# Extending Higher-Order Integral: An Efficient Unified Algorithm of Constructing Integral Distinguishers for Block Ciphers

Wentao Zhang[1], Bozhan Su[2], Wenling Wu[1], Dengguo Feng[2], and Chuankun Wu[1]

[1] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, P.R. China
[2] Institute of Software, Chinese Academy of Sciences, Beijing, P.R. China
zhangwt06@yahoo.com, {subozhan,wwl,feng}@is.iscas.ac.cn, chuankun.wu@gmail.com

**Abstract.** In this paper, we give an extension of the concept of higher-order integral, which can make us design better higher-order integral distinguishers for some block ciphers (structures). Using the new extension, we present a unified algorithm of searching for the best possible higher-order integral distinguishers for block ciphers. We adopt the inside-out approach, trying to predict the behavior of a set of carefully chosen data, not only along encryption direction, but also along decryption direction. Applying the unified algorithm, we search for the best possible higher-order integral distinguishers of Gen-SMS4 structure, Gen-Fourcell structure and Present. For Gen-SMS4 structure and Present, the best higher-order integral distinguishers given by our algorithm are better than the best results known so far. For Gen-Fourcell structure, the best higher-order integral distinguishers given by our algorithm are the same as the best results known so far. We expect that the inside-out method is helpful to understand higher-order integral of block ciphers better, and the unified algorithm presented in this paper can be used as a tool for efficiently evaluating the security of a block cipher against integral cryptanalysis.

**Keywords:** block cipher, integral cryptanalysis, higher-order integral, integral distinguisher, generalized Feistel structure, Present.

## 1 Introduction

Integral cryptanalysis [15] is originally proposed by L.R.Knudsen and D.Wagner as a dedicated attack against Square block cipher [8], so is firstly known as "Square attack". Afterwards, the original idea used in Square attack has been extended and given different names, including saturation attack [18], collision attack [12], multiset attack [5] and integral cryptanalysis [15].

Integral cryptanalysis is of particular significance for its applicability to AES. AES is designed to be resistant to differential cryptanalysis and linear cryptanalysis, and very successful in this aspect, only 6-round AES can be resistant to differential cryptanalysis and linear cryptanalysis. However, 6-round AES can be broken using integral cryptanalysis, only with $6 \cdot 2^{32}$ chosen plaintexts

F. Bao, P. Samarati, and J. Zhou (Eds.): ACNS 2012, LNCS 7341, pp. 117–134, 2012.

and $2^{44}$ time [11]. Up to now, integral cryptanalysis is one of the most effective attacks for round-reduced AES [11,12] and round-reduced IDEA block cipher [4].

Integral cryptanalysis is a chosen-plaintext attack, which considers the propagation of sums of many values. The goal of an attacker is to derive information about the secret key using integral distinguishers. Assume a block cipher has $n$ data subblocks, each data subblock has a length of $m$ bits. When mounting an integral attack, the attacker typically chooses one or several specific subblocks, assume he chooses $d$ subblocks. Then, the attacker chooses $2^{d\times m}$ plaintexts, which take on all possible values in the $d$ subblocks, and have constant values in the other subblocks. The attacker considers these $2^{d\times m}$ chosen plaintexts at a time, trying to predict the properties in some subblock(s) after a certain number of encryption rounds. Customarily, the following four properties are considered:

(1)Constant: The state of a subblock is called "constant" if every data in this subblock has the same constant value.

(2)Active: The state of a subblock is called "active" either if the data in this subblock are all different and have constant values in the other subblocks, or if the data can be divided into some pairwise disjoint subsets and the following condition holds for each subset: the data in this subblock are all different and have constant values in the other subblocks.

(3)Balanced: The state of a subblock is called "balanced" if the XOR of all values is zero.

(4)Unkown: The state of a subblock is called "unknown" if no information is known.

We collectively call the above four states as integral states. Notice that some of the properties are implied by others. For example, a constant or active state is automatically balanced.

The security of a block cipher against integral cryptanalysis depends on several factors, including the length of integral distinguishers, specific input/output forms, the strength of one-round encryption/decryption. Among them, the design of integral distinguishers is the most important. In spite of a long time study of integral cryptanalysis on block ciphers, integral distinguishers have often been designed based on ad hoc approaches and the experience of cryptanalysts. There is no common method of designing integral distinguishers so far.

In this paper, we give an extension of the concept of higher-order integral. Furthermore, based on the new extension, we present an efficient unified algorithm to the design of higher-order integral distinguishers using the method of symbol calculation. The main ideas and contributions are as follows:

• The actual value of a constant state has no influence on the attack, thus all constant states can be denoted as a single letter "C"; A balanced state is usually a sum of some active states. Hence, the state of any subblock can be expressed either as "C", or a sum of some active states and some unknown states. Note that an unknown state is a sum of 0 active state and 1 unknown state. Compared with the customary description, the above expression is more accurate, thus makes the information kept as undamaged as possible.

• Traditionally, integral distinguishers are designed from top to bottom, an attacker tries to predict the behavior of a set of carefully chosen plaintexts after a certain number of encryption rounds. By contrast, we adopt the inside-out approach, trying to predict the behavior of a set of carefully intermediate data, not only after a certain number of encryption rounds, but also after a certain number of decryption rounds. Consequently, we make an extension of the concept of higher-order integral, which can make us design more effective integral distinguishers for some block ciphers (structures).

• Using the matrix method introduced in [13, 14], we propose an efficient unified algorithm of designing the best possible integral distinguishers for block ciphers (structures). The algorithm can be applied widely, not only for byte-oriented block ciphers and some generalized Feistel structures such as AES, Camellia [2], Gen-SMS4 structure [3] and Gen-Fourcell structure [7], but also for bit-oriented block ciphers such as Noekeon [9], Serpent [1] and Present [6]. For Camellia, Gen-SMS4, Noekeon, Serpent and Present, the best integral distinguishers given by our algorithm are better than the best results known so far. For AES and Gen-Fourcell, the best integral distinguishers given by our algorithm are the same as the best results known so far. Hence, we believe that the unified algorithm presented in this paper can be used as a tool for efficiently evaluating the security of block ciphers against integral cryptanalysis.

Due to the length limitation of this paper, we only use Gen-SMS4 structure, Gen-Fourcell structure and Present as 3 typical examples. More examples will be presented in the extended paper.

The focus of this paper is the construction of integral distinguishers for block ciphers. How to design an attack algorithm using these integral distinguishers is out of the scope of this paper, and we leave it for further work.

## 2 Preliminaries

Throughout this paper, we always assume that: (1) A block cipher structure $\mathbb{S}$ has $n$ data subblocks; (2) The round functions $F$ of $\mathbb{S}$ are all bijective; (3) The operation to connect a subblock with another one is $\oplus$, thus the sum in integral cryptanalysis considered in this paper is referred to as "$\oplus$". Although some block ciphers do not satisfy all the above conditions, e.g., IDEA and RC6, yet we believe that the similar idea can also be applied, with some modifications.

### 2.1 Higher-Order Integral

The concept of higher-order integral is proposed by L.R.Knudsen and D.Wagner [15]. Consider a set of $2^m$ elements (representing a set of plaintexts), which differ only in one particular subblock, such that each of the $2^m$ possible values for this particular subblock occurs exactly once, the sum over the elements of this set is called a **first-order integral**. Consider next a set of $2^{d\times m}$ elements, which differ in $d$ subblocks, such that each of the $2^{d\times m}$ possible values for the $d$-tuple of values from these subblocks occurs exactly once, the sum of this set is called a ***dth*-order integral**, and **integral** for short. A $dth$-order integral is called a **higher-order integral** when $d \geq 2$.

Consider a set $\bar{S} = S_1 \cup \cdots \cup S_s$ composed of $s$ sets, where each $S_i$ forms an integral. Then, clearly, if one can determine the sum of the elements of $S_i$ for each $i$, then one can also determine the sum of all elements in $\bar{S}$. This fact is the key point for understanding higher-order integral.

### 2.2 Matrix Characterization of a Block Cipher Structure

Modern block ciphers are designed by iterating a round function certain times. The following gives a matrix characterization of one round of a block cipher.

**Definition 1. [14]** *(**Encryption/Decryption Characteristic Matrix**) For a block cipher structure $\mathbb{S}$, let $(X_0, X_1, \cdots, X_{n-1})$ and $(Y_0, Y_1, \cdots, Y_{n-1})$ respectively denote the input and output of one-round encryption, then the $n \times n$ encryption/decryption characteristic matrix are defined as follows:*
***(1)Encryption characteristic matrix** $\mathcal{E}_{n\times n}$: If $Y_j = X_i \oplus R$, where $R$ is some value [1], the $(i, j)$ entry of $\mathcal{E}$ is set to 1; If $Y_j$ is nonlinearly affected by $X_i$, the $(i, j)$ entry of $\mathcal{E}$ is set to 2; If $Y_j$ is not affected by $X_i$, the $(i, j)$ entry of $\mathcal{E}$ is set to 0.*
***(2)Decryption characteristic matrix** $\mathcal{D}_{n\times n}$: If $X_j = Y_i \oplus T$, where $T$ is some value, the $(i, j)$ entry of $\mathcal{D}$ is set to 1; If $X_j$ is nonlinearly affected by $F(Y_i)$, the $(i, j)$ entry of $\mathcal{D}$ is set to 2; If $X_j$ is not affected by $Y_i$, the $(i, j)$ entry of $\mathcal{D}$ is set to 0.*

In Definition 1, each entry of the encryption/decryption characteristic matrix has only one of the three values: 0, 1, or 2. For byte-oriented block ciphers, such as AES and Camellia, the length of a subblock is chosen to be 8 bits; For bit-oriented block ciphers, such as Noekeon, Serpent and Present, the length of a subblock is chosen to be 1 bit. For some block ciphers, one characteristic matrix is sufficient to describe one-round encryption (decryption); While for some other block ciphers, it needs a composition of two characteristic matrices, i.e., firstly the first matrix $E_1$ ($D_1$), then the second matrix $E_2$ ($D_2$). In the following, we can see that most popular block ciphers can be represented by one characteristic matrix or a composition of two characteristic matrices.

## 3 A Unified Approach for the Design of Integral Distinguishers

### 3.1 A New Representation of the 4 kinds of Integral States

The following observations are very important to our new representation:

(1) For a constant subblock state, it is sufficient to know that it is a constant state, and ignoring its exact value. Thus, we can label all constant subblock states with a single letter "C".

[1] Note that it is a formal expression, $R$ can either be independent of $X_i$, or have a nonlinear relation with $X_i$.

(2) Generally, a balanced subblock state is produced by an XOR sum of some active states. Thus, we can express a balanced subblock state as $\bigoplus_{i \in I_A} A_i$, where $A_i$ denotes an active subblock state, $I_A$ is the index set, note that the constant monomial is ignored. Compared with a single letter "B", it is more accurate to label a balanced state as an XOR sum of some active states.

(3) For an unknown subblock state, we can express it as $(\bigoplus_{i \in I_A} A_i) \oplus (\bigoplus_{j \in I_?} ?_j)$, where $A_i$ denotes an active state, $?_j$ denotes an unknown state, $I_A$ and $I_?$ is the index set respectively, $I_?$ is not empty, similarly the constant monomial is ignored. Compared with a single letter "?", such an expression is more accurate.

**Example 1.** Assume a block cipher has 2 subblocks, the state is $(?_0, ?_0 \oplus A_0)$ at some point, where $?_0$ denotes an unknown state, $A_0$ an active state. Considering XOR sum of the values in the two subblocks, we can get that $?_0 \oplus (?_0 \oplus A_0) = A_0$. However, if we just express the state as $(?, ?)$, we can get nothing. □

Based on the above 3 observations, an integral state is not limited to the 4 types: constant, active, balanced or unknown. An integral state may have much more types, it can be either "C", or an XOR sum of some active states and some unknown states. The following gives a formal description.

**Definition 2.** ***(Integral Form in Subblock)****For a given set of plaintexts or intermediate data blocks, fixing a subblock, define integral form in the subblock as*

$$<< A.set, A.maxs >, < U.set, U.maxs >>$$

*where $A.set$ is a set consisting of some active subblock states, $U.set$ is a set consisting of some unknown subblock states. $A.maxs$ is defined as the maximum subscript in $A.set$ plus 1 (i.e., let $h$ be the maximum subscript among all the elements of $A.set$, then $A.maxs \equiv h + 1$), especially $\emptyset.maxs \equiv 0$ for an empty set $\emptyset$. Similarly, $U.maxs$ is defined.*

Notice that $A.maxs$ ($U.maxs$) is necessary in Definition 2 for the expression of a newly-produced active (unknown) subblock state.

**Example 2.** (**Integral form in sublock**) For a constant subblock state, its integral form is $<< \emptyset, 0 >, < \emptyset, 0 >>$; For an active subblock state $A_0$, its integral form is $<< \{A_0\}, 1 >, < \emptyset, 0 >>$; For an integral subblock state $A_0 \oplus A_2 \oplus ?_1 \oplus ?_5$, its integral form is $<< \{A_0, A_2\}, 3 >, < \{?_1, ?_5\}, 6 >>$. □

On the other hand, let $<< A.set, A.maxs >, < U.set, U.maxs >>$ denote an integral form in subblock, define $Unionset \equiv A.set \cup U.set$, where "∪" is the operation of set union, then the integral state in this subblock is just the XOR sum of all elements in $Unionset$, we will use this representation together with integral form in subblock defined in Definition 2 interchangeably in the following.

Assume a block cipher has $n$ data subblocks, naturally, we can define integral form in block.

**Definition 3.** ***(Integral Form in Block)*** *For a given set of plaintexts or intermediate data blocks, define its integral form as* $(\alpha_0, \alpha_1, \cdots, \alpha_{n-1})$*, where* $\alpha_i$ *is the integral form in subblock corresponding to the* $i$*-th subblock,* $0 \leq i \leq n-1$*.*

We will simply write "integral form" instead of "integral form in (sub)block", when the context is clear.

### 3.2 Rules for Applying Encryption/Decryption Characteristic Matrix to An Integral Form in Block

For a given set of plaintexts or intermediate data blocks, we can determine its integral form in block. Next, we need to define rules to calculate the integral form after one-round encryption/decryption. In the following, we only focus on encryption process, since decryption process can be treated similarly.

Firstly, we define an operator "$\uplus$" between two integral forms in subblock, this operator is something like adding mod 2.

**Definition 4.** *Let* $\mu$ *and* $\nu$ *be two integral forms in subblock, let*

$$\mu =<< (A.set)_\mu, (A.maxs)_\mu >, < (U.set)_\mu, (U.maxs)_\mu >>,$$
$$\nu =<< (A.set)_\nu, (A.maxs)_\nu >, < (U.set)_\nu, (U.maxs)_\nu >>$$

*Define* $\omega = \mu \uplus \nu =<< (A.set)_\omega\, (A.maxs)_\omega >, < (U.set)_\omega, (U.maxs)_\omega >>$*, thereinto,*

$$(A.set)_\omega \equiv ((A.set)_\mu \setminus (A.set)_\nu) \cup ((A.set)_\nu \setminus (A.set)_\mu),$$
$$(U.set)_\omega \equiv ((U.set)_\mu \setminus (U.set)_\nu) \cup ((U.set)_\nu \setminus (U.set)_\mu),$$
$$(A.maxs)_\omega \equiv Maxsubscript((A.set)_\omega) + 1,$$
$$(U.maxs)_\omega \equiv Maxsubscript((U.set)_\omega) + 1$$

*where "*$\setminus$*" is the operation of set minus, "*$\cup$*" is the operation of set union, and* $Maxsubscript(X)$ *function returns the maximum subscript in* $X$*.*

**Example 3.** Here is an example of set minus "$\setminus$". Let $(A.set)_\mu = \{A_0, A_1, A_3\}$ and $(A.set)_\nu = \{A_1, A_2, A_4\}$. Then, $(A.set)_\mu \setminus (A.set)_\nu = \{A_0, A_3\}$, $(A.set)_\nu \setminus (A.set)_\mu = \{A_2, A_4\}$. □

Now, we are ready to present the rules.

**Definition 5.** *Let* $\mathcal{E}_{n\times n} = [e_{ij}]_{n\times n}$ *be the encryption characteristic matrix of a block cipher. For a given set of plaintexts or intermediate data blocks, let* $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_{n-1})$ *be its integral form. Let* $\gamma = \mathcal{E}_{n\times n}(\alpha) = (\gamma_0, \gamma_1, \cdots, \gamma_{n-1})$ *be the integral form of the outputs after one-round encryption, then* $\gamma_i$ *is defined as* $\gamma_i = \biguplus_{j=1}^{n} e_{ij}(\alpha_j)$*, where* $e_{ij}(\alpha_j)$ *means applying the transformation* $e_{ij}$ *to* $\alpha_j$*.*

The entry $e_{ij}$ has 3 possible values, 0, 1, or 2. Thereinto, 0 is the zero transformation, which transforms any integral form $x$ to $<< \emptyset, 0 >, < \emptyset, 0 >>$; 1 is the identical transformation, which transforms any integral form $x$ to $x$ ; 2 is a bijective transformation, which transforms a constant state to a constant state, an active state to a new active state, and any other state to a new unknown

**Table 1.** Rules for Applying $0, 1, 2$ to an Integral Form in Subblock–Along Encryption Direction

| Trans. | input | output |
|---|---|---|
| 0 | $x$ | $<<\emptyset, 0>, <\emptyset, 0>>$ |
| 1 | $x$ | $x$ |
| 2 | $C$ | $C$ |
| | $A_i$ | $A_{emaxsA+1}$ |
| | otherwise | $U_{emaxsU+1}$ |

state. Table 1 summarizes the above rules, where $emaxsA$ ($emaxsU$) denotes the maximum subscript of all active(unknown) states brought forth so far, along encryption direction.

**Example 4.** The structure of SMS4 [3] is a kind of 4-branch generalized Feistel structure (denoted as Gen-SMS4), one round encryption of Gen-SMS4 is described as follows:

$$Y_0 = X_1, Y_1 = X_2, Y_2 = X_3, Y_3 = X_0 \oplus F(X_1 \oplus X_2 \oplus X_3)$$

It needs two characteristic matrices to describe one-round encryption (decryption) of Gen-SMS4, i.e., firstly the first matrix $E1_{GenSMS4}$ ($D1_{GenSMS4}$), then the second matrix $E2_{GenSMS4}$ ($D2_{GenSMS4}$). The uppermost row is the 0-th row, the leftmost column is the 0-th column. The encryption and decryption characteristic matrices of Gen-SMS4 are as follows:

$$E1_{GenSMS4} = \begin{pmatrix} 0,1,1,1 \\ 1,0,0,0 \\ 0,1,0,0 \\ 0,0,1,0 \end{pmatrix}, \qquad E2_{GenSMS4} = \begin{pmatrix} 0,0,1,0 \\ 0,0,0,1 \\ 1,0,1,1 \\ 2,1,0,0 \end{pmatrix}$$

$$D1_{GenSMS4} = \begin{pmatrix} 1,1,1,0 \\ 0,1,1,0 \\ 1,0,1,0 \\ 0,0,0,1 \end{pmatrix}, \qquad D2_{GenSMS4} = \begin{pmatrix} 2,0,0,1 \\ 1,1,0,0 \\ 1,0,1,0 \\ 1,1,1,0 \end{pmatrix}$$

Assume an attacker chooses a set of $2^m$ data, which has the form of $\{(c_0, x \oplus c_1, x \oplus c_2, x \oplus c_3)\}$, where $x$ takes on all the $2^m$ possible values, $c_0, c_1, c_2$ and $c_3$ are 4 constants. The integral form of the data set is $\alpha^0 = (C, A_0, A_0, A_0)$. Applying $E1_{GenSMS4}$ to $\alpha^0$, using Def. 5 and the rules in Table 1, we can get:

$$\begin{pmatrix} 0,1,1,1 \\ 1,0,0,0 \\ 0,1,0,0 \\ 0,0,1,0 \end{pmatrix} \begin{pmatrix} C \\ A_0 \\ A_0 \\ A_0 \end{pmatrix} = \begin{pmatrix} A_0 \\ C \\ A_0 \\ A_0 \end{pmatrix}$$

Next, applying $E2_{GenSMS4}$ to $(A_0, C, A_0, A_0)$, we can get:

$$\begin{pmatrix} 0,0,1,0 \\ 0,0,0,1 \\ 1,0,1,1 \\ 2,1,0,0 \end{pmatrix} \begin{pmatrix} A_0 \\ C \\ A_0 \\ A_0 \end{pmatrix} = \begin{pmatrix} A_0 \\ A_0 \\ A_0 \\ A_1 \end{pmatrix}$$

Hence the integral form of the outputs after one-round encryption is $\alpha^1 = (A_0, A_0, A_0, A_1)$.

Along the encryption direction, let $\alpha^i = (\alpha_0^i, \alpha_1^i, \cdots, \alpha_{n-1}^i)$ denote the integral form of the outputs after $i$-round encryption, $i = 1, 2, \cdots$. Let $\beta^0 = \alpha^0$, along the decryption direction, let $\beta^j = (\beta_0^j, \beta_1^j, \cdots, \beta_{n-1}^j)$ denote the integral form of the outputs after $j$-round decryption. Similarly, we can calculate $\alpha^i$ for $i = 2, 3, \cdots$, and $\beta^j$ for $j = 1, 2, 3, \cdots$. Table 2 presents the results.

**Table 2.** An Example of Gen-SMS4: Application of Def. 5 and Rules in Table 1

| $\chi$ | $\chi_0$ | $\chi_1$ | $\chi_2$ | $\chi_3$ |
|---|---|---|---|---|
| $\beta^5$ | $A(0,2)$ | $A_1$ | $A_0$ | $A_0$ |
| $\beta^4$ | $A_1$ | $A_0$ | $A_0$ | $A_0$ |
| $\beta^3$ | $A_0$ | $A_0$ | $A_0$ | $C$ |
| $\beta^2$ | $A_0$ | $A_0$ | $C$ | $A_0$ |
| $\beta^1$ | $A_0$ | $C$ | $A_0$ | $A_0$ |
| $\alpha^0 = \beta^0$ | $C$ | $A_0$ | $A_0$ | $A_0$ |
| $\alpha^1$ | $A_0$ | $A_0$ | $A_0$ | $A_1$ |
| $\alpha^2$ | $A_0$ | $A_0$ | $A_1$ | $A(0,2)$ |
| $\alpha^3$ | $A_0$ | $A_1$ | $A(0,2)$ | $A_0 \oplus ?_0$ |
| $\alpha^4$ | $A_1$ | $A(0,2)$ | $A_0 \oplus ?_0$ | $A_0 \oplus ?_1$ |
| $\alpha^5$ | $A(0,2)$ | $A_0 \oplus ?_0$ | $A_0 \oplus ?_1$ | $A_1 \oplus ?_2$ |

$\chi_k$ : integral form in the $k$-th subblock of $\alpha^i$ or $\beta^j$, $k = 0, \cdots, n-1$;
$A(i, j, \cdots, k)$ : a simplified expression for $A_i \oplus A_j \oplus \cdots \oplus A_k$;
$?(i, j, \cdots, k)$ : a simplified expression for $?_i \oplus ?_j \oplus \cdots \oplus ?_k$.

In fact, Table 2 presents a 10-round integral distinguisher for Gen-SMS4, we will give more explainations in the following sections. □

### 3.3 Finishing Conditions for Calculus and an Extension of Higher-Order Integral

For a given set of plaintexts or intermediate data blocks, Def. 5 and Table 1 show that we can calculate the integral form of the outputs after one-round encryption. Decryption process can be treated similarly. Theoretically, such a process can be iterated for arbitrary number of rounds, either along encryption direction, or along decryption direction. However, we must give some restrictions to terminate the process for deriving useful integral distinguishers.

**Finishing Condition along Encryption Direction.** Let $(\alpha_0, \alpha_1, \cdots, \alpha_{n-1})$ be an integral form in block. If there exists a subset $I^* \subseteq \{0, 1, \cdots, n-1\}$, such that the *U.set* of $\biguplus_{i \in I^*} \alpha_i$ is empty, that means, $\biguplus_{i \in I^*} \alpha_i$ is either a constant state or an XOR sum of some active states. In either case, the attacker can derive useful information from the corresponding set of data.

On the other hand, if the *U.set* of $\biguplus_{i \in I} \alpha_i$ is non-empty for every subset $I \subseteq \{0, 1, \cdots, n-1\}$, then the attacker can derive nothing from the corresponding set of data. The following gives a formal definition.

**Definition 6.** ***(Integral-Nothing)*** *An integral form in block* $(\alpha_0, \alpha_1, \cdots, \alpha_{n-1})$ *is called integral-nothing, if the U.set of* $\biguplus_{i \in I} \alpha_i$ *is non-empty for every subset* $I \subseteq \{0, 1, \cdots, n-1\}$.

If an integral form is integral-nothing, the attacker can derive nothing from the corresponding set of data. Hence, along encryption direction, when the integral form becomes integral-nothing, the attacker should terminate the process.

**Example 5.** This example comes from a 18-round integral distinguisher of Gen-Fourcell structure, see Table 3 for more details.

We have $\alpha^{15} = (\alpha_0^{15}, \alpha_1^{15}, \alpha_2^{15}, \alpha_3^{15}) = (A(0,1) \oplus ?(0,2), A(0,1,2) \oplus ?(0,1,2,3), A(2,3) \oplus ?(1,3,4), A(3,4) \oplus ?_4)$. Every component of $\alpha^{15}$ has unkown ingredients, but we have $\alpha_0^{15} \uplus \alpha_1^{15} \uplus \alpha_2^{15} \uplus \alpha_3^{15} = A_4$, thus $\alpha^{15}$ is not integral-nothing.

Applying the encryption matrices of Gen-Fourcell to $\alpha_{15}$, we get $\alpha^{16} = (\alpha_0^{16}, \alpha_1^{16}, \alpha_2^{16}, \alpha_3^{16}) = (A(0,1,2) \oplus ?(0,1,2,3), A(2,3) \oplus ?(1,3,4), A(3,4) \oplus ?_4, A(0,1,4) \oplus ?(0,2,5))$, we can verify that $\alpha^{16}$ is integral-nothing. □

**Finishing Condition along Decryption Direction and an Extension of Higher-Order Integral.** The finishing condition is different along decryption direction.

In the original definition [15] (also refer to section 2.1), $d$th-order integral is related to a set of $2^{d \times m}$ elements, which differ only in $d$ subblocks. However, we argue that the linear relations among different subblocks should be taken into account. In the following, we give an extension of higher-order integral, a $d$th-order integral is related to $2^{d \times m}$ elements, but they can differ in $d^*$ subblocks where $d^* \geq d$. To do this, we should firstly define "integral order of an integral form", which takes the linear relations among different subblocks into account.

**Definition 7.** ***(Integral Order of an Integral Form)*** *Given a set of plaintexts or intermediate data blocks, let* $\beta = (\beta_0, \beta_1, \cdots, \beta_{n-1})$ *be its integral form, and let*

$$\beta_i = << (A.set)_i, (A.maxs)_i >, < (U.set)_i, (U.maxs)_i >>$$

*for* $i = 0, 1, \cdots, n-1$. *Let dmaxsA (respectively dmaxsU) denote the maximum subscript among all active (respectively unknown) states brought forth so far along decryption direction (If no unknown state is brought forth, then* $dmaxsU \equiv -1$*), note that they are irrelevant with the process along encryption direction.*

*Denote* $w \equiv (dmaxsA+1)+(dmaxsU+1)$*, construct a* $n \times w$ *matrix* $\mathcal{G}_{n\times w} = (g_{ij})$ *as follows: Each element* $g_{ij}$ *is firstly initialized as* $0$*,* $0 \le i \le n-1$*,* $0 \le j \le w-1$*. Next, if* $A_j \in (A.set)_i$*, then* $g_{ij}$ *is modified to 1, for* $j = 0, 1, \cdots, dmaxsA$*; If* $U_j \in (U.set)_i$*, then* $g_{i(j+dmaxsA+1)}$ *is modified to 1, for* $j = 0, 1, \cdots, dmaxsU$*. Notice that the* $i$*-th row is completely determined by* $\beta_i (i = 0, 1, \cdots, n-1)$*, and the first* $(dmaxsA + 1)$ *columns are corresponding to the active ingredients, the last* $(dmaxsU + 1)$ *columns corresponding to the unknown ingredients. Define the* ***integral order*** *of* $\beta$ *as* $d = rank(\mathcal{G}_{n\times w})$*, here* $rank(\mathcal{G}_{n\times w})$ *is the rank of* $\mathcal{G}_{n\times w}$*, where* $\mathcal{G}$ *is regarded as a matrix over* $GF(2)$*.*

Integral cryptanalysis is a kind of chosen-plaintext attack, the attacker can choose a priori a set of plaintexts and obtain the corresponding ciphertexts. For a successful attack, the amount of the chosen plaintexts must be less than $2^l$, where $l$ $(= n \times m)$ is the block length. An integral form with an integral order $d$ is corresponding to $2^{d\times m}$ data (accordingly, corresponding to $2^{d\times m}$ plaintexts), hence $d$ must satisfy that $d \le n-1$. Therefore, when the integral order $d$ of the integral form satisfies that $d = n$ after some decryption rounds (since there are $n$ subblocks in total, integral order can not be larger than $n$), the attacker should terminate the process.

Along decryption direction, let $d_j$ denote the integral order of the integral form of the outputs after $j$-round decryption, $j = 1, 2, \cdots$. Due to diffusion of the block cipher (structure) along decryption direction, $d_j$ will increase or keep unchanged as $j$ increases. Hence, there must exist a unique $t$ which satisfies that $d_t \le n-1$ and $d_{t+1} = n$. That means, the attacker should terminate the process after $(t+1)$ rounds along decryption direction. We call $d_t$ the integral order of the corresponding integral distinguisher, and the distinguisher is called a $d_t th$-order integral distinguisher. The following gives a formal description.

**Definition 8.** ***(Integral Order of an Integral Distinguisher)*** *Let* $Dis$ *be an integral distinguisher, which is constructed by the above inside-out approach. Let* $d_j$ *denote the integral order of the integral form of the outputs after* $j$*-round decryption, then there must exist a unique integer* $t$ *satisfying* $d_t \le n-1$ *and* $d_{t+1} = n$*.* $d_t$ *is called the integral order of* $Dis$*, and* $Dis$ *is called a* $d_t th$*-order integral distinguisher.*

**Example 6.** Considering the integral distinguisher of Gen-SMS4 in Table 2.

We have $\beta^5 = (A_0 \oplus A_2, A_1, A_0, A_0)$, the corresponding matrix $\mathcal{G}_{\beta^5}$ is :

$$\mathcal{G}_{\beta^5} = \begin{pmatrix} 1,0,1 \\ 0,1,0 \\ 1,0,0 \\ 1,0,0 \end{pmatrix}$$

The rank of $\mathcal{G}_{\beta^5}$ is 3, thus the integral order of $\beta^5$ is 3.

Applying the composition of $D1_{GenSMS4}$ and $D2_{GenSMS4}$ to $\beta^5$, we get $\beta^6 = (A_0 \oplus ?_0, A_0 \oplus A_2, A_1, A_0)$, the corresponding matrix $\mathcal{G}_{\beta^6}$ is :

$$\mathcal{G}_{\beta^6} = \begin{pmatrix} 1,0,0,1 \\ 1,0,1,0 \\ 0,1,0,0 \\ 1,0,0,0 \end{pmatrix}$$

The first 3 columns of $\mathcal{G}_{\beta^6}$ are corresponding to active ingredients, and the last column to unknown ingredients. The rank of $\mathcal{G}_{\beta^6}$ is 4, i.e., the integral order of $\beta^6$ is 4, which is equal to the number of subblocks. Hence, the attacker should terminate the process aftre 6-round decryption along decryption direction. The integral order of this distinguisher is equal to the integral order of $\beta^5$, i.e., 3. □

In Section 4, we will see that better higher-order integral distinguishers can be constructed using our new extension of higher-order integral, including Gen-SMS4 structure and Present.

Let $Dis$ be an integral distinguisher for a block cipher (structure), which is constructed according to the rules and finishing conditions in Section 3.1-3.3. Assume $Dis$ has $w$ rounds along encryption direction, $t$ rounds along decryption direction, let $\beta_j$ denote the integral form of the outputs after $j$-round decryption along decryption direction, and let $d_j$ denote the integral order of $\beta_j$. Now, we present some details about the part of $Dis$ along decryption direction. We will see that $Dis$ is indeed a $(w + t)$-round integral distinguisher.

Note the following facts:

**(1.)** Firstly, considering the outputs after $(j+1)$-round decryption. The attacker can choose $d_{j+1}$ independent subblocks, which take on all possible values (corresponding to $2^{d_{j+1} \times m}$ data); For each of the other $(n - d_{j+1})$ subblocks, the state is either constant, or the value in this subblock can be linearly determined by the values in the chosen $d_{j+1}$ subblocks. Thus, the attacker chooses a set of $2^{d_{j+1} \times m}$ data blocks, this set is denoted as $\Omega^{j+1}$.

**(2.)** Secondly, considering the set of the 1-round encryption outputs of all of the elements of $\Omega^{j+1}$, we will get a new set of $2^{d_{j+1} \times m}$ data blocks, which is denoted as $\Omega^j$. Since $\beta^{j+1}$ and $\beta_j$ are correlated by the decryption characteristic matrices and the calculus rules, also $\Omega^{j+1}$ and $\Omega^j$ are correlated by the one-round encryption function, the $2^{d_{j+1} \times m}$ elements of $\Omega^j$ can be separated into $d_{j+1}/d_i$ groups, which satisfy the following condition: each group has $2^{d_j \times m}$ elements with an integral form of $\beta_j$, thus each group is corresponding to a $(w + j)$-round integral distinguisher. Hence, if the sum of the outputs after $(w + j)$-round decryption is zero (corresponding to $2^{d_j \times m}$ data blocks), then the sum of the outputs after $(w+j+1)$-round decryption is also zero (corresponding to $2^{d_{j+1} \times m}$ data blocks), since the XOR of many zeros is also zero.

**(3.)** For $Dis$, the $w$-round part along encryption direction can be regarded as a traditional integral distinguisher. Then, exucute one-round decryption, and applying the above induction to $j = 0$, we get a $(w + 1)$-round integral distinguisher. The induction can be applied iteratively along decryption direction

for $j = 0, 1, \cdots, t-1$, totally $t$ times. Finally, we get a $(w+t)$-round integral distinguisher, that is to say, $Dis$ is indeed a $(w+t)$-round integral distinguisher.

**Example 7.** Considering the integral distinguisher of Gen-SMS4 in Table 2. We have that $\beta^5 = (A_0 \oplus A_2, A_1, A_0, A_0)$ and $\beta^4 = (A_1, A_0, A_0, A_0)$. Based on $\beta^5$, the attacker chooses a set of $2^{3\times m}$ elements, which has the form of $(u_0 \oplus c_0, u_1 \oplus c_1, u_2 \oplus c_2, u_2 \oplus c_3)$, for each possible $(u_0, u_1, u_2) \in (GF(2^m))^3$, and $c_0, c_1, c_2$ and $c_3$ are $m$-bit constants. For simplicity, let $c_0 = c_1 = c_2 = c_3 = 0$. Then, applying the encryption characteristic matrices to $\beta^5$, we can get that the set of the $2^{3\times m}$ outputs after one-round encryption (this set is denoted by $\Omega$) have the form of $(u_1, u_2, u_2, u_0 \oplus F(u_1 \oplus c_4))$, where $c_4$ is a new constant (which depends on the key). In the following, we will show that the $2^{3\times m}$ elements of $\Omega$ can be divided into $2^m$ groups, each group has $2^{2\times m}$ elements, satisfying that the integral form of each group is equal to $\beta^4$.

Let *const* denote a $m$-bit constant, let $u_0 = u_2 \oplus F(u_1 \oplus c_4) \oplus const$, then we get a subset of $2^{2\times m}$ elements of $\Omega$, which have the form of $(u_1, u_2, u_2, u_2 \oplus const)$, for each possible $(u_1, u_2) \in (GF(2^m))^2$, we use $Group_{const}$ to denote this group. It is easy to see that the integral form of $Group_{const}$ is equal to $\beta^4$. There are $2^m$ possible values of *const*, thus there are $2^m$ disjoint groups, and the union of these $2^m$ groups will cover every element of $\Omega$. □

If an integral distinguisher has an integral order $d_t$, then it is corresponding to $2^{d_t \times m}$ plaintexts. Hence, the integral order of an integral distinguisher reflects the amount of data blocks needed by this integral distinguisher.

### 3.4 A Unified Algorithm of Constructing Integral Distinguishers

Based on the results of Section 3.1-3.3, we are now ready to present a unified algorithm of constructing integral distinguishers for block ciphers.

For a block cipher (structure), let $\{\mathcal{E}_{n\times n}\}/\{\mathcal{D}_{n\times n}\}$ denote its encryption/de -cryption characteristic matrices. Choose a set of data blocks, let $\alpha^0 = (\alpha_0, \alpha_1, \cdots,$
$\alpha_{n-1})$ denote its integral form. Along encryption direction, let $\alpha^i = (\alpha_0^i, \alpha_1^i, \cdots,$ $\alpha_{n-1}^i)$ denote the integral form of the outputs after $i$-round encryption, $i = 1, 2, \cdots$. Let $\beta^0 = \alpha^0$, along decryption direction, let $\beta^j = (\beta_0^j, \beta_1^j, \cdots, \beta_{n-1}^j)$ denote the integral form of the outputs after $j$-round decryption, and let $d_j$ denote the integral order of $\beta^j$, $j = 1, 2, \cdots$.

In integral cryptanalysis, an attacker is usually intended to derive the longest distinguishers. In the following, we present an algorithm to calculate the length of the longest possible integral distinguishers.

Once we get the length of the longest possible integral distinguishers using Algorithm 1, we can backtrack to derive the corresponding distinguishers. Note that there are usually many longest integral distinguishers using Algorithm 1.

In Algorithm 1, Step 2 needs to enumerate all the cases of $\alpha^0$. For some block ciphers (structures), it is impossible to enumerate all the cases due to the computing limitation. We will discuss the selection of initial integral forms $\alpha^0 (= \beta^0)$ for different block ciphers(structures) in Section 4.4.

**Algorithm 1.** Compute the Length of the Longest Possible Integral Distinguishers

**Input:** Encryption characteristic matrices $\{\mathcal{E}_{n\times n}\}$, decryption characteristic matrices $\{\mathcal{D}_{n\times n}\}$.
**Output:** The length of the longest possible integral distinguishers, denoted by $r$, and $r$ is initialized to be 0.

**Step1.** For a chosen integral form $\alpha^0 = \beta^0$, do the following:
- **(1)** Find the largest integer $s$ such that $\alpha^{s+1}$ is integral-nothing and $\alpha^s$ is not integral-nothing.
- **(2)** Find the largest integer $t$ such that $d_{t+1} = n$ and $d_t < n$.
- **(3)** Calculate $h = s + t$, then $h$ is the length of the longest integral distinguisher corresponding to $\alpha^0$. If $h > r$, let $r \leftarrow h$.

**Step2.** Repeat Step 1 until all the cases of $\alpha^0$ are enumerated.
**Step3.** Output $r$.

## 4 Experimental Results – Application to Gen-SMS4, Gen-Fourcell and Present

In this section, we present experimental results of applying Algorithm 1 to Gen-SMS4 structure [3], Gen-Fourcell structure [7] and Present [6].

Assume the round subkey is XORed with the state, and the sum in integral cryptanalysis considered is XOR sum. Thus, subkey addition has no effect on the design of integral distinguishers, we will omit it. For a given block cipher (structure), there are usually many longest possible integral distinguishers applying Algorithm 1. Although some are the same or equivalent, we will not tell them apart.

In the following, let $\chi_k$ denote the integral form in the $k$-th subblock of $\alpha^i$ (or $\beta^j$), where $i = 0, 1, \cdots$, $j = 0, 1, \cdots$ and $k = 0, \cdots, n-1$.

### 4.1 Gen-SMS4

Using Algorithm 1, we found 256 10-round integral distinguishers. Table 2 presents one: 5 rounds along encryption direction, and 5 rounds along decryption direction. From $\beta^5$, it is a $3rd$-order integral distinguisher; The attacker chooses a set of $2^{3\times m}$ elements, which has the form of $(u_0 \oplus c_0, u_1 \oplus c_1, u_2 \oplus c_2, u_2 \oplus c_3)$, for each possible $(u_0, u_1, u_2) \in (GF(2^m))^3$, and $c_0, c_1, c_2$ and $c_3$ are $m$-bit constants. Then, considering the outputs after 10-round encryption. From $\alpha_0^5 = A(0, 2)$, we can get that the XOR sum of all the $2^{3\times m}$ values in the 0th subblock (corresponding to $\chi_0$) is zero. A 8-round integral distinguisher of Gen-SMS4 is given in [17], the 10-round distinguisher in Table 2 truncated from the 3rd round to the 10th round is equivalent to the 8-round distinguisher in [17].

### 4.2 Gen-Fourcell

The structure of Fourcell [7] is also a kind of 4-branch generalized Feistel structure (denoted as Gen-Fourcell), one round of Gen-Fourcell is described as follows:

**Table 3.** A 18-round Integral Distinguisher for Gen-Fourcell

| $\chi$ | $\chi_0$ | $\chi_1$ | $\chi_2$ | $\chi_3$ |
|---|---|---|---|---|
| $\beta^3$ | $?_0$ | $A_2$ | $A_1$ | $C$ |
| $\beta^2$ | $A_2$ | $A_1$ | $C$ | $C$ |
| $\beta^1$ | $A_1$ | $C$ | $C$ | $C$ |
| $\alpha^0 = \beta^0$ | $C$ | $C$ | $C$ | $A_0$ |
| $\alpha^1$ | $C$ | $C$ | $A_0$ | $A_0$ |
| $\alpha^2$ | $C$ | $A_0$ | $A_0$ | $C$ |
| $\alpha^3$ | $A_0$ | $A_0$ | $C$ | $C$ |
| $\alpha^4$ | $A_0$ | $C$ | $C$ | $A(0,1)$ |
| $\alpha^5$ | $C$ | $C$ | $A(0,1)$ | $A(0,1,2)$ |
| $\alpha^6$ | $C$ | $A(0,1)$ | $A(0,1,2)$ | $A_2$ |
| $\alpha^7$ | $A(0,1)$ | $A(0,1,2)$ | $A_2$ | $C$ |
| $\alpha^8$ | $A(0,1,2)$ | $A_2$ | $C$ | $A(0,1) \oplus ?_0$ |
| $\alpha^9$ | $A_2$ | $C$ | $A(0,1) \oplus ?_0$ | $A(0,1,2) \oplus ?(0,1)$ |
| $\alpha^{10}$ | $C$ | $A(0,1) \oplus ?_0$ | $A(0,1,2) \oplus ?(0,1)$ | $A(2,3) \oplus ?_1$ |
| $\alpha^{11}$ | $A(0,1) \oplus ?_0$ | $A(0,1,2) \oplus ?(0,1)$ | $A(2,3) \oplus ?_1$ | $A_3$ |
| $\alpha^{12}$ | $A(0,1,2) \oplus ?(0,1)$ | $A(2,3) \oplus ?_1$ | $A_3$ | $A(0,1) \oplus ?(0,2)$ |
| $\alpha^{13}$ | $A(2,3) \oplus ?_1$ | $A_3$ | $A(0,1) \oplus ?(0,2)$ | $A(0,1,2) \oplus ?(0,1,2,3)$ |
| $\alpha^{14}$ | $A_3$ | $A(0,1) \oplus ?(0,2)$ | $A(0,1,2) \oplus ?(0,1,2,3)$ | $A(2,3) \oplus ?(1,3,4)$ |
| $\alpha^{15}$ | $A(0,1) \oplus ?(0,2)$ | $A(0,1,2) \oplus ?(0,1,2,3)$ | $A(2,3) \oplus ?(1,3,4)$ | $A(3,4) \oplus ?_4$ |

$$Y_0 = X_1, Y_1 = X_2, Y_2 = X_3, Y_3 = F(X_0) \oplus X_1 \oplus X_2 \oplus X_3$$

The encryption and decryption characteristic matrices of Gen-Fourcell are as follows:

$$E_{GenFourcell} = \begin{pmatrix} 0,1,0,0 \\ 0,0,1,0 \\ 0,0,0,1 \\ 2,1,1,1 \end{pmatrix}$$

$$D1_{GenFourcell} = \begin{pmatrix} 1,1,1,1 \\ 0,1,1,1 \\ 1,0,1,1 \\ 1,1,0,1 \end{pmatrix}, \qquad D2_{GenFourcell} = \begin{pmatrix} 2,0,0,0 \\ 1,1,0,0 \\ 1,0,1,0 \\ 1,0,0,1 \end{pmatrix}$$

Note: the uppermost row is the 0-th row, the leftmost column is the 0-th column.

Using Algorithm 1, we found 56 18-round integral distinguishers. Table 3 presents one: 15 rounds along encryption direction, and 3 rounds along decryption direction. From $\beta^3$, it is a $3rd$-order integral distinguisher; The attacker chooses a set of $2^{3\times m}$ elements, which has the form of $(u_0 \oplus c_0, u_1 \oplus c_1, u_2 \oplus c_2, c_3)$, for each possible $(u_0, u_1, u_2) \in (GF(2^m))^3$, $c_0, c_1, c_2$ and $c_3$ are $m$-bit constants. Then, considering the outputs after 18-round encryption. Based on $\alpha^{15}$, considering $\alpha_0^{15} \uplus \alpha_1^{15} \uplus \alpha_2^{15} \uplus \alpha_3^{15}$, we have $(A(0,1)\oplus?(0,2)) \uplus (A(0,1,2)\oplus?(0,1,2,3)) \uplus (A(2,3)\oplus?(1,3,4)) \uplus (A(3,4)\oplus?_4) = A_4$. Hence the XOR sum of all the $2^{3\times m}$ values of the XOR sum of the 4 subblocks is zero. In [16], a 18-round integral distinguisher of Gen-Fourcell is given. The distinguisher in Table 3 is equivalent to that in [16].

### 4.3 Present

Present [6] is a SP-network block cipher, the block length is 64. Since it is bit-oriented, we will treat a bit as a data subblock, then 64 subblocks in total, i.e., $m = 1$ and $n = 64$. Figure 1 gives the bit indexing of a 64-bit data block.

One round of Present [6] is described as $Y = Theta \circ Gamma(X)$, where $Gamma$ is the S-box layer, and $Theta$ is a linear transformation. $Gamma$ operates independently on 16 4-tuple of bits, the first S-box takes bits 0-3 as input, the next S-box takes bits 4-7 as input, and so on. Let $a_i$ denote the $i$-th bit of $a$, $i = 0, 1, \cdots, 63$, then $Theta(a_i) = a_j$, where $j = 16 \times (i \bmod 4) + \lfloor i/4 \rfloor$, $\lfloor x \rfloor$ is the integer portion of $x$.

Present uses a $4 \times 4$ S-box. Let $x = x_3x_2x_1x_0$, where $x_i$ is the $i$-th bit of $x$, $i = 0, 1, 2, 3$. Let $(\Delta x \to \Delta y)$ denote a differential with input difference $\Delta x$ and output difference $\Delta y$. For the S-box of Present, there are 3 truncated differentials with probability 1:

$$(1001 \to ***0), \quad (0001 \to ***1), \quad (1000 \to ***1)$$

where "$*$" denotes an unknown bit.

For Present, the size of the characteristic matrices is $64 \times 64$. It is impractical to use Definition 6 as the finishing condition along encryption direction. However, Present is a SP-network cipher, the outputs of different S-boxes can be regarded as being independent. Thus, Definition 6 can be revised as follows, without any effect on the design of the best possible integral distinguishers for SP-network block ciphers.

**Definition 6' (Integral-Nothing for SP-network Ciphers).** For a SP-network block cipher (structure), an integral form $(\alpha_0, \alpha_1, \cdots, \alpha_{n-1})$ is integral-nothing, if the $U.set$ of $\alpha_i$ is non-empty for every $i \in \{0, 1, \cdots, n-1\}$.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 60 | 56 | 52 | 48 | 44 | 40 | 36 | 32 | 28 | 24 | 20 | 16 | 12 | 8 | 4 | 0 |
| 61 | 57 | 53 | 49 | 45 | 41 | 37 | 33 | 29 | 25 | 21 | 17 | 13 | 9 | 5 | 1 |
| 62 | 58 | 54 | 50 | 46 | 42 | 38 | 34 | 30 | 26 | 22 | 18 | 14 | 10 | 6 | 2 |
| 63 | 59 | 55 | 51 | 47 | 43 | 40 | 35 | 31 | 27 | 23 | 19 | 15 | 11 | 7 | 3 |

**Fig. 1.** 4×16 Bit Indexing of a 64-bit Data Block

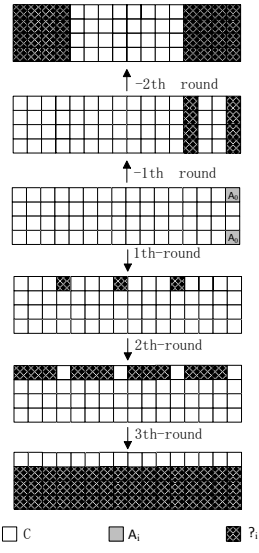


**Fig. 2.** The 5-round integral distinguisher for Present (In the initial integral form, we label $A_0$ both in bit 0 and bit 3, which means that the two bits are linearly dependent, the integral order of the integral form is 1, instead of 2

Using Algorithm 1, we found many 5-round integral distinguishers of Present. Figure 2 illustrates one of the best, which uses $Prob_{Sbox}(1001 \rightarrow ***0) = 1$, 3 rounds along encryption direction, and 2 rounds along decryption direction. It is a $32th$-order integral distinguisher. The attacker chooses a set of $2^{32}$ plaintexts, which have constant values in the following 32 bits: 0-15 and 48-63, while taking all possible values in the other 32 bits. Then, considering the outputs after 5-round encryption, the XOR sum of all the $2^{32}$ values in each of the following 16 bits is zero: 0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56 and 60. Note that this distinguisher uses the new extension concept of higher-order integral, the initial integral form has an integral order of 1, while it is related to 2 bits.

In [20], a 3-round integral distinguisher of Present is given, which has an integral order of 4. The distinguisher in Figure 2 is a 5-round integral distinguisher, furthermore, the truncated 3-round distinguisher of the last 3 rounds has an integral order of 2. Thus, our distinguisher (as illustrated in Figure 2) is more better than that in [20].

### 4.4 Selection of Initial Integral Forms

Let $(\chi_0, \chi_1, \chi_2, \cdots, \chi_{n-1})$ denote an initial integral form in block.

For Gen-Fourcell and Gen-SMS4, $\chi_i$ can be any element or an XOR combination of the elements in the set $\{C, A_0, A_1, A_2\}$, which have $2^4$ possibilities for each subblock. There are 4 subblocks, thus $2^{16} - 1$ possibilities in total. In our experiments, we have tried them exhaustively.

For Present, the integral order of an initial integral form can be $1, 2, \cdots, 63$, it is impractical to search for each case. However, Present is a SP-network cipher,

and the diffusion layer is very simple. Our experiments also show that the smaller the integral order of an initial integral form, the better the integral distinguishers. Hence, we only considered 1st-order initial integral forms.

## 5 Discussion and Conclusion

Our work in this paper is originally inspired by the work of J.Kim et al [13, 14]. In [13, 14], the authors proposed a general tool for finding impossible differentials of block cipher structures using matrix method and meet-in-the-middle approach, and applied their tool to some block ciphers (structures). They also pointed out that the matrix method can be converted into a tool for the Square attack. However, they only considered the 1st-order integral, not considering higher-order integral. In [19], Y.Y.Luo et al. greatly improved the results of [13, 14].

In this paper, we adopted inside-out approach to construct integral distinguishers for block ciphers, and extended the concept of higher-order integral by considering the linear relations among different subblocks. Furthermore, we presented an efficient unified algorithm to the design of the longest possible integral distinguishers for block ciphers. We applied the algorithm to many block ciphers (structures), the experiments showed: For Gen-SMS4 structure and Present, the best integral distinguishers given by our algorithm are better than the best results known so far; For Gen-Fourcell structure, the best integral distinguishers given by our algorithm are the same as the best results known so far.

To sum up, we believe that the inside-out method for designing integral distinguishers and the new extension of higher-order integral are helpful to better understand integral cryptanalysis of block ciphers. Also, the unified algorithm in this paper can be useful as a tool for efficiently evaluating the security of block ciphers against integral cryptanalysis.

**Acknowledgment.** We would like to thank anonymous referees for their helpful comments and suggestions. The research presented in this paper is supported by the National Natural Science Foundation of China (No.60903212), the "Strategic Priority Research Program" of the Chinese Academy of Sciences (No.XDA06010701), and the Knowledge Innovation Project of the Chinese Academy of Sciences.

## References

1. Anderson, R., Biham, E., Knudsen, L.R.: Serpent: A Proposal for the Advanced Encryption Standard. In: NIST AES Proposal (1998), `http://www.cl.cam.ac.uk/`
2. Aoki, K., Ichikawa, T., Kanda, M., Matsui, M., Moriai, S., Nakajima, J., Tokita, T.: *Camellia*: A 128-Bit Block Cipher Suitable for Multiple Platforms - Design and Analysis. In: Stinson, D.R., Tavares, S. (eds.) SAC 2000. LNCS, vol. 2012, pp. 39–56. Springer, Heidelberg (2001)
3. SMS4 Encryption Algorithm for Wireless Networks, English translation of the Chinese document, `http://eprint.iacr.org/2008/329.pdf`

4. Biham, E., Dunkelman, O., Keller, N.: A New Attack on 6-Round IDEA. In: Biryukov, A. (ed.) FSE 2007. LNCS, vol. 4593, pp. 211–224. Springer, Heidelberg (2007)
5. Biryukov, A., Shamir, A.: Structural Cryptanalysis of SASAS. In: Pfitzmann, B. (ed.) EUROCRYPT 2001. LNCS, vol. 2045, pp. 394–405. Springer, Heidelberg (2001)
6. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: An Ultra-Lightweight Block Cipher. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 450–466. Springer, Heidelberg (2007)
7. Choy, J., Chew, G., Khoo, K., Yap, H.: Cryptographic Properties and Application of a Generalized Unbalanced Feistel Network Structure. In: Boyd, C., González Nieto, J. (eds.) ACISP 2009. LNCS, vol. 5594, pp. 73–89. Springer, Heidelberg (2009)
8. Daemen, J., Knudsen, L.R., Rijmen, V.: The Block Cipher SQUARE. In: Biham, E. (ed.) FSE 1997. LNCS, vol. 1267, pp. 149–165. Springer, Heidelberg (1997)
9. Daemen, J., Peeters, M., Assche, G.V., Rijmen, V.: Nessie Proposal: NOEKEON. In: First Open NESSIE Workshop (2000), `http://gro.noekeon.org/`
10. Daemen, J., Rijnmen, V.: AES Proposal: Rijndael, `http://csrc.nist.gov/encryption/aes/rijndael`
11. Ferguson, N., Kelsey, J., Lucks, S., Schneier, B., Stay, M., Wagner, D., Whiting, D.L.: Improved Cryptanalysis of Rijndael. In: Schneier, B. (ed.) FSE 2000. LNCS, vol. 1978, pp. 213–230. Springer, Heidelberg (2001)
12. Gilbert, H., Minier, M.: A collision attack on seven rounds of Rijndael. In: Proceedings of the Third AES Candidate Conference, pp. 230–241
13. Kim, J., Hong, S., Sung, J., Lee, S., Lim, J.: Impossible Differential Cryptanalysis for Block Cipher Structures. In: Johansson, T., Maitra, S. (eds.) INDOCRYPT 2003. LNCS, vol. 2904, pp. 82–96. Springer, Heidelberg (2003)
14. Kim, J., Hong, S., Lim, J.: Impossible differential cryptanalysis using matrix method. Discrete Mathematics 310(5), 988–1002 (2010)
15. Knudsen, L.R., Wagner, D.: Integral Cryptanalysis (Extended Abstract). In: Daemen, J., Rijmen, V. (eds.) FSE 2002. LNCS, vol. 2365, pp. 112–127. Springer, Heidelberg (2002)
16. Li, R., Sun, B., Li, C., Qu, L.: Cryptanalysis of a Generalized Unbalanced Feistel Network Structure. In: Steinfeld, R., Hawkes, P. (eds.) ACISP 2010. LNCS, vol. 6168, pp. 1–18. Springer, Heidelberg (2010)
17. Liu, F., Ji, W., Hu, L., Ding, J., Lv, S., Pyshkin, A., Weinmann, R.-P.: Analysis of the SMS4 Block Cipher. In: Pieprzyk, J., Ghodosi, H., Dawson, E. (eds.) ACISP 2007. LNCS, vol. 4586, pp. 158–170. Springer, Heidelberg (2007)
18. Lucks, S.: Attacking seven rounds of Rijndael under 192-bit and 256-bit keys. In: in Proc. 3rd AES Candidate Conf., pp. 215–229 (2000)
19. Luo, Y., Wu, Z., Lai, X., Gong, G.: A Unified Method for Finding Impossible Differentials of Block Cipher Structures. Cryptology ePrint Archive: Report 2009/627
20. Z'aba, M.R., Raddum, H., Henricksen, M., Dawson, E.: Bit-Pattern Based Integral Attack. In: Nyberg, K. (ed.) FSE 2008. LNCS, vol. 5086, pp. 363–381. Springer, Heidelberg (2008)