Congruent Differential Cluster for Binary SPN Ciphers

Ting Cui, Yiming Mao, Yang Yang, Yi Zhang, Jiyan Zhang, and Chenhui Jin

***Abstract*— This study is focused on the differential clustering effect of the SPN block cipher, which employs a binary matrix as its diffusion layer. We present a novel strategy for differential estimation, named the congruent differential cluster. This method does not guarantee the optimization of each single differential characteristic but gathers a large number of characteristics satisfying a specific condition, i.e., the output differences of active S-boxes are equal. Given a binary SPN cipher, the exact probability of the congruent differential cluster can be obtained with negligible computational resources. Moreover, we consider a popular instance, binary `AES`-like ciphers, since the processing of their column-mixing layer can be divided into several independent parts. Therefore, if we set the output differences of the active S-boxes in the same partition to be equal, we can obtain more differential characteristics in the cluster, known as a semicongruent differential cluster. To demonstrate the application of the proposed method, we apply it to several block ciphers, i.e., `Midori-64`, `CRAFT-64`, `SKINNY-64` and their variants proposed in Todo and Sasaki (2022). Compared with the active S-box counting method, the congruent differential clusters have considerably higher probabilities for most instances. In addition, we find a 7-round semicongruent differential cluster for `Midori-64` with probability $2^{-52.25}$, an 8-round semicongruent differential cluster for `SKINNY-64` with probability $2^{-50.72}$ and a 10-round semicongruent differential cluster for `CRAFT-64` with probability $2^{-42.32}$. To the best of our knowledge, the semicongruent differential clusters we identify for 7-round `Midori-64`, 8-round `SKINNY-64` and 10-round `CRAFT-64` have the highest probabilities thus far among the existing differential clusters with the same rounds. Therefore, we believe that the proposed method is a valuable tool for evaluating the differential security of associated block ciphers.**

***Index Terms*— Differential cryptanalysis, congruent differential cluster, DDT, binary SPN cipher.**

I. Introduction

In THE last several decades, block ciphers have played a central role in the development of cryptology. Open research on block ciphers started with the proposal of the `DES` in 1977 [2], which provided the most important target for academic cryptanalysis. Around the 1990s, differential cryptanalysis [3], [4] and linear cryptanalysis [5] were proposed successively, which have become the two classic analyses of `DES`.

Differential cryptanalysis forced designers to reconsider their design methodologies. The basic idea to ensure robustness against cryptanalysis is to introduce an upper bound on the probability of any *differential* of the cipher. For an upper bound p, the data complexity of the attack is approximately $1/p$ [6]. In the mid 1990's, Luke O'connor [7] studied the differential and linear properties of random permutations for the first time. They give the probabilistic upper bound for a differential characteristic. However, it remains challenging to accurately estimate the probability of the differential of the target cipher.

Another milestone in the development of block ciphers is the proposal of the `AES`. To ensure robustness against differential/linear cryptanalysis, Daemen and Rijimen introduced the wide trail design strategy during the design of Rijndael [8]. This strategy helps designers more easily evaluate the security boundary against differential (and linear) cryptanalysis. Compared with previous designs, the wide trail strategy eliminates the necessity of heavy arguments or programming work in ensuring differential security. Generally, if the cipher has no less than n active S-boxes after an r-round cascade, then there never exists an r-round differential characteristic with a probability greater than p_{max}^n, where p_{max} denotes the highest differential probability of the S-box.

In recent years, many researchers have attempted to design block ciphers under the wide trail strategy framework. Furthermore, with the continuous progress in automated cryptanalysis technologies, increasing efforts are being made to evaluate the resistance of block ciphers against differential cryptanalysis by counting the active S-boxes.

For 4-round `AES`, there are at least 25 active S-boxes and the highest differential probability of `AES`'s S-box is 2^{-6}. Theoretically, there is no differential characteristic with a probability exceeding $2^{-6\times 25} = 2^{-150}$. However, in 2005, Keliher proved that for 4-round `AES`, there exists a differential hold with probability 1.881×2^{-114} [9]. This finding indicates that the probability of the identified characteristic may be considerably smaller than the differential in `AES`. Furthermore, Ankele and Kölbl [10] developed an automated approach for enumerating the characteristics with the highest probability of contributing to a differential based on SMT solvers, and in 2020, Dunkelman et al. [11] noted that counting the minimum

Manuscript received 14 June 2023; revised 20 October 2023 and 1 January 2024; accepted 2 January 2024. Date of publication 5 January 2024; date of current version 11 January 2024. This work was supported in part by the National Natural Science Foundations of China under Grant 62372463 and Grant 62302518 and in part by the Natural Science Foundation of Henan Province under Grant 222300420100. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hossein Pishro-Nik. *(Corresponding authors: Jiyan Zhang; Yang Yang.)*

The authors are with the Department of Applied Mathematics, PLA SSF Information Engineering University, Zhengzhou 450000, China (e-mail: cuiting_1209@126.com; mym220615@163.com; yangyang_wawa@sina.com; yizhang0796@foxmail.com; xdzhangjiyan@126.com; jinchenhui@126.com).

Digital Object Identifier 10.1109/TIFS.2024.3350374

number of active S-boxes may not always be sufficiently accurate. The authors introduced a 4-round Feistel with the SPSPSPSP layer as its round function, and their block cipher had at least 36 active AES S-boxes. However, there exist differentials with a considerably higher probability. Although Dunkelman's block cipher is artificial, and the evidence of the high probability differential is based on a mathematical deduction instead of the real differential, it indicates that a provable secure block cipher (other than the AES) against differential cryptanalysis may still be insecure in practice.

In ASIACRYPT 2021 [12], Leurent et al. considered the clustering effect of differential/linear properties in Simon and Simeck, resulting in stronger distinguishers than previously proposed differential/linear characteristics.

Our Contribution: This study focuses on binary SPN ciphers, i.e., ciphers that use only S-box level XOR as their diffusion layers. This kind of diffusion layer often has better hardware and software implementation as well as lower energy consumption, making them widely used in lightweight cryptography. Typical representatives are SKINNY [13], Midori [14], CRAFT [15], etc. The objective is to establish a novel differential cluster (named *congruent differential cluster* and its variant *semicongruent differential cluster* which we will define later) for such ciphers. The main contributions of this research can be summarized as follows:

- We present the theoretical framework of the congruent differential cluster for binary SPN ciphers. Instead of optimizing the probability of a single differential characteristic, we collect a large number of differential characteristics following several special differential patterns to attain an appreciable probability effect. Moreover, we prove that the exact probability of the congruent differential cluster can be obtained by computing the multiplication of r matrices of scale $2^n \times 2^n$, where n and r indicate the size of the S-box and the number of cascading rounds of the target cipher, respectively. Thus, the computational cost of the proposed approach is negligible for widely used S-boxes.
- For a special type of binary SPN ciphers, i.e., binary AES-like ciphers, we add more differential characteristics to the cluster and obtain the semicongruent differential cluster with higher probabilities. The column-mixing layer of binary AES-like ciphers divides the intermediate state into a number of independent parts. Thus, if we set the output differences of the active S-boxes in the same partition to be equal, we can obtain more differential characteristics. We establish a realistic and feasible algorithm to calculate the probabilities of semicongruent differential clusters. Experimentally, semicongruent differential clusters significantly improve the probabilities of congruent differential clusters.
- To demonstrate the application of the proposed framework, we test the probabilities of semicongruent differential clusters for several typical block cipher instances, namely, SKINNY-64, Midori-64, and CRAFT-64, and the results are summarized in Table I. In most cases, we have achieved the best results so far. However, there are still some results that cannot surpass other methods. And we attribute this to the fact that our clusters have more active S-boxes than theirs at the corresponding rounds. According to our approach, we also find a differential cluster of 15-round SKINNY-128. The probability of the cluster is about $2^{-122.9}$ which has a gain of $2^{9.1}$ compared to the security boundary by counting the number of active S-boxes.

Organization: Section II introduces the basic notations and definitions. Section III and Section IV describe the basic idea and calculation of the probabilities of congruent differential clusters, respectively. Section V describes the development of semicongruent differential clusters and the corresponding probability calculation. Section VI describes the application of the proposed framework to several typical instances. Section VII presents the concluding remarks.

II. Fundamentals

The following symbols are used in this paper.

n	size of the S-box;
m	number of S-boxes in one layer of the SPN cipher;
$\oplus$	XOR operation;
$g \circ f$	composition of f and g, i.e., $g \circ f(x) = g(f(x))$;
$M(i, j)$ or $M_{i,j}$	$i \times j$-th entry of matrix M;
$M^{<r>}$	$M^{<r>}=(m^r_{i,j})$, where $M=(m_{i,j})$;
$wt(\alpha)$	hamming weight of α;
$\#\bullet$	cardinal number of the set $\bullet$;
$\mathbb{GF}(2^n)$	finite field with 2^n elements.

Differential cryptanalysis is a classical cryptanalysis technique introduced by Biham and Shamir [3], [4], which exploits the propagation of differences in the target cipher. This cryptanalysis starts from a carefully chosen differential pair (a, b) such that the probability of $E(x) \oplus E(x \oplus a) = b$ is considerably higher than 2^{-mn}, where E denotes the target cipher.

The difference distribution table (DDT) of a single S-box is a $2^n \times 2^n$ table, where the $a \times b$-th entry is the number of pairs that satisfy the difference $(a \xrightarrow{S} b)$. Notably, in several studies, the $a \times b$-th entry of the DDT is defined as $2^n \times \Pr(a \xrightarrow{S} b)$. For simplicity, in this paper, we omit the constant 2^n, and the $a \times b$-th entry of the DDT is defined as $\mathtt{DDT}_S(a, b) = \Pr(a \xrightarrow{S} b)$.

It is challenging to calculate the exact probability of the differential for large-scale mappings. Fortunately, the probability of a differential characteristic can be derived from difference propagation over several rounds of the cipher. Thus, the differential characteristic with a high probability is usually used to represent a differential. In other words, the probabilities of differential characteristics can quantify the resistance against differential cryptanalysis to some extent.

Lai et al. [18] introduced the Markov cipher $\mathcal{E}^r$, in such a cipher, the probability of an r-round characteristic of $\mathcal{E}^r$ is the product of the probabilities of each round function, i.e.,

$$\Pr(\delta^0 \xrightarrow{\mathcal{E}} \delta^1 \xrightarrow{\mathcal{E}} \cdots \xrightarrow{\mathcal{E}} \delta^r) = \prod_{i=1}^{r} \Pr(\delta^{i-1} \xrightarrow{\mathcal{E}} \delta^i).$$

TABLE I
SUMMARY OF DIFFERENTIAL DISTINGUISHERS FOR SKINNY, MIDORI AND CRAFT IN THE SINGLE-TWEAK MODEL

Cipher	Type of Differential	Round	Probability	Time†	Source
SKINNY-64 [13]	Characteristic	8	2^{-72}	–	[13]
	Cluster by automatic tools	8	$2^{-56.55}$	$23.5h$	[10]
	Semicongruent cluster	8	$2^{-50.72}$	<1min	This paper
SKINNY-64 with improved S-box [1]	Characteristic	8	2^{-76}	–	[1]
	Semicongruent cluster	8	$2^{-60.74}$	<1min	This paper
SKINNY-128 [13]	Characteristic	15	2^{-132}	–	[13]
	Semicongruent cluster	15	$2^{-122.9}$	636s	This paper
Midori-64 [14]	Characteristic	7	2^{-70}	–	[14]
	Cluster by automatic tools	7	$2^{-57.43}$	✘	[10]
	Semicongruent cluster	7	$2^{-52.25}$	<1min	This paper
	Cluster by automatic tools	8	$2^{-60.86}$	✘	[10]
	Semicongruent cluster	8	$2^{-69.68}$	<1min	This paper
Midori-64 with improved S-box [1]	Characteristic	6	2^{-68}	–	[1]
	Cluster by automatic tools	6	2^{-61}	✘	[1]
	Semicongruent cluster	6	$2^{-58.58}$	<1min	This paper
CRAFT-64 [15]	Characteristic	9	2^{-64}	–	[15]
	Cluster with minimum active S-boxes	9	$2^{-54.67}$	✘	[15]
	Cluster by correlation matrices	9	$2^{-40.68}+2^{-48.6}$	–	[16]
	Cluster by automatic tools	9	$2^{-44.37}$	5417s	[17]
	Cluster by partitioning technique	9	$2^{-40.2}$	✘	[17]
	Semicongruent cluster	9	$2^{-39.41}$	<1min	This paper
	Characteristic	10	2^{-72}	–	[15]
	Cluster with minimum active S-boxes	10	$2^{-62.61}$	✘	[15]
	Cluster by automatic tools	10	$2^{-50.2554}$	4 days	[17]
	Cluster by partitioning technique	10	$2^{-44.89}$	✘	[17]
	Semicongruent cluster	10	$2^{-42.32}$	<1min	This paper
	Cluster by partitioning technique	11	$2^{-49.79}$	✘	[17]
	Semicongruent cluster	11	$2^{-49.90}$	<1min	This paper
	Cluster by partitioning technique	12	$2^{-54.48}$	✘	[17]
	Semicongruent cluster	12	$2^{-57.16}$	<1min	This paper
	Cluster by partitioning technique	13	$2^{-59.13}$	✘	[17]
	Cluster by partitioning technique	14	$2^{-63.8}$	✘	[17]

†: The probabilities of these differential characteristics are deduced from the number of active S-boxes. The clusters in [10] and [1] are obtained using the same methods in [10]. When constructing the differential cluster for 8-round SKINNY-64, [10] claims that "*This process took in total 23.5 h on a single core, however after 1 h the estimate for the differential probability improves by less than* $2^{-0.9}$". The clusters in [15] are searched by automatic tools. Each cluster contains about four to six differential characteristics. The platform and time for these results are not mentioned. The clusters in [17] are obtained by automatic tools and partitioning technique, respectively. The authors claim they get the cluster for 9-round CRAFT-64 in 5417s on their personal computer (Intel Core (TM)i-5, 8 Gig RAM, running Ubuntu 18.04 LTS), and get the cluster for 10-round CRAFT-64 in 4 days on a G9 Hp server with 32 Gig RAM and Windows 10 x64 as the operating system. They do not provide the platform and time for the clusters obtained by partitioning technique. We get our clusters on a laptop(Intel Core (TM)i5-10300H CPU @ 2.50GHz, 8 G RAM and Windows 11 x64 as the operating system). '-' means that these results are derived from theoretical derivations, and '✘' means that the authors do not show the platforms or time.

In general, the probability of a differential is the sum over all compatible characteristics, which can be calculated as

$$\Pr(\delta^0 \xrightarrow{\mathcal{E}^r} \delta^r) = \sum_{\delta^1,\delta^2,\cdots,\delta^{r-1}\in\{0,1\}^{mn}} \Pr(\delta^0 \xrightarrow{\mathcal{E}} \delta^1 \xrightarrow{\mathcal{E}} \cdots \xrightarrow{\mathcal{E}} \delta^r).$$

Theoretically, there exist at most $2^{mn\times(r-1)}$ possible differential characteristics compatible with such a differential.

In this study, we explore the clustering effect for binary SPN ciphers by exploring a class of differential characteristics in which the active S-boxes remain in the same positions. Formally, we construct a subset $\Omega \subseteq \{0,1\}^{mn}$ and then estimate the probability of

$$\sum_{\delta^1,\delta^2,\cdots,\delta^{r-1}\in\Omega} \Pr(\delta^0 \xrightarrow{\mathcal{E}} \delta^1 \xrightarrow{\mathcal{E}} \cdots \xrightarrow{\mathcal{E}} \delta^r).$$

If the cardinal number of Ω is adequately large, the probability advantages associated with the best differential characteristic will likely be surpassed.

First, we introduce several basic definitions.

Definition 1 [19]: Let SLayer be a nonlinear transformation on $\{0,1\}^{mn}$ defined by m paralleled n-bit S-boxes, i.e.,

$$\mathtt{SLayer}(x_0,\cdots,x_{m-1}) = (S(x_0),\cdots,S(x_{m-1})),$$

the permutation layer $\mathtt{P} : \{0,1\}^{mn} \to \{0,1\}^{mn}$ is a linear bijection, and the one-round SPN structure $\mathcal{E}$ is defined as

$$\mathcal{E}(x) = \mathtt{P} \circ \mathtt{SLayer}(x \oplus k),$$

where k denotes the subkey.

Accordingly, in an r-round SPN cipher $\mathcal{E}^r$, if $\mathtt{P}$ can be represented by an $m \times m$ binary matrix P, i.e.,

$$\mathtt{P}(x) = P \times x,$$

where P is a binary matrix over $\mathbb{GF}(2^n)$, then $\mathcal{E}$ is termed the round function of a binary SPN cipher. Typical examples of this configuration are SKINNY [13], Midori [14] and CRAFT [15].

Definition 2 [19]: Let $(x_0, \cdots, x_{m-1}) \in \{0,1\}^{mn}$ and $\theta : \{0,1\}^n \to \{0,1\}$ corresponds to the following mapping:

$$\theta(x) = \begin{cases} 0, & \text{if } x = 0; \\ 1, & \text{if } x \neq 0. \end{cases}$$

Then,

$$\chi(x_0, \cdots, x_{m-1}) = (\theta(x_0), \cdots, \theta(x_{m-1}))$$

is said to be the pattern of $(x_0, \cdots, x_{m-1})$.

If we assume that $\delta^0 \xrightarrow{\mathcal{E}} \delta^1 \xrightarrow{\mathcal{E}} \cdots \xrightarrow{\mathcal{E}} \delta^r$ is one differential characteristic of an r-round SPN cipher, then the pattern of the characteristic is calculated as $(\chi(\delta^0), \chi(\delta^1), \cdots, \chi(\delta^{r-1}))$. Instead of searching a single characteristic with the highest probability, we search for a large number of differential characteristics having the same low-weight pattern. The final probability of our differential cluster is the sum of the probabilities of potential characteristics. This tradeoff is expected to facilitate the realization of our objective.

III. Architecture of the Congruent Differential Cluster

Instead of minimizing the number of active S-boxes, our core idea is to increase the number of differential characteristics with the same number of active S-boxes. We focus on the differential behavior of the binary SPN cipher, and the process is initiated at the diffusion layer. First, we present several basic definitions.

Definition 3: Let P be an $m \times m$ binary matrix over $\mathbb{GF}(2^n)$. Then, the $m \times m$ matrix $\mathcal{B}(P)$ over $\mathbb{GF}(2)$ is defined by

$$\mathcal{B}(P) = \begin{cases} 1, & \text{if } P_{i,j} = 1 \quad (\in \mathbb{GF}(2^n)); \\ 0, & \text{else.} \end{cases}$$

which is termed the basic matrix of P.

The partition mapping

$$\mathcal{P}_i : \{0,1\}^{mn} \mapsto \{0,1\}^m$$

for a vector

$$x = [(x_{1,1}, x_{1,2}, \cdots, x_{1,n}), \cdots, (x_{m,1}, x_{m,2}, \cdots, x_{m,n})] \in \{0,1\}^{mn}$$

is defined as

$$\mathcal{P}_i(x) := \widehat{x_i} = (x_{1,i}, x_{2,i}, \cdots, x_{m,i}).$$

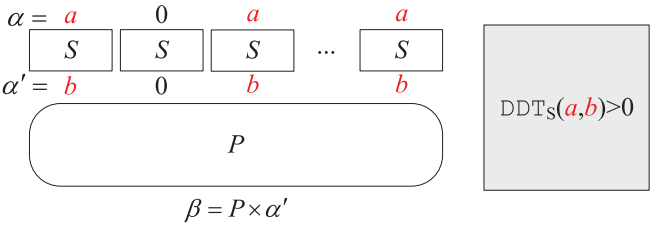

Fig. 1. Differential bypass for one round of a binary SPN.

Inversely, we define the combined mapping $\mathcal{C} : \{0,1\}^{m \times n} \mapsto \{0,1\}^{mn}$ as

$$\mathcal{C}(\widehat{x_1}, \widehat{x_2}, \cdots, \widehat{x_n}) = x.$$

Note that the diffusion layer $\mathtt{P}$ of the binary SPN diffuses the input by n-bitwise XOR. The process of the mn-bit diffusion can be divided into m copies of independent n-bit diffusions $\mathcal{B}(P)$. Specifically,

$$P \times x = \mathcal{C}[\mathcal{B}(P) \times \widehat{x_1}, \mathcal{B}(P) \times \widehat{x_2}, \cdots, \mathcal{B}(P) \times \widehat{x_n}].$$

Since $\mathcal{B}(P) = P$ in form, we use the symbol P to represent both $\mathcal{B}(P)$ and P.

Definition 4: Let $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_{m-1})$ be a vector over $\mathbb{GF}(2^n)$ and $\chi_0 \in \{0,1\}^m$. If all the nonzero entries of α remain the same and $\chi(\alpha) = \chi_0$, then α is termed an *equal extension vector* of χ_0, denoted by $\chi_0 \propto \alpha$.

For any equal extension vector α of a given pattern χ_0 (formally denoted as $\alpha = a \times \chi_0$), we have $P \times \alpha = a \times (P \times \chi_0)$. Consequently, if the equal extension vector α is the input difference of the linear layer $\mathtt{P}$, $P \times \alpha$ is also an equal extension vector of $P \times \chi_0$, and each of the nonzero components remains unchanged. If this property can be inherited by the SLayer, we can combine these two properties to identify a number of differential characteristics (Fig. 1).

Lemma 1: Let P be the matrix representation of the $\mathtt{P}$-layer in a binary SPN cipher's round function $\mathcal{E}$ and S be the S-box. Then, for any $\chi_0 \in \{0,1\}^m$, we can find at least $\#\{\mathtt{DDT}_S(a,b) > 0 : 1 \leq a, b \leq 2^n - 1\}$ of difference pairs $\alpha \xrightarrow{\mathcal{E}} \beta$ such that $\chi(\alpha) = \chi_0$ and $\chi(\beta) = P \times \chi_0$.

Proof: We assume that $\chi_0 = (\lambda_0, \lambda_1, \cdots, \lambda_{m-1})$. For any $(a,b) \in \{0,1\}^{mn} \times \{0,1\}^{mn}$ with any a, b such that $\mathtt{DDT}_S(a,b) > 0$, we construct two differences α and α' by setting

$$\alpha_i = \begin{cases} a, & \text{if } \lambda_i = 1; \\ 0, & \text{else.} \end{cases} \quad \text{and} \quad \alpha'_i = \begin{cases} b, & \text{if } \lambda_i = 1; \\ 0, & \text{else.} \end{cases}$$

In this case, the difference $\alpha \xrightarrow{\mathtt{SLayer}} \alpha'$ is one of the possible differentials of the SLayer. Thus, we conclude that we find at least $\#\{\mathtt{DDT}_S(a,b) > 0 : 1 \leq a, b \leq 2^n - 1\}$ differential pairs.

Let $P \times \alpha' = \beta$, it follows $\beta_i = \bigoplus_{j=0}^{m-1} P_{i,j} \times \alpha'_j$. Since P is a binary matrix, any nonzero $\beta_\bullet$ equals to b, and $\theta(\beta_i) = \bigoplus_{j=0}^{m-1} P_{i,j} \times \lambda_j$.

Q.E.D.

Given an input pattern χ_0, we can obtain a sequence of r-round patterns $(\chi_1, \chi_2, \chi_3, \cdots, \chi_r)$ by computing

$\chi_i = P \times \chi_{i-1}$, or equivalently, we can obtain the pattern by computing a linear systematic code

$$(\chi_0, \chi_1, \cdots, \chi_r) = (E|P|P^2|\cdots|P^r) \times \chi_0.$$

The core idea is to collect the differential characteristics following such a pattern, i.e., we search for some $\chi_0 \propto \delta^0$ such that for the active S-boxes, empirically, the value of $wt[(E|P|P^2|\cdots|P^{r-1}) \times \chi_0]$ is minimized. We compute the sum of probabilities of all characteristics $\delta^0 \to \delta^1 \to \cdots \to \delta^r$, such that $\chi_1 \propto \delta^1, \cdots, \chi_{r-1} \propto \delta^{r-1}$. More formally, we introduce the definition belows.

Definition 5: Let $\chi_i (0 \le i \le r-1)$ be the input pattern of the $(i+1)$-th round of a binary SPN cipher, where $\chi_i = P \times \chi_{i-1}$. The differential cluster $\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r$ that contains all the differential characteristics $(\delta^0, \delta^1, \cdots, \delta^{r-1}, \delta^r)$ satisfying $\chi_1 \propto \delta^1, \cdots, \chi_{r-1} \propto \delta^{r-1}$ is called a *congruent differential cluster*, i.e.,

$$\Pr(\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r) := \sum_{\chi_1 \propto \delta^1, \cdots, \chi_{r-1} \propto \delta^{r-1}} \Pr(\delta^0 \overset{\mathcal{E}}{\to} \delta^1 \overset{\mathcal{E}}{\to} \cdots \overset{\mathcal{E}}{\to} \delta^r).$$

Accordingly, if the input differences of the active S-boxes remain the same, we may assume that all the output differences of the active S-boxes are equal, and $t = \#\{\mathtt{DDT}_S(a,b) > 0 : 1 \le a, b \le 2^n - 1\}$. In this case, we may collect at most t^r differential characteristics compatible with such given r-round pattern. In any characteristic of such a differential cluster, the differences of all the S-boxes remain the same.

IV. Probability of a Congruent Differential Cluster

Next, we examine the probability of our congruent differential cluster. Historically, the probability of a differential characteristic has been used in evaluating the resistance of a cipher against differential cryptanalysis. In [6], Nyberg and Knudsen studied the provable security against differential cryptanalysis for DES-like ciphers, e.g. Serpent [20]. Later, the wide-trail strategy was proposed [8]. According to the wide-trail strategy, the branch number of the linear layer and maximum differential probability of the S-box layer can be used to bound the probability.

The same logic applies to our differential cluster. The following text describes the factors that influence the probability of our differential cluster $\Pr(\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r)$.

Theorem 1: Let P be the matrix representation of the P-layer in an r-round binary SPN cipher $\mathcal{E}^r$, and DDT be the difference distribution table of the S-box. Then, the value of $\Pr(\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r)$ equals the $\rho_0 \times \rho_r$-th entry of the matrix $\prod_{i=1}^{r} \mathtt{DDT}^{<wt(P^{i-1} \times \chi_0)>}$, where $\chi_0 = \chi(\delta_0)$, ρ_0 and ρ_r are the nonzero entries of $\delta^0 = (\delta^0_0, \delta^0_1, \cdots, \delta^0_{m-1})$ and $\delta^r = (\delta^r_0, \delta^r_1, \cdots, \delta^r_{m-1})$, respectively.

Proof: By definition, we assume that $\chi(\delta^0) = \chi_0$, and for $1 \le i \le r$, we set $\chi_i = P \times \chi_{i-1}$. By definition, the probability of the differential cluster is

$$\begin{aligned}
\Pr(\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r) &:= \sum_{\chi_1 \propto \delta^1, \cdots, \chi_{r-1} \propto \delta^{r-1}} \Pr(\delta^0 \overset{\mathcal{E}}{\to} \delta^1 \overset{\mathcal{E}}{\to} \cdots \overset{\mathcal{E}}{\to} \delta^r) \\
&= \sum_{\chi_1 \propto \delta^1, \cdots, \chi_{r-1} \propto \delta^{r-1}} \prod_{i=1}^{r} \Pr(\delta^{i-1} \overset{\mathcal{E}}{\to} \delta^i) \\
&= \sum_{\rho_1 \ne 0, \cdots, \rho_{r-1} \ne 0} \prod_{i=1}^{r} [\Pr(\rho_{i-1} \overset{S}{\to} \rho_i)]^{wt(\chi_{i-1})}.
\end{aligned}$$

Next, we prove that

$$\begin{aligned}
&\sum_{\rho_1 \ne 0, \cdots, \rho_{r-1} \ne 0} \prod_{i=1}^{r} [\Pr(\rho_{i-1} \overset{S}{\to} \rho_i)]^{wt(\chi_{i-1})} \\
&= (\prod_{i=1}^{r} \mathtt{DDT}^{<wt(P^{i-1} \times \chi_0)>})_{\rho_0, \rho_r}
\end{aligned}$$

by performing mathematical induction on round r.

For $r = 2$, we may verify that

$$\begin{aligned}
&\sum_{\rho_1 \ne 0} \Pr(\rho_0 \overset{S}{\to} \rho_1)^{wt(\chi_0)} \times \Pr(\rho_1 \overset{S}{\to} \rho_2)^{wt(\chi_1)} \\
&= (\mathtt{DDT}^{<wt(\chi_0)>} \times \mathtt{DDT}^{<wt(\chi_1)>})_{\rho_0, \rho_2}.
\end{aligned}$$

Assuming that the equation holds for less than the $(r-1)$-round, we check the case of the r-round cascade.

$$\begin{aligned}
&\sum_{\rho_1 \ne 0, \cdots, \rho_{r-1} \ne 0} \prod_{i=1}^{r} [\Pr(\rho_{i-1} \overset{S}{\to} \rho_i)]^{wt(\chi_{i-1})} \\
&= \sum_{\rho_{r-1} \ne 0} \sum_{\rho_1 \ne 0, \cdots, \rho_{r-2} \ne 0} \\
&\left(\prod_{i=1}^{r-1} [\Pr(\rho_{i-1} \overset{S}{\to} \rho_i)]^{wt(\chi_{i-1})} \times [\Pr(\rho_{r-1} \overset{S}{\to} \rho_r)]^{wt(\chi_{r-1})} \right) \\
&= \sum_{\rho_{r-1} \ne 0} [\Pr(\rho_{r-1} \overset{S}{\to} \rho_r)]^{wt(\chi_{r-1})} \\
&\quad \times \sum_{\rho_1 \ne 0, \cdots, \rho_{r-2} \ne 0} \prod_{i=1}^{r-1} [\Pr(\rho_{i-1} \overset{S}{\to} \rho_i)]^{wt(\chi_{i-1})} \\
&= \sum_{\rho_{r-1} \ne 0} (\mathtt{DDT}^{<wt(P^{r-1})>})_{\rho_{r-1}, \rho_r} \\
&\quad \times (\prod_{i=1}^{r-1} \mathtt{DDT}^{<wt(P^{i-1} \times \chi_0)>})_{\rho_0, \rho_{r-1}} \\
&= (\prod_{i=1}^{r} \mathtt{DDT}^{<wt(P^{i-1} \times \chi_0)>})_{\rho_0, \rho_r}
\end{aligned}$$

Q.E.D.

According to this theorem, we can promptly calculate the exact probability of the differential cluster for a given binary SPN cipher. The main cost of the computation is the multiplication of r of $2^n \times 2^n$ matrices, where n is the size of the S-box (typically 4 or 8). Thus, the computational complexity of a practical binary SPN cipher is negligible. In addition, the largest entry of the matrix $\prod_{i=1}^{r} \mathtt{DDT}^{<wt(P^{i-1} \times \chi_0)>}$ indicates the probability of the best congruent differential cluster.

An interesting question follows: Given an r-round binary SPN cipher $\mathcal{E}^r$, can we find a new r-round binary SPN

cipher $\mathcal{E}'^r$ that keeps the best probability of r-round congruent differential clusters unchanged?

Let P_1 and P_2 be two distinct $n \times n$ binary diffusion layers. If for any input pattern χ_1 of P_1, there exists input pattern χ_2 of P_2 such that

$$wt([E|P_1|\cdots|P_1^{r-1}] \times \chi_1) = wt([E|P_2|\cdots|P_2^{r-1}] \times \chi_2),$$

and the multiple set $\{wt(\chi_1), wt(P_1 \times \chi_1), \cdots, wt(P_1^{r-1} \times \chi_1)\}$ is equal to the multiple set $\{wt(\chi_2), wt(P_2 \times \chi_2), \cdots, wt(P_2^{r-1} \times \chi_2)\}$. Then, if we adopt P_1 and P_2 in two binary SPN ciphers and keep the S-box unchanged, the probabilities of the best r-round congruent differential clusters of these two ciphers are identical.

Empirically, it seems that if we replace the current S-box with a new one, the probability of the best r-round congruent differential clusters of the binary SPN cipher is uncontrollable. However, this probability remains unchanged if we replace the S-box with certain affine equivalent S-boxes.

Definition 6 [21]: Let S' and S be two n-bit S-boxes. If there exist two affine mappings A_0 and A_1, such that

$$S'(x) = A_1 \circ S \circ A_0(x),$$

then S and S' are termed affine-equivalent (AE).

Within the AE assumption, most of the basic properties of S-boxes, such as the maximum differential probability, linear correlation properties, and algebraic degree, remain invariant. We may verify the properties between DDT_S and $\mathrm{DDT}_{S'}$ as follows.

Lemma 2: Let M_0, M_1 be two invertible $n \times n$ matrices. If S and S' are two AE S-boxes and $S'(x) = [M_1 \circ S \circ M_0(x \oplus c_0)] \oplus c_1$, then

$$\mathrm{DDT}_{S'}(i, j) = \mathrm{DDT}_S(M_0 \times i, M_1^{-1} \times j),$$

where c_0 and c_1 are two constants.

Theorem 2: Let $\mathcal{E}^r$ and $\mathcal{E}'^r$ be two r-round binary SPN ciphers that employ the unified P-layer. S and S' are the S-boxes of $\mathcal{E}$ and $\mathcal{E}'$, respectively. If $S'(x) := [M \circ S \circ M^{-1}(x \oplus c_0)] \oplus c_1$, then the probabilities of the best congruent differential clusters in these two cipher are equal, where constants $c_0, c_1 \in \{0,1\}^n$ and M denote an invertible $n \times n$ matrix.

Proof: This argument follows the notation used in Theorem 1. We assume that the best congruent differential cluster of $\mathcal{E}$ is $\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r$. According to Theorem 1, $\Pr(\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r) = (\prod_{i=1}^r \mathrm{DDT}^{<wt(P^{i-1} \times \chi_0)>})_{\rho_0, \rho_r}$. Then,

$$(\prod_{i=1}^{r} \mathrm{DDT}_S^{<wt(P^{i-1} \times \chi_0)>})_{\rho_0, \rho_r} = \sum_{\rho_1, \rho_2, \cdots, \rho_{r-1}} \prod_{i=1}^{r} [(\mathrm{DDT}_S^{<wt(P^{i-1} \times \chi_0)>})_{\rho_{i-1}, \rho_i}],$$

Applying Lemma 2, we obtain

$$[(\mathrm{DDT}_S^{<wt(P^{i-1} \times \chi_0)>})_{\rho_{i-1}, \rho_i}] = [(\mathrm{DDT}_{S'}^{<wt(P^{i-1} \times \chi_0)>})_{M \times \rho_{i-1}, M \times \rho_i}],$$

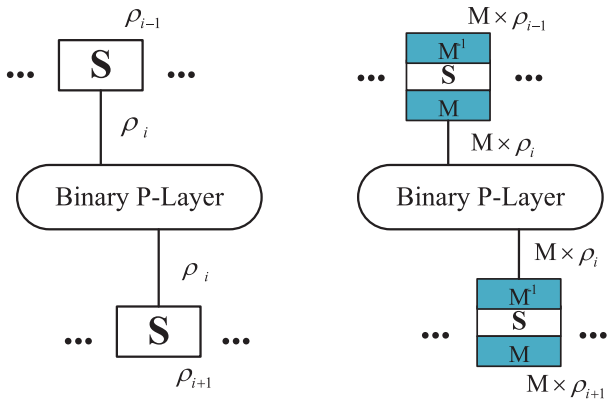

Fig. 2. Differential behavior of S and S' in a congruent differential cluster.

it follows

$$\begin{aligned}
&\Pr(\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r) \\
&= \sum_{\rho_1, \rho_2, \cdots, \rho_{r-1}} \prod_{i=1}^{r} [(\mathrm{DDT}_S^{<wt(P^{i-1} \times \chi_0)>})_{\rho_{i-1}, \rho_i}] \\
&= \sum_{\rho_1, \rho_2, \cdots, \rho_{r-1}} \prod_{i=1}^{r} [(\mathrm{DDT}_{S'}^{<wt(P^{i-1} \times \chi_0)>})_{M \times \rho_{i-1}, M \times \rho_i}] \\
&\overset{\mu_\bullet := M \times \rho_\bullet}{=} \sum_{\mu_1, \mu_2, \cdots, \mu_{r-1}} \prod_{i=1}^{r} [(\mathrm{DDT}_{S'}^{<wt(P^{i-1} \times \chi_0)>})_{\mu_{i-1}, \mu_i}] \\
&= (\prod_{i=1}^{r} \mathrm{DDT}_{S'}^{<wt(P^{i-1} \times \chi_0)>})_{\mu_0, \mu_r} = \Pr(\alpha_0 \overset{\mathcal{E}'^r}{\rightsquigarrow} \alpha_r),
\end{aligned}$$

where $\chi(\delta^0) \propto \alpha_0$ and $\chi(\delta^r) \propto \alpha_r$; and the nonzero components of α_0 and α_r are μ_0 and μ_r, respectively. Thus, we conclude that the best r-round congruent differential cluster of $\mathcal{E}$ is not greater than that of $\mathcal{E}'$ and vice versa.

Q.E.D.

In the design of the SPN cipher, one of the key concerns regarding the S-box is to maintain the efficiency of the decryption. One possible approach is to employ an involutory core function S (for example, the inverse function of the finite field $S(x) := x^{-1}$) and an affine mapping A. In this case we can construct a new involute S-box by

$$S'(x) := A^{-1} \circ S \circ A(x).$$

The result of Theorem 2 indicates that such a modification will not change the probability of the congruent differential cluster.

V. Extension of the Congruent Differential Cluster to Binary AES-Like Cipher

Given any pattern $\chi_0 \in \{0,1\}^m \setminus \{0\}$ and $\chi_0 \propto \delta^0$ for a binary SPN cipher, we can efficiently calculate the congruent differential cluster $\delta^0 \overset{\mathcal{E}^r}{\rightsquigarrow} \delta^r$. Note that the diffusion layer is considered to be a binary matrix without distinction. Therefore, we can harvest more differential characteristics for the cluster if more details of the cipher are considered.

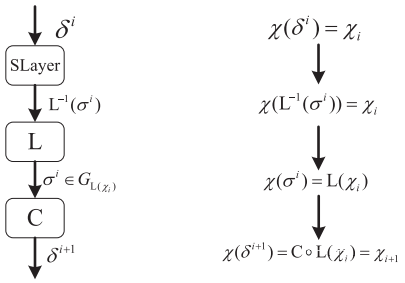

Fig. 3. Development of the difference and the pattern.

A series of recently proposed block ciphers adopt the AES-like SPN structure, and their diffusion layers satisfy certain typical constructions, i.e., a word-width shuffle followed by several independent copies of a binary P-layer. In this case, we may add more differential characteristics to the congruent differential cluster.

Paralleled binary permutation: Let $C = diag(M, M, \cdots, M)$ be a diagonal matrix, and M be a binary submatrix over $\mathbb{GF}(2^n)$. Then, C is called a paralleled binary permutation.

Shuffle matrix: Let L be an $m \times m$ binary submatrix over $\mathbb{GF}(2^n)$ and f be a bijection over $\{0, 1, \cdots, m-1\}$, such that

$$L_{i,j} = \begin{cases} 1, & \text{if } f(i) = j; \\ 0, & \text{else.} \end{cases}$$

Then, L is called a shuffle matrix.

Let $\mathcal{A}^r$ be an r-round binary SPN cipher. If its diffusion layer consists of a shuffle matrix and paralleled binary permutation, i.e.,

$$P(x) = C \circ L(x),$$

then $\mathcal{A}^r$ is an r-round binary AES-like cipher.

Next, we will demonstrate the basic idea of our improvement. For simplicity, in the rest of this section we assume that m is a square number.

Definition 7: Let $\chi' \in \{0,1\}^m$ and $\gamma := (\gamma_0, \gamma_1, \cdots, \gamma_{\sqrt{m}-1}) \in \{0,1\}^{mn}$, where $\gamma_i = (\gamma_{i,0}, \gamma_{i,1}, \cdots, \gamma_{i,\sqrt{m}-1})$ is a vector over $\mathbb{GF}(2^n)$. If $\chi(\gamma) = \chi'$ and for any $0 \le i \le \sqrt{m}-1$, all the nonzero entries of each separate γ_i are equal, then γ is called a *semicongruent vector* of χ', denoted by $\chi' \ltimes \gamma$. The set of all semicongruent vectors of χ' is denoted by $G_{\chi'}$.

Now we take a closer look at the paralleled binary permutation layer. This layer consists of several copies of one binary permutation M, at this time, if the input difference of C is restricted as a semicongruent vector, then the output difference of C always keep the *'semicongruent'* property.

Example 1: We choose $m = 16$ and $n = 4$ for a binary AES-like cipher, i.e., the size of the S-box is 4-bit, and the submatrix M in the C-layer is a 4×4 binary matrix (typical instances include Midori-64, SKINNY-64 and CRAFT-64). Then for an input difference of the paralleled binary permutation layer $\gamma =$ (**0,0xA,0xA,0**, **0x5,0,0x5,0x5**, **0,0,0,0**, **0xB,0xB,0,0**) and a 16-bit pattern $\chi' = (0,1,1,0,1,0,1,1,0,0,0,0,\ 1,1,0,0)$, it could be verified that $\chi(\gamma) = \chi'$, therefore we have $\chi' \ltimes \gamma$.

In the example above, the nonzero entries of each 4-tuple of γ are identical, at the same time, the submatrix M is a 4×4 binary matrix. As a consequence, the *semicongruent* property is inherited by the C-layer.

Proposition 1: Let $\mathcal{A}$ be the round function of a binary AES-like cipher, i.e., $\mathcal{A} = P \circ SLayer = C \circ L \circ SLayer$ and $\alpha \in \{0,1\}^{mn}$ be the input difference of $\mathcal{A}$. If $L \circ SLayer_\Delta(\alpha) \in G_{L(\chi_0)}$, then $\chi(\mathcal{A}(\alpha)) = P(\chi(\alpha))$, where $SLayer_\Delta(\alpha)$ denotes the output difference of α bypass the SLayer.

Proof: Since SLayer does not change the development of patterns, then $\chi(SLayer_\Delta(\alpha)) = \chi(\alpha)$. The shuffle matrix L before the C layer can be treated as a word-level permutation. Thus, $\chi(L \times SLayer_\Delta(\alpha)) = L \times \chi(SLayer_\Delta(\alpha))$. According to the definition of paralleled binary permutation, the mapping C can be treated as $\sqrt{m}$ binary P-layers $M_{\sqrt{m}\times\sqrt{m}}$ (over $\mathbb{GF}(2^n)$) in parallel. Therefore, the input difference of C can be divided into $\sqrt{m}$ independent parts, i.e., $L \times SLayer_\Delta(\alpha) = (x_0, x_1, \cdots, x_{\sqrt{m}-1})$, where $x_i \in \mathbb{GF}(2^n)^{\sqrt{m}}$ for $0 \le i \le \sqrt{m}-1$ and $\chi(L \times SLayer_\Delta(\alpha)) = (\chi(x_0), \chi(x_1), \cdots, \chi(x_{\sqrt{m}-1}))$. Then, the output difference of C can be calculated as $(M \times x_0, M \times x_1, \cdots, M \times x_{\sqrt{m}-1})$, i.e.,

$$C \circ L \circ SLayer_\Delta(\alpha) = (M \times x_0, M \times x_1, \cdots, M \times x_{\sqrt{m}-1}).$$

Since $L \circ SLayer_\Delta(\alpha) \in G_{L(\chi_0)}$ and M is a binary matrix, it follows from the discussion in Section III that $\chi(M \times x_i) = M \times \chi(x_i)$. From the independence of $M \times x_0, M \times x_1, \cdots, M \times x_{\sqrt{m}-1}$, it can be inferred that

$$\begin{aligned} &\chi(M \times x_0, M \times x_1, \cdots, M \times x_{\sqrt{m}-1}) \\ &\quad = (\chi(M \times x_0), \chi(M \times x_1), \cdots, \chi(M \times x_{\sqrt{m}-1})), \end{aligned}$$

and then,

$$\begin{aligned} \chi(\mathcal{A}(\alpha)) &= \chi(C \circ L \circ SLayer_\Delta(\alpha)) \\ &= \chi(M \times x_0, M \times x_1, \cdots, M \times x_{\sqrt{m}-1}) \\ &= (\chi(M \times x_0), \chi(M \times x_1), \cdots, \chi(M \times x_{\sqrt{m}-1})) \\ &= (M \times \chi(x_0), M \times \chi(x_1), \cdots, M \times \chi(x_{\sqrt{m}-1})) \\ &= C(\chi(L \times SLayer_\Delta(\alpha))) \\ &= C(L \times \chi(SLayer_\Delta(\alpha)) \\ &= P(\chi(SLayer_\Delta(\alpha))) \\ &= P(\chi(\alpha)). \end{aligned}$$

Q.E.D.

As a result, given an input difference α, if the shuffle matrix L makes the output difference of the SLayer (also the input difference of the C-layer) a semicongruent vector, i.e., $L \circ SLayer_\Delta(\alpha) \in G_{L(\chi(\alpha))}$, then we predict that:

1) The *semicongruent* property will hold for the output difference of a single round.
2) The output pattern of the round function can be obtained by applying the linear layer to the input pattern, which is the same as that of the congruent differential cluster.

Therefore, we only take consider of the differential characteristics that ensure the input difference of C to be a semicongruent vector in each round, we name such characteristics as semicongruent differential characteristics. In particular,

the congruent differential characteristics are special cases of semicongruent differential characteristics. So if we are able to add more semicongruent differential characteristics into to our congruent differential clusters (Fig. 4: the same color indicates the same difference value or zero difference), the probability of differential cluster could be certainly increased.

Definition 8: Let $\chi_i (0 \le i \le r-1)$ be the input pattern of the $(i+1)$-th round of a binary AES-like cipher, where $\chi_i = \mathtt{P}(\chi_{i-1})$. The differential cluster $\delta^0 \overset{\mathcal{A}^r}{\Rightarrow} \delta^r$ that contains all the differential characteristics $(\delta^0, \delta^1, \cdots, \delta^{r-1}, \delta^r)$ satisfying $\mathtt{L}(\chi_0) \ltimes \mathtt{LoSLayer}_\Delta(\delta^0), \cdots, \mathtt{L}(\chi_{r-1}) \ltimes \mathtt{LoSLayer}_\Delta(\delta^{r-1})$ is called a *semicongruent differential cluster*, i.e.,

$$\begin{aligned}&\Pr(\delta^0 \overset{\mathcal{A}^r}{\Rightarrow} \delta^r)\\ &:= \sum_{\mathtt{L}(\chi_0)\ltimes \mathtt{LoSLayer}_\Delta(\delta^0), \cdots, \mathtt{L}(\chi_{r-1})\ltimes \mathtt{LoSLayer}_\Delta(\delta^{r-1})} \Pr(\delta^0 \overset{\mathcal{A}}{\to} \cdots \overset{\mathcal{A}}{\to} \delta^r).\end{aligned}$$

We use $\mathcal{SC}$ to denote the set of all the differential characteristics of the semicongruent differential cluster, i.e.,

$$\begin{aligned}\mathcal{SC} := \{(\delta^0, \cdots, \delta^r) : &\mathtt{L}(\chi_0) \ltimes \mathtt{L} \circ \mathtt{SLayer}_\Delta(\delta^0),\\ &\cdots, \mathtt{L}(\chi_{r-1}) \ltimes \mathtt{L} \circ \mathtt{SLayer}_\Delta(\delta^{r-1})\}.\end{aligned}$$

Then we have

$$\begin{aligned}&\Pr(\delta^0 \overset{\mathcal{A}^r}{\Rightarrow} \delta^r)\\ &= \sum_{(\delta^0,\delta^1,\cdots,\delta^{r-1},\delta^r)\in\mathcal{SC}} \Pr(\delta^0 \overset{\mathcal{A}}{\to} \delta^1 \overset{\mathcal{A}}{\to} \cdots \overset{\mathcal{A}}{\to} \delta^r)\\ &= \sum_{(\delta^0,\delta^1,\cdots,\delta^{r-1},\delta^r)\in\mathcal{SC}} \Pr(\delta^0 \to \delta^1) \times \cdots \times \Pr(\delta^{r-2} \to \delta^{r-1})\\ &\times \Pr(\delta^{r-1} \to \delta^r)\\ &= \sum_{\substack{\delta^{r-1}:\\ (\delta^0,\delta^1,\cdots,\delta^{r-1},\delta^r)\in\mathcal{SC}}} \Pr(\delta^{r-1} \to \delta^r)\\ &\times \sum_{\substack{(\delta^0,\delta^1,\cdots,\delta^{r-1}):\\ (\delta^0,\delta^1,\cdots,\delta^{r-1},\delta^r)\in\mathcal{SC}}} \Pr(\delta^0 \to \delta^1) \times \cdots \times \Pr(\delta^{r-2} \to \delta^{r-1})\\ &= \sum_{\substack{\delta^{r-1}:\\ (\delta^0,\delta^1,\cdots,\delta^{r-1},\delta^r)\in\mathcal{SC}}} \Pr(\delta^{r-1} \to \delta^r) \times \Pr(\delta^0 \overset{\mathcal{A}^{r-1}}{\Rightarrow} \delta^{r-1}),\end{aligned}$$

so we can calculate the probability of the semicongruent differential cluster round by round. We introduce Algorithm 1 to calculate to probability of the semicongruent differential cluster for a given input and output differences.

Let $\delta^i (0 \le i \le r-1)$ be the input differences of the $(i+1)$-th round and $\sigma^i (0 \le i \le r-1)$ be the input differences of paralleled binary permutation in the $(i+1)$-th round. Let $\chi_i (0 \le i \le r-1)$ be the input pattern of the $(i+1)$-th round. For a semicongruent differential cluster, we have $\chi_i = \mathtt{P}(\chi_{i-1})$ and $\sigma^i \in G_{\mathtt{L}(\chi_i)}$.

To calculate the probabilities of the semicongruent differential clusters $\delta^0 \overset{\mathcal{A}^{i+1}}{\Rightarrow} \delta^{i+1}$, we traverse δ^i and sum up the probabilities $\Pr(\delta^0 \overset{\mathcal{A}^i}{\Rightarrow} \delta^i) \times \Pr(\delta^i \xrightarrow{\mathtt{SLayer}} \mathtt{L}^{-1}(\sigma^i))$, where

Algorithm 1 Calculating the Probability of r-Round Semicongruent Differential Clusters

Input: input difference $\delta^0 = (\delta_0^0, .., \delta_{m-1}^0)$, number of encryption rounds r, round function $\mathcal{E} = \mathtt{C} \circ \mathtt{L} \circ \mathtt{SLayer}$;
Output: $(\delta^r, \Pr_{\delta^r})$;

```
Let χ_i be the input pattern and δ^i = (δ^i_0, ..., δ^i_{m-1}) be the input differences of round i+1. Let σ^i = (σ^i_0, ..., σ^i_{m-1}) be the input differences of the paralleled binary permutation in the (i+1)-th round, then δ^{i+1} = C(σ^i). Denote Pr_{δ^i} the probabilities of the semicongruent differential clusters δ^0 ⇒^{A^i} δ^i.
for all σ^0 ∈ G_{L(χ_0)} do
    Pr_{δ^1} = Pr(δ^0 --SLayer--> L^{-1}(σ^0))
    record (δ^1, Pr_{δ^1}) in an array
end
for 2 ≤ i ≤ r do
    for all σ^{i-1} ∈ G_{L(χ_{i-1})} do
        Algorithm 2
        record (δ^i, Pr_{δ^i}) in an array
    end
end
return (δ^r, Pr_{δ^r})
```

$\sigma^i = \mathtt{C}^{-1}(\delta^{i+1}) \in G_{\mathtt{L}(\chi_i)}$. We calculate round by round like this until obtain the probability of δ^r. From Definition 8, we can deduce that there are theoretically $2^{l_i \times n}$ output differences for the i-th round, where l_i is numbers of the partitions with active S-boxes and n is the size of S-box. The calculation of $\Pr(\delta^i \xrightarrow{\mathtt{SLayer}} \mathtt{L}^{-1}(\sigma^i))$ can be attributed to looking up DDT for t_i times, where t_i is the numbers of active S-boxes in the i-th round. Thus, the computational complexity of the i-th round is $t_i \times 2^{l_{i-1} \times n} \times 2^{l_i \times n}$ times DDT look-up, theoretically.

In programming implementation, we do not simply traverse each (δ^i, σ^i) and then calculate $\Pr(\delta^i \xrightarrow{\mathtt{SLayer}} \mathtt{L}^{-1}(\sigma^i))$ by looking up t_i times DDT. And we introduce Algorithm 2 to improve the efficiency. Firstly, only the δ^i with nonzero probabilities will be recorded. It means that the input differences of active S-boxes in the $(i+1)$-th round may not take all the 2^n values. Then we iterate through the input differences of the active S-boxes one by one. For example, assuming there are two active S-boxes *S1* and *S2* in the current partition, and their input differences are a and b, respectively. To meet the conditions of semicongruent vectors, the input differences a and b must have the same output differences after S-box. If the current input differences a and b cannot lead to a same output difference c, then we do not consider other active S-boxes and can abandon a series of δ^i. An array is needed to record the output differences and their probabilities of the current round. For the i-th round, the array has at most $2^{l_i \times n}$ entries, where l_i is numbers of the partitions with active S-boxes and n is the size of S-box. Thus, the storage complexity is $O(2^{\max(l_i) \times n})$, where $0 \le i \le r-1$. If there exists i that meets $l_i = \sqrt{m}$, then the storage complexity reaches its maximum value $O(2^{\sqrt{m} \times n})$. In addition, we can set a probability threshold and retain δ^i

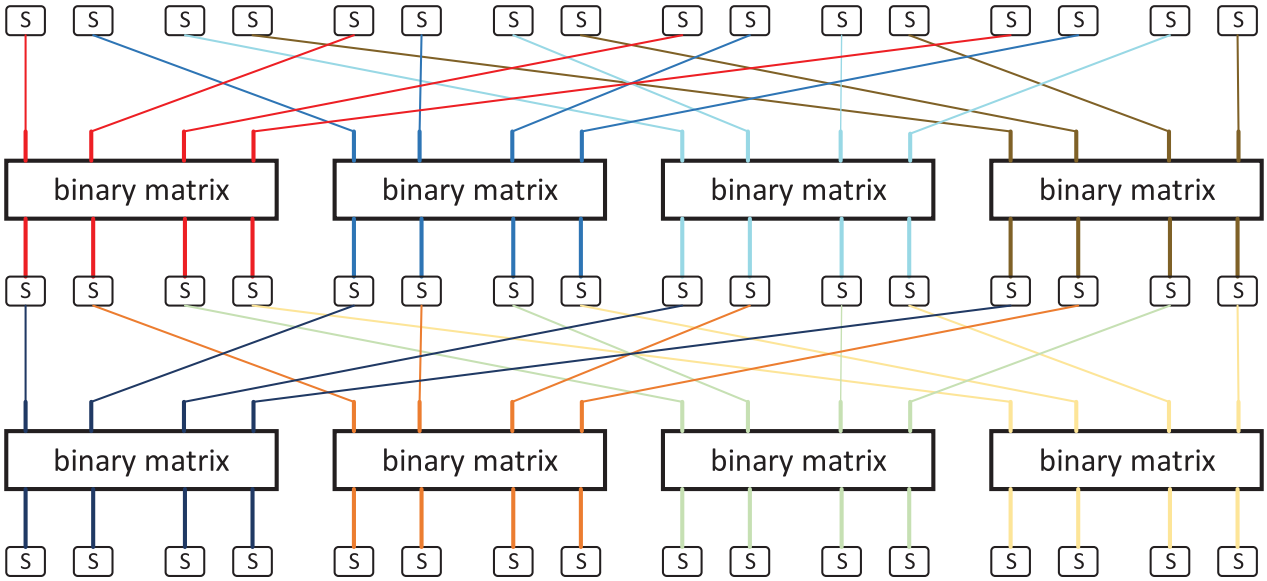

Fig. 4. Differentials compatible with the congruent differential cluster of AES-like construction.

TABLE II
INPUT PATTERNS AND CORRESPONDING NUMBERS OF ACTIVE S-BOXES

Cipher	Input Pattern	Round	Number of Active S-boxes
Midori-64	(0001 0010 0001 0010)	7	36(= 4+4+4+12+4+4+4)
SKINNY-64	(0001 1111 0111 0110)	8	36(= 10+4+2+2+2+2+4+10)
CRAFT-64	(0001 0101 0000 0101)	10	36(= 3+3+4+5+3+3+5+4+3+3)

when $\Pr_{\delta^i}$ exceeds this threshold. This can further improve the efficiency.

VI. EXPERIMENTS AND APPLICATIONS

We apply the theory discussed above to specific block ciphers, i.e, Midori [14], SKINNY [13], and CRAFT [15] (see Appendix).

A. Decide the Input Patterns and Input Differences

First, we traverse all input patterns χ_0 of Midori, SKINNY, and CRAFT and calculate $\chi_1, \ldots, \chi_{r-1}$ by $\mathrm{P}(\chi_0), \ldots, \mathrm{P}(\chi_{r-2})$. As mentioned in Section III, $wt(\chi_i)$ is the number of active S-boxes in round $i+1$. We focus on the input patterns that minimize the total number of active S-boxes after encrypting a certain number of rounds. Table II shows the input patterns we choose and the corresponding numbers of active S-boxes.

Next, it comes to choose the input differences for the active S-boxes of the first round. We empirically believe that clusters containing single characteristic with high probability may have better performance. Therefore, we choose the input differences with the maximum differential transition probabilities for active S-boxes. For example, when 0×2 and 0xa be the input differences of Midori-64's S-box, the probabilities of the possible output differences are the maximum value 0.25. Thus, we choose 0x2 and 0xa to be the input differences for the active S-boxes of the first round.

B. Apply Semicongruent Differential Clusters to Midori, SKINNY, and CRAFT

Based on the above strategies, we obtain the differential cluaters of Midori-64, SKINNY-64, and CRAFT-64. Some clusters are as follows:

$$0x000a00a0000a00a0 \xrightarrow{7-round\ \texttt{Midori-64}} 0xaa0aaa0aaaa0aaa0.$$

$$0x0001111101110110 \xrightarrow{8-round\ \texttt{SKINNY-64}} 0x2022222222020220.$$

$$0x000a0a0a00000a0a \xrightarrow{10-round\ \texttt{CRAFT-64}} 0x0a00000000000a0a.$$

To the best of our knowledge, the probabilities of these three clusters are higher than those of the clusters with the same number of rounds.

We also find a differential cluster for 15-round SKINNY-128 with a probability $2^{-122.9}$. The detail is as follows:

$$0x(21, 21, 00, 00, 00, 00, 00, 00, 00, 21, 21, 00, 21, 00, 21, 00) \rightarrow 0x(00, 00, 00, 00, 04, 00, 00, 04, 04, 00, 04, 00, 00, 00, 04, 04).$$

In the design report, the total number of active S-boxes for 15-round SKINNY-128 is at least 66. The maximum differential transition probability for SKINNY-128's S-box is 2^{-2}, so there is no differential characteristic with a probability greater than 2^{-132}. In our cluster, the total number of active S-boxes is 74 which means the maximum probability of the differential characteristic is 2^{-148}. However, when considering the cluster effect, we have a gain about $2^{9.1}$ compared with the current theoretically optimal differential characteristic.

Algorithm 2 Calculating $\Pr_{\delta^i}$ $(i > 1)$

Input: $\sigma^{i-1} \in G_{\mathrm{L}(\chi_{i-1})}$, all δ^{i-1} where $\Pr_{\delta^{i-1}} \neq 0$;
Output: $(\delta^i, \Pr_{\delta^i})$;

Separate $\mathrm{L}^{-1}(\sigma^{i-1})$ into $\sqrt{m}$ partitions $(\gamma_0^i, \gamma_1^i, \cdots, \gamma_{\sqrt{m}-1}^i) \in \{0,1\}^{\sqrt{m}\times\sqrt{m}\times n}$ where the nonzero entries of each partition are equal. Denote the value of the nonzero entries of $\gamma_j^i \in \{0,1\}^{\sqrt{m}\times n}$ by $\eta_j^i \in \{0,1\}^n$ and the input difference of SLayer correspond to γ_j^i by $\xi_j^{i-1} = (\xi_{j,0}^{i-1}, \xi_{j,1}^{i-1}, \cdots, \xi_{j,\sqrt{m}-1}^{i-1}) \in \{0,1\}^{\sqrt{m}\times n}$.
$\Pr_{\delta^i} = 0$
for *all* ξ_0^{i-1} **do**
  if *all the nonzero entries of* ξ_0^{i-1} *have nonzero probabilities to* η_0^i **then**
    for *all* ξ_1^{i-1} **do**
      ...
      for *all* $\xi_{\sqrt{m}-1}^{i-1}$ **do**
        if *all the nonzero entries of* $\xi_{\sqrt{m}-1}^{i-1}$ *have nonzero probabilities to* $\eta_{\sqrt{m}-1}^i$ **then**
          $\Pr_{\delta^i} = \Pr_{\delta^i} + \Pr_{\delta^{i-1}} \times \Pr(\xi_0^{i-1} \to \gamma_0^i) \times \cdots \times \Pr(\xi_{\sqrt{m}-1}^{i-1} \to \gamma_{\sqrt{m}-1}^i)$
        end
      end
    end
  end
end
/* $Pr(\xi_j^{i-1} \to \gamma_j^i)$ can be obtained by looking up DDT. For example, if ξ_j^{i-1} has two nonzero entries $\xi_{j,0}^{i-1}$ and $\xi_{j,1}^{i-1}$, then $Pr(\xi_j^{i-1} \to \gamma_j^i) = \mathrm{DDT}(\xi_{j,0}^{i-1}, \eta_j^i) \times \mathrm{DDT}(\xi_{j,1}^{i-1}, \eta_j^i)$. */
return $(\delta^i, \Pr_{\delta^i})$

In addition, we apply this method to variants of Midori-64 and SKINNY-64 in [1]. Todo and Sasaki noted the presence of chains of differences $\Delta_1 \to \Delta_2 \to \Delta_3 \to \cdots$ over the S-boxes of Midori-64 and SKINNY-64, in which each transition occurs with a high probability of 2^{-2}. To enhance the resistance of Midori-64 and SKINNY-64 against differential cryptanalysis, Todo and Sasaki designed new S-boxes ensuring that the high-probability chain length is at most 2. Employing the improved S-box, the maximum differential characteristic probability of 6-round Midori-64 decreases from 2^{-60} to 2^{-68}, and the maximum differential characteristic probability of 8-round SKINNY-64 decreases from 2^{-72} to 2^{-76}. Similar to the method reported in [10], Ankele and Kölbl constructed a 6-round differential cluster of Midori-64 using the new S-box with a probability of 2^{-61}. Using Algorithm 1, we construct semicongruent differential clusters of 6-round Midori-64 and 8-round SKINNY-64 using the new S-box, with probabilities of $2^{-58.58}$ and $2^{-60.74}$, respectively.

Finally, we investigate the probability gap between differential clusters and characteristics. The results are summarized in Table III. A large distance can be observed between the differential characteristics and clusters. If only the number of active S-boxes is used as the criterion for evaluating the resistance of the cipher against differential attacks, we may receive a marginal security bound. Therefore, it is necessary to consider the clustering efficiency for the target block ciphers.

C. Gap Between Our Clusters and Real Differentials

In order to enhance the persuasiveness of our work, we verified the differential clusters obtained using our method. Due to limitations in computing resources, it is difficult to verify differential clusters with a large number of rounds. So we used the same method to construct some differential clusters with probabilities around 2^{-25} for verification.

We constructed a 6-round cluster with a probability $2^{-24.41}$ for CRAFT-64, a 4-round cluster with a probability $2^{-23.58}$ for Midori-64 and a 5-round cluster with a probability $2^{-22.9}$ for SKINNY-64. The details are as follows:

$$0x000a0a0a00000a0a \xrightarrow{6-round\ \texttt{CRAFT-64}} 0x000000000aa00a00,$$

$$0x2000200000200020 \xrightarrow{4-round\ \texttt{Midori-64}} 0x0222022222022202,$$

$$0x0000100000010110 \xrightarrow{5-round\ \texttt{SKINNY-64}} 0x5555050000550555.$$

We randomly generated 200 keys, and for each key 2^{31} random plaintext pairs with the given input difference were encrypted. A counter was needed to record the number of ciphertext pairs with the corresponding output difference after encryption for each key. Then we calculated the experimental probability for each cluster according to the average value of the counter. The experimental probabilties are $2^{-23.71}$ for 6-round CRAFT-64's cluster, $2^{-26.13}$ for 4-round Midori-64's cluster and $2^{-21.51}$ for 5-round SKINNY-64's cluster. It can be found that the experimental results are in general agreement with the theory.

VII. Discussion and Conclusion

Although differential cryptanalysis was proposed more than 30 years ago, it still plays an important role in modern cryptanalysis. In recent decades, counting the number of active S-boxes has become the mainstream strategy for evaluating the resistance against such attacks. However, for certain constructions, the use of differential characteristics instead of differentials involves several challenges. Thus, it is of significance to investigate the differential probability for new block cipher design. In particular, an increasing number of cryptographic schemes have been designed based on round-reduced block ciphers, e.g., AEGIS [22], SNOW-V [23], and Rocca [24]. A better understanding of the security of round-reduced block ciphers can provide valuable guidance for future block cipher designers.

TABLE III
PROBABILITY GAP BETWEEN DIFFERENTIAL CLUSTERS AND CHARACTERISTICS

Cipher	Type of Differential	Round	Probability
Midori-64 [14]	The best characteristic	7	2^{-70} [14]
	The best characteristic in our cluster	7	2^{-72}
	Congruent differential cluster	7	$2^{-63.22} = 2^{6.78} \times 2^{-70} = 2^{8.78} \times 2^{-72}$
	Semicongruent differential cluster	7	$2^{-52.25} = 2^{17.75} \times 2^{-70} = 2^{19.75} \times 2^{-72}$
Midori-64 with improved S-box [1]	The best characteristic	6	2^{-68} [1]
	Semicongruent differential cluster	6	$2^{-58.58}$=$2^{9.42} \times 2^{-68}$
CRAFT-64 [15]	The best characteristic	10	2^{-72} [15]
	The best characteristic in our cluster	10	2^{-72}
	Congruent differential cluster	10	$2^{-58.88} = 2^{13.12} \times 2^{-72}$
	Semicongruent differential cluster	10	$2^{-42.32} = 2^{29.68} \times 2^{-72}$
SKINNY-64 [13]	The best characteristic	8	2^{-72} [13]
	The best characteristic in our cluster	8	2^{-72}
	Congruent differential cluster	8	$2^{-62.81} = 2^{9.19} \times 2^{-72}$
	Semicongruent differential cluster	8	$2^{-50.72} = 2^{21.28} \times 2^{-72}$
SKINNY-128 [13]	The best characteristic	15	2^{-132}[13]
	Semicongruent differential cluster	15	$2^{-122.9} = 2^{9.1} \times 2^{-132}$
SKINNY-64 with improved S-box [1]	The best characteristic	8	2^{-76} [1]
	Semicongruent differential cluster	8	$2^{-60.74}$=$2^{15.26} \times 2^{-76}$

This paper proposes a novel technique to estimate the resistance against differential cryptanalysis for binary SPN ciphers by introducing the congruent differential cluster. For a binary AES-like cipher, which is the most popular instance of binary SPN ciphers, we introduce the semicongruent differential cluster and add more characteristics into this cluster. For congruent differential clusters, the probability calculation involves the multiplication of several $2^n \times 2^n$ matrices, where n indicates the size of the S-box. Moreover, we present an efficient algorithm (the source code is available at https://github.com/hahahai123/cluster.git) to calculate the probability of a semicongruent differential cluster. Compared to automatic methods, our approach requires fewer computational resources and often yields results in a shorter period of time. And our approach has better generalization with the help of limited computational resources compared to the theoretical derivation methods. Our method has provided the optimal results for the target ciphers in some rounds. And from both theoretical and experimental viewpoints, our methods are insensitive to the size of the S-boxes and the number of the rounds and can thus serve as an efficient tool for estimating the differential security of the target block ciphers.

We believe that congruent and semicongruent clusters can quickly evaluate the resistance to differential cryptanalysis of binary SPN ciphers and binary AES-like ciphers. Thus, it is interesting to extend these two kinds of clusters to SPN ciphers with bit-level linear layers. The difficulty of this extension lies in how to describe the development of patterns for such ciphers. We consider it as an open problem for our future research.

APPENDIX A
INTRODUCTION TO MIDORI-64

Midori is a family of AES-like ciphers, published at ASIACRYPT 2015 [14]. This cipher has been advertised as one of the first lightweight ciphers optimized in terms of the energy consumed by the circuit per bit in encryption or decryption operations.

TABLE IV
S-BOX OF MIDORI-64

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$S(x)$	C	A	D	3	E	B	F	7	8	9	1	5	0	2	4	6

The round function of Midori-64 consists of the S-layer and P-layer and uses the following 4×4 array named "state" as a data expression.

$$State = \begin{pmatrix} s_0 & s_4 & s_8 & s_{12} \\ s_1 & s_5 & s_9 & s_{13} \\ s_2 & s_6 & s_{10} & s_{14} \\ s_3 & s_7 & s_{11} & s_{15} \end{pmatrix},$$

where the size of each cell is 4 bits for Midori-64. A 64-bit plaintext P is loaded into the state.

The round function of Midori consists of an S-layer SubCell, P-layers ShuffleCell and MixColumn, and a key-addition layer KeyAdd. Each layer updates the 64-bit state as follows.

SubCell: A 4-bit S-box is applied to every 4-bit cell of $State$ in parallel. The 4-bit S-box of Midori-64 is presented in Table IV.

ShuffleCell: Each cell of the state is permuted as follows:

$$(s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}) \rightarrow (s_0, s_{10}, s_5, s_{15}, s_{14}, s_4, s_{11}, s_1, s_9, s_3, s_{12}, s_6, s_7, s_{13}, s_2, s_8).$$

MixColumn: M is applied to every 32-bit column of the state, i.e., for $i = 0, 4, 8, 12$

$$(s_i, s_{i+1}, s_{i+2}, s_{i+3})^\tau \leftarrow M \times (s_i, s_{i+1}, s_{i+2}, s_{i+3})^\tau,$$

TABLE V
S-BOX OF SKINNY-64

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$S(x)$	C	6	9	0	1	A	2	B	3	8	5	D	4	E	7	F

where the binary matrix M is defined as

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

KeyAdd(RK_i, $State$): The round key RK_i is XORed to $State$.

Clearly, Midori-64 is a binary SPN cipher that operates on a 4-bit word.

APPENDIX B
INTRODUCTION TO SKINNY-64

SKINNY is a family of AES-like ciphers, published at CRYPTO 2016 [13]. As a tweakable block cipher, SKINNY has excellent hardware/software implementation performance.

The round function of SKINNY-64 consists of SubCells, AddConstants, Add-RoundTweakey, ShiftRows, and MixColumns. We use the following 4×4 array named "state" as a data expression.

$$State = \begin{pmatrix} s_0 & s_1 & s_2 & s_3 \\ s_4 & s_5 & s_6 & s_7 \\ s_8 & s_9 & s_{10} & s_{11} \\ s_{12} & s_{13} & s_{14} & s_{15} \end{pmatrix},$$

where the size of each cell is 4 bits for SKINNY-64. A 64-bit plaintext P is loaded into the state. Each layer updates the 128-bit state as follows.

SubCell: A 4-bit S-box is applied to every 4-bit cell of the state in parallel. The 4-bit S-box of SKINNY-64 is presented in Table V.

ShiftRows: Each cell of the state is rotated to the right. Specifically, the first, second, third, and fourth cell rows are rotated by 0, 1, 2, and 3 positions to the right, respectively.

MixColumn: Each column of the cipher internal state array is multiplied by the following binary matrix M:

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

APPENDIX C
INTRODUCTION TO CRAFT-64

CRAFT is a family of AES-like ciphers, published at ToSC 2019 [15]. The efficient protection of CRAFT-64 implementations against differential fault analysis (DFA) attacks was one of the main design criteria.

The round function of CRAFT-64 consists of MixColumns, AddConstants, Add-RoundTweakey, PermuteNibbles, and SubCells. We use the following 4×4 array named "state" as a data expression.

$$State = \begin{pmatrix} s_0 & s_1 & s_2 & s_3 \\ s_4 & s_5 & s_6 & s_7 \\ s_8 & s_9 & s_{10} & s_{11} \\ s_{12} & s_{13} & s_{14} & s_{15} \end{pmatrix},$$

where the size of each cell is 4 bits for SKINNY-64. A 64-bit plaintext P is loaded into the state. Each layer updates the 128-bit state as follows.

MixColumn: Each column of the cipher internal state array is multiplied by the following binary matrix M:

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

PermuteNibbles: Each cell of the state is permuted as follows:

$$(s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}) \rightarrow (s_{15}, s_{12}, s_{13}, s_{14}, s_{10}, s_9, s_8, s_{11}, s_6, s_5, s_4, s_7, s_1, s_2, s_3, s_0).$$

SubCell: A 4-bit S-box is applied to every 4-bit cell of the state in parallel. The 4-bit S-box of CRAFT-64 is the same as that of Midori-64.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions. The author Ting Cui thanks Han Gu for the sumptuous dinner on the Chinese New Year of the Rabbit.

REFERENCES

[1] Y. Todo and Y. Sasaki, "Designing s-boxes providing stronger security against differential cryptanalysis for ciphers using byte-wise XOR," in *Proc. 28th Int. Conf. Sel. Areas Cryptogr.* Cham, Switzerland: Springer, 2022, pp. 179–199.

[2] *Data Encryption Standard*, Standard 46, U.S. Dept. Commerce, Federal Inf. Process. Standard (FIPS), Washington, DC, USA, 1977.

[3] E. Biham and A. Shamir, "Differential cryptanalysis of des-like cryptosystems," *J. Cryptol.*, vol. 4, no. 1, pp. 3–72, Jan. 1991.

[4] E. Biham and A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*. New York, NY, USA: Springer, 1993.

[5] M. Matsui, "Linear cryptanalysis method for DES cipher," in *Proc. Workshop Theory Appl. Cryptograph. Techn.*, Lofthus, Norway. Berlin, Germany: Springer, 1994, pp. 386–397.

[6] K. Nyberg and L. R. Knudsen, "Provable security against a differential attack," *J. Cryptol.*, vol. 8, no. 1, pp. 27–37, Dec. 1995.

[7] L. O'Connor, "On the distribution of characteristics in bijective mappings," *J. Cryptol.*, vol. 8, no. 2, pp. 67–86, Mar. 1995.

[8] J. Daemen and V. Rijmen, "The wide trail design strategy," in *The Design of Rijndael* (Information Security and Cryptography). Berlin, Germany: Springer, 2020, pp. 125–147.

[9] L. Keliher, "Refined analysis of bounds related to linear and differential cryptanalysis for the AES," in *Proc. 4th Int. Conf. Adv. Encryption Standard*, Bonn, Germany. Berlin, Germany: Springer, 2005, pp. 42–57.

[10] R. Ankele and S. Kölbl, "Mind the gap—A closer look at the security of block ciphers against differential cryptanalysis," in *Proc. 25th Int. Conf. Sel. Areas Cryptogr.*, Calgary, AB, Canada. Cham, Switzerland: Springer, 2019, pp. 163–190.

[11] O. Dunkelman, A. Kumar, E. Lambooij, and S. K. Sanadhya, "Counting active S-boxes is not enough," in *Proc. 21st Int. Conf. Cryptol. India*, Bangalore, India. Cham, Switzerland: Springer, 2020, pp. 332–344.

[12] G. Leurent, C. Pernot, and A. Schrottenloher, "Clustering effect in SIMON and SIMECK," in *Proc. 27th Int. Conf. Theory Appl. Cryptol. Inf. Secur.*, Singapore. Cham, Switzerland: Springer, 2021, pp. 272–302.
[13] C. Beierle et al., "The skinny family of block ciphers and its low-latency variant mantis," in *Proc. 36th Annu. Int. Cryptol. Conf.*, Santa Barbara, CA, USA. Berlin, Germany: Springer, 2016, pp. 123–153.
[14] S. Banik et al., "Midori: A block cipher for low energy," in *Proc. 21st Int. Conf. Theory Appl. Cryptol. Inf. Secur.*, Auckland, New Zealand. Berlin, Germany: Springer, 2015, pp. 411–436.
[15] C. Beierle, G. Leander, A. Moradi, and S. Rasoolzadeh, "CRAFT: Lightweight tweakable block cipher with efficient protection against DFA attacks," *IACR Trans. Symmetric Cryptol.*, vol. 2019, no. 1, pp. 5–45, Mar. 2019.
[16] H. Liu, W. Zhang, J. Zhang, and X. Sun, "Clustering of differentials in CRAFT with correlation matrices," *Int. J. Intell. Syst.*, vol. 37, no. 12, pp. 12113–12134, Dec. 2022.
[17] H. Hadipour, S. Sadeghi, M. M. Niknam, L. Song, and N. Bagheri, "Comprehensive security analysis of CRAFT," *IACR Trans. Symmetric Cryptol.*, vol. 2019, no. 4, pp. 290–317, Jan. 2020.
[18] X. Lai, J. L. Massey, and S. Murphy, "Markov ciphers and differential cryptanalysis," in *Proc. Workshop Theory Appl. Cryptograph. Techn.*, Brighton, U.K. Berlin, Germany: Springer, 1991, pp. 17–38.
[19] B. Sun, M. Liu, J. Guo, V. Rijmen, and R. Li, "Provable security evaluation of structures against impossible differential and zero correlation linear cryptanalysis," in *Proc. 35th Annu. Int. Conf. Theory Appl. Cryptograph. Techn.*, Vienna, Austria. Berlin, Germany: Springer, 2016, pp. 196–213.
[20] E. Biham, R. Anderson, and L. Knudsen, "Serpent: A new block cipher proposal," in *Proc. Int. Workshop Fast Softw. Encryption*. Berlin, Germany: Springer, 1998, pp. 222–238.
[21] A. Biryukov, C. De Canniere, A. Braeken, and B. Preneel, "A toolbox for cryptanalysis: Linear and affine equivalence algorithms," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, Warsaw, Poland. Berlin, Germany: Springer, 2003, pp. 33–50.
[22] H. Wu and B. Preneel, "AEGIS: A fast authenticated encryption algorithm," in *Proc. 20th Int. Conf. Sel. Areas Cryptogr.*, Burnaby, BC, Canada. Berlin, Germany: Springer, 2014, pp. 185–201.
[23] P. Ekdahl, T. Johansson, A. Maximov, and J. Yang, "A new SNOW stream cipher called SNOW-V," *ToSC*, vol. 2019, no. 3, pp. 1–42, Sep. 2019.
[24] K. Sakamoto, F. Liu, Y. Nakano, S. Kiyomoto, and T. Isobe, "Rocca: An efficient AES-based encryption scheme for beyond 5G," *IACR Trans. Symmetric Cryptol.*, vol. 2021, no. 2, pp. 1–30, Jun. 2021.

Ting Cui received the Ph.D. degree from the Institute of Information Science and Technology, Zhengzhou, China, in 2013. He is currently a Professor with PLA SSF Information Engineering University, Zhengzhou. His current research interests include block cipher designs and cryptanalysis.

Yiming Mao is currently pursuing the master's degree with PLA SSF Information Engineering University, Zhengzhou, China. His research interests include cryptanalysis of block ciphers.

Yang Yang received the Ph.D. degree from the Institute of Information Science and Technology, Zhengzhou, China. She is currently an Associate Professor with PLA SSF Information Engineering University, Zhengzhou. Her current research interests include block cipher designs and cryptanalysis.

Yi Zhang is currently pursuing the master's degree with PLA SSF Information Engineering University, Zhengzhou, China. His research interests include cryptanalysis of block ciphers.

Jiyan Zhang received the Ph.D. degree from PLA SSF Information Engineering University, Zhengzhou, China, in 2022. He is currently a Lecturer with PLA SSF Information Engineering University. He has published articles in highly ranked journals, such as IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and IEEE TRANSACTIONS ON INFORMATION THEORY. His research interests include cryptanalysis of symmetry ciphers and the Internet of Things. Additionally, he is a reviewer of several international journals and conferences.

Chenhui Jin is currently a Professor and a Ph.D. Supervisor with PLA SSF Information Engineering University, Zhengzhou, China. His research interests include cryptography, information security, and cyberspace security.