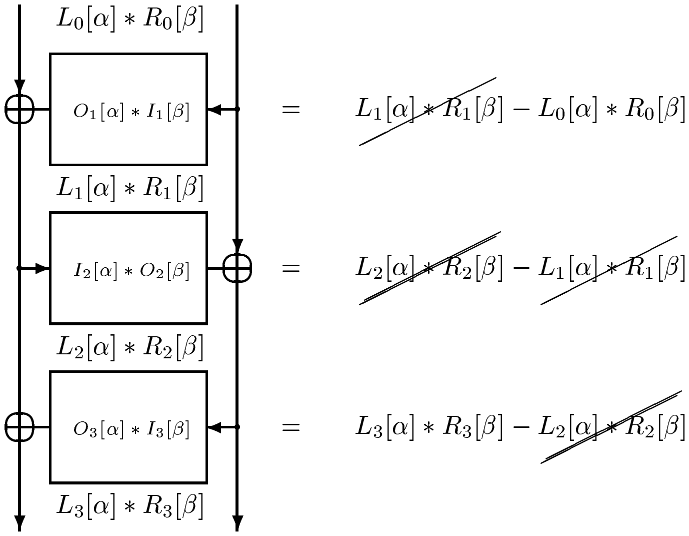

Fig. 1. Fundamental remark: combining bi-linear expressions in a Feistel cipher

In this paper we represent Feistel schemes in a completely "untwisted" way, allowing to see more clearly the part that is not changed in one round. As a consequence, the orientation changes compared to most of the papers and we obtain an apparent (but extremely useful) distinction between odd and even rounds of a Feistel scheme. Otherwise, our notations are very similar to these used for DES in [23, 18]. For example $L_0[\alpha]$ denotes a sum (XOR) of some subset α of bits of the left half of the plaintext. Combinations of inputs (or outputs) of round function number $r = 1, 2, \ldots$ are denoted by $I_r[\alpha]$ (or $O_r[\beta]$). Our exact notations for DES will be explained in more details when needed, in Section 6.1. For the time being, we start with a simple rather self-explaining example (cf. Figure 1) that works for any Feistel cipher.

Proposition 2.1.1 (Combining bi-linear expressions in a Feistel cipher). For all (even unbalanced) Feistel ciphers operating on $n + n'$ bits with arbitrary round functions we have: $\forall \alpha \subset \{1, \ldots, n\}, \forall \beta \subset \{1, \ldots, n'\},\ \forall r \geq 0$:

$$L_r[\alpha]R_r[\beta] \oplus L_0[\alpha]R_0[\beta] = \sum_{i=1}^{\lceil r/2 \rceil} O_{2i-1}[\alpha]I_{2i-1}[\beta] \ \oplus \sum_{i=1}^{\lfloor r/2 \rfloor} I_{2i}[\alpha]O_{2i}[\beta] \qquad \square$$

From one product this fundamental result extends immediately, by linearity, to arbitrary bi-linear expressions. Moreover, we will see that these bi-linear expressions do not necessarily have to be the same in every round, and that they can be freely combined with linear expressions (BLC contains LC).

3 Bi-linear Characteristics

For simplicity let $n = n'$. In this section we construct a completely general bi-linear characteristic for one round of a Feistel cipher. Then we show how it combines for the next round. Here we study bits locally and denote them by A_i, B_j etc. Later for constructing attacks for many rounds of practical Feistel ciphers we will use (again) the notations $L_i[j_1, \ldots, j_k]$ (cf. Section 6.1).

3.1 Constructing a Bi-linear Characteristic for One Round

Let $\mathcal{S}$ be a homogeneous bi-linear Boolean function $GF(2^n) \times GF(2^n) \to GF(2)$. Let $\mathcal{S}(A_1, \ldots, A_n; B_1, \ldots, B_n) = \sum s_{ij} A_i B_j$.

Let f_K be the round function of a Feistel cipher. We assume that there exist two linear combinations u and v such that the function:

$$(B_1, \ldots, B_n) \mapsto \begin{cases} \sum s_{ij} O_i B_j \oplus \sum u_i O_i \oplus \sum v_i B_i \\ \text{with} \ \ (O_1, \ldots, O_n) = f_K(B_1, \ldots, B_n) \end{cases}$$

is biased and equal to 0 with some probability $p \neq 1/2$ with $p = p(K)$ depending in some way on the round key K.

We have $C_i = A_i \oplus O_i$. By bi-linearity (or from Proposition 2.1.1) the following holds:

$$\sum s_{ij} A_i B_j \oplus \sum s_{ij} O_i B_j = \sum s_{ij} C_i B_j$$

From this, for the first round, (could be also any odd-numbered round), we obtain the following characteristic:

$$\left.\begin{array}{l} \sum s_{ij} A_i B_j \oplus \sum u_i A_i \oplus \sum v_i B_i = \\ \sum s_{ij} C_i B_j \oplus \sum u_i C_i \end{array}\right\} \quad \text{with probability} \ \ p(K)$$

Finally, we note that, the part linear in the B_i can be arbitrarily split in two parts: $\sum v_i B_i = \sum v_i^{(1)} B_i \oplus \sum v_i^{(2)} B_i$ with $v_i = v_i^{(1)} \oplus v_i^{(2)}$ for all $i = 1, \ldots, n$.

All this is summarized on the following picture:

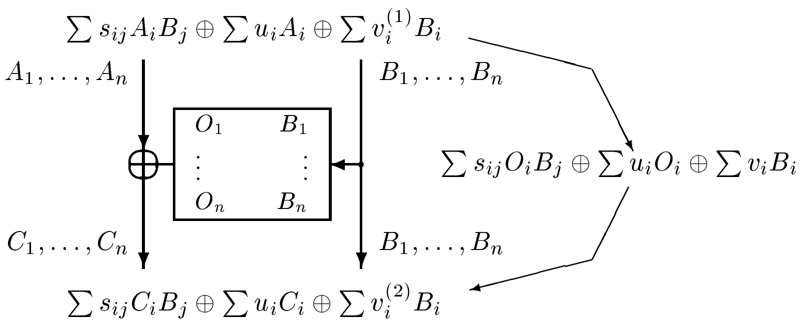

Fig. 2. Constructing a bi-linear characteristic for an odd round of a Feistel cipher

3.2 Application to the Next (Even) Round

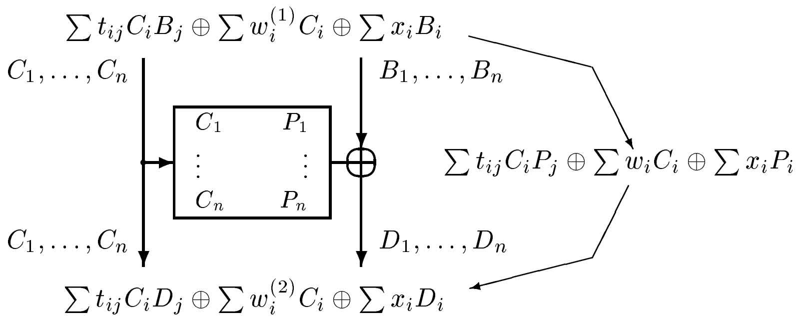

Fig. 3. Constructing a bi-linear characteristic for an even round of a Feistel cipher

The same method can be applied to the next, even, round of a Feistel scheme, with the only difference that the round function is connected in the inverse direction. In this case, to obtain a characteristic true with probability $\neq 1/2$, we need to have a bias in the function:

$$(C_1,\ldots,C_n) \mapsto \left\{ \begin{array}{c} \sum t_{ij}C_iP_j \oplus \sum w_iC_i \oplus \sum x_iP_i \\ \text{with} \ \ (P_1,\ldots,P_n) = f_K(C_1,\ldots,C_n) \end{array} \right.$$

3.3 Combining Approximations to Get a Bi-Linear Attack for an Arbitrary Number of Rounds

It is obvious that such I/O sums as specified above can be combined for an arbitrary number of rounds (contradicting [20] page 226). To combine the two characteristics specified above, we require the following three conditions:

1. We need $u = w^{(1)}$.
2. We need $v^{(2)} = x$.
3. We need the homogenous quadratic parts s et t to be correlated (seen as Boolean functions). They do **not** have to be the same (though in many cases they will). In linear cryptanalysis (LC), a correlation between two linear combinations means that these linear combinations have to be the same. In generalized linear cryptanalysis (GLC) [9], and in particular here, for bi-linear I/O sums, it is no longer true. Correlations between quadratic Boolean functions are frequent, and does not imply that $s = t$. For these reasons the number of possible bi-linear attacks is potentially very large.

Summary: We observe that bi-linear characteristics combine exactly as in LC for their linear parts, and that their quadratic parts should be either identical (with orientation that changes in every other round), or correlated.

4 Predicting the Behaviour of Bi-linear Attacks

The behaviour of LC is simple and the heuristic methods of Matsui [25] are known to be able to predict the behaviour of the attacks with good precision (see below). Some attacks work even better than predicted. As already suggested in [9, 20] the study of generalised linear cryptanalysis is **much harder**.

4.1 Computing the Bias of Combined Approximations

A bi-linear attack will use an I/O sum for the whole cipher, being a sum of I/O sums for each round of the cipher such that the terms in the internal variables do cancel. To compute the probability the resulting equation is true, is in general not obvious. Assuming that the I/O sum uses balanced Boolean functions, (otherwise it will be even harder to analyse) one can apply the Matsui's Piling-up Lemma from [25]. This however **can fail**. It is known from [9] that a sum of two very strongly biased characteristics can have a bias much weaker than expected. The resulting bias can even be exactly zero: an explicit example can be found in Section 6.1. of [9]. Such a problem can arise when the connecting characteristics are not independent. This will happen more frequently in BLC than in LC: two linear Boolean functions are perfectly independent unless equal, for non-linear Boolean functions, correlations are frequent. Accordingly, we do not sum independent random variables and the Matsui's lemma may fail.

At this stage there are two approaches: one can try to define a class of attacks that can be proved to work, and restrict oneself only to studying such attacks, or try to explore all possible attacks, including those that do work experimentally without proof. This first approach is adopted in [9]: the Lemma 6 gives a sufficient condition to guarantee that the Piling-up lemma will apply. For this the probability, that the characteristic is true, for a random partial key, should be independent of the input (e.g. the input of the whole round). This explains why Matsui's attacks indeed work well. In [9] it allows to prove that the proposed family of GLC attacks based on homomorphic properties will work as predicted. We will also use this argument in Section 5.

In this paper we frequently adopt rather the second approach: try find as many working attacks as possible, even if current theory does not allow to predict their behaviour with accuracy. A price to pay for this is that each application of Matsui's Lemma will be systematically questioned and confronted to experimental results.

4.2 Key Dependence in Bi-Linear Attacks

Another important property of bi-linear cryptanalysis is that the existence of a bias for one characteristic does frequently depend on the key. This does not really happen for LC applied DES, because in DES all key bits are combined linearly and a linear equation will be true with probability either p or $1-p$

depending on the key. However it will happen for LC and other ciphers, if key bits are involved in a more complex way, for example for ICE [22].

In bi-linear cryptanalysis, the behaviour becomes complex already when the key bits are combined linearly as in DES. Adding a constant (a key bit) to an input of an S-box, does not only modify the constant part in a bi-linear characteristic, but also the linear part. (We note that for DES only the linear part in the output variables will be modified when the key changes). From this, quite frequently two bi-linear characteristics for two parts of a cipher (e.g. for S-boxes) will only connect together for some keys. Such attacks are still very interesting and frequently also do work, with only a slightly weaker bias, for all the other keys. For simplicity, no key bits are displayed in bi-linear characteristics for one or several rounds of a cipher that are studied/displayed in this paper. The values of biases we will present (unless otherwise stated) are given for the reference key being zero. Yet typically we observed that they exist, and slightly vary in value, also for **any** other key (chosen at random). In rare cases, the bias works well only for a fraction of keys (e.g. 25 %): this happens in Appendix B.1.

4.3 Exploring Bi-linear Cryptanalysis

There are different approaches to finding interesting bi-linear attacks to block ciphers. In few cases one can construct attacks that will provably or arguably work (see [9] and later Section 5). Another method is to construct characteristics "by hand" around some particularly strong bias found for one S-box.

We noted the two major difficulties: predicting the bias of combined characteristics, and huge number of possible characteristics (including fragmentation due to the fact they the bias does in general depend on the key). These make it very difficult to have a systematic method (a computer program) that would compute the best bi-linear characteristic for a given cipher. To check if an attack indeed works requires to be able to generate as many plaintexts as for the real attack. To find the best attack is even much harder. It requires to exhaustively search and reject lots of other combinations that should work well but they don't. Each of them has to be tested on an equally large set of plaintexts.

5 The Killer Example for Bi-Linear Cryptanalysis

We will construct a practical cipher that is very secure w.r.t. all known attacks for block ciphers, in particular for LC, yet broken by BLC. It mixes two group operations: the XOR and the multiplication in $GF(2^n)$ e.g. $n = 32$ or 64. It uses the inverse in $GF(2^n)$ (cf. Rijndael): let $Inv(X) = X^{-1}$ in $GF(2^n)$ when $X \neq 0$ and 0 otherwise. We build a $2n$-bit Feistel cipher with the i-th round function being:

$$f_i(X) = Inv(X) \cdot (K_i \oplus G(X)) \quad \text{in } GF(2^n), \tag{1}$$

with K_i being the partial key, and G being some function with S-boxes and arbitrary components $\{0,1\}^n \to \{0,1\}^n$. In order to get an insecure cipher, we need to assume that some linear combination of outputs of G is biased. For example, let $Y_1 \oplus Y_5 = 0$ with probability 3/4. Building a cipher with G alone would be insecure for LC, however here G is composed by a group operation $\cdot$ with $Inv(X)$. The $Inv(X)$ assures global diffusion and very high non-linearity (cf. [3]). Accordingly our round function has very good resistance to linear and differential cryptanalysis for most G, even when $G = 0$. But not against BLC.

First, we can consider a bi-linear attack with bi-linear equations over $GF(2^n)$: $\forall r \geq 0$:

$$L_r \cdot R_r \oplus L_0 \cdot R_0 = \sum_{i=1}^{\lceil r/2 \rceil} O_{2i-1} \cdot I_{2i-1} \oplus \sum_{i=1}^{\lfloor r/2 \rfloor} I_{2i} \cdot O_{2i} = \sum_{i=1}^{r} I_i \cdot O_i \qquad (2)$$

Let $X \cdot Y = (Z_1, \ldots, Z_n)$ with $Z_k = \sum_{ij} M_k^{ij} X_i Y_j$. From (2), or if we prefer, directly from Proposition 2.1.1 and by symmetry $M_k^{ij} = M_k^{ji}$, we get:

$$\forall k \in \{1, \ldots, n\}, \forall r \geq 0 \quad \sum_{ij} M_k^{ij} \left(L_{ri} R_{rj} \oplus L_{0i} R_{0j} \right) = \sum_{l=1}^{r} \sum_{ij} M_k^{ij} I_{li} O_{lj} \qquad (3)$$

Now, $\forall l \geq 1$, $I_l \cdot O_l = K_l \oplus G(I_l)$ with probability $(1 - 1/2^n)$. We rewrite it:

$$\forall k \in \{1, \ldots, n\}, \forall l \geq 0 \quad \sum_{ij} M_k^{ij} I_{li} O_{lj} = K_{ik} \oplus G_k(I_i) \qquad (4)$$

Then we use the linear output bias of G: $G_1 \oplus G_5 = 0$ with probability $3/4$.

$$\forall l \geq 0 \quad \sum_{ij} M_1^{ij} I_{li} O_{lj} \oplus \sum_{ij} M_5^{ij} I_{li} O_{lj} = K_{i1} \oplus G_1(I_i) \oplus K_{i5} \oplus G_5(I_i) \approx C_l \qquad (5)$$

The last expression is equal to come constant denoted C_l with probability $3/4$. Finally, we combine with (3) (or equivalently sum these bi-linear expressions over the whole cipher with r rounds).

$$\sum_{ij} \left(M_1^{ij} \oplus M_5^{ij} \right) \left(L_{ri} R_{rj} \oplus L_{0i} R_{0j} \right) = \sum_{l=1}^{r} C_l \text{ with probability } \frac{1}{2} + \frac{1}{2^{r+1}} \qquad (6)$$

What we obtained is a biased bi-linear I/O sum for the whole cipher. We can distinguish this cipher from a random permutation given about 2^{2r+2} plaintexts. For example 16 rounds will be broken on a laptop PC.

Does it work as predicted ? In general, as we explain in Section 4.1, it is hard to predict accurately the behaviour of a composed bi-linear attack. However we have little doubt it will work: the $Inv(X)$ should render possible correlation between approximations being combined negligible. In some case we can even prove that this attack works: when $G = 0$, and also when one fixed linear combination of output bits of G is 0, (the other parts can be arbitrary functions). In these cases, dependencies cannot be a problem: we add equations (5) true with probability 1 to get the equation (6) true with probability 1.

Related work: Similar results were previously obtained for some substitution-permutation network (SPN) ciphers. In [9] Harpes, Kramer and Massey give an example of 8-bit SPN that is secure against LC and DC, but insecure for generalised linear cryptanalysis due to a probabilistic homomorphic property of each round relative to quadratic residuosity function modulo 2^8+1. The Jakobsen attack for substitution ciphers that uses probabilistic univariate polynomials from [15] can also be seen as a special case of GLC. However, it is the first time that GLC allows to break a Feistel cipher, which contradicts the impossibility professed by Knudsen and Robshaw [20]. This cipher is built with state-of-art components (inverse in $GF(2^n)$) and can in addition incorporate any additional fashionable component with lots of theory and designer tricks, as a part of G.

Due to G it will not have homomorphic properties. Moreover, by adjusting the bias in G, the security of this cipher against BLC will be freely adjusted between (nearly) zero and infinity. It can therefore be arbitrarily weak for BLC, and this even for a very large number of rounds. Yet, the security against the usual attacks (LC, DC) should remain equally good (due to the big Inv S-box).

6 Bi-Linear Attacks on DES

6.1 Notation

We ignore the initial and final permutations of DES that have no incidence on the attacks. We use the "untwisted method" of representing DES, as on the right-hand figure, page 254 in [28]. The bit numbering is compatible with the FIPS standard [8], and [23, 18], and differs from Biham, Shamir [2] or Matsui [25, 27]. We denote the bits of the left hand side of the plaintext by $L_0[1] \ldots L_0[n]$. The bits of the right hand side are $R_0[1] \ldots R_0[n]$. Similarly, as in other papers, the plaintext after i rounds will be L_i, R_i, except that we felt it necessary to have our notations completely "untwisted" which implies that our L_i and R_i for an odd $i = 1, 3, \ldots$ will be inversed compared to [23, 18, 28], Then, we apply the popular convention $X[i_1, \ldots, i_n]$being $X[i_1] \oplus \ldots \oplus X[i_n]$. For example $L_0[9, 7, 23, 31]$ is the XOR of 4 bits of the left half of the plaintext that are added to the outputs of S1 in the first round. We denote the input bits to the $i-$th round function by $I_i[1], \ldots, I_i[32]$. Similarly the output bits will be $O_i[1], \ldots, O_i[1]$.
For odd i we have $I_i[j] = R_{i-1}[j] = R_i[j]$ and $O_i[j] = L_{i-1}[j] \oplus L_i[j]$.
For even i we have $I_i[j] = L_{i-1}[j] = L_i[j]$ and $O_i[j] = R_{i-1}[j] \oplus R_i[j]$.

For individual S-boxes, we will denote the inputs/outputs by respectively $O[i]$ and $J[j]$ with i, j being directly the numbers 1..32 in the round function of DES. For example $O[8], O[14], O[25], O[3]$ are the outputs of S-box S5, and $J[16], \ldots, J[21]$ are the inputs of this S-box S5. Depending on the key in round i, we have $I_i[k] = J_i[k]$ or $I_i[k] = J_i[k] + 1$. For better readability, we will avoid naming precisely the key bits involved.

6.2 First Example of Bi-Linear Cryptanalysis of DES

Our simulations on DES S-boxes (cf. Appendix A) show that the following two bi-linear characteristics exist for DES S-boxes S1 and S5:

$$O[8, 14, 25, 3] \oplus J[17] \cdot O[3] = 0 \quad \text{for } S5 \text{ with probability } 17/64$$

$$O[17] \oplus J[3] \cdot O[17] = 0 \quad \text{for } S1 \text{ with probability } 47/64$$

From these, acting as if all the key bits were zero ($I_i[k] = J_i[k]$), we deduce the following bi-linear characteristic for two rounds:

$$(*) \left.\begin{array}{l} L_0[3, 8, 14, 25] \oplus L_0[3]R_0[17] \oplus R_0[17] \oplus \\ L_2[3, 8, 14, 25] \oplus L_2[3]R_2[17] \oplus R_2[17] = K[sth] \end{array}\right\} \quad \frac{1}{2} - 1.76 \cdot 2^{-4}$$

The explanation is given on the following picture:

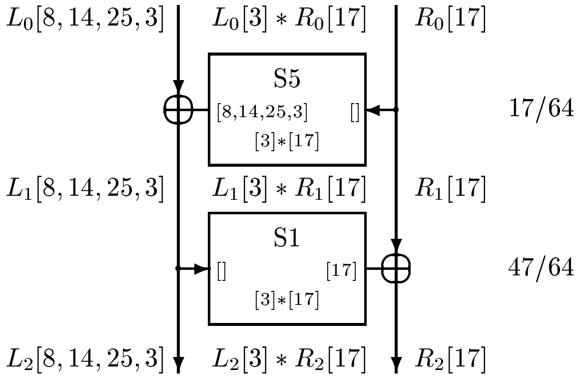

Fig. 4. Our first example - an invariant bi-linear attack on DES $(*)$

We verified this bias experimentally, and the probability is (we were lucky) equal to the probability that is predicted by Matsui's Piling-Up Lemma.

Key Dependence: Very surprisingly, the above equation $(*)$ is biased, not only when all key bits are 0, but for every DES key. This can be seen to come from a couple of other (different) bi-linear characteristics from Appendix A.

More rounds: It is easy to see from the picture, and we verified it experimentally, that $(*)$ is also biased for $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \ldots$ rounds of DES, and all this happens to work about equally well for an arbitrary key.

Relation to LC: The bias of $(*)$ is closely related to some prominent equations of Matsui, see the extended version of this paper.

6.3 Invariant Attacks on DES

The equation $(*)$ is an invariant equation, i.e. the input and the output bi-linear expressions are the same. We have found a simple invariant bi-linear I/O sum for DES that is biased for any key and for any number of rounds. For LC and DES, such simple invariant characteristics do exist, have been found by Biham (page 347 in [1]) in close relation to Davies-Murphy attack. The example $(*)$ above is one of the best we found for DES, and so far it also **the only known** non-linear 1-round invariant attack on DES that works really well in practice. Our invariant on DES is stronger than Biham's. We recall that Biham uses a bias on a sum of some outputs for two successive DES S-boxes. The best bias obtained by Biham (also exhibited by Matsui in [26] and contained unnoticed in the earlier Davies-Murphy attack [6, 7]) is equal to $(35/64 - 1/2)$ for 2 rounds and for S-boxes S7-S8. This gives $1.4 \cdot 2^{-22}$ for 12 rounds. Instead, $(*)$ gives experimentally only about $1.3 \cdot 2^{-18}$. Accordingly, $(*)$ is **the strongest known 1-round invariant attack on DES**.

To break full DES requires a bias for 14 rounds (Matsui's 2R method) and the Biham's invariant requires then 2^{50} plaintexts. Our invariant attack requires about 2^{43} plaintexts (the bias of $(*)$ for 14 rounds is expected to be about 2^{-22}, we did not dispose of a sufficient computing power to compute it exactly).

6.4 How Good is Our First Example, BLC vs. LC

These new properties of DES give a chosen-plaintext attack on an arbitrary number of rounds of DES, somewhat simpler than Matsui's laborious search for the best linear characteristic. If we try here to predict the resulting bias for 14 rounds by applying the Matsui's Piling-up formula, we would get for 14 rounds the bias of: $1.63 \cdot 2^{-17}$ which means an attack on full DES with only $2^{32.6}$ known plaintexts (!?). Unfortunately, unlike for LC in DES, such predictions are frequently not valid for BLC. Starting from 3 rounds, the bias of our invariant does not follow the prediction at all, yet remains significative. For example if we apply Matsui's Piling-Up Lemma to predict the bias for 4 rounds as 2+2 rounds, we obtain $1.55 \cdot 2^{-6}$, while in practice it is about $1.80 \cdot 2^{-8}$. Our invariant attack seems very bad for 4 rounds, and unfortunately with $(*)$ we never get a bias better than obtained by Matsui. Yet, it is the best invariant attack on DES known, and for more than 4 rounds the results are again not so bad. Only slightly worse than Matsui. For example for 12 rounds the best result of Matsui from [25] gives $1.19 \cdot 2^{-17}$, while for $(*)$ and a random key our simulation gives $1.3 \cdot 2^{-18}$, To break full DES Matsui requires about 2^{43} plaintexts, and with $(*)$ we also need about 2^{43} (and both are related). In the full version of this paper we give a heuristic argumentation why for DES (but not in general !) the complexity of the best bi-linear attack should be roughly the same than for LC.

For DES and 1-round invariants attacks extended to an arbitrary number of rounds, BLC gives strictly better results than LC. It is also so for more complex periodic constructions and we are going to see that BLC attacks can also be strictly better than any existing linear attack.

6.5 Second Example of Bi-Linear Cryptanalysis of DES

In order to exhibit biases really better than Matsui we looked what is the best bi-linear characteristic that exists in DES:

$$J[16,20] \oplus O[8,14,25,3] \oplus J[16,17,20] \cdot O[3] = 0 \quad \text{for } S5 \text{ with probability } 61/64.$$

We note that this equation can be seen as "causing" the existence of the Matsui's best equation (A) for S5: their difference is highly biased. Based mainly on this, we constructed a periodic characteristic for 3,7, 11 and more rounds that is strictly better than the best results of Matsui for the same number of rounds.

Proposition 6.5.1 (Our Best Attack on 11 Rounds of DES). For all keys, the following equation is biased for 11 rounds of DES:

$$(**)\quad \left.\begin{array}{l} L_0[3,8,14,25] \oplus L_0[3]R_0[16,17,20] \oplus R_0[17] \oplus \\ L_{11}[3,8,14,25] \oplus L_{11}[3]R_{11}[16,17,20] \oplus R_{11}[17] = \\ K[sth] + K[sth']L_0[3] + K[sth'']L_{11}[3] \end{array}\right\} \quad \frac{1}{2} \pm \text{around } 1.2 \cdot 2^{-15}$$

The exact construction to achieve this is a bit complicated. (cf. Appendix B). The bias of this equation is strictly better than the best linear characteristic for 11 rounds obtained by Matsui (which gives $1.91 \cdot 2^{-16}$ for 11 rounds). It has been verified by computer simulations at every stage. We note also that both are closely related: their difference, is a biased Boolean function.

Our second example allows us to give an attack strictly better than Matsui for 11+2=13 rounds of DES. For the full 16-round DES our results are roughly as good as Matsui (but we hope to improve this soon too). For 17 rounds of DES, as the construction of our second example (**) is periodic, we expect that for 11+4=15 rounds it should also be better than the best bias of Matsui, which would allow to break 15+2=17 rounds of DES faster than by LC. We do not dispose of a sufficient computing power to fully confirm this fact.

7 Conclusion

It was stated that for Feistel ciphers non-linear characteristics cannot be joined together for several rounds, see [20]. In this paper we show that generalised linear cryptanalysis (GLC) is in fact possible for Feistel schemes. To achieve this goal, we introduced bi-linear cryptanalysis (BLC). It gives a new (and the fastest known) 1-round invariant attack on DES. Though more powerful, generalized linear cryptanalysis is unfortunately much harder to study than LC. At present heuristic constructions, to be confirmed (or not) by computer simulations are the only method known to explore it. BLC is related to LC in multiple important ways. It contains LC as a sub-set. LC can be used to construct good bi-linear characteristics and vice-versa. BLC also contains LC as an extension: a combination of biased bi-linear characteristics may extend a concrete combination of biased linear characteristics by adding quadratic polynomials. Yet BLC can be strictly better than any (existing) linear attack. This was demonstrated for 3, 7, 11 and more rounds of DES, and also for s^5DES.

In this paper we only initiate the study of bi-linear cryptanalysis. BLC and GLC extend the role of LC as an essential tool to evaluate the real-life security of many practical ciphers. An interesting contribution of this paper is to point out that, though GLC is excessively general to be systematically explored, the properties of the top-level structure of a cryptographic scheme (e.g. being a Feistel scheme) will determine the type of the attacks (e.g. BLC) that may indeed work. Our new attack can be quite devastating: we constructed a large family of practical ciphers based on big Rijndael-type S-box, that are strongly resistant against LC and all previously known attacks on Feistel ciphers, yet can be broken in practice with BLC for an important number of rounds. Fortunately, for DES, BLC gave only slight improvements over LC and does not cause excessive trouble.

References

1. Eli Biham: *On Matsui's Linear Cryptanalysis,* Eurocrypt'94, LNCS 950, Springer-Verlag pp. 341-355, 1994.
2. Eli Biham and Adi Shamir, *Differential Cryptanalysis of DES-like Cryptosystems.* Journal of Cryptology, vol. 4, pp. 3-72, IACR, 1991.
3. Anne Canteaut, Marion Videau: *Degree of composition of highly nonlinear functions and applications to higher order differential cryptanalysis,* Eurocrypt 2002, LNCS 2332, Springer, 2002.
4. Anne Tardy-Corfdir, Henri Gilbert: *A Known Plaintext Attack of FEAL-4 and FEAL-6,* Crypto'91, LNCS 576, Springer, pp. 172-181, 1992.
5. Nicolas Courtois, Guilhem Castagnos and Louis Goubin: *What do DES S-boxes Say to Each Other ?* Available on `eprint.iacr.org/2003/184/`.

6. D.W. Davies, *Some Regular Properties of the Data Encryption Standard,* Crypto'82, pp. 89-96, Plenum Press, New-York, 1982.
7. D. Davies and S. Murphy, *Pairs and Triplets of DES S-Boxes,* Journal of Cryptology, vol. 8, Nb. 1, pp. 1-25, 1995.
8. *Data Encryption Standard (DES)*, Federal Information Processing Standards Publication (FIPS PUB) 46-3, National Bureau of Standards, Gaithersburg, MD (1999). `http://csrc.nist.gov/publications/fips/fips46-3/fips46-3.pdf`
9. C. Harpes, G. Kramer, and J. Massey: *A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-up Lemma,* Eurocrypt'95, LNCS 921, Springer, pp. 24-38. `http://www.isi.ee.ethz.ch/ harpes/GLClong.ps`
10. Carlo Harpes: *Cryptanalysis of iterated block ciphers,* PhD thesis, No 11625, Swiss Federal Int. of Tech., ETH Series in Information Processing, Ed. J. L. Massey, Hartung-Gorre Verlag Konstanz, 1996, ISBN 3-89649-079-6, ISSN 0942-3044.
11. Carlo Harpes: *Partitioning Cryptanalysis,* Post-Diploma Thesis, Signal and Information Processing Lab., Swiss Federal Institute of Technology, Zurich, March 1995. `http://www.isi.ee.ethz.ch/`∼harpes/pc.ps
12. Thomas Jakobsen, Carlo Harpes: *Non-Uniformity Measures for Generalized Linear Cryptanalysis and Partitioning Cryptanalysis,* Pragocrypt'96, 1996.
13. Thomas Jakobsen: *Correlation Attacks on Block Ciphers,* Master's Thesis, Dept. of Mathematics, Technical University of Denmark, January 1996.
14. Thomas Jakobsen: *Higher-Order Cryptanalysis of Block Ciphers.* Ph.D. thesis, Dept. of Math., Technical University of Denmark, 1999.
15. Thomas Jakobsen: *Cryptanalysis of Block Ciphers with Probabilistic Non-Linear Relations of Low Degree,* Crypto 98, LNCS 1462, Springer, pp. 212-222, 1998.
16. Pascal Junod: *On the complexity of Matsui's attack,* Selected Areas in Cryptography (SAC'01), Toronto, Canada, LNCS 2259, pp. 199-211, Springer, 2001.
17. Burton S. Kaliski Jr, and M.J.B. Robshaw. *Linear Cryptanalysis Using Multiple Approximations,* Crypto'94, LNCS, Springer, pp. 26-39, 1994.
18. Toshinobu Kaneko and Takeshi Shimoyama: *Quadratic Relation of S-box and Its Application to the Linear Attack of Full Round DES,* In Crypto 98, LNCS 1462, p. 200-211, SPringer, 1998.
19. Kwangjo Kim. Sangjin Lee, Sangjoon Park, Daiki Lee: *Securing DES S-boxes against Three Robust Cryptanalysis,* SAC'95, pp.145-157, 1995.
20. Lars R. Knudsen, Matthew J. B. Robshaw: *Non-Linear Characteristics in Linear Cryptoanalysis. Eurocrypt'96, LNCS 1070, Springer, pp. 224-236, 1996.*
21. *Lars R. Knudsen, John Erik Mathiassen: A Chosen-Plaintext Linear Attack on DES.* FSE'2000, LNCS 1978, Springer, pp. 262-272, 2001.
22. Matthew Kwan: *The Design of the ICE Encryption Algorithm,* FSE'97, 4th International Workshop, Haifa, Israel, Springer, LNCS 1267, pp. 69-82, 1997. Available from `http://www.darkside.com.au/ice/ice.ps.gz`.
23. Susan K. Langford, Martin E. Hellman: *Differential-linear cryptanalysis,* Crypto 94, LNCS 839, pp. 17-25, Springer, 1994.
24. Michael Luby, Charles W. Rackoff, *How to construct pseudorandom permutations from pseudorandom functions,* SIAM Journal on Computing, vol. 17, n. 2, pp. 373-386, April 1988.
25. M. Matsui: *Linear Cryptanalysis Method for DES Cipher,* Eurocrypt'93, LNCS 765, Springer, pp. 386-397, 1993.
26. M. Matsui, *On correlation between the order of S-boxes and the strength of DES,* Eurocrypt'94, LNCS 950, pp. 366-375, Springer, 1995.
27. M.Matsui: *The First Experimental Cryptanalysis of the Data Encryption Standard,* Crypto'94, LNCS 839, Springer, pp. 1-11, 1994.

28. Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone: *Handbook of Applied Cryptography*; CRC Press, 1996.
29. J. Patarin, *How to construct pseudorandom and super pseudorandom permutations from one single pseudorandom function. Eurocrypt'92, Springer, pp. 256-266, 1992.*
30. *Adi Shamir: On the security of DES,* Crypto'85, LNCS 218, Springer, pp. 280-281, 1985.

A Selected Bi-Linear Characteristics of DES S-boxes

In this section we give some bi-linear characteristics for DES S-boxes. Our results are not exhaustive: the number of possible bi-linear characteristics is huge and we do not have a fast method to find all interesting characteristics. Accordingly we are not certain to have found the best existing characteristics. It is certain that there is no characteristics true with probability 1, as these are easy to check algebraically. Otherwise we explored all cases that use up to two products and we conjecture that the other does not have practical relevance for the security of DES. We give here some interesting results we have found. More will appear in the extended version of this paper.

Table 1. A few selected bi-linear characteristics for DES S-boxes

		equation			remarks and comments
		input	output	input*output	
$S5$	$12/64$	17	$8,14,25,3$		Matsui's equation A
$S5$	$6/64$	17	$8,14,25,3$	$[17]*[8,14,25,3]$	gets better
$S5$	$58/64$			$[17]*[8,14,25,3]$	
$S5$	$8/64$	17	$8,14,25,3$	$[16,17,20]*[8]$	
$S5$	$8/64$	$16,20$	$8,14,25$	$[16,20]*[8,14,25]$	
$S5$	$\mathbf{61/64}$	$16,20$	$8,14,25,3$	$[16,17,20]*[3]$	the best in DES
$S5$	$47/64$		$8,14,25$	$17*3$	
$S5$	$17/64$		$8,14,25,3$	$17*3$	
$S5$	$47/64$			$17*3$	
$S5$	$49/64$		3	$17*3$	
$S5$	$49/64$	17		$17*3$	
$S5$	$17/64$	17	3	$17*3$	
$S1$	$30/64$	3	17		Matsui's equation C
$S1$	$15/64$	3	17	$3*17$	gets better
$S1$	$47/64$		17	$3*17$	
$S1$	$47/64$	3		$3*17$	
$S1$	$49/64$			$3*17$	
$S2$	$8/64$	5	$13,28,18$	$8*2$	
$S4$	$56/64$			$[12,14,16,17]*[26,1]$	(there are many similar)
$S6$	$38/64$		$11,19$	$21*29$	
$S7$	$11/64$	$25,28$	$32,12,7$	$28*12, 27*22$	
$S8$	$40/64$		$5,27,15$	$29*21$	

B Improved Bi-Linear Attacks for DES

The goal of this section is to find or construct examples where bi-linear cryptanalysis gives strictly better bias on DES than the best Matsui's result.

We look at the best Matsui's characteristic on 3 rounds given at the last page of [25]. By itself, it can be considered as very good, even compared to other Matsui's characteristics: it uses twice the best element (A) of Matsui, and nothing between them. Moreover, this element (A) is in itself the best linear characteristic that exist in DES, first described by Shamir in [30]:

$$\textbf{(A)} \quad J[17] \oplus O[8,14,25,3] = 0 \qquad \text{for } S5 \text{ with probability } \; 12/64$$

From this we get immediately, using Matsui's Piling-Up Lemma from [25], that for 3 rounds, and for any key, the following equation is biased:

$$\left.\begin{array}{l} L_0[8,14,25,3] \oplus R_0[17] \oplus \\ L_3[8,14,25,3] \oplus R_3[17] = K[sth] \end{array}\right\} \quad \frac{1}{2} - 1.56 \cdot 2^{-3}$$

We call Matsui-3 this equation.

B.1 Improving on Matsui-3

We will show that with bi-linear characteristics, there are strictly better equations than Matsui-3. Our simulations looking for the best bi-linear characteristics for DES S-boxes (cf. Appendix A), showed that the best one is the following:

$$J[16,20] \oplus O[8,14,25,3] \oplus J[16,17,20] \cdot O[3] = 0 \quad \text{for } S5 \text{ with probability } \; 61/64$$

Remark: It is clearly related to, and can be seen as "causing" the existence of the Matsui's equation (A): their difference is naturally biased.

We will use this characteristic. Let KS5 denote the combination of the S-box S5 and the key bits XORed to its inputs. It is easy to see that for KS5, if we denote by $K[sth]$ some constant linear combination of key bits, for any key, one of the following equations is always strongly biased:

$$\left\{\begin{array}{l} \textbf{(a1)} \;\; I[16,20] \oplus O[8,14,25,3] \oplus I[16,17,20] \cdot O[3] = K[sth] \\ \qquad\qquad\qquad\qquad\qquad \text{or} \\ \textbf{(a2)} \;\; I[16,20] \oplus O[8,14,25] \oplus I[16,17,20] \cdot O[3] = K[sth] \end{array}\right. \qquad |\text{bias}| = 1/2 - 3/64$$

In our construction, we will use one of the above, and we will also use another, naturally biased equation, which will be one of the following:

$$\left\{\begin{array}{l} \textbf{(b)} \;\; O[16,17,20] \oplus I[3] \cdot O[16,17,20] = 0 \\ \qquad\qquad\qquad\qquad\qquad \text{and} \\ \textbf{(c)} \;\; I[3] \oplus O[16,17,20] \oplus I[3] \cdot O[16,17,20] \cdot O[3] = 0 \end{array}\right. \qquad |\text{bias}| = 1/2 - 1/4$$

Now we are ready to construct characteristics for 3 rounds of DES.

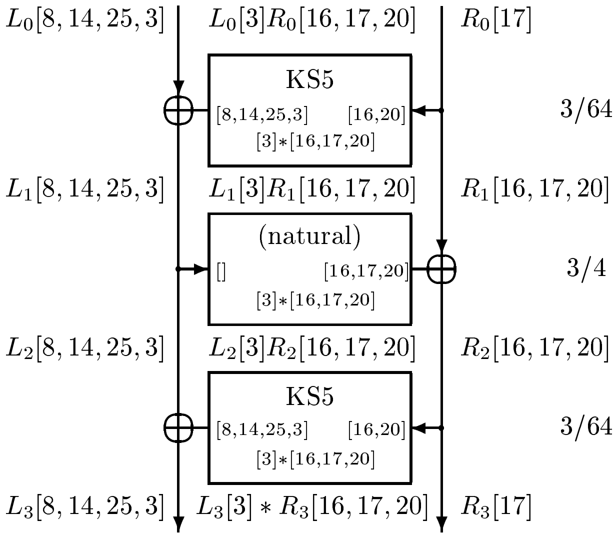

Fig. 5. Combining a1-b-a1 to get a characteristic for 3 rounds of DES

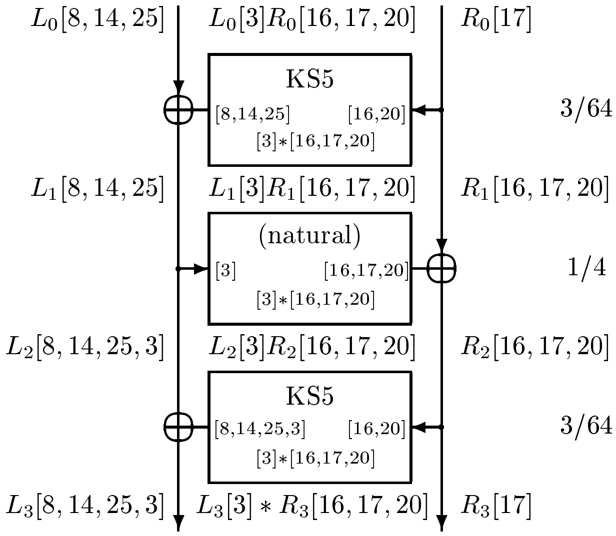

Fig. 6. Combining a2-c-a1 to get a characteristic for 3 rounds of DES

As one should expect, our construction goes as follows:

- ⋄ In round 1 and 3, depending on the key either a1 or a2 is strongly biased.
- ⋄ To connect a1 to a1, or a2 with a2, we can use b, as in Figure 5.
- ⋄ To connect a1 with a2 and the reverse, we use c, as in Figure 6.
- ⋄ For 3 rounds and for any key, we always have a strong bias on one of the four possibilities: a1-b-a1, a1-c-a2, a2-c-a1, a2-b-a2.
- ⋄ From Matsui's Piling-Up Lemma, we expect that the whole characteristic will be true with probability $\frac{1}{2} \pm 1.64 \cdot 2^{-3}$. Our simulations show that it is between $\frac{1}{2} \pm 1.65 \cdot 2^{-3}$ and $\frac{1}{2} \pm 1.67 \cdot 2^{-3}$.
- ⋄ Since, the choice of a1/a2 depends on a linear combination of key bits, We can combine all these into one equation and we get the following result:

Proposition B.1.1 (Our Best Attack on 3 Rounds of DES). For all keys, the following equation is biased for 3 rounds of DES: :

$$(**) \quad \left.\begin{array}{l} L_0[3,8,14,25] \oplus L_0[3]R_0[16,17,20] \oplus R_0[17] \oplus \\ L_3[3,8,14,25] \oplus L_3[3]R_3[16,17,20] \oplus R_3[17] = \\ K[sth] + K[sth']L_0[3] + K[sth'']R_3[3] \end{array}\right\} \quad \frac{1}{2} \pm 1.66 \cdot 2^{-3}$$

In comparison, Matsui-3 gives $\frac{1}{2} - 1.56 \cdot 2^{-3}$. Bi-linear cryptanalysis works better than LC. In the next section we will extend this result (and again beat Matsui) to 7, 11 and more rounds.

Remark: The equation above can be seen as 4 different equations, each of them is highly biased for 1/4 of all keys. We observed that each of the 4 equations is also biased for all DES keys, except that for 3/4 of them the bias is much weaker, we get about $\frac{1}{2} \pm 1.6 \cdot 2^{-7}$.

B.2 Extending the Result for 7, 11 and More Rounds

The idea is to find an element (maybe not very good in itself) that will allow to connect together our (very good) characteristics on 3 rounds. For example, to connect Figure 5 with Figure 6 we use the following element:

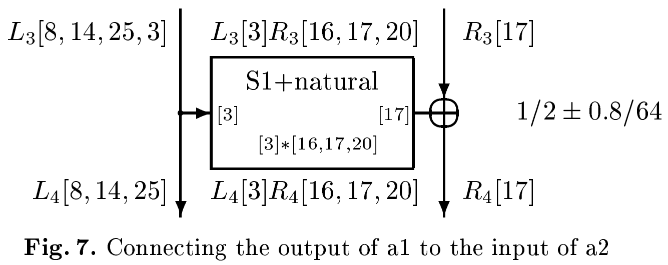

Fig. 7. Connecting the output of a1 to the input of a2

Simulations show that, for any key, this characteristic is true with probability about $1/2 \pm 0.8/64$. The explanation is as follows: the bias is due to to the combination of Matsui's equation (C)

$$\textbf{(C)} \quad J[3] \oplus O[17] = 0 \quad \text{for } S1 \text{ with probability } 30/64$$

and of the fact that $I[3] \cdot O[16,17,20]$ is naturally biased. The same element (Figure 7) does also work to connect a2 to a1.

It remains to be seen how the connection between a1 and a1 or a2 and a2. This is done in a very similar way: we combine (C) with $I[3] \oplus I[3] \cdot O[16, 17, 20]$ that is also naturally biased.

Summary: In every of 4 possible cases, there is a connecting element based on (C). This means that, also for 7 rounds and for any key, again one of the four possibilities is quite biased: a1-b-a1, a1-c-a2, a2-c-a1, a2-b-a2. Again we can recompose it in a single attack:

Proposition B.2.1 (Extension to 7 Rounds of DES). For all keys, the following equation is biased for 7 rounds of DES:

$$\left.\begin{array}{l} L_0[3,8,14,25] \oplus L_0[3]R_0[16,17,20] \oplus R_0[17] \oplus \\ L_7[3,8,14,25] \oplus L_7[3]R_7[16,17,20] \oplus R_3[17] = \\ K[sth] + K[sth']L_0[3] + K[sth'']L_7[3] \end{array}\right\} \quad \frac{1}{2} \pm \text{about } 2^{-9}$$

This bias is, depending on the key, sometimes better, sometimes worse than Matsui-7 that gives $\frac{1}{2} - 1.95 \cdot 2^{-10}$.

Finally, it is now obvious, that our construction works also for 11, 15, 19 rounds etc. We verified experimentally that for 11 rounds we have:

Proposition B.2.2 (Our Best Attack on 11 Rounds of DES). For all keys, the following equation is biased for 11 rounds of DES: :

$$\left.\begin{array}{l} L_0[3,8,14,25] \oplus L_0[3]R_0[16,17,20] \oplus R_0[17] \oplus \\ L_{11}[3,8,14,25] \oplus L_{11}[3]R_{11}[16,17,20] \oplus R_{11}[17] = \\ K[sth] + K[sth']L_0[3] + K[sth'']L_{11}[3] \end{array}\right\} \quad \frac{1}{2} \pm \text{around } 1.2 \cdot 2^{-15}$$

For a few different keys we have tried (long computation on a PC) the bias was **always strictly better than Matsui-11** that gives $\frac{1}{2} - 1.91 \cdot 2^{-16}$.

Remark: The best characteristics found by Matsui for 3 and 11 rounds [25] are closely related to those presented here: their difference is a biased Boolean function. BLC contains LC not only as a subset, but also as an extension allowing to strictly improve the best linear attacks on DES by adding higher degree monomials.

B.3 Beyond Bi-Linear Attacks: Using Cubic Equations

We observed that, for 3 rounds, even better results can be achieved using cubic partially bi-linear characteristics, instead of quadratic bi-linear (**) from Proposition B.1.1. Our simulations show that, for an important fraction of keys:

$$(***) \quad \left.\begin{array}{l} L_0[3,8,14,25] \oplus L_0[3]R_0[16,17,20]R_0[17,18,19,20] \oplus \\ L_3[3,8,14,25] \oplus L_3[3]R_3[16,17,20]R_3[17,18,19,20] \oplus \\ R_0[17] \oplus R_3[17] = K[sth] \end{array}\right\} \quad \frac{1}{2} - 1.82 \cdot 2^{-3}$$

The explanation why this works is quite similar. Though the non-linear part of this equation is not bi-linear, it is well correlated with a truly bi-linear function:

$$L[3]R[16,17,20]R[17,18,19,20] = L[3]R[16,17,20] \quad \text{with probability } 7/8$$

Unfortunately, the bias of $(***)$ is worse for other keys. On average, the best bias we know for 3 rounds remains $(**)$ from Proposition B.1.1. We also observed that that $(***)$ works for any number of DES rounds and for any key, but again the results are not as good as with $(**)$.