SCIENCE CHINA
Information Sciences

• LETTER •

March 2020, Vol. 63 139111:1–139111:3
https://doi.org/10.1007/s11432-018-9495-x

Improved distinguisher search techniques based on parity sets

Xiaofeng XIE & Tian TIAN*

National Digital Switching System Engineering & Technological Research Center, P.O. Box 407, Zhengzhou 450001, China

Received 12 February 2018/Accepted 15 June 2018/Published online 10 February 2020

Citation Xie X F, Tian T. Improved distinguisher search techniques based on parity sets. Sci China Inf Sci, 2020, 63(3): 139111, https://doi.org/10.1007/s11432-018-9495-x

Dear editor,

Division property was a technique proposed by Todo at EUROCRYPT 2015 to search integral distinguishers against block ciphers [1]. Todo [2] applied this technique to perform structural evaluation against both the Feistel and the SPN constructions and attacked the full MISTY1. Subsequently, many improved techniques based on the division property were proposed [3, 4]. At FSE 2016, Todo and Morii [3] introduced the bit-based division property and proved its effectiveness to find distinguishers against non-S-box-based ciphers.

Although more accurate integral distinguishers were found by using the bit-based division property, it could not be applied to ciphers whose block length is more than 32 because of its high time and memory complexities. Based on Todo's work, Xiang et al. [5] converted the distinguisher search algorithm based on the bit-based division property into an MILP problem at ASIACRYPT 2016. With this method, they obtained a series of improved results including a 9-round PRESENT distinguisher with one balanced bit. This distinguisher is one of the best-known distinguishers related to round numbers.

At CRYPTO 2016, Boura and Cauteaut [6] introduced the parity set to study the division property. They utilized the parity set to exploit further properties of the PRESENT S-box and the PRESENT linear layer, leading to several improved distinguishers against reduced-round PRESENT. Because more properties of the S-box and the linear layer are utilized, parity sets can find more accurate integral characteristics. However, although the authors did not point out, a parity set requires higher time and memory complexities than the division property does. Our work aims at reducing time and memory complexities when using parity sets to search integral distinguishers. As a result, we introduce the idea of meet-in-the-middle into the distinguisher search. To illustrate our techniques, we performed extensive experiments on PRESENT and found a 9-round distinguisher with 22 balanced bits.

Notation 1 (Bit product function)**.** Let $\boldsymbol{u}$, $\boldsymbol{x} \in \mathbb{F}_2^n$. Denote

$$\boldsymbol{x}^{\boldsymbol{u}} = \prod_{i=1}^{n} x[i]^{u[i]},$$

and for $\boldsymbol{u}$, $\boldsymbol{x} \in \mathbb{F}_2^{n_1} \times \mathbb{F}_2^{n_2} \times \cdots \times \mathbb{F}_2^{n_m}$, where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$, $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$, define bit product function as

$$\boldsymbol{x}^{\boldsymbol{u}} = \prod_{i=1}^{m} \boldsymbol{x}_i^{\boldsymbol{u}_i}.$$

Notation 2 (Comparison between vectors)**.** For $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{Z}^m$, denote $\boldsymbol{a} \geqslant \boldsymbol{b}$ if $a_i \geqslant b_i$ for all $0 < i \leqslant m$, and denote $\boldsymbol{a} > \boldsymbol{b}$ if $\boldsymbol{a} \geqslant \boldsymbol{b}$ but $\boldsymbol{a} \neq \boldsymbol{b}$.

For $\boldsymbol{u} \in \mathbb{F}_2^n$, let us denote

$$\mathrm{Prec}(\boldsymbol{u}) = \{\boldsymbol{v} \in \mathbb{F}_2^n : \boldsymbol{v} \leqslant \boldsymbol{u}\},$$

$$\mathrm{Succ}(\boldsymbol{u}) = \{\boldsymbol{v} \in \mathbb{F}_2^n : \boldsymbol{u} \leqslant \boldsymbol{v}\}.$$

* Corresponding author (email: tiantian_d@126.com)

Theorem 1. If $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}_2^{nt}$ satisfy $\boldsymbol{u} \succcurlyeq \boldsymbol{v}$, then $W(\boldsymbol{u}) \geqslant W(\boldsymbol{v})$.

Notation 3 (Comparison between sets)**.** Let A and B be two sets whose elements are in $\mathbb{F}_2^n$. Denote $A \succcurlyeq B$ if there exist $\boldsymbol{a} \in A$ and $\boldsymbol{b} \in B$ with $\boldsymbol{a} \succcurlyeq \boldsymbol{b}$, and $A \not\succcurlyeq B$ if none of such couple exists.

Proposition 1. Let A and B be two sets whose elements are in $\mathbb{F}_2^n$ with $A \succcurlyeq B$. If there are $\boldsymbol{a}_1, \boldsymbol{a}_2 \in A, \boldsymbol{b}_1, \boldsymbol{b}_2 \in B$ such that $\boldsymbol{a}_2 \succcurlyeq \boldsymbol{a}_1$ and $\boldsymbol{b}_1 \succcurlyeq \boldsymbol{b}_2$, then $A \setminus \{\boldsymbol{a}_1\} \succcurlyeq B \setminus \{\boldsymbol{b}_1\}$.

Notation 4 (Round function)**.** Let F be a permutation of $\mathbb{F}_2^n$ defined by

$$F : \boldsymbol{x} = (x_1, x_2, \ldots, x_n) \mapsto \boldsymbol{y} = (y_1, y_2, \ldots, y_n).$$

Then every y_i can be seen as a Boolean function on $x_1, x_2, \ldots, x_n$, denoted by $y_i = F_i(\boldsymbol{x})$. For a positive integer r, we denote F^r as a composition of r permutation F.

Definition 1 (Division property [1])**.** Let X be a multiset whose elements belong to $\mathbb{F}_2^n$. Then, X has the division property D_k^n when it fulfills the following conditions: For $\boldsymbol{u} \in \mathbb{F}_2^n$, the parity of $\boldsymbol{x}^{\boldsymbol{u}}$ over all elements in X is always even when $wt(\boldsymbol{u}) < k$. For further study of the division property, please refer to [1, 4] in detail.

Definition 2 (Parity set [6])**.** Let X be a set whose elements take values of $\mathbb{F}_2^n$. The parity set of X is denoted by $\mathcal{U}(X)$ and defined as follows:

$$\mathcal{U}(X) = \left\{ \boldsymbol{u} \in \mathbb{F}_2^n : \bigoplus_{\boldsymbol{x} \in X} \boldsymbol{x}^{\boldsymbol{u}} = 1 \right\}.$$

Remark 1. If the parity set $\mathcal{U}(X)$ of X is known, then the division property of X is given by D_k^n, where

$$k = \min_{\boldsymbol{u} \in \mathcal{U}(X)} wt(\boldsymbol{u}).$$

For the propagation rules of the parity set on SPN, please refer to [1].

For an input set X and a round function E, denote the parity set after r_1-round encryption as $\mathcal{U}(E^{r_1}(X))$, and the algebraic normal form (ANF) of the i-th output bit after r_2-round encryption as $E_i^{r_2}(\boldsymbol{x})$. If all the terms appearing in $E_i^{r_2}(\boldsymbol{x})$ are not divisible by any term in $\{\boldsymbol{x}^{\boldsymbol{u}} : \boldsymbol{u} \in \mathcal{U}(E^r(X))\}$, then the i-th output bit of $(r_1 + r_2)$-round encryption is balanced.

Based on this observation, we improved the integral distinguisher search by utilizing the meet-in-the-middle technique which divides the n-round propagation of parity sets into n_1-round propagation of parity sets and $(n - n_1)$-round propagation of the ANF.

Next, we propose a new concept, which we call term set, to describe the ANF and show the propagation rules of the term set on SPN.

Definition 3 (Term set)**.** Let $f(\boldsymbol{x})$ be an n-variable Boolean function. The term set of $f(\boldsymbol{x})$ denoted by $T(f)$ is the subset of $\mathbb{F}_2^n$ defined by

$$T(f) = \{\boldsymbol{u} \in \mathbb{F}_2^n : \boldsymbol{x}^{\boldsymbol{u}} \text{ appears in the ANF of } f(\boldsymbol{x})\}.$$

Proposition 2. Let S be an S-box over $\mathbb{F}_2^m$. Denote

$$Ts(\boldsymbol{u}) = \{\boldsymbol{v} \in \mathbb{F}_2^m : \boldsymbol{x}^{\boldsymbol{v}} \text{ appears in the ANF of } S^{\boldsymbol{u}}(\boldsymbol{x})\}.$$

Then for an m-variable Boolean function f with the term set $T(f)$, we have

$$T(f(S(\boldsymbol{x}))) \subseteq \bigcup_{\boldsymbol{u} \in T(f)} Ts(\boldsymbol{u}).$$

Proposition 3. Let $\mathcal{S}$ be a permutation of $\mathbb{F}_2^{mt}$ which consists of t parallel independent S-boxes over $\mathbb{F}_2^m$, namely, $\mathcal{S}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) = (S(\boldsymbol{x}_1), \ldots, S(\boldsymbol{x}_t))$. For an mt-variable Boolean function f with the term set $T(f)$, we have

$$T(f(\mathcal{S})) \subseteq \bigcup_{(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_t) \in T(f)} Ts_1(\boldsymbol{u}_1) \times \ldots \times Ts_t(\boldsymbol{u}_t).$$

Proposition 4. Let f be an n-variable Boolean function with the term set $T(f)$. For any $\boldsymbol{k} \in \mathbb{F}_2^n$, the term set of $f(\boldsymbol{k} \oplus \boldsymbol{x}) = (x_1 \oplus k_1, \ldots, x_n \oplus k_n)$ satisfies

$$T(f(\boldsymbol{k} \oplus \boldsymbol{x})) \subseteq \bigcup_{\boldsymbol{u} \in T(f)} \mathrm{Prec}(\boldsymbol{u}).$$

Then, the term set after one round encryption can be deduced by Propositions 2 and 4, i.e.,

$$T(f(S(\boldsymbol{x} \oplus \boldsymbol{k}))) \subseteq \bigcup_{\boldsymbol{u} \in T(f)} \bigcup_{\boldsymbol{v} \in Ts(\boldsymbol{u})} \mathrm{Prec}(\boldsymbol{v}), \text{ for } \boldsymbol{k} \in \mathbb{F}_2^n.$$

The proofs of these propositions could be found through https://eprint.iacr.org/2018/447.

We can also search distinguishers by term sets only. If there exists a $\boldsymbol{u} \in \mathbb{F}_2^n$ satisfying $\mathrm{Succ}(\boldsymbol{u}) \bigcap T(E_i^r) = \emptyset$, then a r-round distinguisher whose input set is $\mathrm{Prec}(\boldsymbol{u})$ is found. However, the time and memory complexities will be very high. Thus, we took advantage of the meet-in-the-middle technique so that the term set and the parity set could be combined to reduce time and memory complexities.

In order to find a distinguisher, we need to compare $T(E_i^{r_2})$ with $\mathcal{U}(E^{r_1}(X))$ and verify whether $T(E_i^{r_2}) \not\succcurlyeq \mathcal{U}(E^{r_1}(X))$. Our distinguisher search algorithm consists of five steps, which can be described as follows.

Step 1. Choose the propagation round numbers r_1 and r_2 for the parity set and the term set respectively, where $r_1 + r_2 = r$.

Step 2. Choose an input set X.

Step 3. Calculate the parity set $\mathcal{U}(E^{r_1}(X))$.

Step 4. Calculate the term sets $T(E_i^{r_2})$ for $1 \leqslant i \leqslant n$.

Step 5. Compare $\mathcal{U}(E^{r_1}(X))$ with $T(E_i^{r_2})$ for $1 \leqslant i \leqslant n$. If $\mathcal{U}(E^{r_1}(X)) \not\geqslant T(E_i^{r_2})$, then the i-th output bit in r-round encryption is balanced. If none of such intersections is empty, then choose another X and go to Step 2.

We also propose some novel techniques to make our algorithm more efficient.

Size reduce operation. For the term set $T(E_i^r(\boldsymbol{x}))$, the size reduce operation R^t removes all the elements $\boldsymbol{v} \in T(E_i^r(x))$ such that there is an element $\boldsymbol{v}' \in T(E_i^r(\boldsymbol{x}))$ with $\boldsymbol{v}' \geqslant \boldsymbol{v}$. As for a parity set, the operation R^u removes all the elements $\boldsymbol{u} \in \mathcal{U}(E^{r_1}(X))$ such that there is an element $\boldsymbol{u}' \in \mathcal{U}(E^{r_1}(X))$ with $\boldsymbol{u} \geqslant \boldsymbol{u}'$. It can be deduced from Proposition 1 that the comparison result of $T(E_i^r(\boldsymbol{x}))$ and $\mathcal{U}(E^{r_1}(X))$ is the same as the comparison result of $R^t(T(E_i^r(\boldsymbol{x})))$ and $R^u(\mathcal{U}(E^{r_1}(X)))$.

Observation 1. The PRESENT super S-boxes can work independently in the 2-round encryption.

Reducing look-up table. Based on Observation 1, we can easily construct a 2-round propagation table for the super S-box by calculating

$$R^u\big(\mathcal{U}(\mathcal{S}(P(\mathcal{S}(X)))))\big)$$

for all possible inputs, where $\mathcal{S}$ is a permutation of $\mathbb{F}_2^{4n}$ consisting of four PRESENT S-boxes

$$\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) = (S(\boldsymbol{x}_1), S(\boldsymbol{x}_2), S(\boldsymbol{x}_3), S(\boldsymbol{x}_4)).$$

Multiple comparison. This technique attempts to remove the terms that have no multiple in $\mathcal{U}$; if no term is divisible by a vector in $\mathcal{U}$, then it is clear that the output bit is balanced. We tried to judge such divisibility in terms of degree order and alphabet order. For the details of this technique, refer to https://eprint.iacr.org/2018/447.

To illustrate our techniques, we apply our algorithm to the PRESENT distinguisher search.

Observation 2. The cubic terms in the ANFs of the second and fourth coordinates of the PRESENT S-box (say S_2 and S_4) are the same [7].

As a result, the xor of these two coordinates

$$S_2 \oplus S_4 = 1 \oplus x_1 \oplus x_2x_3 \oplus x_2x_4 \oplus x_3x_4$$

has only degree 2. Moreover, every term in $S_2 \oplus S_4$ has a multiple in S_2 and S_4 respectively. Hence, $S_2 \oplus S_4$ may be balanced even if S_2 and S_4 are unbalanced.

We tried to find 10-round PRESENT distinguishers first, but the result of the rightmost output bit is unbalanced for all the input sets with dimension 63. It seems that the ANF of this output bit is the simplest among 64 output bits, and therefore, our results show that the PRESENT probably has no 10-round integral distinguishers by only using the division property. Then, we focus on the 9-round PRESENT, and find a distinguisher with 22 balanced output bits.

Input:
(*aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaac*),
Output:
(????????????b_3?b_3b ????????????b_2?b_2b ????????????b_1?b_1b *bbbbbbbbbbbbbbbb*),
where "*c*" means a constant bit, "*a*" means an active bit, "?"means an unknown bit, and "*b*" means a balanced bit. In addition, the presence of bits with the same notation b_i means their addition is balanced.

Conclusion. In this study, we proposed a concept called the term set to propagate information of the ANF. With term sets, we improved the distinguisher search method based on the parity set in terms of both memory and time complexities. From the relation between the parity set and the bit-based division property, it was found that the term set could also be applied to improve the distinguisher search method based on the bit-based division property.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61672533).

References

1 Todo Y. Structural evaluation by generalized integral property. Lect Notes Comput Sci, 2015, 9056: 287–314
2 Todo Y. Integral cryptanalysis on full MISTY1. J Cryptol, 2017, 30: 920–959
3 Todo Y, Morii M. Bit-based division property and application to simon family. Lect Notes Comput Sci, 2016, 9783: 357–377
4 Sun L, Wang W, Wang M Q. Automatic search of bit-based division property for ARX ciphers and word-based division property. Lect Notes Comput Sci, 2017, 10624: 128–157
5 Xiang Z J, Zhang W T, Bao Z Z, et al. Applying MILP method to searching integral distinguishers based on division property for 6 lightweight block ciphers. Lect Notes Comput Sci, 2016, 10031: 648–678
6 Boura C, Canteaut A. Another view of the division property. Lect Notes Comput Sci, 2016, 9814: 654–682
7 Bogdanov A, Knudsen L R, Leander G, et al. PRESENT: an ultra-lightweight block cipher. Lect Notes Comput Sci, 2007, 4727: 450–466