New Insights on Impossible Differential Cryptanalysis

Charles Bouillaguet[1], Orr Dunkelman[2,3],
Pierre-Alain Fouque[1], and Gaëtan Leurent[4]

[1] Département d'Informatique
École normale supérieure
45 Rue d'Ulm
75320 Paris, France
{charles.bouillaguet, pierre-alain.fouque}@ens.fr
[2] Computer Science Department
University of Haifa
Haifa 31905, Israel
orrd@cs.haifa.ac.il
[3] Faculty of Mathematics and Computer Science
Weizmann Institute of Science
P.O. Box 26, Rehovot 76100, Israel
[4] Faculty of Science, Technology and Communications
University of Luxembourg
6 Rue Richard Coudenhove-Kalergi
L-1359 Luxembourg
gaetan.leurent@uni.lu

Abstract. Since its introduction, impossible differential cryptanalysis has been applied to many ciphers. Besides the specific application of the technique in various instances, there are some very basic results which apply to generic structures of ciphers, e.g., the well known 5-round impossible differential of Feistel ciphers with bijective round functions.
In this paper we present a new approach for the construction and the usage of impossible differentials for Generalized Feistel structures. The results allow to extend some of the previous impossible differentials by one round (or more), answer an open problem about the ability to perform this kind of analysis, and tackle, for the first time the case of non-bijective round functions.
Keywords: Impossible differential cryptanalysis, Miss in the middle, Generalized Feistel, Matrix method.

1 Introduction

Impossible differential attack [3] is a method of using differential concepts in cryptanalytic attacks. While regular differential cryptanalysis [5] exploits differentials with as high probability as possible, impossible differential cryptanalysis exploits differentials that cannot happen, i.e., have probability of zero. The actual

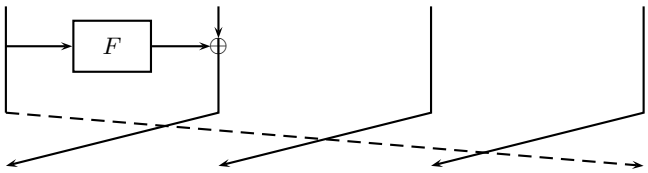


Fig. 1. CAST-like Structure with Four Threads

use of the impossible differential resembles the one of a high probability differentials: given a pair that may "satisfy" the differential, the adversary obtains the subkey(s) suggested by the pair. Unlike differential cryptanalysis, where such a subkey is more likely to be the right subkey, in impossible differential cryptanalysis, once a subkey is suggested by a candidate pair, it is necessarily a wrong one (and thus discarded).

To start an impossible differential attack, the adversary has to identify such impossible differentials. Most of these differentials are constructed in a miss-in-the-middle approach [4]. The approach is based on combining two probability 1 truncated differentials that cannot coexist. For example, there is a generic 5-round impossible differential for Feistel constructions with a bijective round function (first identified in [12]) of the form $(0, \alpha) \not\rightarrow (0, \alpha)$ (depicted in Figure 2).

A method for finding such impossible differentials is presented in [11] under the name $\mathcal{U}$*-method*. In this method, one can construct probability 1 truncated differentials, which in turn leads to finding contradictions. An automated version of the method is presented in [10]. The tool (called *the matrix method*). The automated analysis shows several results for generalizations of the Feistel cipher (the Generalized Feistel Network of [14], MARS-like constructions [6], and CAST-like constructions [1]).

As an example, consider a CAST-like construction (depicted in Figure 1). The matrix method suggests an impossible differential of $n^2 - 1$ rounds for $n \geq 3$ threads assuming that the round function is bijective. The impossible differential has the form of $(0, 0, \ldots, 0, \alpha) \not\rightarrow (0, 0, \ldots, 0, \omega)$ for any non-zero α and ω, and is based on the fact that the $(n-1)$-round truncated differential starting at $(0, 0, \ldots, 0, \alpha)$ predicts a zero difference in the one before last word, while the $n(n-1)$-round truncated differential ending at $(0, 0, \ldots, 0, \omega)$ predicts that the same word has a non-zero difference.

The $\mathcal{U}$-method was later improved in [13] to incorporate a much larger set of contradictions. Such new set of contradictions may include the use of specific differences in the input and the output (rather than truncated differences) or conditions on XORing a few words together.

In this paper we take a deeper look into the construction of impossible differentials. We start the analysis by considering a slightly different approach for the analysis, a one which does not classify the state of the word as part of a small

Structure	Number of		Round	Source
	Words	Rounds	Function	
Feistel	2	5	bijective	[12]
Generalized Feistel Network	2	7	bijective	[10]
Generalized Feistel Network	$2n$	$3n+2$	bijective	[10]
CAST-like	n	n^2-1	bijective	[11]
CAST-like	n	n^2+3	bijective	[7, 13]
MARS-like	n	$2n-1$	bijective	[10]
MARS-like	n	$2n+3$	bijective	[13]
RC6-like	$2n$	$4n+1$	bijective	[10]
CAST-like	n	n^2	any	Sect. 4
MARS-like	n	$2n$	any	Sect. 4

Table 1. Comparison of Impossible Differentials for Feistel Networks

set of values.[1] Instead, we try to look at the specific differences that may form a contradiction, taking the structure of the round function into account. The main property we use is the existence of impossible differentials in the round function.

This allows us to extend the impossible differentials by an additional round, leading to improved attacks on some structures of block ciphers. Moreover, following the new point of view, one can even reduce the requirements from the round function. For example, as part of our analysis, we can offer n^2-round impossible differentials for CAST-like ciphers, even if their round function is not bijective. We note that our results contradict a claim made in [16], which claims that "generic" impossible differentials for this structure exist only up to n^2-1 rounds. We compare the previously known results with our new results in Table 1.

We continue and define the *differential expansion rate* of a round function for a (set of) input difference(s). The rate tries to measure the speed in which the set of possible differences evolves through invocations of the round function. To some extent, it is the equivalent of the expanding rate of a graph.

We then study how to use our new impossible differential in an actual attack, and how useful is the new impossible differential. We describe attacks using our new extended impossible differentials, with the same time complexity as previous attacks (under some natural conditions on the round function), and covering more rounds.

The structure of this paper is as follows: In Section 2 we cover the basics of differential cryptanalysis and impossible differential cryptanalysis. Section 3 discusses the previous results and the matrix method. In Section 4 we suggest a new approach for constructing impossible differentials, and in Section 5 we show that impossible differential attacks that use the previous impossible differentials can be extended to more rounds when instantiated with our newly found im-

[1] The matrix method classifies the state of a word as one of five states: zero difference, fixed difference, unknown non-zero difference, the XOR of a fixed difference with an unknown non-zero difference, or unknown.

possible differentials (almost with no additional complexity). Finally, Section 6 concludes this paper.

2 Preliminaries

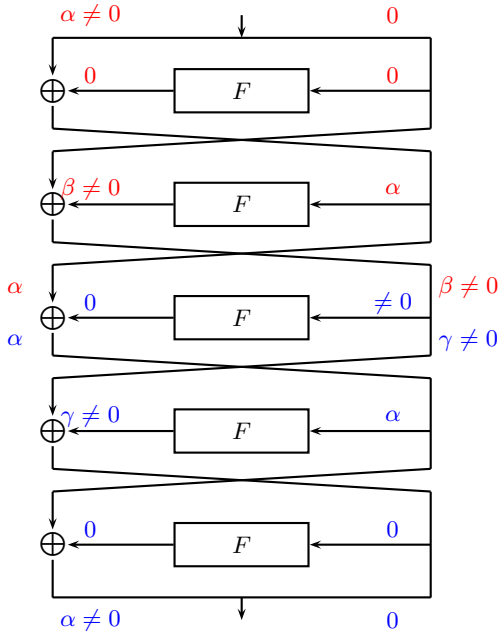
Differential cryptanalysis [5] is one of the corner stones of modern cryptanalytic techniques. It was used to successfully attack block ciphers, hash functions, and even stream ciphers. The core idea behind differential cryptanalysis is to look at the development of differences through the encryption function rather than at the actual values directly. This approach leads to a much stronger knowledge of the adversary concerning the encryption process, as it allows "replacing" the key addition with probabilistic behavior in the nonlinear parts of the cipher.

For sake of simplicity we shall concentrate on differential cryptanalysis used for the cryptanalysis of block ciphers. In such a case, the adversary first finds a differential (or a differential characteristic) of high probability p, e.g., $\Delta_{IN} \rightarrow \Delta_{OUT}$ with probability p. The differential can be for the full cipher, but in many cases, a slightly shorter differential is used. After identifying the differential (characteristic), the adversary asks for the encryption of $O(1/p)$ pairs of plaintexts with input difference Δ_{IN} and collects the corresponding ciphertexts. Then, the adversary tries to identify the subkey used in the last rounds of the cipher, by partially decrypting the ciphertext pairs, or by analyzing the last rounds of the differential characteristics. For the right subkey, it is expected that a few pairs with difference Δ_{OUT} appear, while the number of pairs with difference Δ_{OUT} is expected to be significantly lower for wrong guesses.

As the data complexity (and consequently, the time complexity) of differential attacks are proportional to $1/p$, the existence of high probability differentials is considered a weakness of the block cipher. Hence, many block cipher designers suggest methodologies to ensure that there are no differentials with high probability for (almost) the entire cipher. For example, in the case of AES [19], it can be shown that any 4-round differential characteristic has probability not higher than 2^{-150} [8] and that no 4-round differential with probability higher than 2^{-113} exists [9].

At that point, it was observed that differential cryptanalysis, as a statistical attack, uses the fact that the number of pairs counted for the right subkey guess and the wrong subkey guesses differs. The standard differential attack assumes that the number of pairs suggested by right subkey is higher than for a wrong subkey, but it is also possible to mount an attack when the number of pairs suggested by the right subkey is lower. This led to the introduction of impossible differential attacks (first at [12] as a dedicated attack on the DEAL cipher, and then as a general cryptanalytic tool at [3, 4]). These attacks are based on finding differentials whose probability is 0. Namely, for the right subkey guess *no pairs* with output difference Δ_{OUT} exist, while for wrong subkey guesses, such pairs may be "discovered".

Hence, the impossible differential attack is based on taking a set of plaintext pairs, asking for their encryption, and then partially decrypting these pairs



The miss-in-the-middle follows the fact that the input and output differences force the output difference of the third round to be 0. At the same time, due to the bijectiveness of the round function the input difference of the third round is necessarily non-zero. The two cannot coexist, as the round function is bijective.

Fig. 2. A Generic 5-Round Impossible Differential for Feistel Ciphers with a Bijective Round Function

under the subkey candidates. Once a subkey candidate suggests that a specific ciphertext pair "satisfies" the differential, i.e., that $\Delta_{IN} \rightarrow \Delta_{OUT}$ has occurred, we can be assured that this subkey is wrong and discard it.

The most successful method for constructing impossible differentials is the *miss in the middle* method. In this method, a probability one truncated differential $\Delta_{IN} \rightarrow \Delta_A$ and a probability one truncated differential in the backward direction $\Delta_B \leftarrow \Delta_{OUT}$ are identified, such that Δ_A and Δ_B cannot coexist simultaneously. For example, Figure 2 describes a 5-round Feistel construction with a bijective round function, for which $(\alpha, 0) \not\rightarrow (\alpha, 0)$ is an impossible differential.

2.1 Notations

In this paper we use the following notations:

- n — denotes the number of threads in a given structure.
- w — denotes the size (in bits) of a given thread.
- $\alpha, \beta, \ldots$ — denotes a non-zero difference.

– 0 — denotes a zero difference (in a thread).
– ? — denotes an unknown difference.
– $\rightarrow_i$, $\leftarrow_i$ — denotes the propagation of a (truncated) difference for i rounds in the encryption/decryption direction.
– $\alpha \rightsquigarrow \beta$ — denotes the event that an input difference α to a round function F may result in an output difference β, i.e., $Pr_x[F(x) \oplus F(x \oplus \alpha) = \beta] > 0$.

3 Previous Results and the Matrix Method

Similarly to looking for good differentials, the search for impossible differentials is not an easy task. While good impossible differentials were found by hand (e.g., the one of Figure 2 or the 24-round impossible differential of SKIPJACK from [3]), it was suggested to try and automate the process of finding these impossible differentials [10, 11, 13].

One tool, based on the $\mathcal{U}$-method (of [11]), is the matrix method [10], a mechanism to identify truncated differentials of probability 1. The intermediate encryption value is divided into words (or threads), where each such word can be in one of five states, each associated with a number: a zero difference (denoted by 0), a fixed non-zero difference (denoted by 1), a non-fixed non-zero difference (denoted by 1^*), the XOR of a fixed non-zero difference with a non-fixed one (denoted by 2), and an unknown difference (denoted by 2^* or any other number larger than 3, with or without *).

The tool targets mostly ciphers whose round function contains one nonlinear function, which is in itself a bijective function. The round function is represented as a matrix composed of 0's, 1's, and at most one special entry denoted by 1_F. The automated search starts with a vector $\{0,1\}^n$, which is multiplied by the special matrix, repeatedly, to reveal the properties of the truncated difference after each round.

The main idea behind the matrix multiplication, is to represent the possible transitions in a way that conveys the actual difference in the state. The matrix has size of $n \times n$ (for an n-thread construction). If thread i does not affect thread j, then entry (i, j) of the matrix is 0. If thread i affects thread j, then the entry is 1 (if thread i is XORed/copied into thread j) or 1_F (if thread i is sent through the nonlinear function F and the output is XORed or copied to thread j).

Now, one can define the arithmetics. For example, if the thread has state 0, then it has no affect on other threads (independent of the operation). A thread i whose state is 1, contributes 0, 1, or 1^*, to thread j when the corresponding entry in the matrix is 0,1, or 1_F, respectively. A thread i whose state is 1^*, contributes 0, 1^*, or 1^*, when the corresponding entry in the matrix is 0,1, or 1_F, respectively. Other states, α (or x^*), contribute 0, α (respectively, x^*), and $x + 1$, when the matrix is 0,1, or 1_F, respectively.

Then, each new thread is summed under the following addition rules: $0+x = x$, $1+1 = 1$, $1+1^* = 2$, $1+x = 2^*$ (for $x > 1$ or $x^* > 1^*$), and any other addition just sums the actual numbers (and maintains the *). This gives the new state after the round function, and one can compute as many rounds as possible, until

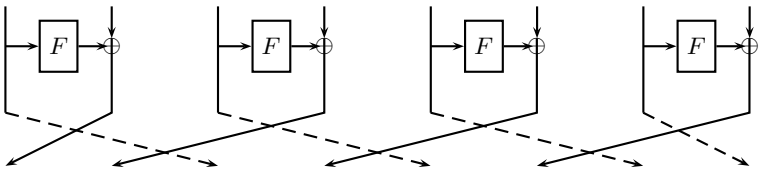


Fig. 3. A Generalized Feistel Network (GFN_4) with 8 Threads

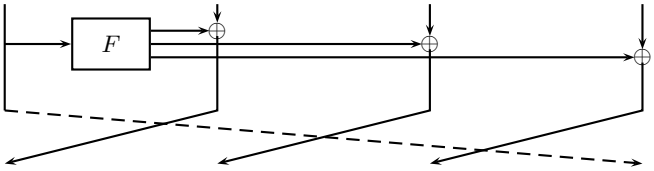


Fig. 4. A MARS-like Cipher

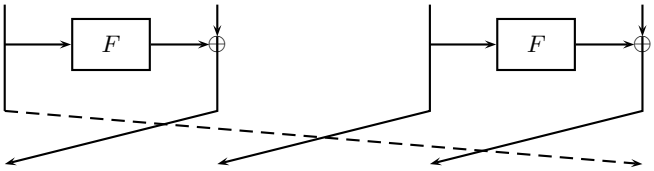


Fig. 5. An RC6-like Cipher

the entire intermediate encryption value is composed of only 2^* and $x > 2$ (with or without *), which denote the longest truncated differential of probability 1 that can be found and that conveys some "useful" characteristics.

It is also possible to run the same algorithm in the backward direction (with the corresponding matrix), to obtain the longest truncated differential or probability 1 of that direction.

Given two truncated differentials $\Delta_{IN} \rightarrow \Delta_A$ and $\Delta_B \leftarrow \Delta_{OUT}$, one can scan Δ_A and Δ_B to find inconsistencies. For example, if a word has a a non-zero difference (fixed or not) in Δ_A but a zero difference in Δ_B, both Δ_A and Δ_B cannot coexist, which leads to the miss in the middle differential $\Delta_{IN} \not\rightarrow \Delta_{OUT}$. This fact is described by the matrix method as pairs of contradicting states, e.g., 0 and 1 (or 0 and 1^*) or 1 and 2.

The method was applied for several constructions: Generalized Feistel Networks (introduced in [14]), CAST-like ciphers (based on the CAST-256 block cipher [1]), MARS-like ciphers (based on MARS [6]), RC6-like ciphers (based on RC6 [17]), and various variants of SKIPJACK-like ciphers [18]. We outline the structure of the first four structures in Figures 3, 1, 4, and 5, respectively.

For GFN with 4 threads (called GFN_2), there exist several 7-round impossible differentials assuming that the round function is bijective. For example, the input difference $(0, 0, 0, \alpha)$ becomes after 6 rounds $(?, ?, ?, \delta)$, while the output difference $(\beta_1, 0, \beta_2, 0)$ is decrypted by one round to the difference $(\beta_2, ?, 0, 0)$ which

cannot coexist. For sake of simplicity we use the notation $(0,0,0,\alpha) \rightarrow_6 (?,?,?,\delta)$ and $(\beta_2,?,0,0) \leftarrow_1 (\beta_1,0,\beta_2,0)$ to denote these truncated differentials. Combining these two truncated differentials we obtain that $(0,0,0,\alpha) \not\rightarrow_7 (\beta_1,0,\beta_2,0)$.

Similarly, for GFN_n (with $2n$ threads) there exists a $(3n+2)$-round impossible differential of the form $(0,0,\ldots,0,\alpha) \not\rightarrow_{3n+2} (\beta_1,0,\beta_2,0,0,\ldots,0)$ following the truncated differentials $(0,0,\ldots,0,\alpha) \rightarrow_{2n} (?,?,\ldots,?,\delta,?,?,?,?)$ and $(?,?,\ldots,?,0,?,?,?,?) \leftarrow_{n+2} (\beta_1,0,\beta_2,0,0,\ldots,0)$.[2]

For an n-thread CAST-like construction (for $n \geq 3$), there exists an (n^2-1)-round impossible differential $(0,0,\ldots,0,\alpha) \not\rightarrow_{n^2-1} (\beta,0,0,\ldots,0)$ following the truncated differentials $(0,0,\ldots,0,\alpha) \rightarrow_{3n-3} (?,?,\ldots,?,\delta)$ and $(?,?,\ldots,?,0) \leftarrow_{n^2-3n+2} (\beta,0,0,\ldots,0)$.

For an n-thread MARS-like construction (again, for $n \geq 3$), the two truncated differentials $(0,0,\ldots,0,\alpha) \rightarrow_{n+1} (?,?,\ldots,?,\delta)$ and $(?,?,\ldots,?,0) \leftarrow_{n-2} (\beta,0,0,\ldots,0)$ are combined into an $2n-1$-round impossible differential of the form $(0,0,\ldots,0,\alpha) \not\rightarrow_{2n-1} (\beta,0,0,\ldots,0)$.

In the case of RC6-like structure with n-threads, the impossible differential is $(0,0,\ldots,0,\alpha_i,0,\ldots,0) \not\rightarrow_{4n+1} (0,0,\ldots,0,\beta_{i+1},0,\ldots,0)$, where $\alpha_i = \beta_{i+1}$ and α_i is positioned in the ith thread (and β_{i+1} in the $(i+1)$'th thread) for some odd i.

For details concerning the SKIPJACK-like variants, we refer the interested reader to [10].

The UID-method of [13] is a generalization of the $\mathcal{U}$-method. In this variant, each word is not associated with a mere state, but its history (of the actual difference) is tracked. Using this history, it is possible to compose longer impossible differentials, as one may look at the XOR of a few words at some point (which may still contain non-trivial state information even after all words of the state become "?"). We note that this method still relies on the fact that the round function is bijective.

4 New Impossible Differentials

Our new impossible differentials on CAST-like and MARS-like ciphers follow a more subtle analysis.

4.1 CAST-like ciphers

We first consider a 4-thread CAST-like cipher. In such a cipher, there is a 4-round truncated differential with probability 1 of the form $(0,0,0,\alpha) \rightarrow_4 (\beta,0,0,\alpha)$ for non-zero α and some β (which may be zero if F is not bijective). At the same time, there exists a 12-round truncated differential in the decryption direction of the form $(\omega,?,?,\phi) \leftarrow_{12} (\omega,0,0,0)$ for non-zero ω and some ϕ (which may be zero if the round function is not bijective). We outline the differentials in Table 2.

[2] We note that in [10], a small typo suggests that the word which causes the contradiction is the fourth from the right while it is actually the fifth.

Round	Difference	Round	Difference
Input (0)	$(0,0,0,\alpha)$	Output (16)	$(\omega,0,0,0)$
1	$(0,0,\alpha,0)$	15	$(0,\omega,0,0)$
2	$(0,\alpha,0,0)$	14	$(0,0,\omega,0)$
3	$(\alpha,0,0,0)$	13	$(0,0,0,\omega)$
4	$(\beta,0,0,\alpha)$	12	$(\omega,\psi,0,0)$
		11	$(0,\omega,\psi,0)$
		10	$(0,0,\omega,\psi)$
		9	$(\psi,\chi,0,\omega)$
		8	$(\omega,?,\chi,0)$
		7	$(0,\omega,?,\chi)$
		6	$(\chi,\phi,\omega,?)$
		5	$(?,?,\phi,\omega)$
		4	$(\omega,?,?,\phi)$

Differences are given after the round.

Table 2. The Two Truncated Differentials Used in Our New 16-Round Impossible Differential on 4-Thread CAST-like Ciphers

Observation 1 *We note that the above two truncated differentials can coexist if and only if $\beta = \omega$. Hence, if an input difference α to the round function may not cause an ω difference at the output, i.e., if $\alpha \not\rightsquigarrow \omega$ is an impossible differential for F, then these two differentials cannot coexist, and we obtain a 16-round impossible differential of the form $(0,0,0,\alpha) \not\rightarrow_{16} (\omega,0,0,0)$ for the cipher.*

Given the structure of the round function F, it is possible to determine whether $\alpha \rightsquigarrow \omega$ through F. Consider for example the round function of DES, for which determining whether $\alpha \rightsquigarrow \omega$ can be easily done by checking each of the 8 S-boxes separately. We note that in DES' round function, given a pair of random input/output differences from an S-box, there is an 80% chance of the transition being possible. Hence, for a random α and ω, the probability of $x \rightsquigarrow a$ is only $0.8^8 \approx 0.17$.[3]

In the more general case, where the form of the round function is $F_k(x) = G(x \oplus k)$, one can exhaustively try all possible pairs with input difference α, and see if any of them leads to ω output difference. For a w-bit $G(\cdot)$ this takes 2^w invocations of $G(\cdot)$, even if we only have a black box access to $G(\cdot)$ (but not to $F_k(\cdot)$). Of course, when the description of $G(\cdot)$ is known, this verification is expected to be significantly faster. As we show in Section 5, even under the worst case assumption, i.e., when $G(\cdot)$ is unknown, this has no real effect on the actual attack that uses this impossible differential.

[3] Even though the actual inputs to the different S-boxes are independent of each other, assuming the key is chosen randomly, the differences are not. Hence, the actual value of the full round function may be slightly different.

Moreover, we note that for a function $G(\cdot)$ of this form, the probability that $\alpha \rightsquigarrow \omega$ is at most 0.5 for a random[4] α and ω (following the fact that the row corresponding to α in the difference distribution table has at most half of its entries as non-zero). If we assume that $G(\cdot)$ is a random function, then according to [15] we can determine that about 60.6% of the possible (α, ω) pairs yield an impossible differential.

An interesting point concerning the truncated differentials suggested above is the fact that their existence is independent of the actual differential properties of the round functions. Namely, in the case $F_k(\cdot)$ is not bijective, the above truncated differentials still hold, and thus, also the impossible differential. More precisely, even if different round functions are used, the only one of interest is the one of round 4.

Now, one can easily generalize the above impossible differential, and can easily see that for an n-thread CAST-like block cipher, the following is an impossible differential: $(0, 0, 0, \ldots, \alpha) \rightarrow_{n^2} (\omega, 0, \ldots, 0)$ if $\alpha \not\rightsquigarrow \omega$ following the n-round truncated differential $(0, 0, 0, \ldots, 0, \alpha) \rightarrow_n (\beta, 0, 0, \ldots, 0, \alpha)$ and the $n(n-1)$-round truncated differential $(\omega, ?, \ldots, ?, \phi) \leftarrow_{n(n-1)} (\phi, 0, \ldots, 0)$.

4.2 MARS-like ciphers

The same approach can also be used to extend the impossible differential suggested for a MARS-like structure. As before, we start with a 4-thread example, and then generalize it. In such a cipher, there is a 5-round truncated differential $(0, 0, 0, \alpha) \rightarrow_5 (?, ?, ?, \beta)$ and a 3-round truncated differential $(\omega, 0, 0, 0) \leftarrow_3 (0, 0, 0, \omega)$. As β is the output difference caused by an α input difference, the two can coexist if and only if $\alpha \rightsquigarrow \omega$ through the corresponding $F(\cdot)$. We outline the differentials in Table 3.[5]

We can of course generalize the above truncated differentials for the case of an n-thread MARS-like cipher. The backwards differential is the same, i.e.,

[4] Most impossible differential attacks face a large amount of (α, ω) pairs which are generated in a random manner.

[5] We note that the differentials presented in Table 3 assume that the differences that are XORed into each of the three threads is different (as in the real MARS there are three different functions). When the same output is XORed into all the three threads (in the real MARS, additions and subtractions are also used) then one can construct a longer impossible differential for 9 rounds. In the forward direction we use the following 5-round differential:

$$(0, 0, 0, \alpha) \rightarrow_4 (\beta, \beta, \beta, \alpha) \rightarrow (\gamma, \gamma, \delta, \beta)$$

where $\delta = \alpha \oplus \beta \oplus \gamma \neq \gamma$ (whenever $\alpha \not\rightsquigarrow \alpha$ through $F(\cdot)$), and in the backward direction we use the following 4-round differential:

$$(\omega, \psi, \psi, \psi) \leftarrow (0, 0, 0, \omega) \leftarrow_3 (\omega, 0, 0, 0)$$

and it is easy to see that the two cannot coexist, as the XOR of the two intermediate words cannot be the same.

Round	Difference	Round	Difference
Input (0)	$(0,0,0,\alpha)$	Output (8)	$(\omega,0,0,0)$
1	$(0,0,\alpha,0)$	7	$(0,\omega,0,0)$
2	$(0,\alpha,0,0)$	6	$(0,0,\omega,0)$
3	$(\alpha,0,0,0)$	5	$(0,0,0,\omega)$
4	$(\beta,\gamma,\delta,\alpha)$		
5	$(?,?,?,\beta)$		

Differences are given after the round.

Table 3. The Two Truncated Differentials Used in Our New 8-Round Impossible Differential on 4-Thread MARS-like Ciphers

an $(n-1)$-round differential of the form $(0,0,\ldots,0,\omega) \leftarrow_{n-1} (\omega,0,\ldots,0)$ and the forward differential is of the form $(0,0,\ldots,0,\alpha) \rightarrow_{n+1} (?,?,\ldots,?,\beta)$ which cannot coexist if $\alpha \not\rightsquigarrow \omega$ through the corresponding $F(\cdot)$.

4.3 A Larger Class of Impossible Differentials

We can extend the above impossible differentials by taking an even closer look into the round function. Instead of looking for impossible differential in the round function, we now look for impossible differentials in the iterated round function. We can do this more delicate analysis based on the following definition of the output difference set of an unkeyed function $F(\cdot)$ and a set S of input difference:

Definition 1. *For a function $F(\cdot)$ and a set ΔS of input differences, we define the output difference set $\Delta F(\Delta S)$ to be the set containing all the output differences that are feasible by an input difference in ΔS.*

Now, we can define the differential expansion rate of an unkeyed function $f(\cdot)$:

Definition 2. *The* differential expansion rate *of a function $F(\cdot)$ is*

$$\max_{|\Delta S|>0} \frac{|\Delta F(\Delta S)|}{|\Delta S|},$$

i.e., the maximal increase in the size of a difference set through the round function.

We first note that the above definitions are set for unkeyed functions. However, for round functions of the form $F_k(x) = G(x \oplus k)$, one can disregard the key addition, and use the same results. Moreover, once the key is fixed, this is the case for any round function. For the following discussion, we shall assume that indeed $F(\cdot)$ is of that form.

Now, if the differential expansion rate of a function is small, then $\Delta F(\Delta F(\{\alpha\}))$ for a fixed input difference α may not be large enough to cover all possible differences. Assume that this is indeed the case for a round function $F(\cdot)$ (we later

describe an example of such a round function), then one can easily extend the 16-round impossible differential for CAST-like structure with 4 threads by one round by using the following truncated differentials: $(0,0,0,\alpha) \rightarrow_5 (\gamma,0,\alpha,\beta)$ and $(\omega,?,?,\phi) \leftarrow_{12} (\omega,0,0,0)$. If $\omega \notin \Delta F(\Delta F(\{\alpha\}))$, one can easily see that $(0,0,0,\alpha) \not\rightarrow_{17} (\omega,0,0,0)$.

More generally, if the differential expansion rate is $c < 2^{w/2}$ then $|\Delta F(\Delta F(\{x\}))| < 2^w$, which means that there are many ω values for which $\omega \notin \Delta F(\Delta F(\{\alpha\}))$. The smaller the value of c is, there is a larger set of differences which are not in $|\Delta F(\Delta F(\{\alpha\}))|$, and thus, allow for longer impossible differentials.

These results can be generalized. If $c < 2^{w/3}$, then the above arguments can be repeated and the forward differential can be extended to a 6-round differential $(0,0,0,\alpha) \rightarrow_6 (\delta,\alpha,\beta,\gamma)$ where $\delta \in \Delta F(\Delta F(\Delta F(\{\alpha\})))$, and if $\omega \notin \Delta F(\Delta F(\Delta F(\{\alpha\})))$, then obviously both differentials cannot coexist. Extending this analysis to more rounds is feasible, by taking into consideration that the next set of differences is XORed with a difference α (which affects the difference set, but not its size).

We note that even when the differential expansion rate is large, there are still cases where we can extend the 16-round impossible differential. This follows the fact that the differential expansion rate may be determined by a special set of differences that are not relevant for the impossible differential.

Consider for example a CAST-like structure with 4 threads, whose round function is from 64 bits to 64 bits, and is based on applying eight 8-bit to 8-bit S-boxes in parallel, accompanied by a linear transformation L (e.g., the round function of Camellia [2]). We can even assume that this linear transformation has a branch number of 9, which ensures that a difference in one S-box affects all output bytes. Following the properties of differential cryptanalysis, consider an input difference α with one active byte, where all other bytes have a zero difference. $\Delta F(\{\alpha\})$, thus, contains at most 128 possible differences, each with all the bytes active. For each such difference, applying F again, can yield at most $128^8 = 2^{56}$ possible differences. This implies that the size of $\Delta F(\Delta F(\{\alpha\}))$ is upper bounded by 2^{63}, which allows the extension of the impossible differential to 17 rounds.

If the linear transformation L does not have a maximal branch number b (e.g., the actual round function of Camellia uses a linear transformation with branch number of 4), then we can extend the attack to 18 rounds. Indeed, given values ω and θ with a single active S-box and $\omega = L\theta$, we have $\omega \notin \Delta F(\Delta F(\Delta F(\{\alpha\})))$ for most choices of ω and u. This comes from the fact that $\alpha \not\rightarrow \omega$ is an impossible differential for $F \circ F \circ F$, following the miss in the middle principle. The differences in $\Delta F(\{\omega\})$ all have the same pattern of b active S-boxes with some inactive S-boxes. On the other hand, the differences in $\Delta F^{-1}(\{\omega\})$ have a single S-box (the same as in θ), therefore the differences in $\Delta F^{-1}(\Delta F^{-1}(\{\omega\}))$ all have the same pattern of at least b active S-boxes, with some inactive ones. If ω and α are chosen so that the pattern are incompatible, we have $\alpha \not\rightarrow \omega$ for $F \circ F \circ F$.

There are two issues that need to be addressed using this new extension. The first is what is the ratio of impossible combinations. In the first example given

above, the probability that for a random ω, and a random α of the form suggested, the probability that $\omega \notin \Delta F(\Delta F(\{\alpha\}))$ is indeed at least 0.5, which still offers a high probability of contradiction (which is needed to form the impossible event).

The second concern is the ability to check whether a given ω is in $\Delta F(\Delta F(\{\alpha\}))$. Unfortunately, at the moment, even if $F(\cdot)$ is of the form $F_k(x) = G(x \oplus k)$, for an unknown $G(\cdot)$, we are not aware of any algorithm, besides enumerating all possible differences. At the same time, if the structure of the round function is known, it can be used to offer an easy way to check whether $\omega \in \Delta F(\Delta F(\{\alpha\}))$

For example, in the above example (with a Camellia round function), it is possible to use a meet-in-the-middle approach. First, apply the inverse linear transformation on ω, and obtain the required output differences in every S-box of the second $F(\cdot)$. Then, by trying all 128 values of $\Delta F(\{\alpha\})$ one can check whether the difference distribution table of the S-box offers this transition.

4.4 Changes to the Matrix Method

We note that it is possible to extend the matrix method of [10] such that it would suggest impossible differentials of the above structure. The main idea behind the change is to know for each non-fixed input difference the size of the set of possible differences.

The simplest change would be to store for each initial state the size of possible differences (which is 1 for each word, either active or not). Then, when an active word passes through the round function, the size of the set is increased by a factor of c, the differential expansion rate of the round function. Finally, when XORing two active thread, one with t_1 options, and one with t_2 options, the number of possible differences in the output is at most $t_1 \cdot t_2$.

In the step when we look for contradictions, we first search for the previous class of contradictions. Then, we also look for pairs of words, one in Δ_A (with t_1 options) and one in Δ_B (with t_2 options), such that $t_1 \cdot t_2 < 2^w$, as for such words, it is probable that the differences cannot coexist (the probability for contradiction is $1 - t_1 \cdot t_2/2^w$).

4.5 A 7-Round Impossible Differentials for Feistel Block Ciphers with Small Differential Expansion Rate

For some block ciphers with small differential expansion rate (or whose round function allows selecting such differences), it is possible to suggest a 7-round impossible differential. The impossible differential is based on two truncated differentials of three rounds each $(\alpha, 0) \rightarrow_3 (\{X\}, ?)$ and $(\{Y\}, ?) \leftarrow_3 (\omega, 0)$, where

$$\{X\} = \{\alpha \oplus \beta | \beta \in \Delta F(\Delta F(\{\alpha\}))\}$$

and

$$\{Y\} = \{\omega \oplus \psi | \psi \in \Delta F(\Delta F(\{\omega\}))\}.$$

If the differential expansion rate of the round function is smaller than $2^{w/4}$, then it is expected that $|\{X\}|\cdot|\{Y\}| < 2^n$, which means that there are combinations of X and Y which cannot coexist. We note that this impossible differential does not assume that the round function is bijective. We note that the 7-round impossible differential of DES mentioned in [4] can be found using this approach (when starting with α and ω for which there is only one active S-box).

5 New Attacks Using the New Impossible Differentials

Given the new impossible differentials, we need to show that they can be used for the analysis of block ciphers. As mentioned before, our impossible differentials are more restricted than the previous ones, and thus they may be of a lesser usage in attacks.

To show the contrary, we consider an attack which uses the 16-round impossible differential on 4-thread CAST-like structure and compare it to a similar attack the uses the original 15-round impossible differential. As before, for simplicity, we shall consider round functions of the form $F_k(x) = G(x \oplus k)$, which are very common in block ciphers.

We first note that both the 16-round impossible differential and the 15-round impossible differential share the same structure, i.e., $(0,0,0,\alpha) \not\rightarrow (\omega,0,0,0)$. Hence, the use of structures and early abort in the attacks is (almost) the same.

In Figure 6 we compare the 16-round attacks using the 15-round impossible differential (there are two variants, in one the additional round is before the impossible differential, and in the second it is after the impossible differential) with the 17-round attacks using the 16-round impossible differential. As can be seen, the attack algorithms are very similar, and the analysis of them is also very similar. For example, the data complexity of the 16-round attack with an additional round after the impossible differential is $2w \cdot 2^{2w}$ chosen plaintexts, while for the equivalent 17-round attack, the data complexity is $4w \cdot 2^{2w}$ chosen plaintexts.[6] We compare the complexities of these attacks in Table 4.

Rounds in Imp. Diff.	Attacked Round	Size of Structures	Complexity Data	Complexity Time	Complexity Memory
15	After	2^w	$2w \cdot 2^{2w}$	$2w \cdot 2^{2w}$	2^w
	Before	2^{2w}	$w \cdot 2^{2w}$	$w \cdot 2^{2w}$	2^{2w}
16	After	2^w	$4w \cdot 2^{2w}$	$4w \cdot 2^{2w}$	2^w
	Before	2^{2w}	$2w \cdot 2^{2w}$	$2w \cdot 2^{2w}$	2^{2w}

Table 4. Comparison of the complexities of the $(n+1)$-round attacks

[6] This assumption is made under the worst case assumption, where the function $G(\cdot)$ is an almost perfect nonlinear permutation (for which half of the input/output differences α and ω satisfy that $\omega \in \Delta G(\{\alpha\})$).

16-Round Attacks	17-Round Attacks
– Pick structures of the form $(A_i, B_i, C_i, \star)$ (where A_i, B_i, C_i are fixed in the structure), and ask for their encryption. – Locate in each structure (independently) ciphertext pairs whose difference is $(\psi, 0, 0, \omega)$. – For each remaining pair, discard any subkey K_{16} that suggest that the difference before the last round is $(\omega, 0, 0, 0)$.	– Pick structures of the form $(A, B, C, \star)$ (where A_i, B_i, C_i are fixed in the structure), and ask for their encryption. – Locate in each structure (independently) ciphertext pairs whose difference is $(\psi, 0, 0, \omega)$, and denote their plaintext difference by $(0, 0, 0, \alpha)$. – If $\omega \in \Delta F(\{\alpha\})$, discard the pair. – For each remaining pair, discard any subkey K_{17} that suggest that the difference before the last round is $(\omega, 0, 0, 0)$.
– Pick structures of the form $(\star, \star, C_i, D_i)$ (where C_i, D_i are fixed in the structure), and ask for their encryption. – Locate in each structure (independently) ciphertext pairs whose difference is $(\omega, 0, 0, 0)$, and their plaintext difference is $(\alpha, \beta, 0, 0)$. – For each remaining pair, discard any subkey K_1 that suggest that the difference after the first round is $(0, 0, 0, \alpha)$.	– Pick structures of the form $(\star, \star, C_i, D_i)$ (where C_i, D_i are fixed in the structure), and ask for their encryption. – Locate in each structure (independently) ciphertext pairs whose difference is $(\omega, 0, 0, 0)$, and their plaintext difference is $(\alpha, \beta, 0, 0)$. – If $\omega \in \Delta F(\{\alpha\})$, discard the pair. – For each remaining pair, discard any subkey K_1 that suggest that the difference after the first round is $(0, 0, 0, \alpha)$.

Fig. 6. Attacks of $(n + 1)$ rounds using an n-round impossible differentials

We note that the attack, requires the identification whether $\omega \in \Delta G(\{\alpha\})$. We note that in the worst case, this requires calling $G(\cdot)$ about 2^w times (with all 2^{w-1} pairs of distinct pairs with input difference α). However, the number of candidate α and ω's is about $O(w \cdot 2^w)$ (depending on the attack), whose evaluation is faster than evaluating the full block cipher. Moreover, by collecting the pairs of α, ω, one can check several pairs using the same invocations of $G(\cdot)$, thus incurring very little overhead to the attack.

In cases where α and ω are known, one can discard the pairs for which $\alpha \rightsquigarrow \omega$ beforehand, and repeat the same steps as in the original attack. This allows extending previous attacks by one more round, in exchange for at most twice the data and time. In other cases, where there are more candidate pairs,

and when α and ω cannot be determined directly from the plaintext/ciphertext pairs, one can postpone the verification whether $\omega \notin \Delta G(\{\alpha\})$, to the step just before discarding the full subkey. If such an attack uses an early abort approach (i.e., stops the analysis of a pair immediately as it found to be useless), it is possible to show that performing this check only when ω and α are both known, again increases the data and time by a factor of two at most.

We conclude that the new attacks are indeed one round longer (when the impossible differential is one round longer), and can be made longer, depending on the exact impossible differential. At the same time, the data and time complexities of the attacks increase by at most factor of two (the accurate increase is the $1/p$ where p is the ratio of non-zero entries in the difference distribution of $G(\cdot)$).

Finally, we note that when more complex impossible differentials are used, the same results apply, as long as the differential expansion rate of $G(\cdot)$ is small enough, or in the cases where the structure of $G(\cdot)$ allows quick verification of the existence of contradiction.

6 Summary and Conclusions

In this paper we show how to extend several impossible differentials for generalized Feistel schemes by one or more round, using a more subtle analysis of the round function. We follow and show that attacks which are based on these new impossible differentials require almost the same data and time complexity as the previous attacks, which proves that these impossible differentials are not only of theoretical interest, but can also be used in the analysis of block ciphers.

The new measure we introduced, the differential expansion rate of a round function, is expected to motivate block ciphers designers to re-think some of the basic approaches in block cipher design. For example, it is commonly believed that even if only a small amount of nonlinearity is used in the round function, then the cipher can still be secure. While this belief is not necessarily contradicted by our findings, we do show that it is possible to exploit this small nonlinearity in more complex attacks, such as impossible differential attacks, a combination that was not suggested before.

Additionally, our results may suggest that constructions which take the opposite approach than MARS, i.e., strong outer rounds with weaker inner rounds, may be susceptible to impossible differential attacks. This follows the fact that the development of difference sets that interest us, happen not in the outer rounds, but instead in the inner rounds.

Acknowledgements

We are grateful to the *Lesamnta* team, and especially to Hirotaka Yoshida, for helping us with this research. We would also like to thank Nathan Keller and Adi Shamir for the fruitful discussions and comments.

References

1. Adams, C., Heys, H., Tavares, S., Wiener, M.: The CAST-256 Encryption Algorithm (1998) AES submission.
2. Aoki, K., Ichikawa, T., Kanda, M., Matsui, M., Moriai, S., Nakajima, J., Tokita, T.: Camellia: A 128-Bit Block Cipher Suitable for Multiple Platforms - Design and Analysis. In Stinson, D.R., Tavares, S.E., eds.: Selected Areas in Cryptography. Volume 2012 of Lecture Notes in Computer Science., Springer (2000) 39–56
3. Biham, E., Biryukov, A., Shamir, A.: Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In: EUROCRYPT. (1999) 12–23
4. Biham, E., Biryukov, A., Shamir, A.: Miss in the Middle Attacks on IDEA and Khufu. In Knudsen, L.R., ed.: FSE. Volume 1636 of Lecture Notes in Computer Science., Springer (1999) 124–138
5. Biham, E., Shamir, A.: Differential Cryptanalysis of the Data Encryption Standard. Springer (1993)
6. Burwick, C., Coppersmith, D., DAvignon, E., Gennaro, R., Halevi, S., Jutla, C., Jr., S.M.M., OConnor, L., Peyravian, M., Safford, D., Zunic, N.: MARS - a candidate cipher for AES (1998) AES submission.
7. Choy, J., Yap, H.: Impossible Boomerang Attack for Block Cipher Structures. In Takagi, T., Mambo, M., eds.: IWSEC. Volume 5824 of Lecture Notes in Computer Science., Springer (2009) 22–37
8. Daemen, J., Rijmen, V.: AES Proposal: Rijndael (1998) NIST AES proposal.
9. Keliher, L., Sui, J.: Exact Maximum Expected Differential and Linear Probability for 2-Round Advanced Encryption Standard (AES) (2005) IACR ePrint report 2005/321.
10. Kim, J., Hong, S., Lim, J.: Impossible differential cryptanalysis using matrix method. Discrete Mathematics **310**(5) (2010) 988–1002
11. Kim, J., Hong, S., Sung, J., Lee, C., Lee, S.: Impossible Differential Cryptanalysis for Block Cipher Structures. In Johansson, T., Maitra, S., eds.: INDOCRYPT. Volume 2904 of Lecture Notes in Computer Science., Springer (2003) 82–96
12. Knudsen, L.R.: Deal — A 128-bit Block Cipher (1998) AES submission.
13. Luo, Y., Wu, Z., Lai, X., Gong, G.: A Unified Method for Finding Impossible Differentials of Block Cipher Structures (2009) IACR ePrint report 2009/627.
14. Nyberg, K.: Generalized Feistel Networks. In Kim, K., Matsumoto, T., eds.: ASIACRYPT. Volume 1163 of Lecture Notes in Computer Science., Springer (1996) 91–104
15. O'Connor, L.: On the Distribution of Characteristics in Bijective Mappings. In Helleseth, T., ed.: EUROCRYPT. Volume 765 of Lecture Notes in Computer Science., Springer (1993) 360–370
16. Pudovkina, M.: On Impossible Truncated Differentials of Generalized Feistel and Skipjack Ciphers (2009) presented at the rump session of the FSE 2009 workshop, available online at http://fse2009rump.cr.yp.to/e31bba5d1227eac5ef0daa6bcbf66f27.pdf.
17. Rivest, R.L., Robshaw, M.J., Sidney, R., Yin, Y.L.: The RC6 Block Cipher (1998) AES submission.
18. US Government: SKIPJACK and KEA Algorithm Specification (1998)
19. US National Institute of Standards and Technology: Advanced Encryption Standard (2001) Federal Information Processing Standards Publications No. 197.