Superposition Meet-in-the-Middle Attacks: Updates on Fundamental Security of AES-like Hashing

Zhenzhen Bao[2,3], Jian Guo[2], Danping Shi[1,4(✉)], and Yi Tu[2]

[1] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China shidanping@iie.ac.cn
[2] School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
baozhenzhen10@gmail.com, guojian@ntu.edu.sg, TUYI0002@e.ntu.edu.sg
[3] Institute for Network Sciences and Cyberspace, BNRist, Tsinghua University, Beijing, China
[4] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

Abstract. The Meet-in-the-Middle approach is one of the most powerful cryptanalysis techniques, demonstrated by its applications in preimage attacks on the full MD4, MD5, Tiger, HAVAL, and Haraka-512 v2 hash functions, and key recovery of the full block cipher KTANTAN. The success relies on the separation of a primitive into two independent chunks, where each active cell of the state is used to represent only one chunk or is otherwise considered unusable once mixed. We observe that some of such cells are linearly mixed and can be as useful as the independent ones. This leads to the introduction of superposition states and a whole suite of accompanied techniques, which we incorporate into the MILP-based search framework proposed by Bao *et al.* at EUROCRYPT 2021 and Dong *et al.* at CRYPTO 2021, and find applications on a wide range of AES-like hash functions and block ciphers.

Keywords: Whirlpool, Grøstl, AES hashing modes, MITM, MILP

1 Introduction

Hash function is a function mapping a document of arbitrary length into a short fixed-length digest. For a cryptographically secure hash function, it should fulfill three basic security requirements: collision resistance, preimage resistance, and second-preimage resistance. In this paper, we focus on the security notion of preimage and collision resistance, i.e., it should be computationally difficult to invert the function or find two inputs map to the same digest. Specially, for an ideal hash function H with n-bit digest and a target T given at random, it should cost no less than 2^n compression function evaluations to find an input x such that $H(x) = T$. Preimage attack refers to an algorithm achieving this in lesser evaluations. Collision attack refers to an algorithm finding different x and x' such that $H(x) = H(x')$ with lesser evaluations than birthday attack, *i.e.*, $2^{n/2}$ computations.

Traditionally, there are two common methods to construct cryptographic hash functions. One is to convert from block ciphers through mode of operations, and the other is to build from scratch. There are 12 secure PGV modes [23], which enjoy the proof of security reduction of the hash function to the underlying block cipher. This method is especially useful when a block cipher like AES [10] has long-standing security against intensive cryptanalysis. When it is already implemented for other purposes like encryption, the same implementation can be re-used to construct a hash function by implementing the additional mode only. In this way it also leads to performance merits. Particularly, hash function constructed from AES through PGV modes are called AES hashing, and they have been standardized by Zigbee [1] and also suggested by ISO/IEC [17]. Due to the well understood security and high performance, many dedicated block ciphers and hash functions built from scratch follow similar design strategy by using an AES-like round function, such as Whirlpool [7] and Grøstl [13].

The Meet-in-the-Middle (MITM) attack has a long history and an important role in various cryptanalysis on various primitives. It was introduced to preimage attacks on hash functions by Aumasson *et al.* [3] and Sasaki *et al.* [25] in 2008. Since then, MITM preimage attack has shown its power on many MD/SHA families of hash functions. That includes the full versions of MD4 [15], MD5 [27], Tiger [15], and HAVAL [25], as well as lightweight block cipher KTANTAN [8].

The basic MITM idea is to find ways to split the cipher into two computational chunks and find the so-called neutral bits from each side, independent of the computation of the state of the other side. Hence, the two chunks can be computed independently but end up at a common state, where previously independent computations are finally pairwise matched. The critical point is that the independence between the two chunks allows their results to be pairwise matchable, enabling this MITM procedure to require fewer computations than a trivial enumeration.

In 2011, MITM was introduced by Sasaki [24] for the first time to preimage attack on AES hashing, invalidating the preimage resistance of the 7-round reduced version. To avoid dealing with the key schedule, the key value of AES was pre-set to a constant, and hence the same number of rounds was attacked for AES hashing based on all three versions of AES (AES-128, AES-192, and AES-256). In 2019, the attack was revisited in [4], and it was found that the degree of freedoms from the key values can be utilized for at least one side of the computation. This observation led to the attack complexity improvements over 7-round AES-128 hashing, and increased the number of attacked rounds from 7 to 8 for the AES-hashing based on AES-192 and AES-256. In 2020, the MITM preimage attack was introduced to meet a popular automatic tool, the Mixed-Integer-Linear-Programming (MILP) [5]. This match with MILP enables the previously invented enhancements for MITM, *e.g.*, the initial structure, the selection and cancellation of neutral bits, to evolve to a more generalized form. The MILP models characterizing the generalized formalization of MITM, produced many improved preimage attacks on AES hashing, penetrated 8-round

AES-128 hashing and full rounds of Haraka v2-512. In 2021, MITM-MILP models find more applications not only in preimage attacks but also in key-recovery and collision attacks on AES-like ciphers [11].

A common practice in all these MITM attacks is the byte-oriented decomposition of states, *i.e.*, each useful byte carries the influence of neutral bits from at most one direction. Once influences of neutral bits from both directions reach the same byte, this byte will be considered unusable for the either chunk. An immediate consequence of this practice is that the information from unusable bytes is lost, and any byte in the subsequent round computed from it becomes unusable too.

1.1 Our Contribution

Superposition Meet-in-the-Middle Attack Framework. In this work, different from the byte-oriented decomposition, we propose Superposition MITM attack framework with the help of Superposition State (SupP), under which every byte is viewed as a combination of two virtual bytes separately representing the influence of the neutral bits from one direction each. It becomes obvious to see that, the advantage of this representation is the preservation of linearity, *i.e.*, any state which is a linear combination of neutral bits from both directions will be kept so when linear operations such as `MixColumns` and `AddRoundKey` are applied. SupP opens up the possibility of new local collisions between/within the encryption and key states, thus maximizes the chances of canceling impacts on independence and enables new MITM attack configurations. Enabled by SupP, a suite of techniques is added into the MILP-based search framework proposed by Bao *et al.* [5] and Dong *et al.* [11], which greatly enlarged the solution space and led to rich results on AES-like ciphers.

Guess-and-Determine (GnD) is a popular technique and has countless applications in cryptanalysis against symmetric-key primitives such as stream ciphers [31] and block ciphers [9]. The basic idea is that, in the process of some attack, the gain of guessing some state or key bits is higher than the price, *i.e.*, the guess itself comes with a probability p then the attack needs to repeat at least $1/p$ times in order to have a correct guess. This technique has also been used in the MITM preimage attacks [15, 28, 30], where the guess allows further computation of more state/message bits which help either extend the attack to more rounds or lower the time complexities. It is noted that GnD has never been incorporated into the MILP models for the MITM attacks in [5, 11].

Bi-direction Attribute-Propagation and Cancellation (BiDir). In previous MITM attacks [4, 5, 11, 24, 28, 30], each computation chunk propagates towards a single direction. Although the idea of adding constraints to some neutral bits in order to cancel their impact on the opposite computation has already appeared, such cancellations are only allowed in one direction in each chunk. In this work, we allow cancellation between neutral bytes in both directions. This led us to attack configurations with lower time complexities.

Multiple Ways of `AddRoundKey` (MulAK). The identical encryption and key-schedule of Whirlpool allows the `AddRoundKey` to be moved around the `MixColumns` using an equivalent key state already involved in the key-schedule. We add the flexibility of key-addition at different positions in each round into the model, and make it possible to save the double consumption of freedom degrees between the encryption and key-schedule. This simple yet efficient strategy makes it possible to attack one more round on reduced Whirlpool.

Application Results. The superposition MITM attack framework aided by the additional techniques shows its effectiveness when applied to the popular AES-like hashing Whirlpool (an ISO/IEC standard), Grøstl (a finalist of the SHA-3 competition), AES-hashing modes (widely used, *e.g.*, AES-MMO in Zigbee protocols), and tweakable block cipher SKINNY (in its way to be included in ISO/IEC 18033). Broad improvements upon the previous best results on preimage, collision, and key-recovery attacks were obtained. The updates and a comparison with the state-of-the-art results in literature are summarized in Table 1.

For Whirlpool, the preimage resistance of its 7-round reduced version out of the total 10 rounds gets challenged for the first time, a decade after MITM itself challenging its 5-round [24] for the first time, or GnD joining MITM challenging its 6-rounds [28]. Meanwhile, the complexities of preimage attacks on the 5- and 6-round reduced versions are significantly improved. The new attack on 7-round Whirlpool relies on multiple modeling enhancements, including GnD, BiDir, and MulAK. Noticeably, in terms of preimage resistance, the presented attack reduces one round security margin out of the remaining four for this important target.

For Grøstl, there are two main instances, Grøstl-256 and Grøstl-512, named after the size of their digests. The preimage resistance of their longest attacked variants, 6-round Grøstl-256 and 8-round Grøstl-512, get updated eight years after [32]. Improvements with larger complexity reductions are also found on variants with lesser rounds. The improved attacks are achieved by allowing BiDir, in addition to the GnD. Such improvement is not possible using only MITM-GnD, and MITM-GnD is already considered in [22, 32].

In addition, the MILP models for searching preimage attacks can be directly transformed into models searching for collision attacks on hash functions and key-recovery attacks on block ciphers. With the superposition MITM-GnD models, immediate improvements were obtained on many more targets: For Grøstl-256, the collision resistance of its output transformation's longest attacked version, the 6-round, gets updated; For AES-hashing, the collision resistance of hash functions based on AES-128 reduced to 7-round get challenged, more than a decade after rebound attacks challenging its 6-round [14, 19], or two years after quantum attacks challenging its 7-round version [16]. For SKINNY, the security in terms of key-recovery attack in the single-key setting of the longest attacked version, 23-round SKINNY-n-$3n$, gets updated. Remarkably, data complexity is drastically reduced, *e.g.*, from 2^{52} to 2^{28} for SKINNY-64-192.

Organization. Section 2 briefly introduces the AES-like hashing and MITM preimage attacks. Section 3 describes the details of our enhanced MILP model. The application to Whirlpool and Grøstl is given in Section 4 and 5, respectively. Section 6 briefly introduces the applications to collision and key-recovery attack. Please refer to the full version [6] for more details. Relevant source codes can be found via https://github.com/MITM-AES-like.

2 AES-like Hashing and MITM Preimage Attacks

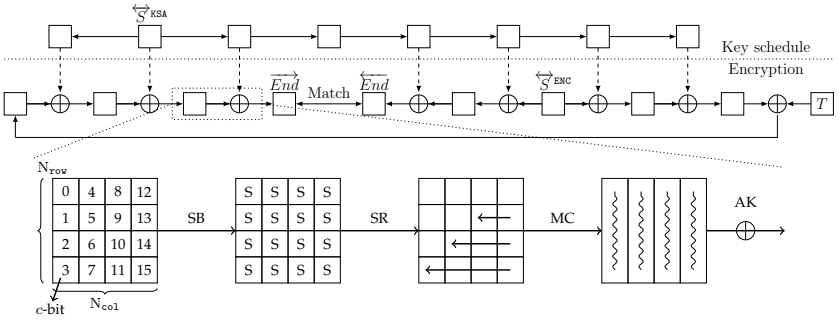


Fig. 1: Overview of AES-like hashing [5]

The AES-like hashing in our context refers to those hash functions whose compression function (CF) or output transformation (OT) uses an AES-like round function as depicted in Figure 1, where the state can be viewed as a $\mathtt{N_{row}} \times \mathtt{N_{col}}$ matrix of c-bit cells. There are 4 general operations in order:

- `SubBytes` (SB) applies a non-linear substitution-box operation to each cell.
- `ShiftRows` (SR) cyclically shifts each row by a pre-defined number of positions.
- `MixColumns` (MC) mixes every column, *e.g.*, by multiplying an (MDS) matrix.
- `AddRoundKey` (AK) adds the round key (or round message).

For some designs, the SR and MC may work on the transpose of the matrix, *i.e.*, SR on columns and MC on rows; the SR can be a cell-permutation.

The MITM Preimage Attacks generally cut the entire encryption process into two chunks (*forward* and *backward chunks*), and there are few *neutral bits* from each side, so that the computation of each chunk can be done independently starting from the cutting point (*starting point*). The two computations end at a common intermediate (partial) state (*matching point*), and the computed results from the two sides can be pairwise matched via a linear relation.

Since MITM has been applied to preimage attacks on MD4, MD5, and HAVAL [3, 20,25,26], it has been developed further and applied to preimage attacks on many

Table 1: Updated results on (pseudo-) preimage attacks

(Pseudo-) Preimage						
Cipher (Target)	#R	Time-1	Time-2	$(\overrightarrow{d_b}, \overleftarrow{d_r}, \overrightarrow{\leftarrow m}, \overrightarrow{d_{g_b}}, \overleftarrow{d_{g_r}})$	Critical Tech.	Ref.
Whirlpool (Hash)	5/10	2^{416}	2^{448}	$(16, 12, 16, 0, 0)$	Dedicated	[28]
	5/10	$\mathbf{2^{352}}$	$\mathbf{2^{433}}$	$(20, 20, 20, 0, 0)$	MILP, BiDir, MulAK	*Fig. 13
	6/10	2^{448}	2^{481}	$(32, 8, 32, 0, 24)$	Dedicated, GnD	[28]
	6/10	$\mathbf{2^{440}}$	$\mathbf{2^{477}}$	$(9, 24, 24, 15, 0)$	MILP, GnD	*Fig. 12
	7/10	$\mathbf{2^{480}}$	$\mathbf{2^{497}}$	$(16, 4, 16, 0, 12)$	MILP, GnD, MulAK	*Fig. 3, 11
Grøstl-256 (CF+OT)	5/10	2^{192}	$2^{234.67}$	$(8, 8, 8, 0, 0)$	Dedicated	‡ [22, 32]
	5/10	$\mathbf{2^{184}}$	$\mathbf{2^{232}}$	$(9, 9, 16, 0, 0)$	MILP, BiDir	*Fig. 14, 15
	6/10	2^{240}	2^{252}	$(8, 2, 8, 0, 6)$	Dedicated, GnD	‡ [22, 32]
	6/10	$\mathbf{2^{224}}$	$\mathbf{2^{245.33}}$	$(4, 20, 16, 12, 0)$	MILP, GnD, BiDir	*Fig. 5, 6
Grøstl-512 (CF+OT)	7/14	2^{416}	2^{480}	$(19, 12, 19, 0, 7)$	MILP, GnD, BiDir	*Fig. 16, 17
	8/14	2^{472}	2^{504}	$(10, 10, 18, 5, 5)$	Dedicated	† [32]
	8/14	2^{472}	$\mathbf{2^{500}}$	$(9, 5, 10, 0, 4)$	MILP, GnD, BiDir	*Fig. 7, 8
AES-192 Hashing	9/12	2^{120}	2^{125}	$(1, 1, 1, 0, 0)$	MILP	[5]
	9/12	$\mathbf{2^{112}}$	$\mathbf{2^{121}}$	$(2, 2, 2, 0, 0)$	MILP, SupP, BiDir	*Fig. 18
Kiasu-BC Hashing	8/10	2^{120}	2^{123}	$(1, 4, 4, 0, 0)$	Dedicated	[4]
	9/10	$\mathbf{2^{120}}$	$\mathbf{2^{125}}$	$(1, 1, 1, 0, 0)$	MILP, SupP, BiDir	*Fig. 19

(Free-start) Collision						
Cipher (Target)	#R	Time	Mem	Setting & Type	Critical Tech.	Ref.
Grøstl-256 (OT)	6/10	2^{124}	2^{124}	classic collision	MILP	[11]
	6/10	$\mathbf{2^{116}}$	$\mathbf{2^{116}}$	classic collision	MILP, BiDir	*Fig. 20
Grøstl-512 (OT)	8/14	2^{248}	2^{248}	classic collision	MILP	[11]
	8/14	$\mathbf{2^{244}}$	$\mathbf{2^{244}}$	classic collision	MILP, BiDir	*Fig. 21
AES-128 Hashing	6/10	2^{56}	2^{32}	classic collision	Dedicated	[14, 19]
	7/10	$2^{42.5}$	(2^{48})	quantum collision	Dedicated	[16]
	7/10	$\mathbf{2^{56}}$	$\mathbf{2^{56}}$	classic free-start	MILP, BiDir	*Fig. 22

Key-recovery						
Cipher (Target)	#R	Time	Mem	Data	Critical Tech.	Ref.
SKINNY-64-192	23/40	2^{188}	2^{4}	2^{52}	MILP	[11]
	23/40	$\mathbf{2^{184}}$	2^{8}	2^{60}	MILP, SupP	*Fig. 23
	23/40	2^{188}	2^{4}	$\mathbf{2^{28}}$	MILP, SupP	*Fig. 24
SKINNY-128-384	23/56	2^{376}	2^{8}	2^{104}	MILP	[11]
	23/56	$\mathbf{2^{368}}$	2^{16}	2^{120}	MILP, SupP	*Fig. 23
	23/56	2^{376}	2^{8}	$\mathbf{2^{56}}$	MILP, SupP	*Fig. 24

– CF: Compression Function; OT: Output Transformation; Dedicated: Dedicated method;
– Time-1 and Time-2 are complexities of pseudo-preimage and preimage attacks following the notions in [4] when the target is a hash function, and complexities of inverting OT and CF+OT (pseudo-preimage) following the notions in [30] for Grøstl, respectively.
– * Please refer to the full version of this paper [6].
– † The presented complexities for the attacks in [32] are recomputed by removing constant factors (*e.g.*, the cost C_{TL} for lookup table is replaced by 1) and replacing $C_2(2n, b)$ that is lower bounded by $b/2$ in [30] with $\overline{C_2(2n, b)}$ that can be 2^0 considering the amortized complexity. Thus, all complexities are computed follow the same way.
– ‡ The 5- and 6-round attacks in [22] are on the OT of Grøstl-256. To convert to pseudo-preimage, we used the results for the case with no truncation in [22]. However, for the 6-round attack, in which guessing are required, it cannot be directly used. Thus, we combined the attack on OT in [22] with the best previous attack on the CF in [32].

other hash functions. Advanced techniques in MITM preimage attacks were developed, including the *splice-and-cut* [2], (probabilistic) *initial structure* [15, 27], and (indirect) *partial matching* techniques [2, 27].

The *splice-and-cut* technique [2] views the input and output of the compression function *connected* through the feedforward operation. The *initial structure* is a few consecutive starting steps, where the two chunks overlap and two sets of neutral words (denoted by $\overrightarrow{\mathcal{N}}$ and $\overleftarrow{\mathcal{N}}$) appear simultaneously at these steps. Typically, one adds *constraints* to the values of $\overrightarrow{\mathcal{N}}$ and $\overleftarrow{\mathcal{N}}$ to limit their impact on the opposite chunk such that steps after the initial structure (forward chunk) can be computed independently of $\overleftarrow{\mathcal{N}}$ and steps before the initial structure (backward chunk) can be computed independently of $\overrightarrow{\mathcal{N}}$. The *(indirect-) partial matching* exploits any easily determined relations between the computed ending states from the two computation chunks instead of requiring values of full states. There may or may not be any *neutral bits* from the message of the compression function (or key of the underlying block cipher).

The Attack Framework. The procedure and complexity of the MITM attack depend on how the computation is decomposed into independent chunks, how neutral bits are selected and constrained, and how to match. Once these configurations have been determined, the basic MITM pseudo-preimage attack on a CF goes as follows (with initial structure and partial matching).

1. Assign arbitrary compatible values to all bytes except those that depend on the neutral bytes.
2. Obtain possible values of neutral bytes $\overrightarrow{\mathcal{N}}$ and $\overleftarrow{\mathcal{N}}$ under the constraints on them. Suppose there are 2^{d_1} values for $\overrightarrow{\mathcal{N}}$, and 2^{d_2} for $\overleftarrow{\mathcal{N}}$.
3. For all 2^{d_1} values of $\overrightarrow{\mathcal{N}}$, compute forward from the initial structure to the matching point to get a table $\overrightarrow{\mathcal{L}}$, whose indices are the values for matching, and the elements are the values of $\overrightarrow{\mathcal{N}}$.
4. For all 2^{d_2} values of $\overleftarrow{\mathcal{N}}$, compute backward from the initial structure to the matching point to get a table $\overleftarrow{\mathcal{L}}$, whose indices are the values for matching, and the elements are the values of $\overleftarrow{\mathcal{N}}$.
5. Check whether there is a match on indices between $\overrightarrow{\mathcal{L}}$ and $\overleftarrow{\mathcal{L}}$.
6. In case of partial-matching exist in the above step, for the surviving pairs, check for a full-state match. In case none of them are fully matched, repeat the procedure by changing values assigned in Step 1 till find a full match.

The Attack Complexity. Denote the size of the internal state by n, the degree of freedom in the forward and backward chunks by d_1 and d_2, and the number of bits for (partial-) matching by m, the time complexity of the attack is [4]:

$$2^{n-(d_1+d_2)} \cdot (2^{\max(d_1,d_2)} + 2^{d_1+d_2-m}) \simeq 2^{n-\min(d_1,d_2,m)}. \tag{1}$$

3 MILP Model for the Configuration Search

3.1 Basic MILP Model for MITM

For searching MITM attacks using MILP, one should characterize valid attack configurations in MILP language.

To generate MILP models and search for the best MITM attack, the high-level workflow is as follows. Given a targeted cipher, enumerate all possible combinations of starting and matching points (detailed in **Searching Framework.**). For each combination, build the MILP model as follows: 1) claim variables indicating the attribute of each state cell in the input/output of each operation in each round, introduce necessary auxiliary variables; 2) define linear equations and inequalities to express relations and constraints among those variables according to the cipher specification and attack principles; 3) write the objective function corresponding to the optimal computational complexity. For each generated MILP model, use an MILP solver to search its optimal solution; Among solutions of various MILP models, select the best one; parse the solution into the attack.

The challenging parts of the modeling are to formalize new attack settings and techniques into explicit rules and then translate into linear inequalities/equations. The linear inequalities/equations should be exact such that the solutions can one-to-one correspond to valid attack configurations.

In the following, we describe the basic modeling method following the framework in [5]. As introducing the basic model, we directly introduce some natural generalizations, while other generalizations and enhancements are deferred to the sequel subsections.

Notations and Encoding. For the ease of notation, the conversational coloring scheme for describing and visualizing MITM attacks [4,5,11,24,28,30] is adopted here to characterize the attributes of state cells. As in the MILP modeling in [5], the attribute of individual state cell is encoded using two binary variables (x, y) and defined as follows. A cell is

- `Gray` (■) if and only if its value is a predefined constant, thus is known and fixed in *both* forward and backward chunks; Indicating variables $(x, y) = (1, 1)$;
- `Blue` (■) if and only if its value is determined by forward neutral bytes and predefined constants, thus is known and active in the *forward* chunk but unknown in the backward; Indicating variables $(x, y) = (1, 0)$;
- `Red` (■) if and only if its value is determined by backward neutral bytes and predefined constants, thus is known and active in the *backward* chunk but unknown in the forward; Indicating variables $(x, y) = (0, 1)$;
- `White` (□) if and only if its value is determined by both forward and backward neutral bytes, thus can be computed independently in *neither* chunk; Indicating variables $(x, y) = (0, 0)$.

For convenience, `Black` (■) is used to represent any of the 4 cells (■, ■, ■, □). Additionally, indicating variables β and ω are defined as follows.

- $\beta = \begin{cases} 1 & \text{is inactive, do not contain degree of freedom; is } \texttt{Gray}; \\ 0 & \text{is active, is } \texttt{Blue}, \texttt{Red} \text{ or } \texttt{White}. \end{cases}$
- $\omega = \begin{cases} 1 & \text{cannot be computed in both forward and backward chunks; is } \texttt{White}; \\ 0 & \text{can be computed in at least one direction; is } \texttt{Blue}, \texttt{Red}, \text{ or } \texttt{Gray}. \end{cases}$

Searching Framework. The top-level of a search for the optimal MITM attacks is to enumerate all high-level configurations. A high-level configuration is determined by four parameters, including $\texttt{total}_r$, $\texttt{init}_r^{\mathrm{E}}$, $\texttt{init}_r^{\mathrm{K}}$, and $\texttt{match}_r$, representing the total number of targeted rounds, the location of the initial state in encryption, the location of the initial state in the key-schedule, and the location of the matching point, respectively. For the complete search of $\texttt{total}_r$-round attacks, all possible combinations of values of $\texttt{init}_r^{\mathrm{E}}$, $\texttt{init}_r^{\mathrm{K}}$, and $\texttt{match}_r$ should be tried; each combination corresponds to an independent MILP model; The following description of the modeling method is for an individual model with a fixed ($\texttt{total}_r$, $\texttt{init}_r^{\mathrm{E}}$, $\texttt{init}_r^{\mathrm{K}}$, $\texttt{match}_r$).

When building an MILP model, constraints are imposed on propagation of attributes starting from some initial states (*i.e.*, $\overleftrightarrow{S}^{\texttt{ENC}}$ and $\overleftrightarrow{S}^{\texttt{KSA}}$ in round $\texttt{init}_r^{\mathrm{E}}$ and $\texttt{init}_r^{\mathrm{K}}$) and terminating at some ending states (*i.e.*, $\overrightarrow{End}$ and $\overleftarrow{End}$ in round $\texttt{match}_r$) from two directions.

- In all states in both encryption and key-schedule data-paths, constraints are imposed on attribute-indicating variables for state cells. The constraints indicate the relations of cells between consecutive states intra-round and inter-round. The change of attributes of cells from state to state is attribute propagation.
- In the initial states in encryption and key-schedule (*i.e.*, in $\overleftrightarrow{S}^{\texttt{ENC}}$ and $\overleftrightarrow{S}^{\texttt{KSA}}$), the attribute of each cell is constrained to be non-`White`. Thus, its indicating variables can take three assignments, *i.e.*, $(x_i, y_i) \in \{(1,0),(0,1),(1,1)\}$ for $\forall\ i \in \mathcal{N}$, where $\mathcal{N} = \{0, 1, \cdots, \mathrm{N}_{\texttt{row}} \cdot \mathrm{N}_{\texttt{col}} - 1\}$. The states in the starting round are called initial because initial degree of freedoms are all contained in these states. Denote the initial degree of freedom for the forward by $\overrightarrow{\iota}$, and that for backward by $\overleftarrow{\iota}$. Accordingly, one has the equations for $\overrightarrow{\iota}$ and $\overleftarrow{\iota}$ as in Eq. (2), where $|\mathcal{BL}^{\texttt{ENC}}|$, $|\mathcal{RD}^{\texttt{ENC}}|$, $|\mathcal{BL}^{\texttt{KSA}}|$, and $|\mathcal{RD}^{\texttt{KSA}}|$ denote the number of `Blue`, `Red` cells in $\overleftrightarrow{S}^{\texttt{ENC}}$, and `Blue`, `Red` cells $\overleftrightarrow{S}^{\texttt{KSA}}$, respectively.
- In the ending states at the matching round (*i.e.*, $\overrightarrow{End}$ and $\overleftarrow{End}$), each column is associated with a general variable since the matching is performed column-wise. Specifically, for each pair of input/output columns of the states before and after MixColumns, the associated variable $\overrightarrow{m}^{\leftarrow}_i$ indicates the degree of matching in column i and is constrained by the numbers of `Blue`, `Red`, and `Gray` cells. The total degree of matching of the attack, denoted as $\overrightarrow{m}^{\leftarrow}$, is the sum of the degrees of matching from all columns, as shown in Eq. (4).

In order for `Blue`- and `Red`-attribute to independently propagate from initial states to ending states and remain matchable, special constraints might be occasionally imposed to indicate whether to cancel impact by consuming degrees

of freedom. Concrete constraints on how attributes propagate and how freedom be consumed will be introduced shortly. At this moment, let's denote the accumulated consumed degree of freedom of forward by $\overrightarrow{\sigma}$ and that of backward by $\overleftarrow{\sigma}$. The remaining degrees of freedom in forward and backward (denoted by $\overrightarrow{d_b}$ and $\overleftarrow{d_r}$, respectively) after the occasionally freedom-consuming during attribute-propagation are constrained as in Eq. (3). Note that the $\overrightarrow{d_b}$ and $\overleftarrow{d_r}$ are the essential degrees of freedom for the attack, which determine the attack complexity.

According to the attack complexity in Formula 1, the search for a *valid* attack corresponds to the search for a valid attribute propagation with $\min\{\overrightarrow{d_b}, \overleftarrow{d_r}, \overrightarrow{\overleftarrow{m}}\} \geq 1$; the search of the *optimal* attack corresponds to the search with maximized $\min\{\overrightarrow{d_b}, \overleftarrow{d_r}, \overrightarrow{\overleftarrow{m}}\}$. Thus, the objective of the search model is to maximize a variable $\tau_{\texttt{Obj}}$, which is constrained by Eq. (5).

$$\begin{cases} \overrightarrow{\iota} = |\mathcal{BL}^{\texttt{ENC}}| + |\mathcal{BL}^{\texttt{KSA}}|, \\ \overleftarrow{\iota} = |\mathcal{RD}^{\texttt{ENC}}| + |\mathcal{RD}^{\texttt{KSA}}|. \end{cases} \quad (2) \qquad \begin{cases} \overrightarrow{d_b} = \overrightarrow{\iota} - \overrightarrow{\sigma}, \\ \overleftarrow{d_r} = \overleftarrow{\iota} - \overleftarrow{\sigma}. \end{cases} \quad (3) \qquad \overrightarrow{\overleftarrow{m}} = \sum_{i=0}^{\mathrm{N_{col}}-1} \overrightarrow{\overleftarrow{m}}_i \,. \quad (4) \qquad \begin{cases} \tau_{\texttt{Obj}} \leq \overrightarrow{d_b}, \\ \tau_{\texttt{Obj}} \leq \overleftarrow{d_r}, \\ \tau_{\texttt{Obj}} \leq \overrightarrow{\overleftarrow{m}} \end{cases} . \quad (5)$$

Basic Rules for Attributes to Propagate and to Match. The attribute propagation and the matching are governed by two types of constraints. The first type is due to the specification of the targeted cipher; the second type is due to the principle of the attack. If attacks are technically improved, the second type constraints should be adapted to the improvement. In turn, when the second type constraints are properly relaxed, the attack space is expanded so that improved attacks might be discovered.

Remark 1 (Bi-direction attribute-propagation and cancellation (BiDir)). Previously [4, 5, 11, 24, 28, 30], the constraints are imposed such that the propagation of `Red`-attribute can "make a concession" [5] to the propagation of `Blue`, but `Blue` never "gives in" to `Red` for forward computation, and vice versa. Unlike the previous work, in this work, the constraints are relaxed such that the propagation of `Blue` or `Red` attribute can make a concession (cancel its impact by consuming its degree of freedom) to the propagation of the opposite attribute in both directions. The reasons are as follows.

[5] Here, the use of phrases "make a concession" or "give in" is due to a view of the forward computation and backward computation be in a competition for being able to be propagated unaffected. In previous attacks, for forward computation, propagation of `Blue`-attribute is of high priority. When unaffected propagation of `Blue`-attribute becomes not straightforward due to the existence of cells of `Red`-attribute, we may try to cancel the impact by consuming the freedom of backward to ensure the propagation of `Blue`-attribute. We say such cancellation of impact by consuming freedom "concede", "make a concession " or "give in".

– In the modeling, we introduce neutral bytes into key states as well as in encryption states to bring more degree of freedom. In attribute propagation through encryption, letting `Blue`-propagation concede such that a local `Red`-cell be reserved, that might enable a remote `Red`-cell introduced from the key state be canceled such that `Blue`-propagation be possible in that remote point. That also applies to the case of `Red`-propagation.
– Besides, letting `Blue` to concede and reserve a local `Red`-cell might enable this local `Red`-cell to propagate and combine with other `Red`-cells at a remote point such that their impacts on a certain cell be mutually canceled through MixColumns and benefit `Blue`-propagation.
– In addition, once an attribute of `Blue` or `Red` propagate to the ending states no matter from which direction, it provides source of degree of matching.

Relaxing the model in this way results in attacks with bi-direction attribute-propagation; the attack space is expanded so that improved attacks are possible. However, this relaxation caused the models to be solved less efficiently. When the efficiency is acceptable, we used such relaxed model for better attacks. Otherwise, we restricted the models such that in certain rounds of the cipher, one attribute can only propagate in one direction and concede in the other direction.

In the following, we describe the MILP modeling with the relaxed model as the basic setting. The difference between this basic modeling and the modeling in [5] and how to get the restricted models will be indicated alongside.

Modeling of the Attribute Propagation through SubBytes and ShiftRows. The SubBytes operation does not change the attribute of the cells, thus is not involved when building the basic model. The ShiftRows operation permutes the state cells, thus can be modeled by a set of equations or hardcoded variable substitution.

Modeling of the Attribute Propagation through AddRoundKey and `XOR` *(`XOR-RULE`).* The AddRoundKey operation is involved in the model when the cipher has KeySchedule (message schedule), and the attack exploits freedom from the key state. Basically, the attribute propagation through AddRoundKey is governed by a set of cell-wise constraints under the name `XOR-RULE`. The high-level principle is that `White` is the dominant attribute, `Gray` is the recessive attribute, `Blue` and `Red` are mutually exclusive attributes. The concrete rules are as follows.

– A `White` cell `XOR`ed with a cell of any attribute results in a `White` cell, *i.e.*, (□ ⊕ ■) → □;
– a `Gray` cell `XOR`ed with a cell of any attribute results in the cell of the same attribute, *i.e.*, (■ ⊕ ■) → ■;
– a couple of `Blue` and `Red` cells results in a cell deteriorated to `White`, *i.e.*, (■ ⊕ ■) → □;
– a couple of `Blue` cells can keep the attributes without consuming or evolve to `Gray` by consuming a degree of freedom of `Blue`, *i.e.*, (■ ⊕ ■) → ■ or (■ ⊕ ■) $\xrightarrow{-1\times\blacksquare}$ ■;

– a couple of `Red` cells can keep the attributes without consuming or evolve to `Gray` by consuming a degree of freedom of `Red`, *i.e.*,
($\blacksquare \oplus \blacksquare$) $\rightarrow \blacksquare$ or ($\blacksquare \oplus \blacksquare$) $\xrightarrow{-1\times\blacksquare}$ $\blacksquare$.

These propagation rules can be described using only a few variables, thus can be easily translated into inequalities using convex hull computations [29].

Modeling of the Attribute Propagation through MixColumns (MC-RULE). The attribute propagation through MixColumns is governed by a set of column-wise constraints under the name `MC-RULE`. The `MC-RULE` constraints are mostly governed by the branch number ($\mathtt{Br_n}$) of the MixColumns. Again, the high-level principle is that `White` is the dominant attribute, `Gray` is the recessive attribute, `Blue` and `Red` are mutually exclusive attributes. The concrete rules are as follows:

– any `White` cell in an input column results in all cells in the output column deteriorated to `White`, *i.e.*,
$(i \times \square,\ j \times \blacksquare) \rightarrow (\mathrm{N_{row}} \times \square)$, where $i \geq 1, i + j = \mathrm{N_{row}}$;
– the `Gray` attribute inherits to the output without consuming degrees of freedom only if all cells in the input column are `Gray`, *i.e.*,
$(\mathrm{N_{row}} \times \blacksquare) \rightarrow (\mathrm{N_{row}} \times \blacksquare)$;
– existing no `White` cell, a column of i `Blue`, j `Red`, and k `Gray` cells propagate to a column of i' `Blue`, j' `Red`, k' `Gray`, and ℓ' `White` cells by consuming $j' + k'$ degree of freedom from `Blue`, and $i' + k'$ from `Red`, *i.e.*,
$(i \times \blacksquare,\ j \times \blacksquare,\ k \times \blacksquare) \xrightarrow[-(i'+k')\times\blacksquare \text{ if } j\neq 0]{-(j'+k')\times\blacksquare \text{ if } i\neq 0} (i' \times \blacksquare,\ j' \times \blacksquare,\ k' \times \blacksquare,\ \ell' \times \square)$, where
$i + j + k = i' + j' + k' + \ell' = \mathrm{N_{row}}$ and $\begin{cases} j' + k' < i \leq \mathrm{N_{row}} & \text{if } i \neq 0 \\ j' + k' = \mathrm{N_{row}} & \text{otherwise} \end{cases}$,
$\begin{cases} i' + k' < j \leq \mathrm{N_{row}} & \text{if } j \neq 0 \\ i' + k' = \mathrm{N_{row}} & \text{otherwise} \end{cases}$. Note that when $i \neq 0$, $j' + k' < i \Leftrightarrow \mathrm{N_{row}} - i' - l' < i \Leftrightarrow i + i' + l' >= \mathrm{N_{row}} + 1$, which is due to the branch number; similarly, $i' + k' < j$ when $j \neq 0$ is due to the branch number.

To formalize these propagation rules into a system of inequalities, the involved number of variables is not small. Concretely, the involved variables include the binary variables that indicate the attribute of each cell in the input and output columns, *i.e.*, (x_i^I, y_i^I), (x_i^O, y_i^O), and ω_i^I for $i \in \{0, 1, \cdots, \mathrm{N_{row}} - 1\}$, the general variables $c_{\vec{x}}$ and $c_{\vec{y}}$ representing the consumed degree of freedom from `Blue` and `Red`, respectively. Apart from those variables, three auxiliary binary variables are introduced to indicate the following attributes of the input column:

$$\vec{\omega} = \begin{cases} 1 & \text{exists } \mathtt{White} \text{ cell}, \\ 0 & \text{otherwise}. \end{cases} \quad (6) \qquad \vec{x} = \begin{cases} 1 & \text{all are } \mathtt{Blue/Gray}, \\ 0 & \text{exists } \mathtt{Red/White}. \end{cases} \quad (7) \qquad \vec{y} = \begin{cases} 1 & \text{all are } \mathtt{Red/Gray}, \\ 0 & \text{exists } \mathtt{Blue/White}. \end{cases} \quad (8)$$

The constraints can then be formalized using inequalities listed in Eq. (15, 16, 17) in the full version [6].

Modeling of the Matching through MixColumns and AddRoundKey (MATCH-RULE). The modeling for matching also focuses on the MixColumns and AddRoundKey operations. The matching through MixColumns and AddRoundKey is governed by a set of column-wise constraints under the name `MATCH-RULE`.

The involved states are those around MixColumns at the matching round, including $\overrightarrow{End}$ and $\overleftarrow{End}$ in encryption and $\overrightarrow{End}^{\mathrm{KMC}}$ or $\overleftarrow{End}^{\mathrm{K}}$ in key-schedule. Note that $\overrightarrow{End}$ and $\overleftarrow{End}$ are states that have not been added with key-state $\overrightarrow{End}^{\mathrm{KMC}}$ or $\overleftarrow{End}^{\mathrm{K}}$. Since AddRoundKey is linear, the influence of AddRoundKey for matching can be formalized using simple rules. Concretely, only a `White` cell in key state destroy the match-ability of the corresponding cell in encryption state. The `Blue` and `Red` cells in key state do not impact the match-ability but on the contrary might provide degree of matching.

Remark 2. For specific targeted cipher whose key-schedule is almost identical to the encryption, *e.g.*, Whirlpool, one can use $\overrightarrow{End} \oplus \overrightarrow{End}^{\mathrm{KMC}}$ as an equivalent key addition to $\overleftarrow{End} \oplus \overleftarrow{End}^{\mathrm{K}}$. The color pattern (most importantly, the distribution of `White` cells) of $\overrightarrow{End}^{\mathrm{KMC}}$ and $\overleftarrow{End}^{\mathrm{K}}$ are different in most cases. Thus, adding $\overrightarrow{End}^{\mathrm{KMC}}$ or $\overleftarrow{End}^{\mathrm{K}}$, these two ways may have different effects on the degree of matching. To find the optimal solution, one should consider both ways. However, we can simply choose to use the key state with fewer `White` cells. That is because, known the propagation direction of the key-schedule (*i.e.*, $\mathtt{init}_r^{\mathrm{K}}$), between the two states $\overrightarrow{End}^{\mathrm{KMC}}$ and $\overleftarrow{End}^{\mathrm{K}}$, the one that is near to the initial key state must have set of `White` cells be subset of that in the remote state, thus has less impact.

The condition for the i-th column to have $\overrightarrow{\overleftarrow{m}}_i$ degree of matching is as follows: denote the number of known cells (*i.e.*, except `White` cells) in the input and output columns by $\overline{m_k}_i$; when $\overline{m_k}_i > \mathrm{N}_{\mathtt{row}}$, $\overrightarrow{\overleftarrow{m}}_i = \overline{m_k}_i - \mathrm{N}_{\mathtt{row}}$; otherwise, $\overrightarrow{\overleftarrow{m}}_i = 0$. Denote the number of white cells by $\overline{m_w}_i$; Since $\overline{m_k}_i = 2{\cdot}\mathrm{N}_{\mathtt{row}} - \overline{m_w}_i$, one have $\overrightarrow{\overleftarrow{m}}_i = \max(0, \mathrm{N}_{\mathtt{row}} - \overline{m_w}_i)$. Accordingly, the concrete system of inequalities modeling `MATCH-RULE` can be obtained as explained in Sect. A and as listed in Eq. (18) in the full version [6].

3.2 Superposition States and Separate Attribute-Propagation

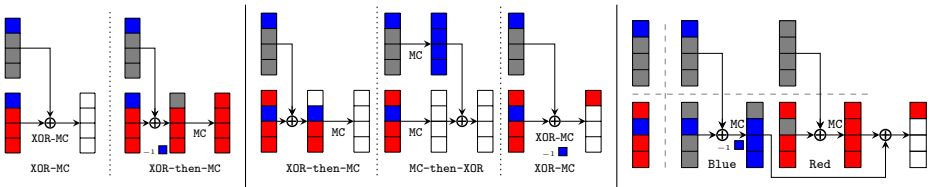
Except for allowing bi-direction attribute-propagation, the above basic modeling is in line with previous work [4, 5, 11, 24, 28, 30], where `Blue` and `Red` attributes propagate exclusively and competitively through each operation. Once the two exclusive attribute propagate into the same cell position, this cell is considered being `White`, cannot be independently computed in either forward nor backward.

However, such modeling might miss valid attacks as will be discussed shortly. In our final modeling, the two exclusive attributes `Blue` and `Red` propagate independently as long as it is possible. This is achieved by introducing *superposition states*. Superposition states are all intermediate states in encryption and key-schedule being separated into two virtual states. One virtual state carries one attribute propagation independently of the other attribute propagation. The two

virtual states are combined only when going through non-linear operations. In this way, two exclusive attributes can simultaneously propagate through all linear operations in both encryption and key-schedule. See Fig. 3 for an example. This superposition setting captures the independence of the computation in a more essential way, allows new ways of local collisions between/within the encryption and key states, thus, maximize the chances of canceling impacts on independence and enables new ways of MITM decomposition. Concretely, the reasons and benefits of separating propagation with superposition states are as follows.

- As mentioned above, under certain constraints, `Blue` propagation can make a concession so that impacts are canceled and `Red` can propagate unaffected, and vice versa. To cancel impact, it requires consuming degrees of freedom. To be able to consume degrees of freedom, different types of operations impose different requirements on the attribute of input states. Concretely, the `XOR-RULE` and `MC-RULE` requires differently for canceling impacts. Attributes of state cells might not meet the individual requirements but it is actually possible to cancel impacts when combining these two linear operations.
  In [5], the propagation through the combination of AddRoundKey and MixColumns is characterized using the set of `XOR-MC-RULE`. In `XOR-MC-RULE`, the `OR` of the attribute indicating variables of encryption and key-state cell is used to indicate the attribute of the input cell of MixColumns. In that way, the group of cells of the same attribute in both encryption and key states can jointly cancel their impacts on certain output cells.
  However, using only `XOR-MC-RULE`, the possibility of the following scenario is missed. That is, an attribute can be completely canceled via `XOR-RULE` before propagating through MixColumns (refer to Fig. 2a for an illustration). Thus, to find optimal attack configurations, applying `XOR-RULE`-then-`MC-RULE` and `XOR-MC-RULE` should be both considered in the models.
  In this work, we model the combination of AddRoundKey and MixColumns by considering the separation of (`Blue` and `Red`) attribute propagation with superposition states. Note that AddRoundKey and MixColumns are linear. Essentially, for linear operations, in the same state, the attributes of `Blue` and `Red` can separately propagate through them and then combine by cell-wise `XOR` upon the non-linear operation (*i.e.*, SubBytes).
  Due to this separation with superposition states, `XOR-RULE` and `MC-RULE` without `XOR-MC-RULE` are sufficient (refer to Fig. 2 for an illustration of the separation of attribute-propagation through AddRoundKey and MixColumns).
- Additionally, the key/message-schedule of the ciphers also has linear operations. It is possible that before going through the non-linear operation in the key-schedule, the attribute of one cell in the round-key is a linear combination of `Blue` and `Red`. If not be separately considered, such a linear combination of `Blue` and `Red` becomes `White`. Separately, the `Blue` component in the linear combination in a key state can be used to cancel impact from the `Blue` component in another key state or in an encryption state, and same for the `Red` component. Consequently, impacts that cannot be canceled

in previous models can be canceled now, thus independent propagation gets more chances (see Fig. 22 in the full version [6] for an example.)
– Moreover, at the matching point, if a key state cell is in superposition (a linear combination of `Blue` and `Red`), it does not impact the matching ability of the corresponding state cell, and instead, the `Blue` component and `Red` component may provide degrees of matching (see Fig. 18 in the full version [6] for an example.)



(a) `XOR-MC-RULE` does not cover a propagation

(b) `XOR-MC-RULE` is necessary if not in superposition

(c) `XOR-RULE` and `MC-RULE` are sufficient once in superposition

– In 2a, using `XOR-MC-RULE`, a result of full `Red` attribute can not be achieved, which can be obtained by `XOR-then-MC-RULE`.
– In 2b, without superposition, (`Blue` and `Red`) attribute-propagation ruled by `XOR-MC-RULE` cannot be obtained directly by `XOR-then-MC-RULE` and `MC-then-XOR-RULE`.
– In 2c, with (`Blue` and `Red`) attribute-propagation in superposition, `XOR-RULE` and `MC-RULE` are sufficient (2c achieve the same results as achieved by `XOR-MC-RULE` 2b).

Fig. 2: Combination of linear operations and superposition attribute-propagation

3.3 Multiple Ways of `AddRoundKey` (MulAK)

In Whirlpool, the key-schedule shares the same operations with the encryption except for AddRoundKey. This identity between encryption and key-schedule enables that in encryption, the AddRoundKey can be easily moved around MixColumns. Moving around MixColumns is simply switching between adding $\#\mathbf{KMC}$ or adding $\boldsymbol{k}$, where $\#\mathbf{KMC}$ and $\boldsymbol{k}$ are the states before and after MixColumns in the round function of key-schedule, that is, a switch between adding the real round-key $\boldsymbol{k}$ or an equivalent (up to MixColumns) round-key $\#\mathbf{KMC}$.

Moving AddRoundKey before MixColumns and using $\#\mathbf{KMC}$ can bring more advantages in some cases. Take the case where we need to reserve `Blue` by consuming `Red` for example. Let's focus on one column of the state. Suppose there is one `Red` cell in that column of $\#\mathbf{MC}$ and one `Red` cell in the same cell-position in $\#\mathbf{KMC}$; the influence of these single `Red` cells will be diffused into the whole columns in both states if there is no constraint. In such case, adding $\#\mathbf{KMC}$ with $\#\mathbf{MC}$ and letting the `Red` cell in $\#\mathbf{MC}$ be canceled by the `Red` cell in $\#\mathbf{KMC}$ achieve the same cancellation effect but consume fewer degrees of freedom than adding $\#\mathbf{AK}$ with $\boldsymbol{k}$ (take using $\#\mathbf{MC}^5 \oplus \#\mathbf{KMC}^4$ vs. $\#\mathbf{AK}^5 \oplus \boldsymbol{k}_5$ in Fig. 3 for an example). Similarly, it is also possible that adding $\#\mathbf{AK}$ with $\boldsymbol{k}$ after the MixColumns has more advantages than adding $\#\mathbf{MC}$ with $\#\mathbf{KMC}$ (take using $\#\mathbf{SB}^4 \oplus \boldsymbol{k}_3$ vs. $\#\mathbf{MC}^5 \oplus \boldsymbol{k}_2$ in Fig. 3 for an example).

Thus, to find optimal attack configurations, both scenarios should be considered. The essential difference between the scenarios is to either firstly use freedom in the key state to directly cancel impacts before diffusion or to postpone the insertion of the key state in order to postpone impacts from the key. We name the choice of applying the first scenario by `AK-MC-RULE` and the second scenario by `MC-AK-RULE`. The integration of the two scenarios into one model is in the form of indicator constraints that is available in Gurobi. Note that, for forward computation, `AK-MC-RULE` corresponds to using $\#\mathbf{MC} \oplus \#\mathbf{KMC}$, `MC-AK-RULE` corresponds to using $\#\mathbf{AC} \oplus \boldsymbol{k}$; for backward computation, `AK-MC-RULE` corresponds to using $\#\mathbf{SB} \oplus \boldsymbol{k}$, and `MC-AK-RULE` corresponds to using $\#\mathbf{MC} \oplus \#\mathbf{KMC}$;

In addition, since the MixColumns is column-wise, different columns can apply different ways (apply `AK-MC-RULE` or `MC-AK-RULE`) independently. Besides, since MixColumns is linear, with superposition states, different attributes of `Blue` and `Red` can independently apply `AK-MC-RULE` or `MC-AK-RULE`.

Integrating such flexibility of choice into the MILP models expand the covered attack space but make the solving less efficient. So, to decide in which way to proceed for each column, we use the heuristic that when the key-schedule has already consumed some degree of freedom in that column, we apply `MC-AK-RULE`; otherwise, apply `AK-MC-RULE`.

3.4 Enhanced Model with Guess-and-Determine (GnD)

The Guess-and-Determine Strategy. In MITM attacks, one unfavorable situation is that a single or a few unknown cells in the input column of `MC` makes all cells in the output column unknown (refer to point 1 of `MC-RULE`). This is inevitable due to the diffusion of MixColumns, especially the property of the MDS matrix.

To turn things around in such situations, Sasaki *et al.* in [28] invent the following guess-and-determine strategy. That is, guess values of the few unknown cells to continue the propagation of attribute to reach the matching point, after (partial) matching, check the consistency of the few guessed cells. If the gained degree of matching is sufficient and the required guesswork is very little, one can still achieve a better attack complexity than a brute-force attack.

Concretely, denote 2^c by ς (where c is the number of bits in each cell). Let:

- $\overrightarrow{d_{g_b}}$ be the number of cells only guessed to be `Blue` (forward computation);
- $\overleftarrow{d_{g_r}}$ be the number of cells only guessed to be `Red` (backward computation);
- $\overleftrightarrow{d_{g_{br}}}$ be the number of cells guessed to be both `Blue` and `Red`. [6]

The framework of the MITM attack with GnD is as follows:

1. Assign arbitrary compatible values to all cells except for those depending on the neutral bits, and assign arbitrary values to the constants in pre-defined constraints on neutral bits;

[6] Since we allow bi-direction attribute propagation in superposition states, it might bring benefit to guess a superposition cell to be simultaneously `Blue` and `Red`. Thus, here is a slight generalization of the previous GnD strategy.

2. Compute values $\{\overrightarrow{v}_i\}$ of forward neutral bits and values $\{\overleftarrow{v}_i\}$ of backward neutral bits fulfilling pre-defined constraints.
3. For all $\varsigma^{\overrightarrow{d_b}}$ values $\{\overrightarrow{v}_i\}$ of forward neutral bits, and $\varsigma^{(\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}})}$ guessed values $\{\overrightarrow{v}_g\}$ for forward, compute forward to the matching point and store all $\varsigma^{\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}}$ partial matching values $\{\overrightarrow{v}_m\}$ in a look up table $\overrightarrow{\mathcal{T}}$ (the values are $\overrightarrow{v}_i$ and $\overrightarrow{v}_g$, and the index is $\overrightarrow{v}_m$).
4. For all $\varsigma^{\overleftarrow{d_r}}$ values $\{\overleftarrow{v}_i\}$ of backward neutral bits, and $\varsigma^{(\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}})}$ guessed values $\{\overleftarrow{v}_g\}$ for backward, compute backward to the matching point, obtain the partial matching values $\overleftarrow{v}_m$.
5. For all values of $\overrightarrow{v}_i$ and $\overrightarrow{v}_g$ in entry $\overrightarrow{\mathcal{T}}[\overleftarrow{v}_m]$, use $\overrightarrow{v}_i$ and $\overleftarrow{v}_i$ to compute and check if the guessed values $\overrightarrow{v}_g$ and $\overleftarrow{v}_g$ are compatible. For compatible guesses, compute to the matching point to check if it is a full-state match. If so, use $\overrightarrow{v}_i$ and $\overleftarrow{v}_i$ to compute the preimage, output it, and return; otherwise, repeat from Step 1 by changing values assigned at that step.

In the above framework of the MITM attack with GnD,

- the total degree of freedom for `Blue` with guessing is $\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}$;
- the total degree of freedom for `Red` with guessing is $\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}$;
- the expected number of matched pairs is $\varsigma^{\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}+\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}}}$;
- the required number of repetitions to get a full match at the guessing cells and a full-state match is $\varsigma^{-(\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}+\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}})}\cdot\varsigma^{\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}+\overleftarrow{d_{g_r}}}\cdot\varsigma^{(n-\overrightarrow{\overleftarrow{m}})}$, which equals $\varsigma^{n-\overrightarrow{d_b}-\overleftarrow{d_r}-\overleftrightarrow{d_{g_{br}}}}$;

Thus, the complexity of the attack is $\varsigma^{n-\overrightarrow{d_b}-\overleftarrow{d_r}-\overleftrightarrow{d_{g_{br}}}}\cdot(\varsigma^{\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}}+\varsigma^{\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}}+\varsigma^{\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}+\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}}})$, which equals

$$\begin{aligned}&\varsigma^n\cdot(\varsigma^{-(\overleftarrow{d_r}-\overrightarrow{d_{g_b}})}+\varsigma^{-(\overrightarrow{d_b}-\overleftarrow{d_{g_r}})}+\varsigma^{-(\overrightarrow{\overleftarrow{m}}-\overrightarrow{d_{g_b}}-\overleftarrow{d_{g_r}}-\overleftrightarrow{d_{g_{br}}})})\\ \simeq\ &\varsigma^n\cdot\max(\varsigma^{-(\overleftarrow{d_r}-\overrightarrow{d_{g_b}})},\varsigma^{-(\overrightarrow{d_b}-\overleftarrow{d_{g_r}})},\varsigma^{-(\overrightarrow{\overleftarrow{m}}-\overrightarrow{d_{g_b}}-\overleftarrow{d_{g_r}}-\overleftrightarrow{d_{g_{br}}})})\end{aligned}\tag{9}$$

Accordingly, the complexity is determined by

$$\min(\overleftarrow{d_r}-\overrightarrow{d_{g_b}},\overrightarrow{d_b}-\overleftarrow{d_{g_r}},\overrightarrow{\overleftarrow{m}}-\overrightarrow{d_{g_b}}-\overleftarrow{d_{g_r}}-\overleftrightarrow{d_{g_{br}}}).$$

Building Model for Guessing. To model the mechanism of GnD, three binary variables, g_b, g_r, g_{br}, are introduced for each cell in the input state of MixColumns (invMixColumns for the backward computation).

The variables indicate whether the values of the cells should be guessed to be of one attribute. Concretely, $g_b = 1$ for guessing one `White` cell to be `Blue`. $g_r = 1$ for guessing one `White` cell to be `Red`. $g_{br} = 1$ for guessing one `White` cell to be `Blue` (for forward propagation) and also `Red` (for backward propagation).

With these guess-indicating variables, attribute-indicating variables of each cell in the input of MixColumns are thus constrained together with attribute-indicating variables of the cell in output state of last operations (*e.g.*, ShiftRows

or AddRoundKey). Besides, according to the complexity Eq. (9) of the attack with GnD, the objective should turn from $\min(\overrightarrow{d_b}, \overleftarrow{d_r}, \overrightarrow{\overleftarrow{m}})$ to $\min(\overrightarrow{d_b} - \overleftarrow{d_{g_r}}, \overleftarrow{d_r} - \overrightarrow{d_{g_b}}, \overrightarrow{\overleftarrow{m}} - \overrightarrow{d_{g_b}} - \overleftarrow{d_{g_r}} - \overleftrightarrow{d_{g_{br}}})$. Thus, the variable that to be maximized is constrained as in Eq. (10) and (11).

$$\begin{cases} \overrightarrow{d_{g_b}} = \sum_{r=0,i=0}^{\mathtt{total}_r-1,n-1} {g_b}_i^r, \\ \overleftarrow{d_{g_r}} = \sum_{r=0,i=0}^{\mathtt{total}_r-1,n-1} {g_r}_i^r, \\ \overleftrightarrow{d_{g_{br}}} = \sum_{r=0,i=0}^{\mathtt{total}_r-1,n-1} {g_{br}}_i^r, \end{cases} \tag{10}$$

$$\begin{cases} \tau_{\mathtt{Obj}} \le \overrightarrow{d_b} - \overleftarrow{d_{g_r}}, \\ \tau_{\mathtt{Obj}} \le \overleftarrow{d_r} - \overrightarrow{d_{g_b}}, \\ \tau_{\mathtt{Obj}} \le \overrightarrow{\overleftarrow{m}} - \overrightarrow{d_{g_b}} - \overleftarrow{d_{g_r}} - \overleftrightarrow{d_{g_{br}}}. \end{cases} \tag{11}$$

Note that in the starting round, all cells are known and in the matching round, guessing brought no advantage. Thus, for these two rounds, such constraints on guessing can be omitted. Besides, generally, guessing are only required around the matching point. Thus, when trade-off between search quality and search efficiency is needed, these guessing constraints can be added only for rounds around the matching point.

3.5 Transforming to Models for Searching for Collision Attacks

A MITM partial target preimage attack whose matching point is at the last round can be transformed into a collision attack, as described by Li *et al.* in [21]. Thus, the searching of MITM preimage attacks can be constraint to search for such partial target preimage attacks, and then translate into valid collision attack. Concretely, one have the follows.

Definition 1 ((t-bit) partial target pseudo preimage attack on CF [21]). *Given the value v of t bits in T, find (h', m') such that t bits in $T' := CF(h', m')$ at the same position as the t bits in T, take value v.*

Let $\mathcal{A}$ be a random algorithm that can find a t-bit partial target pseudo preimage with a complexity of 2^s. This complexity can be in the average sense, which means that $\mathcal{A}$ can generate 2^r different (h', m', T') in one time with a complexity 2^{r+s}. Additionally, assume $\mathcal{A}$ output different (h', m', T')'s for different calls. Then, a free-start collision attack goes as follows.

- Set t-bit arbitrary value d' to the target. Call $\mathcal{A}$ with d' as the partial target and run it to obtain $2^{(n-t)/2}$ values of (h', m', T').
- From the $2^{(n-t)/2}$ values of (h', m', T'), find a collision on the remaining $(n-t)$ bits of T'.

According to the birthday paradox, with a high probability, the above procedure produces a valid collision pair. The total complexity of this attack is $2^{s+(n-t)/2}$.

The above complexity analysis misses the details of the MITM complexity. In the following, we re-formalize the complexity analysis of the above collision attack using its correspondence with the MITM-GnD preimage attack.

Essentially, the $\mathcal{A}$ can be the core of a MITM pseudo preimage attack on CF. The t-bit partial target corresponds to $c\times \overrightarrow{\overleftarrow{m}}$ bits for partial matching in the MITM attack. Essentially, the t bits for partial matching can be a fixed t-bit linear relations on ℓ bits for $\ell \geq t$, which restrict the values of ℓ bits to a subspace of dimension $\ell - t$. Thus, matching through `MixColumns` instead of matching at exact same bit positions does not have essential influence on the effectiveness of the attack. The 2^s computational complexity of $\mathcal{A}$ corresponds to $\varsigma^{\max\left\{(\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}),(\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}),(\overrightarrow{d_b}+\overrightarrow{d_{g_b}}+\overleftrightarrow{d_{g_{br}}}+\overleftarrow{d_r}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}})\right\}}/\varsigma^{\overrightarrow{d_b}+\overleftarrow{d_r}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}}}$, which equals to $\varsigma^{\max\left\{(\overrightarrow{d_{g_b}}-\overleftarrow{d_r}),(\overleftarrow{d_{g_r}}-\overrightarrow{d_b}),(\overrightarrow{d_{g_b}}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}})\right\}}/\varsigma^{-\overrightarrow{\overleftarrow{m}}}$. Thus, the total complexity is $\varsigma^{\max\left\{(\overrightarrow{d_{g_b}}-\overleftarrow{d_r}),(\overleftarrow{d_{g_r}}-\overrightarrow{d_b}),(\overrightarrow{d_{g_b}}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}}-\overrightarrow{\overleftarrow{m}})\right\}}/\varsigma^{-\overrightarrow{\overleftarrow{m}}} \times \varsigma^{\frac{n-\overrightarrow{\overleftarrow{m}}}{2}}$, *i.e.*,

$$\varsigma^{-\min\left\{(\overleftarrow{d_r}-\overrightarrow{d_{g_b}}),(\overrightarrow{d_b}-\overleftarrow{d_{g_r}}),(\overrightarrow{\overleftarrow{m}}-\overrightarrow{d_{g_b}}-\overleftarrow{d_{g_r}}-\overleftrightarrow{d_{g_{br}}})\right\}} \times \varsigma^{\frac{n+\overrightarrow{\overleftarrow{m}}}{2}}. \tag{12}$$

For a valid attack (better than birthday attack), the following should be fulfilled

$$\left\{\overrightarrow{d_b}-\overleftarrow{d_{g_r}} > \overrightarrow{\overleftarrow{m}}/2, \quad \overleftarrow{d_r}-\overrightarrow{d_{g_b}} > \overrightarrow{\overleftarrow{m}}/2, \quad \overrightarrow{d_{g_b}}+\overleftarrow{d_{g_r}}+\overleftrightarrow{d_{g_{br}}} < \overrightarrow{\overleftarrow{m}}/2\right\}. \tag{13}$$

For searching for the best attack, the objective function is the same as that for preimage attack, *i.e.*,

$$\max\left\{\min\left\{\overrightarrow{d_b}-\overleftarrow{d_{g_r}}, \overleftarrow{d_r}-\overrightarrow{d_{g_b}}, \overrightarrow{\overleftarrow{m}}-\overrightarrow{d_{g_b}}-\overleftarrow{d_{g_r}}-\overleftrightarrow{d_{g_{br}}}\right\}\right\}. \tag{14}$$

Remark 3. A solution to the MILP model only corresponds to a valid attack configuration but does not formally imply a valid attack. For the attack complexity in Eq. 1, 9, and 12 to be valid, the attacker should be able to generate each value of neutral bytes with (amortized) computational complexity $O(1)$. In some obtained attack configurations, the neutral bytes are constrained in such a sophisticated manner that it is not trivial to efficiently generate their values. For such non-trivial cases, we propose to use local meet-in-the-middle procedures to solve the problem (as done for the pseudo-preimage attack on 7-round Whirlpool in Sect. 4.1). Such local meet-in-the-middle procedures might be found by manual analysis or aided by automatic tools, such as the Automatic-tool from [9]. Sometimes, to achieve amortized computational complexity $O(1)$, it is necessary to pre-compute values of neutral bytes for many fixed bytes (enumerated in the outermost loop of the attack) at once. However, for some very complex cases, even aided by automatic-tools and considering amortized complexity, it might be difficult to find pre-computation procedures with total complexity lower than the main procedure. As a theoretical problem for all attacks under this framework, this problem of efficiently generating values of neutral bits stays open.

3.6 Exploit Symmetry of the Ciphers

Integrating that many technical generalizations, the search space is greatly enlarged but the search is slow down. One needs to make a trade-off between the

quality of the searching result and the efficiency of the search. Apart from the ways of trade-off mentioned alongside the introduction of the modeling, one can reduce the problem scale using symmetry of the ciphers. Specifically, in many AES-like designs, the states and operations have symmetric structures and parameters. The symmetry allows projecting attacks on small-size versions to that on large-size versions. Concretely, a state of 8×8 cells can be viewed as a 2×2 matrix of state of 4×4 cells, or a 1×2 matrix of state of 8×4 cells. Suppose the ShiftRows parameters of the 8×8 state version are $\{p_0, p_1, \ldots, p_7\}$ and $(p_{i+4} \mod 4) = p_i$ for $i \in \{0, \ldots, 3\}$, and 2 times the branch number of MixColumns of a 4×4 state version is no less than that of the 8×8 state version. Then, obtaining an attack on the 4×4 state version, cloning the state patterns four times and placing them in a 2×2 matrix, this will result in an attack on the 8×8 state version. Exploiting such symmetry of the ciphers, the search can be efficient, while might lose asymmetric attack configurations.

4 Application to Preimage Attacks on Whirlpool

Whirlpool [7] is a block-cipher based secure hash function designed by Rijmen and Barreto in 2000 and has been adopted as an ISO/IEC standard. It produces a 512-bit hash value using Miyaguchi-Preneel-mode (MP-mode) CF. The CF is defined as $\mathsf{CF}(H_i, M_i) = E_{H_i}(M_i) \oplus M_i \oplus H_i$, where E is a 10-round AES-like block cipher with 8×8-byte internal states. This underlying block cipher takes the 512-bit chaining value H_i as the key material and the 512-bit message block M_i as the input of the encryption. Both the encryption and key-schedule use round functions consisting of four operations:

- SubBytes (SB) applies the Substitution-Box to each byte.
- ShiftColumns (SC) cyclically shifts the j-column downwards by j bytes.
- MixRows (MR) multiplies each row of the state by an MDS matrix.
- AddRoundKey (AK or AddRoundConstants AC) XOR the round key (or round constants in key-schedule) to the state.

Note that the last round is a complete round which is unlike AES; a whiting key is added before the first round of encryption. However, the effect of whitening key will be canceled in the splice-and-cut MITM attacks due to the feed-forward mechanism of MP-mode. Please refer to [7] for a detailed description of Whirlpool.

4.1 New Attacks Resulted from Applying the MILP Modeling

Applying the MILP modeling in Sect. 3 on Whirlpool, improved attacks are found for 5- and 6-round, and first attacks are found for 7-round.

For 5-round attacks, guess-and-determine is not required, but allowing bi-direction attribute-propagation/cancellation is essential to achieve the best complexity. For 6-round attacks, guess-and-determine is the critical technique that enables the improved results. For 7-round attacks, guess-and-determine and

multiple ways of `AddRoundKey` (flexible choices from applying `AK-MC-RULE` or `MC-AK-RULE`) are the two critical points.

The remaining of this section describes how to use one of the resulted attack configurations to launch a concrete attack on 7-round Whirlpool. A brief description of the improved 6-round attack is then followed. In the description, ShiftRows and MixColumns instead of ShiftColumns and MixRows are used. Thus, the states should be transposed to correspond with the specification of Whirlpool. This transposition does not influence the attacks.

Please refer to Figures 11, 12, and 13 in the full version [6] for visualizations of configurations of the 7-, 6-, and 5-round attacks, and refer to Sect. B [6] for a summary of notations.

The attack on 7-round Whirlpool can be obtained by searching on a small version with $\mathrm{N}_{\mathtt{row}} \times \mathrm{N}_{\mathtt{col}} = 4 \times 4$ states and then projecting to attack on the 8×8-state version. In the sequel, to facilitate readers to find the correspondence between the text description and the code implementation for experimental verification of the attack, we describe it using the small version with $\mathrm{N}_{\mathtt{row}} \times \mathrm{N}_{\mathtt{col}} = 4 \times 4$ states that is depicted in Fig. 3. One can quickly project this attack on the small version to the real version of Whirlpool (refer to Sect. 3.6 and the correspondence between Figures 10 and 11 in the full version [6] for illustrations). The attack complexity on the real version is the fourth power of that on the small version.

The attack configuration in Fig. 3a is found with the MILP models. With this configuration, one can conceive an equivalent configuration shown in Fig. 3b. Following both configurations, one can devise the attack, with different procedures to compute initial values of backward neutral bytes. The latter (Fig. 3b) is more direct for the computation; thus, it will be used in the following description. In the following description, all referred states are actual states that are the combination of two virtual states.

Compute initial values of neutral bytes in ***`Red`****.* To get the initial values of backward neutral bytes (in `Red`), one only needs to enumerate all possible values of cell $k_3[15]$, and fix values of cells in the main anti-diagonal of $\#\mathbf{SB}^4$. That is due to the following observation. Fixing the values of cells in the first column of $\#\mathbf{MC}^4$ to be 0 (equivalently, fixing the values of cells in the main anti-diagonal of $\#\mathbf{SB}^4$ to be $\mathrm{Sbox}^{-1}[0]$), the values of the first column of $\#\mathbf{SB}^5$ equals that of k_4. Note that the operations of the round function in encryption and key-schedule are exactly the same, excepting the former using AddRoundKey and the latter using AddRoundConstants. Thus, the first anti-diagonal of $\#\mathbf{MC}^5$ will equal that of $\#\mathbf{KMC}^4$. Consequently, the impact brought by adding `Red` cells in the 4^{th} round (k_4) will be canceled in the next round by adding $\#\mathbf{KMC}^4$ without consuming degrees of freedom.

Compute initial values of neutral bytes in ***`Blue`****.* To get the initial values of forward neutral bytes (in `Blue`), one focuses on the constraints among states $\{\#\mathbf{MC}^2, \#\mathbf{AK}^2, \#\mathbf{SB}^3, \#\mathbf{MC}^3, \#\mathbf{AK}^3\}$ (refer to Fig. 3b).

There are five degrees of freedom (in bytes) for the forward (Blue). Using four out of the five, one can keep the same attack complexity because the degree of freedom for backward is the bottleneck. Thus, we fix the value of one Blue cell in $\#\mathbf{AK}^3$, *i.e.*, $\#\mathbf{AK}^3[3]$, as indicated by c.

Note that values of Blue cells are constrained such that they have no impact on the last anti-diagonal of $\#\mathbf{MC}^2$ and the first diagonal of $\#\mathbf{AK}^3$ (are mapped to constants on these state cells). Denote the constants by $\#\mathbf{MC}^2_{\mathtt{C}}[3,6,9,12]$ and $\#\mathbf{AK}^3_{\mathtt{C}}[0,5,10]$. The initial values of forward neutral bytes can be computed using a *local meet-in-the-middle procedure* as shown in Algorithm 1 in the full version [6]. In Algorithm 1 [6], the procedure starts from guessing two free cells in the first column of $\#\mathbf{AK}^3$ and one cell in each of the columns in $\#\mathbf{SB}^3$, computes other undetermined cells column-by-column using predetermined constant-impacts on cells in $\#\mathbf{MC}^2$, and compute back pair-wisely to match at constant cells in $\#\mathbf{AK}^3$. From Algorithm 1 [6], the computational complexity for obtaining the initial values of neutral bytes in Blue is 2^{32} and the memory required is 2^{32} (blocks).

The Main Procedure (refer to Fig. 3b). Assign arbitrary values to those constant impacts of Blue on Red in $\#\mathbf{MC}^2$ (*i.e.*, $\#\mathbf{MC}^2_{\mathtt{C}}[3,6,9,12]$) and to $\#\mathbf{AK}^3_{\mathtt{C}}[3]$. Set $\#\mathbf{AK}^3[0,5,10,12,13,14]$ be 0, $k_3[0,5,10]$ and $\#\mathbf{SB}_4[0,5,10,15]$ be $\mathrm{Sbox}^{-1}[0]$.

Compute the initial values of backward neutral bytes as described above and as shown in Algorithm 1 [6], store the result in $\overrightarrow{\mathcal{T}_{\mathtt{Init}}}$.

1. For a new value of 12 Gray bytes $k_3[1,2,3,4,6,7,8,9,11,12,13,14]$, calculate the values of Gray bytes in k_4 and $\#\mathbf{KMC}^4$ (this is expected to repeat ς^{11} times before terminating from Step 1(b)iiiE according to Sect. 3.4).
   (a) For each value $\overrightarrow{v}_i$ of Blue cells in $\#\mathbf{MC}^3$ stored in $\overrightarrow{\mathcal{T}_{\mathtt{Init}}}$,
      i. start from $\#\mathbf{MC}^3$, compute forward (only cells in Blue) with the values of Gray bytes in k_3 and k_4 to $\#\mathbf{MC}^5$;
      ii. XOR $\#\mathbf{MC}^5$ with the values of Gray bytes in $\#\mathbf{KMC}^4$;
      iii. set the first anti-diagonal of $\#\mathbf{MC}^5$ to be zero, and compute forward to $\#\mathbf{MC}^6$ (without AddRoundKey with k_5 but need to XOR the round constant RC[5], because $\#\mathbf{KMC}^4$ is used instead);
      iv. compute MC($\#\mathbf{MC}^6$) and get the value $\overrightarrow{v}_m$ of the main diagonal
      v. store $\overrightarrow{v}_i$ in a look-up table $\overrightarrow{\mathcal{T}}$ with $\overrightarrow{v}_m$ as index;
   (b) For each value $\overleftarrow{v}_i$ of the Red cell in k_3, with the value of Gray cells,
      i. compute all values of the round keys, *i.e.*, $k_2, k_1, k_0, k_m, k_4, k_5, k_6$, where k_m is the master key;
      ii. set the Red cell $\#\mathbf{AK}^3[15]$ to be $\overleftarrow{v}_i \oplus \mathrm{Sbox}^{-1}[0]$, combine with the constant value in $\#\mathbf{AK}^3$, compute backward up to $\#\mathbf{AK}^0$;
      iii. For each value $\overleftarrow{v}_g$ of the Pink cells in $\#\mathbf{AK}^0$,
         A. with the value of Red cell in the first column of $\#\mathbf{AK}^0$, compute backward through feed-forward, XOR the given target T, compute $\#\mathbf{AK}^6$;
         B. get the value $\overleftarrow{v}_m$ of the main diagonal of $\#\mathbf{AK}^6$.

C. for each value $\overrightarrow{v}_i$ of `Blue` cells in $\#\mathbf{MC}^3$ stored in $\overrightarrow{\mathcal{T}}[\overleftarrow{v}_m]$, combine the values of `Red` cells, restart the computation from $\#\mathbf{AK}^3$ up to $\#\mathbf{AK}^0$;
D. If the newly computed value $\overrightarrow{v}'_g$ of the `Pink` cells in $\#\mathbf{AK}^0$ does not equal $\overleftarrow{v}_g$, go to Step 1(b)iii. Else, compute to $\#\mathbf{AK}^6$, denote the value by $\overleftarrow{v}$.
E. Start from $\#\mathbf{AK}^3$ with the combined knowledge of values of both `Blue` and `Red` cells, compute to $\mathtt{MC}(\#\mathbf{MC}^6)$, if its value $\overrightarrow{v}$ equals $\overleftarrow{v}$, a full-state match is found, output the state $\#\mathbf{SB}^0$ and the key state k_m, and *terminate*. Otherwise, go to Step 1.

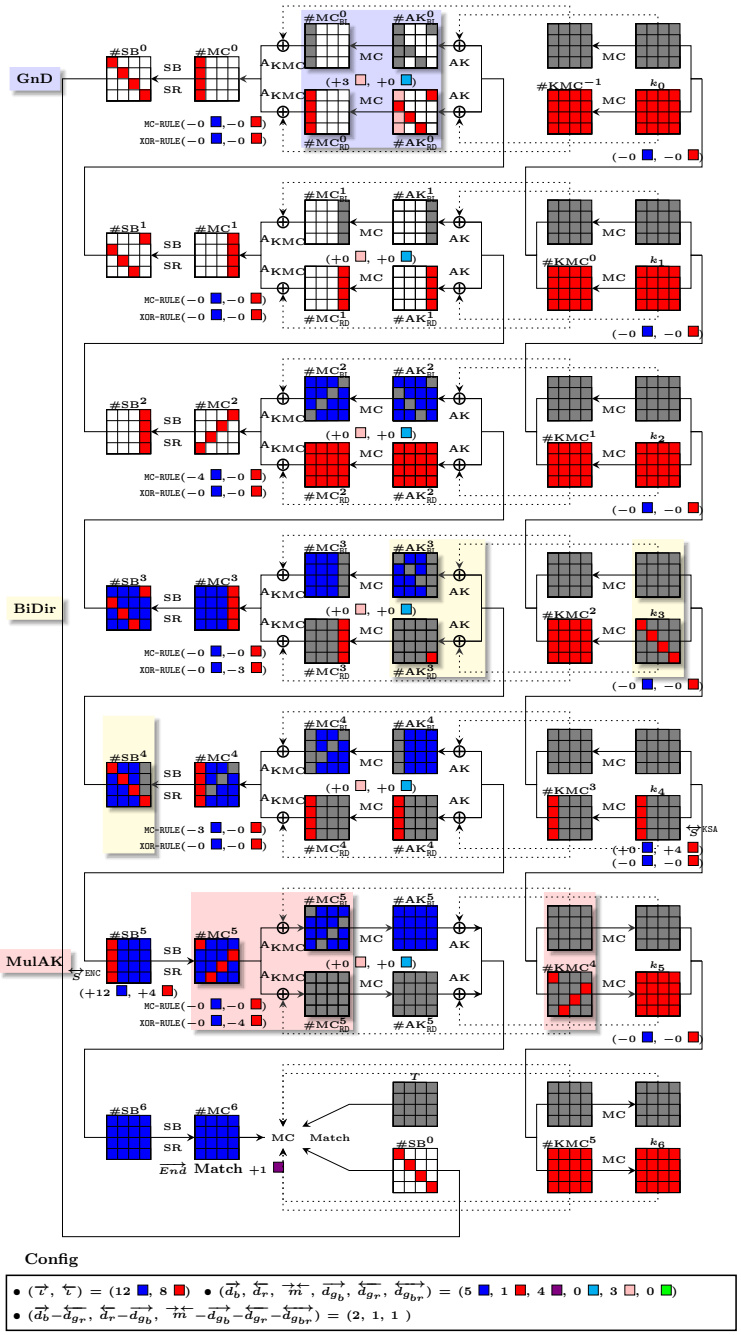
Complexity. As analyzed above, the computational and memory complexity of the precomputation of the initial value of forward neutral bytes is 2^{32}. As for the complexity of the main procedure, the attack configuration on the small version ($n = \mathrm{N}_{\mathtt{row}} \times \mathrm{N}_{\mathtt{col}} = 4 \times 4 = 16$) is, $\overrightarrow{d_b} = 4$, $\overleftarrow{d_r} = 1$, $\overleftarrow{d_{g_r}} = 3$, $\overrightarrow{\leftarrow m} = 4$, $\varsigma = 2^8$, and $\overrightarrow{d_{g_b}} = \overleftrightarrow{d_{g_{br}}} = 0$. According to Eq. (9), the complexity of the whole attack on the small version is therefore $\varsigma^n \cdot (\varsigma^{-(\overleftarrow{d_r} - \overrightarrow{d_{g_b}})} + \varsigma^{-(\overrightarrow{d_b} - \overleftarrow{d_{g_r}})} + \varsigma^{-(\overrightarrow{\leftarrow m} - \overrightarrow{d_{g_b}} - \overleftarrow{d_{g_r}} - \overleftrightarrow{d_{g_{br}}})}) = \varsigma^{16} \cdot \max(\varsigma^{-1} + \varsigma^{-(4-3)} + \varsigma^{-(4-3)}) = \varsigma^{15} = 2^{120}$

Projecting to the real version of Whirlpool ($n = \mathrm{N}_{\mathtt{row}} \times \mathrm{N}_{\mathtt{col}} = 16 \cdot 4 = 64$), the complexity will be $(2^{120})^4 = 2^{480}$. Concretely, the attack configuration will be $\overrightarrow{d_b} = 4 \cdot 4 = 16$, $\overleftarrow{d_r} = 1 \cdot 4 = 4$, $\overleftarrow{d_{g_r}} = 3 \cdot 4 = 12$, $\overrightarrow{\leftarrow m} = 4 \cdot 4 = 16$, $\varsigma = 2^8$, and $\overrightarrow{d_{g_b}} = \overleftrightarrow{d_{g_{br}}} = 0$. Accordingly, the attack complexity on the real version of 7-round Whirlpool is $\varsigma^{64} \cdot \max(\varsigma^{-4} + \varsigma^{-(16-12)} + \varsigma^{-(16-12)}) = 2^{480}$.

The memory required by the main procedure is the memory taken by $\overrightarrow{\mathcal{T}}$, whose size is limited by the degree of freedom for forward, that is 2^{32} for the small version and 2^{128} for the real version.

To further verify the attack and its complexity analysis, we implemented the attack on the small version with 4×4 states. The round functions of encryption and key-schedule of small Whirlpool CF is simulated using the round function of AES. To make the verification practical, the experiments aim for partial matching instead of full-state matching, while preserving the complexity gain. Concretely, the goal is to match m bits with complexity $\max\{2^{32}, 2^{m-8}\}$. Please refer to https://github.com/MITM-AES-like/Whirlpool_7R for results on $m \in \{36, 40, 44, 48\}$.

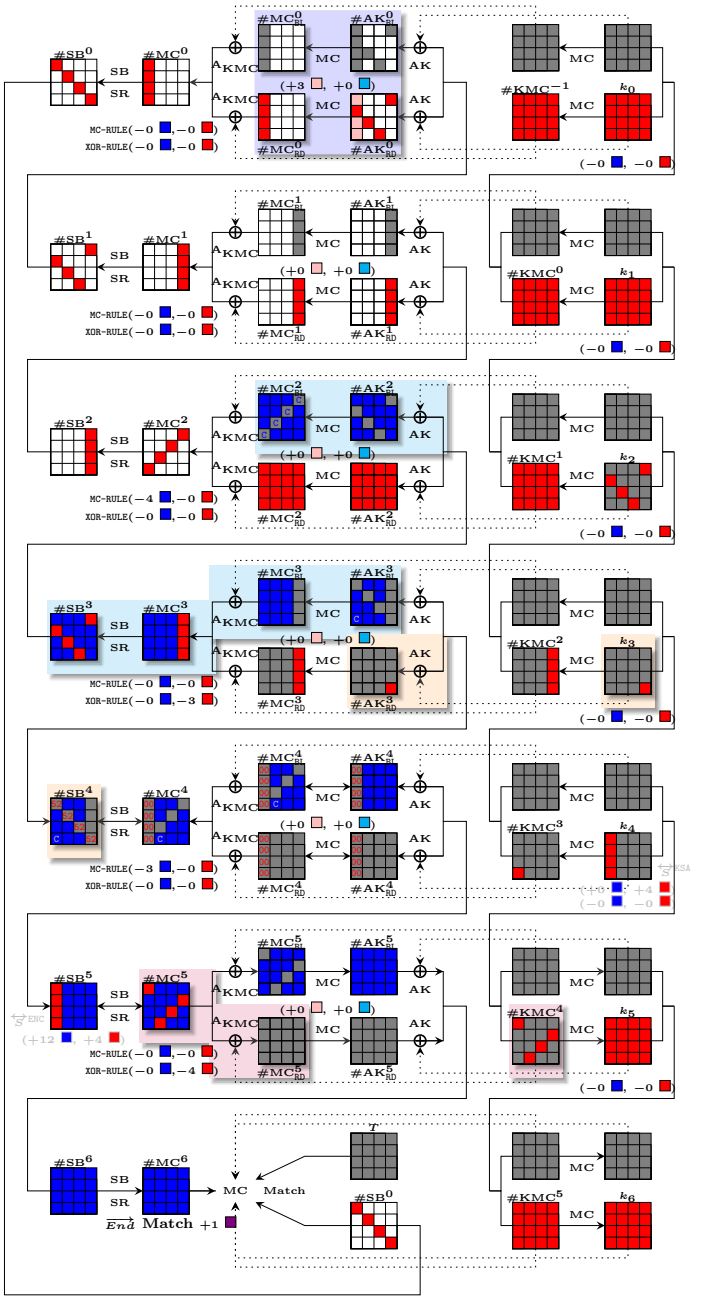
The attack on 6-round Whirlpool. When searching on the full-size version ($\mathrm{N}_{\mathtt{row}} \times \mathrm{N}_{\mathtt{col}} = 8 \times 8$), results with asymmetric patterns were found. One example is depicted in Fig. 12 [6]. Following the configuration, one can devise an attack on 6-round of Whirlpool with better complexity than previous ones. The concrete attack configuration is $\overrightarrow{d_b} = 9$, $\overleftarrow{d_r} = 24$, $\overrightarrow{d_{g_b}} = 15$, $\overrightarrow{\leftarrow m} = 24$, $\varsigma = 2^8$, and $\overleftarrow{d_{g_r}} = \overleftrightarrow{d_{g_{br}}} = 0$. Accordingly, the attack complexity on the full-size version of 6-round Whirlpool is $\varsigma^{64} \cdot \max(\varsigma^{-24-15} + \varsigma^{-(9)} + \varsigma^{-(24-15)}) = 2^{440}$. The memory required is $2^{24 \times 8} = 2^{192}$. The procedures to compute the initial values of neutral bytes for both directions are relatively simpler than that of the above attack on



(a) Configuration automatically found

This implies an attack on the full version (8×8) as shown in Fig. 11 with configuration $((\overrightarrow{d_b} - \overleftarrow{d_{g_r}},\ \overleftarrow{d_r} - \overrightarrow{d_{g_b}},\ \overrightarrow{\overleftarrow{m}} - \overrightarrow{d_{g_b}} - \overleftarrow{d_{g_r}} - \overleftrightarrow{d_{g_{br}}}) = (8,\ 4,\ 4))$

Fig. 3: An example of using $(2 \times 2$ of $4 \times 4)$ to search the MITM attack on 7-round Whirlpool and an equivalent configuration used in the experimental verification



(b) Equivalent configuration used in the experimental verification

Fig. 3: An example of using $(2 \times 2$ of $4 \times 4)$ to search the MITM attack on 7-round Whirlpool and an equivalent configuration used in the experimental verification (cont.)

7-round. That is because, the cancellation constraints on the Blue is at a single point, *i.e.*, $(\#\mathbf{MC}^2, \#\mathbf{AK}^2)$. The cancellation constraints on the Red is also at a single point, *i.e.*, $(\#\mathbf{MC}^4, \#\mathbf{AK}^4)$. As for such constraints, column-by-column independent computations can be used to derive the initial values of neutral bytes for both directions.

Conversion from Pseudo-Preimage to Preimage Attacks has been discussed in previous works. Here, we follow the *Two Types of Last Block Attacks* from [28,30]. Denote the time complexity of inverting the reduced-Whirlpool compression function as 2^{ℓ}. A random message fulfills the padding rule of Whirlpool with probability 2^{-9}, hence it costs $2^{\ell+9}$ to find a right last block. Then an unbalanced meet-in-the-middle is carried out between the initial value and the input chaining value of the last block, which costs $2^{(512+\ell)/2}$ and sums to $2^{(512+\ell)/2} + 2^{\ell+9}$ to find a long and full preimage. Detailed conversion results are summarized in Table 1. We note further complexity optimizations are possible, by finding pseudo-preimages under multi-target scenarios and utilizing them in the "unbalanced meet-in-the-middle" phase as discussed in [15].

4.2 Discussions on the New Attacks

The previous best attack on Whirlpool is up to 6-round [28]. In the 6-round attack in [28], although the freedom degree in the key states is exploited, the computational chunks between the key schedule and the encryption data-path are designed to be almost identical due to the limitation of manual analysis. In contrast, in the new attack, the computational chunks between the key schedule and the encryption data-path are largely different. Thus, the degree of freedom in Blue and Red can be relatively more balanced, and the required number of guessing bytes is relatively less. Consequently, the complexity is better.

Attacking one more round than the previously best attacked, new strategies are required on top of those appeared in [28]. In the new 7-round attack, apart from GnD, flexibly using of equivalent round-keys $\#\mathbf{KMC}$ or the real round keys k (MulAK) is critical to save degrees of freedom. Moreover, complex non-linear constraints must be imposed on Blue cells to let Blue propagates towards backward besides forward (BiDir), and then cancel their impacts on Red-attribute propagation. Thus, efficient procedure (*e.g.*, the *local meet-in-the-middle procedure*) for obtaining initial values of neutral cells fulfilling the non-linear constraints is necessary here.

5 Application to Preimage Attacks on Grøstl

Grøstl, proposed by Gauravaram *et al.* in [13], is one of the five finalists of SHA-3 competition hosted by NIST. Grøstl adopts a double-pipe design, *i.e.*, the size of the chaining value, which is $2n$-bit, is twice as the hash size, which is n-bit. For Grøstl-256, the hash size is 256 bits, and for Grøstl-512, it is 512 bits. Two $2n$-bit AES-like permutations P and Q are employed to build the CF and OT.

5.1 New Attacks Resulted from Applying the MILP Modeling

Applying the MILP modeling approach described in Sect. 3 on the OT (and on the $P(H_i) \oplus H_i$ part of the CF) of Grøstl-256 and Grøstl-512, new attacks are found on 6-round and 8-round, respectively. Besides, many efficient attacks on shorter rounds are found.

For the 6-round attack on the OT of Grøstl-256, GnD is the essential that enables to cover one more attacked round than previous 5-round attacks [30]; BiDir is the essential that enables better complexity than previous 6-round attacks [22]. Besides, many inferior attacks than the presented best one on 6-round Grøstl-256 are found. For those inferior attacks, the computation of initial values of neutral bytes is relatively easier, but the complexity of the entire attack is higher. Thus, a non-trivial procedure to compute the initial structure (initial values of neutral bytes) is essential for achieving the best complexity.

For the 8-round attack on the OT of Grøstl-512, GnD is the essential that enables to achieve better complexity than the previous 8-round attack [30]. Besides, compared to the 8-round attack in [30], in the presented attack, the initial structure covers one more round (4 rounds) by allowing one attribute-propagation conceding to the opposite attribute-propagation in both directions.

The configurations of various attacks can be found in Figures 5, 6, 7, 8, 16, 17, 14, 15 in the full version [6]. The concrete attack procedures on 6-round OT of Grøstl-256 and 8-round OT of Grøstl-512 are presented in Sect. C [6].

Conversion to Pseudo-Preimage Attacks. The attacks on the OT of Grøstl can be converted into pseudo-preimage attacks on Grøstl combining with similar attacks on the $P(H) \oplus H$ of the CF using the conversion method in [30]. The complexity of the converted pseudo-preimage attacks are summarized in Table 1. More details about the conversion can be found in Sect. F of the full version [6].

5.2 Discussions on the New Attacks

An interesting feature of the best attack on 6-round Grøstl-256 depicted in Fig. 5 [6] is that, with necessary guessing, the computation of `Blue` covers the full 6-round. That is, the `Blue` propagates to backward besides forward and contributes to degrees of matching from both sides. Besides, like the attack on 7-round Whirlpool, it requires a non-trivial local meet-in-the-middle procedure to compute initial values of neutral bytes.

Note that, obtaining the previous best attacks on Grøstl, the work in [30] has already been assisted with automatic searching, and the 5-round attack on Grøstl-256 in [30] was claimed to be optimal. However, the optimality hold only in a restricted search space, apart from lacking the consideration of the GnD technique. In contrast, the presented 5-round attacks on Grøstl-256 in Figures 14 and 15 [6] achieve better complexity than that in [30] and is optimal in an expanded search space where BiDir and GnD are considered.

6 Applications to Collision and Key-recovery Attacks

The MILP models for searching for the best preimage attacks can be directly transformed to search for the best collision attacks on hash functions and key-recovery attacks on block ciphers. For collision attacks, according to the analysis in Sect. 3.5, what needs to be done is to simply restrict the matching point to be at the last round and add constraints shown in Eq. 13. Applying to Grøstl's OT, AES-hashing, Kiasu-BC-hashing [7], new and improved attacks were found.

For key-recovery attacks, upon the MILP models for preimage attack, one simply needs to constraint that the degrees of freedom in both forward and backward only source from the key states, relax the degrees of matching such that it is not included in the objective but simply be non-zero [8], and constraint that the plaintext or ciphertext contains only `Red` and `Gray` cells or only `Blue` and `Gray` cells. Besides, the objective can be set to maximize the number of `Gray` cells in the plaintext or ciphertext, which can optimize data complexity. Applying to SKINNY-n-$3n$, improvements in terms of time complexity (see Fig. 23 of the full version [6]) or data complexity (see Fig. 24 [6]) upon the attack in [11] on 23-round reduced version were obtained. In Sect. F of the full version, visualizations of representative attack configurations are presented. The complexity of the attacks are summarized in Table 1.

Acknowledgements

We thank anonymous reviewers for their valuable comments. This research is partially supported by the National Natural Science Foundation of China (Grants No. 62172410, 61802400, 61732021, 61802399, 61961146004), the National Key R&D Program of China (Grants No. 2018YFA0704704, 2018YFA0704701), the Youth Innovation Promotion Association of Chinese Academy of Sciences; Nanyang Technological University in Singapore under Grant 04INS000397C230, Ministry of Education in Singapore under Grants RG91/20 and MOE2019-T2-1-060; the Gopalakrishnan – NTU Presidential Postdoctoral Fellowship 2020; the Major Program of Guangdong Basic and Applied Research (Grant No. 2019B030302008), Shandong Province Key R&D Project (Nos. 2020ZLYS09 and 2019JZZY010133).

References

1. Alliance, ZigBee. ZigBee 2007 specification. *Online: http://www.zigbee.org/*, 2007.
2. K. Aoki and Y. Sasaki. Preimage attacks on one-block MD4, 63-step MD5 and more. In R. M. Avanzi, L. Keliher, and F. Sica, editors, *SAC 2008*, volume 5381 of *LNCS*, pages 103–119. Springer, Heidelberg, Aug. 2009.

[7] Kiasu-BC [18] is a tweakable block cipher, the only difference with AES-128 is XORing a 64-bit tweak value to the first two rows of the state after each AddRoundKey.

[8] In MITM key-recovery attack, the degree of matching can be efficiently increased using simultaneous matching with multiple plaintext/ciphertext pairs [12].

3. J.-P. Aumasson, W. Meier, and F. Mendel. Preimage attacks on 3-pass HAVAL and step-reduced MD5. In R. M. Avanzi, L. Keliher, and F. Sica, editors, *SAC 2008*, volume 5381 of *LNCS*, pages 120–135. Springer, Heidelberg, Aug. 2009.
4. Z. Bao, L. Ding, J. Guo, H. Wang, and W. Zhang. Improved meet-in-the-middle preimage attacks against AES hashing modes. *IACR Trans. Symm. Cryptol.*, 2019(4):318–347, 2019.
5. Z. Bao, X. Dong, J. Guo, Z. Li, D. Shi, S. Sun, and X. Wang. Automatic search of meet-in-the-middle preimage attacks on AES-like hashing. In A. Canteaut and F.-X. Standaert, editors, *EUROCRYPT 2021, Part I*, volume 12696 of *LNCS*, pages 771–804. Springer, Heidelberg, Oct. 2021.
6. Z. Bao, J. Guo, D. Shi, and Y. Tu. Superposition meet-in-the-middle attacks: Updates on fundamental security of AES-like hashing. Cryptology ePrint Archive, Report 2021/575, 2021. https://eprint.iacr.org/2021/575.
7. P. S. L. M. Barreto and V. Rijmen. The WHIRLPOOL Hashing Function. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.529.3184&rep=rep1&type=pdf, 2000. Revised in 2003.
8. A. Bogdanov and C. Rechberger. A 3-subset meet-in-the-middle attack: Cryptanalysis of the lightweight block cipher KTANTAN. In A. Biryukov, G. Gong, and D. R. Stinson, editors, *SAC 2010*, volume 6544 of *LNCS*, pages 229–240. Springer, Heidelberg, Aug. 2011.
9. C. Bouillaguet, P. Derbez, and P.-A. Fouque. Automatic search of attacks on round-reduced AES and applications. In P. Rogaway, editor, *CRYPTO 2011*, volume 6841 of *LNCS*, pages 169–187. Springer, Heidelberg, Aug. 2011.
10. J. Daemen and V. Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard*. Information Security and Cryptography. Springer, 2002.
11. X. Dong, J. Hua, S. Sun, Z. Li, X. Wang, and L. Hu. Meet-in-the-middle attacks revisited: Key-recovery, collision, and preimage attacks. In T. Malkin and C. Peikert, editors, *CRYPTO 2021, Part III*, volume 12827 of *LNCS*, pages 278–308, Virtual Event, Aug. 2021. Springer, Heidelberg.
12. T. Fuhr and B. Minaud. Match box meet-in-the-middle attack against KATAN. In C. Cid and C. Rechberger, editors, *FSE 2014*, volume 8540 of *LNCS*, pages 61–81. Springer, Heidelberg, Mar. 2015.
13. P. Gauravaram, L. R. Knudsen, K. Matusiewicz, F. Mendel, C. Rechberger, M. Schläffer, and S. S. Thomsen. Grøstl a SHA-3 candidate. http://www.groestl.info/Groestl.pdf, March 2011.
14. H. Gilbert and T. Peyrin. Super-sbox cryptanalysis: Improved attacks for AES-like permutations. In S. Hong and T. Iwata, editors, *FSE 2010*, volume 6147 of *LNCS*, pages 365–383. Springer, Heidelberg, Feb. 2010.
15. J. Guo, S. Ling, C. Rechberger, and H. Wang. Advanced meet-in-the-middle preimage attacks: First results on full Tiger, and improved results on MD4 and SHA-2. In M. Abe, editor, *ASIACRYPT 2010*, volume 6477 of *LNCS*, pages 56–75. Springer, Heidelberg, Dec. 2010.
16. A. Hosoyamada and Y. Sasaki. Finding hash collisions with quantum computers by using differential trails with smaller probability than birthday bound. In A. Canteaut and Y. Ishai, editors, *EUROCRYPT 2020, Part II*, volume 12106 of *LNCS*, pages 249–279. Springer, Heidelberg, May 2020.
17. ISO/IEC. 10118-2:2010 Information technology Security techniques – Hash-functions – Part 2: Hash-functions using an n-bit block cipher. 3rd ed., International Organization for Standardization, Geneve, Switzerland, October, 2010.

18. J. Jean, I. Nikolic, and T. Peyrin. Tweaks and keys for block ciphers: The TWEAKEY framework. In P. Sarkar and T. Iwata, editors, *ASIACRYPT 2014, Part II*, volume 8874 of *LNCS*, pages 274–288. Springer, Heidelberg, Dec. 2014.
19. M. Lamberger, F. Mendel, C. Rechberger, V. Rijmen, and M. Schläffer. Rebound distinguishers: Results on the full Whirlpool compression function. In M. Matsui, editor, *ASIACRYPT 2009*, volume 5912 of *LNCS*, pages 126–143. Springer, Heidelberg, Dec. 2009.
20. G. Leurent. MD4 is not one-way. In K. Nyberg, editor, *FSE 2008*, volume 5086 of *LNCS*, pages 412–428. Springer, Heidelberg, Feb. 2008.
21. J. Li, T. Isobe, and K. Shibutani. Converting meet-in-the-middle preimage attack into pseudo collision attack: Application to SHA-2. In A. Canteaut, editor, *FSE 2012*, volume 7549 of *LNCS*, pages 264–286. Springer, Heidelberg, Mar. 2012.
22. B. Ma, B. Li, R. Hao, and X. Li. Improved (pseudo) preimage attacks on reduced-round GOST and Grøstl-256 and studies on several truncation patterns for AES-like compression functions. In K. Tanaka and Y. Suga, editors, *IWSEC 15*, volume 9241 of *LNCS*, pages 79–96. Springer, Heidelberg, Aug. 2015.
23. B. Preneel, R. Govaerts, and J. Vandewalle. Hash functions based on block ciphers: A synthetic approach. In D. R. Stinson, editor, *CRYPTO'93*, volume 773 of *LNCS*, pages 368–378. Springer, Heidelberg, Aug. 1994.
24. Y. Sasaki. Meet-in-the-middle preimage attacks on AES hashing modes and an application to Whirlpool. In A. Joux, editor, *FSE 2011*, volume 6733 of *LNCS*, pages 378–396. Springer, Heidelberg, Feb. 2011.
25. Y. Sasaki and K. Aoki. Preimage attacks on 3, 4, and 5-pass HAVAL. In J. Pieprzyk, editor, *ASIACRYPT 2008*, volume 5350 of *LNCS*, pages 253–271. Springer, Heidelberg, Dec. 2008.
26. Y. Sasaki and K. Aoki. Preimage attacks on step-reduced MD5. In Y. Mu, W. Susilo, and J. Seberry, editors, *ACISP 08*, volume 5107 of *LNCS*, pages 282–296. Springer, Heidelberg, July 2008.
27. Y. Sasaki and K. Aoki. Finding preimages in full MD5 faster than exhaustive search. In A. Joux, editor, *EUROCRYPT 2009*, volume 5479 of *LNCS*, pages 134–152. Springer, Heidelberg, Apr. 2009.
28. Y. Sasaki, L. Wang, S. Wu, and W. Wu. Investigating fundamental security requirements on Whirlpool: Improved preimage and collision attacks. In X. Wang and K. Sako, editors, *ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 562–579. Springer, Heidelberg, Dec. 2012.
29. S. Sun, L. Hu, P. Wang, K. Qiao, X. Ma, and L. Song. Automatic security evaluation and (related-key) differential characteristic search: Application to SIMON, PRESENT, LBlock, DES(L) and other bit-oriented block ciphers. In P. Sarkar and T. Iwata, editors, *ASIACRYPT 2014, Part I*, volume 8873 of *LNCS*, pages 158–178. Springer, Heidelberg, Dec. 2014.
30. S. Wu, D. Feng, W. Wu, J. Guo, L. Dong, and J. Zou. (Pseudo) preimage attack on round-reduced Grøstl hash function and others. In A. Canteaut, editor, *FSE 2012*, volume 7549 of *LNCS*, pages 127–145. Springer, Heidelberg, Mar. 2012.
31. B. Zhang and D. Feng. New guess-and-determine attack on the self-shrinking generator. In X. Lai and K. Chen, editors, *ASIACRYPT 2006*, volume 4284 of *LNCS*, pages 54–68. Springer, Heidelberg, Dec. 2006.
32. J. Zou, W. Wu, S. Wu, and L. Dong. Improved (Pseudo) Preimage Attack and Second Preimage Attack on Round-Reduced Grostl Hash Function. *J. Inf. Sci. Eng.*, 30(6):1789–1806, 2014.