IACR Transactions on Symmetric Cryptology
ISSN 2519-173X, Vol. 2022, No. 2, pp. 141–160. DOI:10.46586/tosc.v2022.i2.141-160

Truncated Differential Attacks on Contracting Feistel Ciphers

Tim Beyne[1] and Yunwen Liu[2]

[1] imec - Computer Security and Industrial Cryptography (COSIC) Research Group, Department of Electrical Engineering (ESAT), KU Leuven, Belgium
tim.beyne@esat.kuleuven.be
[2] Independent researcher, China
univerlyw@hotmail.com

Abstract. We improve truncated differential attacks on t-branch contracting Feistel ciphers with a domain size of N^t. Based on new truncated differentials, a generic distinguisher for $t^2 + t - 2$ rounds using $O(N^{t-1})$ data and time is obtained. In addition, we obtain a key-recovery attack on $t^2 + 1$ rounds with $\widetilde{O}(N^{t-2})$ data and $\widetilde{O}(N^{t-1})$ time. Compared to previous results by Guo *et al.* (ToSC 2016), our attacks cover more rounds with a lower data-complexity. Applications of the generic truncated differential to concrete ciphers include full-round attacks on some instances of GMiMC-crf, and the best-known key-recovery attack on 17 rounds of the Chinese block cipher standard SM4. In addition, we propose an automated search method for truncated differentials using SMT, which is effective even for trails with probability below the probability of the truncated differential for a random permutation.

Keywords: Truncated differentials · Contracting Feistel ciphers · SMT · GMiMC · SM-4

1 Introduction

Following its inception by Horst Feistel in the 1970s, the Feistel structure has become one of the most prominent architectures in modern block cipher design. One of its most eminent applications is undoubtedly the former American block cipher standard DES. Hence, it is not unexpected that the design and analysis of variants of the Feistel structure has become a significant research topic with valuable applications.

Following the widespread use of Feistel ciphers, many variations on the original structure were proposed. One of the main directions of this research has been the exploration of Feistel-like structures with more than two branches. Examples include the family of *generalized Feistel ciphers* [Nyb96, ZMI90] and the unbalanced Feistel ciphers discussed by Schneier and Kelsey [SK96]. The family of unbalanced Feistel structures can be further subdivided into expanding and contracting constructions. This paper is concerned with the security of the latter structure. Figure 1 shows a single Feistel round of a *contracting Feistel cipher* with $t = 4$ branches.

Examples of contracting Feistel ciphers include the algebraic cipher GMiMC-crf [AGP+19] and the general-purpose block cipher SM4 [Dt08]. The latter example is particularly important, as SM4 is the Chinese commercial block cipher standard (GB/T 32907-2016). In addition, it has been standardized by ISO/IEC under the reference number 18033-3:2010.

Given their widespread application, it is not surprising that the security analysis of Feistel ciphers and their variants has been an industrious area of research. Luby and

Tim Beyne is supported by a PhD Fellowship from the Research Foundation – Flanders (FWO).

Received: 2021-11-23 Revised: 2022-03-01 Accepted: 2022-05-01 Published: 2022-06-10

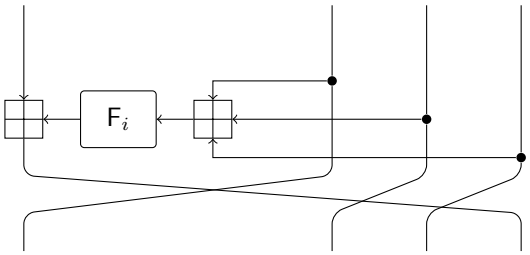

Figure 1: One round of a contracting Feistel cipher with $t = 4$ branches. The function F_i is a permutation and possibly key-dependent.

Rackoff [LR86] were first to prove the indistinguishability of three-round Feistel ciphers with uniform random round functions. Yun, Park and Lee [YPL11] proved the birthday-bound security of t-branch[1] contracting Feistel ciphers with $2t - 1$ rounds. However, from a practical point of view, optimal security is expected and desired if the number of rounds is large enough. Hence, several works have proposed generic attacks on contracting Feistel ciphers – thereby lower bounding the number of rounds necessary for security. In particular, Guo *et al.* [GJNS16] describe meet-in-the-middle attacks on contracting Feistel ciphers. Patarin, Nachef and Berbain [PNB06] analyze a more general contracting structure.

Differential cryptanalysis [BS91] has proven to be one of the most successful tools in the security analysis of both concrete and generic Feistel structures. For example, the generic attacks of Patarin [Pat04] on ordinary Feistel ciphers are based on differential cryptanalysis. The differential attack itself has also been extended and generalized in several ways. At FSE 1994, Knudsen [Knu95] introduced a powerful extension known as *truncated differential cryptanalysis*.

In this paper, we analyze the security of generic contracting Feistel ciphers using truncated differentials. Our motivation for doing so is twofold. On the one hand, from the viewpoint of block cipher design, it is important to know the baseline number of rounds required for security. In other words: how many rounds does a t-branch contracting Feistel cipher need? On the other hand, new generic attacks may have an impact on the security of concrete ciphers such as GMiMC-crf and SM4. SM4 in particular has received a significant amount of dedicated cryptanalysis and given its status as both a domestic and international standard, further advances in its analysis would be of interest to observers.

Contributions. Improved generic attacks on contracting Feistel ciphers are obtained using truncated differential cryptanalysis. In particular, when the security-level is equal to the block size, we obtain distinguishers and key-recovery attacks on more rounds than previous works for any number of branches. As an immediate consequence, full-round attacks on some instances of GMiMC-crf are obtained. In addition, our attacks lead to the best-known key-recovery attack on 17-round SM4.

With respect to generic attacks, the main result of this paper is a distinguisher on $t^2 + t - 2$ rounds of a contracting Feistel cipher with t branches and a domain of size N^t using $O(N^{t-1})$ data and time. In addition, we obtain a key-recovery attack on $t^2 + 1$ rounds requiring $\widetilde{O}(N^{t-2})$ data and $\widetilde{O}(N^{t-1})$ time. This is a significant improvement over the results of Guo *et al.* [GJNS16]. In particular, the key-recovery attacks of Guo *et al.* cover at most $5t - 4$ rounds (assuming the key length is equal to the block length).

The starting point for our attacks is an iterated truncated differential for generic contracting Feistel ciphers. It is similar to a truncated differential for the expanding Feistel cipher GMiMC-erf [AGP+19] that was presented at CRYPTO 2020 [BCD+20]. The basic iterated differential covers at most $t^2 - t - 2$ rounds. Our final $(t^2 + t - 2)$-

[1] The number of branches t must be even in order to avoid a trivial differential distinguisher.

round distinguisher is based on a different truncated differential that relies on several improvements not considered in previous work on contracting or expanding Feistel ciphers. In particular, we consider truncated differential trails whose probability p_{trail} is lower than their ideal probability p_{ideal}, take advantage of relations between input and output differences, and optimize the trade-off between the size of input structures and other parameters such as p_{trail} and p_{ideal}.

When applied to the algebraic cipher GMiMC-crf, the aforementioned attacks result in full-round distinguishers and key-recovery attacks for some instances. However, the practical relevance of these attacks may be relatively limited because most applications of GMiMC-crf use a relatively small number of branches t but a large N. In such cases, algebraic attacks become the dominant threat vector.

The data- and time-complexity of our 18-round distinguisher on SM4 are approximately 2^{96}. For 17-round SM4, we obtain a key-recovery attack using 2^{70} chosen plaintexts and 2^{99} encryption operations. Although dedicated attacks on SM4 reach up to 23 rounds, their data- and time-complexity is extremely large. As will be argued in Section 6, our key-recovery attack is the best-known attack for 17 rounds. This is remarkable given the fact that our attacks do not use any details about the round function of SM4.

Finally, we show how the propagation of truncated differentials in a contracting Feistel cipher can be modelled as a Satisfiability Modulo Theories (SMT) problem. This allows us to show that the distinguishers we obtain for $t = 4$ have optimal data-complexity. Importantly, our SMT model is able to analyze truncated differentials with $p_{\mathsf{trail}} \ll p_{\mathsf{ideal}}$. This is in contrast to previous methods, such as the MILP-based method from [BCD+20]. In addition, our model allows for dependencies between the input and output differences.

Outline. The main preliminaries are introduced in Section 2. The generic distinguishers are gradually built-up throughout the paper. We begin by exhibiting a new iterated truncated differential and deriving a basic distinguisher for $t^2 - t - 1$ rounds in Section 3. In Section 4, improved truncated differential distinguishers are constructed. We optimize the selection of the input structure using dependencies between the input and output differences, and obtain distinguishers for $t^2 + t - 2$ and t^2 rounds. We show that our 16-round trail for $t = 4$ is optimal using SMT in Section 4.2, and verify the distinguishers experimentally in Section 4.3. Key recovery attacks are discussed in Section 5. Section 6 concludes by discussing the application of our attacks to GMiMC-crf and SM4.

2 Preliminaries

Throughout this paper, we let U be a finite-dimensional vector space over a finite field. Furthermore, let $N = |U|$ denote the cardinality of the set U. That is, $N = p^n$ with p a prime and n a positive integer.

Contracting Feistel ciphers. Contracting Feistel ciphers are a type of generalized unbalanced Feistel structure. As illustrated in Figure 1 for $t = 4$, a Feistel round of a contracting Feistel cipher $\mathsf{R} : U^t \to U^t$ with t branches is defined as $\mathsf{R} : (x_1, x_2, \ldots, x_t) \mapsto (y_1, y_2, \ldots, y_t)$ where

$$y_i = x_{i+1} \text{ for } i = 1, 2, \ldots, t-1\,,$$
$$y_t = x_1 + \mathsf{F}(x_2 + x_3 + \ldots + x_t)\,.$$

Here, F is called the round function of the contracting Feistel cipher and is often key-dependent. The round function F can take various forms. For instance, the round function of SM4 has a SHARK-like structure consisting of an S-box layer followed by a multiplication with an MDS matrix [Dt08]. For GMiMC-crf [AGP+19], $\mathsf{F}(x) = x^3$, assuming U is a finite

field. Since the attacks in this paper are generic and do not exploit the inner structure of the round function and key schedule, we omit further details.

Differentials. For a function $\mathsf{E} : U^t \to U^t$ and input and output differences $a, b \in U^t$, the probability of a differential propagation from a to b through F is defined as (recall that $|U| = N$)

$$\Pr[a \xrightarrow{\mathsf{E}} b] = |\{x \in U^t \mid \mathsf{E}(x+a) - \mathsf{E}(x) = b\}|/N^t\,.$$

The propagation $a \to b$ is called a differential over F, and it gives a distinguishing property if the probability $\Pr[a \to b]$ is significantly larger than $1/N^t$. Roughly speaking, one needs $1/\Pr[a \to b]$ queries to F to distinguish it from a random permutation.

Truncated differentials. An important extension of differential cryptanalysis is the so-called truncated differential attack, first proposed by Knudsen [Knu95]. In the following, we describe the most general form of truncated differentials. Let A and B be subsets of U^t. The probability of the truncated differential for E with input set A and output set B is defined by

$$\Pr[A \xrightarrow{\mathsf{E}} B] = \Pr[\mathsf{F}(\mathbf{x}) - \mathsf{F}(\mathbf{y}) \in B \mid \mathbf{x} - \mathbf{y} \in A]\,,$$

where $\mathbf{x}$ and $\mathbf{y}$ are independent uniform random variables on U^t. Equivalently,

$$\Pr[A \xrightarrow{\mathsf{E}} B] = \frac{1}{|A|} \sum_{a \in A} \Pr[\mathsf{E}(\mathbf{x}+a) - \mathsf{E}(\mathbf{x}) \in B]\,.$$

A truncated differential with $A = B$ will be called iterative or iterated.

Markov ciphers. Directly evaluating the probability of a (truncated) differential over a block cipher is usually not feasible. However, using the iterated structure of most block ciphers and the Markov cipher assumption, one can approximate the probability of differentials by the probability of differential characteristics. Let E_k denote an r round block cipher with key $k = (k_1, \ldots, k_r)$ and round functions $\mathsf{R}^1_{k_1}, \mathsf{R}^2_{k_2}, \ldots, \mathsf{R}^r_{k_r}$. That is,

$$\mathsf{E}_k = \mathsf{R}^r_{k_r} \circ \cdots \circ \mathsf{R}^2_{k_2} \circ \mathsf{R}^1_{k_1}\,.$$

The block cipher $\mathsf{E}_\mathbf{k}$ with a random key $\mathbf{k}$ is called a Markov cipher [LMM91] if for any i and $a, b, x, y \in U^t$,

$$\Pr[\mathsf{R}^i_{\mathbf{k}_i}(x+a) - \mathsf{R}^i_{\mathbf{k}_i}(x) = b] = \Pr[\mathsf{R}^i_{\mathbf{k}_i}(y+a) - \mathsf{R}^i_{\mathbf{k}_i}(y) = b]\,,$$

where the probabilities are only with respect to the random key $\mathbf{k}_i$. This implies that the sequence of intermediate differences for a fixed input pair under $\mathsf{E}_\mathbf{k}$ form a homogeneous Markov chain. Contracting Feistel ciphers such as SM4 and GMiMC-crf satisfy this property *when instantiated with independent round keys.*

A differential characteristic for $\mathsf{E}_\mathbf{k}$ is a sequence of intermediate differences $a_1 \to a_2 \to \cdots \to a_{r+1}$ such that $a_i \to a_{i+1}$ is a differential for $\mathsf{R}_{\mathbf{k}_i}$ for $i = 1, \ldots, r$. If $\mathsf{E}_\mathbf{k}$ is a Markov cipher, then the probability of the differential characteristic $a_1 \to a_{r+1}$ satisfies

$$\Pr[a_1 \to a_2 \to \cdots \to a_{r+1}] = \prod_{i=1}^{r} \Pr[a_i \to a_{i+1}].$$

Note that the above probability is with respect to the random key $\mathbf{k}$. Equivalently, it is equal to the *key-averaged* probability of the differential characteristic. For the sake of completeness, we mention that the key-averaged probability of a differential $a_1 \to a_{r+1}$

for a Markov cipher $\mathsf{E}_\mathbf{k}$ is the sum of the probabilities of all characteristics with input difference a_1 and output difference a_{r+1}:

$$\Pr[a_1 \xrightarrow{\mathsf{E}_\mathbf{k}} a_{r+1}] = \sum_{a_2,\dots,a_r \in U^t} \prod_{i=1}^{r} \Pr[a_i \to a_{i+1}] \,.$$

Input-output dependencies. When dealing with truncated differentials, it is sometimes convenient to use dependencies between input and output differences. A simple example is the property that any input difference from a set A results in *the same* output difference, rather than just any difference in the set A.

A convenient way to describe such properties without leaving the usual framework for truncated differentials from above, is to consider the input-extended cipher $\overline{\mathsf{E}} : U^t \times U^t \to U^t \times U^t$ defined by $(x, y) \mapsto (x, \mathsf{E}(y))$. Indeed, if $\overline{A} = \{(a, a) \mid a \in A\}$ and $B \subseteq U^t \times U^t$, then we have

$$\Pr\left[\overline{A} \xrightarrow{\overline{\mathsf{E}}} B\right] = \Pr[(\mathbf{x} - \mathbf{y}, \mathsf{E}(\mathbf{x}) - \mathsf{E}(\mathbf{y})) \in B \mid \mathbf{x} - \mathbf{y} \in A] \,,$$

with $\mathbf{x}$ and $\mathbf{y}$ uniform random on U^t. The right-hand side above is indeed the desired probability. Ordinary truncated differentials correspond to the case $B = A \times C$ for some output difference set C.

3 Basic Truncated Differential Distinguisher

In Section 3.1, we exhibit a first t-round iterated truncated differential for generic contracting Feistel ciphers and show that it leads to interesting distinguishers. This truncated differential bears similarity to the truncated differential from [BCD+20, §5.2] for the expanding Feistel cipher GMiMC-erf. A comparison with the latter attack and a previous truncated differential attack on GMiMC-crf is given in Section 3.2.

When iterated too many times, the probability of the aforementioned truncated differential trail drops below the probability of the truncated differential for uniform random permutations. However, it can be argued that it should still be possible to obtain a distinguisher as long as enough pairs are available. This observation is used in Section 3.3 to show that the distinguisher from Section 3.1 can cover more rounds.

3.1 An Iterated Truncated Differential Trail

Figure 2 shows an iterated truncated differential $A \to A$ with $A = \{(a, -a, 0, 0) \mid a \in U\}$ for four rounds of a contracting Feistel cipher with $t = 4$ branches. The input difference is represented symbolically on each branch. For instance, the label a corresponds to an arbitrary nonzero input difference.

In the first round, the probability is one since the output difference b of F_i is arbitrary. The probability for the second round is $1/(N-1) \sim 1/N$ on average, assuming that F_{i+1} is a uniform random permutation. Finally, the truncated differences in the third and fourth rounds propagate with probability one since $a + b - a - b = 0$.

A similar trail exists for any number of branches $t \geq 4$. In particular, one can simply set the rightmost $t - 2$ branches to zero. Since the trail in Figure 2 has the same input and output set, it can be iterated. For r divisible by t, we obtain an r round trail with probability $p_\mathsf{trail} = 1/(N-1)^{r/t} \sim 1/N^{r/t}$.

For a random permutation, however, the probability of $A \to A$ is $p_\mathsf{ideal} = (N-1)/(N^t - 1) \sim 1/N^{t-1}$. Hence, if $p_\mathsf{ideal} = o(p_\mathsf{trail})$, one obtains an r-round distinguisher using approximately $1/p_\mathsf{trail} = N^{r/t}$ data. It follows that a t-branch contracting Feistel cipher

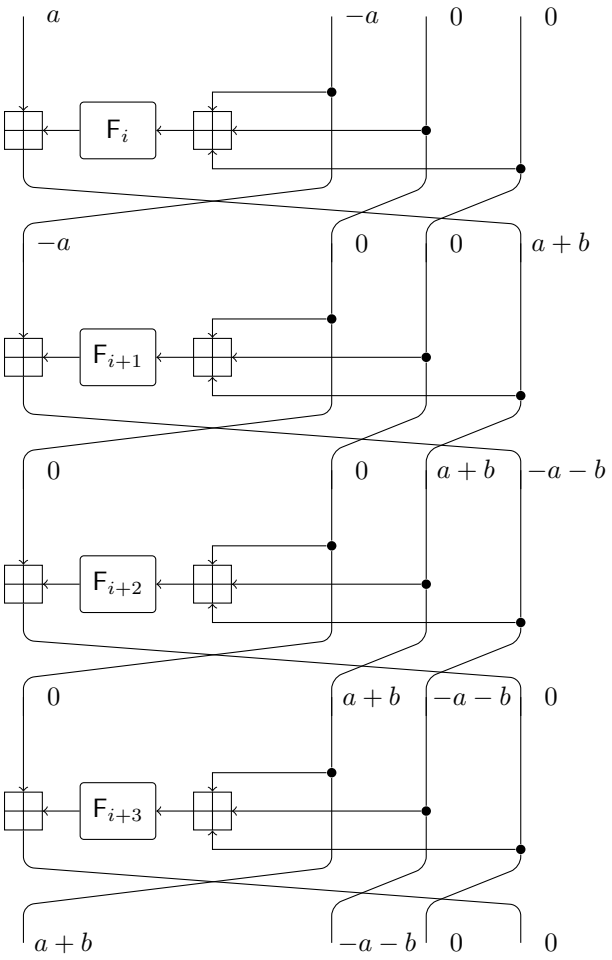

Figure 2: Truncated differential for a contracting generalized Feistel cipher with $t = 4$ branches. The probability of this trail is $1/N$. In characteristic two, the minus signs may be dropped.

must have $r > t^2 - 2t$ rounds to be secure. Furthermore, the attack on $t^2 - 2t$ rounds requires N^{t-2} data.

In fact, the above can be improved by prepending $t - 2$ rounds to the trail as shown in Figure 3 for $t = 4$. Since this modification does not affect p_{real} or p_{ideal}, one obtains a distinguisher on $t^2 - t - 2$ rounds with N^{t-2} data.

A further extension by appending at most $t-2$ rounds to the trail is possible. However, appending s rounds increases p_{ideal} to $(N-1)/(N^{t-s}-1) \sim 1/N^{t-s-1}$. Hence, appending rounds does not lead to an attack on more rounds. Nevertheless, for a smaller number of rounds, appending $t-2$ rounds may lead to a lower data-complexity. Optimizing for the number of rounds, we obtain the following result.

Result 1. *A generic contracting Feistel cipher with t branches and t^2-t-2 rounds can be distinguished from a uniform random permutation with advantage $\Theta(1)$ using N^{t-2} data.*

Application to generic contracting Feistel ciphers. Result 1 implies that the number of rounds of a contracting Feistel cipher must scale quadratically with the number of branches. For a large enough number of branches, this is a significant improvement over the attacks by Patarin *et al.* [PNB06] and Guo *et al.* [GJNS16], who showed that the

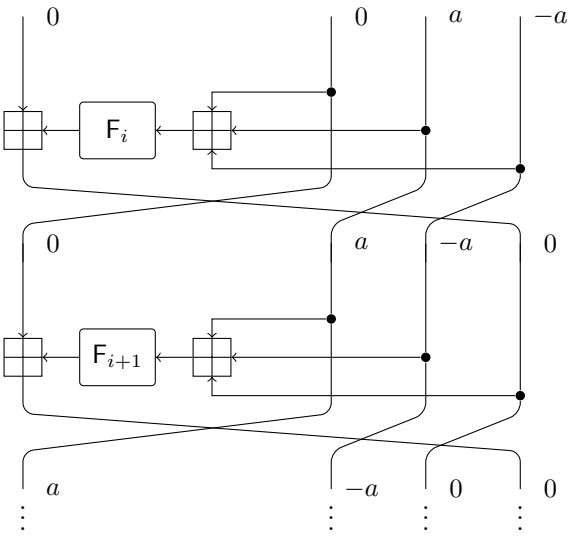

Figure 3: Prepending $t-2$ rounds to the trail from Figure 2 with probability one. In characteristic two, the minus signs may be dropped.

number of rounds must scale linearly with the number of branches. However, note that for the most interesting applications small values of t are of particular importance. Hence, a more detailed comparison is necessary.

The distinguishers of Patarin *et al.* [PNB06] consider a more general form of contracting Feistel ciphers, but cover at most $2t-1$ rounds. Hence, Result 1 improves over this for all $t \geq 4$. Guo *et al.* [GJNS16] describe key-recovery attacks up to $5t-4$ rounds when the key length is $t \log_2 N$ bits. By guessing the last round key, The distinguisher above is easily adapted to a key-recovery attack on t^2-t-1 rounds with time-complexity $\widetilde{O}(N^{t-1})$. Hence, Result 1 improves over the attacks of Guo *et al.* [GJNS16] for $t \geq 6$.

Application to GMiMC-crf. Result 1 yields a full-round attack on some instances of GMiMC-crf [AGP+19] with $t \log_2 N$-bit keys. Indeed, GMiMC-crf has roughly $t(t+3)/2+1$ rounds when t is large compared to $\log_2 N$. To the best of our knowledge, this is the first result violating the security claims of GMiMC-crf – albeit not for the most practically relevant instances. In fact, in this setting, the number of rounds of GMiMC-crf is roughly determined by a truncated differential attack from [AGP+19]. An in-depth analysis of and comparison with the truncated attack from [AGP+19] is given in Section 3.2. The instances of GMiMC-crf with $\log_2 N$-bit keys are less interesting, since they are vulnerable to a simple birthday-bound attack [Bon19]. However, our attacks are also applicable to these instances and would be the best-known attacks if the key-schedule of GMiMC-crf is modified to thwart the attack from [Bon19].

Result 1 leaves significant room for improvements. Importantly, even extensions by a number of rounds linear in t are relevant, since important examples such as SM4 have a small number of branches. A first improvement is the use of input structures. An affine space of dimension d over U contains $N^d(N^d-1)/2$ pairs. This allows reducing the amount of data. For example, for the truncated differential used in Result 1, one could reduce the data-complexity to $2N^{t-3}$ in this manner. However, the number of rounds that can be distinguished does not increase as this is determined by the condition $p_{\mathsf{ideal}} = o(p_{\mathsf{trail}})$. A more careful selection of the truncated differential that takes into account the effect of input structures will lead to further improvements in Section 4.1. A further improvement

is introduced in Section 3.3 below, where a distinguisher with $p_{\mathsf{trail}} \leq p_{\mathsf{ideal}}$ is proposed.

3.2 Comparison with Previous Distinguishers on GMiMC-crf/erf

As mentioned above, the designers of GMiMC-crf also proposed a truncated differential attack on a smaller number of rounds [AGP+19, §4.2.1]. Below, we re-evaluate the number of rounds covered by the latter attack and discuss how the distinguisher from Section 3.1 improves over it. In addition, as discussed in the introduction, the iterated truncated differential from Section 3.1 is similar to the truncated differential distinguisher on GMiMC-erf from [BCD+20, §5.2]. The second paragraph below discusses the relation between this distinguisher and the one from Section 3.1.

GMiMC-crf. The truncated differential trail proposed by the designers of GMiMC-crf [AGP+19, §4.2.1] attempts to minimize the number of active S-boxes. Based on their analysis, the designers conclude that the trail extends to at most $(t+1)\lceil t/2 \times n/(n-1)\rceil \approx t(t+1)/2$ rounds with $n = \log_2 N$. An additional $t-2$ rounds can be appended without decreasing p_{trail} or p_{ideal}. However, as shown below, this estimate is optimistic.

The truncated differential from [AGP+19] is obtained by iterating the $(t+1)$-round truncated differential $B \to B$ with $B = \{(0, 0, \ldots, 0, b, -b) \mid b \in U \setminus \{0\}\}$. The authors analyzed the case $U = \mathbb{F}_2^n$ and obtained a $(t+1)$-round probability of $2/N^2$. Their analysis is not quite generic because it assumes that the round function is differentially 2-uniform. For a generic contracting Feistel cipher, the probability is $\sim 1/N^2$ instead. Making suitable adjustments, the analysis in [AGP+19] suggests that the truncated differential can be iterated $t/2$ times since $(N^2)^{t/2} \leq N^t$. However, this is optimistic because $p_{\mathsf{ideal}} = 1/N^{t-1}$ rather than $1/N^t$. In other words, only $t/2-1$ iterations are possible if we require that $p_{\mathsf{ideal}} = o(p_{\mathsf{trail}})$. Appending $t-2$ rounds without decreasing p_{trail} or increasing p_{ideal} results in a distinguisher on $(t/2-1)\times(t+1)+t-2 = (t+3)(t-2)/2$ rounds. Using a distinguisher with $p_{\mathsf{trail}} \leq p_{\mathsf{ideal}}$ as in Section 3.3 below and by taking advantage of input structures, an extension to $t/2 \times (t+1) + t - 2 = (t+4)(t-1)/2$ rounds might be feasible. However, such a distinguisher requires N^t data and time and is not likely to achieve a high advantage.

For a large number of branches t, the generic distinguisher from Section 3.1 covers twice as many rounds. This is despite the fact that it activates more S-boxes for the same number of rounds. However, the trail in Section 3.1 benefits from the fact that it takes advantage of probabilistic conditions on the output differences of two consecutive round functions instead of just one.

GMiMC-erf. The truncated differential trail from Section 3.1 for GMiMC-crf is similar to that for GMiMC-erf from [BCD+20, §5.5]. GMiMC-erf is an expanding Feistel cipher, where the round function is applied to the first branch and the result is added to all other branches. The analysis in [BCD+20] was specific to GMiMC-erf, but it can be generalized to generic expanding Feistel ciphers. Both truncated differentials are iterative and rely on the condition that the output differences of two consecutive round functions are opposite. The t-round probability is $1/N$ in both cases. However, there are important differences in terms of the input structure size and p_{ideal}.

It is worth noting that expanding and contracting Feistel ciphers are dual constructions. Due to this, a truncated differential on one construction directly yields a multidimensional linear approximation for the other. The average-case duality between multidimensional linear and truncated differential cryptanalysis in turn results in a corresponding truncated differential for the same construction. However, it can be checked that the truncated differentials from Section 3.1 and [BCD+20] are not dual to each other.

In the remainder of this paper, several improvements to the basic truncated differential from Section 3.1 will be introduced. This includes the extension of the distinguisher to

the setting with $p_{\mathsf{trail}} < p_{\mathsf{ideal}}$ in Section 3.3. In Section 4 further improvements will be made, including taking more advantage of input structures and using dependencies between input and output differences. We believe that similar improvements might be useful for expanding Feistel ciphers. However, due to the lack of applications beyond GMiMC-erf, only contracting Feistel ciphers are considered in this paper.

3.3 Extended Distinguisher with $p_{\mathsf{trail}} \leq p_{\mathsf{ideal}}$

Even when the probability of a truncated differential trail is much lower than the ideal probability of the corresponding truncated differential, it is sometimes possible to obtain a distinguisher. Heuristically, the idea is that wrong pairs for a truncated differential trail behave as if they were encrypted under a uniform random permutation. Hence, one can argue that the true probability p_{real} of the truncated differential satisfies the folklore approximation

$$p_{\mathsf{real}} \approx p_{\mathsf{trail}} + p_{\mathsf{ideal}}(1 - p_{\mathsf{trail}}) = p_{\mathsf{ideal}} + p_{\mathsf{trail}}(1 - p_{\mathsf{ideal}}) \approx p_{\mathsf{ideal}} + p_{\mathsf{trail}}\,. \tag{1}$$

That is, one expects slightly more right pairs for the cipher than for a random permutation.

We now consider the data-complexity of a distinguisher with $p_{\mathsf{trail}} \ll p_{\mathsf{ideal}}$, and derive a distinguisher for more than $t^2 - t - 2$ rounds based on exactly the same iterated truncated differential as in Section 3.1. This is possible because, as was just argued, this truncated differential has

$$p_{\mathsf{real}} - p_{\mathsf{ideal}} \approx p_{\mathsf{trail}} \sim 1/N^{r/t}\,. \tag{2}$$

Suppose one encrypts D plaintext pairs with differences in the input set of the truncated differential. After encrypting these pairs under the cipher, we expect to obtain an average number of Dp_{real} pairs with a difference in the output set of the truncated differential. For a random permutation, the expected number of pairs is instead Dp_{ideal}. Moreover, the distribution of the number of right pairs under a random permutation is binomial with variance $p_{\mathsf{ideal}}(1 - p_{\mathsf{ideal}})D \sim p_{\mathsf{ideal}}D$ since $p_{\mathsf{ideal}} \ll 1$. To obtain a distinguisher with advantage $\Theta(1)$, we require that the difference between the means of the real an ideal distribution of the number of valid pairs exceeds the standard deviation of the ideal distribution:

$$D\,(p_{\mathsf{real}} - p_{\mathsf{ideal}}) \gg \sqrt{D\,p_{\mathsf{ideal}}}\,.$$

Rewriting the above, we obtain the estimate

$$D \gg p_{\mathsf{ideal}}/(p_{\mathsf{real}} - p_{\mathsf{ideal}})^2\,. \tag{3}$$

For a more detailed derivation of this result including a proof that this is optimal, we refer the reader to Blondeau and Gérard [BG09].

By Equation 2 and Equation 3, we get $D \gg p_{\mathsf{ideal}}\,N^{2r/t}$ with $p_{\mathsf{ideal}} \sim 1/N^{t-1}$. Hence, if the trail is iterated $t-1$ times (once more than in Section 3.1), we must have $D = N^{t-1}$ pairs. After prepending $t-2$ rounds with probability one, a distinguisher on $t(t-1)+t-2 = t^2-2$ rounds is obtained. In fact, it is possible to improve upon this by appending one round at the end. This increases the ideal probability to approximately $1/N^{t-2}$, so that $t^2 - 1$ rounds can be distinguished using N^t pairs. Using an input structure of size N, these pairs can be obtained from roughly $2N^{t-1}$ plaintexts. Note that iterating the truncated differential t times or appending one more round at the end of the trail is not worthwhile, since that would lead to a data-complexity of N^t.

Result 2. *A generic contracting Feistel cipher with t branches and $t^2 - 1$ rounds can be distinguished from a uniform random permutation with advantage $\Theta(1)$ using N^{t-1} data.*

Compared to Result 1, the distinguisher with $p_{\mathsf{trail}} \ll p_{\mathsf{ideal}}$ covers $t+1$ more rounds. Unlike in Section 3.1, the limiting factor in further improvements is now the amount of pairs which can be obtained from the input space. Indeed, provided that one has a sufficiently large input structure, it would be possible to use more than N^t. However, the trail we considered here had an input structure of size only N. In Section 4.1, truncated differentials that allow for bigger input structures will be introduced.

Finally, note that for $t = 4$ (as for SM4), we now obtain a 15 round distinguisher with a data-complexity of N^3. This may be compared with the 16 round key-recovery attack of Guo *et al.* [GJNS16] with a similar data-complexity. The distinguisher from Result 1 can also be extended to a 16 round key-recovery attack, but it requires N^4 partial decryption operations and hence offers only marginal advantage over exhaustive search.

4 Improved Truncated Differential Distinguishers

This section develops our final truncated differential attacks on generic Feistel ciphers. In Section 4.1, improvements to the distinguisher from Section 3 are obtained by taking into account input structures and by allowing for dependencies between the input and output differences. As a result, we obtain distinguishers for $t-1$ additional rounds with the same-data complexity (Result 3). In Section 4.2, we develop an SMT model to show that (for $t = 4$), these distinguishers are indeed optimal. We report on the experimental verification of our results in Section 4.3.

4.1 Input Structures and Input-Output Dependencies

As discussed at the end of Section 3.3, the number of rounds that can be distinguished using the truncated differential from Section 3 is primarily limited by the dimension of the input space. Indeed, if the dimension d of the input structure is large enough, then the number of pairs used in the attack can exceed N^t while keeping the data- and time-complexity below N^t. In particular, one can obtain up to $N^d(N^d-1)/2 \sim N^{2d}/2$ pairs for each structure of size N^d. A larger dimension d leads to a distinguisher for more rounds, *ceteris paribus*. In principle d can be up to $t-1$, but the trade-off with the probability p_{trail} of the trail as well as the ideal probability p_{ideal} should be kept in mind. Note that when using structures, the time-complexity of the distinguisher is still equal to the data-complexity. Indeed, one can count the number of occurrences of the relevant parts of the output and store them in a table. After sorting, the number of valid pairs can be determined by iterating through the table once.

Iterative truncated differential with larger d. In Figure 4, an iterative truncated differential for $t = 4$ is shown. Whereas the truncated differential from Section 3 had input structures of dimension one, the truncated differential in Figure 4 has $d = 2$. Importantly, this is achieved by allowing dependencies between the input and output differences. Recall from Section 2, section 2, that this can be described formally by considering the input-extended cipher. The probability of the four-round trail in Figure 4 is $\sim 1/N$, and the ideal probability is $p_{\mathsf{ideal}} \sim 1/N^3$.

The trail from Figure 4 can be generalized to t branches by considering the following input difference structure:

$$(a_1, a_2, \ldots, a_{t-2}, b, b) \text{ such that } \sum_{i=1}^{t-2} a_i = -b\,,$$

with $a_1, \ldots, a_{t-2}, b \in U$ not all zero. Like the trail from Section 3, this iterated trail covers r rounds with $p_{\mathsf{trail}} \sim 1/N^{r/t}$ for r a multiple of t. Furthermore, the input structure has dimension $d = t-2$ and $p_{\mathsf{ideal}} \sim 1/N^3$.

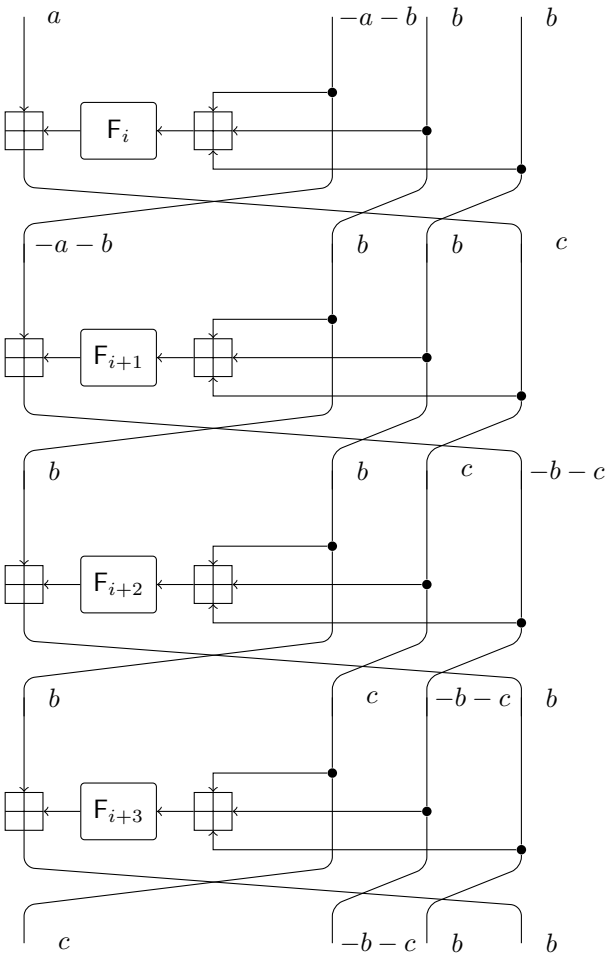

Figure 4: Truncated differential for a contracting generalized Feistel cipher with $t = 4$ branches. The probability of this trail is $1/N$. In characteristic two, the minus signs may be dropped.

There are several ways to extend the iterative trail from above by additional rounds, such as prepending two rounds or appending $t-2$ rounds. However, extending the number of rounds is not necessarily optimal as it may lead to a smaller d or a higher ideal probability p_{ideal}. The next paragraph analyzes the available trade-offs in detail.

Trade-off analysis. Suppose we iterate the trail from Figure 4 m times, covering $m\,t$ rounds. Further assume that when the trail is deterministically extended by s rounds, the input structure dimension is d and let i be an integer such that $p_{\mathsf{ideal}} \sim 1/N^i$. As discussed in Section 3.3 on eq. (3), the number of pairs D required for the attack is then

$$D \sim p_{\mathsf{ideal}}/(p_{\mathsf{real}} - p_{\mathsf{ideal}})^2 = N^{2m-i}\,,$$

since $p_{\mathsf{real}} - p_{\mathsf{ideal}} \sim 1/N^m$. Since the maximum number of pairs that can be obtained is $N^{t-d}\,N^d(N^d-1)/2 \sim N^{t+d}/2$, we must have $D \ll N^{t+d}$. Hence, $2m - i \leq t + d$ or equivalently $m \leq \lfloor (t+d+i)/2 \rfloor$. It follows that the number of rounds r that can be distinguished satisfies

$$\boxed{r \leq t \left\lfloor \frac{t+d+i}{2} \right\rfloor + s\,.} \tag{4}$$

This bound is tight. If $2m \geq 2d+i$, then the corresponding data-complexity is $N^d\,N^{2m-i-2d} = N^{2m-d-i}$. Otherwise, the data-complexity is approximately $N^{m-i/2}$.

We now consider the possible trade-offs for the iterative trail introduced above. Note that it is always possible to prepend two rounds to the trail, without affecting the trail probability or p_{ideal}. If no further rounds are appended, then $i = 3$ as discussed above. It then follows from Equation 4 with $s = 2$ and $d = t-2$ that $r = t\lfloor t + 1/2 \rfloor + 2 = t^2 + 2$ rounds can be distinguished using N^{t-1} data. If instead an additional $t-2$ rounds are appended, then $i = 2$ and $s = t$. Hence, Equation 4 yields a distinguisher on $r = t^2 + t$ rounds with N^t data. This data-complexity is only marginally acceptable. If we choose $m = t-1$ instead of $m = t$, then a distinguisher for t^2 rounds with N^{t-2} data is obtained. It is also possible to append $t-1$ rounds, but this yields $i = 1$ and is not worthwhile.

Alternatively, it is possible to use a slightly larger input structure. Indeed, consider input differences of the following form:

$$(a_1, a_2, \ldots, a_{t-2}, b, b)\,,$$

with $a_1, \ldots, a_{t-2}, b \in U$ not all zero. This is an input structure of dimension $d = t-1$. Importantly, this can be connected to the iterative trail from above with probability $\sim 1/N$. With the input structure above, it is not possible to prepend rounds without decreasing d. If no rounds are appended, then $i = 3$ and $s = 0$. Hence, Equation 4 yields that there is a distinguisher on $t\lfloor t+1 \rfloor = t^2 + t$ rounds with N^t data. Again, the data-complexity of this distinguisher is only marginally acceptable. Choosing $m = t$ instead, a t^2-round distinguisher with lower data-complexity is obtained. However, since $2t < 2(t-1) + 3$, the data-complexity is $N^{t-1.5}$ – higher than for the t^2-round distinguisher from above. Finally, suppose we append $t-2$ rounds such that $i = 2$ and $s = t-2$. By Equation 4, one can then distinguish up to $t\lfloor t + 1/2 \rfloor + t - 2 = t^2 + t - 2$ rounds with N^{t-1} data.

Overview of the best distinguishers. Summarizing the results from the trade-off analysis, we obtain Result 3. Note that these distinguishers cover more rounds than those mentioned in Results 1 and 2. More importantly, they improve over previous generic attacks on contracting Feistel ciphers for any number of branches.

Result 3. *For a generic contracting Feistel cipher with t branches, we have the following distinguishers from a uniform random permutation:*

- $t^2 + t - 2$ *rounds using* N^{t-1} *data,*
- t^2 *rounds using* N^{t-2} *data.*

Each of these distinguishers achieves an advantage of $\Theta(1)$.

The case $t = 4$ is of particular relevance, since the corresponding results yields a distinguisher on 16 rounds of SM4 with 2^{64} data and time and on 18 rounds with 2^{96} data and time. In Section 5, it will be discussed how the distinguishers in Result 3 can be turned into key-recovery attacks on slightly more rounds. It will be demonstrated in Section 6 that this leads to the best-known key-recovery attack on 17-round SM4.

4.2 Modelling Truncated Differentials using SMT

In this section, we show how to model the propagation of truncated differentials through a generic contracting Feistel cipher as a Satisfiability Modulo Theories (SMT) problem. For simplicity, we restrict ourselves to the case with base field $\mathbb{F}_2$. An important feature of the model is that it can be used to find distinguishers with $p_{\mathsf{trail}} \ll p_{\mathsf{ideal}}$. In addition, relations between the input and output variables are accounted for. This is important to verify the distinguishers from Section 4.1. To automate the process of finding truncated differentials by SMT solving, we need to model the truncated differences and the corresponding transition rules by properly defined variables and constraints. Our implementation is based on PyBoolector [NPB14] and is available at https://homes.esat.kuleuven.be/~tbeyne/contracting-feistels.zip.

Variables. For each nonzero truncated difference in our model, it is either a new variable or a linear combination of previous variables. In order to simplify checking linear (in)dependence, we use a bitvector variable to represent the truncated difference on each branch. The zero bitvector represents the zero difference. However, nonzero bitvectors do not correspond to a specific difference and should be thought of as symbolic variables.

Specifically, a bitvector with Hamming weight one represents a free variable, *i.e.* one that is not a linear combination of other variables. Linearly independent truncated differences are represented by distinct bitvectors. Truncated differences that are linear combinations of other differences (with coefficients zero or one, as we work over $\mathbb{F}_2$) can then be represented by a bitvector with Hamming weight two or higher.

The length of the bitvectors is determined by the maximum number of free variables. Specifically, the truncated differences for an r-round t-branch contracting Feistel structure contain at most $r + t$ independent variables, including the input differences and the output differences of the round functions F_i with $i = 1, \ldots, r$. Hence, bitvectors of length $r + t$ are sufficient.

Finally, the model keeps track of the probabilities p_{trail} and p_{ideal} and represents them by their integer weights $\mathsf{wt}(p_{\mathsf{trail}})$ and $\mathsf{wt}(p_{\mathsf{ideal}})$ such that $p_{\mathsf{trail}} \sim 1/N^{\mathsf{wt}(p_{\mathsf{trail}})}$ and $p_{\mathsf{ideal}} \sim 1/N^{\mathsf{wt}(p_{\mathsf{ideal}})}$. In addition, the probability p_i of the truncated differential in round i of the trail has weight $\mathsf{wt}(p_i)$. If a probability is zero, we formally denote its weight by ∞. Within the SMT model, infinite weights are excluded by appropriate constraints.

Constraints. Under the Markov cipher assumption, the average trail probability satisfies $p_{\mathsf{trail}} = \prod_{i=1}^{r} p_i$. Equivalently, the weights must satisfy the constraint

$$\mathsf{wt}(p_{\mathsf{trail}}) = \sum_{i=1}^{r} \mathsf{wt}(p_i)\,.$$

Based on the above, additional constraints for $\mathsf{wt}(p_{\mathsf{trail}}) \neq \infty$ are relatively easy to deduce. To ensure that $p_{\mathsf{trail}} \neq 0$, the first $t - 1$ branches of each output difference must equal the

last $t-1$ branches of the output difference. Furthermore, since the round function is a permutation, the output difference of the round function is zero if and only if the input difference, *i.e.* the exclusive or of the bitvectors representing the rightmost $t-1$ input branches, is zero. The weight $\mathsf{wt}(p_{\mathsf{trail}})$ is then equal to the number of round function output differences that are zero or have Hamming weight at least two (a linear combination of other variables).

If $p_{\mathsf{trail}} < p_{\mathsf{ideal}}$, additional constraints are necessary to avoid trivially invalid trails. In particular, we require that at least one branch of the differences in each round is a linear combinations of the differences in preceding branch differences in that round or the input-branch differences. Linear dependence is modelled recursively.

Finally, we add suitable constraints for $\mathsf{wt}(p_{\mathsf{ideal}})$ by recursively determining the number of output variables that are independent of the input variables and previous output variables.

Proving optimality using SMT. Using the SMT model introduced above, we are able to verify the correctness of the differential distinguishers from Section 4.1. To this end, we place a constraint on the trail weight for fixed values of the input structure dimension and the ideal weight and iteratively increase its value until the problem is found to be satisfiable. Alternatively, it is possible to optimize the overall weight directly, by modelling the data-complexity formula from Section 4.1 within the SMT problem.

For $t = 4$ and $r = 16$, we obtain the best possible truncated differential distinguishers (in terms of data-complexity) for all possible values of the input structure size $d \in \{1, 2, 3\}$ and ideal weight $i \in \{1, 2, 3\}$ within 100 minutes on a standard personal computer. The distinguisher from Result 3 was one of several solutions with data-complexity N^2. No distinguishers with a lower data-complexity were found.

4.3 Experimental Verification

In this section we experimentally verify the generic distinguishers from Result 3 for $t = 4$ and $N = 2^8$. Let $\lambda = p_{\mathsf{ideal}} D$ when the distinguisher (implicitly) generates D pairs. Let $\mathbf{X}$ be a random variable counting the number of right pairs when the distinguisher is evaluated on a random permutation. If the distinguisher uses a threshold value $\tau\sqrt{\lambda}$, then the false-positive rate is

$$P_{\mathsf{F}} = \Pr\left[\mathbf{X} \geq \lambda + \tau\sqrt{\lambda}\right] = \Pr\left[\mathbf{X} \geq (1 + \tau/\sqrt{\lambda})\,\lambda\right].$$

Since P_{F} is the sum of D independent Bernoulli random variables with probability of success p_{ideal}, it follows from the multiplicative Chernoff bound that for $\tau \leq \sqrt{\lambda}$,

$$P_{\mathsf{F}} \leq e^{-\tau^2/3}. \tag{5}$$

Choosing $\tau = 2$, we find that the false-positive rate satisfies $P_{\mathsf{F}} \leq 0.26$. For $\tau = 3/2$, we get $P_{\mathsf{F}} \leq 0.47$.

Figure 5 shows the results of the experiments for $t = 4$ and $N = 2^8$. The estimated success probabilities are shown for $\tau = 2$ (for $r = 16$) or $\tau = 3/2$ (for $r = 18$), *i.e.* a false-positive rate which is at most 26% or 47%. For $r = 16$, each datapoint is based on 1000 evaluations of the attack on a contracting Feistel cipher with uniform random round functions. For $r = 18$, each estimate is based on 100 experiments.

As expected, the success probability gradually increases when more structures are used. The experiments show that achieving a high success probability requires slightly more than N^2 (for $r = 16$) or N^3 (for $r = 18$) data. Note that the success probabilities shown in Figure 5 do not represent the maximal advantage that can be achieved using these distinguishers, since the trade-off between P_{F} and the success probability was not optimized for these experiments.

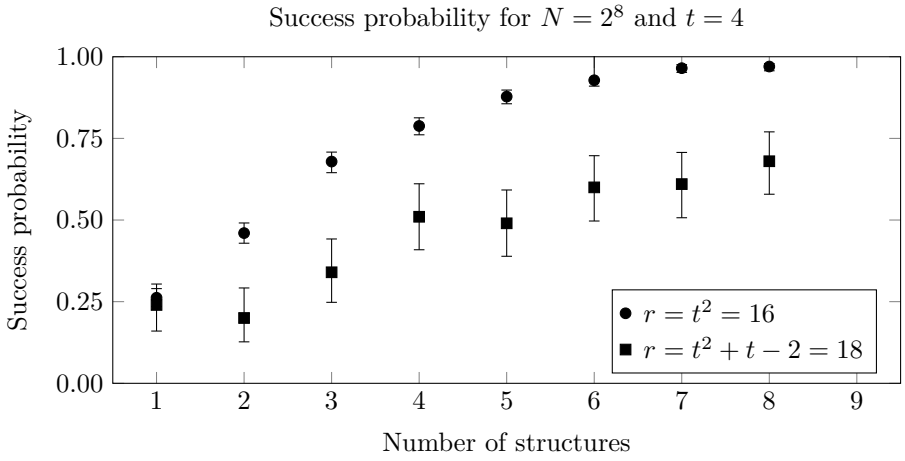

Figure 5: Estimates of the success probability of the distinguishers from Result 3 with $t = 4$ and $\tau = 2$ ($r = 16$) or $\tau = 3/2$ ($r = 18$) as a function of the data-complexity (number of structures). The error bars correspond to 95% confidence intervals computed using the Clopper-Pearson method. Source code to reproduce this figure is available at https://homes.esat.kuleuven.be/~tbeyne/contracting-feistels.zip.

5 Key-Recovery Attacks

If the rounds functions $\mathsf{F}_1, \ldots, \mathsf{F}_r$ of a contracting Feistel cipher are keyed permutations rather than random permutations, then it is of interest to consider key-recovery attacks in addition to distinguishers. For simplicity, assume that the last round key can take N possible values and the total key length is equal to the block size. This is the case for both SM4 and the instances of GMiMC-crf that are not vulnerable to the trivial attack from [Bon19]. The time-complexity of any key-recovery attack can then be at most around N^t encryption operations.

The distinguisher on $t^2 + t - 2$ rounds from Result 3 could in theory be extended to a key-recovery attack on $t^2 + t - 1$ rounds with data-complexity slightly larger (to ensure P_{F} is low enough) than N^{t-1} by guessing the last round key. However, the time-complexity of this attack would be slightly above N^t *partial* encryptions, which is only a marginal improvement over brute-force in the most optimistic case.

More realistically, the t^2-round distinguisher from Result 3 can be extended to a key-recovery attack on $t^2 + 1$ rounds with data-complexity close to N^{t-2} and time-complexity close to N^{t-1}. Again, the attack is based on guessing the last round key and partially decrypting the set of N^{t-2} ciphertexts. Suppose that we wish to reduce the number of candidates for the last round key by a fraction $1/N^{1-\delta}$. By Equation 5 in Section 4.3, this can be achieved by choosing the distinguisher's threshold τ such that $\exp(-\tau^2/3) \leq 1/N^{1-\delta}$. Equivalently,

$$\tau \geq \sqrt{3(1-\delta)\log N}\,,$$

where log denotes the natural logarithm.

By a similar reasoning as in the derivation of Equation (3), the number of required pairs D must satisfy $D(p_{\mathsf{real}} - p_{\mathsf{ideal}}) \geq \tau\sqrt{D p_{\mathsf{ideal}}}$. Since $p_{\mathsf{ideal}} \sim 1/N^2$ and $p_{\mathsf{real}} - p_{\mathsf{ideal}} \sim N^{t-2}$, it follows that

$$D \geq \tau^2 p_{\mathsf{ideal}}/(p_{\mathsf{real}} - p_{\mathsf{ideal}})^2 \approx \tau^2 N^{2t-4}\,.$$

Since the input structures have dimension $t - 2$, the data-complexity becomes $\tau^2 N^{t-2} = 3(1-\delta)\,(\log N)\,N^{t-2}$. The overall time-complexity T of the attack is then

$$T \approx N^{t-1+\delta} + 3\epsilon(1-\delta)(\log N)\,N^{t-1}\,,$$

assuming partial decryption takes ϵ times the time of encryption. The first term is the remaining guessing cost and the second term is due to the partial decryption of the data. To minimize the time-complexity, the parameter $\delta \in [0, 1)$ should be chosen to balance the terms. For instance, if $N = 2^{32}$ and $t = 4$ (the case of SM4), then with $\delta = 0.06140$ we get

$$T \approx 2^{97.96} + 2^{97.96} = 2^{98.96}\,.$$

This estimate assumes $\epsilon = 1/16$. The corresponding data-complexity is $2^{69.96}$.

Alternatively, one could guess more than one round key and rely on a distinguisher for a smaller number of rounds with a lower data-complexity. However, when optimizing for the number of rounds covered by the attack, this is typically not worthwhile because guessing one round key increases the time-complexity by an equal amount as increasing the length of the truncated differential by t rounds. Nevertheless, this approach is interesting in the low-data setting. Optimal trails for a smaller number of rounds can be obtained using the SMT model from Section 4.2, but we leave a detailed analysis of such attacks as future work.

6 Application to GMiMC-crf and SM4

We now briefly consider the impact of the generic distinguishers and key-recovery attacks from Section 4 on GMiMC-crf and SM4. In both cases, we obtain improvements over the state of the art.

GMiMC-crf. As mentioned in Section 3, Result 1 already implies a full-round distinguisher for some instances of GMiMC-crf. Using Result 3, we obtain a distinguisher for $t^2 + t - 2$ rounds. We conclude that for large values of t (relative to $\log_2 N$), the number of rounds of GMiMC-crf must be at least doubled. However, GMiMC-crf is typically instantiated with $t \ll \log_2 N$ so that algebraic attacks are dominant. For these instances of the GMiMC-crf cipher, we do not obtain full-round attacks.

SM4. From Result 3, truncated differential distinguishers on 16- and 18-round SM4 can be obtained using 2^{64} and 2^{96} data respectively. As discussed in Section 5, the generic 16-round distinguisher can be converted into a 17-round key recovery attack with $2^{69.96}$ data and $2^{98.96}$ time by guessing the last round key. We summarize the attacks on SM4 from our paper and compare them to the main attacks from the literature in Table 1.

In terms of the number of rounds covered, the best attacks are differential and linear type covering up to 24-round SM4. However, those attacks require a large amount of data and time. For instance, the 24-round linear attack requires 2^{127} data and 2^{127} time (as measured in arithmetic operations), which is close to the full codebook and the cost of a brute-force key search. Previous attacks on SM4 aiming at lower data- and time-complexity were presented by Guo *et al.* [GJNS16], who give a 16-round key-recovery attack with data and time complexity of 2^{99} using a generic meet-in-the-middle approach. Our 16-round truncated differential distinguisher only requires 2^{64} data, which significantly improves over their attack. Our 17-round key-recovery attack has a similar time-complexity, but a much lower data-complexity.

As far as we know, there is no direct analysis of differential or linear attacks on 16- or 17-round SM4. To make a reasonable comparison, we consider previous differential and linear attacks with a reduced number of rounds. We claim that our 17-round key-recovery attack improves over reduced-round variants of previous work, for the same or similar data-complexity. This claim is motivated by the analysis below. For brevity, we assume the reader is familiar with previous attacks on SM4.

The differential attack from Zhao *et al.* [ZLW18] is very similar to that of Su *et al.* [SWZ11], so the latter will be used for reference below. Both attacks are based on

Table 1: An overview of attacks on the SM4 block cipher. Attacks marked by † are distinguishers, the others are key-recovery attacks.

Attack Type	Rounds	Data	Time	Reference
Differential	12	2^{67}	2^{67}	[SWZ11]†
	21	2^{118}	2^{127}	[ZZW08a]
	22	2^{117}	2^{112}	[ZWFS09]
	23	2^{118}	2^{127}	[SWZ11]
Multiple differential	21	2^{104}	2^{114}	[SG16]
	23	2^{114}	2^{127}	[ZLW18]
Linear	22	2^{117}	2^{112}	[ER09]
	23	2^{120}	2^{122}	[LLWW17]
	24	2^{127}	2^{127}	[LLWW17]
Multiple linear	22	2^{112}	2^{124}	[LGZ10]
	23	2^{127}	2^{127}	[CN11]
Multidimensional linear	23	2^{123}	2^{123}	[LC14]
Boomerang	18	2^{120}	2^{117}	[KKHS08]
Rectangle	16	2^{125}	2^{116}	[ZZW08b]
	18	2^{124}	2^{113}	[KKHS08]
	18	2^{127}	2^{104}	[KWX13]
Impossible differential	16	2^{105}	2^{107}	[Lu08]
	16	2^{117}	2^{132}	[TD08]
	17	2^{117}	2^{132}	[Wan10]
	18	2^{117}	2^{132}	[SWX12]
Meet-in-the-middle	16	2^{99}	2^{99}	[GJNS16]
Truncated differential	16	2^{64}	2^{64}	**Ours**†
	17	2^{70}	2^{99}	**Ours**
	18	2^{96}	2^{96}	**Ours**†

multiple 19-round differentials with the same output difference. The key-recovery appends four rounds. If the 19 round differentials are restricted to 12 rounds, they have probabilities between 2^{-84} and 2^{-82}. As each structure of 2^{33} plaintexts contains 2^{46} pairs, the resulting data-complexity would be around 2^{70}. However, following [SWZ11, §5.1], appending five rounds for the key-recovery attack would have a time-complexity larger than 2^{99}, in particular because there are very little conditions that can be used to filter pairs in the last round.

More generally, we cannot use any known 13-round differentials because their probability is too low. There exist other 12-round characteristics with higher probability (optimally 2^{-67}, according to [SWZ11]), but the key-recovery heavily depends on the structure of the output differences so the analysis from [SWZ11, ZLW18] is then not directly applicable. In any case, a five-round extension by key-recovery with a time-complexity below 2^{99} is questionable.

The other attacks in Table 1 covering more than 18 rounds are linear attacks. Liu *et al.* [LLWW17] propose to use a three-round iterative approximation with absolute correlation 2^{-3r} for r rounds. For 19 rounds this gives an absolute correlation of 2^{-57}, and key-recovery extends this by four rounds. To set up a round-reduced variant of this attack with $\leq 2^{70}$ data, the approximation can be extended to at most 11 rounds (absolute correlation 2^{-33}). However, the key-recovery should then cover 6 rounds, which is not

realistic since 80 bits already have to be guessed for just four rounds.

The work by Liu *et al.* [LC14] is a multidimensional linear attack, but it only uses 25 linear approximations (extended to 64 in order to apply a multidimensional analysis) and their absolute correlations are lower than those from [LLWW17]. The key-recovery appends four rounds and extending this would drive up the time-complexity even more.

Cho and Nyberg [CN11] rely on the 5-round iterative approximations from [ER09]. These have absolute correlation $2^{-18.4}$ in the last two rounds. Hence, for 13 rounds, the absolute correlation would be $2^{-36.8}$. This gives a data-complexity of around $2^{73.6}$. Using multiple approximations as in [CN11], a rough estimate suggests a data-complexity similar to that of our 17-round key-recovery attack. However, this improvement will only be achieved if some internal roundkey bits are guessed (signs of the correlations must be guessed). Due to this, the key-recovery strategy of [CN11] only covers three rounds. In particular, they guess 88 key bits from the initial and final rounds as well as 34 internal roundkey bits. Hence, only a 16 round key-recovery attack is obtained and with a time-complexity above 2^{99}.

References

[AGP+19] Martin R. Albrecht, Lorenzo Grassi, Léo Perrin, Sebastian Ramacher, Christian Rechberger, Dragos Rotaru, Arnab Roy, and Markus Schofnegger. Feistel structures for MPC, and more. In Kazue Sako, Steve Schneider, and Peter Y. A. Ryan, editors, *ESORICS 2019, Part II*, volume 11736 of *LNCS*, pages 151–171. Springer, Heidelberg, September 2019.

[BCD+20] Tim Beyne, Anne Canteaut, Itai Dinur, Maria Eichlseder, Gregor Leander, Gaëtan Leurent, María Naya-Plasencia, Léo Perrin, Yu Sasaki, Yosuke Todo, and Friedrich Wiemer. Out of oddity - new cryptanalytic techniques against symmetric primitives optimized for integrity proof systems. In Daniele Micciancio and Thomas Ristenpart, editors, *CRYPTO 2020, Part III*, volume 12172 of *LNCS*, pages 299–328. Springer, Heidelberg, August 2020.

[BG09] Céline Blondeau and Benoît Gérard. On the data complexity of statistical attacks against block ciphers (full version). Cryptology ePrint Archive, Report 2009/064, 2009. https://eprint.iacr.org/2009/064.

[Bon19] Xavier Bonnetain. Collisions on Feistel-MiMC and univariate GMiMC. Cryptology ePrint Archive, Report 2019/951, 2019. https://eprint.iacr.org/2019/951.

[BS91] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems. In Alfred J. Menezes and Scott A. Vanstone, editors, *CRYPTO'90*, volume 537 of *LNCS*, pages 2–21. Springer, Heidelberg, August 1991.

[CN11] Joo Yeon Cho and Kaisa Nyberg. Improved linear cryptanalysis of SMS4 block cipher. In *Workshop Record of SKEW 2011*, 2011.

[Dt08] Whitfield Diffie and George Ledin (translators). SMS4 encryption algorithm for wireless networks. Cryptology ePrint Archive, Report 2008/329, 2008. https://eprint.iacr.org/2008/329.

[ER09] Jonathan Etrog and Matthew J. B. Robshaw. The cryptanalysis of reduced-round SMS4. In Roberto Maria Avanzi, Liam Keliher, and Francesco Sica, editors, *SAC 2008*, volume 5381 of *LNCS*, pages 51–65. Springer, Heidelberg, August 2009.

[GJNS16] Jian Guo, Jérémy Jean, Ivica Nikolic, and Yu Sasaki. Meet-in-the-middle attacks on classes of contracting and expanding Feistel constructions. *IACR Trans. Symm. Cryptol.*, 2016(2):307–337, 2016. https://tosc.iacr.org/index.php/ToSC/article/view/576.

[KKHS08] Taehyun Kim, Jongsung Kim, Seokhie Hong, and Jaechul Sung. Linear and differential cryptanalysis of reduced SMS4 block cipher. Cryptology ePrint Archive, Report 2008/281, 2008. https://eprint.iacr.org/2008/281.

[Knu95] Lars R. Knudsen. Truncated and higher order differentials. In Bart Preneel, editor, *FSE'94*, volume 1008 of *LNCS*, pages 196–211. Springer, Heidelberg, December 1995.

[KWX13] Xianglong Kong, Wei Wang, and Qiuliang Xu. Improved rectangle attack on SMS4 reduced to 18 rounds. In *Ninth International Conference on Computational Intelligence and Security, CIS 2013, Emei Mountain, Sichan Province, China, December 14-15, 2013*, pages 575–578. IEEE Computer Society, 2013.

[LC14] Mingjie Liu and Jiazhe Chen. Improved linear attacks on the chinese block cipher standard. *J. Comput. Sci. Technol.*, 29(6):1123–1133, 2014.

[LGZ10] Zhiqiang Liu, Dawu Gu, and Jing Zhang. Multiple linear cryptanalysis of reduced-round SMS4 block cipher. *Chinese Journal of Electronics*, 19-3:389–393, 2010.

[LLWW17] Yu Liu, Huicong Liang, Wei Wang, and Meiqin Wang. New linear cryptanalysis of chinese commercial block cipher standard SM4. *Secur. Commun. Networks*, 2017:1461520:1–1461520:10, 2017.

[LMM91] Xuejia Lai, James L. Massey, and Sean Murphy. Markov ciphers and differential cryptanalysis. In Donald W. Davies, editor, *EUROCRYPT'91*, volume 547 of *LNCS*, pages 17–38. Springer, Heidelberg, April 1991.

[LR86] Michael Luby and Charles Rackoff. How to construct pseudo-random permutations from pseudo-random functions (abstract). In Hugh C. Williams, editor, *CRYPTO'85*, volume 218 of *LNCS*, page 447. Springer, Heidelberg, August 1986.

[Lu08] Jiqiang Lu. Attacking reduced-round versions of the SMS4 block cipher in the Chinese WAPI standard. In Sihan Qing, Hideki Imai, and Guilin Wang, editors, *ICICS 07*, volume 4861 of *LNCS*, pages 306–318. Springer, Heidelberg, December 2008.

[NPB14] Aina Niemetz, Mathias Preiner, and Armin Biere. Boolector 2.0. *J. Satisf. Boolean Model. Comput.*, 9(1):53–58, 2014.

[Nyb96] Kaisa Nyberg. Generalized Feistel networks. In Kwangjo Kim and Tsutomu Matsumoto, editors, *ASIACRYPT'96*, volume 1163 of *LNCS*, pages 91–104. Springer, Heidelberg, November 1996.

[Pat04] Jacques Patarin. Security of random Feistel schemes with 5 or more rounds. In Matthew Franklin, editor, *CRYPTO 2004*, volume 3152 of *LNCS*, pages 106–122. Springer, Heidelberg, August 2004.

[PNB06] Jacques Patarin, Valérie Nachef, and Côme Berbain. Generic attacks on unbalanced Feistel schemes with contracting functions. In Xuejia Lai and Kefei Chen, editors, *ASIACRYPT 2006*, volume 4284 of *LNCS*, pages 396–411. Springer, Heidelberg, December 2006.

[SG16] Heyingxiu Song and Haiying Gao. Multiple differential attack on 21-round sms4 (in Chinese). *Journal of Cryptologic Research*, 2016(3(6)):584–595, 2016.

[SK96] Bruce Schneier and John Kelsey. Unbalanced Feistel networks and block cipher design. In Dieter Gollmann, editor, *FSE'96*, volume 1039 of *LNCS*, pages 121–144. Springer, Heidelberg, February 1996.

[SWX12] Tao Shi, Wei Wang, and Qiuliang Xu. Improved impossible differential cryptanalysis of SMS4. In *Eighth International Conference on Computational Intelligence and Security, CIS 2012, Guangzhou, China, November 17-18, 2012*, pages 492–496. IEEE Computer Society, 2012.

[SWZ11] Bozhan Su, Wenling Wu, and Wentao Zhang. Security of the SMS4 block cipher against differential cryptanalysis. *J. Comput. Sci. Technol.*, 26(1):130–138, 2011.

[TD08] Deniz Toz and Orr Dunkelman. Analysis of two attacks on reduced-round versions of the SMS4. In Liqun Chen, Mark Dermot Ryan, and Guilin Wang, editors, *ICICS 08*, volume 5308 of *LNCS*, pages 141–156. Springer, Heidelberg, October 2008.

[Wan10] Gaoli Wang. Improved impossible differential crytanalysis on SMS4. In *Communications and Intelligence Information Security (ICCIIS) 2010*, pages 105–108, 2010.

[YPL11] Aaram Yun, Je Hong Park, and Jooyoung Lee. On Lai–Massey and quasi-Feistel ciphers. *Designs, Codes and Cryptography*, 58(1):45–72, 2011.

[ZLW18] Yanmin Zhao, Yu Liu, and Meiqin Wang. Improved differential attack on 23-round SMS4 (in Chinese). *Journal of Software*, 2018(29(9)):2821–2828, 2018.

[ZMI90] Yuliang Zheng, Tsutomu Matsumoto, and Hideki Imai. On the construction of block ciphers provably secure and not relying on any unproved hypotheses. In Gilles Brassard, editor, *Advances in Cryptology — CRYPTO' 89 Proceedings*, pages 461–480, New York, NY, 1990. Springer New York.

[ZWFS09] Wentao Zhang, Wenling Wu, Dengguo Feng, and Bozhan Su. Some new observations on the SMS4 block cipher in the chinese WAPI standard. In Feng Bao, Hui Li, and Guilin Wang, editors, *Information Security Practice and Experience, 5th International Conference, ISPEC 2009, Xi'an, China, April 13-15, 2009, Proceedings*, volume 5451 of *Lecture Notes in Computer Science*, pages 324–335. Springer, 2009.

[ZZW08a] Lei Zhang, Wentao Zhang, and Wenling Wu. Cryptanalysis of reduced-round SMS4 block cipher. In Yi Mu, Willy Susilo, and Jennifer Seberry, editors, *ACISP 08*, volume 5107 of *LNCS*, pages 216–229. Springer, Heidelberg, July 2008.

[ZZW08b] Lei Zhang, Wentao Zhang, and Wenling Wu. Cryptanalysis of reduced-round SMS4 block cipher. In Yi Mu, Willy Susilo, and Jennifer Seberry, editors, *Information Security and Privacy, 13th Australasian Conference, ACISP 2008, Wollongong, Australia, July 7-9, 2008, Proceedings*, volume 5107 of *Lecture Notes in Computer Science*, pages 216–229. Springer, 2008.