Cryptogr. Commun. (2018) 10:755–776
https://doi.org/10.1007/s12095-018-0286-5

Statistical integral distinguisher with multi-structure and its application on AES-like ciphers

Tingting Cui[1,2] · Huaifeng Chen[1] · Sihem Mesnager[3] · Ling Sun[1] · Meiqin Wang[1]

Received: 29 June 2017 / Accepted: 21 February 2018 / Published online: 3 March 2018

Abstract Integral attack is one of the most powerful tools in the field of symmetric ciphers. In order to reduce the time complexity of original integral one, Wang et al. firstly proposed a statistical integral distinguisher at FSE'16. However, they don't consider the cases that there are several integral properties on output and multiple structures of data should be used at the same time. In terms of such cases, we put forward a new statistical integral distinguisher, which enables us to reduce the data complexity comparing to the traditional integral ones under multiple structures. As illustrations, we use it into the known-key distinguishers on AES-like ciphers including AES and the permutations of Whirlpool, PHOTON and Grøstl-256 hash functions based on the Gilbert's work at ASIACRYPT'14. These new distinguishers are the best ones comparing with previous ones under known-key setting. Moreover, we propose a secret-key distinguisher on 5-round AES under chosen-ciphertext

This article is part of the Topical Collection on *Special Issue on Statistics in Design and Analysis of Symmetric Ciphers*

This is an extended version of [6] presented at ACISP 2017. In [6] we proposed a statistical integral distinguisher with multiple structures model and used it directly into known-key distinguishers on AES. In this paper, besides the content of [6], we generalize the known-key distinguisher on AES-like cipher in Section 4, and apply it not only on AES but also on other AES-like ciphers such as Whirlpool, PHOTON and Grøstl-256 in Section 5. The construction of the whole paper is changed comparing with [6].

✉ Meiqin Wang
mqwang@sdu.edu.cn

[1] Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Shandong University, Jinan 250100, China

[2] Nanyang Technological University, Singapore, Singapore

[3] Department of Mathematics, University of Paris VIII, Saint-Denis, France

mode. Its data, time and memory complexities are $2^{114.32}$ chosen ciphertexts, 2^{110} encryptions and $2^{33.32}$ blocks. This is the best integral distinguisher on AES with secret S-box under secret-key setting so far.

Keywords Statistical integral model · Multi-structure · Secret S-box · Secret-key · Known-key · AES-like cipher

Mathematics Subject Classification (2010) 94-XX · 94A60

1 Introduction

Integral attack is an important cryptanalytic technique for symmetric-key ciphers, which was firstly put forward by Daemen et al. in [7], then unified as integral attack by Knudsen and Wagner in [18]. In an integral distinguisher, in order to distinguish an actual cipher from any random permutation, one chooses plaintexts by taking all values of part bits and fixing other bits as constant to check whether the values on partial bits of the ciphertext are uniformly distributed (integral property) or not. If one additional linear layer is considered, the property will be that the XOR of all possible values on the specific part bits of ciphertext becomes zero, which is referred as zero-sum property [2]. With the purpose of reducing the data complexity, Wang et al. applied statistical techniques on original integral distinguisher to build a statistical integral distinguisher at FSE'16 in [25]. As a result, this statistical integral distinguisher requires less data complexity than that of the original integral one. However, Wang et al. only considered the case that there was only one integral property on output, they didn't discuss the cases that several integral properties existed on output and multiple structures of data should be used at the same time. These limit the effect of certain integral attacks on block ciphers, especially for known-key distinguishing attack. We want to extend the statistical integral distinguisher for more situations. This is the first motivation for this work.

Block ciphers, because of their security and simplicity, are often adopted as components of hash functions by designers, such as Whirlpool [3] and PHOTON [14]. Since the attacker can fully control the inter behaviour of a hash function, the block cipher used in it shall resist known-key and chosen-key attacks. The first known-key security model was proposed by Knudsen and Rijmen for block cipher in [17] where the secret key is known to the attacker. Its goal is to distinguish the block cipher from a random permutation by constructing a set of plaintext/ciphertext pairs satisfying a special property. Such property is easy to check but impossible to achieve for any random permutation with the same complexity and a non-negligible probability by using oracle accesses to this random permutation and its inverse. Since the establishment of known-key model, several types of known-key distinguishers have been proposed, such as distinguishers with integral property [1, 10, 17, 22], subspace distinguishers [19, 20], (multiple) limited-birthday distinguishers [11, 16], and the known-key distinguisher for PRESENT by combining meet-in-the-middle technique with truncated differential [5]. Moreover, the chosen-key distinguishing attack on the full AES-256 has been proposed in [4].

The best known-key distinguisher on AES was proposed by Gilbert at ASIACRYPT'14 in [10]. He utilized 2^{64} middle-texts belonging to 2^{32} structures and integral property to put forward a known-key distinguisher on full-round AES-128. If we can use the statistical integral property into this known-key distinguisher, we can reduce the time complexity so

as to improve it. Furthermore, we can implement improved known-key distinguishers on all AES-like ciphers which adopt the wide trail strategy as AES. This is the second motivation of this work.

Our Contributions In this paper, we propose a statistical integral distinguisher with multiple structures on input and integral properties on output. We found that in some situations of integral attacks such as known-key distinguishing attack on AES, multiple structures of inputs have to be used. For instance, N_s structures of inputs are needed. In each structure, all 2^s possible values on s input bits (out of n bits) are taken and the corresponding outputs on b different t bits are uniformly distributed respectively. The statistical integral distinguisher in [25] can reduce the data complexity from $\mathcal{O}(2^s)$ to $\mathcal{O}(2^{s-t/2})$ by using one t-bit integral property if only one structure is used. However, if there are N_s structures involved, the model in [25] cannot be applied. For the sake of reducing the data requirements, we construct a new statistical integral distinguisher. In our new distinguisher, the data complexity is

$$\mathcal{O}(\sqrt{N_s/b} \cdot 2^{s-\frac{t}{2}}),$$

while the data complexity of the original distinguisher is

$$\mathcal{O}(N_s \cdot 2^s).$$

In order to verify our theoretical model, we implement the experiments for mini version of AES. It shows that the experimental results are in good accordance with the theoretical results.

As an application, we put our statistical integral method into known-key distinguishing attack and generalize the new known-key distinguishers on 8-round and 10-round AES-like ciphers. Taking AES, the permutations of Whirlpool, PHOTON and Grøstl as examples (regard these AES-like permutations as block ciphers), we show the actual distinguishers and compare them with previous results. As far as we know, our distinguishers on AES-like ciphers are the best ones under known-key setting. The results are summarized in Table 1.

Besides that, we propose the improved secret-key integral distinguisher on AES with secret S-box, which has a great improvement comparing with the previous integral one proposed by Sun et al. at CRYPTO'16 in [23]. The results are summarized in Table 2.

Outline of This Paper In Section 2, some preliminaries are given. Then we present a statistical integral model with multiple structures on input and several integral properties on output in Section 3. In Section 4 new general known-key distinguishers on AES-like ciphers are given and are applied on several actual ciphers in Section 5. Then a new secret-key statistical integral distinguisher on AES is put forward in Section 6. At last, we conclude this paper in Section 7.

2 Preliminaries

2.1 Notations

We define some notations in this part which are used through this paper.

(1) $A \diamond B$: implement A operation firstly, then B operation;
(2) $A \circ B$: implement B operation firstly, then A operation;
(3) $X_{(i)}$: the i-th cell of state X;
(4) $X_{(i\sim j)}$: the cells of X from the i-th one to j-th one.

Table 1 Summary of KK and CK distinguishers on AES-like Ciphers. Note that we only consider the security of the AES-like permutation used in hash function and such permutation is regarded as block cipher

Cipher	Type	Rounds	Time	Memory	Source
AES	KK dist.	7	2^{56}	–	[17]
	KK dist.	7	2^{24}	–	[21]
	KK dist.	8	2^{48}	2^{35}	[11]
	KK dist.	8	2^{44}	2^{35}	[16]
	KK dist.	8	2^{64}	–	[10]
	KK dist.	**8**	$\mathbf{2^{42.61}}$	$\mathbf{2^{13}}$	[6], **Section 5.1**
	KK dist.	10	2^{64}	–	[10]
	KK dist.	**10**	$\mathbf{2^{59.60}}$	$\mathbf{2^{58.84}}$	[6], **Section 5.1**
Whirlpool	KK dist.	8	2^{122}	2^{67}	[20]
	KK dist.	**8**	$\mathbf{2^{85.31}}$	$\mathbf{2^{16.32}}$	**Section 5.2**
	CK dist.	10	2^{188}	2^{11}	[19]
	CK dist.	10	$2^{115.7}$	2^{11}	[16]
	KK dist.	**10**	$\mathbf{2^{113.90}}$	$\mathbf{2^{102.92}}$	**Section 5.2**
Grøstl-256	KK dist.	9	2^{368}	2^{67}	[15]
	KK dist.	9	2^{362}	2^{67}	[16]
	KK dist.	**10**	$\mathbf{2^{113.90}}$	$\mathbf{2^{102.92}}$	**Section 5.3**
PHOTON-80/20/16	KK dist.	8	2^{8}	$2^{5.58}$	[14]
	KK dist.	8	$2^{3.4}$	$2^{5.58}$	[16]
	KK dist.	**10**	$\mathbf{2^{40.71}}$	$\mathbf{2^{37.00}}$	**Section 5.4**
PHOTON-128/16/16	KK dist.	8	2^{8}	$2^{5.58}$	[14]
	KK dist.	8	$2^{2.8}$	$2^{5.58}$	[16]
	KK dist.	**10**	$\mathbf{2^{49.80}}$	$\mathbf{2^{40.50}}$	**Section 5.4**
PHOTON-160/36/36	KK dist.	8	2^{8}	2^{6}	[14]
	KK dist.	8	2^{2}	2^{6}	[16]
	KK dist.	**10**	$\mathbf{2^{58.94}}$	$\mathbf{2^{43.77}}$	**Section 5.4**
PHOTON-224/32/32	KK dist.	9	2^{184}	2^{34}	[15]
	KK dist.	9	2^{178}	2^{34}	[16]
	KK dist.	**10**	$\mathbf{2^{68.09}}$	$\mathbf{2^{46.96}}$	**Section 5.4**
PHOTON-256/32/32	KK dist.	8	2^{16}	$2^{10.58}$	[14]
	KK dist.	8	$2^{10.8}$	$2^{10.58}$	[16]
	KK dist.	**10**	$\mathbf{2^{96.59}}$	$\mathbf{2^{70.46}}$	**Section 5.4**

CK dist.: Chosen-key distinguisher;

KK dist.: Known-key distinguisher

Bold emphasis means this is our new work in this paper

Table 2 Summary of secret-key distinguishers on AES

Type	Rounds	Data	Time	Memory	Source
Integral	5	2^{128} CC	2^{128}	–	[23]
Impossible differential	5	$2^{98.2}$ CP	$2^{103.2}$	–	[12]
Subspace trail	5	2^{32} CP	$2^{35.6}$	2^{36}	[13]
Statistical integral	**5**	$\mathbf{2^{114.32}}$**CC**	$\mathbf{2^{110}}$	$\mathbf{2^{33.32}}$	[6], **Section 6**

CC: Chosen-cipertext; CP: Chosen-Plaintext

Bold emphasis means this is our new work in this paper

To describe the integral distinguishers, we use C, A, $(A_1, A_2, \ldots, A_n)$ and $(A_1^j, A_2^j, \ldots, A_n^j)$ to denote different properties as follows.

(1) C: a constant byte where all values in a set are fixed as one constant;
(2) A: a active byte where all values in a set are distributed uniformly;
(3) $(A_1, A_2, \ldots, A_n)$: n active bytes where all values in a set are not only distributed uniformly in each byte but also distributed uniformly on n bytes;
(4) $(A_1^j, A_2^j, \ldots, A_n^j)$: n active bytes in one column of a state where all values in a set are not only distributed uniformly in each byte but also distributed uniformly on n bytes.

2.2 Description of AES-like cipher

AES-like cipher adopts wide trail strategy which is used in the design of AES cipher [8]. Such cipher is constructed by iterating Even-Mansour construction for target rounds. Its internal state can be viewed as square matrix of m rows and m columns. Each of these m^2 cells is c bits. The round function includes 4 components:

– AddRoundKey (AK): Xor with a subkey, sometimes a constant.
– SubBytes (SB): A nonlinear bijective mapping for each cell of state;
– ShiftRows (SR): Left rotate the i-th row by i cells, where $i = 0, 1, \ldots, m-1$;
– MixColumns (MC): Left multiply with an MDS matrix on each column;

So a r-round AES-like cipher can be described as follows:

$$AES_r^{like} = (AK \diamond SB \diamond SR \diamond MC)^r \diamond AK. \tag{1}$$

Usually the MC operation is omitted in the last round.

Note that this description for AES-like cipher follows the design of AES. Sometimes a cipher is regarded as an AES-like cipher as well if its SR operation is slightly modified and MC operation does not adopt an MDS matrix. In our paper, we also consider the ciphers whose SR and MC are replaced by ShiftColumns (SC) and MixRows (MR) as AES-like ciphers.

In [10], Gilbert proposed a new representation of AES. Firstly, he defined two operations T and CC as follows, then built two special byte permutations $P = SR \diamond T \diamond SR^{-1}$ and $Q = SR^{-1} \diamond T \diamond SR \diamond CC$. With these two permutations, Gilbert proposed a new observation as follows.

Observation 1 (From [10]) Transformations $S = Q^{-1} \diamond SB \diamond MC \diamond AK \diamond SB \diamond P^{-1}$ and $R = P \diamond SR \diamond MC \diamond AK \diamond SR \diamond Q$ operate on columns and rows respectively.

$$T : \begin{pmatrix} a_0 & a_4 & a_8 & a_{12} \\ a_1 & a_5 & a_9 & a_{13} \\ a_2 & a_6 & a_{10} & a_{14} \\ a_3 & a_7 & a_{11} & a_{15} \end{pmatrix} \mapsto \begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 & a_7 \\ a_8 & a_9 & a_{10} & a_{11} \\ a_{12} & a_{13} & a_{14} & a_{15} \end{pmatrix}$$

$$CC : \begin{pmatrix} a_0 & a_4 & a_8 & a_{12} \\ a_1 & a_5 & a_9 & a_{13} \\ a_2 & a_6 & a_{10} & a_{14} \\ a_3 & a_7 & a_{11} & a_{15} \end{pmatrix} \mapsto \begin{pmatrix} a_0 & a_{12} & a_8 & a_4 \\ a_1 & a_{13} & a_9 & a_5 \\ a_2 & a_{14} & a_{10} & a_6 \\ a_3 & a_{15} & a_{11} & a_7 \end{pmatrix}$$

With this new observation, r-round AES has three equivalent representations:

$$\begin{aligned} AES_r &= AK \diamond SR \diamond Q \diamond (S \diamond R)^{r/2-1} \diamond S \diamond P \diamond SR \diamond AK, && (2)\\ &= AK \diamond P^{-1} \diamond SB \diamond R \diamond (S \diamond R)^{r/2-1} \diamond SB \diamond Q^{-1} \diamond AK, && (3)\\ &= AK \diamond SB \diamond SR \diamond MC \diamond AES_{r/2-2} \diamond AK^{-1} \diamond MC \diamond AK \diamond SB \diamond SR \diamond AK. && (4) \end{aligned}$$

Easily, these representations can be directly applied on AES-like ciphers by choosing suitable T and CC operations. But note that the MC operation in the last round function is omitted in above three representations.

2.3 Brief description of known-key distinguishers on AES in [10]

In this subsection, we briefly recall the known-key distinguishers for 8-round and 10-round AES proposed by Gilbert at ASIACRYPT'14 [10].

In order to mount a known-key distinguisher for AES_8, Gilbert firstly proposed two integral distinguishers shown in Fig. 1. Then given 2^{64} data $\mathcal{Z} = \{R(x,0,0,0) \oplus (y,0,0,0) | x, y \in \{0,1\}^{32}\}$, this set $\mathcal{Z}$ can be divided into 2^{32} structures according to different values of x, and each structure takes all 2^{32} values on the first column and constants on other columns. So the set $\mathcal{Z}$ satisfies the first integral distinguisher in Fig. 1. Since R operation is an affine mapping, $R^{-1}(\mathcal{Z}) = \{(x,0,0,0) \oplus R^{-1}(y,0,0,0)\}$ can be divided into 2^{32} structures according to different values of y, thus the set $R^{-1}(\mathcal{Z})$ satisfies the second integral distinguisher in Fig. 1.

Combining these two integral distinguishers with R operation, a known-key distinguisher on AES_8 is built so that all input and output bytes resulted from 2^{64} middle texts $\mathcal{Z}$ are uniformly distributed. However, for random permutations, the upper bound of the probability satisfying the uniformly distributed property for each byte is $\frac{1}{2^{128-1}}$ with $q \leq N = 2^{64}$ oracle queries.

Furthermore, with the representation of (3), Gilbert mounted a known-key distinguisher for AES_{10}. This distinguisher is implemented by extending one round on each side based on the distinguisher for AES_8. The same 2^{64} middle texts $\mathcal{Z}$ as for the known-key distinguisher on AES_8 are used. For the corresponding input-output pairs (p_i, c_i), $i = 1, \ldots, 2^{64}$, The adversary can find at least one value (Δ, Γ), where $\Delta, \Gamma \in \{0,1\}^{128}$, to make each byte of $R \circ SB(P^{-1}(p_i) \oplus \Delta)$ and $R^{-1} \circ SB^{-1}(Q(c_i) \oplus \Gamma)$ uniformly distributed within time complexity 2^{64}. However, for a random permutation, the upper bound of the probability satisfying the uniformly distributed property for each byte is $2^{-16.5}$ with $q \leq N = 2^{64}$ oracle queries.

$$\begin{bmatrix} A_1^0CCC \\ A_2^0CCC \\ A_3^0CCC \\ A_4^0CCC \end{bmatrix} \xrightarrow{S} \begin{bmatrix} A_1^0CCC \\ A_2^0CCC \\ A_3^0CCC \\ A_4^0CCC \end{bmatrix} \xrightarrow{R} \begin{bmatrix} A_1^1A_1^2A_1^3A_1^4 \\ A_2^1A_2^2A_2^3A_2^4 \\ A_3^1A_3^2A_3^3A_3^4 \\ A_4^1A_4^2A_4^3A_4^4 \end{bmatrix} \xrightarrow{S} \begin{bmatrix} A_1^1A_1^2A_1^3A_1^4 \\ A_2^1A_2^2A_2^3A_2^4 \\ A_3^1A_3^2A_3^3A_3^4 \\ A_4^1A_4^2A_4^3A_4^4 \end{bmatrix} \xrightarrow{P \circ SR \diamond AK} \begin{bmatrix} AAAA \\ AAAA \\ AAAA \\ AAAA \end{bmatrix}$$

$$\begin{bmatrix} A_1^0CCC \\ A_2^0CCC \\ A_3^0CCC \\ A_4^0CCC \end{bmatrix} \xrightarrow{S^{-1}} \begin{bmatrix} A_1^0CCC \\ A_2^0CCC \\ A_3^0CCC \\ A_4^0CCC \end{bmatrix} \xrightarrow{R^{-1}} \begin{bmatrix} A_1^1A_1^2A_1^3A_1^4 \\ A_2^1A_2^2A_2^3A_2^4 \\ A_3^1A_3^2A_3^3A_3^4 \\ A_4^1A_4^2A_4^3A_4^4 \end{bmatrix} \xrightarrow{S^{-1}} \begin{bmatrix} A_1^1A_1^2A_1^3A_1^4 \\ A_2^1A_2^2A_2^3A_2^4 \\ A_3^1A_3^2A_3^3A_3^4 \\ A_4^1A_4^2A_4^3A_4^4 \end{bmatrix} \xrightarrow{(AK \diamond SR \diamond Q)^{-1}} \begin{bmatrix} AAAA \\ AAAA \\ AAAA \\ AAAA \end{bmatrix}$$

Fig. 1 Two integral distinguishers under the new representation of AES in [10]

Since Gilbert's work is based on the integral distinguisher and uses the active property[1], if we can improve the statistical integral model proposed by Wang et al. in [25], we can further improve Gilbert's work and widely utilize the new method to all AES-like ciphers. With the improved known-key distinguishers, 10-round AES-like ciphers cannot be regarded as ideal random permutations, and the time complexities of new distinguishers are less than previous ones.

2.4 Statistical integral distinguisher

In this subsection, we recall the statistical integral distinguisher proposed by Wang et al. in [25].

Assume that $H : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is a part of a block cipher, its input and output both can be splitted into two parts as follows:

$$H : \mathbb{F}_2^r \times \mathbb{F}_2^s \to \mathbb{F}_2^t \times \mathbb{F}_2^u, H(x, y) = \begin{pmatrix} H_1(x, y) \\ H_2(x, y) \end{pmatrix}.$$

If the first r bits of input are fixed as a constant λ and only the first t bits of output are considered, then the function H can be denoted as T_λ:

$$T_\lambda : \mathbb{F}_2^s \to \mathbb{F}_2^t, T_\lambda(y) = H_1(\lambda, y).$$

When y takes over all possible values, the outputs $T_\lambda(y)$ are uniformly distributed, then an integral distinguisher is constructed.

If the adversary only takes $N < 2^s$ different y, sets a counter vector $V[T_\lambda(y)]$, $T_\lambda(y) \in \mathbb{F}_2^t$ to keep track of the number of each value $T_\lambda(y)$ and initializes this counter as zero, a statistical integral distinguisher can be constructed by investigating the distribution of the statistic as follows:

$$T = \sum_{T_\lambda(y)=0}^{2^t-1} \frac{(V[T_\lambda(y)] - N \cdot 2^{-t})^2}{N \cdot 2^{-t}}. \tag{5}$$

For the right key guess (the target cipher), the statistic T follows a χ^2 distribution with mean $\mu_0 = (2^t - 1)\frac{2^s-N}{2^s-1}$ and variance $\sigma^2 = 2(2^t - 1)(\frac{2^s-N}{2^s-1})^2$, but for the wrong key guess (a random permutation), it follows a χ^2 distribution with mean $\mu_0 = (2^t - 1)$ and variance $\sigma^2 = 2(2^t - 1)$. The relation of data complexity, type-I error probability α_0 and type-II error probability α_1 is as follows

$$N = \frac{(2^s - 1)(q_{1-\alpha_0} + q_{1-\alpha_1})}{\sqrt{(2^t - 1)/2} + q_{1-\alpha_1}} + 1, \tag{6}$$

where $q_{1-\alpha_0}$ and $q_{1-\alpha_1}$ are the respective quantiles of the standard normal distribution.

3 Statistical integral distinguisher with multiple structures on input and integral properties on output

In some integral distinguishers, there are b groups of t output bits with the active property. If we can utilize all properties at the same time, the data complexity can be further reduced. What's more, in some attack settings, N_s structures, i.e. that N_s different λ, should be used

[1] Active property means that the values on target bits are uniform distributed.

together. For these special settings, we construct the new statistical integral distinguisher in this section.

Firstly, we split the input into two parts and output into $b+1$ parts.

$$H : \mathbb{F}_2^r \times \mathbb{F}_2^s \to \mathbb{F}_2^t \times \mathbb{F}_2^t \times \ldots \times \mathbb{F}_2^t \times \mathbb{F}_2^u,\ H(x,y) = \begin{pmatrix} H_1(x,y) \\ H_2(x,y) \\ \ldots \\ H_{b+1}(x,y) \end{pmatrix}.$$

Then we use T_λ^i to denote the function H_i where the first r bits of its input are fixed to the value λ and b outputs H_i, $1 \le i \le b$, are considered:

$$T_\lambda^i : \mathbb{F}_2^s \to \mathbb{F}_2^t,\ T_\lambda^i(y) = H_i(\lambda, y), i = 1, 2, \ldots, b.$$

For a special integral distinguisher, when y iterates all possible values of $\mathbb{F}_2^s$, $T_\lambda^i(y), i = 1, 2, \ldots, b$ are all uniformly distributed with probability one. Further more, if we take N_s values for λ, *i.e.* N_s structures and in each structure y iterates all possible values of $\mathbb{F}_2^s$, the integral properties on output are satisfied as well.

Now assume we need $N < 2^s$ values of y under each structure and we use N_s structures which are independent. $T_\lambda^i(y) \in \mathbb{F}_2^t, i = 1, 2, \ldots, b$ are computed for each y and we allocate a counter vector $V_i[T_\lambda^i(y)]$ to store the occurrences of $T_\lambda^i(y)$. Then we investigate the distribution of the following statistic:

$$C = \sum_{\lambda=1}^{N_s} \sum_{i=1}^{b} \sum_{T_\lambda^i(y)=0}^{2^t-1} \frac{(V_i[T_\lambda^i(y)] - N \cdot 2^{-t})^2}{N \cdot 2^{-t}}. \tag{7}$$

The statistic C follows different distributions determined by whether we are dealing with an actual cipher or a random permutation.

Proposition 1 *For sufficiently large N and t, the statistic $\frac{2^s-1}{2^s-N} C_{cipher}$ (C_{cipher} is the statistic C for cipher) follows a χ^2-distribution with degree of freedom $b \cdot N_s \cdot (2^t - 1)$, which means that C_{cipher} approximately follows a normal distribution with mean and variance*

$$\mu_0 = Exp(C_{cipher}) = b \cdot N_s \cdot (2^t - 1)\frac{2^s - N}{2^s - 1},$$
$$\sigma_0^2 = Var(C_{cipher}) = 2b \cdot N_s \cdot (2^t - 1)(\frac{2^s - N}{2^s - 1})^2.$$

The statistic C_{random} (C_{random} is the statistic C for randomly drawn permutation) follows a χ^2-distribution with degree of freedom $b \cdot N_s \cdot (2^t - 1)$, which means that C_{random} approximately follows a normal distribution with mean and variance

$$\mu_1 = Exp(C_{random}) = b \cdot N_s \cdot (2^t - 1) \text{ and } \sigma_1^2 = Var(C_{random}) = 2b \cdot N_s \cdot (2^t - 1).$$

Proof Deduced from Proposition 1 in [25], for a randomly drawn permutation, the statistic $\sum_{T_\lambda^i(y)=0}^{2^t-1} \frac{(V_i[T_\lambda^i(y)]-N\cdot 2^{-t})^2}{N\cdot 2^{-t}}$ follows a χ^2-distribution with degree of freedom $2^t - 1$ for any λ and i. Then the statistic C'_{random} for the randomly drawn permutation

$$C_{random} = \sum_{\lambda=1}^{N_s} \sum_{i=1}^{b} \sum_{T_\lambda^i(y)=0}^{2^t-1} \frac{(V_i[T_\lambda^i(y)] - N \cdot 2^{-t})^2}{N \cdot 2^{-t}}$$

is the sum of $N_s \cdot b$ independent χ^2 statistics with degree of freedom $2^t - 1$, so the statistic C_{random} follows a χ^2-distribution with degree of freedom $b \cdot N_s \cdot (2^t - 1)$. [2] Then for sufficiently large N and t, C_{random} approximately follows a normal distribution with the expected value and variance:

$$Exp(C_{random}) = b \cdot N_s \cdot (2^t - 1) \text{ and } Var(C_{random}) = 2b \cdot N_s \cdot (2^t - 1).$$

Since the statistic for the cipher $\frac{2^s-1}{2^s-N} \sum_{T_\lambda^i(y)=0}^{2^t-1} \frac{(V_i[T_\lambda^i(y)]-N\cdot 2^{-t})^2}{N\cdot 2^{-t}}$, for any λ and i, follows a χ^2-distribution with degree of freedom $2^t - 1$ deduced from [25]. Then the statistic $\frac{2^s-1}{2^s-N} C'_{cipher}$ for the cipher

$$\frac{2^s - 1}{2^s - N} C_{cipher} = \sum_{\lambda=1}^{N_s} \sum_{i=1}^{b} \frac{2^s - 1}{2^s - N} \sum_{T_\lambda^i(y)=0}^{2^t-1} \frac{(V_i[T_\lambda^i(y)] - N \cdot 2^{-t})^2}{N \cdot 2^{-t}}$$

is the sum of $N_s \cdot b$ independent χ^2 statistics with degree of freedom $2^t - 1$, so the statistic $\frac{2^s-1}{2^s-N} C_{cipher}$ follows a χ^2-distribution with degree of freedom $b \cdot N_s \cdot (2^t - 1)$. Then for sufficiently large N and t, C_{cipher} approximately follows a normal distribution with the expected value and variance:

$$Exp(C_{cipher}) = b \cdot N_s \cdot (2^t - 1) \cdot \frac{2^s - 1}{2^s - N} \text{ and } Var(C_{cipher}) = 2b \cdot N_s \cdot (2^t - 1) \cdot (\frac{2^s - 1}{2^s - N})^2.$$

□

Corollary 1 *Under the assumption of Proposition 1, for type-I error probability α_0 (the probability to wrongfully discard the cipher), and type-II error probability α_1 (the probability to wrongfully accept a randomly chosen permutation as the cipher), to distinguish a cipher and a random permutation based on b independent t-bit outputs when randomly choosing N_s values for r-bit inputs and N values for s-bit inputs, then the following equation holds.*

$$N = \frac{(2^s - 1)(q_{1-\alpha_0} + q_{1-\alpha_1})}{\sqrt{(b \cdot N_s \cdot (2^t - 1))/2 + q_{1-\alpha_0}}} + 1, \tag{8}$$

where $q_{1-\alpha_0}$ and $q_{1-\alpha_1}$ are the respective quantiles of the standard normal distribution.

Corollary 1 is obtained from the equation about the decision threshold $\tau = \mu_0 + \sigma_0 q_{1-\alpha_0} = \mu_1 - \sigma_1 q_{1-\alpha_1}$. While the statistic test is also based on the decision threshold τ: if $C \leq \tau$, the test outputs 'cipher'; Otherwise, the test outputs 'random'. Note that in this statistical method the success probability $Ps = 1 - \alpha_0$ and the relation between α_1 and the advantage of the attack a is $\alpha_1 = 2^{-a}$.

In order to verify the theoretical model in Corollary 1, we implement the experiments for mini version of AES in Appendix A.1. It shows that the experimental results are in good accordance with the theoretical results.

From (8), we know that the data complexity for the statistical distinguisher is $N \cdot N_s$. For the given values of $n, s, t, \alpha_0, \alpha_1$, the ratio of the data complexity with N_s structures to that with one structure is $\sqrt{N_s}$. It means that more structures will result in high data

[2] Here is an underlying assumption that all $T_\lambda^i(y) = H_i(\lambda, y)$ are i.i.d.. If $\{H_i(\lambda, y)\}$ are simple and have strong relationship with each other, this assumption is incorrect. However, in actual ciphers, integral distinguishers often include so many rounds that $\{H_i(\lambda, y)\}$ are complicated and have enough randomness. So this assumption here is suitable in practice, which is also verified by experiments in Appendix A.1.

complexity, so we should avoid to utilize more structures. However, for the known-key integral distinguisher for AES etc., we have to use enough structures to make the plaintexts and the ciphertexts satisfying the desired properties simultaneously. Moreover, if b is increased, the data complexity can be reduced, but as b increases, the time complexity in some situations will be increased accordingly. Thus, we should take the proper value for b according to the time-data tradeoff.

4 Known-key distinguishers on AES-like cipher

In this section, we follow the definition of known-key distinguisher in [10] and use our new statistical integral model to put forward known-key distinguishers on AES-like ciphers based on Gilbert's work at ASIACRYPT'14.[3]

4.1 Definition of known-key distinguisher

Known-key model was introduced by Knudsen and Rijmen in [17] to learn something about the security margin of a cipher. There exist differences between secret-key model and known-key model. In secret-key model, the key used in the cipher is secret to the adversary so that the cipher is more like a black box. The adversary needs to recover the key or distinguish this cipher from any random permutation by accessing to this black box. While in known-key model, the adversary knows the key used in the cipher. The goal of such adversary is to achieve an input-output correlation which she can not achieve with the inputs and outputs obtained by accessing to any random permutation. In this paper, we follow the definition of known-key distinguisher in [10] below.

Definition 1 (T-Intractable Relation [10]) Let $E : (K, X) \in \{0,1\}^k \times \{0,1\}^n \rightarrow E_K(X) \in \{0,1\}^n$ denote a block cipher of block size n bits. Let $N \geq 1$ and $\mathcal{R}$ denote an integer and any relation over the set S of N-tuples of n-bit blocks. $\mathcal{R}$ is said to be T-intractable relatively to E if, given any algorithm $\mathcal{A}'$ that is given an oracle access to a perfect random permutation Π of $\{0,1\}^n$ and its inverse, it is impossible for $\mathcal{A}'$ to construct in time $T' \leq T$ two N-tuples $\mathcal{X}' = (X'_i)$ and $\mathcal{Y}' = (Y'_i)$ such that $Y'_i = \Pi(X'_i), i = 1 \ldots N$ and $\mathcal{X}'\mathcal{R}\mathcal{Y}'$ with a success probability $p' \geq \frac{1}{2}$ over Π and the random choices of $\mathcal{A}'$. The computing time T' of $\mathcal{A}'$ is measured as an equivalent number of computations of E, with the convention that the time needed for one oracle query to Π or Π^{-1} is equal to 1. Thus if q' denotes the number of queries of $\mathcal{A}'$ to Π or Π^{-1}, $q' \leq T'$.

Definition 2 (Known-Key Distinguisher [10]) Let $E : (K, X) \in \{0,1\}^k \times \{0,1\}^n \rightarrow E_K(X) \in \{0,1\}^n$ denote a block cipher of block size n bits. A known-key distinguisher $(\mathcal{R}, \mathcal{A})$ of order $N \geq 1$ consists of (1) a relation $\mathcal{R}$ over the N-tuples of n-bit blocks (2) an algorithm $\mathcal{A}$ that on input a k-bit key K produces in time $T_{\mathcal{A}}$, i.e. in time equivalent with $T_{\mathcal{A}}$ computations of E, an N-tuple $\mathcal{X} = (X_i), i = 1 \ldots N$ of plaintext blocks and an N-tuple $\mathcal{Y} = (Y_i), i = 1 \ldots N$ of ciphertext blocks related by $Y'_i = E(X'_i)$, The two following conditions must be met:

– The relation $\mathcal{R}$ must be $T_{\mathcal{A}}$-intractable relatively to E.

[3] These improved known-key distinguishers on AES-like cipher in this paper follow the idea in Gilbert' work at ASIACRYPT'14, but we adopt statistical integral method instead of integral method and more delicate processes to reduce the data and time complexities.

– The validity of $\mathcal{R}$ must be efficiently checkable: we formalize this requirement by incorporating the time for checking whether two N-tuples are related by $\mathcal{R}$ in the computing time $T_{\mathcal{A}}$ of algorithm $\mathcal{A}$.

This definition can be understood as follows. Firstly, the adversary can derive an N-tuples of input and output blocks satisfying a relation $\mathcal{R}$ with time T, while she cannot derive from oracle queries to any random permutation and its inverse an N-tuples of input/output pairs satisfying the same relation with time less than T. Secondly, this relation $\mathcal{R}$ must be achieved under each key from the key space and must be checkable without knowing K. In this paper, we use statistical integral property to implement known-key distinguisher on AES-like cipher instead of integral property, so we slightly modify the definition that an N-tuples of input and output blocks satisfies a relation $\mathcal{R}$ with a high probability while the same relation exists on random permutation with a very low probability.

4.2 New known-key distinguisher on 8-round AES-like cipher

As described in Section 2.3, the known-key distinguisher for AES_8 is based on the uniformly distributed integral property with 2^{32} structures and each structure takes 2^{32} texts. This integral property can be transformed into a statistical integral property for AES-like cipher by using Proposition 1. So in our known-key distinguisher on AES_8^{like}, we utilize the statistical integral property on each cell of input and output to distinguish the actual cipher from random permutation. In this way, the required number of structures and texts in one structure can be reduced. The process to distinguish the actual cipher AES_8^{like} from random permutation is described in Algorithm 1.

Algorithm 1 New known-key distinguisher on AES_8^{like}

Initialize the statistic C' and C'' as zero;
for *all N values of* $x \in (0,1)^{m\times c}$ **do**
  Initialize the counter vector $V[m^2][2^c]$ to zero;
  for *all N values of* $y \in (0,1)^{m\times c}$ **do**
    Compute m^2 cells of input $p_{(l)}, l = 0, \ldots, m^2-1$ from $Z = (x, 0, \ldots, 0) \oplus R(y, 0, \ldots, 0)$;
    Increment the corresponding counter $V[l][p_{(l)}]$ by one;
  $C' = C' + \sum_{l=0}^{m^2-1}\sum_{p_{(l)}=0}^{2^c-1}[\frac{(V[l][p_{(l)}]-N\times 2^{-c})^2}{N\times 2^{-c}}]$;
if $C' > \tau$ **then**
  return $\perp$; `// The distinguishing attack is failed.`
for *all N values of* $y \in (0,1)^{m\times c}$ **do**
  Initialize the counter vector $V[m^2][2^c]$ to zero;
  for *all N values of* $x \in (0,1)^{m\times c}$ **do**
    Compute m^2 cells of output $c_{(l)}, l = 0, \ldots, m^2-1$ from $Z = (x, 0, \ldots, 0) \oplus R(y, 0, \ldots, 0)$;
    Increment the corresponding counter $V[l][c_{(l)}]$ by one;
  $C'' = C'' + \sum_{l=0}^{m^2-1}\sum_{c_{(l)}=0}^{2^c-1}[\frac{(V[l][c_{(l)}]-N\times 2^{-c})^2}{N\times 2^{-c}}]$;
For AES_8^{like}, $C'' \le \tau$;
For any random permutation, $C'' > \tau$.

Since we use the statistical integral distinguisher twice, both type I error probabilities are set to be α_0, for the case of AES_8^{like}, the probability to wrongly regard AES_8 as a random permutation is $\alpha_0+(1-\alpha_0)\alpha_0$, which means the success probability of correctly identifying AES-like cipher is about $(1-\alpha_0)^2$.

While for the case of random permutation, the adversary can implement encryption and decryption oracle queries to the cipher and random permutation. But statistical integral property (exploit χ^2 distribution) is different from traditional integral property (utilize uniform distribution). At the best of times the adversary chooses the data which automatically satisfy the statistical property on the input, but satisfy the statistical property on the output with probability α_1. In order to satisfy the statistical properties both on the input and output, the probability of wrongly regarding this random permutation as AES cipher is $1 \times \alpha_1$.

To summarize, the relation $\mathcal{R}$ used in this known-key distinguisher is that N^2-tuples of input and output satisfy statistical integral property respectively. For the AES cipher, this relation exists with probability of $(1-\alpha_0)^2$, while the same relation can be achieved with probability of only α_1 by oracle queries to any random permutation. The advantage to distinguish AES cipher from random permutation is $(1-\alpha_0)^2-\alpha_1$, which should be not negligible. The total time complexity of this known-key distinguisher is about $2 \times N^2$ computations. The memory requirements are about $m^2 \times 2^c$ used for storing the counter vector $V[m^2][2^c]$.

4.3 New known-key distinguisher on 10-round AES-like cipher

The statistical integral distinguisher on AES_{10} is based on the distinguishing property of AES_{10} in [10], which is represented according to (4), see Fig. 2.

Along with the idea within the distinguisher on AES_{10} in [10], in our known-key distinguisher on AES_{10}^{like}, we use $N_s < 2^{m \times c}$ structures, each of which takes $N = N_s$ middle texts, to obtain N^2 input/output pairs. For AES-like cipher, there is one value for (Δ, Γ) to let each byte of $MC \circ SR \circ SB(input \oplus \Delta)$ and $MC^{-1} \circ SB^{-1} \circ SR^{-1}(output \oplus \Gamma)$ satisfy the statistical integral property with a high probability. But for any random permutation, the probability of having one solution for (Δ, Γ) to obtain the same property is very low.

However, just simplely in above way, the distinguisher has high time complexity. In order to reduce the time complexity, we implement the distinguisher in the following way. As N_s structures are used, we divide them into N_s/n_s groups and each group has n_s structures.

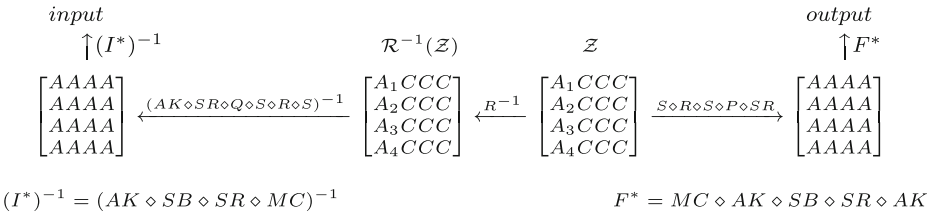

Fig. 2 Take known-key distinguisher for AES_{10} as an example

Then we compute the statistic value for each group. There is one value (Δ, Γ) to make all statistics for N_s/n_s groups on both states $Input' = MC \circ SR \circ SB(input \oplus \Delta)$ and $Output' = MC^{-1} \circ SB^{-1} \circ SR^{-1}(output \oplus \Gamma)$ less than the given threshold τ for AES_{10}. Meanwhile, for any random permutation, even if the attacker can carefully choose the inputs to find one value of Δ to satisfy the statistical property on the state $Input'$ with probability one, the probability of finding one value Γ to satisfy the statistical property on the state $Output'$ is very low.

In order to further reduce the time complexity in condition of ensuring a non-negligible distinguishing advantage, we only focus on statistics on m-cell states – $Input'_{(0\sim m-1)}$ and $Output'_{(0\sim m-1)}$. So we only need to find two $(m \times c)$-bit values for Δ' and Γ'. The detailed process for this known-key distinguisher on AES_{10}^{like} is described in Algorithm 2.

In Algorithm 2, we filter out the wrong values for Δ' with the statistics on $Input'_{(0\sim m-1)}$ one by one. At last, the probability that one wrong Δ' is remained after all N_s/n_s filtering processes is about $(2^{m\times c} - 1) \cdot \alpha_1^{m\times N_s/n_s}$, while the probability that the right candidate Δ cannot pass the filtering process is $1 - (1 - \alpha_0)^{m\times N_s/n_s}$.

In the similar way, we filter out the wrong values for Γ' with the statistics for $Output'_{(0\sim m-1)}$ one by one. Finally, the probability that one wrong Γ' can pass the filtering process is also $(2^{m\times c} - 1) \cdot \alpha_1^{m\times N_s/n_s}$, while the probability that the right Γ' cannot pass the filtering process is $1 - (1 - \alpha_0)^{m\times N_s/n_s}$ as well. Therefore, for the case of AES_{10}, the probability to correctly identify the AES_{10} cipher is about $(1 - (1 - \alpha_0)^{m\times N_s/n_s})^2$.

While for the case of random permutation, at the best of the times the adversary can choose the inputs that there is always at least one value of Δ' remaining after the filtering process, but the probability that at least one Γ' survives after the filtering process is about 0.

To summarize, the relation $\mathcal{R}$ used in this known-key distinguisher is that there exists one pair (Δ, Γ) such that N^2-tuples of $MC \circ SR \circ SB(input \oplus \Delta)$ and $MC^{-1} \circ SB^{-1} \circ SR^{-1}(output \oplus \Gamma)$ satisfy the statistical integral property respectively. For AES cipher, this relation exists with probability of $(1 - (1 - \alpha_0)^{4N_s/n_s})^2$, while the same relation can be achieved with probability of about 0 for any random permutation. So the success probability of this distinguisher is about $(1 - (1 - \alpha_0)^{4N_s/n_s})^2$ which is the distinguishing advantage as well.

The time complexity of Steps $2 \sim 3$ is N^2 full round encryptions. Then the time complexity of Steps $4 \sim 9$ is about $2^{m\times c/2} \times n_s \times N$ memory accesses (MA). Steps $10 \sim 15$ take $m/2 \times 2^{(m+2)\times c} \times n_s$ MA, Steps $19 \sim 26$ take $2^{m\times c} \times \alpha_1 \times n_s \times N$ MA and Step 27 needs about $2^{m\times c} \times (\alpha_1^2 + \alpha_1^3 + \ldots + \alpha_1^{m-1}) \times n_s \times N$ MA. The time of Step 27 depends on the value of α_1. Usually we only need to calculate one or two items. After one filter process, the number of candidates for Δ is about 1. Consequently by filtering with other $N/n_s - 1$ groups of structures, the time complexity of Step 28 is $(N/n_s - 1) \times n_s \times N$ MA. Since Step 34 also takes the same times of encryptions as Steps $4 \sim 28$, if we roughly set that one access to a table is equivalent to one full round encryption, the total time complexity of the whole attack is about $N^2 + 2(2^{m\times c/2} \times n_s \times N + m/2 \times 2^{(m+2)\times c} \times n_s + 2^{m\times c} \times (\alpha_1 + \alpha_1^2 + \ldots + \alpha_1^{m-1}) \times n_s \times N + (N/n_s - 1) \times n_s \times N)$ full round encryptions. In addition, the dominant memory requirements happen on $V[N][N]$ and $V'[N][N]$, which need about $2 \times m \times N^2$.

Algorithm 2 New known-key distinguisher on AES_{10}^{like}

Allocate vectors $V[N][N]$, $V'[N][N]$;
for *all* N^2 *values of* (y_i, x_j), $0 \le i, j < N$ **do**
  Calculate input p and output c from $Z = (x_j, 0, \ldots, 0) \oplus R(y_i, 0, \ldots, 0)$ and let $V[j][i] = p'$, $V'[i][j] = c'$; // p' and c' are m-cell input and output related to $Input'_{0\sim m-1}$ and $Output'_{0\sim m-1}$ respectively.
// Steps $4 \sim 28$ proceed the first group with n_s structures.
// MC operation: $y_i = a_0^i x_0 + a_0^i x_0 + \ldots + a_{m-1}^i x_{m-1}, i = 0, 1, \ldots, m-1$.
for *all* $2^{m\times c/2}$ *values of* $(\Delta'_{(0)}, \ldots, \Delta'_{(m/2-1)})$ **do**
  Allocate vectors $V_1[n_s][2^{(m/2+1)\times c}]$;
  for *all* n_s *values of* j *and* N *values of* i **do**
    Get $(p'_{(0\sim m/2-1)}$ from V[j][i];
    Compute $W_{m/2-1} = a_0^0 \cdot SB(p'_{(0)} \oplus \Delta_{(0)}) \oplus \ldots \oplus a_{m/2-1}^0 \cdot SB(p'_{(m/2-1)} \oplus \Delta_{(m/2-1)})$;
    Let $V_1[j][W_{m/2-1} \| p'_{(m/2\sim m-1)}]$ increase one;
  for *all remained* $m/2$*-cell* $\Delta'_{(i)}$, $i = m/2, \ldots, m-1$ **do**
    for *all* $2^{(i-m/2+1)\times c}$ *values of* $\Delta'_{(m/2\sim i)}$ **do**
      Allocate a counter vectors $V_i[n_s][2^{(m-i)\times c}]$, and initialize to zero;
      for *all* n_s *values of* j *and all* $2^{(m-i+1)\times c}$ *values of* $W_i \| p'_{(i\sim m-1)}$ **do**
        Compute $W_i = W_{i-1} \oplus (a_i^0 \cdot SB(p'_{(i)} \oplus \Delta'_{(i)}))$;
        Let $V_i[j][W_i \| p_{(i+1\sim m-1)}]$ += $V_{i-1}[j][W_i \| p_{(i\sim m-1)}]$;
  $C_1 = \sum_{j=0}^{n_s-1} \sum_{W=0}^{2^c-1} \frac{(V_{m-1}[j][W] - N\times 2^{-c})^2}{N\times 2^{-c}}$;
  if $C_1 \le \tau$ **then**
    Put Δ' into V_k. // About remain $2^{m\times c} \cdot \alpha_1$ values.
for *all values of* $\Delta' \in V_k$ **do**
  Allocate counter vectors $V_m[n_s][2^c]$, and initialize to zero;
  for *all* n_s *values of* j *and* N *values of* i **do**
    Get m-cell p' from V[j][i] and compute $Input'_{(1)} = a_0^1 \cdot SB(p_{(0)} \oplus \Delta_{(0)}) \oplus \ldots \oplus a_{m-1}^1 \cdot SB(p'_{(m-1)} \oplus \Delta'_{(m-1)})$;
    Increment $V_m[j][Input'_{(1)}]$ by one;
  $C_2 = \sum_{j=0}^{n_s-1} \sum_{W=0}^{2^c-1} \frac{(V_m[j][W] - N\times 2^{-c})^2}{N\times 2^{-c}}$;
  if $C_2 \le \tau$ **then**
    Put Δ' into $V_{k_1}[\cdot]$. // About $2^{m\times c} \cdot \alpha_1^2$ values are remained.
Proceed the similar steps as $19 \sim 26$ for the other $m-2$ cells of $Input'_{(2\sim m-1)}$.
  // About 1 value is remained.
Check if this Δ' satisfies the other $N/n_s - 1$ groups of n_s structures;
if *there is no solution for* Δ' *remained* **then**
  return $\perp$. // The distinguishing attack is failed.
Proceed Steps $4 \sim 28$ with $V'[N][N]$ to compute the distributions on $Output'_{(0\sim m-1)}$ by guessing Γ';
For AES_{10}^{like}, there exists one solution for Γ';
For any random permutation, there is no solution for Γ'.

5 Application on AES-like ciphers

In this section, we apply the known-key distinguishers on actually AES-like ciphers including AES, Whirlpool, Grøstl and PHOTON permutations. These distinguishers are the best ones under known-key setting.

5.1 Application to AES

Advanced Encryption Standard (AES) [8], published by NIST, is the first target for our method. Take AES-128 as an example. Its block size is 128 bits and it totally has 10 rounds. By applying known-key distinguishers on AES, we have the following results.

8-round known-key distinguisher According to Proposition 1, we have $s = 32$, $t = 8$, $b = 16$ and $N = N_s$. If we set the error probabilities $\alpha_0 = 2^{-128}$ and $\alpha_1 = 2^{-128}$ (the values of α_0 and α_1 can be different and take any suitable values), then $q_{1-\alpha_0} = q_{1-\alpha_1} \approx 13.06$. Calculating with (8), $N = N_s \approx 2^{20.81}$ and the threshold value $\tau \approx 7478730631.39$. Following the algorithm in Section 4.2, this distinguisher can be implemented successfully with $2\times(2^{20.81})^2 = 2^{42.61}$ computations and $16\times 2^8\times 2 = 2^{13}$ bytes of memory. Meanwhile, for the case of AES_8, the probability of correctly identifying AES cipher is about $1-2^{-127}$, but for the case of random permutation, the probability of wrongly regarding this random permutation as AES cipher is 2^{-128}. So the distinguishing advantage is $1-2^{-127}-2^{-128} \approx 1$, which is not negligible.

10-round known-key distinguisher In this setting, by applying Proposition 1, we have $s = 32$, $t = 8$, $b = 1$ and $n_s = 2^8$. If we set the error probabilities $\alpha_0 = 2^{-50}$ and $\alpha_1 = 2^{-10.51}$, then $N = 2^{27.92}$ and $\tau = 64123.53$ according to (8). Following the algorithm in Section 4.3, this known-key distinguisher needs $2^{59.60}$ full-round encryptions and $2^{58.84}$ bytes of memory. The distinguishing advantage is about $1-2^{-27.08}$ which is non-ignorable.

The best previous known-key distinguisher on AES-128 is 10 rounds and was introduced by Gilbert [10], in which the time complexity is 2^{64}. Comparing with this distinguisher, ours are the best one according to the time complexity.

5.2 Application to Whirlpool

Whirlpool hash function [3] is an ISO/IEC standard hash function designed by Barreto and Rijmen in 2000. It processes 512-bit message blocks and produces a 512-bit hash value. Its compression function is based on an AES-like block cipher E with the Miyaguchi-Preneel mode: $H_j = h(H_{j-1}, M_j) = E_{H_{j-1}}(M_j) \oplus M_j \oplus H_{j-1}$, where H_j is the chaining value and M_j is the j-th message block. This block cipher E employs two similar 10-round permutations: one takes the chaining value as input to produce 11 subkeys for the second permutation; the second one updates 8×8 bytes message block with the subkeys from the first permutation with 10 rounds. The round transformation includes 4 operations: SubBytes(SB), ShiftColumns(SC), MixRows(MR) and AddRoundKey(AK) or AddRoundConstant(AC) in key schedule, which is similar to AES. For more detail, please refer to [3]. Known-key distinguishers on block cipher E (regard this permutation as block cipher) can be obtained by following the method in Section 4. The only difference is that R and S operate on columns and rows respectively.

8-round known-key distinguisher By applying Proposition 1 on the known-key distinguisher, we have $s = 64$, $t = 8$, $b = 64$ and $N = N_s$. If we set the error probabilities $\alpha_0 = 2^{-512}$ and $\alpha_1 = 2^{-512}$, then $q_{1-\alpha_0} = q_{1-\alpha_1} \approx 26.48$. According to (8), $N = N_s \approx 2^{42.15}$ and the threshold value $\tau \approx 79819582776240513.99$. According to the complexity analysis in Section 4.2, the known-key distinguisher on 8-round E needs $2^{85.31}$ computations and $2^{16.32}$ bytes of memory. The distinguishing advantage is about $1 - 2^{-511} - 2^{-512} \approx 1$.

10-round known-key distinguisher In this setting, by applying Proposition 1, $s = 64$, $t = 8$, $b = 1$ and $n_s = 2^{30}$. If we set the error probabilities $\alpha_0 = 2^{-64}$ and $\alpha_1 = 2^{-34}$, then $N = 2^{49.46}$ according to (8). According to Section 4.3, the time and memory complexities are $2^{113.90}$ encryptions and $2^{102.92}$ bytes respectively. The distinguishing advantage is $1 - 2^{-40.54}$ which is non-ignorable.

The best known-key distinguisher on Whirlpool was proposed by Lamberger et al. in [20] and is only 8 rounds. But there are chosen-key distinguishers for 10-round Whirlpool in [16, 20]. Comparing with the previous results, our known-key distinguisher on 8-round E block cipher is the best one according to the time complexity and on 10-round E is the first published one under known-key distinguishing setting.

5.3 Application to Grøstl-256

Grøstl [9] is a SHA-3 candidate proposal. It is an iterated hash function with a compression function built from two distinct permutations. These permutations are constructed using wide trail strategy so that they are AES-like ciphers. Grøstl has two versions Grøstl-256 and Grøstl-512. The permutation used in Grøstl-256 has 10 rounds. Its internal state is 512 bits (8×8 cells). The known-key distinguishers on 8-round and 10-round of this permutation are similar to that for Whirlpool except that R and S operations are implemented on rows and columns respectively. As a result, the distinguisher on 8-round permutation of Grøstl-256 needs $2^{85.31}$ encryptions and $2^{16.32}$ bytes of memory and on 10-round permutation needs $2^{113.90}$ encryptions and $2^{102.92}$ bytes. The latter is the best one so far under known-key setting compared with the best previous ones for 9-round Grøstl proposed in [15, 16].

5.4 Application to PHOTON

PHOTON, introduced by Guo et al. at CRYPTO'11 in [14], is a family of lightweight hash functions. It is designed based on AES-like block cipher in a sponge construction. The security of PHOTON family directly depend on the security of the internal permutations. It is necessary to study the distinguishers for this component. There are five different permutations P_{100}, P_{144}, P_{196}, P_{256} and P_{288} being used for PHOTON-80/20/16, PHOTON-128/16/16, PHOTON-160/36/36, PHOTON-224/32/32 and PHOTON-256/32/32. The number of rounds are always 12 rounds. Taking P_{100} as an example, we propose the known-key distinguishers on it. Since the distinct among these permutations is the size of state, similar method is applied on other 4 ones. We show the results directly in Table 1.

10-round known-key distinguisher In this setting, by applying Proposition 1, $s = 20$, $t = 5$, $b = 1$ and $n_s = 2^{10}$ for P_{100}. If we set the error probabilities $\alpha_0 = 2^{-32}$ and $\alpha_1 = 2^{-64}$ (We set the same α_0 and α_1 and $n_s = 2^{15}$, 2^{20}, 2^{25} and 2^{30} sequentially for other four permutations), then $N = 2^{17.34}$ according to (8). According to Section 4.3, the time and memory complexities are $2^{40.71}$ encryptions and $2^{37.00}$ bytes respectively. The distinguishing advantage is $1 - 2^{-21.66}$ which is non-ignorable.

Compared with the best previous results for up to 9-round permutations of PHOTON in [16], our known-key distinguishers on 10-round permutations are the best ones under known-key distinguishing setting.

6 Secret-key statistical integral distinguisher on 5-round AES

In this section, we propose a secret-key distinguisher on 5-round AES with our statistical integral model based on the work of Sun et al. in [23]. In this distinguisher, the S-box used in AES is secret.

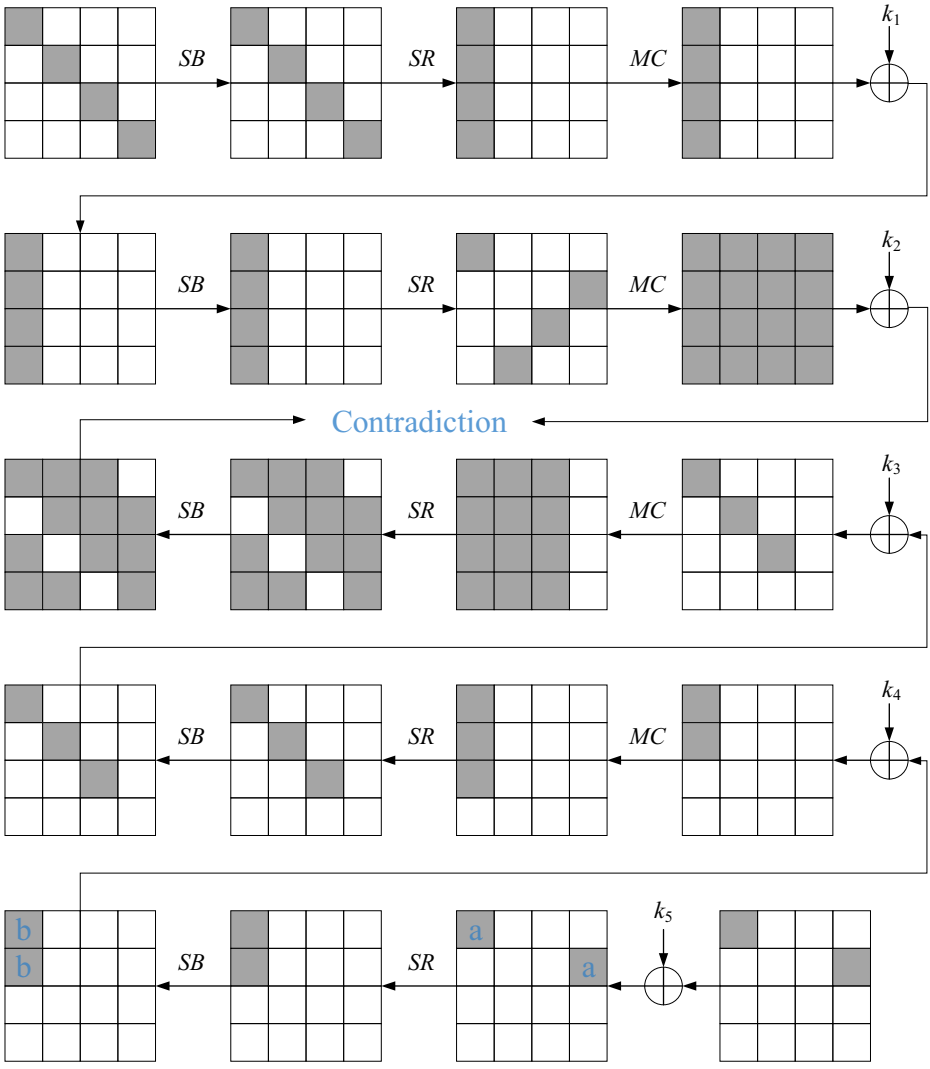

Fig. 3 Zero-correlation linear hull on 5-round AES with secret S-box under secret-key setting. Gray and white cells denote nonzero and zero masks respectively. the two cells with a or b are exactly the same mask

Firstly, we slightly modify the zero-correlation linear hull for 5-round decryption of AES under chosen-ciphertext mode proposed by Sun et al. in [23] (Lemma 3). Let $V = \{(x_{(i)}) \in F_{2^8}^{16} | x_{(0)} \oplus x_{(13)} = (k_5)_{(0)} \oplus (k_5)_{(13)}\}$, and assume that the input mask $\Gamma_I = (a_{(i)})_{0 \le i \le 15}$ and output mask $\Gamma_O^0 = (\beta_{(i)})_{0 \le i \le 15}$ satisfy:

$$a_{(i)} = \begin{cases} a, i = 0, 13; \\ 0, otherwise. \end{cases} \quad \beta_{(j)} = \begin{cases} nonzero, j = \{0, 5, 10, 15\}; \\ 0, otherwise. \end{cases}$$

Then the correlation for $\Gamma_I \to \Gamma_O^0$ on V is always 0. Note that there are three other zero-correlation linear hulls as well, when $j = \{1, 6, 11, 12\}, \{2, 7, 8, 13\}, \{3, 4, 9, 14\}$. The corresponding output masks are Γ_O^1, Γ_O^2 and Γ_O^3 respectively. One of the four cases is shown in Fig. 3.

With the technique proposed by Sun et al. in [24], these four zero-correlation linear hulls can be transformed into integral ones. Taking the linear hull $\Gamma_I \to \Gamma_O^0$ as an example, the corresponding integral distinguisher is that if the adversary takes over 2^{120} different values of ciphertexts c satisfying $c_{(0)} \oplus c_{(13)} = (k_5)_{(0)} \oplus (k_5)_{(13)}$, then the values on 4 bytes of plaintext $(p_{(0)}, p_{(5)}, p_{(10)}, p_{(15)})$ are uniformly distributed.

Based on these integral distinguishers, we can implement a statistical integral distinguisher for each candidate $\Delta = (k_5)_{(0)} \oplus (k_5)_{(13)}$, where $s = 120$ and $t = 32$. In order to have the success probability $(1 - \alpha_0)^{2^8} = (1 - \alpha_1)^{2^8} = 95\%$, we set $\alpha_0 = \alpha_1 = 0.0002$, then $q_{1-\alpha_0} = q_{1-\alpha_1} \approx 3.54$. Meanwhile, we can use these four integral distinguishers above together within one structure, so $b = 4$ and $N_s = 1$. Thus by (8), $N = 2^{106.32}$ chosen ciphertexts. The decision threshold is about $\tau \approx 17179212992.15$. As there are 2^8 different values of Δ, the total data complexity of this distinguisher is $N' = 2^{106.32} \times 2^8 = 2^{114.32}$ chosen ciphertexts.

What's more, we can see from algorithm 3, the main time complexity happens on Step 5, which is about $2^8 \times 2^{106.32} \times 4 \times 1/16 \times 1/5 \approx 2^{110}$ encryptions, if we regard one simple operation as $\frac{1}{16}$ one round encryption. Besides that, memory requirements are about $4 \times 2^{32} \times 10 \approx 2^{37.32}$ bytes $= 2^{33.32}$ blocks.

Algorithm 3 Secret-key statistical integral distinguisher on 5-round AES with secret S-box

1 **for** 2^8 *candidates of* Δ **do**
2 Set a counter $V[4][2^{32}]$ and initialize it to zero;
3 **for** N *chosen ciphertext/plaintext pairs* (c, p) **do**
   `// Consider those four integrals together.`
4 **for** $i \leftarrow 0 \sim 3$ **do**
5 Increment counter $V[i][c^i_{part}]$ by one according to the related 4 bytes $c^i_{part} \in (0, 1)^{32}$ of ciphertext c;
6 Calculate the statistic $T_\Delta = \sum_{b=0}^{3} \sum_{z=0}^{2^{32}-1} \frac{(V[b][z] - N \cdot 2^{-32})^2}{N \cdot 2^{-32}}$;
7 **if** *Only one* Δ *such that* $T_\Delta < \tau$ **then**
8 **return** AES;
9 **return** random permution;

As far as we know, this distinguisher is the best secret-key integral one on 5-round AES with secret S-box.

7 Conclusion

In this paper, we propose a statistical integral distinguisher with multiple structures on input and multiple integral properties on output based the work of Wang et al. at FSE'16. With this distinguisher, we give the known-key distinguishing attack on 8-round and 10-round AES-like ciphers such as AES, the permutation of Whirlpool, PHOTON and Grøstl-256 hash funtions based on Gilbert's work at ASIACRYPT'14, which are the best known-key distinguishers so far. Besides that, we present a secret-key statistical integral distinguisher on 5-round AES under chosen-ciphertext mode. This is the best integral distinguisher on AES with secret S-box under secret-key setting. As a future work, we try to apply more statistical techniques to the field of symmetric cipher and find improved attack on AES and AES-like ciphers.

Acknowledgements This work has been supported by NSFC Projects (No. 61572293, No. 61502276, No. 61692276), National Cryptography Development Fund (MMJJ20170102), National Natural Science Foundation of Shandong Province, China (ZR2016FM22), Fundamental Research Fund of Shandong Academy of Sciences (NO.2018:12-16), Major Scientific and Technological Innovation Projects of Shandong Province, China (2017CXGC0704).

Appendix

A.1 Experimental results

In order to verify the theoretical model of statistical integral distinguisher in Section 3, we implement the distinguishing attack in Section 4.2 on a mini variant of AES with the block size 64-bit denoted as AES* here. The round function of AES* is similar to that of AES, including four operations, *i.e.,* SB, SR, MC and AK. 64-bit block is partitioned into 16 nibbles and SB uses S-box S_0 in LBlock. SR is same as that of AES, and the matrix used in MC is

$$M = \begin{pmatrix} 1 & 1 & 4 & 9 \\ 9 & 1 & 1 & 4 \\ 4 & 9 & 1 & 1 \\ 1 & 4 & 9 & 1 \end{pmatrix},$$

which is defined over $GF(2^4)$. For the multiplication, each nibble and value in M are considered as a polynomial over $GF(2)$ and then the nibble is multiplied modulo $x^4 + x + 1$ by the value in M. The addition is simply XOR operation. The subkeys are XORed with the nibbles in AK operation.

There is similar known-key integral distinguisher for 8-round AES* since its similarity to AES, see Fig. 1. Given a set of data $\mathcal{Z} = \{(x, 0, 0, 0) \oplus R(y, 0, 0, 0) | x \in (0, 1)^{16}\}$ for fixed y, *i.e.,* the first column of $\mathcal{Z}$ takes all 2^{16} possible values and other columns are fixed to some constants, after $S \diamond R \diamond S$ operation, each column of output v is active, i.e. that 2^{16} values are uniformly distributed on each column of ouput. Since $R^{-1}(\mathcal{Z}) = \{R^{-1}(x, 0, 0, 0) \oplus (y, 0, 0, 0)\}$ has 2^{16} structures that each one takes all 2^{16} possible values on the first columns and constants on other columns, after $(S \diamond R \diamond S)^{-1}$ operation, each column of output u is active.

In our experiment, we consider the distributions of four 8-bit values in v including the first and second nibble in each column of v. Here $s = 16, t = 8$ and $b = 4$. If we set $\alpha_0 = 0.2$ and take different values for N and N_s, α_1 and τ can be computed using (8). By

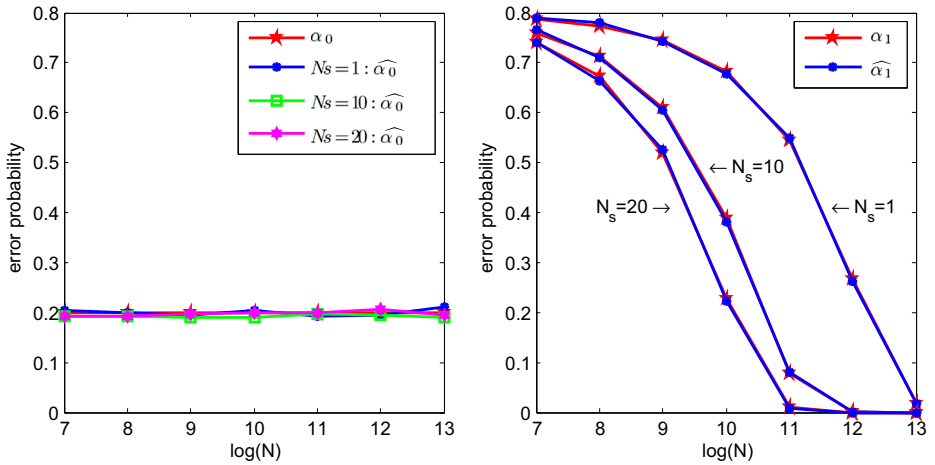

Fig. 4 Experimental results for AES* considering four input bytes. In detail, set the value of α_0 and change the values of N and N_s, the theoretical and empirical α_0 are shown in the left part of figure, corresponding α_1 calculated and tested by equation (5) are shown in the right part of figure

randomly choosing N_s values for y and N values for x, we proceed the experiment to compute the statistics C' for AES* and random permutations. With 2000 times of experiments, we can obtain the empirical error probabilities $\widehat{\alpha_0}$ and $\widehat{\alpha_1}$. The experimental results for $\widehat{\alpha_0}$ and $\widehat{\alpha_1}$ are compared with the theoretical values α_0 and α_1 in Fig. 4.

Moreover, we implement the second experiment where we set b=4 including two bytes of u and two bytes of v. We set $\alpha_0 = 0.2$ and let $N = N_s$, the empirical error probabilities are obtained from 1000 times of experiments. The experimental results for $\widehat{\alpha_0}$ and $\widehat{\alpha_1}$ are compared with the theoretical values α_0 and α_1 in Fig. 5.

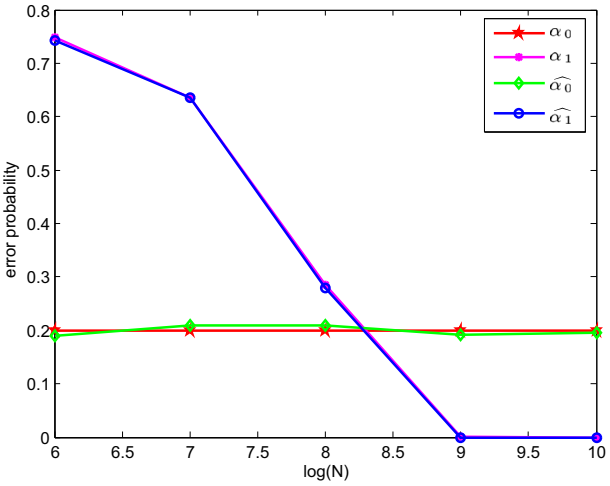

Fig. 5 Experimental results for AES* considering two input and output bytes. In detail, set the theoretical $\alpha_0 = 0.2$ and change the values of N, then the corresponding theoretical α_1 and empirical α_0 and α_1 are calculated and tested by equation (5) in this figure

Figures 4 and 5 show that the test results for the error probabilities are in good accordance with those for theoretical model.

References

1. Aoki, K.: A middletext distinguisher for full CLEFIA-128. In: Proceedings of the international symposium on information theory and its applications, ISITA 2012, Honolulu, October 28-31, 2012, pp. 521–525. IEEE, Piscataway (2012)
2. Aumasson, J.-P., Meier, W.: Zero-sum distinguishers for reduced keccak-f and for the core functions of luffa and hamsi 01 (2018)
3. Barreto, P.S.L.M., Rijmen, V.: Whirlpool. In: van Tilborg, H.C.A., Jajodia, S. (eds.) Encyclopedia of cryptography and security. 2nd edn., pp. 1384–1385. Springer, Berlin (2011)
4. Biryukov, A., Khovratovich, D., Nikolic, I.: Distinguisher and related-key attack on the full AES-256. In: Halevi, S. (ed.) Advances in cryptology - CRYPTO 2009, 29th annual international cryptology conference, Santa Barbara, August 16-20, 2009, Proceedings, vol. 5677 of Lecture Notes in Computer Science, pp. 231–249. Springer, Berlin (2009)
5. Blondeau, C., Peyrin, T., Wang, L.: Known-key distinguisher on full PRESENT. In: Gennaro, R., Robshaw, M. (eds.): Advances in cryptology - CRYPTO 2015 - 35th annual cryptology conference, Santa Barbara, August 16-20, 2015, Proceedings, Part I, vol. 9215 of lecture notes in computer science, pp. 455–474. Springer, Berlin (2015)
6. Cui, T., Sun, L., Chen, H., Wang, M.: Statistical integral distinguisher with multi-structure and its application on AES. In: Pieprzyk, J., Suriadi, S. (eds.) Information security and privacy - 22nd Australasian conference, ACISP 2017, Auckland, July 3-5, 2017, Proceedings, Part I, vol. 10342 of lecture notes in computer science, pp. 402–420. Springer, Berlin (2017)
7. Daemen, J., Knudsen, L.R., Rijmen, V.: The block cipher square. In: Biham, E. (ed.) Fast software encryption, 4th international workshop, FSE '97, Haifa, Israel, January 20-22, 2017 proceedings, vol. 1267 of lecture notes in computer science, pp. 149–165. Springer, Berlin (1997)
8. Daemen, J., Rijmen, V.: The design of Rijndael: AES - the advanced encryption standard. Information Security and Cryptography. Springer, Berlin (2002)
9. Gauravaram, P., Knudsen, L.R., Matusiewicz, K., Mendel, F., Rechberger, C., Schläffer, M., Thomsen, S.S.: Grøstl - a SHA-3 candidate. In: Handschuh, H., Lucks, S., Preneel, B., Rogaway, P. (eds.) Symmetric Cryptography, 11.01. – 16.01.2009, vol. 09031 of Dagstuhl seminar proceedings. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany (2009)
10. Gilbert, H.: A simplified representation of AES. In: Sarkar, P., Iwata, T. (eds.) Advances in cryptology - ASIACRYPT 2014 - 20th international conference on the theory and application of cryptology and information security, Kaoshiung, R.O.C., December 7-11, 2014, Proceedings, Part I, vol. 8873 of lecture notes in computer science, pp. 200–222. Springer, Berlin (2014)
11. Gilbert, H., Peyrin, T.: Super-sbox cryptanalysis: Improved attacks for aes-like permutations. In: Hong, S., Iwata, T. (eds.) Fast software encryption, 17th international workshop, FSE 2010, Seoul, February 7-10, 2010, Revised Selected Papers, vol. 6147 of lecture notes in computer science, pp. 365–383. Springer, Berlin (2010)
12. Grassi, L., Rechberger, C., Rønjom, S.: Subspace trail cryptanalysis and its applications to AES. IACR Trans Symmetric Cryptol **2016**(2), 192–225 (2016)
13. Grassi, L., Rechberger, C., Rønjom, S.: A new structural-differential property of 5-round AES. In: Coron, J., Nielsen, J.B. (eds.) Advances in cryptology - EUROCRYPT 2017 - 36th annual international conference on the theory and applications of cryptographic techniques, Paris, April 30 - May 4, 2017, Proceedings, Part II, volume 10211 of lecture notes in computer science, pp. 289–317 (2017)
14. Guo, J., Peyrin, T., Poschmann, A.: The PHOTON family of lightweight hash functions. In: Rogaway, P. (ed.) Advances in cryptology - CRYPTO 2011 - 31st annual cryptology conference, Santa Barbara, August 14-18, 2011 proceedings, vol. 6841 of lecture notes in computer science, pp. 222–239. Springer, Berlin (2011)
15. Jean, J., Naya-Plasencia, M., Peyrin, T.: Improved rebound attack on the finalist grøstl. In: Canteaut, A. (ed.) Fast software encryption - 19th international workshop, FSE 2012, Washington, March 19-21, 2012, Revised Selected papers, vol. 7549 of lecture notes in computer science, pp. 110–126. Springer, Berlin (2012)

16. Jean, J., Naya-Plasencia, M., Peyrin, T.: Multiple limited-birthday distinguishers and applications. In: Lange, T., Lauter, K.E., Lisonek, P. (eds.) Selected areas in cryptography - SAC 2013 - 20th international conference, Burnaby, August 14-16, 2013, Revised Selected papers, vol. 8282 of lecture notes in computer science, pp. 533–550. Springer, Berlin (2013)
17. Knudsen, L.R., Rijmen, V.: Known-key distinguishers for some block ciphers. In: Kurosawa, K. (ed.) Advances in cryptology - ASIACRYPT 2007, 13th international conference on the theory and application of cryptology and information security, Kuching, December 2-6, 2007, Proceedings, vol. 4833 of lecture notes in computer science, pp. 315–324. Springer, Berlin (2007)
18. Knudsen, L.R., Wagner, D.A.: Integral cryptanalysis. In: Daemen, J., Rijmen, V. (eds.) Fast software encryption, 9th international workshop, FSE 2002, Leuven, February 4-6, 2002, revised papers, vol. 2365 of lecture notes in computer science, pp. 112–127. Springer, Berlin (2002)
19. Lamberger, M., Mendel, F., Rechberger, C., Rijmen, V., Schläffer, M.: Rebound distinguishers: Results on the full whirlpool compression function. In: Matsui, M. (ed.) Advances in cryptology - ASIACRYPT 2009, 15th international conference on the theory and application of cryptology and information security, Tokyo, December 6-10, 2009. Proceedings, vol. 5912 of lecture notes in computer science, pp. 126–143. Springer, Berlin (2009)
20. Lamberger, M., Mendel, F., Schläffer, M., Rechberger, C., Rijmen, V.: The rebound attack and subspace distinguishers: Application to whirlpool. J. Cryptology **28**(2), 257–296 (2015)
21. Mendel, F., Peyrin, T., Rechberger, C., Schläffer, M.: Improved cryptanalysis of the reduced grøstl compression function, ECHO permutation and AES block cipher. In: Rijmen, JrM.J.J.V., Safavi-Naini, R. (eds.) Selected areas in cryptography, 16th annual international workshop, SAC 2009, Calgary, August 13-14, 2009, revised selected papers, vol. 5867 of lecture notes in computer science, pp. 16–35. Springer, Berlin (2009)
22. Minier, M., Phan, R.C., Pousse, B.: Distinguishers for ciphers and known key attack against rijndael with large blocks. In: Preneel, B. (ed.) Progress in cryptology - AFRICACRYPT 2009, Second international conference on cryptology in Africa, Gammarth, June 21-25, 2009, Proceedings, vol. 5580 of lecture notes in computer science, pp. 60–76. Springer, Berlin (2009)
23. Sun, B., Liu, M., Guo, J., Qu, L., Rijmen, V.: New insights on aes-like SPN ciphers. In: Robshaw, M., Katz, J. (eds.) Advances in cryptology - CRYPTO 2016 - 36th annual international cryptology conference, Santa Barbara, August 14-18, 2016, Proceedings, Part I, vol. 9814 of lecture notes in computer science, pp. 605–624. Springer, Berlin (2016)
24. Sun, B., Liu, Z., Rijmen, V., Li, R., Cheng, L., Wang, Q., AlKhzaimi, H., Li, C.: Links among impossible differential, integral and zero correlation linear cryptanalysis. In: Gennaro, R., Robshaw, M. (eds.): Advances in cryptology - CRYPTO 2015 - 35th annual cryptology conference, Santa Barbara, August 16-20, 2015, Proceedings, Part I, vol. 9215 of lecture notes in computer science, pp. 95–115. Springer, Berlin (2015)
25. Wang, M., Cui, T., Chen, H., Sun, L., Wen, L., Bogdanov, A.: Integrals go statistical: Cryptanalysis of full skipjack variants. In: Peyrin, T. (ed.) Fast software encryption - 23rd international conference, FSE 2016, Bochum, March 20-23, 2016, revised selected papers, vol. 9783 of lecture notes in computer science, pp. 399–415. Springer, Berlin (2016)