# EasyBC: A Cryptography-Specific Language for Security Analysis of Block Ciphers against Differential Cryptanalysis

PU SUN, ShanghaiTech University, China
FU SONG*, Institute of Software at Chinese Academy of Sciences, China and University of Chinese Academy of Sciences, China
YUQI CHEN, ShanghaiTech University, China
TAOLUE CHEN, Birkbeck, University of London, UK

Differential cryptanalysis is a powerful algorithmic-level attack, playing a central role in evaluating the security of symmetric cryptographic primitives. In general, the resistance against differential cryptanalysis can be characterized by the maximum expected differential characteristic probability. In this paper, we present generic and extensible approaches based on mixed integer linear programming (MILP) to bound such probability. We design a high-level cryptography-specific language EasyBC tailored for block ciphers and provide various rigorous procedures as differential denotational semantics, to automate the generation of MILP from block ciphers written in EasyBC. We implement an open-sourced tool that provides support for fully automated resistance evaluation of block ciphers against differential cryptanalysis. The tool is extensively evaluated on 23 real-life cryptographic primitives including all the 10 finalists of the NIST lightweight cryptography standardization process. The experiments confirm the expressivity of EasyBC and show that the tool can effectively prove the resistance against differential cryptanalysis for all block ciphers under consideration. EasyBC makes resistance evaluation against differential cryptanalysis easily accessible to cryptographers.

CCS Concepts: • **Theory of computation**; • **Security and privacy**;

Additional Key Words and Phrases: Cryptography-Specific Language, Block Ciphers, Differential Cryptanalysis

**ACM Reference Format:**
Pu Sun, Fu Song, Yuqi Chen, and Taolue Chen. 2024. EasyBC: A Cryptography-Specific Language for Security Analysis of Block Ciphers against Differential Cryptanalysis. *Proc. ACM Program. Lang.* 8, POPL, Article 29 (January 2024), 49 pages. https://doi.org/10.1145/3632871

## 1 INTRODUCTION

A block cipher is a symmetric cryptographic technique that uses the same key to encrypt and decrypt data in fixed-size blocks. Guided by the design principles of block ciphers [Shannon 1949], a vast number of block ciphers have been proposed, varying in, e.g., network structures and block sizes. Many of them have been standardized and are widely used in daily life to provide confidentiality, integrity, and authentication. Differential cryptanalysis, proposed by [Biham and Shamir 1990], is a powerful algorithmic-level attack against block ciphers by analyzing the effect of particular differences in input pairs on the differences of pairs of intermediate states under

*Corresponding author

Authors' addresses: Pu Sun, ShanghaiTech University, Shanghai, China, sunpu@shanghaitech.edu.cn; Fu Song, Institute of Software at Chinese Academy of Sciences, Beijing, China and University of Chinese Academy of Sciences, Beijing, China, songfu@ios.ac.cn; Yuqi Chen, ShanghaiTech University, Shanghai, China, chenyq@shanghaitech.edu.cn; Taolue Chen, Birkbeck, University of London, London, UK, t.chen@bbk.ac.uk.

ACM 2475-1421/2024/1-ART29
https://doi.org/10.1145/3632871

the same key. It has proved to be a very effective attack, which has broken block ciphers such as DES [Biham and Shamir 1990], FEAL [Aoki et al. 1997], WARP [Teh and Biryukov 2022], and K-Cipher [Mahzoun et al. 2022]. As a result, a provable guarantee of the resistance of block ciphers against differential cryptanalysis has become a standard criterion for new block ciphers and indeed a basic requirement for them to be standardized [Katz and Lindell 2014]. In light of the diversity and wide deployment of block ciphers, a generic, and ideally automated, approach which can be used to evaluate their resistance against differential cryptanalysis, becomes indispensable.

In general, the resistance of block ciphers against differential cryptanalysis is commonly characterized by the maximum expected differential characteristic probability (MaxEDCP for short, the definition of which is fairly standard but technical, and will be given in Section 2.1). A block cipher is considered to be resistant against differential cryptanalysis if the MaxEDCP is no greater than $2^{-\mathscr{b}}$, where $\mathscr{b}$ is the block size of the block cipher [Heys 2002a; Lai et al. 1991]. In light of this, the central task of security analysis for block ciphers against differential cryptanalysis is reduced to computing such probability. To this end, [Matsui 1994] proposed a branch-and-bound searching algorithm that traverses differential characteristics in a depth-first manner and computes their probabilities during traversal. However, it becomes inefficient with the increasing of candidate differential characteristics. Various heuristics are then proposed to improve the efficiency [Aoki et al. 1997; Bao et al. 2014; Biryukov and Nikolić 2010; Ji et al. 2021], but they harness cipher-specific optimizations and thus require sophisticated programming skills.

Several alternative methods are introduced, which reduce to mixed integer linear program (MILP [Mouha et al. 2011]), Boolean satisfiability problem (SAT [Mouha and Preneel 2013]), or satisfiability modulo theory (SMT [Aumasson et al. 2014]). These methods allow cryptanalysts to specify the problem for each cipher using the input language of MILP/SAT/SMT solvers so that respective solvers can be harnessed. Plenty of modeling methods for cryptography-specific operations such as substitution-box (S-box), Exclusive-OR (XOR), and linear transformations, as well as heuristics, are proposed to improve efficiency and accuracy. However, insofar cryptanalysts have to manually model each cipher in the tool they choose to use, or at best write a model generation script for each cipher, which is usually intricate, error-prone, and laborious, as it requires the cryptanalysts to be familiar with the specific tool and a wide range of modeling methods. There appears to be a lack of language support, unified computational approaches, and fully automated tools for evaluating the resistance of block ciphers against differential cryptanalysis.

*Contributions.* In this work, our primary aim is to develop a generic and automated approach for evaluating the resistance of block ciphers against differential cryptanalysis. To achieve this goal, we begin by designing a novel high-level statically-typed, C-like cryptography-specific language, EasyBC (**Easy B**lock **C**ipher), tailored for block ciphers. Besides standard types and operations, EasyBC provides cryptography-specific types (e.g., S-boxes, P-boxes) and operations (e.g., substitution via S-box, linear transformation via P-box, or matrix-vector product), to facilitate the implementation of block ciphers. (Note that an S-box takes $m$ input bits and transforms them into $n$ output bits and P-box, short for Permutation-box, is an array specifying a permutation of inputs.) The language is fully specified by a formal grammar together with typing rules and operational semantics, enabling further automatic analysis.

Concretely speaking, the analysis of block cipher resistance against differential cryptanalysis primarily involves computing the MaxEDCP. However, calculating MaxEDCP precisely is practically infeasible. As a result, we employ a strategy to calculate a tight upper bound that is sufficient to demonstrate resistance. Specifically, we adopt the typical MILP-based approach where linear constraints are used to characterize the dependency (i.e., feasibility) between input and output differences of each operation, and the optimization objective is to minimize an upper bound. In particular, we give a rigorous procedure, formalized as a differential denotational semantics, to automate the generation of MILP from EasyBC programs, which not only unifies and optimizes the existing but also discovers new generation processes. Our approach is of generic nature, thanks to the expressivity of the EasyBC langauge, namely, a multitude of block ciphers can be handled in a unified way.

Technically, the generation of MILP can be done either at the word or bit level. The former is less involved and generates fewer constraints, but is limited to certain block ciphers; the latter approach is more fine-grained and has wider applicability. In both cases, the general strategy is to determine the lower bound of the minimum number of (differentially) active S-boxes, i.e., S-boxes whose input differences are nonzero under two executions [Biryukov and Nikolić 2010; Heys 2002b], from which the upper bound of the MaxEDCP can be deduced according to [Heys 2002b; Sun et al. 2014a]. While this strategy is efficient, it is important to note that the obtained upper bound may not always be sufficiently tight and may not be applicable for certain ciphers. To address this limitation, we introduce an extended bit-wise approach that directly bounds the MaxEDCP by encoding probabilities using additional Boolean variables. We have successfully implemented our approach as the first fully automated tool for evaluating the resistance of block ciphers against differential cryptanalysis. This tool eliminates the need for cryptanalysts to possess knowledge of MILP generation for cryptographic operations. Instead, they can simply write a program in EasyBC for a block cipher. Moreover, the generation of MILP from EasyBC program is modular, i.e., each cryptographic operation in EasyBC is associated with its own MILP generation rule and new generation rules could be easily added by implementing designated APIs. As a result, our approach exhibits excellent extensibility for new block ciphers, enabling a wider range of applicability.

To evaluate the tool, we implement 23 realistic cryptographic primitives with EasyBC, including all the 10 finalists of the NIST lightweight cryptography standardization process [NIST 2023] and other commonly used block ciphers, covering both substitution-permutation network (SPN) based ciphers (e.g., AES, PRESENT, and GIFT) and balanced Feistel networks (BFN) based ciphers (e.g., DES, LBLOCK, and TWINE). It turns out that EasyBC can express these block ciphers in a considerably more succinct way, demonstrating our language's expressiveness. Moreover, it turns out that our tool is able to effectively handle all realistic block ciphers under consideration, showcasing its capability in proving the resistance of block ciphers against differential cryptanalysis.

In addition, we compare various alternative MILP generation methods for cryptographic operations. Interestingly, we observe that certain methods may generate fewer constraints and variables. However, it is worth noting that such reductions may actually have a negative impact on the overall MILP-solving process. For instance, the recent S-box modeling method proposed by [Udovenko 2021] produces the fewest constraints, but also exhibits the least performance to solve those constraints. (Overall, it is significantly less efficient than some alternatives.) Our findings shed light on the selection of generation methods among various alternatives for practical applications.

We summarize the main contributions as follows.

- We design a high-level cryptography-specific language EasyBC tailored for block ciphers, enabling further automatic analysis.
- We give generic and extensible approaches for automated resistance evaluation of block ciphers written in EasyBC against differential cryptanalysis.
- We implement and extensively evaluate an open-sourced prototype of EasyBC, which confirms the expressiveness of EasyBC and the effectiveness of our approach.

*Structure.* The rest of the paper is organized as follows. Section 2 presents the background of block ciphers and differential cryptanalysis. Section 3 introduces EasyBC and an overview of our approach. Section 4 presents three key utilities used in our MILP generation. Section 5 and Section 6 describe the word-wise and bit-wise approach. Section 7 describes the extended bit-wise approach. Section 8 reports the experimental results. We discuss related work in Section 9 and conclude this work in Section 10.

## 2 BACKGROUND

Throughout this paper, $\mathbb{B}$ denotes the Boolean domain $\{0, 1\}$, and $\mathbb{N}$ denotes the set of non-negative integers. Boolean values are treated as integers in arithmetic computations. Given a vector/array $\vec{x}$, $\vec{x}_i$ denotes the $(i + 1)$-th entry. Given a matrix $M$, $M_i$ denotes the $(i + 1)$-th row, and $M_{i,j}$ denotes the $(j + 1)$-th entry of $M_i$. An *$m$-bitstream* is Boolean vector $\vec{b}$ with $m$ entries. We denote by $\oplus$ the bit-wise XOR operator. $\vec{b}\|\vec{b}' = (b_0, \cdots, b_m, b'_0, \cdots, b'_n)$ is the concatenation of two bitstreams $\vec{b} = (b_0, \cdots, b_m)$ and $\vec{b}' = (b'_0, \cdots, b'_n)$. We denote by $\mathsf{bin}(x)$ the binary representation of an unsigned integer $x$ as a bitstream.

### 2.1 Block ciphers

Block ciphers are a type of symmetric cryptography, which encrypts and decrypts data in fixed-size (e.g., 64 or 128 bits) blocks using the same key [Bogdanov 2010; Knudsen 1998].

*Definition 2.1.* A *block cipher* is a function $\mathsf{Enc} : \mathbb{B}^{\mathscr{k}} \times \mathbb{B}^{\mathscr{b}} \longrightarrow \mathbb{B}^{\mathscr{b}}$ such that for every key $K \in \mathbb{B}^{\mathscr{k}}$, $\mathsf{Enc}(K, \cdot)$ is a bijective function, where $\mathscr{b}$ is the block size and $\mathscr{k}$ is the key size.

Intuitively, $\mathsf{Enc}(K, \cdot)$ is a keyed-permutation that maps an input block to an output block, where a block is a $\mathscr{b}$-bitstream, the key $K$ determines which permutation to perform. The input and output of $\mathsf{Enc}(K, \cdot)$ are called *plaintext* and *ciphertext*, respectively. A block cipher $\mathsf{Enc}$ is *ideal* if it is defined by assigning a uniformly drawn permutation to each of the $2^{\mathscr{k}}$ keyed-permutations. An ideal block cipher is commonly considered to be computationally secure if the key size $\mathscr{k}$ is large enough since the brute-force attack requires $O(2^{\mathscr{k}})$ time. However, it is extremely difficult to implement an ideal block cipher for practical block sizes (e.g., 64 or 128), as one randomly drawn permutation $\mathsf{Enc}(K, \cdot)$ has to be stored for each given key $K$.

To be efficient yet strong, modern block ciphers apply several (possibly distinct) keyed permutations, where one keyed permutation is a round and implemented by a round function.

*Definition 2.2.* A *$\mathscr{r}$-round iterative block cipher* ($\mathscr{r}$-IBC) is a function $\mathsf{Enc} : \mathbb{B}^{\mathscr{k}} \times \mathbb{B}^{\mathscr{b}} \longrightarrow \mathbb{B}^{\mathscr{b}}$ such that for every key $K \in \mathbb{B}^{\mathscr{k}}$,

$$\mathsf{Enc}(K, \cdot) = \mathsf{Enc}_{\mathscr{r}}(K^{\mathscr{r}}, \cdot) \circ \cdots \circ \mathsf{Enc}_1(K^1, \cdot),$$

where $\mathsf{Enc}_i : \mathbb{B}^{\mathscr{k}} \times \mathbb{B}^{\mathscr{b}} \longrightarrow \mathbb{B}^{\mathscr{b}}$ is the $i$-th round with its subkey $K^i$ for $1 \leq i \leq \mathscr{r}$, symbol $\circ$ denotes function composition, and the subkeys are generated via a key schedule algorithm $g : \mathbb{B}^{\mathscr{k}} \longrightarrow (\mathbb{B}^{\mathscr{k}})^{\mathscr{r}}$, i.e., $g(K) = (K^1, \cdots, K^{\mathscr{r}})$.

Given a key $K \in \mathbb{B}^{\mathscr{k}}$, $\mathsf{Enc}(K, \cdot)$ is used for encryption and $\mathsf{Dec}(K, \cdot)$ is used for decryption, i.e., $\mathsf{Dec}(K, \cdot) = \mathsf{Enc}_1^{-1}(K^1, \cdot) \circ \cdots \circ \mathsf{Enc}_{\mathscr{r}}^{-1}(K^{\mathscr{r}}, \cdot)$, where $K^i$ is the $i$-th round subkey in Definition 2.2.

Block ciphers can be built in various ways following iterative cipher schemes. The two most widely used are substitution-permutation networks (SPN) [Kam and Davida 1979] and balanced Feistel networks (BFN) [Nyberg 1996]. Note that BFN is used in the former U.S. encryption standard (DES-Data Encryption Standard) [Fox 2000] and SPN is used in the current one (AES-Advanced Encryption Standard) [Daemen and Rijmen 1999]. A brief introduction of BFN and SPN is given in [Sun et al. 2023, Section A].

In the sequel, we fix a $\mathscr{r}$-IBC $\mathsf{Enc} : \mathbb{B}^{\mathscr{k}} \times \mathbb{B}^{\mathscr{b}} \longrightarrow \mathbb{B}^{\mathscr{b}}$ with rounds $\mathsf{Enc}_1, \cdots, \mathsf{Enc}_{\mathscr{r}}$. We assume that the attacker knows all dethe tails of the encryption and decryption except for the secret key.

## 2.2 Differential Cryptanalysis

Differential cryptanalysis recovers the secret key by exploiting the fact that the probability of some output differences of rounds in a *non-ideal* block cipher is higher than the expected value (i.e., $2^{-\mathscr{b}}$) for certain input differences [Biham and Shamir 1990]. We review the related concepts below.

**Difference and differential**. Given two $\mathscr{b}$-bitstreams $X \in \mathbb{B}^{\mathscr{b}}$ and $X' \in \mathbb{B}^{\mathscr{b}}$, their (XOR-) *difference* $\Delta X$ is defined by $\Delta X = X \oplus X'$. Note that for any fixed difference $\Delta X$, there are exactly $2^{\mathscr{b}}$ pairs $(X, X')$ such that $X \oplus X' = \Delta X$. A *differential* is defined to be a pair of differences $(\Delta X, \Delta Y)$.

Given a pair of inputs $(X, X') \in \mathbb{B}^n \times \mathbb{B}^n$ for a (deterministic) function $f : \mathbb{B}^n \longrightarrow \mathbb{B}^m$, the *input difference* of the function $f$ is the difference $\Delta X = X \oplus X'$ of the inputs $(X, X')$, and the *output difference* of $f$ is the difference $f(X) \oplus f(X')$ of the outputs $(f(X), f(X'))$. Clearly, when the input difference is fixed to be $\Delta X$, the output difference of an input $X$ is $f(X) \oplus f(X \oplus \Delta X)$.

The *probability* $\mathsf{Pr}_f(\Delta X, \Delta Y)$ of a given differential $(\Delta X, \Delta Y)$ for the function $f$ is the proportion of inputs $X \in \mathbb{B}^n$ such that the output difference $f(X) \oplus f(X \oplus \Delta X)$ is equal to $\Delta Y$, i.e.,

$$\mathsf{Pr}_f(\Delta X, \Delta Y) = \frac{|\{X \in \mathbb{B}^n \mid f(X) \oplus f(X \oplus \Delta X) = \Delta Y\}|}{2^n}.$$

In particular, for an $i$-th round $\mathsf{Enc}_i : \mathbb{B}^{\mathscr{k}} \times \mathbb{B}^{\mathscr{b}} \longrightarrow \mathbb{B}^{\mathscr{b}}$, with fixed subkey $K^i$, $\mathsf{Pr}_{\mathsf{Enc}_i(K^i,\cdot)}(\Delta X^{i-1}, \Delta X^i)$ is the probability of a differential $(\Delta X^{i-1}, \Delta X^i)$ for the function $\mathsf{Enc}_i(K^i, \cdot)$.

It is worth mentioning that in differential cryptanalysis, a key assumption is that the output difference $\mathsf{Enc}_i(K^i, X^{i-1}) \oplus \mathsf{Enc}_i(K^i, X^{i-1} \oplus \Delta X^{i-1})$ of the $i$-th round is independent of the subkey $K^i$ for any fixed input $X^{i-1}$ and input difference $\Delta X^{i-1}$. As a result, for clarity, $\mathsf{Pr}_{\mathsf{Enc}_i(K^i,\cdot)}(\cdot)$ is simply written as $\mathsf{Pr}_{\mathsf{Enc}_i}(\cdot)$.

**Differential characteristic**. An *s-round differential characteristic* is a vector $(\Delta X^0, \cdots, \Delta X^s)$ of differences, where

- $\Delta X^0$ a nonzero input difference to the cipher $\mathsf{Enc} : \mathbb{B}^{\mathscr{k}} \times \mathbb{B}^{\mathscr{b}} \longrightarrow \mathbb{B}^{\mathscr{b}}$,
- for $1 \leq i \leq s$, $(\Delta X^{i-1}, \Delta X^i)$ is a differential of the $i$-th round $\mathsf{Enc}_i(K^i, \cdot)$.

*Definition 2.3.* The *differential characteristic probability* $\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\Delta X^0, \cdots, \Delta X^s)$ of an $s$-round differential characteristic $(\Delta X^0, \cdots, \Delta X^s)$ is the proportion of inputs $X^0 \in \mathbb{B}^{\mathscr{b}}$ to the cipher $\mathsf{Enc}$ such that the output difference of the $i$-th round $\mathsf{Enc}_i$ is $\Delta X^i$ for every $1 \leq i \leq s$ in the two executions of $\mathsf{Enc}$ under the two inputs $(K, X^0)$ and $(K, X^0 \oplus \Delta X^0)$, namely,

$$\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\Delta X^0, \cdots, \Delta X^s) = \frac{\left|\left\{X^0 \in \mathbb{B}^{\mathscr{b}} \mid \forall i.1 \leq i \leq s.\mathsf{Enc}_{\leq i}(X_0) \oplus \mathsf{Enc}_{\leq i}(X_0 \oplus \Delta X^0) = \Delta X^i\right\}\right|}{2^{\mathscr{b}}}$$

where $\mathsf{Enc}_{\leq i} := \mathsf{Enc}_i(K^i, \cdot) \circ \cdots \circ \mathsf{Enc}_1(K^1, \cdot)$ for $1 \leq i \leq s$.

Recall that we assumed that the output difference $\mathsf{Enc}_i(K^i, X^{i-1}) \oplus \mathsf{Enc}_i(K^i, X^{i-1} \oplus \Delta X^{i-1})$ of the $i$-th round is independent upon the subkey $K^i$ for any fixed input $X^{i-1}$ and input difference $\Delta X^{i-1}$. Thus, the probability $\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\Delta X^0, \cdots, \Delta X^s)$ does not depend on the $s$-round subkey $K^s$, but depends on the other subkeys $K^1, \cdots, K^{s-1}$. It is known that $\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\Delta X^0, \cdots, \Delta X^s)$ is upper bound by $\prod_{i=1}^{s} \mathsf{Pr}_{\mathsf{Enc}_i}(\Delta X^{i-1}, \Delta X^i)$ [Heys and Tavares 1996], and they are the same if $\mathsf{Enc}$ is a Markov cipher and its round subkeys are independent [Lai et al. 1991].

In a resistant block cipher, for any fixed key $K$, the probability $\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\Delta X^0, \cdots, \Delta X^s)$ should be small enough for any differential characteristic $(\Delta X^0, \cdots, \Delta X^s)$ if the input $X^0$ is sampled uniformly. However, in practice, $\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\Delta X^0, \cdots, \Delta X^s)$ may be higher than $2^{-\mathscr{b}}$ based on which an attacker can efficiently recover the subkey $K^{s+1}$. The differential characteristic $(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^s)$ is said to be *optimal* if it attains the greatest expected differential characteristic probability among all the $s$-round differential characteristics. We remark that optimal differential characteristics are commonly assumed to be identical for different keys $K$ during the attack, as the actual key is unknown to the adversary before attacking. In the sequel, for simplicity, $\mathsf{Pr}_{\mathsf{Enc}(K,\cdot)}(\cdot)$ is written as $\mathsf{Pr}_{\mathsf{Enc}}(\cdot)$ and $\mathsf{Pr}_{\mathsf{Enc}}(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^s)$ is refer to as the *maximum expected differential characteristic probability (*$\mathsf{MaxEDCP}$*).* The key recovering procedure is given in [Sun et al. 2023, Section B], where the number of plaintexts required to infer the $(s+1)$-round subkey $K^{s+1}$ is proportional to $\frac{1}{\mathsf{Pr}_{\mathsf{Enc}}(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^s)}$ [Heys 2002b], i.e., the reciprocal of the $\mathsf{MaxEDCP}$ of $s$-round differential characteristics. As a result, if one could show that an upper bound of $\mathsf{Pr}_{\mathsf{Enc}}(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^s)$ is no greater than $2^{-\mathscr{b}}$, one would conclude the resistance of the block cipher against such differential cryptanalysis.

### 2.3 Active S-box, Differential Distribution Table and Branch Number

We introduce some notions of active S-box, differential distribution table and branch number which will be used in our approach.

**Active S-boxes**. The number of active S-boxes can be used to upper bound the $\mathsf{MaxEDCP}$. Assume S-boxes are distinct in the cipher $\mathsf{Enc}$.

*Definition 2.4.* [Heys 2002b] Given a key $K$, an input $X^0$ and an input difference $\Delta X^0$ to the cipher $\mathsf{Enc}$, an S-box $\mathcal{S}$ is *active* if the two inputs to $\mathcal{S}$ are distinct in the two executions of the cipher $\mathsf{Enc}$ under two inputs $(K, X^0)$ and $(K, X^0 \oplus \Delta X^0)$, otherwise it is *inactive*.

We denote by $\mathcal{N}_{\mathtt{diff}}$ the minimum number of the active S-boxes in all the possible pairs of executions. The probability of optimal $s$-round differential characteristics is bounded from above by $p^{\mathcal{N}_{\mathtt{diff}}}$ [Heys 2002b], where $p$ denotes the maximum probability $\mathsf{Pr}_{\mathcal{S}}(\Delta X, \Delta Y)$ among all the nonzero differentials $(\Delta X, \Delta Y)$ for any S-box $\mathcal{S}$ which is active in the $s$-round differential characteristics.

**Differential distribution table**. A differential distribution table (DDT) is a data structure to represent the distribution $\mathsf{Pr}_f$ of a function $f$ for all possible differentials. It also explicitly expresses the dependency (i.e., feasibility) between input and output differences of the function $f$.

*Definition 2.5.* Given a function $f : \mathbb{B}^{n_1} \times \cdots \times \mathbb{B}^{n_i} \longrightarrow \mathbb{B}^{m_1} \times \cdots \times \mathbb{B}^{m_j}$, its DDT $\mathcal{D}_f$ is table such that for every vector of input differences $(\Delta X^1, \cdots, \Delta X^i)$ and every vector of output differences $(\Delta Y^1, \cdots, \Delta Y^j)$, the entry $\mathcal{D}_f(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ gives the number of vectors of inputs $(X^1, \cdots, X^i) \in \mathbb{B}^{n_1} \times \cdots \times \mathbb{B}^{n_i}$ such that

$$f(X^1, \cdots, X^i) \oplus f(X^1 \oplus \Delta X^1, \cdots, X^i \oplus \Delta X^i) = (\Delta Y^1, \cdots, \Delta Y^j).$$

The probability $\mathsf{Pr}_f(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ can be deduced from the DDT $\mathcal{D}_f$:

$$\mathsf{Pr}_f(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j) = \frac{\mathcal{D}_f(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)}{2^{n_1 + \cdots + n_i}}.$$

We say the the vector of input and output differences $(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ is *feasible* for $f$, if the probability $\mathsf{Pr}_f(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ is nonzero, otherwise it is *infeasible*. When the input space of the function $f$ is small, its DDT $\mathcal{D}_f$ can be computed by enumeration.

*Example 2.6.* Consider the AND operation $\wedge : \mathbb{B} \times \mathbb{B} \longrightarrow \mathbb{B}$. We have that

$$\Delta Y = (X^1 \wedge X^2) \oplus ((X^1 \oplus \Delta X^1) \wedge (X^2 \oplus \Delta X^2)).$$

The DDT $\mathcal{D}_\wedge$ is shown in Table 1, e.g., $\mathsf{Pr}_\wedge(0, 0, 0) = \frac{4}{4} = 1$, and $\mathsf{Pr}_\wedge(\Delta X^1, \Delta X^2, \Delta Y) = \frac{2}{4} = \frac{1}{2}$ if $\Delta X^1 = 1$ or/and $\Delta X^2 = 1$ for any fixed $\Delta Y$.

The DDT $\mathcal{D}_\vee$ of the OR operation $\vee : \mathbb{B} \times \mathbb{B} \longrightarrow \mathbb{B}$ is the same as $\mathcal{D}_\wedge$. We can observe that the input difference $(0, 0)$ cannot lead to the output difference $0$ for both the AND and OR operations. □

**Branch number**. The (differential) branch number of a function is also used to characterize the dependency between input and output differences of the function [Daemen and Rijmen 1999].

Table 1. The DDT ($\mathcal{D}_\wedge$, $\mathcal{D}_\vee$) for $\wedge/\vee : \mathbb{B} \times \mathbb{B} \longrightarrow \mathbb{B}$.

| $(\Delta X^1, \Delta X^2)$ | (0,0) | (0,1) | (1,0) | (1,1) |
|---|---|---|---|---|
| $\Delta Y = 0$ | 4 | 2 | 2 | 2 |
| $\Delta Y = 1$ | 0 | 2 | 2 | 2 |

Table 2. Branch numbers of +, −, ∧, ∨, ⊕.

| | + | − | ∧ | ∨ | ⊕ |
|---|---|---|---|---|---|
| $\mathcal{B}^{\min}_{\text{ww}}$ and $\mathcal{B}^{\min}_{\text{bw}}$ | 2 | 2 | 1 | 1 | 2 |
| $\mathcal{B}^{\max}_{\text{ww}}$ | 3 | 3 | 3 | 3 | 3 |
| $\mathcal{B}^{\max}_{\text{bw}}$ | 3n-1 | 3n-1 | 3n | 3n | 2n |

*Definition 2.7.* Given a function $f : \mathbb{B}^{n_1} \times \cdots \times \mathbb{B}^{n_i} \longrightarrow \mathbb{B}^{m_1} \times \cdots \times \mathbb{B}^{m_j}$, its *minimum (resp. maximum) word-wise branch number* $\mathcal{B}^{\min}_{\text{ww}}(f)$ (resp. $\mathcal{B}^{\max}_{\text{ww}}(f)$) is defined as

$$\begin{aligned}\mathcal{B}^{\max}_{\text{ww}}(f) &= \max\left\{\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j) \mid \forall 1 \le \ell \le n.\ X^\ell, \Delta X^\ell \in \mathbb{B}^{n_\ell}.\mathsf{BNCond}(f)\right\}\\ \mathcal{B}^{\min}_{\text{ww}}(f) &= \min\left\{\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j) \mid \forall 1 \le \ell \le n.\ X^\ell, \Delta X^\ell \in \mathbb{B}^{n_\ell}.\mathsf{BNCond}(f)\right\}\end{aligned}$$

where $\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ counts the number of nonzero entries in the vector $(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ and the branch-number condition $\mathsf{BNCond}(f)$ is:

$$\mathsf{BNCond}(f) = \begin{pmatrix} (\Delta X^1 \neq 0 \vee \cdots \vee \Delta X^i \neq 0) \wedge (Y^1, \cdots, Y^j) = f(X^1, \cdots, X^i) \\ \wedge (Y^1 \oplus \Delta Y^1, \cdots, Y^j \oplus \Delta Y^j) = f(X^1 \oplus \Delta X^1, \cdots, X^i \oplus \Delta X^i) \end{pmatrix}.$$

Likewise, the *minimum (resp. maximum) bit-wise branch number* $\mathcal{B}^{\min}_{\text{bw}}(f)$ (resp. $\mathcal{B}^{\max}_{\text{bw}}(f)$) of the function $f$ is defined except that $\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ counts the number of 1 bits in the bitstream $\Delta X^1 \| \cdots \| \Delta X^i \| \Delta Y^1 \| \cdots \| \Delta Y^j$, i.e., Hamming weight.

Intuitively, when the input differences of the function $f$ are not all 0 bits (i.e., $\Delta X^1 \neq 0 \vee \cdots \vee \Delta X^i \neq 0$), the number of nonzero entries in any input and output differences $(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ of the function $f$ ranges from $\mathcal{B}^{\min}_{\text{ww}}(f)$ to $\mathcal{B}^{\max}_{\text{ww}}(f)$, and the Hamming weight of any bitstream $\Delta X^1 \| \cdots \| \Delta X^i \| \Delta Y^1 \| \cdots \| \Delta Y^j$ ranges from $\mathcal{B}^{\min}_{\text{bw}}(f)$ to $\mathcal{B}^{\max}_{\text{bw}}(f)$.

*Example 2.8.* Consider the function $f_\oplus : \mathbb{B}^n \times \mathbb{B}^n \longrightarrow \mathbb{B}^n$ such that $f_\oplus(X^1, X^2) = X^1 \oplus X^2$. We have: $\mathcal{B}^{\min}_{\text{ww}}(f_\oplus) = \mathcal{B}^{\max}_{\text{ww}}(f_\oplus) = \mathcal{B}^{\min}_{\text{bw}}(f_\oplus) = 2$ and $\mathcal{B}^{\max}_{\text{bw}}(f_\oplus) = 2n$. From $\mathcal{B}^{\min}_{\text{ww}}(f_\oplus) = 2$, we can deduce that at least two of $(X^1, X^2, X^1 \oplus X^2)$ are nonzero if some of $(X^1, X^2)$ is nonzero.

Similarly, $\mathcal{B}^{\min}_{\text{ww}}(f_\odot), \mathcal{B}^{\max}_{\text{ww}}(f_\odot), \mathcal{B}^{\min}_{\text{bw}}(f_\odot)$ and $\mathcal{B}^{\max}_{\text{bw}}(f_\odot)$ for each bit-wise operation $\odot \in \{+, -, \wedge, \vee\}$ can be defined, whose values are given in Table 2, where for clarity, we denote by $\mathcal{B}^{\min}_{\text{ww},\odot}$, $\mathcal{B}^{\max}_{\text{ww},\odot}$, $\mathcal{B}^{\min}_{\text{bw},\odot}$ and $\mathcal{B}^{\max}_{\text{bw},\odot}$ the corresponding numbers of the bit-wise operation $\odot \in \{+, -, \wedge, \vee, \oplus\}$. Furthermore, for a given matrix $M$ and S-box $\mathcal{S}$, $\mathcal{B}^{\min}_{\text{ww},M}$, $\mathcal{B}^{\max}_{\text{ww},M}$, $\mathcal{B}^{\min}_{\text{bw},M}$, $\mathcal{B}^{\max}_{\text{bw},M}$, $\mathcal{B}^{\min}_{\text{ww},\mathcal{S}}$, $\mathcal{B}^{\max}_{\text{ww},\mathcal{S}}$, $\mathcal{B}^{\min}_{\text{bw},\mathcal{S}}$ and $\mathcal{B}^{\max}_{\text{bw},\mathcal{S}}$ are defined accordingly for the linear transformation $f_M(x) = M * x$ and substitution $f_{\mathcal{S}} = \mathcal{S}(x)$. □

## 3 THE DESIGN OF EASYBC

EasyBC is a high-level, statically-typed, C-like cryptography-specific language, designed for conveniently describing block ciphers but without complicating the subsequent automated security analysis, so that cryptographers can quickly implement and analyze a block cipher especially during the design phase.

### 3.1 Syntax

The syntax of EasyBC is given in Figure 1.

**Boxes**. EasyBC features one standard array type and four cryptography-specific array types decorated by `sbox`, `pbox`, `pbox`$_\mathtt{m}$ and `ffm`, respectively, that are commonly used for implementing block ciphers. The cryptography-specific arrays are global and immutable, thus called boxes in this paper. In particular, `sbox` defines an array that acts as a lookup-table based S-box for transforming $m$ input bits to $n$ output bits; `pbox` defines an array for performing permutation; `pbox`$_\mathtt{m}$ defines a matrix for linear transformation via matrix-vector product. (Note that `pbox` can be implemented via `pbox`$_\mathtt{m}$, we provide both for convenience.) `ffm` describes the finite-field multiplication (⊗) which may vary in block ciphers, thus should be defined by users. An `ffm` box is required *only* for performing

$$\begin{array}{rl}
\textsc{Operation} & \odot \in \{+, -, \wedge, \vee, \oplus\} \\
\textsc{Operation}_2 & \star \in \{, -, *, /, \%\} \\
\textsc{Width} & s ::= 1 \mid 4 \mid 6 \mid 8 \mid 16 \mid \cdots \\
\textsc{Constant} & n ::= 0 \mid 1 \mid 2 \mid 3 \mid \cdots \\
\textsc{Base Type} & \tau ::= \texttt{uint}s \mid \texttt{uint}s[n] \mid \texttt{uint} \\
\textsc{Position} & \xi ::= n \mid x \mid \xi \star \xi \\
\textsc{Expression} & e ::= n \mid x \mid e_1 \odot e_2 \mid\sim e \mid M * e \\
 & \quad \mid x\langle\cdot e\cdot\rangle \mid x\langle e\rangle \mid \texttt{View}(e, \xi_1, \xi_2) \\
 & \quad \mid \texttt{touint}(e_0, \cdots, e_{s-1}) \mid \texttt{touint}(e) \mid e \lll \xi \mid e \ggg \xi \mid e[\xi]
\end{array}$$

$$\begin{array}{rll}
\text{STMT} & S ::= \tau\, x & \text{Declaration} \\
 & \mid x = e; & \text{Assignment} \\
 & \mid x[\xi] = e; & \text{Array put} \\
 & \mid x = f(e_1, \cdots, e_n); & \text{Function call} \\
 & \mid \tau\, x = e; & \text{Decl-Init} \\
 & \mid \texttt{for}\ (x\ \texttt{from}\ n_1\ \texttt{to}\ n_2)\{S^+\} & \text{Range-for}
\end{array}$$

$$\begin{array}{rl}
\textsc{Round\_fn} & \text{rnd} ::= \texttt{r\_fn}\ \texttt{uint}s[n]\ \ f(\texttt{uint}\ r, \texttt{uint}s[n_1]\ sk, \texttt{uint}s[n]\ p)\{S^+\ \texttt{return}\ y;\} \\
\textsc{Sbox\_fn} & \text{sbox} ::= \texttt{s\_fn}\ \texttt{uint}s_1\ \ f(\texttt{uint}s_2\ x)\{S^+\ \texttt{return}\ y;\} \\
\textsc{Box} & \text{box} ::= \texttt{uint}s[n]\ x = \{n_0, \cdots, n_m\} \\
 & \quad \mid \texttt{sbox}\ \texttt{uint}s[n]\ x = \{n_0, \cdots, n_m\} \mid \texttt{pbox}\ \texttt{uint}[n]\ x = \{n_0, \cdots, n_m\} \\
 & \quad \mid \texttt{pbox}_{\texttt{m}}\ \texttt{uint}s[n][n]\ M = \{\ \{n_{0,0}, \cdots, n_{0,m}\}, \ \cdots, \ \{n_{m,0}, \cdots, n_{m,m}\}\ \} \\
 & \quad \mid \texttt{ffm}\ \texttt{uint}s[n][n]\ M = \{\ \{n_{0,0}, \cdots, n_{0,m}\}, \ \cdots, \ \{n_{t,0}, \cdots, n_{t,m}\}\ \} \\
\textsc{Def} & \text{defs} ::= \text{box} \mid \text{sbox} \mid \text{rnd} \mid \text{defs defs} \\
\textsc{Program} & \text{P} ::= \texttt{@cipher}\ \text{name defs}\ \texttt{fn}\ \texttt{uint}s[n]\ f(\texttt{uint}s[n_1]\ k,\ \texttt{uint}s[n]\ p)\{S^+\ \texttt{return}\ y;\}
\end{array}$$

Fig. 1. Syntax of EasyBC.

linear transformations using $\texttt{pbox}_{\texttt{m}}$, i.e., evaluating expressions of the form $M * x$, thus cannot be explicitly involved in any statements. In this work, we assume a finite field of characteristic 2, so the finite-field addition is the bit-wise XOR ($\oplus$).

**Positions.** Position $\xi$ is used to express array indices for array get ($e[\xi]$), array slice ($\texttt{View}(e, \xi_1, \xi_2)$), array put ($x[\xi] = e$) and array left/right-rotation ($e \lll \xi$ and $e \ggg \xi$) via the common operations $\{+, -, *, /, \%\}$, whose values can be statically determined after preprocessing (i.e., independent of inputs). After preprocessing, all the positions $\xi$ will be constants.

**Expressions.** Expressions are defined as usual, including modular addition (+), modular substitution (−), bit-wise AND ($\wedge$), bit-wise OR ($\vee$), bit-wise NOT ($\sim$) and bit-wise XOR ($\oplus$), as well as common cryptography-specific operations.

$M * e$ is a matrix-vector product using finite-field multiplication ($\otimes$) and addition ($\oplus$), where the matrix $M$ must be defined as an array of type $\texttt{pbox}_{\texttt{m}}\ \texttt{uint}s[n][n]$ and the vector $e$ should be an array of type $\texttt{uint}s[n]$. $x\langle\cdot e\cdot\rangle$ is provided for performing permutation, where $x$ is a P-box. Similarly, $x\langle e\rangle$ is provided for performing substitution, where $x$ is an S-box. $\texttt{View}(e, \xi_1, \xi_2)$ is a slice of the array $e$ starting at the index $\xi_1$ and ending at the index $\xi_2$ (inclusive), e.g., $\texttt{View}((a, b, c, d), 1, 3)$ is $(b, c, d)$. $\texttt{touint}(e_0, \cdots, e_{s-1})$ transforms the $s$-bitstream $(e_0, \cdots, e_{s-1})$ into an $s$-bit unsigned integer $x$ such that $\text{bin}(x) = (e_0, \cdots, e_{s-1})$, e.g., $\texttt{touint}(1, 1, 0, 1)$ is the 4-bit unsigned integer 13. $\texttt{touint}(e)$ is the same as $\texttt{touint}(e[0], \cdots, e[n-1])$ for the array $e$ of type $\texttt{uint}1[n]$. An array $e$ can be left (resp. right) rotated $\xi$ positions via $e \lll \xi$ (resp. $e \ggg \xi$), which can be seen special length-preserving permutations. $e[\xi]$ is array get, the same as $\texttt{View}(e, \xi, \xi)$.

**Statements.** Statements in EasyBC can be declarations, assignments, array puts, returns and round function calls. Note that EasyBC does not support branching statements (e.g., if-then-else), because current EasyBC suffices to implement both encryption and decryption processes of block ciphers while branching statements will make modeling complicated when analyzing security. We will evaluate the expressive capability of EasyBC in Section 8.1.

Statement $\tau\ x = e$ is a syntactic sugar of $\tau\ x; x = e$. Statement $\texttt{for}\ (x\ \texttt{from}\ n_1\ \texttt{to}\ n_2)\{S\}$ is a range-for loop. Note that the range $[n_1, n_2]$ is limited to constants, which suffices to express block ciphers. The range variable $x$ should be typed as $\texttt{uint}$, thus could be used in computing indices $\xi$.

**Functions**. EasyBC has three types of functions decorated by `r_fn`, `s_fn` or `fn`. A function decorated by `r_fn` is a round function, whose formal parameters are fixed to be the round number $r$, subkey $sk$, and text $txt$. Note that the subkey $sk$ and input text $txt$ should have the same element type `uint`$s$ but may differ in number of elements, and the input text $txt$ and output of a round function should have the same type `uint`$s[n]$. An `s_fn` function is an alternative way to perform substitution instead of using arrays. Small S-boxes (e.g., 4-bit S-box in PRESENT [Bogdanov et al. 2007]) can be easily expressed as arrays, which can facilitate the follow-up security analysis. However, it would be infeasible to express large S-boxes as arrays (e.g., 64-bit S-box [Beierle et al. 2020]), for which `s_fn` functions can be used. An `fn` function defines the encryption (resp. decryption) process of a block cipher. Its parameters include the key $k$ and plaintext (resp. ciphertext) $txt$. The function body comprises declarations and round function calls for computing the ciphertext (resp. plaintext).

**Programs**. A program $P$ in EasyBC consists of a cipher name (decorated by @`cipher`), definitions of global boxes and functions, and a definition of an `fn` function describing the cipher.

In this work, following [Mouha et al. 2011; Wu and Wang 2012; Zhang et al. 2018; Zhou et al. 2019], we consider *single-key* differential cryptanalysis, namely, the key is fixed and has no difference in any pair of executions, thus key schedule algorithms are omitted in EasyBC programs. Nevertheless, EasyBC can be easily extended to *related-key* differential cryptanalysis [Biryukov and Nikolić 2010] where the key may differ in some pairs of executions, by introducing a key schedule function. We leave this as interesting future work.

**Core language**. The (full) language of EasyBC is designed for conveniently describing block ciphers with rich but redundant constructs. To ease the automation of the subsequent security analysis, we have identified a subset of EasyBC as the core language (cf. Figure 1), i.e., in the places highlighted in yellow, positions $\xi$ and expressions $e$ are limited to constants and variables, respectively; in the places highlighted in grey constructs will not be present (i.e., they are to be eliminated by preprocessing the program written in the full language). A program in the core language is obtained by performing loop unrolling, constant-folding, constant propagation and dead-code elimination, on which type checking and security analysis are performed.

*Example 3.1.* Figure 2 shows a snippet of the 64-bit block cipher PRESENT in EasyBC. The array `s` is an S-box for substitution, the array `p` is the P-box for permutation. The round function `f1` is invoked during the 1-st to the 31-st round. Given the subkey `sk` and text `t`, `f1`($i$,`sk`,`t`) for $1 \leq i \leq 31$ produces the output `rtn` of $i$-th round. In detail, the input `t` is XORed with the subkey `sk` and results in the array `nt`, then the second range-for loop slices `nt` into 16 arrays via calling `View`, each of which is substituted via the S-box `s`, resulting in the array `s_out`. The array `s_out` is processed by applying the array `p` to perform permutation. Finally, the result `rtn` returned.

**Remarks on the design choice of** EasyBC. In EasyBC, we introduce high-level constructs `View`, `touint` and permutations (i.e., $x\langle\cdot e\cdot\rangle$ and $M * e$) to ease the implementation of block ciphers. The inputs and outputs of round functions are fixed-size blocks which can be implemented by arrays (e.g., `t` and `rtn` in Figure 2). Typically, a block is to be split into small ones on which a look-up table based S-box (e.g., array in EasyBC) of suitable size is applied (e.g., 4-bit S-box `s` in Figure 2). The outputs of S-boxes will be juxtaposed to form a large block on which permutations are performed via either matrix-vector product or P-boxes (e.g., `p1`$\langle\cdot$`s_out`$\cdot\rangle$ in Figure 2). On the other hand, to ease the automation of the subsequent security analysis, most indices are limited to positions $\xi$ whose values can be statically determined (e.g., $e[\xi]$ and `View`$(e, \xi_1, \xi_2)$), and thus become constants after preprocessing. The loop-up table based S-boxes (i.e., $x\langle e\rangle$) are an exception as they require a complicated modeling method. It remains open how to handle generic array access with variable indices although currently there appears no such need to specify block ciphers.

```
@cipher PRESENT
sbox uint4[16] s = {12,5,6,11,9,0,10,13,3,14,15,8,4,7,1,2};
pbox uint[64] p = {0,16,32,48,1,17,33,...15,31,47,63};
r_fn uint1[64] f1(uint r, uint1[64] sk, uint1[64] t){
    uint1[64] nt;
    for(i from 0 to 63){ nt[i] = t[i] ^ sk[i]; }
    uint1[64] s_out;
    for(i from 0 to 15) {
        uint1[4] temp = View(nt, i*4, i*4+3);
        uint4 sbox_in = touint(temp[0], temp[1], temp[2], temp[3]);
        uint4 sbox_out = s1⟨sbox_in⟩      # substitution via S-box
        s_out[i*4]=sbox_out[0];    s_out[i*4+1]=sbox_out[1];
        s_out[i*4+2]=sbox_out[2];  s_out[i*4+3]=sbox_out[3];
    }
    uint1[64] rtn = p1⟨·s_out·⟩;           # permutation via P-box
    return rtn; }
fn uint1[64] enc(uint1[2048] key, uint1[64] plaintext){
    uint1[64] text= plaintext;
    for(i from 1 to 31){  text=f1(i,View(key,(i-1)*64,i*64-1),text);  }
    ....          # execute the last rounds
    return text; }
```

Fig. 2. Code snippet of the 64-bit block cipher PRESENT in EasyBC.

### 3.2 Operational Semantics

Let $\mathbb{X}$ denote a set of variables. An *(evaluation) context* $\sigma : \mathbb{X} \longrightarrow \bigcup_{i\geq 1} \mathbb{N}^i$ is a mapping from variables to values, where a value can be a (fixed-width) non-negative integer or an array. Let $\sigma[x \mapsto v]$ be the context such that $\sigma[x \mapsto v](y) = v$ if $x = y$, otherwise $\sigma[x \mapsto v](y) = \sigma(y)$.

The evaluation judgement is in the form of

$$\sigma \models e : v$$

meaning that the expression $e$ evaluates to the value $v$ under the evaluation context $\sigma$.

The evaluation rules are given in Figure 3 (top-part), most of which are standard. Rule (S-Box) states that $x\langle y\rangle$ is a substitution, namely, the entry of the S-box $\sigma(x)$ at the index $\sigma(y)$. Rule (P-Box$_1$) states that $M * x$ is the matrix-vector product of the matrix $M$ and the array $\sigma(x)$. Rule (P-Box$_2$) states that $x\langle\cdot y\cdot\rangle$ is a permutation of the array $\sigma(y)$ according to the indices given by the P-box $\sigma(x) = (j_0, \cdots, j_{n-1})$, where its entry at the index $i$ is the entry of the array $\sigma(y)$ at the index $j_i$.

The operational semantics of statements is defined as transition rules of the form

$$(\sigma, S) \Rightarrow \sigma'$$

meaning that the execution of the statement $S$ from the state $\sigma$ results in the state $\sigma'$. For a sequence of statements $S_1; S_2; \cdots S_n;$, $(\sigma_0, S_1; S_2; \cdots S_n;) \Rightarrow^+ \sigma_n$ denotes the transitive transition of $\Rightarrow$, i.e., $(\sigma_0, S_1) \Rightarrow \sigma_1$, $(\sigma_1, S_2) \Rightarrow \sigma_2, \cdots, (\sigma_{n-1}, S_n) \Rightarrow \sigma_n$. The transition rules of EasyBC are listed in Figure 3 (bottom-part), which are standard. We denote by $\sigma_0 S_1 \sigma_1 S_2 \sigma_2 \cdots S_n \sigma_n$ the execution of the program $P$ starting from the state $\sigma_0$ and ending at the state $\sigma_n$, and $(\sigma_{i-1}, S_i) \Rightarrow \sigma_i$ for $1 \leq i \leq n$.

For an execution of an $r$-IBC, we have

(1) round functions can only be invoked in the fn function,
(2) the first arguments in the invoked round functions are the round numbers $1, 2, 3, \cdots, r$, and
(3) the input and output of $i$-th round are the third argument and return value of the invoked round function whose first argument is $i$.

$$\frac{}{\sigma \models n : n}\ \text{(Const)} \qquad \frac{\sigma(x) = v}{\sigma \models x : v}\ \text{(Var)} \qquad \frac{\sigma(x) = v \quad v' = \sim v}{\sigma \models \sim x : v'}\ \text{(Not)}$$

$$\frac{\sigma(x_1) = n_1 \quad \sigma(x_2) = n_2 \quad n = n_1 \odot n_2}{\sigma \models x_1 \odot x_2 : n}\ \text{(Op)} \qquad \frac{\sigma(x) = (v_0, \cdots, v_{n-1}) \quad \sigma(y) = i}{\sigma \models x\langle y \rangle : v_i}\ \text{(S-Box)}$$

$$\frac{\sigma(x) = (v_0, \cdots, v_{n-1}), \quad \vec{v} = \left(\bigoplus_{j=0}^{n-1}(M_{0,j} \otimes v_j), \cdots, \bigoplus_{j=0}^{n-1}(M_{n-1,j} \otimes v_j)\right)}{\sigma \models M * x : \vec{v}}\ \text{(P-Box}_1\text{)}$$

$$\frac{\sigma(x) = (j_0, \cdots, j_{n-1}) \quad \sigma(y) = (v_0, \cdots, v_{n-1}) \quad \vec{v} = (v_{j_0}, \cdots, v_{j_{n-1}})}{\sigma \models x\langle \cdot y \cdot \rangle : \vec{v}}\ \text{(P-Box}_2\text{)}$$

$$\frac{\sigma(x) = (v_0, \cdots, v_{n-1}) \quad \vec{v} = (v_i, \cdots, v_j)}{\sigma \models \mathtt{View}(x, i, j) : \vec{v}}\ \text{(View)} \qquad \frac{\mathrm{bin}(n) = (b_0, \cdots, b_{-1})}{\sigma \models \mathtt{touint}(b_0, \cdots, b_{-1}) : n}\ \text{(ToUint)}$$

$$\frac{}{(\sigma, \tau\ x) \Rightarrow \sigma}\ \text{(Decl)} \qquad \frac{\sigma \models e : v \quad \sigma' = \sigma[x \mapsto v]}{(\sigma, x = e) \Rightarrow \sigma'}\ \text{(Ass)}$$

$$\frac{\sigma(x) = (v_0, \cdots, v_{n-1}) \quad \sigma(y) = v \quad \sigma' = \sigma[x \mapsto (v_0, \cdots, v_{i-1}, v, v_{i+1}, \cdots, v_{n-1})]}{(\sigma, x[i] = y) \Rightarrow \sigma'}\ \text{(Arr-Put)}$$

$$\frac{\begin{array}{c}\tau_0\ \ f(\tau_1\ x_1, \cdots, \tau_m\ x_m)\{S^+\ \mathtt{return}\ y;\} \quad v_1 = \sigma(y_1), \cdots, v_m = \sigma(y_m) \\ \sigma_{\mathtt{in}} = \sigma[x_1 \mapsto v_1] \cdots [x_m \mapsto v_m] \quad (\sigma_{\mathtt{in}}, S^+) \Rightarrow^+ \sigma_{\mathtt{out}} \quad \sigma_{\mathtt{out}}(y) = v \quad \sigma_{\mathtt{ret}} = \sigma[x \mapsto v]\end{array}}{(\sigma, x = f(y_1, \cdots, y_m)) \Rightarrow \sigma_{\mathtt{in}} \qquad (\sigma_{\mathtt{out}}, \mathtt{return}\ y) \Rightarrow \sigma_{\mathtt{ret}}}\ \text{(Call-Ret)}$$

Fig. 3. The operational semantics of core EasyBC, where $\odot \in \{+, -, \oplus, \wedge, \vee\}$.

### 3.3 Type System of Core EasyBC

The type system of core EasyBC is designed to disallow certain kinds of illegal programs and provide type information for security analysis.

EasyBC supports the following types, i.e.,

$$\beta ::= \ \mathtt{uint}s \mid \mathtt{uint} \mid \mathtt{uint}s[n] \mid \mathtt{sbox}\ \mathtt{uint}s[n] \mid \mathtt{pbox}\ \mathtt{uint}[n] \mid \mathtt{pbox}_{\mathtt{m}}\ \mathtt{uint}s[n][n].$$

Here, $\mathtt{uint}s$ is for $s$-bit unsigned integers, $\mathtt{uint}s[n]$ is for vectors (or arrays) of $s$-bit unsigned integers, $\mathtt{uint}$ is for unsigned integers, and $\mathtt{uint}[n]$ is for vectors (or arrays) of unsigned integers. Note that $\mathtt{uint}s$ is identical to $\mathtt{uint}1[s]$ and $\mathtt{uint}s[1]$. $\mathtt{uint}$ and $\mathtt{uint}[n]$ are used only when the variables under typing are independent of inputs.

**Typing expressions**. The typing judgement is of the form of

$$T_g, T_l \vdash e : \beta,$$

where $(T_g, T_l)$ is a typing context, $e$ is an expression under typing, and $\beta$ is a type. The global environment $T_g$ is a mapping from global variables to their types and from function names to function signatures $(\beta_0, \cdots, \beta_n)$ where $\beta_0$ is the return type and $\beta_1, \cdots, \beta_n$ are the types of the formal parameters. The local environment $T_l$ is a mapping from local variables to their types. The typing judgement $T_g, T_l \vdash e : \beta$ is valid if $e$ has type $\beta$ under the typing context $(T_g, T_l)$.

Figure 4 (top-part) gives typing rules for expressions. Rule (T-Uints) express that a non-negative integer $n$ can be typed as $\mathtt{uint}s$ if $n \leq 2^s - 1$. Rules (T-Var) and (T-Not) are defined as usual. Rule (T-Op) ensures that the two operands and result of the operation $\odot \in \{+, -, \oplus, \wedge, \vee\}$ have the same type $\mathtt{uint}s$. Rule (T-PBox$_1$) ensures that the array $x$ has suitable type w.r.t. the type of the matrix $M$ for matrix-vector product. Rule (T-PBox$_2$) requires that the elements in the array $x\langle \cdot y \cdot \rangle$ and the operand $y$ have the same type, as the P-box $x$ only specifies the element order for the permutation. Rule (T-SBox) requires that the S-box has a sufficient number of elements (i.e., $n \geq 2^{s_2}$) and the

$$\frac{0 \le n \le 2^s - 1}{T_g, T_l \vdash n : \mathtt{uint}s}\ (\text{T-}\textsc{Uints}) \qquad \frac{\odot \in \{+, -, \oplus, \wedge, \vee\} \quad T_g, T_l \vdash x_i : \mathtt{uint}s \text{ for } i = 1, 2}{T_g, T_l \vdash x_1 \odot x_2 : \mathtt{uint}s}\ (\text{T-}\textsc{Op})$$

$$\frac{T = (x \in \mathbb{X}_f\ ?\ T_l : T_g)}{T_g, T_l \vdash x : T(x)}\ (\text{T-}\textsc{Var}) \qquad \frac{T_g(M) = \mathtt{pbox}_{\mathtt{m}}\ \mathtt{uint}s[n][n] \quad T_g, T_l \vdash x : \mathtt{uint}s[n]}{T_g, T_l \vdash M * x : \mathtt{uint}s[n]}\ (\text{T-}\textsc{Pbox}_1)$$

$$\frac{T_g, T_l \vdash e : \tau}{T_g, T_l \vdash \sim e : \tau}\ (\text{T-}\textsc{Not}) \qquad \frac{T_g(x) = \mathtt{pbox}\ \mathtt{uint}[n_1] \quad T_g, T_l \vdash y : \mathtt{uint}s[n_2]}{T_g, T_l \vdash x\langle\cdot y \cdot\rangle : \mathtt{uint}s[n_1]}\ (\text{T-}\textsc{Pbox}_2)$$

$$\frac{T_g(x) = \mathtt{sbox}\ \mathtt{uint}s_1[n] \quad T_g, T_l \vdash y : \mathtt{uint}s_2 \quad n \ge 2^{s_2}}{T_g, T_l \vdash x\langle y \rangle : \mathtt{uint}s_1}\ (\text{T-}\textsc{Sbox})$$

$$\frac{T_g, T_l \vdash x : \mathtt{uint}s[n] \quad 0 \le n_1 \le n_2 < n \quad n' = n_2 - n_1 + 1}{T_g, T_l \vdash \mathtt{View}(x, n_1, n_2) : \mathtt{uint}s[n']}\ (\text{T-}\textsc{View})$$

$$\frac{T_g, T_l \vdash x_i : \mathtt{uint}1 \text{ for } 0 \le i < s}{T_g, T_l \vdash \mathtt{touint}(x_0, ..., x_{s-1}) : \mathtt{uint}s}\ (\text{T-}\textsc{Touint})$$

$$\frac{T_l(x) = \tau}{T_g, T_l, f \vdash \tau\ x}\ (\text{T-}\textsc{Decl}) \qquad \frac{T_l(x) = \mathtt{uint}s[n] \quad T_g, T_l \vdash y : \mathtt{uint}s \quad 0 \le i < n}{T_g, T_l, f \vdash x[i] = y}\ (\text{T-}\textsc{Arr-Put})$$

$$\frac{\begin{array}{c} T_l(x) = \tau \\ T_g, T_l \vdash e : \tau \end{array}}{T_g, T_l, f \vdash\ x = e}\ (\text{T-}\textsc{Ass}) \qquad \frac{\begin{array}{c} T_g(f') = (\tau_0, \cdots, \tau_m) \\ T_g, T_l \vdash x : \tau_0 \quad T_g, T_l \vdash x_i : \tau_i \text{ for } 1 \le i \le m \end{array}}{T_g, T_l, f \vdash x = f'(x_1, \cdots, x_m)}\ (\text{T-}\textsc{Call})$$

$$\frac{\begin{array}{c} \ell \in \{\mathtt{fn}, \mathtt{r\_fn}, \mathtt{s\_fn}\} \qquad T_g(f) = (\tau_0, \cdots, \tau_m) \\ T_g, T_l[p_1 \mapsto \tau_1, \cdots, p_m \mapsto t_m, y \mapsto \tau_0], f \vdash S_i \text{ for } 1 \le i \le n \end{array}}{T_g, T_l \vdash \ell\ \tau_0\ \ f(\tau_1\ p_1, \cdots, \tau_m\ p_m)\{S_1; \cdots; S_n;\ \mathtt{return}\ y;\ \}}\ (\text{T-}\textsc{Fn-Def})$$

Fig. 4. The typing rules of core EasyBC.

result $x\langle y \rangle$ has the same type as the elements in the S-box $x$. Note that both S-boxes and P-boxes do not necessarily preserve the length which may occur, e.g., DES. Rule (T-View) requires that the indices $n_1$ and $n_2$ are within the bounds of the array $x$, moreover, the slice $\mathtt{View}(x, n_1, n_2)$ is an array with length $n_2 - n_1 + 1$ and its elements have the same type as the elements in the array $x$. Rule (T-Toint) requires that all the operands $x_i$ have type $\mathtt{uint}1$ and the result has type $\mathtt{uint}s$ where $s$ is the number of operands.

**Typing statements**. The typing judgement of a statement is in the form of

$$T_g, T_l, f \vdash S$$

where $(T_g, T_l)$ is a typing context, $S$ is the statement under typing in the function $f$. We write $T_g, T_l, f \vdash S$ is valid if $S$ is well-typed.

The typing rules are given in Figure 4 (middle-part). Rule (T-Decl) is defined as usual. Rule (T-Ass) requires that the type of the expression $e$ conforms to the declared type of the variable $x$. Rule (T-Arr-Put) requires that the index $i$ is within the bounds of the array $x$ and the operand $y$ has the same type as the elements in the array $x$. Rule (T-Call) requires that the types of actual arguments and return conform to the corresponding function signature $T_g(f')$.

**Typing programs**. Each program $P$ is typed by iteratively typing each function definition. The program is well-typed if all the function definitions are well-typed. The typing judgement of a function definition `fn_def` is in the form of

$$T_g, T_l \vdash \mathtt{fn\_def},$$

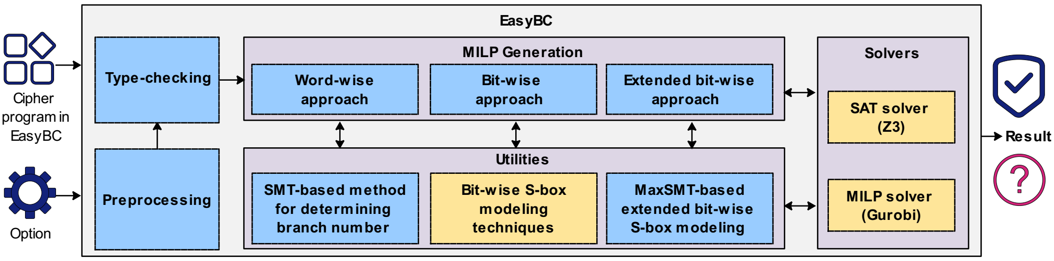


Fig. 5. Overview of our approach.

where $(T_g, T_l)$ is a typing context and fn_def is a function definition under typing. The typing judgement $T_g, T_l \vdash$ fn_def is valid if the function definition fn_def is well-typed. The typing rule (T-Fn-Def) is given in Figure 4 (bottom-part), which enforces the well-typed function body when the formal parameters and return have declared types.

### 3.4 Compilation

We have implemented an interpreter in C++ for EasyBC to test the operational semantics of programs, making sure that they are consistent with the execution of reference block ciphers. In particular, we compare the output of EasyBC programs with that of running the binary executable compiled from C/C++ programs by GNU C++ compiler (G++). For each cryptographic primitive, we randomly generate inputs and then run the EasyBC program (with our interpreter) and the binary executable. We record their output, as well as the execution time for analysis. The results are given in Section 8.1.

### 3.5 Overview of Analysis

Recall that we are interested in evaluating the resistance of block ciphers against differential cryptanalysis by bounding the MaxEDCP. A block cipher is considered to be resistant to differential cryptanalysis if MaxEDCP is no greater than $O(2^{\mathscr{b}})$ for the block size $\mathscr{b}$.

Figure 5 gives an overview of our approach. Given a program in full EasyBC together with an option for selecting a particular MILP generation approach and an S-box modeling technique, EasyBC computes an upper bound of the MaxEDCP. The result is conclusive if this upper bound is sufficient to show the resistance of the program. Our approach is not necessarily complete (e.g., in most cases we only compute an upper bound of MaxEDCP), so it may fail to prove the resistance of some programs, although this does not happen in our evaluation (cf. Section 8).

First, the input program is preprocessed to eliminate range-for loops and positional variables by performing loop unrolling, constant-folding, constant propagation and dead-code elimination. The final program will be in the core language of EasyBC. Hereafter, we assume that the given EasyBC program has been preprocessed.

Next, the program is type-checked to disallow certain kinds of illegal programs, e.g., the types of operands in expressions, formal parameters in function definitions and actual arguments in function calls are proper. It also provides type information for security analysis, in particular, the lengths of arrays, the type and the bit widths of array elements, which are used for MILP generation.

After type-checking, we reduce the problem of bounding the MaxEDCP to MILP. The key insight of the reduction is to characterize the dependency (i.e., feasibility) between input and output differences of each operation using integer linear (IL) constraints and bound the MaxEDCP by minimizing an objective function subject to the IL constraints. By utilizing an MILP solver (e.g., Gurobi [Gurobi Optimization 2018]), we can obtain an upper bound of the MaxEDCP. In practice, one may be only interested in proving the resistance against differential cryptanalysis. Hence we also verify whether the MaxEDCP is no greater than a given threshold, an affirmative answer to which would be sufficient to show that the given cipher is resistant against differential cryptanalysis. This strategy is very effective in practice, as few rounds are often sufficient to prove the resistance by leveraging the decomposition approach (cf. Proposition 5.3 and Proposition 7.3).

To this end, we present two different approaches for reducing to MILP. The first one is by determining the lower bound of the minimum number $\mathcal{N}_{\mathtt{diff}}$ of active S-boxes in either word-wise (Section 5) or bit-wise (Section 6) manner, because the MaxEDCP of $s$-round differential characteristics is bounded from above by $p^{\mathcal{N}_{\mathtt{diff}}}$ [Heys 2002b; Sun et al. 2014a], where $p$ denotes the maximum probability $\mathtt{Pr}_{\mathcal{S}}(\Delta X, \Delta Y)$ among all the nonzero differentials $(\Delta X, \Delta Y)$ for any active S-box. Intuitively, the word-wise one models the difference of an $s$-bitstream under two executions by only one Boolean variable, thus is less involved and produces fewer constraints, but is limited to certain block ciphers e.g., it cannot be directly applied to bit-oriented block ciphers such as PRESENT). In contrast, the bit-wise approach models the difference of each bit by one Boolean variable, thus is more fine-grained and has wider applicability. One may understand that the bit-wise approach implicitly bit-blasts the program and then generates MILP similar to the word-wise approach. The MILP generation in this approach requires the (maximum/minimum word-/bit-wise) branch numbers of some operations and representing S-boxes as IL constraints, for which we propose novel SMT-based methods to automatically determine branch numbers for each operation and implement some recent promising bit-wise S-box modeling techniques.

The first approach is efficient and often effective, but the obtained upper bound may not be sufficiently tight and is not applicable for some ciphers. We provide an extended bit-wise approach to *directly* bound the MaxEDCP (Section 7). In the extended bit-wise approach, the probabilities between input and output differences for each operation are further encoded into IL constraints using additional Boolean variables. This approach may be less efficient but is more accurate than the first approach. We also propose a novel Maximal Satisfiability Modulo Theories (MaxSMT) [Bjørner and Phan 2014] based extended bit-wise S-box modeling method which guarantees that the least number of Boolean variables is used for encoding differential probabilities of S-boxes.

## 4 UTILITIES

In this section, we present the three key utilities used in our MILP generation.

### 4.1 SMT-based Method for Determining Branch Numbers

The branch numbers of some operations are required in MILP generation. However, to our best knowledge, existing work usually relies on manual analysis. In this paper, we propose to determine branch numbers of a given operation/function, by reducing to the optimization problem modulo bit-vector theory, which can be solved by off-the-shelf optimizing SMT solver, e.g., Z3 [Bjørner et al. 2015].

Given a function $f : \mathbb{B}^{n_1} \times \cdots \times \mathbb{B}^{n_i} \longrightarrow \mathbb{B}^{m_1} \times \cdots \times \mathbb{B}^{m_j}$, to compute its minimum (resp. maximum) word-wise branch number $\mathcal{B}^{\min}_{\mathsf{ww}}(f)$ (resp. $\mathcal{B}^{\max}_{\mathsf{ww}}(f)$), by Definition 2.7, we express the condition $\mathtt{BNCond}(f)$ and the additional condition $d = \mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ as a quantifier-free SMT formula $\phi_f$ in the bit-vector theory, and use Z3 to minimize (resp. maximize) the variable $d$ subject to the SMT formula $\phi_f$. The optimized value of $d$ is $\mathcal{B}^{\min}_{\mathsf{ww}}(f)$ (resp. $\mathcal{B}^{\max}_{\mathsf{ww}}(f)$). An illustrating example is given in [Sun et al. 2023, Section C.1].

The minimum (resp. maximum) bit-wise branch number $\mathcal{B}^{\min}_{\mathsf{bw}}(f)$ (resp. $\mathcal{B}^{\max}_{\mathsf{bw}}(f)$) can be computed the same as above, except that $\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ counts the number of 1 bits in the bitstream $\Delta X^1 \| \cdots \| \Delta X^i \| \Delta Y^1 \| \cdots \| \Delta Y^j$.

Proposition 4.1. *The minimized (maximized) value of $d$ is*

- $\mathcal{B}_{\text{ww}}^{\min}(f)$ *(resp.* $\mathcal{B}_{\text{ww}}^{\max}(f)$*) when* $\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ *counts the number of nonzero entries in the vector* $(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$,
- $\mathcal{B}_{\text{bw}}^{\min}(f)$ *(resp.* $\mathcal{B}_{\text{bw}}^{\max}(f)$*) when* $\mathtt{cnt}(\Delta X^1, \cdots, \Delta X^i, \Delta Y^1, \cdots, \Delta Y^j)$ *counts the number of* 1 *bits in the bitstream* $\Delta X^1 \| \cdots \| \Delta X^i \| \Delta Y^1 \| \cdots \| \Delta Y^j$.

## 4.2 Bit-wise S-box Modeling

Consider an expression $\mathcal{S}\langle X\rangle$ where $\mathcal{S} : \mathbb{B}^n \to \mathbb{B}^m$ is a lookup-table based S-box. It is non-trivial to specify the bit-level dependency of differences between the input $X$ and the result of $\mathcal{S}\langle X\rangle$ (denoted by $Y$), as the S-box provides only input and output pairs. To resolve this issue, we first compute its DDT $\mathcal{D}_{\mathcal{S}} : \mathbb{B}^n \times \mathbb{B}^m \to \mathbb{N}$ from which IL constraints are generated to characterize the bit-level dependency of differences between $X$ and $Y$. Recall that for every differential $(\Delta X, \Delta Y) \in \mathbb{B}^n \times \mathbb{B}^m$, $\mathcal{D}_{\mathcal{S}}(\Delta X, \Delta Y)$ gives the number of inputs $X \in \mathbb{B}^n$ such that $\mathcal{S}(X) \oplus \mathcal{S}(X \oplus \Delta X) = \Delta Y$.

Let $\vec{b}$ be the input difference $\Delta X$ and $\vec{b}'$ be the output difference $\Delta Y$ of the S-box $\mathcal{S}$. We have

- $\sum_{i=0}^{n-1} \vec{b}_i = 0 \Rightarrow \sum_{i=0}^{m-1} \vec{b}'_i = 0$, i.e., no output difference if no input difference, namely, if two inputs to $\mathcal{S}$ are the same, then the two outputs are the same. This condition can be characterized by the IL constraint $m \cdot \sum_{i=0}^{n-1} \vec{b}_i \geq \sum_{i=0}^{m-1} \vec{b}'_i$, denoted by $\Psi_{\mathcal{S}}^1$.
- $\sum_{i=0}^{m-1} \vec{b}'_i = 0 \Rightarrow \sum_{i=0}^{n-1} \vec{b}_i = 0$ if $\mathcal{S}$ is injective, i.e., no input difference if no output difference, namely, if two outputs of $\mathcal{S}$ are the same, then the two inputs must be the same. This condition can be exactly characterized by the IL constraint $n \cdot \sum_{i=0}^{m-1} \vec{b}'_i \geq \sum_{i=0}^{n-1} \vec{b}_i$, denoted by $\Psi_{\mathcal{S}}^2$.
- $\mathcal{D}_{\mathcal{S}}(\vec{b}, \vec{b}') \neq 0$, namely, $(\vec{b}, \vec{b}')$ should be feasible for $\mathcal{S}$. We implement and compare the promising techniques [Abdelkhalek et al. 2017; Boura and Coggia 2020; Li and Sun 2022; Sasaki and Todo 2017; Sun et al. 2014b; Udovenko 2021] that can characterize $\mathcal{D}_{\mathcal{S}}(\vec{b}, \vec{b}') \neq 0$ by IL constraints. Hereafter, we denote by $\Psi_{\mathcal{S}}^3$ the set of IL constraints such that $(\vec{b}, \vec{b}')$ is a solution of $\Psi_{\mathcal{S}}^3$ iff $\mathcal{D}_{\mathcal{S}}(\vec{b}, \vec{b}') \neq 0$.

Proposition 4.2. *$(\vec{b}, \vec{b}')$ is feasible input and output differences of $\mathcal{S}$ iff $(\vec{b}, \vec{b}')$ is a solution of $\Psi_{\mathcal{S}}^3$.*

We denote by $\Psi_{\mathcal{S}}$ the set $\Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^2 \cup \Psi_{\mathcal{S}}^3$ if the S-box $\mathcal{S}$ is injective, otherwise $\Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^3$.

## 4.3 MaxSMT-based Extended Bit-wise S-box Modeling

The above bit-wise S-box modeling method is able to characterize all the feasible differentials $(\Delta X, \Delta Y) \in \mathbb{B}^n \times \mathbb{B}^m$ of the S-box $\mathcal{S}$, but the probability $\mathtt{Pr}_{\mathcal{S}}(\Delta X, \Delta Y) = \frac{\mathcal{D}_{\mathcal{S}}(\Delta X, \Delta Y)}{2^n}$ of differentials is not present in the IL constraints, so it is impossible to bound the MaxEDCP *directly*. We propose a novel MaxSMT-based method which guarantees that the least number of Boolean variables is used for encoding probabilities of differentials.

*Definition 4.3.* Given two sets of constraints $(\Phi_1, \Phi_2)$, the MaxSMT problem is to find a solution that satisfies all the constraints in $\Phi_1$ and maximizes the number of satisfied constraints in $\Phi_2$.

Let $V = \{v_1, \cdots, v_h\}$ be the set of nonzero probabilities $\mathtt{Pr}_{\mathcal{S}}(\Delta X, \Delta Y)$ for $(\Delta X, \Delta Y) \in \mathbb{B}^n \times \mathbb{B}^m$. We define the MaxSMT problem $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$ for the S-box $\mathcal{S}$, where

$$\Phi_1^{\mathcal{S}} = \{\textstyle\sum_{i=1}^{h} c_i \cdot p_{i,j} = -\log_2 v_j \mid 1 \leq j \leq h\} \text{ and } \Phi_2^{\mathcal{S}} = \{c_1 = 0, \cdots, c_h = 0\},$$

for $1 \leq i, j \leq h$, $c_i$ is a variable over real numbers and $p_{i,j}$ is a Boolean variable. Clearly, for every $1 \leq j \leq h$, $2^{-\sum_{i=1}^{h} c_i \cdot p_{i,j}}$ retains the probability $v_j$. Since the Boolean variable $p_{i,j}$ can be treated as a variable over real numbers by adding $p_{i,j} = 0 \vee p_{i,j} = 1$ to $\Phi_1^{\mathcal{S}}$, and $v_j$ (hence $\log_2 v_j$) is a constant for each $1 \leq j \leq h$, the problem $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$ is a MaxSMT problem (modulo the theory of real numbers).

A solution of the MaxSMT problem $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$ assigns values to the variables $c_i$'s and $p_{i,j}$'s from which the probability $v_j$ for every $1 \leq j \leq h$ can be obtained. Suppose the solution assigns the values $\{b_{1,j}, \cdots, b_{h,j}\}$ to the Boolean variables $\{p_{1,j}, \cdots, p_{h,j}\}$ and the values $\{t_1, \cdots, t_h\}$ to the variables $\{c_1, \cdots, c_h\}$, we have: $2^{-\sum_{i=1}^{h} t_i \cdot b_{i,j}} = v_j$. Note that $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$ is always satisfiable.

We can observe that if $t_i = 0$, the value $b_{i,j}$ of the Boolean variable $p_{i,j}$ for $1 \leq j \leq h$ can be omitted for retaining all the probabilities in $V$. We will see later that the values $\{b_{1,j}, \cdots, b_{h,j}\}$ of the Boolean variables $p_{i,j}$'s will be used to encode the probabilities in $V$, we define $\Phi_2^{\mathcal{S}}$ as $\{c_1 = 0, \cdots, c_h = 0\}$ so that a solution of the MaxSMT problem $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$ maximizes the number of 0 bits in the values $\{t_1, \cdots, t_h\}$ of the variables $\{c_1, \cdots, c_h\}$, thus minimizing the number of additional Boolean variables used for encoding the probabilities in $V$.

Let $\{i_1, \cdots, i_k\}$ be the set of indices of the nonzero values in $\{t_1, \cdots, t_h\}$. To encode all the probabilities in $V$, we define an extended DDT $\mathcal{D}_{\mathcal{S}}^{\dagger}$ of the S-box $\mathcal{S}$ as follows:

$$\forall(\Delta X, \Delta Y) \in \mathbb{B}^n \times \mathbb{B}^m . 1 \leq j \leq h.\ \mathcal{D}_{\mathcal{S}}^{\dagger}(\Delta X, \Delta Y, b_{i_1,j}, \cdots, b_{i_k,j}) \neq 0 \text{ iff } \mathsf{Pr}_{\mathcal{S}}(\Delta X, \Delta Y) = v_j.$$

A set $\Psi_{\mathcal{S}}^4$ of constraints over the Boolean variables $\vec{b}, \vec{b'}, p_{i_1}, \cdots, p_{i_k}$ can be generated from the extended DDT $\mathcal{D}_{\mathcal{S}}^{\dagger}$ (cf. Section 4.2) such that

$$\mathcal{D}_{\mathcal{S}}^{\dagger}(\Delta X, \Delta Y, b_{i_1}, \cdots, b_{i_k}) \neq 0 \text{ iff } (\Delta X, \Delta Y, b_{i_1}, \cdots, b_{i_k}) \text{ is a solution of } \Psi_{\mathcal{S}}^4.$$

Proposition 4.4. *For any solution* $(\Delta X, \Delta Y, b_{i_1}, \cdots, b_{i_k})$ *of* $\Psi_{\mathcal{S}}^4$, $\mathsf{Pr}_{\mathcal{S}}(\Delta X, \Delta Y) = 2^{-\sum_{j=1}^{k} t_{i_j} \cdot b_{i_j}}$.

We denote by $\Psi_{\mathcal{S}}^{\dagger}$ the set $\Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^2 \cup \Psi_{\mathcal{S}}^4$ if the S-box $\mathcal{S}$ is injective, otherwise $\Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^4$. An illustrating example is given in [Sun et al. 2023, Section C.2].

## 5 WORD-WISE APPROACH

In this section, we present an approach for determining the lower bound of the minimum number of active S-boxes by reducing to MILP in a word-wise fashion. It works for programs $P$ where types are uint$s[n]$ for a fixed bit size $s$ of the involved S-boxes and individual bits of any entry in an array cannot be changed. Our tool can automatically check if the word-wise approach is applicable. Recall that uint$s$, uint1$[s]$ and uint$s[1]$ are identical, and we shall use uint$s[1]$ hereafter. Note that, in this setting, the program $P$ is free of touint expressions.

**High-level intuition.** Each entry of an array variable $x$ in the program $P$ is modeled as a Boolean variable $b$, where $b = 1$ in the MILP solution indicates that the entry has a difference when $P$ is executed under two distinct inputs for some fixed key. Thus, a variable of type uint$s[n]$ is modeled by a vector $\vec{b}$ of $n$ Boolean variables. Each statement is modeled by a set of IL constraints over the Boolean variables which characterize the propagation of differences through the statement. Furthermore, each S-box $\mathcal{S}$ is associated with a unique Boolean variable $b_{\mathcal{S}}$ such that the S-box $\mathcal{S}$ is active under two execution if $b_{\mathcal{S}} = 1$ (cf. Definition 2.4). The objective function is to minimize the sum of Boolean variables $b_{\mathcal{S}}$ for all the S-boxes $\mathcal{S}$, subject to the extracted IL constraints. The MILP solution gives the lower bound of the minimum number of active S-boxes in the program.

The word-wise MILP generation rules are given by a word-wise differential denotational semantics of EasyBC. We present the denotational semantics for expressions in Section 5.1 and the denotational semantics for statements in Section 5.2.

### 5.1 Word-wise Differential Denotational Semantics for Expressions

W denotes by $\mathbb{X}_f$ the set of its local variables of a function $f$, and by $\mathbb{K}_f \subseteq \mathbb{X}_f$ the set of variables that are subkeys. We denote by $|x| = n$ if $x$ has type uint$s[n]$.

$$\llbracket \texttt{View}(x,i,j) \rrbracket^{\mathsf{W}}_{\gamma} = \big(\emptyset, (\vec{b}_i, \cdots, \vec{b}_j)\big), \text{ where } \vec{b} = \gamma(x) \qquad \llbracket \sim x \rrbracket^{\mathsf{W}}_{\gamma} = \big(\emptyset, \gamma(x)\big)$$

$$\llbracket x_1 + x_2 \rrbracket^{\mathsf{W}}_{\gamma} = \llbracket x_1 - x_2 \rrbracket^{\mathsf{W}}_{\gamma} = \llbracket x_1 \oplus x_2 \rrbracket^{\mathsf{W}}_{\gamma} = \big(\Psi^i_{2,3}(b_0,b_1,b_2), b_0\big), \text{ where } i = 1,2,\ b_0 = \mathsf{newBV}()$$
$$\Psi^1_{2,3}(b_0,b_1,b_2) = \big\{b_1 + b_2 \geq b_0, b_0 + b_1 \geq b_2,\ b_0 + b_2 \geq b_1\big\} \text{ [Li et al. 2019]}, b_1 = \gamma(x_1), b_2 = \gamma(x_2)$$
$$\Psi^2_{2,3}(b_0,b_1,b_2) = \big\{b' \geq b_0, b' \geq b_1, b' \geq b_2, \textstyle\sum_{i=0}^{2} b_i \geq 2b'\big\} \text{ [Mouha et al. 2011]},\ b' = \mathsf{newBV}()$$

$$\llbracket x_1 \wedge x_2 \rrbracket^{\mathsf{W}}_{\gamma} = \llbracket x_1 \vee x_2 \rrbracket^{\mathsf{W}}_{\gamma} = \big(\{b_1 + b_2 \geq b_0\}, b_0\big), \text{ where } b_0 = \mathsf{newBV}(), b_1 = \gamma(x_1), b_2 = \gamma(x_2)$$

$$\llbracket M * x \rrbracket^{\mathsf{W}}_{\gamma} = (\Psi^i_M(\vec{b}, \vec{b}'), \vec{b}'), \text{ where } i = 1,2, \vec{b} = \gamma(x), \vec{b}' = \mathsf{newBV}(),\ b'' = \mathsf{newBV}()$$
$$\Psi^1_M(\vec{b}, \vec{b}') = \big\{\mathcal{B}^{\min}_{\mathsf{ww},M} \cdot b'' \leq \textstyle\sum_{i=0}^{|x|-1}(\vec{b}_i + \vec{b}'_i) \leq \mathcal{B}^{\max}_{\mathsf{ww},M},\ 2|x| \cdot b'' \geq \textstyle\sum_{i=0}^{|x|-1}(\vec{b}_i + \vec{b}'_i)\big\}$$
$$\Psi^2_M(\vec{b}, \vec{b}') = \big\{\mathcal{B}^{\min}_{\mathsf{ww},M} \cdot b'' \leq \textstyle\sum_{i=0}^{|x|-1}(\vec{b}_i + \vec{b}'_i) \leq \mathcal{B}^{\max}_{\mathsf{ww},M},\ b'' \geq \vec{b}_0, b'' \geq \vec{b}'_0, \cdots, b'' \geq \vec{b}_{|x|-1}, b'' \geq \vec{b}'_{|x|-1}\big\}$$

$$\llbracket x\langle\cdot y\cdot\rangle \rrbracket^{\mathsf{W}}_{\gamma} = \big(\emptyset, (\vec{b}_{j_0}, \cdots, \vec{b}_{j_{n-1}})\big), \text{ where } x \text{ is P-box } (j_0, \cdots, j_{n-1}) \text{ and } \vec{b} = \gamma(y)$$

$$\llbracket x\langle y\rangle \rrbracket^{\mathsf{W}}_{\gamma} = \big(\{b \diamond b'\}, b'\big), \text{ where } b = \gamma(y), b' = \mathsf{newBV}(), \diamond \text{ is } = \text{ if } x \text{ is injective, otherwise } \geq$$

Fig. 6. The word-wise differential denotational semantic rules for expressions.

**State.** A state $\gamma$ is a mapping from array entries $(x, i) \in (\mathbb{X}_f \setminus \mathbb{K}_f) \times \mathbb{N}$ to Boolean variables that model the differences of array entries under two executions. For each variable $x \in \mathbb{X}_f \setminus \mathbb{K}_f$,

- $\gamma(x)$ gives the sequence of Boolean variables $\gamma(x, 0), \cdots, \gamma(x, |x| - 1)$ of the array $x$.
- $\gamma[(x, i) \mapsto b]$ denotes the update of $\gamma$ by mapping $(x, i)$ to the Boolean variable $b$, and $\gamma[x \mapsto \vec{b}]$ denotes the update $\gamma[(x, 0) \mapsto \vec{b}_0] \cdots [(x, |x| - 1) \mapsto \vec{b}_{|x|-1}]$ for a Boolean vector $\vec{b}$ with $|\vec{b}| = |x|$.

By abuse of notation, $\gamma(x)$ gives the vector $\vec{0}$ with $|\vec{0}| = |x|$ if $x$ is a constant or subkey variable in $\mathbb{K}_f$. Note that $\gamma(x)$ is $\gamma(x, 0)$ if $x$ has type $\texttt{uints}[1]$. We denote by $\Gamma$ the set of states.

**Denotational semantics**. The (word-wise differential) denotational semantics of an expression $e$ is given by $\llbracket e \rrbracket^{\mathsf{W}}$ that maps each state $\gamma \in \Gamma$ to a pair $(\Psi, \vec{b})$, denoted by $\llbracket e \rrbracket^{\mathsf{W}}_{\gamma}$, where

- $\Psi$ is a set of IL constraints over Boolean variables characterizing the dependency/feasiblity of the differences between the support variables and result of $e$ such that the differences is a solution of $\Psi$ iff these differences are feasible for support variables and result of $e$;
- $\vec{b}$ such that $|\vec{b}| = |e|$ is a vector of the Boolean variables/values that models the difference of the result of $e$ under two executions ($\vec{b}$ may be written as $b$ if $|\vec{b}| = 1$ and $\vec{b}_0 = b$).

**Denotational semantic rules.** The semantics for expressions in EasyBC is shown in Figure 6, where the function $\mathsf{newBV}()$ returns a fresh Boolean variable $b$ or a vector $\vec{b}$ of fresh Boolean variables according to the context. The semantic rules $\llbracket \sim x \rrbracket^{\mathsf{W}}_{\gamma}$, $\llbracket \texttt{View}(x, i, j) \rrbracket^{\mathsf{W}}_{\gamma}$ and $\llbracket x\langle\cdot y\cdot\rangle \rrbracket^{\mathsf{W}}_{\gamma}$ are straightforward according to their operational semantics. We explain the others below.

- $\llbracket x_1 \odot x_2 \rrbracket^{\mathsf{W}}_{\gamma}$ for $\odot \in \{+, -, \oplus, \wedge, \vee\}$ gives a pair $\big(\Psi(b_0, b_1, b_2), b_0\big)$, where $\Psi(b_0, b_1, b_2)$ characterizes the dependency of the differences $b_0$, $b_1$ and $b_2$ between $x_1 \odot x_2$, $x_1$ and $x_2$ according to the maximum and minimum word-wise branch numbers of $\odot$ (cf. Definition 2.7), and $b_0 = 0$ if $b_1 = b_2 = 0$. For instance, $\mathcal{B}^{\min}_{\mathsf{ww},+} = 2$ and $\mathcal{B}^{\max}_{\mathsf{ww},+} = 3$ (cf. Table 2), meaning that either $b_0 = b_1 = b_2 = 0$ or at least two of them are 1 (i.e., $2 \leq b_0 + b_1 + b_2 \leq 3$). For easy reference, these IL constraints are denoted by $\Psi^1_{2,3}$ or $\Psi^2_{2,3}$. Note that the auxiliary Boolean variable $b'$ in $\Psi^2_{2,3}$ is 0 iff $b_0 + b_1 + b_2 = 0$.
- $\llbracket M * x \rrbracket^{\mathsf{W}}_{\gamma}$ gives the pair $\big(\Psi^i_M(\vec{b}, \vec{b}'), \vec{b}'\big)$, where $\Psi^i_M(\vec{b}, \vec{b}')$ characterizes the dependency of the differences $\vec{b}$ and $\vec{b}'$ between the entries in the arrays $x$ and $M \odot x$ according to the maximum and minimum word-wise branch numbers of the linear transformation $M \odot x$. Indeed, $\Psi^i_M(\vec{b}, \vec{b}')$ enforces that the sum of their input and output differences $\sum_{i=0}^{|x|-1}(\vec{b}_i + \vec{b}'_i)$ either ranges from

$$\frac{}{\llbracket \tau\ x \rrbracket^{\mathsf{W}}_{\gamma} = (\emptyset, \gamma, \emptyset)} \qquad \frac{\gamma' = \gamma[(x,i) \mapsto \gamma(y)]}{\llbracket x[i] = y \rrbracket^{\mathsf{W}}_{\gamma} = (\emptyset, \gamma', \emptyset)} \qquad \frac{\llbracket e \rrbracket^{\mathsf{W}}_{\gamma} = (\Psi, \vec{b}) \qquad \gamma' = \gamma[x \mapsto \vec{b}] \quad \Theta = \big(e \text{ is } x'\langle y \rangle ?\ \gamma(y) : \emptyset\big)}{\llbracket x = e \rrbracket^{\mathsf{W}}_{\gamma} = (\Psi, \gamma', \Theta)}$$

$$\frac{\begin{array}{c}\tau_0\ \ f(\tau_1\ x_1, \cdots, \tau_m\ x_m)\{S_1; \cdots; S_n; \texttt{return}\ y;\} \qquad \gamma_0 = \gamma[x_1 \mapsto \gamma(y_1)] \cdots [x_m \mapsto \gamma(y_m)] \\ \llbracket S_1 \rrbracket^{\mathsf{W}}_{\gamma_0} = (\Psi_1, \gamma_1, \Theta_1), \cdots, \llbracket S_n \rrbracket^{\mathsf{W}}_{\gamma_{n-1}} = (\Psi_n, \gamma_n, \Theta_n) \quad \Psi = \bigcup_{i=1}^{n} \Psi_i \quad \gamma' = \gamma[x \mapsto \gamma_n(y)] \quad \Theta = \bigcup_{i=1}^{n} \Theta_i\end{array}}{\llbracket x = g(y_1, \cdots, y_m) \rrbracket^{\mathsf{W}}_{\gamma} = (\Psi,\ \gamma', \Theta)}$$

Fig. 7. The word-wise differential denotational semantic rules for statements.

$\mathcal{B}^{\min}_{\mathsf{ww},M}$ to $\mathcal{B}^{\max}_{\mathsf{ww},M}$ or is 0 in two alternative ways $\Psi^1_M(\vec{b}, \vec{b'})$ and $\Psi^2_M(\vec{b}, \vec{b'})$, where the auxiliary Boolean variable $b''$ in an MILP solution is 0 iff $\sum_{i=0}^{|x|-1}(\vec{b}_i + \vec{b}'_i) = 0$. Note that $\mathcal{B}^{\min}_{\mathsf{ww},M} \geq 1$.

- $\llbracket x\langle y \rangle \rrbracket^{\mathsf{W}}_{\gamma}$ for S-box $x$ depends upon whether $x$ is injective, which is determined by checking whether some constant in the array appears more than once if the S-box is given by an array, or by checking the satisfiability of the constraint $x \neq x' \wedge f(x) = f(x')$ (via SMT solving) if the S-box is defined by an `s_fn` function $f$. If it is injective, $\llbracket x\langle y \rangle \rrbracket^{\mathsf{W}}_{\gamma}$ gives the pair $\big(\{b = b'\}, b'\big)$, where the Boolean variable $b$ models the difference of $y$; the fresh Boolean variable $b'$ models the difference of the result $x\langle y \rangle$; and the constraint $b = b'$ ensures that $b = 1$ iff $b' = 1$. If it is non-injective, the constraint $b \geq b'$ is imposed instead of $b = b'$, as $x\langle y \rangle$ *may* differ in two executions *only* if $y$ differs in the two executions.

Lemma 5.1. *Suppose* $\llbracket e \rrbracket^{\mathsf{W}}_{\gamma} = (\Psi, \vec{b})$ *with* $\Psi \neq \emptyset$. $\vec{b} = (b_1, \cdots, b_m)$ *is a solution of* $\Psi$ *if and only if* $(b_1, \cdots, b_i)$ *is feasible differences of the operands and result of* $e$*, where* $(b_{i+1}, \cdots, b_m)$ *is for the possible auxiliary Boolean variables.*

## 5.2 Word-wise Differential Denotational Semantics for Statements

**Denotational semantics for statements**. The (word-wise differential) denotational semantics of a statement $S$ is given by $\llbracket S \rrbracket^{\mathsf{W}}$ that maps each state $\gamma \in \Gamma$ to a triple $(\Psi, \gamma', \Theta)$, denoted by $\llbracket S \rrbracket^{\mathsf{W}}_{\gamma}$, where $\Psi$ is defined as above (i.e., set of IL constraints), $\gamma'$ is the updated state, and $\Theta$ is a set of Boolean variables each of which models the input difference of an S-box under two executions.

**Denotational semantic rules.** The semantics for statements in EasyBC is shown in Figure 7.

The semantic rules $\llbracket \tau\ x \rrbracket^{\mathsf{W}}_{\gamma}$, $\llbracket x[i] = y \rrbracket^{\mathsf{W}}_{\gamma}$ and $\llbracket x = e \rrbracket^{\mathsf{W}}_{\gamma}$ for declaration $\tau\ x$, array put $x[i] = y$, assignment $x = e$ are straightforward, which updates the state $\gamma$ accordingly to track the mapping from variables $x$ to Boolean variables $\gamma(x)$ that models the difference of $x$ under two executions, the set of constraints $\Psi$ is collected from the semantics $\llbracket e \rrbracket^{\mathsf{W}}_{\gamma}$ of the expression $e$, and moreover, the Boolean variable $\gamma(y)$ modeling the difference of the input $y$ of an S-box is recorded in $\Theta$.

The semantic rule $\llbracket x = g(y_1, \cdots, y_m) \rrbracket^{\mathsf{W}}_{\gamma}$ for a function call $x = g(y_1, \cdots, y_m)$ follows its operational semantics. We first pass the Boolean variables $\gamma(y_i)$ for $1 \leq i \leq m$ that model the differences of the actual arguments $y_i$ to the formal parameters $x_i$, then iteratively evaluate each statement $S_i$ in its function body, and finally maps the variable $x$ to the Boolean variable $\gamma(y)$ that models the difference of the return $y$. The set $\Psi$ of IL constraints and the set $\Theta$ of Boolean variables modeling the input differences of S-boxes are collected from them of the statements, i.e., $\Psi_i$'s and $\Theta_i$'s

## 5.3 Word-wise Resistance Evaluation

To evaluate the resistance of the program $P$, we define the semantics $\llbracket P \rrbracket^{\mathsf{W}}$ of the program $P$ as the semantics of its `fn` function as follows:

$$\llbracket P \rrbracket^{\mathsf{W}} = \llbracket \texttt{uints}[n]\ f(\texttt{uints}[n_1]\ k,\ \texttt{uints}[n]\ txt)\{S_1; \cdots; S_n; \texttt{return}\ y;\} \rrbracket^{\mathsf{W}} = (\Psi, \gamma_n, \Theta)$$

where $\Psi = \bigcup_{i=1}^{n} \Psi_i$, $\Theta = \bigcup_{i=1}^{n} \Theta_i$, $[\![S_1]\!]^{\mathtt{W}}_{\gamma_0} = (\Psi_1, \gamma_1, \Theta_1), \cdots, [\![S_n]\!]^{\mathtt{W}}_{\gamma_{n-1}} = (\Psi_n, \gamma_n, \Theta_n)$ and $\gamma_0$ is an initial state mapping each array element of the formal parameter $txt$ to a fresh Boolean variable.

Clearly, $\Phi$ is the set of IL constraints imposed by the dependency of differences between support variables and results of all the operations in the program $P$, and $\sum_{b\in\Theta} b$ gives the sum of the number of active S-boxes under two executions of the program $P$. Determining the lower bound of the minimum number of active S-boxes under two executions of $P$ amounts to minimizing $\sum_{b\in\Theta} b$ subject to $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$, where $(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1$ ensures that the input difference of text $txt$ is nonzero, otherwise $\sum_{b\in\Theta} b$ would trivially be 0.

Recall from Section 3.5 that $\mathcal{N}_{\mathtt{diff}}$ denotes the minimum number of active S-boxes in all the possible pairs of executions.

Theorem 5.2. *Let* $[\![P]\!]^{\mathtt{W}} = (\Phi, \gamma, \Theta)$ *and* $N$ *be the minimum value of the objective function* $\sum_{b\in\Theta} b$ *subject to* $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$*. We have that* $N \leq \mathcal{N}_{\mathtt{diff}}$.

When the MILP program cannot be solved efficiently with large round number $s$, one can turn to the following decomposition approach.

Proposition 5.3. *Let* $n_1$ *and* $n_2$ *be the minimum number of the active S-boxes of the first* $r_1$*-round and the subsequent* $r_2$*-round differential characteristics respectively, then* $n_1 + n_2$ *is a lower bound of the minimum number of the active S-boxes of the* $(r_1 + r_2)$*-round differential characteristics.*

In practice, one may be only interested in proving the resistance against differential cryptanalysis instead of computing a bound. In this case, it suffices to prove that $\mathcal{N}_{\mathtt{diff}} \geq \frac{-\mathscr{b}}{\log_2 p}$, where $p$ denotes the maximum probability $\mathtt{Pr}_{\mathcal{S}}(\Delta X, \Delta Y)$ among all the nonzero differentials $(\Delta X, \Delta Y)$ for any S-box $\mathcal{S}$ that is active in $s$-round differential characteristics and $\mathscr{b}$ is the block size of the cipher. By Theorem 5.2, we only need to verify if $\{\sum_{b\in\Theta} b < \frac{-\mathscr{b}}{\log_2 p}\} \cup \Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ is unsatisfiable.

Corollary 5.4. *Let* $[\![P]\!]^{\mathtt{W}} = (\Phi, \gamma, \Theta)$*. If* $\{\sum_{b\in\Theta} b < \frac{-\mathscr{b}}{\log_2 p}\} \cup \Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ *is unsatisfiable, then the program* $P$ *with block size* $\mathscr{b}$ *is resistant against differential cryptanalysis.*

If the $s$-round cipher $P$ can be partitioned into $\frac{s}{s'}$ identical $s'$-round ciphers $P'$, by Proposition 5.3 and Corollary 5.4, we can conclude that the cipher $P$ is resistant against differential cryptanalysis if the number of active S-boxes of the cipher $P'$ is no less than $\frac{-s' \cdot \mathscr{b}}{s \cdot \log_2 p}$.

Corollary 5.5. *Let* $[\![P']\!]^{\mathtt{W}} = (\Phi, \gamma, \Theta)$*. If* $\{\sum_{b\in\Theta} b < \frac{-s' \cdot \mathscr{b}}{s \cdot \log_2 p}\} \cup \Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ *is unsatisfiable, then the program* $P$ *with block size* $\mathscr{b}$ *is resistant against differential cryptanalysis.*

## 6 BIT-WISE APPROACH

In this section, we present a bit-wise approach which, as the word-wise approach, determines the lower bound of the minimum number of active S-boxes, but lifts its limitation requiring that a program uses types $\mathtt{uint}s[n]$ for a fixed bit size $s$ of involved S-boxes and individual bits of any entry in an array cannot be changed.

**High-level intuition.** Fix a program $P$, which we normally assume cannot be handled by the word-wise approach. A straightforward idea is to transform the program $P$ to its Boolean counterpart $P'$ by bit-blasting. However, this would introduce a large number of variables and statements, resulting in a prohibitively large MILP problem. In this work, we adopt a strategy to "implicitly" bit-blast, meaning that bit-blasting is performed in the generation of MILP. To this end, each bit of a variable in the program $P$ is modeled by one Boolean variable $b$, where $b = 1$ in an MILP solution indicates that the corresponding bit differs in $P$ when executed under two distinct inputs for some fixed key. Thus, a variable of type $\mathtt{uint}s[n]$ is modeled by a vector $\vec{b}$ of $s \cdot n$ Boolean variables. The word-wise differential denotational semantics is then lifted to the bit-wise one.

### 6.1 Bit-wise Differential Denotational Semantics for Expressions

**State**. We first lift the state $\gamma$ from word-wise to bit-wise. Let $\|x\| = s \cdot n$ for a variable $x$ of the type $\texttt{uint}s[n]$. A state $\gamma$ now maps each pair $(x, i) \in (\mathbb{X}_f \setminus \mathbb{K}_f) \times \mathbb{N}$ to a Boolean variable, where

- for each variable $x$ of type $\texttt{uint}s[n]$, $\gamma(x, i \cdot s + j)$ gives a Boolean variable modeling the difference of the $(j + 1)$-th most significant bit of the $(i + 1)$-th entry $x_i$ in the array $x$;
- $\gamma(x)$ denotes the sequence $\gamma(x, 0), \gamma(x, 1), \cdots, \gamma(x, \|x\| - 1)$, $\gamma[(x, i) \mapsto b]$ denotes the update of the state $\gamma$ by mapping $(x, i)$ to the Boolean variable $b$, and $\gamma[x \mapsto \vec{b}]$ denotes the update $\gamma[(x, 0) \mapsto \vec{b}_0] \cdots [(x, \|x\| - 1) \mapsto \vec{b}_{\|x\|-1}]$ for a Boolean vector $\vec{b}$ with $\|x\| = \|\vec{b}\|$.

**Denotational semantics**. The (bit-wise differential) denotational semantics of an expression $e$ is given by $[\![e]\!]^{\mathsf{B}}$ that maps each state $\gamma \in \Gamma$ to a pair $(\Psi, \vec{b})$, denoted by $[\![e]\!]^{\mathsf{B}}_{\gamma}$, where

- $\Psi$ is a set of IL constraints over Boolean variables characterizing the bit-level dependency of differences between the support variables and result of $e$ such that the differences are a solution of $\Psi$ iff these bit-level differences are feasible for support variables and result of $e$;
- $\vec{b}$ such that $\|\vec{b}\| = \|e\|$ is a Boolean vector modeling the bit-level differences of the result of $e$.

**Denotational semantic rules.** The semantics for expressions in EasyBC is shown in Figure 8. The semantic rules $[\![\sim x]\!]^{\mathsf{B}}_{\gamma}$, $[\![\texttt{View}(x, i, j)]\!]^{\mathsf{B}}_{\gamma}$, $[\![\texttt{touint}(x_1, \cdots, x_m)]\!]^{\mathsf{B}}_{\gamma}$ and $[\![x\langle\cdot y\cdot\rangle]\!]^{\mathsf{W}}_{\gamma}$ are trivial according to their operational semantics. Below, we explain the other non-trivial ones.

- The semantic rule $[\![x \odot y]\!]^{\mathsf{B}}_{\gamma}$ for $\odot \in \{\wedge, \vee\}$ gives the pair $\big(\{\vec{b}^1_i + \vec{b}^2_i \geq \vec{b}^0_i \mid 0 \leq i < \|x\|\}, \vec{b}^0\big)$, where for every $0 \leq i < \|x\|$, $\vec{b}^1_i + \vec{b}^2_i \geq \vec{b}^0_i$ characterizes that if the $(i + 1)$-th bits of the operands $x$ and $y$ have no differences (i.e., $\vec{b}^1_i = \vec{b}^2_i = 0$), then the $(i + 1)$-th bit of the result $x \odot y$ has no differences (i.e., $\vec{b}^0_i = 0$). Otherwise, it may have differences (with the probability of $\frac{1}{2}$).
- The semantic rule $[\![x \oplus y]\!]^{\mathsf{B}}_{\gamma}$ gives the pair $\big(\bigcup_{i=0}^{\|x\|-1} \psi^j_{\oplus}(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i), \vec{b}^0\big)$, where for each $0 \leq i < \|x\|$, $\psi^j_{\oplus}(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$ characterizes that either the $(i+1)$-th bits of the operands $x$, $y$ and result $x \oplus y$ have no differences (i.e., $\vec{b}^0_i = \vec{b}^1_i = \vec{b}^2_i = 0$) or exactly two of them have differences (i.e., $\vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i = 2$).
- The semantic rule $[\![x \odot y]\!]^{\mathsf{B}}_{\gamma}$ for $\odot \in \{+, -\}$ gives the pair $\big(\bigcup_{i=0}^{\|x\|-1} \Psi_i, \vec{b}^0\big)$, where $\Psi_i$ characterizes the dependency of the differences between the $i$-th and $(i + 1)$-th bits of the operands $x$, $y$ and result $x \odot y$. The dependency is obtained by bit-blasting $x + y$ via a ripple-carry adder, i.e.,
  - $\mathsf{bin}_i(x + y) = \mathsf{bin}_i(x) \oplus \mathsf{bin}_i(y) \oplus \vec{c}_i$, for every $0 \leq i < \|x\|$;
  - the carry bit $\vec{c}_i = 1$ iff $\mathsf{bin}_{i-1}(x) + \mathsf{bin}_{i-1}(y) + \vec{c}_{i-1} \geq 2$, for every $1 \leq i < \|x\|$, with $\vec{c}_0 = 0$,

  where $\mathsf{bin}_i(x)$ denotes the $(i + 1)$-th most significant bit of $x$. Clearly, the difference $\vec{b}^0_i$ of the bit $\mathsf{bin}_i(x + y)$ depends upon the differences $\vec{b}^1_{i-1}$, $\vec{b}^2_{i-1}$, $\vec{b}^1_i$ and $\vec{b}^2_i$ of the bits $\mathsf{bin}_{i-1}(x)$, $\mathsf{bin}_{i-1}(y)$, $\mathsf{bin}_i(x)$ and $\mathsf{bin}_i(y)$. Indeed, we can deduce the dependency:
  - if $\vec{b}^0_{i-1} = \vec{b}^1_{i-1} = \vec{b}^2_{i-1} = 1$, then $\vec{c}_{i-1} \oplus \vec{c}'_{i-1} = \vec{b}^3_{i-1} = \vec{b}^3_i = 1$ and $\vec{b}^0_i = \neg(\vec{b}^1_i \oplus \vec{b}^2_i)$,
  - if $\vec{b}^0_{i-1} = \vec{b}^1_{i-1} = \vec{b}^2_{i-1} = 0$, then $\vec{c}_{i-1} \oplus \vec{c}'_{i-1} = \vec{b}^3_{i-1} = \vec{b}^3_i = 0$ and $\vec{b}^0_i = \vec{b}^1_i \oplus \vec{b}^2_i$,
  - otherwise $1 \leq \vec{b}^0_{i-1} + \vec{b}^1_{i-1} + \vec{b}^2_{i-1} \leq 2$. Indeed, the probability of $\vec{b}^0_i = \neg(\vec{b}^1_i \oplus \vec{b}^2_i)$, (resp. $\vec{b}^0_i = \vec{b}^1_i \oplus \vec{b}^2_i$ and $\vec{b}^0_i = 1$) is $\frac{1}{2}$.

  The above dependency is characterized by the set of IL constraints $\Psi_i$. Furthermore, $\Psi_0$ and $\Psi_1$ can be simplified by $\vec{c}_0 = 0$. The semantic rule $[\![x - y]\!]^{\mathsf{B}}_{\gamma}$ is defined the same as $[\![x + y]\!]^{\mathsf{B}}_{\gamma}$ because $z = x - y$ iff $x = z + y$, and the Boolean variables $\vec{b}^0_i$, $\vec{b}^1_i$ and $\vec{b}^2_i$ in $\Psi_i$ are symmetric.
- The semantic rule $[\![M * x]\!]^{\mathsf{B}}_{\gamma}$ gives the pair $\big(\bigcup_{i=0}^{|x|-1} \bigcup_{h=0}^{s-1} \Psi^{v}_{\mathtt{M,i,h}}, \vec{b'}\big)$, where for each $0 \leq i < |x|$ and each $0 \leq h < s$, $\Psi^{v}_{\mathtt{M,i,h}}$ characterizes the bit-level dependency of the differences between the array $x$ and the $(h + 1)$-th bit $\mathsf{bin}_h(y_i)$ of the $(i + 1)$-th entry $y_i$ in the resulting array $y = M * x$.

$$[\![\sim x]\!]^{\mathsf{B}}_{\gamma} = \big(\emptyset, \gamma(x)\big) \qquad [\![\mathtt{touint}(x_1, \cdots, x_m)]\!]^{\mathsf{B}}_{\gamma} = \big(\emptyset, (\gamma(x_1), \cdots, \gamma(x_m))\big)$$

$$[\![\mathtt{View}(x, i, j)]\!]^{\mathsf{B}}_{\gamma} = \big(\emptyset, (\vec{b}_{i\cdot s+0}, \cdots, \vec{b}_{i\cdot s+s-1}, \cdots, \vec{b}_{j\cdot s+0}, \cdots, \vec{b}_{j\cdot s+s-1})\big), \text{ where } \vec{b} = \gamma(x)$$

$$[\![x \wedge y]\!]^{\mathsf{B}}_{\gamma} = [\![x \vee y]\!]^{\mathsf{B}}_{\gamma} = \big(\{\vec{b}^1_i + \vec{b}^2_i \geq \vec{b}^0_i \mid 0 \leq i < \|x\|\}, \vec{b}^0\big), \text{ where } \vec{b}^1 = \gamma(x),\ \vec{b}^2 = \gamma(y),\ \vec{b}^0 = \mathsf{newBV}()$$

$$[\![x \oplus y]\!]^{\mathsf{B}}_{\gamma} = \big(\textstyle\bigcup_{i=0}^{\|x\|-1} \psi^j_{\oplus}(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i), \vec{b}^0\big), \text{ where } \vec{b}^1 = \gamma(x),\ \vec{b}^2 = \gamma(y),\ \vec{b}^0 = \mathsf{newBV}(), b' = \mathsf{newBV}()$$

$\psi^1_{\oplus}(b^1, b^2, b^0) = \{2 \geq b^0 + b^1 + b^2 \geq 2b',\ b' \geq b^0,\ b' \geq b^1,\ b' \geq b^2\}$ [Sun et al. 2014a]
$\psi^2_{\oplus}(b^1, b^2, b^0) = \{b^0 + b^1 + b^2 \leq 2,\ b^1 + b^2 \geq b^0,\ b^0 + b^1 \geq b^2,\ b^0 + b^2 \geq b^1\}$ [Sasaki and Todo 2017]
$\psi^3_{\oplus}(b^1, b^2, b^0) = \{b^0 + b^1 + b^2 = 2b'\}$ [Cui et al. 2016]

$$[\![x + y]\!]^{\mathsf{B}}_{\gamma} = [\![x - y]\!]^{\mathsf{B}}_{\gamma} = \big(\textstyle\bigcup_{i=0}^{\|x\|-1} \Psi_i, \vec{b}^0\big), \text{ where } \vec{b}^1 = \gamma(x),\ \vec{b}^2 = \gamma(y),\ \vec{b}^0 = \mathsf{newBV}(),$$

$$\Psi_0 = \Psi^i_{\oplus}(\vec{b}^0_0, \vec{b}^1_0, \vec{b}^2_0) \qquad \Psi_1 = \left\{ \begin{array}{cc} \vec{b}^0_1 + \vec{b}^1_1 + \vec{b}^2_1 \leq \vec{b}^1_0 + \vec{b}^2_0 + 2, & -\vec{b}^0_1 + \vec{b}^1_1 - \vec{b}^2_1 \leq \vec{b}^1_0 + \vec{b}^2_0, \\ -\vec{b}^0_1 - \vec{b}^1_1 + \vec{b}^2_1 \leq \vec{b}^1_0 + \vec{b}^2_0, & \vec{b}^0_1 - \vec{b}^1_1 - \vec{b}^2_1 \leq \vec{b}^1_0 + \vec{b}^2_0 \end{array} \right\}$$

$$\forall 2 \leq i < \|x\|.\ \Psi_i = \left\{ \begin{array}{ll} 4 \geq \sum_{j=0}^2 \vec{b}^j_{i-1} - \vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \geq 0, & 4 \geq \sum_{j=0}^2 \vec{b}^j_{i-1} + \vec{b}^0_i + \vec{b}^1_i - \vec{b}^2_i \geq 0, \\ 4 \geq \sum_{j=0}^2 \vec{b}^j_{i-1} + \vec{b}^0_i - \vec{b}^1_i + \vec{b}^2_i \geq 0, & \sum_{j=0}^2 \vec{b}^j_{i-1} + 2 \geq \vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \geq \sum_{j=0}^2 \vec{b}^j_{i-1} - 2 \end{array} \right\}$$

$$[\![M * x]\!]^{\mathsf{B}}_{\gamma} = \big(\textstyle\bigcup_{i=0}^{|x|-1} \bigcup_{h=0}^{s-1} \Psi^v_{\mathtt{M,i,h}}, \vec{b}'\big), \text{ where } \vec{b} = \gamma(x), \vec{b}' = \mathsf{newBV}(), b^i_{\mathsf{new}} = \mathsf{newBV}(), v \in \{1, 2\}$$

$$\Psi^v_{\mathtt{M,i,h}} = \psi^v_{\oplus}(b^0, b^1, b^1_{\mathsf{new}}) \cup \psi^v_{\oplus}(b^1_{\mathsf{new}}, b^2, b^2_{\mathsf{new}}) \cup \cdots \cup \psi^v_{\oplus}(b^{m-2}_{\mathsf{new}}, b^{m-1}, b^{m-1}_{\mathsf{new}}) \cup \psi^v_{\oplus}(b^{m-1}_{\mathsf{new}}, b^m, \vec{b}'_{i\cdot s+h})$$

$\Psi^3_{\mathtt{M,i,h}} = \{\vec{b}'_{i\cdot s+h} + \sum_{j=0}^m b^j = 2d,\ 0 \leq d \leq \lfloor \frac{m+2}{2} \rfloor\}$, where $d$ is a fresh integer variable, $\{b^0, \cdots, b^m\}$ is the set of support variables of $\Big(\bigoplus_{0\leq j<|x|,\ h\leq k<s,\ M_{i,j,k}=1} \vec{b}_{j\cdot s+k}\Big) \oplus \Big(\vec{c}_h \wedge \bigoplus_{0\leq j<|x|,\ 0\leq k<\lfloor \frac{s}{2} \rfloor} \vec{b}_{j\cdot s+2k}\Big)$

$[\![x\langle\cdot y\cdot\rangle]\!]^{\mathsf{B}}_{\gamma} = (\emptyset,\ \vec{b}^0 || \cdots || \vec{b}^{|x|-1})$, where $\vec{b} = \gamma(y)$, $x$ is P-box $(j_0, \cdots, j_{n-1})$, and
$\vec{b}^i = (\vec{b}_{j_i\cdot s+0}, \cdots, \vec{b}_{j_i\cdot s+s-1})$ for $0 \leq i \leq |x| - 1$

$[\![x\langle y\rangle]\!]^{\mathsf{B}}_{\gamma} = (\Psi_{\mathcal{S}} \cup \Psi^{\mathsf{bn}}_{\mathcal{S}},\ \vec{b}')$, where $\vec{b} = \gamma(y)$, $x : \mathbb{B}^n \rightarrow \mathbb{B}^m$ is an S-box $\mathcal{S}$, $\vec{b}' = \mathsf{newBV}()$, $b = \mathsf{newBV}()$,
and $\Psi^{\mathsf{bn}}_{\mathcal{S}} = \{\mathcal{B}^{\min}_{\mathsf{bw},x} \cdot b \leq \sum_{i=0}^{n-1} \vec{b}_i + \sum_{i=0}^{m-1} \vec{b}'_i \leq \mathcal{B}^{\max}_{\mathsf{bw},x},\ b \geq \vec{b}_i,\ b \geq \vec{b}'_j \mid 0 \leq i < n,\ 0 \leq j < m\}$

Fig. 8. The bit-wise differential denotational semantic rules for expressions, where $s = \frac{\|x\|}{|x|}$ denotes the bit width of the entries of the array $x$, and $\vec{c} = \mathsf{bin}(2 \otimes 2^{s-1})$ is the $s$-bitstream corresponding to the coefficients of the irreducible polynomial for the underlying finite-field.

The dependency is obtained by expanding the matrix-vector product as follows:

$$M * x = \Big(\bigoplus_{j=0}^{|x|-1}(M_{0,j} \otimes x_j), \cdots, \bigoplus_{j=0}^{|x|-1}(M_{|x|-1,j} \otimes x_j)\Big).$$

Clearly, the $(i+1)$-th entry $y_i$ is $\bigoplus_{j=0}^{|x|-1}(M_{i,j} \otimes x_j)$ and thus the difference $\vec{b}'_{i\cdot s+h}$ of its $(h+1)$-th bit $\mathsf{bin}_h(y_i)$ is the parity of the differences of the $(h+1)$-th bits in $M_{i,j} \otimes x_j$ for $0 \leq j < h$. The expression $M_{i,j} \otimes x_j$ is bit-blasted by expanding the finite-field multiplication ($\otimes$) using a series of modular left shifts, XOR operations and the coefficients $\vec{c}$ of the irreducible polynomial for the underlying finite field, where $\vec{c} = \mathsf{bin}(2 \otimes 2^{s-1})$. We finally can deduce that the parity of the differences of the $(h+1)$-th bits in $M_{i,j} \otimes x_j$ for $0 \leq j < h$ is

$$\Big(\bigoplus_{0\leq j<|x|,\ h\leq k<s,\ M_{i,j,k}=1} \vec{b}_{j\cdot s+k}\Big) \oplus \Big(\vec{c}_h \wedge \bigoplus_{0\leq j<|x|,\ 0\leq k<\lfloor \frac{s}{2} \rfloor} \vec{b}_{j\cdot s+2k}\Big).$$

Let $\{b^0, \cdots, b^m\}$ be the set of support variables of the above expression. We have:

$$\vec{b}'_{i\cdot s+h} = \textstyle\bigoplus_{t=0}^m b_t,$$

which can be alternatively characterized by $\Psi^{v}_{\mathsf{M,i,h}}$ for $v \in \{1, 2, 3\}$.

- The semantic rule $\llbracket x\langle y\rangle \rrbracket^{\mathsf{W}}_{\gamma}$ for S-box $x = \mathcal{S}$ gives the pair $\left(\Psi_{\mathcal{S}} \cup \Psi^{\mathsf{bn}}_{\mathcal{S}},\ \vec{b}'\right)$, where $\Psi_{\mathcal{S}}$ is a set of IL constraints characterizing the bit-level dependency of differences between the input $y$ and the result $\mathcal{S}(y)$ (cf. Section 4.2) and $\Psi^{\mathsf{bn}}_{\mathcal{S}}$ enforces that the Hamming weight of the bitstream $\vec{b}\|\vec{b}'$ ranges from $\mathcal{B}^{\mathsf{min}}_{\mathsf{bw},\mathcal{S}}$ to $\mathcal{B}^{\mathsf{max}}_{\mathsf{bw},\mathcal{S}}$. Though $\Psi^{\mathsf{bn}}_{\mathcal{S}}$ is redundant, it often boosts MILP solving.

Lemma 6.1. *Suppose $\llbracket e \rrbracket^{\mathsf{B}}_{\gamma} = (\Psi, \vec{b})$ with $\Psi \neq \emptyset$. The assignment $(b_1, \cdots, b_m)$ is a solution of $\Psi$ if and only if $(b_1, \cdots, b_i)$ is feasible bit-level differences of the operands and result of $e$, where $(b_{i+1}, \cdots, b_m)$ is for the possible auxiliary Boolean variables.*

### 6.2 Bit-wise Differential Denotational Semantics for Statements

The (bit-wise differential) denotational semantics of a statement $S$ is given by $\llbracket S \rrbracket^{\mathsf{B}}$ that maps each state $\gamma \in \Gamma$ to a triple $(\Psi, \gamma', \Theta)$, denoted by $\llbracket S \rrbracket^{\mathsf{B}}_{\gamma}$, where $\Psi$, $\gamma'$ and $\Theta$ are the same as above.

The bit-wise differential semantic rules for statements in EasyBC are the same as those word-wise ones except for array put and S-box access, which are given below:

$$\frac{\vec{b} = \gamma(y) \qquad x \text{ has type } \texttt{uints}[n] \qquad \gamma' = \gamma[(x, i \cdot s + 0) \mapsto \vec{b}_0] \cdots [(x, i \cdot s + s - 1) \mapsto \vec{b}_{s-1}]}{\llbracket x[i] = y \rrbracket^{\mathsf{B}}_{\gamma} = (\emptyset, \gamma', \emptyset)}$$

$$\frac{\begin{array}{c}\llbracket x'\langle y\rangle \rrbracket^{\mathsf{B}}_{\gamma} = (\Psi, \vec{b}) \qquad \vec{b}' = \gamma(y) \qquad b_x = \mathsf{newBV}() \\ \Psi' = \Psi \cup \{\sum_{j=0}^{\|y\|-1} \vec{b}'_j \geq b_x \geq \vec{b}'_i \mid 0 \leq i < \|y\| \qquad \gamma' = \gamma[x \mapsto \vec{b}]\}\end{array}}{\llbracket x = x'\langle y\rangle \rrbracket^{\mathsf{B}}_{\gamma} = (\Psi', \gamma', \{b_x\})}$$

Intuitively, the semantic rule $\llbracket x[i] = y \rrbracket^{\mathsf{B}}_{\gamma}$ updates the state $\gamma$ accordingly by mapping the bits of the $(i + 1)$-th entry in the array $x$ to the Boolean variables $\vec{b}$ that model the differences of the bits of the result $e$. The semantic rule $\llbracket x = x'\langle y\rangle \rrbracket^{\mathsf{B}}_{\gamma}$ adds a fresh Boolean variable $b_x$, where if $b_x = 1$, then the S-box is active, i.e., some bit $\vec{b}'_i$ that models the difference of one bit of input $y$ is nonzero.

### 6.3 Bit-wise Resistance Evaluation

To evaluate the resistance of the program $P$ in a bit-wise manner, similar to $\llbracket P \rrbracket^{\mathsf{W}}$ (cf. Section 5.3), we define the (bit-wise) semantics $\llbracket P \rrbracket^{\mathsf{B}}$ of the program $P$ using its $\texttt{fn}$ function $f$ as follows.

$$\llbracket P \rrbracket^{\mathsf{B}} = \llbracket \texttt{uints}[n]\ f(\texttt{uints}[n_1]\ k,\ \texttt{uints}[n]\ txt)\{S_1; \cdots; S_n; \texttt{return}\ y;\} \rrbracket^{\mathsf{B}} = (\Psi, \gamma_n, \Theta)$$

where $\Psi = \bigcup_{i=1}^{n} \Psi_i$, $\Theta = \bigcup_{i=1}^{n} \Theta_i$, $\llbracket S_1 \rrbracket^{\mathsf{B}}_{\gamma_0} = (\Psi_1, \gamma_1, \Theta_1), \cdots, \llbracket S_n \rrbracket^{\mathsf{B}}_{\gamma_{n-1}} = (\Psi_n, \gamma_n, \Theta_n)$ and $\gamma_0$ is an initial state mapping each bit of array elements of $txt$ to a fresh Boolean variable. We get that:

Theorem 6.2. *Let $\llbracket P \rrbracket^{\mathsf{B}} = (\Phi, \gamma, \Theta)$ and $N$ be the minimum value of the objective function $\sum_{b \in \Theta} b$ subject to the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s \cdot n-1} \gamma(txt, i) \geq 1\}$. We have that $N \leq \mathcal{N}_{\mathtt{diff}}$.*

Corollary 5.4 and Corollary 5.5 still hold when $\llbracket P \rrbracket^{\mathsf{W}} = (\Phi, \gamma, \Theta)$ is replaced by $\llbracket P \rrbracket^{\mathsf{B}} = (\Phi, \gamma, \Theta)$.

## 7 EXTENDED BIT-WISE APPROACH

The lower bound of the minimum number of the active S-boxes is often effective, but it may not be sufficiently tight to prove the resistance [Sun et al. 2014b], as the probabilities between input and output differences for the operations ∧, ∨, +, − and S-boxes are not fully addressed. Furthermore, the bit-wise approach is not applicable if non-linear layers are implemented in other operations than S-boxes (e.g., SIMON), or S-boxes are too large to be given as arrays (e.g., SPARKLE). In this

$$[\![x \wedge y]\!]^{\mathsf{EB}}_{\gamma} = [\![x \vee y]\!]^{\mathsf{EB}}_{\gamma} = \big(\textstyle\bigcup_{i=0}^{\|x\|-1} \Psi_i,\ \vec{b}^0,\ \sum_{i=0}^{\|x\|-1} \vec{p}_i\big), \text{ where } \vec{b}^1 = \Gamma(x),\ \vec{b}^2 = \Gamma(y), \vec{b}^0 = \mathsf{newBV}()$$

$$\Psi_i = \big\{\vec{b}^1_i + \vec{b}^2_i \geq \vec{p}_i,\ \vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \leq 3\vec{p}_i\big\}$$

$$[\![x + y]\!]^{\mathsf{EB}}_{\gamma} = [\![x - y]\!]^{\mathsf{EB}}_{\gamma} = \big(\textstyle\bigcup_{i=0}^{\|x\|-1} \Psi_i,\ \vec{b}^0,\ \sum_{i=1}^{\|x\|-1} \vec{p}_i\big), \text{ where } \vec{b}^1 = \Gamma(x),\ \vec{b}^2 = \Gamma(y),\ \vec{b}^0 = \mathsf{newBV}()$$

$$\Psi_0 = \Psi^i_{\oplus}\big(\vec{b}^0_0, \vec{b}^1_0, \vec{b}^2_0\big) \qquad \Psi_1 = \left\{\begin{array}{c} \vec{b}^0_1 + \vec{b}^1_1 + \vec{b}^2_1 \leq \vec{p}_1 + 2,\ -\vec{b}^0_1 + \vec{b}^1_1 - \vec{b}^2_1 \leq \vec{p}_1,\ -\vec{b}^0_1 - \vec{b}^1_1 + \vec{b}^2_1 \leq \vec{p}_1, \\ \vec{b}^0_1 - \vec{b}^1_1 - \vec{b}^2_1 \leq \vec{p}_1,\ \ \vec{p}_1 \leq \vec{b}^1_0 + \vec{b}^2_0 \leq 2\vec{p}_1 \end{array}\right\}$$

$$\forall 2 \leq i < \|x\|.\ \Psi_i = \Psi^1_i \cup \Psi^2_i \qquad \Psi^1_i = \left\{\begin{array}{r} 3 - \vec{p}_i \geq \sum_{j=0}^{2} \vec{b}^j_{i-1} \geq \vec{p}_i,\ \ \vec{p}_i \geq \vec{b}^0_{i-1} - \vec{b}^1_{i-1}, \\ \vec{p}_i \geq \vec{b}^1_{i-1} - \vec{b}^2_{i-1},\ \ \vec{p}_i \geq \vec{b}^2_{i-1} - \vec{b}^0_{i-1} \end{array}\right\}$$

$$\Psi^2_i = \left\{\begin{array}{ll} 2 - \vec{b}^1_{i-1} + \vec{p}_i \geq -\vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \geq -\vec{b}^1_{i-1} - \vec{p}_i, & 2 - \vec{b}^1_{i-1} + \vec{p}_i \geq \vec{b}^0_i + \vec{b}^1_i - \vec{b}^2_i \geq -\vec{b}^1_{i-1} - \vec{p}_i, \\ 2 - \vec{b}^1_{i-1} + \vec{p}_i \geq \vec{b}^0_i - \vec{b}^1_i + \vec{b}^2_i \geq -\vec{b}^1_{i-1} - \vec{p}_i, & 2 + \vec{b}^1_{i-1} + \vec{p}_i \geq \vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \geq \vec{b}^1_{i-1} - \vec{p}_i \end{array}\right\}$$

$$[\![x\langle y\rangle]\!]^{\mathsf{EB}}_{\gamma} = \big(\Psi^{\dagger}_{\mathcal{S}} \cup \Psi^{\mathsf{bn}}_{\mathcal{S}},\ \vec{b}',\ \textstyle\sum_{j=1}^{k} t_{i_j} \cdot p_{i_j}\big), \text{ where } x \text{ is an S-box } \mathcal{S} : \mathbb{B}^n \longrightarrow \mathbb{B}^m$$

Fig. 9. The extended bit-wise differential denotational semantic rules for expressions, where $\vec{p} = \mathsf{newBV}()$ is a vector of fresh Boolean variables used to encode probabilities, $\Psi^i_{\oplus}\big(\vec{b}^0_0, \vec{b}^1_0, \vec{b}^2_0\big)$ for $i \in \{1, 2, 3\}$ and $\Psi^{\mathsf{bn}}_{\mathcal{S}}$ are defined in Figure 8, $\Psi^{\dagger}_{\mathcal{S}}$ and $\sum_{j=1}^{k} t_{i_j} \cdot p_{i_j}$ are defined in Section 4.3.

section, we extend the bit-wise approach to directly bound the MaxEDCP rather than by bounding the minimum number of the active S-boxes.

**High-level intuition.** Apart from the encoding presented in Section 4.3 for S-boxes, we further encode the probabilities between input and output differences for the operations $\wedge, \vee, +, -$ into IL constraints using additional Boolean variables. The weighted sum $\varrho$ of these additional Boolean variables retains the probability $2^{-\varrho}$. Thus, the objective function is designed to minimize $\varrho$ instead of the number of the active S-boxes.

### 7.1 Extended Bit-wise Differential Denotational Semantics for Expressions

**Denotational semantics**. The (extended bit-wise differential) denotational semantics of an expression $e$ is given by $[\![e]\!]^{\mathsf{EB}}$ that maps each state $\gamma \in \Gamma$ to a triple $(\Psi, \vec{b}, \varrho)$, denoted by $[\![e]\!]^{\mathsf{EB}}_{\gamma}$, where the state $\gamma$, set of IL constraint $\Psi$ and Boolean vector $\vec{b}$ are the same as in Section 6.1, and the expression $\varrho$ is a weighted sum of Boolean variables encoding the probability $2^{-\varrho}$.

**Denotational semantic rules.** The semantic rules $[\![e]\!]^{\mathsf{EB}}_{\gamma}$ for $\wedge, \vee, +, -$ and S-boxes are shown in Figure 9, otherwise $[\![e]\!]^{\mathsf{EB}}_{\gamma} = (\Psi, \vec{b}, 0)$ if $[\![e]\!]^{\mathsf{B}}_{\gamma} = (\Psi, \vec{b})$, meaning that the probability between the input and output differences characterized by $\Psi$ is 1 (i.e., $2^{-0}$).

- The semantic rule $[\![x \odot y]\!]^{\mathsf{EB}}_{\gamma}$ for $\odot \in \{\wedge, \vee\}$ gives the triple $\big(\bigcup_{i=0}^{\|x\|-1} \Psi_i,\ \vec{b}^0,\ \sum_{i=0}^{\|x\|-1} \vec{p}_i\big)$, where for every $0 \leq i < \|x\|$, $\Psi_i$ ensures that the probability of the difference $\vec{b}^0_i$ of the $(i+1)$-th bit in the result $x \odot y$ is $2^{-\vec{p}_i}$ when the differences of the $(i+1)$-th bits of the operands $x$ and $y$ are $\vec{b}^1_i$ and $\vec{b}^2_i$, respectively. Indeed, the probability of $\vec{b}^0_i = 0$ is $2^{-0}$ when $\vec{b}^1_i = \vec{b}^2_i = 0$, then $\vec{p}_i$ must be 0. The probability of $\vec{b}^0_i = 1$ is $2^{-1}$ when $\vec{b}^1_i + \vec{b}^2_i \geq 1$, then $\vec{p}_i$ must be 1.
- The semantic rule $[\![x \odot y]\!]^{\mathsf{EB}}_{\gamma}$ for $\odot \in \{+, -\}$ gives the triple $\big(\bigcup_{i=0}^{\|x\|-1} \Psi_i,\ \vec{b}^0,\ \sum_{i=1}^{\|x\|-1} \vec{p}_i\big)$, where for every $0 \leq i < \|x\|$, $\Psi_i$ ensures that for any fixed $\vec{b}^1_i$ and $\vec{b}^2_i$,
  - if $\vec{b}^0_{i-1} = \vec{b}^1_{i-1} = \vec{b}^2_{i-1}$, $\vec{p}_i = 0$ (i.e., the probability $2^{-\vec{p}_i}$ of $\vec{b}^0_i = \neg(\vec{b}^1_i \oplus \vec{b}^2_i)$ or $\vec{b}^0_i = \vec{b}^1_i \oplus \vec{b}^2_i$ is 1),
  - if $1 \leq \sum_{j=0}^{2} \vec{b}^j_{i-1} \leq 2$, $\vec{p}_i = 1$ (i.e., the probability $2^{-\vec{p}_i}$ of $\vec{b}^0_i = 1$ or $\vec{b}^0_i = 0$ is $\frac{1}{2}$).

We remark that $\Psi_0$ and $\Psi_1$ can be simplified by $\vec{c}_0 = 0$, and the semantic rule $[\![x - y]\!]^{\mathsf{EB}}_\gamma$ is defined the same as $[\![x + y]\!]^{\mathsf{EB}}_\gamma$ because $z = x - y$ iff $x = z + y$.

- The semantic rule $[\![x\langle y\rangle]\!]^{\mathsf{EB}}_\gamma$ follows the result given in Section 4.3, namely, $(\vec{b}, \vec{b}', p_{i_1}, \cdots, p_{i_k})$ is a solution of $\Psi^\dagger_{\mathcal{S}}$ iff the probability $\mathsf{Pr}_{\mathcal{S}}(\vec{b}, \vec{b}')$ is $2^{-\sum_{j=1}^{k} t_{i_j} \cdot p_{i_j}}$. Note that $t_{i_j}$'s are constants and $\Psi^{\mathsf{bn}}_{\mathcal{S}}$ is added to boost MILP solving.

Lemma 7.1. *Suppose $[\![e]\!]^{\mathsf{EB}}_\gamma = (\Psi, \vec{b}, \varrho)$ with $\Psi \neq \emptyset$. The assignment $\{b_1, \cdots, b_m, p_1, \cdots, p_n\}$ is a solution of $\Psi$ iff the probability of $\{b_1, \cdots, b_i\}$ being bit-level differences of the operands and result of $e$ is $2^{-\varrho[p_1,\cdots,p_n]}$, where $\varrho[p_1, \cdots, p_n]$ denotes the value of $\varrho$ under the assignment $\{p_1, \cdots, p_n\}$ of the Boolean variables for encoding probabilities, and $\{b_{i+1}, \cdots, b_m\}$ is the assignment of the auxiliary Boolean variables if exist.*

### 7.2 Extended Bit-wise Differential Denotational Semantics for Statements

The (extended bit-wise differential) denotational semantics of a statement $S$ is given by $[\![S]\!]^{\mathsf{EB}}$ that maps each state $\gamma \in \Gamma$ to a triple $(\Psi, \gamma', \varrho)$, denoted by $[\![S]\!]^{\mathsf{EB}}_\gamma$, where $\Psi$, $\gamma'$ and $\varrho$ are the same as above. The extended bit-wise differential semantic rules for statements in EasyBC are similar to those word-wise ones, where $[\![S]\!]^{\mathsf{EB}}_\gamma = (\Psi, \gamma', 0)$ if $[\![S]\!]^{\mathsf{B}}_\gamma = (\Psi, \gamma', \emptyset)$, except for

$$\frac{[\![e]\!]^{\mathsf{EB}}_\gamma = (\Psi, \vec{b}, \varrho)}{[\![x = e]\!]^{\mathsf{EB}}_\gamma = (\Psi, \gamma[x/\vec{b}], \varrho)}$$

$$\frac{\begin{array}{c}\tau_0\ \ f(\tau_1\ x_1, \cdots, \tau_m\ x_m)\{S_1; \cdots; S_n; \texttt{return}\ y;\} \qquad \gamma_0 = \gamma[x_1/\gamma(y_1)] \cdots [x_m/\gamma(y_m)] \\ [\![S_1]\!]^{\mathsf{EB}}_{\gamma_0} = (\Psi_1, \gamma_1, \varrho_1),\ \cdots,\ [\![S_n]\!]^{\mathsf{EB}}_{\gamma_{n-1}} = (\Psi_n, \gamma_n, \varrho_n)\end{array}}{[\![x = g(y_1, \cdots, y_m)]\!]^{\mathsf{EB}}_\gamma = (\bigcup_{i=1}^n \Psi_i,\ \gamma[x/\gamma_n(y)],\ \sum_{i=1}^n \varrho_i)}$$

Intuitively, the semantic rule $[\![x = g(y_1, \cdots, y_m)]\!]^{\mathsf{EB}}_\gamma$ sums up the expressions $\varrho_i$'s of the statements $S_i$'s in the function body because of $\prod_{i=1}^n 2^{-\varrho_i} = 2^{-\sum_{i=1}^n \varrho_i}$.

### 7.3 Extended Bit-wise Resistance Evaluation

To evaluate the resistance of the program $P$ in an extended bit-wise manner, we define the (extended bit-wise) semantics $[\![P]\!]^{\mathsf{EB}}$ of the program $P$ using its `fn` function $f$ as follows:

$$[\![P]\!]^{\mathsf{EB}} = [\![\texttt{uints}[n]\ f(\texttt{uints}[n_1]\ k,\ \texttt{uints}[n]\ txt)\{S_1; \cdots; S_n; \texttt{return}\ y;\}]\!]^{\mathsf{EB}} = (\Psi, \gamma_n, \varrho)$$

where $\Psi = \bigcup_{i=1}^n \Psi_i$, $\varrho = \sum_{i=1}^n \varrho_i$, $[\![S_1]\!]^{\mathsf{B}}_{\gamma_0} = (\Psi_1, \gamma_1, \varrho_1), \cdots, [\![S_n]\!]^{\mathsf{B}}_{\gamma_{n-1}} = (\Psi_n, \gamma_n, \varrho_n)$ and $\gamma_0$ is an initial state mapping each bit of array elements of $txt$ to a fresh Boolean variable. We get that:

Theorem 7.2. *Let $[\![P]\!]^{\mathsf{EB}} = (\Phi, \gamma, \varrho)$ and $u$ be the minimum value of $\varrho$ subject to the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s \cdot n-1} \gamma(txt, i) \geq 1\}$. The* MaxEDCP *of the program $P$ is no greater than $2^{-u}$.*

Similar to the decomposition approach given in Proposition 5.3, we have

Proposition 7.3. *Let $u_1$ and $u_2$ be the* MaxEDCP *of the first $s_1$-round and the subsequent $s_2$-round differential characteristics. Then, $u_1 \cdot u_2$ is an upper bound of the* MaxEDCP *of $(s_1 + s_2)$-round differential characteristics.*

By Theorem 7.2 and Proposition 7.3, we have the following corollaries.

Corollary 7.4. *Let $[\![P]\!]^{\mathsf{EB}} = (\Phi, \gamma, \varrho)$. If $\{\varrho < \mathscr{b}\} \cup \Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ is unsatisfiable, then the program $P$ with block size $\mathscr{b}$ is resistant against differential cryptanalysis.*

Table 3. Statistics of NIST candidates.

| Name | EasyBC | | | C/C++ | |
|---|---|---|---|---|---|
| | $\text{LOC}_{\text{bw}}$ | $\text{LOC}_{\text{ww}}$ | Time | LOC | Time |
| ASCON($p^a$) [Dobraunig et al. 2016]♦ | 55 | N/A | 27.7 | **42** | 1.4 |
| ELEPHANT [Beyne et al. 2020]♦ | **37** | N/A | 65.3 | 69 | 6.6 |
| GIFT-COFB-128 [Banik et al. 2020] | **34** | N/A | 23.9 | 81 | 3.0 |
| GRAIN(AEAD) [Hell et al. 2021]⋆ | **99** | N/A | 0.1 | 146 | 0.1 |
| ISAP(v2.0) [Dobraunig et al. 2020]♦ | **72** | N/A | 141.8 | 94 | 0.0 |
| PHOTON(Beetle) [Bao et al. 2019]♦ | - | **51** | 24.4 | 104 | 1.8 |
| ROMULUS-128 [Iwata et al. 2020] | **57** | N/A | 32.3 | 113 | 0.2 |
| SPARKLE256(slim) [Beierle et al. 2019]♦ | 39 | N/A | 0.3 | **19** | 0.0 |
| TinyJAMBU [Wu and Huang 2019]♦ | **15** | N/A | 143.3 | 18 | 0.5 |
| XOODYAK [Daemen et al. 2020]♦ | **38** | N/A | 21.1 | 63 | 0.1 |

Time is in milliseconds. N/A: word-wise is not applicable. -$\mathscr{b}$ gives the block size required for security evaluation of block ciphers. $\text{LOC}_{\text{bw}}$/$\text{LOC}_{\text{ww}}$: LOC of bit-/word-wise implementation. ♦ indicates key-less permutations and ⋆ indicates stream cipher where no block size is given.

Table 4. Statistics of other block ciphers.

| Name | EasyBC | | | C/C++ | |
|---|---|---|---|---|---|
| | $\text{LOC}_{\text{bw}}$ | $\text{LOC}_{\text{ww}}$ | Time | LOC | Time |
| AES-128 [Daemen and Rijmen 1999] | - | **71** | 9.0 | 106 | 12.5 |
| DES-64 [Fox 2000] | **50** | N/A | 4.0 | 77 | 2.4 |
| GIFT-64 [Banik et al. 2017] | **34** | N/A | 6.5 | 63 | 1.9 |
| KLEIN-64 [Gong et al. 2011] | - | **64** | 8.1 | 97 | 0.5 |
| LBLOCK-64 [Wu and Zhang 2011] | - | **44** | 6.0 | 78 | 1.7 |
| MIBS-64 [Izadi et al. 2009] | - | **37** | 13.9 | 69 | 0.3 |
| PICCOLO-64 [Shibutani et al. 2011] | - | **85** | 7.4 | 90 | 84.5 |
| PRESENT-64 [Bogdanov et al. 2007] | **28** | N/A | 8.1 | 87 | 0.6 |
| RECTANGLE-64 [Zhang et al. 2015] | **28** | N/A | 6.6 | 80 | 1.6 |
| SIMON-32 [Beaulieu et al. 2015] | **35** | N/A | 3.0 | 46 | 0.8 |
| SIMON-48 [Beaulieu et al. 2015] | **35** | N/A | 5.9 | 46 | 7.1 |
| SKINNY-64 [Beierle et al. 2016] | **38** | N/A | 10.7 | 67 | 0.3 |
| TWINE-64 [Suzaki et al. 2012] | - | **43** | 9.8 | 61 | 4.4 |

Corollary 7.5. *If the $s$-round cipher $P$ can be partitioned into $\frac{s}{s'}$ identical $s'$-round ciphers $P'$ such that $[\![P']\!]^{\text{EB}} = (\Phi, \gamma, \varrho)$ and $\{\varrho < \frac{s' \cdot \mathscr{b}}{s}\} \cup \Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ is unsatisfiable, then the program $P$ with block size $\mathscr{b}$ is resistant against differential cryptanalysis.*

## 8 EVALUATION

Our approach is implemented as an open-source tool. As shown in Figure 5, it utilizes the SMT solver Z3 for computing the branch number and solving MaxSMT problems and Gurobi [Gurobi Optimization 2018] for solving MILP. In general, EasyBC iteratively increases the round number from 1 (cf. Corollary 5.4 and Corollary 7.4). It also partitions an $s$-round cipher to the maximum number of identical $s'$-round ciphers and iteratively increases the round number $s'$ (cf. Corollary 5.5 and Corollary 7.5), until the resistance is proved. The tool is designed to be modular and extensible, where each semantic rule has an API wrapper and alternative generation methods can be easily chosen and added. We also incorporate the bounding condition [Matsui 1994; Zhang et al. 2018] into MILP to prune search space which often improves the overall MILP solving. (The detail is given in [Sun et al. 2023, Section E.4].)

All the experiments were conducted on a machine with two Intel Xeon Gold 5118 CPUs (12 cores, 2.30GHz), 64-bit Ubuntu 20.04 LTS, and 128GB RAM. The number of threads for Gurobi is set to 16.

### 8.1 Expressiveness of EasyBC

To evaluate the expressiveness of EasyBC, we implement 23 realistic cryptographic primitives with EasyBC, consisting of all the 10 finalists of the NIST lightweight cryptography standardization process [NIST 2023] and 13 commonly used block ciphers, covering both SPN ciphers (e.g., AES, PRESENT, and GIFT-COFB) and BFN ciphers (e.g., DES, LBLOCK, TWINE). We stress that we focus on the underlying primitives (i.e., block ciphers and key-less permutations) rather than the full authenticated encryption, message authentication, or hash protocols in the NIST finalists.

The physical source lines of code (LOC [Nguyen et al. 2007]) counted by cloc [Danial 2021] are reported in Tables 3 and 4. We report LOC of word-wise (when available, or otherwise bit-wise) EasyBC implementations. As a comparison, we also report LOC of the C/C++ reference implementations of the NIST finalists and (randomly selected) GitHub open-source C/C++ implementations of other block ciphers. To some extent, a smaller number (highlighted in **bold**) indicates that the language is more succinct in implementing cryptographic primitives.

We observe that EasyBC is sufficiently expressive to easily implement all the 23 cryptographic primitives either in word-wise or bit-wise fashion. More specifically, when cryptographic primitives can be implemented in a word-wise fashion (i.e., PHOTON, AES, KLEIN, LBLOCK, MIBS,

Table 5. Results of the word-wise approach, where ($n$) indicates the number of entire rounds of the cipher and #AS denotes the lower bound of the minimum number of active S-boxes.

| **Rounds** | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AES (14) | **#AS** | 1 | 5 | 9 | **25** | 26 | 30 | 34 | 50 | 51 | 55 | 59 | 75 | 76 | 80 | N/A | | | | | | | | | | |
| | **Time** | 0s | 0s | 0s | **0s** | 0s | 0s | 1s | 0s | 0s | 0s | 1s | 0s | 0s | 0s | N/A | | | | | | | | | | |
| KLEIN (12) | **#AS** | 1 | 5 | 8 | 15 | 16 | 20 | 23 | 30 | 31 | **35** | 38 | 45 | N/A | | | | | | | | | | | | |
| | **Time** | 1s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | **0s** | 0s | 0s | N/A | | | | | | | | | | | | |
| LBLOCK (32) | **#AS** | 0 | 1 | 2 | 3 | 4 | 6 | 8 | 11 | 14 | 18 | 22 | 24 | 27 | 30 | **32** | 35 | 36 | 39 | 41 | 44 | 45 | 48 | 50 | 53 | 54 |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | **1s** | 1s | 1s | 3s | 3s | 2s | 6s | 5s | 62s | 62s | 80s |
| MIBS (32) | **#AS** | 0 | 1 | 2 | 5 | 6 | 7 | 8 | 11 | 12 | 13 | 14 | 17 | 18 | 19 | 20 | 23 | 24 | 25 | 26 | 29 | 30 | 31 | **32** | 35 | 36 |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | 0s | 1s | 1s | 1s | 3s | 4s | 2s | 4s | 2s | 3s | 2s | 3s | 4s | **5s** | 4s | 7s |
| PHOTON (12) | **#AS** | 1 | 9 | 17 | 81 | 82 | 90 | 98 | 162 | 163 | 171 | 179 | 243 | N/A | | | | | | | | | | | | |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | N/A | | | | | | | | | | | | |
| PICCOLO (25) | **#AS** | 0 | 5 | 10 | 15 | 20 | 30 | **35** | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 | 105 | 110 | 115 | 120 | 125 |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | **0s** | 1s | 0s | 1s | 0s | 1s | 3s | 6s | 4s | 6s | 28s | 30s | 31s | 38s | 187s | 93s | 122s | 289s | 560s |
| TWINE (36) | **#AS** | 0 | 1 | 2 | 3 | 4 | 6 | 8 | 11 | 14 | 18 | 22 | 24 | 27 | 30 | **32** | 35 | 36 | 39 | 41 | 44 | 45 | 48 | 50 | 53 | 54 |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | 0s | 0s | 1s | **0s** | 1s | 2s | 2s | 3s | 2s | 5s | 6s | 26s | 79s | 72s |

PICCOLO, and TWINE), their EasyBC implementations always require considerably less code than the baselines. The main reason is that EasyBC provides high-level constructs of P-box and matrix-vector products for permutations and linear transformations. For cryptographic primitives that are implemented in a bit-wise fashion with EasyBC, their EasyBC implementations also require significantly less code than their baselines except for ASCON and SPARKLE, due to the following reasons. (1) One 320-bit block is stored in five 64-bit variables in the reference implementation of ASCON each of which is permuted, while one 320-bit block is stored in one Boolean array in the EasyBC implementation so that the 320-bit Boolean array has to be split before the permutation and merged after permutation. (2) The C++ reference implementation of SPARKLE uses function-like macro definitions and thus is more succinct.

**Results of** EasyBC **interpreter.** For each realistic cryptographic primitive, we randomly generate 100 inputs to EasyBC programs and binary executables of C/C++ programs. We run the EasyBC program (using our interpreter) and the binary executable for each input and record the output and the execution time. The outputs of the respective programs have been compared with 100% match, which validates the semantics and the interpreter of EasyBC, as well as EasyBC programs. The average execution times over 100 inputs are reported in Tables 3 and 4. While executing via our interpreter is less efficient, it is acceptable for testing EasyBC programs.

## 8.2 Effectiveness of EasyBC

To evaluate the effectiveness of EasyBC, we first compare the performance of various alternative methods to generate MILP. According to our experiment results (cf. [Sun et al. 2023, Section E]), we select the optimal modeling methods for cryptographic primitives, i.e., $\Psi^1_{2,3}$ and $\Psi^1_M$ are used for modeling the modular addition, substitution, XOR and matrix-vector product respectively in the word-wise approach (cf. Figure 6); $\psi^2_{\oplus}$, $\Psi^2_{\mathtt{M,i,h}}$ and the technique of [Boura and Coggia 2020, Alg. 2, Alg. 3 and Proposition 3] are used for modeling XOR, matrix-vector product and constructing $\Psi^3_S$ in the bit-wise approach (cf. Figure 8 and Section 4.2); the technique of [Sasaki and Todo 2017] is used for constructing $\Psi^4_S$ in the extended bit-wise approach (cf. Section 4.3). We remark that there is no consensus on the security of key-less permutations yet, and they are used to evaluate the performance of EasyBC instead of security evaluation.

*8.2.1 Word-wise Approach.* The word-wise approach is evaluated on all the word-wise implementations. The results are reported in Table 5 up to 25 rounds which suffice to prove security. The results for the entire rounds are given in [Sun et al. 2023, Section E.1].

Table 6. Results of the bit-wise approach with the bounding condition, where Timeout is 24 hours.

| Rounds | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ASCON (12) | **#AS** | 1 | 4 | 15 | N/A | | | | | | | | | | | |
| | **Time** | 0s | 1061s | 1573s | Timeout | | | | | | | | | | | |
| DES (16) | **#AS** | 0 | 1 | 2 | 4 | 6 | 8 | 9 | 12 | 12 | N/A | | | | | |
| | **Time** | 0s | 0s | 0s | 6s | 61s | 355s | 1073s | 27294s | 57344s | Timeout | | | | | |
| ELEPHANT (80) | **#AS** | 1 | 2 | 4 | 6 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | N/A | | |
| | **Time** | 0s | 0s | 4s | 5s | 50s | 394s | 1653s | 1592s | 2994s | 4353s | 18294s | 20404s | Timeout | | |
| GIFT-COFB (40) | **#AS** | 1 | 2 | 3 | 5 | 7 | 10 | 13 | 17 | 19 | 21 | N/A | | | | |
| | **Time** | 0s | 0s | 1s | 7s | 7s | 38s | 775s | 2222s | 6725s | 77870s | Timeout | | | | |
| GIFT (28) | **#AS** | 1 | 2 | 3 | 5 | 7 | **10** | 13 | 16 | 18 | 20 | 22 | 24 | 26 | N/A | |
| | **Time** | 0s | 0s | 1s | 1s | 2s | **3s** | 7s | 52s | 43s | 136s | 518s | 846s | 25561s | Timeout | |
| PRESENT (31) | **#AS** | 1 | 2 | **4** | 6 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | N/A |
| | **Time** | 0s | 0s | **0s** | 1s | 6s | 8s | 15s | 65s | 95s | 539s | 1884s | 10271s | 38907s | 50931s | Timeout |
| RECTANGLE (25) | **#AS** | 1 | 2 | 3 | 4 | 6 | **8** | 11 | 13 | 15 | 17 | 19 | 21 | 23 | 25 | 27 |
| | **Time** | 1s | 0s | 0s | 1s | 8s | **21s** | 6s | 436s | 63s | 603s | 1135s | 1221s | 2841s | 36333s | 37860s |
| ROMULUS (40) | **#AS** | 1 | 2 | **5** | 8 | 12 | 16 | 26 | 36 | 41 | N/A | | | | | |
| | **Time** | 0s | 0s | **5s** | 48s | 87s | 378s | 1526s | 17266s | 4043s | Timeout | | | | | |
| SKINNY (36) | **#AS** | **1** | 2 | 5 | 8 | 12 | 16 | 26 | 36 | 41 | 46 | 51 | 55 | N/A | | |
| | **Time** | **0s** | 0s | 1s | 2s | 15s | 24s | 78s | 967s | 2041s | 3610s | 15502s | 26989s | Timeout | | |
| SPARKLE (7) | **#AS** | 1 | 2 | 5 | 6 | 7 | N/A | | | | | | | | | |
| | **Time** | 2s | 44s | 1624s | 4416s | 17592s | Timeout | | | | | | | | | |

We observe that the execution time (i.e., the MILP solving time) in seconds (s) increases with the round number. The execution time of PICCOLO is considerably higher than that of the others when the round number is large (e.g., ≥13), because PICCOLO generates more constraints. For instance, the number of constraints for the 20-round LBLOCK and TWINE are both 481, while it is 641 for the 20-round PICCOLO. However, the large round number is not always necessary, as the security may have been proved with a small round number (see below).

For block ciphers, $p = 2^{-6}$ and $\mathscr{b} = 128$ for AES; $p = 2^{-2}$ and $\mathscr{b} = 64$ for KLEIN, LBLOCK, MIBS, PICCOLO and TWINE; $p = 2^{-2}$ and $\mathscr{b} = 128$ for PHOTON, where the maximum probability $p$ of the involved S-boxes that are arrays is computed by enumeration. By Corollary 5.4, EasyBC proved that AES (resp. KLEIN, LBLOCK, MIBS, PHOTON, PICCOLO and TWINE) is resistant when the round number $s$ is 4 (resp. 10, 15, 23, 4, 7 and 15), as highlighted in **boldface** in Table 5.

*8.2.2 Bit-wise Approach.* The bit-wise approach is evaluated on all bit-wise implementations with S-boxes. (The word-wise implementations are excluded.) The results are given in Table 6 up to 15 rounds within 24 hours, where the execution time is the MILP solving time.

Unsurprisingly, we observe that the execution time increases very quickly with the round number due to the blow-up of constraints. For instance, the numbers of constraints and involved variables of ASCON are 6,913 and 1,408 respectively for 1 round, but become 13,825 and 2,496 (resp. 20,737 and 3,584) for 2 (resp. 3) rounds.

For block ciphers, $p = 2^{-2}$ and $\mathscr{b} = 64$ for DES, PRESENT, RECTANGLE and SKINNY; $p = 2^{-1.415}$ and $\mathscr{b} = 128$ for GIFT-COFB; $p = 2^{-1.415}$ and $\mathscr{b} = 64$ for GIFT; and $p = 2^{-2}$ and $\mathscr{b} = 128$ for ROMULUS. We found that by Corollary 5.4, EasyBC cannot prove that DES (resp. GIFT-COFB, GIFT, PRESENT, RECTANGLE, ROMULUS and SKINNY) is resistant using the lower bounds of numbers of active S-boxes reported in Table 6. Fortunately, by Corollary 5.5, EasyBC proved that GIFT (resp. PRESENT, RECTANGLE, ROMULUS and SKINNY) is resistant with 6 (resp. 3, 6, 3 and 1) rounds, as highlighted in **boldface** in Table 6. Note that the resistance of GIFT-COFB cannot be proved here which will be done by applying the extended bit-wise approach later, while DES is indeed vulnerable to differential cryptanalysis [Biham and Shamir 1990] and EasyBC can return the differential characteristics up to 9 rounds within 24-hour time limit.

We also compared the MILP solving time with/without bounding conditions. Adding the bounding condition often (5 out of 8) improves the efficiency by 1 to 3 times, but does not necessarily improve (and sometimes even worsens) the efficiency (3 out of 8). The results without the bounding condition are given in [Sun et al. 2023, Section E.5].

Table 7. Results of the extended bit-wise approach with the bounding condition, where Pr denotes the upper bound of the probability of optimal differential characteristics, and Timeout is 24 hours.

| **Rounds** $r$ | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GIFT-COFB (40) | Pr | $2^{-1.415}$ | $2^{-3.415}$ | $2^{-7}$ | $2^{-11.415}$ | $\mathbf{2^{-17}}$ | $2^{-22.415}$ | $2^{-28.415}$ | N/A | | | | | | | |
| | **Time** | 0s | 6s | 9s | 124s | **1297s** | 5811s | 11538s | Timeout | | | | | | | |
| SIMON-32 (32) | Pr | $2^{0}$ | $\mathbf{2^{-2}}$ | $2^{-4}$ | $2^{-6}$ | $2^{-8}$ | $2^{-12}$ | $2^{-14}$ | $2^{-18}$ | $2^{-20}$ | $2^{-26}$ | $2^{-30}$ | $2^{-34}$ | $2^{-36}$ | $2^{-38}$ | $2^{-40}$ |
| | **Time** | 0s | **0s** | 0s | 0s | 0s | 0s | 1s | 4s | 3s | 7s | 84s | 57s | 820s | 191s | 6327s |
| SIMON-48 (36) | Pr | $2^{0}$ | $2^{-2}$ | $\mathbf{2^{-4}}$ | $2^{-6}$ | $2^{-8}$ | $2^{-12}$ | $2^{-14}$ | $2^{-18}$ | $2^{-20}$ | $2^{-26}$ | $2^{-30}$ | $2^{-36}$ | $2^{-38}$ | $2^{-44}$ | $2^{-46}$ |
| | **Time** | 0s | 0s | **1s** | 0s | 0s | 1s | 3s | 5s | 14s | 16s | 101s | 95s | 402s | 505s | 53920s |
| Alzette (12) | Pr | $2^{0}$ | $2^{-1}$ | $2^{-2}$ | $2^{-6}$ | $2^{-10}$ | $2^{-18}$ | N/A | | | | | | | | |
| | **Time** | 0s | 1s | 2s | 75s | 221s | 3897s | Timeout | | | | | | | | |

*8.2.3 Extended Bit-wise Approach.* The capability of the extended bit-wise approach is evaluated on all the bit-wise implementations of block ciphers and the S-box of SPARKLE (i.e., 4-round Alzette) that cannot be proved or analyzed before. The results are given in Table 7 up to 15 rounds within 24 hours, where the execution time is the MILP solving time.

Unsurprisingly, the execution time of the extended bit-wise approach is longer than that of the bit-wise approach. For instance, on the 7-round GIFT-COFB, the execution time of the extended bit-wise approach is 11,538s while the execution time of the bit-wise approach is 1,074s. This is because probabilities are explicitly encoded using additional Boolean variables in the extended bit-wise approach, resulting in more difficult MILP instances.

Note the block size: GIFT-COFB $\mathscr{b} = 128$, SIMON-32 $\mathscr{b} = 32$, and SIMON-48 $\mathscr{b} = 48$. By Corollary 7.5, EasyBC prove that GIFT-COFB (resp. SIMON-32 and SIMON-48) is resistant with 5 (resp. 2 and 3) rounds.

## 9 RELATED WORK

[Matsui 1994] proposed the first algorithm for automated resistance analysis of block ciphers against differential cryptanalysis. To further enhance efficiency, various heuristics have been proposed to reduce the search space [Aoki et al. 1997; Bao et al. 2014; Biryukov and Nikolić 2010; Ji et al. 2021]. However, these heuristics generally rely on cipher-specific optimizations, necessitating sophisticated programming skills. Additionally, creating highly reusable code that can be easily adapted for different ciphers is non-trivial [Zhang et al. 2018].

In recent years, a more promising approach based on MILP has been developed. [Mouha et al. 2011] proposed to determine the lower bound of the minimum number of active S-boxes via MILP solving. As an early attempt, it only considered the word-wise modeling for the XOR operation, S-box, and linear transformation. To partially lift this limitation, [Sun et al. 2013] introduced the bit-wise modeling. To precisely characterize S-boxes in MILP, [Sun et al. 2014a] proposed to construct IL constraints from the H-representation of the S-box DDT and reduced the number of constraints via a greedy algorithm. Later, Sun *et al.* [Sun et al. 2014b] proposed a bit-wise modeling method for the AND operation and extended the method of [Sun et al. 2014a] to directly bound the MaxEDCP by encoding the probabilities between input and output differences of S-boxes in IL constraints.

Since then, plenty of modeling methods for specific operations have been proposed, aimed at improving efficiency and applicability. [Sasaki and Todo 2017] proposed an MILP-based algorithm to reduce the number of constraints obtained from the H-representation of the S-box DDT. [Abdelkhalek et al. 2017] proposed to construct IL constraints from the minimized product-of-sum representation of S-box DDT instead of the H-representation. Along this line, [Boura and Coggia 2020; Li and Sun 2022; Sun 2021; Udovenko 2021] generate more diverse constraints from the S-box DDT, allowing the number of the resulting constraints to be further reduced by applying greedy or MILP-based algorithms; [Cui et al. 2016; Li et al. 2019; Sasaki and Todo 2017; Yin et al. 2017] proposed a new modeling for the XOR operation; [Boura and Coggia 2020; Ilter and Selçuk 2021;

Zhang and Zhang 2018] proposed another modeling for linear transformation; [Fu et al. 2016] proposed a modeling method for the modular addition operation; and [Chen et al. 2015; Liu et al. 2017; Wang et al. 2018] proposed a modeling method for the rotation-AND operation. Besides new modeling methods for specific operations, optimization strategies have also been proposed. [Zhang et al. 2018] incorporated the bounding condition of [Matsui 1994] into the MILP-based method; [Zhou et al. 2019] proposed partitioning all possible differential characteristics into subsets, each of which is analyzed by the MILP-based method.

Another direction is to resort to SAT/SMT solving. [Mouha and Preneel 2013] proposed the first SAT/SMT modeling for the XOR, modular addition, and rotation operations. It has been extended to handle a specific permutation [Aumasson et al. 2014], the AND operation and rotation with constants [Kölbl et al. 2015], independent modular addition [Song et al. 2016], S-boxes [Liu et al. 2021; Sun et al. 2018], and modular addition with constants [Azimi et al. 2022]. In contrast to the MILP-based method which determines the lower bound of the minimum number of active S-boxes or the upper bound of MaxEDCP, SAT/SMT-based methods can only verify whether a given number is a bound. Recently, both the bounding condition of [Matsui 1994] and MILP-based method have been combined with the SAT/SMT-based method [Makarim and Rohit 2022; Sun et al. 2021] to improve the efficiency. To facilitate the comparison of all the above MILP/SAT/SMT-based approaches, a summary is given in [Sun et al. 2023, Section F].

Despite significant progress in the field, existing works primarily concentrate on ad-hoc modeling methods designed for specific operations, but do not offer systematic methods for determining word-wise or bit-wise branch numbers and encoding probabilities between input and output differences. There is a lack of language support, unified computational approaches and full automation. This limitation forces cryptanalysts to individually model each cipher within the tool or create a model generation script for every individual cipher, resulting in a complex, error-prone, and time-consuming process. This work fills this significant gap and makes resistance evaluation against differential cryptanalysis easily accessible to cryptographers.

There are other cryptography-specific languages such as SAW [Carter et al. 2013], Jasmin [Almeida et al. 2017], Vale [Bond et al. 2017], Usuba [Mercadier and Dagand 2019], FaCT [Cauligi et al. 2019], QMVerif [Gao et al. 2022], HOME [Gao et al. 2021], and FISCHER [Liu et al. 2023]. They are designed to ensure functional correctness and/or side-channel security of cryptographic algorithms or implementations, which are considerably different from EasyBC.

## 10 CONCLUSION

We have designed a high-level cryptography-specific language EasyBC for describing block ciphers and presented a rigorous MILP generation procedure from EasyBC programs in the form of differential denotational semantics, leading to a generic and extensible approach for automatically evaluating the resistance of block ciphers written in EasyBC against differential cryptanalysis. We have implemented our approach in an open-source tool and extensively evaluate it on a set of realistic cryptographic primitives, demonstrating its expressivity and capability. In particular, experimental results show that realistic cryptographic primitives can be easily described in EasyBC and their resistance against differential cryptanalysis can be efficiently and effectively proved using our tool. Our tool enables cryptanalysts to easily assess the resistance of block ciphers against differential cryptanalysis in a fully automatic way.

For future research, it would be interesting to improve efficiency by combining recent optimization strategies and to develop analysis approaches for other powerful cryptanalysis (e.g., linear cryptanalysis [Biryukov and De Cannière 2011], impossible differential cryptanalysis [Kim et al. 2003]) based on EasyBC.

## ACKNOWLEDGEMENT

We thank the reviewers of POPL'24 for their constructive and insightful comments. This work is supported by the National Natural Science Foundation of China (NSFC) under Grants No. 62072309 and No. 61872340, CAS Project for Young Scientists in Basic Research (YSBR-040), ISCAS New Cultivation Project (ISCAS-PYFX-202201), overseas grants from the State Key Laboratory of Novel Software Technology, Nanjing University (KFKT2023A04), and Birkbeck BEI School Project (EFFECT).

## DATA AVAILABILITY STATEMENT

The full version of the paper [Sun et al. 2023] contains missing proofs and more experimental results. Our tool is available at https://github.com/S3L-official/EasyBC.

## REFERENCES

Ahmed Abdelkhalek, Yu Sasaki, Yosuke Todo, Mohamed Tolba, and Amr M Youssef. 2017. MILP modeling for (large) s-boxes to optimize probability of differential characteristics. *IACR Transactions on Symmetric Cryptology* (2017), 99–129.

José Bacelar Almeida, Manuel Barbosa, Gilles Barthe, Arthur Blot, Benjamin Grégoire, Vincent Laporte, Tiago Oliveira, Hugo Pacheco, Benedikt Schmidt, and Pierre-Yves Strub. 2017. Jasmin: High-Assurance and High-Speed Cryptography. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1807–1823. https://doi.org/10.1145/3133956.3134078

Kazumaro Aoki, Kunio Kobayashi, and Shiho Moriai. 1997. Best differential characteristic search of FEAL. In *Proceedings of the International Workshop on Fast Software Encryption*. 41–53.

Jean-Philippe Aumasson, Philipp Jovanovic, and Samuel Neves. 2014. Analysis of NORX: Investigating Differential and Rotational Properties. In *Proceedings of the 3rd International Conference on Cryptology and Information Security in Latin America*. 306–324.

Seyyed Arash Azimi, Adrián Ranea, Mahmoud Salmasizadeh, Javad Mohajeri, Mohammad Reza Aref, and Vincent Rijmen. 2022. A bit-vector differential model for the modular addition by a constant and its applications to differential and impossible-differential cryptanalysis. *Des. Codes Cryptogr.* 90, 8 (2022), 1797–1855.

Subhadeep Banik, Avik Chakraborti, Tetsu Iwata, Kazuhiko Minematsu, Mridul Nandi, Thomas Peyrin, Yu Sasaki, Siang Meng Sim, and Yosuke Todo. 2020. GIFT-COFB. *IACR Cryptol. ePrint Arch.* (2020), 738.

Subhadeep Banik, Sumit Kumar Pandey, Thomas Peyrin, Yu Sasaki, Siang Meng Sim, and Yosuke Todo. 2017. GIFT: A Small Present - Towards Reaching the Limit of Lightweight Encryption. In *Proceedings of the 19th International Conference on Cryptographic Hardware and Embedded Systems*. 321–345.

Zhenzhen Bao, Avik Chakraborti, Nilanjan Datta, Jian Guo, Mridul Nandi, Thomas Peyrin, and Kan Yasuda. 2019. PHOTON-beetle authenticated encryption and hash family. *NIST Lightweight Compet. Round* (2019), 115.

Zhenzhen Bao, Wentao Zhang, and Dongdai Lin. 2014. Speeding up the search algorithm for the best differential and best linear trails. In *Proceedings of the International Conference on Information Security and Cryptology*. 259–285.

Ray Beaulieu, Douglas Shors, Jason Smith, Stefan Treatman-Clark, Bryan Weeks, and Louis Wingers. 2015. The SIMON and SPECK lightweight block ciphers. In *Proceedings of the 52nd Annual Design Automation Conference*. 175:1–175:6.

Christof Beierle, Alex Biryukov, Luan Cardoso dos Santos, Johann Großschädl, Léo Perrin, Aleksei Udovenko, Vesselin Velichkov, and Qingju Wang. 2020. Alzette: A 64-Bit ARX-box - (Feat. CRAX and TRAX). In *Proceedings of 40th Annual InternationalCryptology Conference*. 419–448.

Christof Beierle, Alex Biryukov, Luan Cardoso dos Santos, Johann Großschädl, Léo Perrin, Aleksei Udovenko, Vesselin Velichkov, Qingju Wang, and Alex Biryukov. 2019. Schwaemm and Esch: lightweight authenticated encryption and hashing using the sparkle permutation family. *NIST round* 2 (2019).

Christof Beierle, Jérémy Jean, Stefan Kölbl, Gregor Leander, Amir Moradi, Thomas Peyrin, Yu Sasaki, Pascal Sasdrich, and Siang Meng Sim. 2016. The SKINNY family of block ciphers and its low-latency variant MANTIS. In *Proceedings of the Annual International Cryptology Conference*. 123–153.

Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. 2020. Dumbo, Jumbo, and Delirium: Parallel Authenticated Encryption for the Lightweight Circus. *IACR Trans. Symmetric Cryptol.* (2020), 5–30.

Eli Biham and Adi Shamir. 1990. Differential Cryptanalysis of DES-like Cryptosystems. In *Proceedings of the 10th Annual International Cryptology Conference*. 2–21.

Alex Biryukov and Christophe De Cannière. 2011. Linear cryptanalysis for block ciphers. *Encyclopedia of cryptography and security* (2011), 722–725.

Alex Biryukov and Ivica Nikolić. 2010. Automatic search for related-key differential characteristics in byte-oriented block ciphers: Application to AES, Camellia, Khazad and others. In *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*. 322–344.

Nikolaj S. Bjørner and Anh-Dung Phan. 2014. $\nu$Z - Maximal Satisfaction with Z3. In *Proceedings of the 6th International Symposium on Symbolic Computation in Software Science*. 1–9.

Nikolaj S. Bjørner, Anh-Dung Phan, and Lars Fleckenstein. 2015. $\nu$Z - An Optimizing SMT Solver. In *Proceedings of the 21st International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. 194–199.

Andrey Bogdanov. 2010. *Analysis and design of block cipher constructions.* Ph. D. Dissertation. Ruhr University Bochum.

Andrey Bogdanov, Lars R Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew JB Robshaw, Yannick Seurin, and Charlotte Vikkelsoe. 2007. PRESENT: An ultra-lightweight block cipher. In *Proceedings of the International workshop on cryptographic hardware and embedded systems*. 450–466.

Barry Bond, Chris Hawblitzel, Manos Kapritsos, K. Rustan M. Leino, Jacob R. Lorch, Bryan Parno, Ashay Rane, Srinath T. V. Setty, and Laure Thompson. 2017. Vale: Verifying High-Performance Cryptographic Assembly Code. In *Proceedings of the 26th USENIX Security Symposium*, Engin Kirda and Thomas Ristenpart (Eds.). 917–934.

Christina Boura and Daniel Coggia. 2020. Efficient MILP modelings for Sboxes and linear layers of SPN ciphers. *IACR Transactions on Symmetric Cryptology* (2020), 327–361.

Kyle Carter, Adam Foltzer, Joe Hendrix, Brian Huffman, and Aaron Tomb. 2013. SAW: the software analysis workbench. In *Proceedings of the 2013 ACM SIGAda annual conference on High integrity language technology*, Jeff Boleng and S. Tucker Taft (Eds.). ACM, 15–18. https://doi.org/10.1145/2527269.2527277

Sunjay Cauligi, Gary Soeller, Brian Johannesmeyer, Fraser Brown, Riad S. Wahby, John Renner, Benjamin Grégoire, Gilles Barthe, Ranjit Jhala, and Deian Stefan. 2019. FaCT: a DSL for timing-sensitive computation. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*, Kathryn S. McKinley and Kathleen Fisher (Eds.). ACM, 174–189. https://doi.org/10.1145/3314221.3314605

Zhan Chen, Ning Wang, and Xiaoyun Wang. 2015. Impossible Differential Cryptanalysis of Reduced Round SIMON. *IACR Cryptol. ePrint Arch.* (2015), 286.

Tingting Cui, Keting Jia, Kai Fu, Shiyao Chen, and Meiqin Wang. 2016. New Automatic Search Tool for Impossible Differentials and Zero-Correlation Linear Approximations. *IACR Cryptol. ePrint Arch.* (2016), 689.

Joan Daemen, Seth Hoffert, Michaël Peeters, Gilles Van Assche, and Ronny Van Keer. 2020. Xoodyak, a lightweight cryptographic scheme. *IACR Trans. Symmetric Cryptol.* (2020), 60–87.

Joan Daemen and Vincent Rijmen. 1999. AES proposal: Rijndael. (1999).

Albert Danial. 2021. *cloc: v1.92.* https://doi.org/10.5281/zenodo.5760077

Christoph Dobraunig, Maria Eichlseder, Stefan Mangard, Florian Mendel, Bart Mennink, Robert Primas, and Thomas Unterluggauer. 2020. ISAP v2.0. *IACR Trans. Symmetric Cryptol.* (2020), 390–416.

Christoph Dobraunig, Maria Eichlseder, Florian Mendel, and Martin Schläffer. 2016. ASCON v1. 2. *Submission to the CAESAR Competition* (2016).

Dirk Fox. 2000. Data Encryption Standard (DES). *Datenschutz und Datensicherheit* (2000).

Kai Fu, Meiqin Wang, Yinghua Guo, Siwei Sun, and Lei Hu. 2016. MILP-based automatic search algorithms for differential and linear trails for speck. In *Proceedings of the International Conference on Fast Software Encryption*. 268–288.

Pengfei Gao, Hongyi Xie, Fu Song, and Taolue Chen. 2021. A Hybrid Approach to Formal Verification of Higher-Order Masked Arithmetic Programs. *ACM Trans. Softw. Eng. Methodol.* 30, 3 (2021), 26:1–26:42. https://doi.org/10.1145/3428015

Pengfei Gao, Hongyi Xie, Pu Sun, Jun Zhang, Fu Song, and Taolue Chen. 2022. Formal Verification of Masking Countermeasures for Arithmetic Programs. *IEEE Trans. Software Eng.* 48, 3 (2022), 973–1000. https://doi.org/10.1109/TSE.2020.3008852

Zheng Gong, Svetla Nikova, and Yee Wei Law. 2011. KLEIN: a new family of lightweight block ciphers. In *Proceedings of the International Workshop on Radio Frequency Identification: Security and Privacy Issues*. 1–18.

LLC Gurobi Optimization. 2018. Gurobi optimizer reference manual.

Martin Hell, Thomas Johansson, Alexander Maximov, Willi Meier, and Hirotaka Yoshida. 2021. Grain-128AEADv2: Strengthening the Initialization Against Key Reconstruction. In *Proceedings of the 20th International Conference on Cryptology and Network Security*. 24–41.

Howard M Heys. 2002a. A tutorial on linear and differential cryptanalysis. *Cryptologia* (2002), 189–221.

Howard M. Heys. 2002b. A Tutorial on Linear and Differential Cryptanalysis. *Cryptologia* 26, 3 (2002), 189–221.

Howard M. Heys and Stafford E. Tavares. 1996. Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis. *J. Cryptol.* (1996), 1–19.

Murat Burhan Ilter and Ali Aydin Selçuk. 2021. A New MILP Model for Matrix Multiplications with Applications to KLEIN and PRINCE. In *Proceedings of the 18th International Conference on Security and Cryptography*. 420–427.

Tetsu Iwata, Mustafa Khairallah, Kazuhiko Minematsu, and Thomas Peyrin. 2020. Duel of the Titans: The Romulus and Remus Families of Lightweight AEAD Algorithms. *IACR Trans. Symmetric Cryptol.* (2020), 43–120.

Maryam Izadi, Babak Sadeghiyan, Seyed Saeed Sadeghian, and Hossein Arabnezhad Khanooki. 2009. MIBS: A New Lightweight Block Cipher. In *Proceedings of the 8th International Conference on Cryptology and Network Security*. 334–348.

Fulei Ji, Wentao Zhang, and Tianyou Ding. 2021. Improving matsui's search algorithm for the best differential/linear trails and its applications for DES, DESL and GIFT. *Comput. J.* 64, 4 (2021), 610–627.

John B. Kam and George I. Davida. 1979. Structured design of substitution-permutation encryption networks. *IEEE Trans. Comput.* 28, 10 (1979), 747–753.

Jonathan Katz and Yehuda Lindell. 2014. *Introduction to Modern Cryptography, 2nd Edition.* CRC Press.

Jongsung Kim, Seokhie Hong, Jaechul Sung, Sangjin Lee, Jongin Lim, and Soohak Sung. 2003. Impossible differential cryptanalysis for block cipher structures. In *Proceedings of the International Conference on Cryptology in India*. 82–96.

Lars R Knudsen. 1998. Block Ciphers: a survey. In *State of the art in applied cryptography*. Springer, 18–48.

Stefan Kölbl, Gregor Leander, and Tyge Tiessen. 2015. Observations on the SIMON Block Cipher Family. In *Proceedings of the 35th Annual Cryptology Conference*. 161–185.

Xuejia Lai and James L. Massey. 1990. A Proposal for a New Block Encryption Standard. In *Proceedings of the Workshop on the Theory and Application of of Cryptographic Techniques (EUROCRYPT)*. 389–404.

Xuejia Lai, James L Massey, and Sean Murphy. 1991. Markov ciphers and differential cryptanalysis. In *Proceedings of the Workshop on the Theory and Application of of Cryptographic Techniques*. 17–38.

Lingchen Li, Wenling Wu, Yafei Zheng, and Lei Zhang. 2019. The Relationship between the Construction and Solution of the MILP Models and Applications. *IACR Cryptol. ePrint Arch.* (2019), 49.

Ting Li and Yao Sun. 2022. SuperBall: A New Approach for MILP Modelings of Boolean Functions. *IACR Transactions on Symmetric Cryptology* (2022), 341–367.

Mingyang Liu, Fu Song, and Taolue Chen. 2023. Automated Verification of Correctness for Masked Arithmetic Programs. In *Proceedings of the 35th International Conference on Computer Aided Verification (CAV), Part III (Lecture Notes in Computer Science, Vol. 13966)*, Constantin Enea and Akash Lal (Eds.). Springer, 255–280. https://doi.org/10.1007/978-3-031-37709-9_13

Yu Liu, Huicong Liang, Muzhou Li, Luning Huang, Kai Hu, Chenhe Yang, and Meiqin Wang. 2021. STP models of optimal differential and linear trail for S-box based ciphers. *Science China Information Sciences* 64, 5 (2021).

Zhengbin Liu, Yongqiang Li, and Mingsheng Wang. 2017. Optimal Differential Trails in SIMON-like Ciphers. *IACR Trans. Symmetric Cryptol.* 2017 (2017), 358–379.

Mohammad Mahzoun, Liliya Kraleva, Raluca Posteuca, and Tomer Ashur. 2022. Differential Cryptanalysis of K-Cipher. In *IEEE Symposium on Computers and Communications*. 1–7.

Rusydi H Makarim and Raghvendra Rohit. 2022. Towards Tight Differential Bounds of ASCON: A Hybrid Usage of SMT and MILP. *IACR Transactions on Symmetric Cryptology* (2022), 303–340.

Mitsuru Matsui. 1994. On correlation between the order of S-boxes and the strength of DES. In *Proceedings of the Workshop on the Theory and Application of of Cryptographic Techniques*. 366–375.

Darius Mercadier and Pierre-Évariste Dagand. 2019. Usuba: high-throughput and constant-time ciphers, by construction. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*, Kathryn S. McKinley and Kathleen Fisher (Eds.). ACM, 157–173. https://doi.org/10.1145/3314221.3314636

Nicky Mouha and Bart Preneel. 2013. Towards finding optimal differential characteristics for ARX: Application to Salsa20. *Cryptology ePrint Archive* (2013).

Nicky Mouha, Qingju Wang, Dawu Gu, and Bart Preneel. 2011. Differential and linear cryptanalysis using mixed-integer linear programming. In *Proceedings of the International Conference on Information Security and Cryptology*. 57–76.

Vu Nguyen, Sophia Deeds-Rubin, Thomas Tan, and Barry Boehm. 2007. A SLOC counting standard. In *Cocomo ii forum*, Vol. 2007. 1–16.

NIST. 2023. Finalists of NIST lightweight cryptography standardization process. https://csrc.nist.gov/Projects/lightweight-cryptography/finalists.

Kaisa Nyberg. 1996. Generalized feistel networks. In *Proceedings of the International conference on the theory and application of cryptology and information security*. 91–104.

Yu Sasaki and Yosuke Todo. 2017. New algorithm for modeling S-box in MILP based differential and division trail search. In *Proceedings of the International Conference for Information Technology and Communications*. 150–165.

Claude E. Shannon. 1949. Communication theory of secrecy systems. *Bell System Technical Journal* 28, 4 (1949), 656–715.

Kyoji Shibutani, Takanori Isobe, Harunaga Hiwatari, Atsushi Mitsuda, Toru Akishita, and Taizo Shirai. 2011. Piccolo: an ultra-lightweight blockcipher. In *Proceedings of the International workshop on cryptographic hardware and embedded systems*. 342–357.

Taizo Shirai, Kyoji Shibutani, Toru Akishita, Shiho Moriai, and Tetsu Iwata. 2007. The 128-Bit Blockcipher CLEFIA (Extended Abstract). In *Proceedings of the 14th International Workshop on Fast Software Encryption (FSE), Revised Selected Papers*. 181–195.

Ling Song, Zhangjie Huang, and Qianqian Yang. 2016. Automatic Differential Analysis of ARX Block Ciphers with Application to SPECK and LEA. In *Proceedings of the 21st Australasian Conference on Information Security and Privacy.* 379–394.

Ling Sun, Wei Wang, and Meiqin Wang. 2018. More Accurate Differential Properties of LED64 and Midori64. *IACR Trans. Symmetric Cryptol.* 2018, 3 (2018), 93–123.

Ling Sun, Wei Wang, and Meiqin Wang. 2021. Accelerating the Search of Differential and Linear Characteristics with the SAT Method. *IACR Trans. Symmetric Cryptol.* 2021, 1 (2021), 269–315.

Pu Sun, Fu Song, Yuqi Chen, and Taolue Chen. 2023. *EasyBC: A Cryptography-Specific Language for Security Analysis of Block Ciphers against Differential Cryptanalysis (Full version).* Technical Report. https://github.com/S3L-official/EasyBC.

Siwei Sun, Lei Hu, Ling Song, Yonghong Xie, and Peng Wang. 2013. Automatic security evaluation of block ciphers with S-bP structures against related-key differential attacks. In *Proceedings of the International Conference on Information Security and Cryptology.* 39–51.

Siwei Sun, Lei Hu, Meiqin Wang, Peng Wang, Kexin Qiao, Xiaoshuang Ma, Danping Shi, Ling Song, and Kai Fu. 2014b. Towards finding the best characteristics of some bit-oriented block ciphers and automatic enumeration of (related-key) differential and linear characteristics with predefined properties. *Cryptology ePrint Archive* (2014).

Siwei Sun, Lei Hu, Peng Wang, Kexin Qiao, Xiaoshuang Ma, and Ling Song. 2014a. Automatic security evaluation and (related-key) differential characteristic search: application to SIMON, PRESENT, LBlock, DES(L) and other bit-oriented block ciphers. In *Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security.* 158–178.

Yao Sun. 2021. Towards the Least Inequalities for Describing a Subset in Z$_2$n. *IACR Cryptol. ePrint Arch.* (2021), 1084.

Tomoyasu Suzaki, Kazuhiko Minematsu, Sumio Morioka, and Eita Kobayashi. 2012. TWINE: A Lightweight Block Cipher for Multiple Platforms. In *Proceedings of the International Conference on Selected Areas in Cryptography.* 339–354.

Je Sen Teh and Alex Biryukov. 2022. Differential cryptanalysis of WARP. *J. Inf. Secur. Appl.* 70 (2022), 103316.

Aleksei Udovenko. 2021. MILP modeling of Boolean functions by minimum number of inequalities. *Cryptology ePrint Archive* (2021).

Xuzi Wang, Baofeng Wu, Lin Hou, and Dongdai Lin. 2018. Automatic Search for Related-Key Differential Trails in SIMON-like Block Ciphers Based on MILP. In *Proceedings of the 21st International Conference on Information Security*, Liqun Chen, Mark Manulis, and Steve A. Schneider (Eds.). 116–131.

Hongjun Wu and Tao Huang. 2019. TinyJAMBU: A family of lightweight authenticated encryption algorithms. *Submission to the NIST Lightweight Cryptography Standardization Process* (2019).

Shengbao Wu and Mingsheng Wang. 2012. Automatic search of truncated impossible differentials for word-oriented block ciphers. In *Proceedings of the International Conference on Cryptology in India.* 283–302.

Wenling Wu and Lei Zhang. 2011. LBlock: a lightweight block cipher. In *Proceedings of the International conference on applied cryptography and network security.* 327–344.

Jun Yin, Chuyan Ma, Lijun Lyu, Jian Song, Guang Zeng, Chuangui Ma, and Fushan Wei. 2017. Improved cryptanalysis of an ISO standard lightweight block cipher with refined MILP modelling. In *Proceedings of the International Conference on Information Security and Cryptology.* 404–426.

Pei Zhang and Wenying Zhang. 2018. Differential cryptanalysis on block cipher skinny with MILP program. *Security and Communication Networks* 2018 (2018).

Wentao Zhang, Zhenzhen Bao, Dongdai Lin, Vincent Rijmen, Bohan Yang, and Ingrid Verbauwhede. 2015. RECTANGLE: a bit-slice lightweight block cipher suitable for multiple platforms. *Science China Information Sciences* 58, 12 (2015), 1–15.

Yingjie Zhang, Siwei Sun, Jiahao Cai, and Lei Hu. 2018. Speeding up MILP aided differential characteristic search with Matsui's strategy. In *Proceedings of the International Conference on Information Security.* 101–115.

Chunning Zhou, Wentao Zhang, Tianyou Ding, and Zejun Xiang. 2019. Improving the MILP-based security evaluation algorithm against differential/linear cryptanalysis using a divide-and-conquer approach. *IACR Transactions on Symmetric Cryptology* (2019), 438–469.

## A A BRIEF INTRODUCTION OF BFN AND SPN

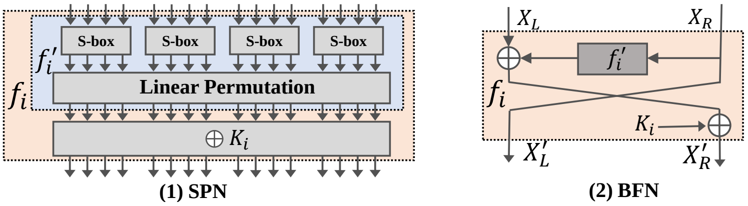
**SPN cipher**. Assume $\mathscr{b} = m\ell$ for some positive integers $m$ and $\ell$. An SPN cipher is built from permutations $\pi_\mathrm{s} : \mathbb{B}^\ell \rightarrow \mathbb{B}^\ell$ and $\pi_\mathrm{p} : \mathbb{B}^\mathscr{b} \rightarrow \mathbb{B}^\mathscr{b}$ where $\pi_\mathrm{s}$ (a.k.a. S-box) is a non-linear transformation to perform a substitution for $\ell$-bitstreams, and $\pi_\mathrm{p}$ (a.k.a. mixing permutation) is a linear transformation. More specifically, for every $1 \leq i \leq \mathscr{r}$ and every input $X = Y^1 \| \cdots \| Y^m$, $\mathsf{Enc}_i(K^i, X)$ is given by $\mathsf{Enc}'_i(X) \oplus K^i$, where $\mathsf{Enc}'_i$ typically is the identify function (used as the first round) or a function built from $\pi_\mathrm{p}$ and $\pi_\mathrm{s}$ such as $\mathsf{Enc}'_i(Y^1 \| \cdots \| Y^m) = \pi_\mathrm{p}(\pi_\mathrm{s}(Y^1) \| \cdots \| \pi_\mathrm{s}(Y^m))$ and $\mathsf{Enc}'_i(Y^1 \| \cdots \| Y^m) = \pi_\mathrm{s}(Y^1) \| \cdots \| \pi_\mathrm{s}(Y^m)$ (used as the final round). For SPN ciphers, S-boxes must

Table 9. The DDT $\mathcal{D}_{\mathcal{S}}$ of the 4-bit S-box in Table 8.

| $\Delta X$ \ $\Delta Y$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 4 | 0 | 4 | 0 | 0 | 0 | 4 | 0 | 0 |
| 2 | 0 | 0 | 0 | 2 | 0 | 4 | 2 | 0 | 0 | 0 | 2 | 0 | 2 | 2 | 2 | 0 |
| 3 | 0 | 2 | 0 | 2 | 2 | 0 | 4 | 2 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 4 | 2 | 2 | 0 | 2 | 2 | 0 | 2 | 0 | 2 | 0 |
| 5 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 4 | 2 | 0 | 0 |
| 6 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 4 | 2 | 0 | 0 | 4 |
| 7 | 0 | 4 | 2 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 4 |
| 8 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 4 | 0 | 2 | 0 | 4 |
| 9 | 0 | 0 | 2 | 0 | 4 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 4 | 0 |
| 10 | 0 | 0 | 2 | 2 | 0 | 4 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 2 | 2 | 0 |
| 11 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 4 | 2 | 2 | 2 | 0 | 2 | 0 | 0 |
| 12 | 0 | 0 | 2 | 0 | 0 | 4 | 0 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 2 | 0 |
| 13 | 0 | 2 | 4 | 2 | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | 2 | 2 | 0 | 0 | 2 | 2 | 0 | 0 |
| 15 | 0 | 4 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 |

be invertible (i.e., $\pi_{\mathrm{s}}$ should be bijective) and linear transformations $\pi_{\mathrm{p}}$ are given by invertible matrices or reordering of $\ell$-bitstreams), thus $\mathsf{Enc}_i^{-1}(K^i, \cdot)$ (hence $\mathsf{Dec}$) can be built trivially.

Figure 10(1) shows one internal round of an SPN cipher. Table 8 shows a 4-bit S-box of PRESENT [Bogdanov et al. 2007] whose DDT is shown in Table 9.



(1) SPN

(2) BFN

Fig. 10. One internal round of SPN and BFN.

Table 8. A 4-bit S-box of PRESENT [Bogdanov et al. 2007].

| input | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| output | 12 | 5 | 6 | 11 | 9 | 0 | 10 | 13 | 3 | 14 | 15 | 8 | 4 | 7 | 1 | 2 |

**BFN cipher**. BFN ciphers have the same low-level building blocks (S-boxes, permutations and key schedule) as SPN ciphers, but differ in the high-level design. Furthermore, although BFN ciphers should be invertible, the involved S-boxes are not necessarily so. More specifically, an input $X \in \mathbb{B}^{\ell}$ to a BFN cipher is separated into two halves, $X_L \in \mathbb{B}^{\frac{\ell}{2}}$ and $X_R \in \mathbb{B}^{\frac{\ell}{2}}$, and $\mathsf{Enc}_i(K^i, X_L \| X_R)$ is given by $X_R \| (\mathsf{Enc}_i'(X_R) \oplus X_L \oplus K^i)$, where $\mathsf{Enc}_i'$ is typically implemented using S-boxes and mixing permutations, similar to SPN ciphers. Figure 10(2) shows one internal round of a BFN cipher. While $\mathsf{Enc}_i'(Y)$ may not be invertible, $\mathsf{Enc}_i^{-1}(K^i, X_L \| X_R)$ can be computed via $X_R \oplus \mathsf{Enc}_i'(X_L) \oplus K^i \| X_L$.

Note that there are other design approaches of iterative block ciphers, e.g., the Lai-Massey scheme [Lai and Massey 1990] and generalized Feistel schemes [Shirai et al. 2007]. Though they are less popular than the SPN and BFN schemes [Bogdanov 2010], our approach is of generic nature and could be applied to block ciphers in the other schemes, thanks to the expressivity of EasyBC.

## B KEY RECOVERING FROM OPTIMAL DIFFERENTIAL CHARACTERISTICS

Consider the optimal $(r - 1)$-round differential characteristic $(\Delta X^0, \cdots, \Delta X^{r-1})$. The attacker first can randomly select a plaintext $X^0$, computes $X^0 \oplus \Delta X^0$ and two ciphertexts $\mathsf{Enc}(K, X^0)$ and $\mathsf{Enc}(K, X^0 \oplus \Delta X^0)$. Then, the attacker will try to solve the following equation for the subkey $K^r$:

$$\mathsf{Enc}_r^{-1}(K^r, \mathsf{Enc}(K, X^0)) \oplus \mathsf{Enc}_r^{-1}(K^r, \mathsf{Enc}(K, X^0 \oplus \Delta X^0)) = \Delta X^{r-1}.$$

The solutions of the equation are candidates of the subkey $K^r$. The attacker can uniformly sample a number of plaintexts and increases the counter of each candidate subkey fulfilling the above equation. The candidate being counted significantly more often than others is regarded as the actual subkey $K^r$, as the differential characteristic $(\Delta X^0, \cdots, \Delta X^{r-1})$ is optimal.

After recovering $K^r$, the attacker can iteratively recover the subkey $K^{i+1}$ according to the optimal $i$-th round differential characteristic $(\Delta X^0, \cdots, \Delta X^i)$ for $i < r - 1$ by solving the following equation,

$$\mathsf{Enc}_{i+1}^{-1}(K^{i+1}, X^{i+1})) \oplus \mathsf{Enc}_{i+1}^{-1}(K^{i+1}, X'^{i+1}) = \Delta X^i,$$

where the outputs $X^{i+1}$ and $X'^{i+1}$ of the $(i + 1)$-round, i.e., the inputs of the $(i + 2)$-round, can be obtained from the ciphertexts $\mathsf{Enc}(K, X^0)$ and $\mathsf{Enc}(K, X^0 \oplus \Delta X^0)$ using the recovered subkeys $K^{i+2}, \cdots, K^r$. The number of plaintexts required to determine the $(i+1)$-round subkey $K^{i+1}$ is proportional to $\frac{1}{\mathsf{Pr}(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^i)}$ for the optimal $i$-th round differential characteristic $(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^i)$ [Heys 2002b].

## C MISSING EXAMPLES

### C.1 Illustrating Example for Section 4.1

Consider the function $f_\oplus : \mathbb{B}^n \times \mathbb{B}^n \to \mathbb{B}^n$ such that $f_\oplus(X^1, X^2) = X^1 \oplus X^2$. We will construct the SMT formula $\phi_{f_\oplus} := \phi_1 \wedge \phi_2$, where

$$\phi_1 := \big((\Delta X^1 \neq 0 \vee \Delta X^2 \neq 0) \wedge Y = X^1 \oplus X^2 \wedge Y \oplus \Delta Y = (X^1 \oplus \Delta X^1) \oplus (X^2 \oplus \Delta X^2)\big),$$

$$\psi_2 := \left\{ \begin{array}{c} (d = c_1 + c_2 + c_3) \\ \wedge \\ \big((\Delta X^1 \neq 0 \wedge c_1 = 1) \vee (\Delta X^1 = 0 \wedge c_1 = 0)\big) \\ \wedge \\ \big((\Delta X^2 \neq 0 \wedge c_2 = 1) \vee (\Delta X^2 = 0 \wedge c_2 = 0)\big) \\ \wedge \\ \big((\Delta Y \neq 0 \wedge c_3 = 1) \vee (\Delta Y = 0 \wedge c_3 = 0)\big) \end{array} \right\},$$

$X^1, X^2, Y, \Delta X^1, \Delta X^2, \Delta Y$ are n-bit unsigned integer variables, and $d, c_1, c_2, c_2$ are 2-bit unsigned integer variables.

The minimized (resp. maximized) value of $d$ subject to the SMT formula $\phi_{f_\oplus}$ is $\mathcal{B}_{\mathsf{ww}}^{\min}(f_\oplus)$ (resp. $\mathcal{B}_{\mathsf{ww}}^{\max}(f_\oplus)$), that is 2 (resp. 3).

### C.2 Illustrating Example for Section 4.3

Consider the S-box $\mathcal{S}$ shown in Table 8 whose DDT is shown in Table 9. There are three nonzero probabilities $\mathsf{Pr}_{\mathcal{S}}(\Delta X, \Delta Y)$ for $(\Delta X, \Delta Y) \in \mathbb{B}^4 \times \mathbb{B}^4$, namely, $2^{-0}$ (i.e., 16/16), $2^{-3}$ (i.e., 2/16) and $2^{-2}$ (i.e., 4/16). Thus, $V = \{2^{-0}, 2^{-3}, 2^{-2}\}$.

Table 10. Part of the extended DDT $\mathcal{D}^{\dagger}_{\mathcal{S}}$ of the 4-bit S-box in Table 8.

| $\Delta X$ $b_1$ $b_2$ \ $\Delta Y$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 0 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 0 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 1 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 1 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 0 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 0 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 1 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 1 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 0 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 0 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 1 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 1 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| 3 0 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 0 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 1 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 1 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 4 0 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 0 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 1 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 1 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |

The MaxSMT problem for the S-box $\mathcal{S}$ is defined as $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$, where

$$\Phi_1^{\mathcal{S}} := \left\{ \begin{array}{c} c_1 \cdot p_{1,1} + c_2 \cdot p_{2,1} + c_3 \cdot p_{3,1} = 0 \\ \wedge \\ c_1 \cdot p_{1,2} + c_2 \cdot p_{2,2} + c_3 \cdot p_{3,2} = 3 \\ \wedge \\ c_1 \cdot p_{1,3} + c_2 \cdot p_{2,3} + c_3 \cdot p_{3,3} = 2 \end{array} \right\},$$

$$\Phi_2^{\mathcal{S}} := \{c_1 = 0, c_2 = 0, c_3 = 0\}.$$

We can see that the following assginment is a solution of the MaxSMT problem $(\Phi_1^{\mathcal{S}}, \Phi_2^{\mathcal{S}})$:

$$\left\{ \begin{array}{c} p_{1,1} = 0,\ p_{2,1} = 0,\ p_{3,1} = 0, \\ p_{1,2} = 1,\ p_{2,2} = 1,\ p_{3,2} = 0, \\ p_{1,3} = 0,\ p_{2,3} = 1,\ p_{3,3} = 0, \\ c_1 = 1,\ c_2 = 2,\ c_3 = 0 \end{array} \right\}.$$

By introducing two addition Boolean variables, $b_1$ and $b_2$, we can construct the extended DDT $\mathcal{D}^{\dagger}_{\mathcal{S}}$ of the S-box $\mathcal{S}$, part of which is shown in Table 10.

## D MISSING PROOFS

### D.1 Proof of Lemma 5.1

Lemma 5.1. *Suppose* $[\![e]\!]^{\mathtt{W}}_{\gamma} = (\Psi, \vec{b})$ *with* $\Psi \neq \emptyset$. *The assignment* $\{b_1, \cdots, b_m\}$ *is a solution of* $\Psi$ *iff* $\{b_1, \cdots, b_i\}$ *is feasible differences of the operands and result of* $e$, *where* $\{b_{i+1}, \cdots, b_m\}$ *is the assignment of the auxiliary Boolean variables if exist..*

Proof. Suppose $[\![e]\!]^{\mathtt{W}}_{\gamma} = (\Psi, \vec{b})$. If $e$ is of the form $\mathtt{View}(x, i, j)$, $\sim x$ or $x\langle \cdot y \cdot \rangle$, $\Psi = \emptyset$. Below, we consider the other cases.

- $e$ is $x_1 \odot x_2$ for $\odot \in \{+, -, \oplus\}$. Suppose $\gamma(x_1) = b_1$ and $\gamma(x_2) = b_2$, namely, $b_1$ and $b_2$ model the differences of the operands $x_1$ and $x_2$ respectively. Let $b_0$ be the Boolean variable that models the difference of the result of $x_1 \odot x_2$. Recall that $\Psi$ can be two equivalent sets of IL constraints $\Psi^1_{2,3}(b_0, b_1, b_2)$ and $\Psi^2_{2,3}(b_0, b_1, b_2)$.

Since $\mathcal{B}^{\max}_{\mathsf{ww},\odot} = 3$ and $\mathcal{B}^{\min}_{\mathsf{ww},\odot} = 2$ (cf. Table 2), we get that
– either at least two of $b_0, b_1, b_2$ are 1,
– or $b_0 = b_1 = b_2 = 0$.
It is easy to see that $(b_0, b_1, b_2)$ is feasible differences of operands $x_1, x_2$ and result of $x_1 \odot x_2$ iff $(b_0, b_1, b_2)$ satisfies $\Psi^1_{2,3}(b_0, b_1, b_2)$. For $\Psi^2_{2,3}(b_0, b_1, b_2)$, the auxiliary Boolean variable $b'$ is 0 iff $b_0 = b_1 = b_2 = 0$. The constraint $\sum_{i=0}^{2} b_i \geq 2b'$ exactly characterizes that at least two of $b_0, b_1, b_2$ are 1 if $b' = 1$. Thus, $(b_0, b_1, b_2)$ is feasible differences of the operands $x_1, x_2$ and result of $x_1 \odot x_2$ iff $(b_0, b_1, b_2, b')$ satisfies $\Psi^2_{2,3}(b_0, b_1, b_2)$, where $b' = 0$ iff $b_0 = b_1 = b_2 = 0$.

- $e$ is $x_1 \odot x_2$ for $\odot \in \{\wedge, \vee\}$. Suppose $\gamma(x_1) = b_1$ and $\gamma(x_2) = b_2$, namely, $b_1$ and $b_2$ model the differences of the operands $x_1$ and $x_2$ respectively. Let $b_0$ be the Boolean variable that models the difference of the result of $x_1 \odot x_2$.

  Since $\mathcal{B}^{\max}_{\mathsf{ww},\odot} = 3$ and $\mathcal{B}^{\min}_{\mathsf{ww},\odot} = 1$ (cf. Table 2), we have that
  – either at least one of $b_0, b_1, b_2$ is 1,
  – or $b_0 = b_1 = b_2 = 0$.
  Moreover, $b_0 = 0$ if $b_1 = b_2 = 0$. The constraint $b_1 + b_2 \geq b_0$ exactly characterizes that $b_0 = 0$ if both $b_1$ and $b_2$ are 0, otherwise $b_0$ can be 1 or 0. Thus, $(b_0, b_1, b_2)$ is feasible differences of the operands $x_1, x_2$ and result of $x_1 \odot x_2$ iff $(b_0, b_1, b_2)$ satisfies $b_1 + b_2 \geq b_0$.

- $e$ is $M * x$. Suppose $\gamma(x) = \vec{b}$, i.e., $\vec{b}$ is a vector of Boolean variables each of which models the difference of an entry in the array $x$. Let $\vec{b'}$ be the vector of Boolean variables each of which models the difference of an entry in the resulting array $M \odot x$. Recall that $\Psi$ can be two equivalent sets of IL constraints $\Psi^1_M(\vec{b}, \vec{b'})$ and $\Psi^2_M(\vec{b}, \vec{b'})$, and $\mathcal{B}^{\min}_{\mathsf{ww},M} \geq 1$.

  We observe that
  – either the sum of their differences $(\vec{b}, \vec{b'})$ ranges from $\mathcal{B}^{\min}_{\mathsf{ww},M}$ to $\mathcal{B}^{\max}_{\mathsf{ww},M}$, namely,

  $$\mathcal{B}^{\min}_{\mathsf{ww},M} \leq \sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) \leq \mathcal{B}^{\max}_{\mathsf{ww},M},$$

  – or both $x$ and $M * x$ have no differences (i.e., $\sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) = 0$).
  These two conditions are equivalent to $\Psi^i_M(\vec{b}, \vec{b'})$ for $i \in \{1, 2\}$. Indeed, in $\Psi^1_M(\vec{b}, \vec{b'})$, $b''$ is an auxiliary Boolean variable,
  – $2|x| \cdot b'' \geq \sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i)$ exactly characterizes that $b'' = 1$ if $\sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) \neq 0$;
  – $\mathcal{B}^{\min}_{\mathsf{ww},M} \cdot b'' \leq \sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) \leq \mathcal{B}^{\max}_{\mathsf{ww},M}$ exactly characterizes $\mathcal{B}^{\min}_{\mathsf{ww},M} \leq \sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) \leq \mathcal{B}^{\max}_{\mathsf{ww},M}$ if $b'' = 1$, and $b'' = 0$ if $\sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) = 0$.

  The set $\Psi^2_M(\vec{b}, \vec{b'})$ is similar to $\Psi^1_M(\vec{b}, \vec{b'})$ except that $2|x| \cdot b'' \geq \sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i)$ is equivalently expressed by the constraints $b'' \geq \vec{b}_0, b'' \geq \vec{b}'_0, \cdots, b'' \geq \vec{b}_{|x|-1}, b'' \geq \vec{b}'_{|x|-1}$. Thus, $(\vec{b}, \vec{b'})$ is feasible differences of the operand $x$ and result of $M * x$ iff $(\vec{b}, \vec{b'}, b'')$ satisfies $\Psi^i_M(\vec{b}, \vec{b'})$, where $b'' = 0$ iff $\sum_{i=0}^{|x|-1} (\vec{b}_i + \vec{b}'_i) = 0$.

- $e$ is $x\langle y \rangle$. Let $b = \gamma(y)$ and $b'$ be the Boolean variable that models the difference of the result of $x\langle y \rangle$. We observe that
  – if the S-box $x$ is injective, the result of $x\langle y \rangle$ has some differences (i.e., $b' = 1$) iff $y$ has some differences (i.e., $b = 1$), which is equivalent to $b = b'$;

– if the S-box $x$ is non-injective, $x\langle y\rangle$ *may* differ in two executions *only* if $y$ differs in the two executions, which is equivalent to $b \geq b'$.

Thus, $(b, b')$ is feasible differences of the operand $y$ and the result of $x\langle y\rangle$ iff $b = b'$ if the S-box $x$ is injective, otherwise $b \geq b'$.

□

## D.2 Proof of Theorem 5.2

Theorem 5.2. *Let* $[\![P]\!]^{\mathsf{W}} = (\Phi, \gamma, \Theta)$ *and* $N$ *be the minimum value of the objective function* $\sum_{b\in\Theta} b$ *subject to* $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$. *We have that* $N \leq \mathcal{N}_{\texttt{diff}}$.

Proof. Suppose $[\![P]\!]^{\mathsf{W}} = (\Phi, \gamma, \Theta)$ for the given program $P$ and $N$ is the minimum value of the objective function $\sum_{b\in\Theta} b$ subject to $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$. Given two pairs of inputs $(K, X)$ and $(K, X')$ such that $X \neq X'$, $P$ executes the same sequence of statements under the inputs $(K, X)$ and $(K, X')$. (Note that $P$ is branching-free.) Let $\sigma_0 S_1 \sigma_1 S_2 \sigma_2 \cdots S_m \sigma_m$ and $\sigma'_0 S_1 \sigma'_1 S_2 \sigma'_2 \cdots S_m \sigma'_m$ be the sequences of states and statements of the executions under the inputs $(K, X)$ and $(K, X')$, respectively. Let $b_x$ be the corresponding Boolean variable of each scalar variable $x$ in a state $\sigma_i$, and let $b_{x,j}$ be the corresponding Boolean variable of the $(j+1)$-th entry (i.e., $x_j$) in the array $x$ in a state $\sigma_i$.

By Lemma 5.1 and induction on $m$ and syntactic structure of statements and expressions, we can get that if $b_x = 1$ (resp. $b_{x,j} = 1$) in any solution of $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ that corresponds to the two executions $\sigma_0 S_1 \sigma_1 S_2 \sigma_2 \cdots S_m \sigma_m$ and $\sigma'_0 S_1 \sigma'_1 S_2 \sigma'_2 \cdots S_m \sigma'_m$, then $\sigma_i(x) \oplus \sigma'_i(x) \neq 0$ (resp. $\sigma_i(x_j) \oplus \sigma'_i(x_j) \neq 0$). Thus, for any S-box $\mathcal{S}$ in $P$ under those two executions, if the Boolean variable $b_{\mathcal{S}}$ that models the input difference of $\mathcal{S}$ is 1 in the solution of $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$, then $\mathcal{S}$ must have different inputs, indicating that $\mathcal{S}$ is active. Therefore, we can deduce that $N \leq \mathcal{N}_{\texttt{diff}}$.

We remark that some solutions of $\Phi \cup \{(\sum_{i=0}^{n-1} \gamma(txt, i)) \geq 1\}$ may yield invalid differential characteristics, thus we cannot conclude that $N = \mathcal{N}_{\texttt{diff}}$. □

## D.3 Proof of Lemma 6.1

Lemma 6.1. *Suppose* $[\![e]\!]_{\gamma}^{\mathsf{B}} = (\Psi, \vec{b})$ *with* $\Psi \neq \emptyset$. *The assignment* $\{b_1, \cdots, b_m\}$ *is a solution of* $\Psi$ *iff* $\{b_1, \cdots, b_i\}$ *is feasible bit-level differences of the operands and result of* $e$, *where* $\{b_{i+1}, \cdots, b_m\}$ *is the assignment of the auxiliary Boolean variables if exist.*

Proof. Suppose $[\![e]\!]_{\gamma}^{\mathsf{B}} = (\Psi, \vec{b})$. If $e$ is of the form $\texttt{View}(x, i, j)$, $\sim x$, $\texttt{touint}(x_1, \cdots, x_m)$ or $x\langle\cdot y\cdot\rangle$, $\Psi = \emptyset$. If $e$ is $x\langle y\rangle$, the result follows from Proposition 4.2. Below, we consider the other cases.

- $e$ is $x_1 \odot x_2$ for $\odot \in \{\wedge, \vee\}$. Suppose $\gamma(x_1) = \vec{b}^1$ and $\gamma(x_2) = \vec{b}^2$, namely, $\vec{b}^1$ and $\vec{b}^2$ model the bit-level differences of of the operands $x_1$ and $x_2$ respectively. Let $\vec{b}^0$ be the vector of Boolean variables each of which models the difference of a bit in the result of $x_1 \odot x_2$.

  For each $0 \leq i < \|x\|$, for any fixed $\vec{b}^1_i$ and $\vec{b}^2_i$, we observe that:
  – if $\vec{b}^1_i = \vec{b}^2_i = 0$, then $\vec{b}^0_i = 0$;
  – otherwise $\vec{b}^0_i$ could be 0 and 1 (with the probability of $\frac{1}{2}$).
  The constraint $\vec{b}^1_i + \vec{b}^2_i \geq \vec{b}^0_i$ exactly characterizes that $\vec{b}^0_i = 0$ if $\vec{b}^1_i = \vec{b}^2_i = 0$ otherwise $\vec{b}^0_i$ could be 0 and 1. Thus, $\{\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i\}$ satisfies $\vec{b}^1_i + \vec{b}^2_i \geq \vec{b}^0_i$ iff $\{\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i\}$ is feasible differences of $(i+1)$-th bits of the operands $x_1, x_2$ and result of $x_1 \odot x_2$. The result immediately follows.

- $e$ is $x_1 \oplus x_2$. Suppose $\gamma(x_1) = \vec{b}^1$ and $\gamma(x_2) = \vec{b}^2$, namely, $\vec{b}^1$ and $\vec{b}^2$ model the bit-level differences of the operands $x_1$ and $x_2$ respectively. Let $\vec{b}^0$ be the vector of Boolean variables each of which models the difference of a bit in the result of $x_1 \oplus x_2$. Recall that $\Psi$ can be three equivalent sets of IL constraints, i.e., $\bigcup_{i=0}^{\|x\|-1} \psi^1_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$, $\bigcup_{i=0}^{\|x\|-1} \psi^2_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$ and $\bigcup_{i=0}^{\|x\|-1} \psi^3_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$.

For each $0 \le i < \|x\|$, for any fixed $\vec{b}^1_i$ and $\vec{b}^2_i$, we observe that:

– either $\vec{b}^0_i = \vec{b}^1_i = \vec{b}^2_i = 0$,
– or exactly two of $\vec{b}^0_i, \vec{b}^1_i, \vec{b}^2_i$ are 1 (i.e., $\vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i = 2$).

The above two conditions are exactly characterized by $\psi^j_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$ for any $j \in \{1, 2, 3\}$. Indeed, we deduce that $\{\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i\}$ is feasible differences of $(i+1)$-th bits of the operands $x_1, x_2$ and result of $x_1 \odot x_2$ iff

– $\{\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i, b'\}$ satisfies $\psi^1_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$, where $b' = 0$ iff $\vec{b}^0_i = \vec{b}^1_i = \vec{b}^2_i = 0$;
– $\{\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i\}$ satisfies $\psi^2_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$;
– $\{\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i\}$ satisfies $\psi^3_\oplus(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i)$, where $b' = 0$ iff $\vec{b}^0_i = \vec{b}^1_i = \vec{b}^2_i = 0$.

Thus, the result follows.

- $e$ is $x_1 \odot x_2$ for $\odot \in \{+, -\}$. Suppose $\gamma(x_1) = \vec{b}^1$ and $\gamma(x_2) = \vec{b}^2$, namely, $\vec{b}^1$ and $\vec{b}^2$ model the bit-level differences of the operands $x_1$ and $x_2$ respectively. Let $\vec{b}^0$ be the vector of Boolean variables each of which models the difference of a bit in the result of $x_1 \odot x_2$.

Consider $y = x_1 + x_2$. Clearly,

– $\mathsf{bin}_i(y) = \mathsf{bin}_i(x_1) \oplus \mathsf{bin}_i(x_2) \oplus \vec{c}_i$, for every $0 \le i < \|y\|$;
– the carry bit $\vec{c}_i = 1$ iff $\mathsf{bin}_{i-1}(x_1) + \mathsf{bin}_{i-1}(x_2) + \vec{c}_{i-1} \ge 2$, for every $1 \le i < \|y\|$, with $\vec{c}_0 = 0$.

Suppose $\vec{b}^0 = \mathsf{bin}(y) \oplus \mathsf{bin}(y') = \mathsf{bin}(x_1 + x_2) \oplus \mathsf{bin}(x'_1 + x'_2)$ where $\mathsf{bin}(x_1) \oplus \mathsf{bin}(x'_1) = \vec{b}^1$ and $\mathsf{bin}(x_2) \oplus \mathsf{bin}(x'_2) = \vec{b}^2$. Let $\vec{b}^3_i$ be the difference of the carry bit $\vec{c}_i \oplus \vec{c}'_i$ for every $0 \le i < \|y\|$.

We first consider the case $i \ge 2$. Clearly,

$$\vec{b}^0_i = \big(\mathsf{bin}_i(x_1) \oplus \mathsf{bin}_i(x_2) \oplus \vec{c}_i\big) \oplus \big(\mathsf{bin}_i(x'_1) \oplus \mathsf{bin}_i(x'_2) \oplus \vec{c}'_i\big) = \vec{b}^1_i \oplus \vec{b}^2_i \oplus \vec{b}^3_i.$$

Observe that $\vec{b}^3_i = 1$ iff exactly one of the follows holds:

– $\mathsf{bin}_{i-1}(x_1) + \mathsf{bin}_{i-1}(x_2) + \vec{c}_{i-1} \ge 2$, or
– $\mathsf{bin}_{i-1}(x'_1) + \mathsf{bin}_{i-1}(x'_2) + \vec{c}'_{i-1} \ge 2$.

Since $\vec{c}_{i-1} \oplus \vec{c}'_{i-1} = \vec{b}^3_{i-1} = \vec{b}^0_{i-1} \oplus \vec{b}^1_{i-1} \oplus \vec{b}^2_{i-1}$, we can deduce that

– if $\vec{b}^0_{i-1} = \vec{b}^1_{i-1} = \vec{b}^2_{i-1} = 1$, then $\vec{c}_{i-1} \oplus \vec{c}'_{i-1} = \vec{b}^3_{i-1} = \vec{b}^3_i = 1$ and $\vec{b}^0_i = \neg(\vec{b}^1_i \oplus \vec{b}^2_i)$
– if $\vec{b}^0_{i-1} = \vec{b}^1_{i-1} = \vec{b}^2_{i-1} = 0$, then $\vec{c}_{i-1} \oplus \vec{c}'_{i-1} = \vec{b}^3_{i-1} = \vec{b}^3_i = 0$ and $\vec{b}^0_i = \vec{b}^1_i \oplus \vec{b}^2_i$,
– otherwise $1 \le \vec{b}^0_{i-1} + \vec{b}^1_{i-1} + \vec{b}^2_{i-1} \le 2$. Indeed, the probability of $\vec{b}^3_i = 1$ is $\frac{1}{2}$, implying that the probability of $\vec{b}^0_i = \neg(\vec{b}^1_i \oplus \vec{b}^2_i)$, (resp. $\vec{b}^0_i = \vec{b}^1_i \oplus \vec{b}^2_i$ and $\vec{b}^0_i = 1$) is $\frac{1}{2}$.

The above three conditions are exactly characterized by the set of IL constraints $\Psi_i$ for $i \ge 2$, where

– if $\sum_{j=0}^{2} \vec{b}^j_{i-1} = 3$, $\Psi_i$ becomes $\big\{\vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \ge 1, \vec{b}^0_i + \vec{b}^1_i \le \vec{b}^2_i + 1, \vec{b}^0_i + \vec{b}^2_i \le \vec{b}^1_i + 1, \vec{b}^1_i + \vec{b}^2_i \le \vec{b}^0_i + 1\big\}$ which is equivalent to $\vec{b}^0_i = \neg(\vec{b}^1_i \oplus \vec{b}^2_i)$;
– if $\sum_{j=0}^{2} \vec{b}^j_{i-1} = 0$, $\Psi_i$ becomes $\big\{-\vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i \ge 0, \vec{b}^0_i + \vec{b}^1_i - \vec{b}^2_i \ge 0, \vec{b}^0_i - \vec{b}^1_i + \vec{b}^2_i \ge 0, 2 \ge \vec{b}^0_i + \vec{b}^1_i + \vec{b}^2_i\big\}$, which is equivalent to $\vec{b}^0_i = \vec{b}^1_i \oplus \vec{b}^2_i$;
– if $1 \le \sum_{j=0}^{2} \vec{b}^j_{i-1} \le 2$, $\Psi_i$ always holds.

Furthermore, $\Psi_0$ and $\Psi_1$ can be defined similarly (with simplification by $\vec{b}^3_0 = \vec{c}_0 = \vec{c}'_0 = 0$). Thus, $\{\vec{b}^1, \vec{b}^2, \vec{b}^0\}$ is feasible bit-level differences of the operands $x_1, x_2$ and result of $x_1 + x_2$ iff $\{\vec{b}^1, \vec{b}^2, \vec{b}^0, b'\}$ satisfies $\bigcup_{i=0}^{\|x\|-1} \Psi_i$ if $b'$ is used in $\Psi_0$ (for which $b'$ is defined the same as for $\oplus$), otherwise $\{\vec{b}^1, \vec{b}^2, \vec{b}^0\}$ satisfies $\bigcup_{i=0}^{\|x\|-1} \Psi_i$.

The result of $x_1 - x_2$ follows because $y = x_1 - x_2$ iff $x_2 = x_1 + y$, and the Boolean variables $\vec{b}_i^0$, $\vec{b}_i^1$ and $\vec{b}_i^2$ in $\Psi_i$ are symmetric.

- $e$ is $M * x$. Suppose $y = M * x$ and $s = \frac{\|x\|}{|x|}$ (i.e., the bit width of entries in the arrays $x$ and $y$). Let $\gamma(x) = \vec{b}$, i.e., for every $0 \leq i < |x|$, $0 \leq h < s$, $\vec{b}_{i \cdot s+h}$ models the difference of the $(h+1)$-th bit $\mathsf{bin}_h(x_i)$ of the $(i+1)$-the entry $x_i$ in the array $x$. Similarly, let $\vec{b}$ be the vector of Boolean variables such that for every $0 \leq i < |x|$, $0 \leq h < s$, $\vec{b}'_{i \cdot s+h}$ models the difference of the $(h+1)$-th bit $\mathsf{bin}_h(y_i)$ of the $(i+1)$-the entry $y_i$ in the array $y$. By expanding the matrix-vector product, we have:

$$M * x = \Big( \bigoplus_{j=0}^{|x|-1} (M_{0,j} \otimes x_j), \cdots, \bigoplus_{j=0}^{|x|-1} (M_{|x|-1,j} \otimes x_j) \Big).$$

Clearly, the $(i+1)$-th entry $y_i$ of the array $y$ is $\bigoplus_{j=0}^{|x|-1} (M_{i,j} \otimes x_j)$ and thus the difference $\vec{b}'_{i \cdot s+h}$ of the $(h+1)$-th bit $\mathsf{bin}_h(y_i)$ of the $(i+1)$-th entry $y_i$ is the parity of the differences of the $(h+1)$-th bits in $M_{i,j} \otimes x_j$ for $0 \leq j < h$.

Let $x'$ be another input such that for every $0 \leq i < |x|$, $0 \leq h < s$, $\vec{b}_{i \cdot s+h} = \mathsf{bin}_h(x_i) \oplus \mathsf{bin}_h(x'_i)$. Let $y' = M * x'$. By expanding the finite-field multiplication ($\otimes$) using a series of modular left shifts, we have that

$$\begin{array}{rcl}
\vec{b}'_{i \cdot s+h} & = & \mathsf{bin}_h(y_i) \oplus \mathsf{bin}_h(y'_i) \\
& = & \mathsf{bin}_h(\bigoplus_{j=0}^{|x|-1} (M_{i,j} \otimes x_j)) \oplus \mathsf{bin}_h(\bigoplus_{j=0}^{|x|-1} (M_{i,j} \otimes x'_j)) \\
& = & \mathsf{bin}_h(\bigoplus_{j=0}^{|x|-1} ((M_{i,j} \otimes x_j) \oplus (M_{i,j} \otimes x'_j))) \\
& = & \bigoplus_{j=0}^{|x|-1} \mathsf{bin}_h((M_{i,j} \otimes (x_j \oplus x'_j))) \\
& = & \bigoplus_{j=0}^{|x|-1} \mathsf{bin}_h(\bigoplus_{k=0}^{s-1} \mathsf{MSL}_{\vec{c}}(\mathsf{bin}(x_j) \oplus \mathsf{bin}(x'_j), M_{i,j,k}, k)) \\
& = & \bigoplus_{j=0}^{|x|-1} \mathsf{bin}_h(\bigoplus_{k=0}^{s-1} \mathsf{MSL}_{\vec{c}}((\vec{b}_{j \cdot s+0}, \cdots, \vec{b}_{j \cdot s+s-1}), M_{i,j,k}, k))
\end{array}$$

where the function $\mathsf{MSL}_{\vec{c}}$ is defined as

$$\mathsf{MSL}_{\vec{c}}(b_0, \cdots, b_{s-1}, b, k) = \begin{cases} \big(0, \cdots, 0\big), & \text{if } b = 0; \\ \big(b_{s-1-k}, b_{s-k}, \cdots, b_{s-1}, 0, \cdots, 0\big) \oplus \big( \bigoplus_{\imath=0}^{s-2-k} b_\imath^j \cdot \vec{c}\big), & \text{otherwise.} \end{cases}$$

$\vec{c} = (\vec{c}_0, \cdots, \vec{c}_{s-1}) = \mathsf{bin}(2 \otimes 2^{s-1})$ is the $s$-bitstream corresponding to the coefficients of the irreducible polynomial for the underlying finite-field (e.g., $\vec{c} = (0, 0, 0, 1, 1, 0, 1, 1)$ for $\mathbb{GF}(2^8)$ whose irreducible polynomial is $X^8 + X^4 + X^3 + X + 1$ in AES).
Thus, $\vec{b}'_{i \cdot s+h}$ can be re-formulated as:

$$\begin{array}{rcl}
\vec{b}'_{i \cdot s+h} & = & \bigoplus_{j=0}^{|x|-1} \mathsf{bin}_h(\bigoplus_{k=0}^{s-1} \mathsf{MSL}_{\vec{c}}((\vec{b}_{j \cdot s+0}, \cdots, \vec{b}_{j \cdot s+s-1}), M_{i,j,k}, k)) \\
& = & \bigoplus_{j=0}^{|x|-1} \Big( \bigoplus\limits_{h \leq k < s} \vec{b}_{j \cdot s+k} \wedge M_{i,j,k} \Big) \oplus \Big( \vec{c}_h \wedge \bigoplus\limits_{0 \leq k < \lfloor \frac{s}{2} \rfloor} \vec{b}_{j \cdot s+2k} \Big) \\
& = & \Big( \bigoplus\limits_{0 \leq j < |x|,\ h \leq k < s,\ M_{i,j,k}=1} \vec{b}_{j \cdot s+k} \Big) \oplus \Big( \vec{c}_h \wedge \bigoplus\limits_{0 \leq j < |x|,\ 0 \leq k < \lfloor \frac{s}{2} \rfloor} \vec{b}_{j \cdot s+2k} \Big).
\end{array}$$

Let $\{b^0, \cdots, b^m\}$ be the set of support variables of the following formula

$$\Big( \bigoplus_{0 \leq j < |x|,\ h \leq k < s,\ M_{i,j,k}=1} \vec{b}_{j \cdot s+k} \Big) \oplus \Big( \vec{c}_h \wedge \bigoplus_{0 \leq j < |x|,\ 0 \leq k < \lfloor \frac{s}{2} \rfloor} \vec{b}_{j \cdot s+2k} \Big).$$

We have $\vec{b}'_{i \cdot s+h} = \bigoplus_{t=0}^{m} b_t$, which is equivalent to $\Psi^{v}_{\mathsf{M,i,h}}$ for any $v \in \{1, 2, 3\}$. Thus, the result follows.

□

### D.4 Proof of Theorem 6.2

Theorem 6.2. *Let $[\![P]\!]^{\mathsf{B}} = (\Phi, \gamma, \Theta)$ and $N$ be the minimum value of the objective function $\sum_{b\in\Theta} b$ subject to the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$. We have that $N \leq \mathcal{N}_{\mathtt{diff}}$.*

Proof. The proof follows the lines of the proof of Theorem 5.2. Given a given program $P$, suppose $[\![P]\!]^{\mathsf{B}} = (\Phi, \gamma, \Theta)$ and $N$ is the minimum value of the objective function $\sum_{b\in\Theta} b$ subject to the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$. Given two pairs of inputs $(K, X)$ and $(K, X')$ such that $X \neq X'$, let $\sigma_0 S_1 \sigma_1 S_2 \sigma_2 \cdots S_m \sigma_m$ and $\sigma'_0 S_1 \sigma'_1 S_2 \sigma'_2 \cdots S_m \sigma'_m$ be the sequences of states and statements of the executions under the inputs $(K, X)$ and $(K, X')$, respectively. For each scalar variable $x$ in a state $\sigma_i$, let $b_{x,j}$ be the Boolean variable modeling the difference of the $(j+1)$-th most significant bit of $\mathsf{bin}(x)$ of $x$ in $\Phi$, and for each array variable $x$ in a state $\sigma_i$, let $b_{x,k,j}$ be the Boolean variable modeling the difference of the $(j+1)$-th most significant bit of the $(k+1)$-th entry (i.e., $\mathsf{bin}(x_k)$) of the array $x$ in $\Phi$.

By Lemma 6.1 and induction on $m$ and syntactic structure of statements and expressions, we can get that if $b_x = 1$ (resp. $b_{x,j} = 1$) in any solution of $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$ that corresponds to the two executions $\sigma_0 S_1 \sigma_1 S_2 \sigma_2 \cdots S_m \sigma_m$ and $\sigma'_0 S_1 \sigma'_1 S_2 \sigma'_2 \cdots S_m \sigma'_m$, the $(j+1)$-th most significant bits of $\mathsf{bin}(\sigma_i(x))$ and $\mathsf{bin}(\sigma'_i(x))$ (resp. $(k+1)$-th entries of $\mathsf{bin}(\sigma_i(x_k))$ and $\mathsf{bin}(\sigma'_i(x_k))$) are different. Thus, for any S-box $\mathcal{S}$ in the program $P$ under those two executions, if the Boolean variable $b_{\mathcal{S}}$ that models the input difference of the S-box $\mathcal{S}$ is 1 in the solution of $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$, then the S-box $\mathcal{S}$ must have different inputs, indicating that the S-box $\mathcal{S}$ is active. Therefore, we deduce that $N \leq \mathcal{N}_{\mathtt{diff}}$. □

### D.5 Proof of Lemma 7.1

Lemma 7.1. *Suppose $[\![e]\!]^{\mathsf{EB}}_{\gamma} = (\Psi, \vec{b}, \varrho)$ with $\Psi \neq \emptyset$. The assignment $\{b_1, \cdots, b_m, p_1, \cdots, p_n\}$ is a solution of $\Psi$ iff the probability of $\{b_1, \cdots, b_i\}$ being bit-level differences of the operands and result of $e$ is $2^{-\varrho[p_1,\cdots,p_n]}$, where $\varrho[p_1, \cdots, p_n]$ denotes the value of $\varrho$ under the assignment $\{p_1, \cdots, p_n\}$ of the Boolean variables for encoding probabilities, and $\{b_{i+1}, \cdots, b_m\}$ is the assignment of the auxiliary Boolean variables if exist.*

Proof. Suppose $[\![e]\!]^{\mathsf{EB}}_{\gamma} = (\Psi, \vec{b}, \varrho)$. If $e$ is of the form $\mathtt{View}(x, i, j)$, $\sim x$, $\mathtt{touint}(x_1, \cdots, x_m)$ or $x\langle\cdot y\cdot\rangle$, $\Psi = \emptyset$. If $e$ is of the form $M * x$ or $x_1 \oplus x_2$, by Lemma 6.1, $\{b_1, \cdots, b_m\}$ satisfies $\Psi$ iff the probability of $\{b_1, \cdots, b_i\}$ being bit-level differences of the operands and result of $e$ is 1. The result immediately follows from the fact that $\varrho = 0$. If $e$ is $x\langle y\rangle$ for S-box $x$, the result follows from Proposition 4.4. Below, we consider the other cases.

- $e$ is $x_1 \odot x_2$ for $\odot \in \{\wedge, \vee\}$. Suppose $\gamma(x_1) = \vec{b}^1$ and $\gamma(x_2) = \vec{b}^2$, namely, $\vec{b}^1$ and $\vec{b}^2$ model the bit-level differences of the operands $x_1$ and $x_2$ respectively. Let $\vec{b}^0$ be the vector of Boolean variables each of which models the difference of a bit in the result of $x_1 \odot x_2$.

  For every $0 \leq i < \|x\|$, for any fixed $\vec{b}^1_i$ and $\vec{b}^2_i$, we observe that:
  – if $\vec{b}^1_i = \vec{b}^2_i = 0$, then $\vec{b}^0_i = 0$;
  – otherwise $\vec{b}^0_i$ could be 0 and 1 with the probability of $\frac{1}{2}$.
  Thus, for every $0 \leq i < \|x\|$,
  – the probability of $\vec{b}^0_i = 0$ is $2^{-0}$ when $\vec{b}^1_i = \vec{b}^2_i = 0$, then $\vec{p}_i$ must be 0.
  – the probability of $\vec{b}^0_i = 1$ is $2^{-1}$ when $\vec{b}^1_i + \vec{b}^2_i \geq 1$, then $\vec{p}_i$ must be 1.

  The above two conditions are exactly characterized by $\Psi_i$, i.e., $(\vec{b}^1_i, \vec{b}^2_i, \vec{b}^0_i, \vec{p}_i)$ is a solution of $\Psi_i$ iff the probability of the difference $\vec{b}^0_i$ of the $(i+1)$-th bit in the result of $x \odot y$ is $2^{-\vec{p}_i}$ when the differences of the $(i+1)$-th bits of the operands $x$ and $y$ are $\vec{b}_i^1$ and $\vec{b}_i^2$, respectively. Thus, the result immediately follows.

- $e$ is $x_1 \odot x_2$ for $\odot \in \{+, -\}$. Suppose $\gamma(x_1) = \vec{b}^1$ and $\gamma(x_2) = \vec{b}^2$, namely, $\vec{b}^1$ and $\vec{b}^2$ model the bit-level differences of the operands $x_1$ and $x_2$ respectively. Let $\vec{b}^0$ be the vector of Boolean variables each of which models the difference of a bit in the result of $x_1 + x_2$.

  Following the lines of the proof of Lemma 7.1, we have:
  - if $\vec{b}_{i-1}^0 = \vec{b}_{i-1}^1 = \vec{b}_{i-1}^2$, then $\vec{p}_i$ should be 0 (i.e., the probability $2^{-\vec{p}_i}$ of $\vec{b}_i^0 = \neg(\vec{b}_i^1 \oplus \vec{b}_i^2)$ or $\vec{b}_i^0 = \vec{b}_i^1 \oplus \vec{b}_i^2$ is 1),
  - if $1 \leq \sum_{j=0}^{2} \vec{b}_{i-1}^j \leq 2$, then $\vec{p}_i$ should be 1 (i.e., the probability $2^{-\vec{p}_i}$ of $\vec{b}_i^0 = 1$ or $\vec{b}_i^0 = 0$ is $\frac{1}{2}$).

  The above two conditions are exactly characterized by $\Psi_i^1 \cup \Psi_i^2$, where $\Psi_i^1$ ensures that $\vec{p}_i = 0$ if $\vec{b}_{i-1}^0 = \vec{b}_{i-1}^1 = \vec{b}_{i-1}^2$, and $\vec{p}_i = 1$ if $1 \leq \sum_{j=0}^{2} \vec{b}_{i-1}^j \leq 2$; $\Psi_i^2$ ensures that $\vec{b}_i^0 = \neg(\vec{b}_i^1 \oplus \vec{b}_i^2)$ if $\sum_{j=0}^{2} \vec{b}_{i-1}^j = 3$, and $\vec{b}_i^0 = \vec{b}_i^1 \oplus \vec{b}_i^2$ if $\sum_{j=0}^{2} \vec{b}_{i-1}^j = 0$. Thus, the result follows from the fact that for each solution $(\vec{b}^0, \vec{b}^1, \vec{b}^2, \vec{p})$ of the IL constraints $\bigcup_{i=0}^{\|x\|-1} \Psi_i$, the probability of the output difference $\vec{b}^0$ of the modular addition (+) for the given input differences $\vec{b}^1$ and $\vec{b}^2$ is $2^{-\sum_{i=0}^{\|x\|-1} \vec{p}_i}$.

  For $x_1 - x_2$, it is easy to deduce that $[\![x_1 - x_2]\!]_\gamma^{\mathsf{EB}} = [\![x_1 + x_2]\!]_\gamma^{\mathsf{EB}}$, because if $y = x_1 - x_2$, then $x_2 = x_1 + y$, and the Boolean variables $\vec{b}_i^0$, $\vec{b}_i^1$ and $\vec{b}_i^2$ in all the IL constraints of $[\![x_1 + x_2]\!]_\gamma^{\mathsf{EB}}$ are symmetric except for $\vec{p}_i \geq \vec{b}_{i-1}^0 - \vec{b}_{i-1}^1$, $\vec{p}_i \geq \vec{b}_{i-1}^1 - \vec{b}_{i-1}^2$ and $\vec{p}_i \geq \vec{b}_{i-1}^2 - \vec{b}_{i-1}^0$ which still work in $[\![x_1 - x_2]\!]_\gamma^{\mathsf{EB}}$.

□

### D.6 Proof of Theorem 7.2

Theorem 7.2. *Let $[\![P]\!]^{\mathsf{EB}} = (\Phi, \gamma, \varrho)$ and $u$ be the minimum value of the objective function $\varrho$ subject to the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$. The maximum differential characteristic probability of the program $P$ is no greater than $2^{-u}$.*

Proof. Suppose $[\![P]\!]^{\mathsf{EB}} = (\Phi, \gamma, \varrho)$ for the given program $P$. For any solution $\{b_1, \cdots, b_m, p_1, \cdots, p_n\}$ of $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$ that corresponds an $s$-round differential characteristic $(\Delta X^0, \cdots, \Delta X^s)$, by Lemma 7.1, the differential characteristic probability $\mathsf{Pr}_{\mathsf{Enc}}(\Delta X^0, \cdots, \Delta X^s)$ is no greater than $2^{-\sum_{i=0}^{n} p_i}$.

Let $u$ be the minimum value of the objective function $\varrho$ subject to the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$. Clearly, the differential characteristic probability $\mathsf{Pr}_{\mathsf{Enc}}(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^s)$ of any optimal $s$-round differential characteristic $(\widetilde{\Delta X}^0, \cdots, \widetilde{\Delta X}^s)$ is no greater than $2^u$. □

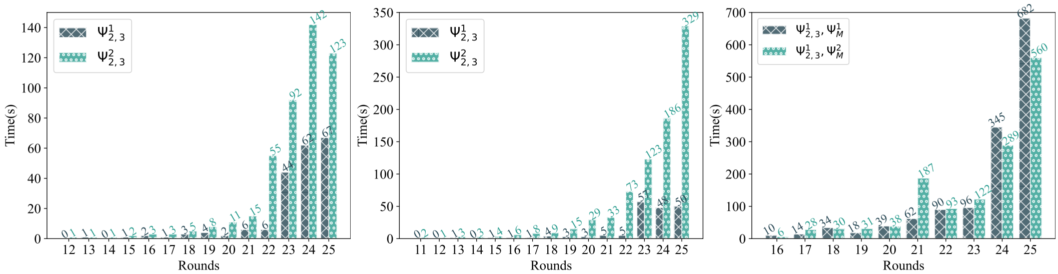
## E FURTHER EXPERIMENTAL RESULTS

In this section, we compare the performance of various alternative methods to generate MILP. In summary, we find that (below, $\Psi$ with super/subscripts refer to IL constraints specified in the semantic rules from Sections 4–7):

(1) $\Psi_{2,3}^1$ performs much better than $\Psi_{2,3}^2$, though 3 out of 4 constraints in $\Psi_{2,3}^2$ are much simpler;
(2) $\Psi_M^1$ and $\Psi_M^2$ are almost comparable though $\Psi_M^1$ contains significantly fewer constraints;
(3) removing the redundant constraints $\Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^2$ and/or $\Psi_{\mathcal{S}}^{\mathsf{bn}}$ from the (extended) bit-wise modeling of S-boxes often significantly degrades the performance;
(4) $\Psi_\oplus^2$ performs much better than $\Psi_\oplus^1$ and $\Psi_\oplus^3$ though $\Psi_\oplus^2$ contains more constraints than $\Psi_\oplus^3$;
(5) $\Psi_{\mathtt{M,i,h}}^2$ outperforms $\Psi_{\mathtt{M,i,h}}^1$ and $\Psi_{\mathtt{M,i,h}}^3$ though $\Psi_{\mathtt{M,i,h}}^2$ contains more constraints than $\Psi_{\mathtt{M,i,h}}^3$;

Table 11. Results of the word-wise approach, where #AS denotes the minimum number of active S-boxes.

| **Rounds** | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AES | **#AS** | 1 | 5 | 9 | **25** | 26 | 30 | 34 | 50 | 51 | 55 | 59 | 75 | 76 | 80 | N/A | | | | | | | | | | | | | | | | | | | | | |
| | **Time** | 0s | 0s | 0s | **0s** | 0s | 0s | 1s | 0s | 0s | 0s | 1s | 0s | 0s | 0s | N/A | | | | | | | | | | | | | | | | | | | | | |
| KLEIN | **#AS** | 1 | 5 | 8 | 15 | 16 | 20 | 23 | 30 | 31 | **35** | 38 | 45 | N/A | | | | | | | | | | | | | | | | | | | | | | | |
| | **Time** | 1s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | **0s** | 0s | 0s | N/A | | | | | | | | | | | | | | | | | | | | | | | |
| LBLOCK | **#AS** | 0 | 1 | 2 | 3 | 4 | 6 | 8 | 11 | 14 | 18 | 22 | 24 | 27 | 30 | **32** | 35 | 36 | 39 | 41 | 44 | 45 | 48 | 50 | 53 | 54 | 57 | 59 | 62 | 63 | 66 | 68 | 71 | N/A | | | |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | **1s** | 1s | 1s | 3s | 3s | 2s | 6s | 5s | 62s | 62s | 80s | 385s | 273s | 492s | 359s | 1036s | 3607s | 8312s | N/A | | | |
| MIBS | **#AS** | 0 | 1 | 2 | 5 | 6 | 7 | 8 | 11 | 12 | 13 | 14 | 17 | 18 | 19 | 20 | 23 | 24 | 25 | 26 | 29 | 30 | 31 | **32** | 35 | 36 | 37 | 38 | 41 | 42 | 43 | 44 | 47 | N/A | | | |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | 0s | 1s | 1s | 1s | 3s | 4s | 2s | 4s | 2s | 3s | 2s | 3s | 4s | **5s** | 4s | 7s | 4s | 5s | 9s | 11s | 9s | 8s | 20s | N/A | | | |
| PHOTON | **#AS** | 1 | 9 | 17 | 81 | 82 | 90 | 98 | 162 | 163 | 171 | 179 | 243 | N/A | | | | | | | | | | | | | | | | | | | | | | | |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | N/A | | | | | | | | | | | | | | | | | | | | | | | |
| PICCOLO | **#AS** | 0 | 5 | 10 | 15 | 20 | 30 | **35** | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 | 105 | 110 | 115 | 120 | 125 | N/A | | | | | | | | | | |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | **0s** | 1s | 0s | 1s | 0s | 1s | 3s | 6s | 4s | 6s | 28s | 30s | 31s | 38s | 187s | 93s | 122s | 289s | 560s | N/A | | | | | | | | | | |
| TWINE | **#AS** | 0 | 1 | 2 | 3 | 4 | 6 | 8 | 11 | 14 | 18 | 22 | 24 | 27 | 30 | **32** | 35 | 36 | 39 | 41 | 44 | 45 | 48 | 50 | 53 | 54 | 57 | 59 | 62 | 63 | 66 | 68 | 71 | 72 | 75 | 77 | 80 |
| | **Time** | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 0s | 1s | 0s | 0s | 0s | 1s | **0s** | 1s | 2s | 2s | 3s | 2s | 5s | 6s | 26s | 79s | 72s | 196s | 166s | 173s | 786s | 858s | 5395s | 6669s | 8168s | 26348s | 19771s | 18743s |



(a) LBLOCK (b) TWINE (c) PICCOLO

Fig. 11. Comparison of alternative modeling methods in the word-wise approach.

(6) the recent promising techniques of [Li and Sun 2022; Udovenko 2021] for constructing $\Psi^3_S$ produce the fewest constraints with the same number of Boolean variables, but exhibit the least efficiency in determining the lower bound of the minimum number of active S-boxes, instead, the techniques of [Boura and Coggia 2020] in general outperform the others;

(7) the techniques of [Sasaki and Todo 2017; Sun et al. 2014b] are comparable for constructing $\Psi^4_S$ and bounding the maximum differential characteristic probability, and outperform the others including [Boura and Coggia 2020; Li and Sun 2022; Udovenko 2021].

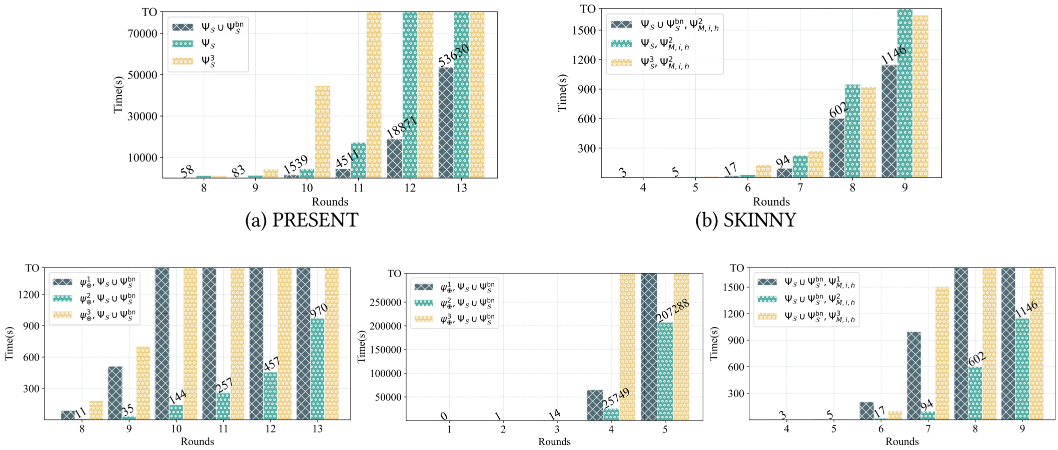
## E.1 More Results of the Word-wise Approach

The minimum number of active S-boxes of the entire rounds of all the word-wise implementations are reported in Table 11.

*E.1.1 Comparison of Alternative Word-wise Modeling Methods.* In our word-wise approach (cf. Figure 6),

- $\Psi^1_{2,3}$ and $\Psi^2_{2,3}$ are two alternative methods for modeling any binary operator (e.g., $+$, $-$, $\oplus$) whose minimum and maximum word-wise branch numbers are 2 and 3, respectively;
- $\Psi^1_M$ and $\Psi^2_M$ are two alternative methods for modeling the matrix-vector product $M * x$.

**Results of $\Psi^1_{2,3}$ vs. $\Psi^2_{2,3}$.** We conduct experiments on LBLOCK and TWINE (up to 25 rounds) that use the XOR operator. The results are depicted in Figures 11a and 11b, respectively, where the number on top of the bar is the execution time in seconds (s). Note that the results for small rounds are negligible, thus omitted. We can observe that $\Psi^1_{2,3}$ performs significantly better than $\Psi^2_{2,3}$, in particular, for large rounds. It is because $\Psi^1_{2,3}$ contains fewer constraints and Boolean variables, although 3 out of 4 constraints in $\Psi^2_{2,3}$ are much simple.



(a) PRESENT

(b) SKINNY

(c) LBLOCK

(d) KLEIN

(e) SKINNY

Fig. 12. Comparison of alternative modeling methods in the bit-wise approach.

**Results of $\Psi_M^1$ vs. $\Psi_M^2$.** We conduct experiments on AES, KLEIN, MIBS, PHOTON and PICCOLO (up to 25 rounds) that use the matrix-vector product, where $\Psi_{2,3}^1$ is used if the XOR operator is used. $\Psi_M^1$ and $\Psi_M^2$ have almost the same performance on AES, KLEIN, MIBS and PHOTON, thus are not reported. The results of PICCOLO are depicted in Figure 11c. We found that $\Psi_M^1$ performs better than $\Psi_M^2$ for some rounds (e.g., 21, 22, 23), but worse than $\Psi_M^2$ for some other rounds (e.g., 20, 24, 25). Recall that $\Psi_M^1$ contains fewer constraints but one more complex constraint than $\Psi_M^2$. It seems that fewer constraints do not always yield better performance. Note that the results of the combinations between $\Psi_{2,3}^1$ vs. $\Psi_{2,3}^2$ and $\Psi_M^1$ vs. $\Psi_M^2$ are not reported, because similar conclusions can be drawn.

## E.2 Comparison of Alternative Bit-wise Modeling Methods

In our bit-wise approach (cf. Figure 8),

- $\psi_{\oplus}^1$, $\psi_{\oplus}^2$ and $\psi_{\oplus}^3$ are three alternative methods for modeling the $\oplus$ operator;
- $\Psi_{\mathtt{M,i,h}}^1$, $\Psi_{\mathtt{M,i,h}}^2$ and $\Psi_{\mathtt{M,i,h}}^3$ are three alternative methods for modeling $M * x$;
- $\Psi_{\mathcal{S}} \cup \Psi_{\mathcal{S}}^{\mathsf{bn}}$ for modeling S-boxes can be safely simplified to $\Psi_{\mathcal{S}}$ and $\Psi_{\mathcal{S}}^3$. Recall that $\Psi_{\mathcal{S}} = \Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^2 \cup \Psi_{\mathcal{S}}^3$ if the S-box $\mathcal{S}$ is injective, otherwise $\Psi_{\mathcal{S}} = \Psi_{\mathcal{S}}^1 \cup \Psi_{\mathcal{S}}^3$, where $\Psi_{\mathcal{S}}^1$, $\Psi_{\mathcal{S}}^2$ and $\Psi_{\mathcal{S}}^{\mathsf{bn}}$ are redundant.
- $\Psi_{\mathcal{S}}^3$ can be constructed from the DDT $\mathcal{D}_{\mathcal{S}}$ of an S-box $\mathcal{S}$ using different techniques.

**Results of $\Psi_{\mathcal{S}} \cup \Psi_{\mathcal{S}}^{\mathsf{bn}}$ vs. $\Psi_{\mathcal{S}}$ vs. $\Psi_{\mathcal{S}}^3$.** We conduct experiments on PRESENT (up to 13 rounds) and SKINNY (up to 13 rounds) that use S-boxes where $\Psi_{\mathcal{S}}^3$ is constructed using the technique from [Sun et al. 2014b] and $\Psi_{\mathtt{M,i,h}}^2$ is used for the matrix-vector product in SKINNY. The results are depicted in Figure 12a and Figure 12b, where timeout (TO) is set to 1.5 times of the best one for the 13 round (i.e., 53630s × 1.5). The results for small rounds are negligible, thus are omitted here. We can observe that $\Psi_{\mathcal{S}} \cup \Psi_{\mathcal{S}}^{\mathsf{bn}}$ achieves the best performance, although it contains more IL constraints than $\Psi_{\mathcal{S}}$ and $\Psi_{\mathcal{S}}^3$. It seems that more constraints provide more information to improve the efficiency of MILP solving.

**Results of $\Psi_{\oplus}^1$ vs. $\Psi_{\oplus}^2$ vs. $\Psi_{\oplus}^3$.** We conduct experiments on LBLOCK (up to 13 rounds) and KLEIN (up to 5 rounds) that use the XOR operator and S-boxes for which $\Psi_{\mathcal{S}} \cup \Psi_{\mathcal{S}}^{\mathsf{bn}}$ and the technique from [Sun et al. 2014b] are used. The results are depicted in Figure 12c and Figure 12d, where

Table 12. Comparison of techniques for constructing $\Psi^3_{\mathcal{S}}$ of individual S-boxes

| **Size** | **S-box** | **$T_1$** | | **$T_2$** | | **$T_3$** | | **$T_4$** | | **$T_5$** | | **$T_6$** | | **$T_7$** | | **$T_8$** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^3_{\mathcal{S}}\rvert$ | **Time** |
| 4-bit | Elephant | 26 | 0s | 23 | 1s | 40 | 0s | 19 | 149s | 38 | 0s | 19 | 3s | **17** | 5s | **17** | 0s |
| | GIFT | 25 | 0s | 21 | 0s | 34 | 0s | 17 | 176s | 33 | 2s | 18 | 7s | 17 | 30s | **16** | 0s |
| | LBLOCK S0 | 27 | 0s | 24 | 0s | 30 | 0s | 17 | 108s | 30 | 1s | 17 | 6s | **15** | 30s | **15** | 0s |
| | LBLOCK S1 | 26 | 0s | 24 | 0s | 30 | 0s | 17 | 107s | 30 | 2s | 17 | 9s | **15** | 31s | **15** | 0s |
| | LBLOCK S2 | 27 | 0s | 24 | 0s | 30 | 0s | 17 | 131s | 30 | 3s | 17 | 10s | **15** | 30s | **15** | 0s |
| | LBLOCK S3 | 27 | 0s | 24 | 1s | 31 | 0s | 17 | 143s | 30 | 2s | 17 | 10s | **15** | 28s | **15** | 0s |
| | LBLOCK S4 | 27 | 0s | 24 | 0s | 30 | 0s | 17 | 144s | 30 | 3s | 17 | 10s | **15** | 30s | **15** | 0s |
| | LBLOCK S5 | 29 | 0s | 24 | 0s | 30 | 0s | 17 | 143s | 30 | 2s | 17 | 10s | **15** | 32s | **15** | 0s |
| | LBLOCK S6 | 26 | 0s | 24 | 0s | 30 | 0s | 17 | 140s | 30 | 2s | 17 | 10s | **15** | 31s | **15** | 0s |
| | LBLOCK S7 | 26 | 0s | 24 | 0s | 30 | 0s | 17 | 140s | 30 | 2s | 17 | 10s | **15** | 32s | **15** | 0s |
| | LBLOCK S8 | 26 | 0s | 24 | 1s | 31 | 0s | 17 | 139s | 30 | 2s | 17 | 9s | **15** | 30s | **15** | 0s |
| | LBLOCK S9 | 26 | 0s | 24 | 0s | 31 | 0s | 17 | 133s | 30 | 1s | 17 | 6s | **15** | 29s | **15** | 0s |
| | PICCOLO | 24 | 0s | 21 | 0s | 31 | 0s | 16 | 133s | 31 | 2s | 16 | 8s | **14** | 23s | **14** | 0s |
| | PRESENT | 24 | 0s | 21 | 0s | 39 | 0s | 17 | 274s | 36 | 1s | 17 | 5s | 17 | 19s | **16** | 0s |
| | RECTANGLE | 23 | 0s | 21 | 0s | 31 | 0s | 17 | 246s | 30 | 2s | 17 | 9s | **15** | 34s | **15** | 0s |
| | SKINNY | 24 | 0s | 21 | 0s | 31 | 0s | 16 | 135s | 31 | 2s | 16 | 7s | **14** | 35s | **14** | 0s |
| | TWINE | 27 | 0s | 23 | 0s | 47 | 0s | 20 | 291s | 45 | 2s | **19** | 5s | **19** | 7s | **19** | 0s |
| 5-bit | ASCON | 50 | 3s | 40 | 19s | 60 | 0s | 31 | 267535s | 59 | 130s | 48 | 14776s | 28 | 2022s | **27** | 72s |

timeout is set the same as above. We can observe that $\Psi^2_\oplus$ achieves the best performance, although it contains more IL constraints than $\Psi^3_\oplus$. It may be because $\Psi^2_\oplus$ does not introduce additional variables, while $\Psi^1_\oplus$ and $\Psi^3_\oplus$ introduce one additional Boolean variable for one bit.

**Results of $\Psi^1_{\mathtt{M,i,h}}$ vs. $\Psi^2_{\mathtt{M,i,h}}$ vs. $\Psi^3_{\mathtt{M,i,h}}$.** We conduct experiments on SKINNY (up to 9 rounds) that uses the matrix-vector product and S-boxes for which $\Psi_{\mathcal{S}} \cup \Psi^{\mathsf{bn}}_{\mathcal{S}}$ and the technique from [Sun et al. 2014b] are used. The results are depicted in Figure 12e, where timeout is set the same as above. We can observe that $\Psi^2_{\mathtt{M,i,h}}$ achieves the best performance, which is consistent to the comparison of $\Psi^1_\oplus$, $\Psi^2_\oplus$ and $\Psi^3_\oplus$ from which $\Psi^1_{\mathtt{M,i,h}}$, $\Psi^2_{\mathtt{M,i,h}}$ and $\Psi^3_{\mathtt{M,i,h}}$ are derived. Surprisingly, $\Psi^3_{\mathtt{M,i,h}}$ contains significantly fewer constraints, but achieves the worst performance.

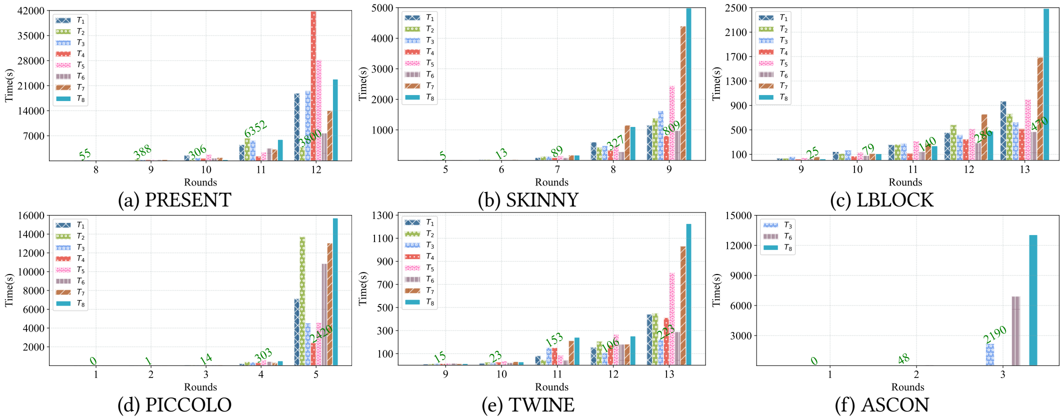
**Results of techniques for constructing $\Psi^3_{\mathcal{S}}$.** The techniques are named as follows for simplifying presentation: $\mathbf{T}_1$=[Sun et al. 2014b], $\mathbf{T}_2$=[Sasaki and Todo 2017], $\mathbf{T}_3$=[Abdelkhalek et al. 2017], $\mathbf{T}_4$=Alg. 1 of [Boura and Coggia 2020], $\mathbf{T}_5$=Alg. 2 of [Boura and Coggia 2020], $\mathbf{T}_6$=Alg. 2, Alg. 3 and Proposition 3 of [Boura and Coggia 2020], $\mathbf{T}_7$=[Li and Sun 2022], and $\mathbf{T}_8$=[Udovenko 2021].

We first compare their performance for constructing $\Psi^3_{\mathcal{S}}$ of 4-bit and 5-bit S-boxes from randomly chosen ciphers in terms of number of constraints and execution time. The results are reported in Table 12. We observe that

- $\mathbf{T}_8$ can efficiently produce the fewest constraints with the same number of Boolean variables,
- $\mathbf{T}_3$ is the most efficient but produces the most constraints,
- $\mathbf{T}_4$ takes the longest time. Note that the constraint coefficients produced by $\mathbf{T}_3$ are limited to $\{-1, 0, 1\}$ while the others do not.

We compare the performance for the overall security analysis in terms of execution time on randomly chosen ciphers using $\psi^2_\oplus$, $\Psi_{\mathcal{S}} \cup \Psi^{\mathsf{bn}}_{\mathcal{S}}$ and $\Psi^2_{\mathtt{M,i,h}}$. The results are depicted in Figure 13, where execution time of constructing $\Psi^3_{\mathcal{S}}$ is excluded as it is only computed once for each S-box. Overall, the performance varies with ciphers and round numbers, and no technique always outperforms the others. Interestingly, we found that

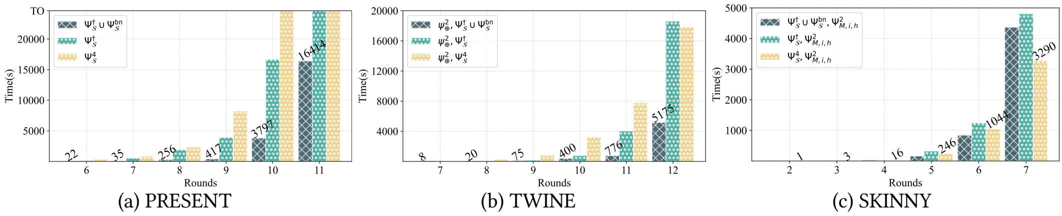
- though the recent promising techniques $\mathbf{T}_7$ and $\mathbf{T}_8$ produce the fewest constraints, they often performs worse than the others (e.g., on LBLOCK, PICCOLO and TWINE);
- though $\mathbf{T}_3$ produces the most constraints, it performs moderately on all the ciphers, because the generated constraints are much simpler;



(a) PRESENT (b) SKINNY (c) LBLOCK

(d) PICCOLO (e) TWINE (f) ASCON

Fig. 13. Comparison of techniques for constructing $\Psi^3_{\mathcal{S}}$ in the overall security analysis.



(a) PRESENT (b) TWINE (c) SKINNY

Fig. 14. Comparison of S-boxes modeling methods in the extended bit-wise approach.

Table 13. Comparison of techniques for constructing $\Psi^4_{\mathcal{S}}$ of individual S-boxes, where timeout (TO) is 1 hour.

| **S-box** | **T₁** | | **T₂** | | **T₃** | | **T₄** | | **T₅** | | **T₆** | | **T₇** | | **T₈** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** | $\lvert\Psi^4_{\mathcal{S}}\rvert$ | **Time** |
| GIFT | 27 | 1s | 19 | 3s | 50 | 0s | TO | | TO | | TO | | TO | | N/A | |

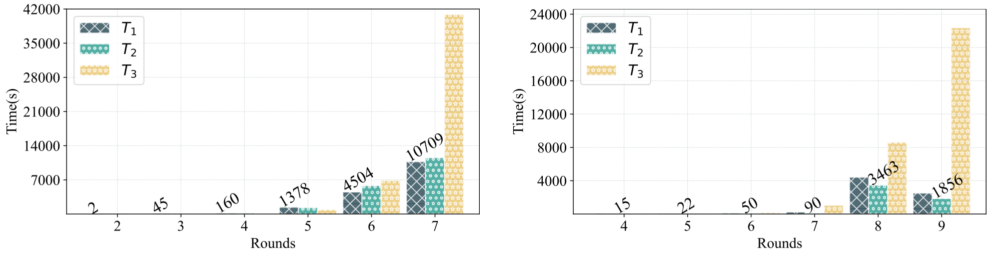
- though $\mathbf{T}_4$ and $\mathbf{T}_6$ take the longest time to construct $\Psi^3_{\mathcal{S}}$ and do not produce the fewest constraints, they in general perform well except for PRESENT with 12 rounds and PICCOLO with 5 rounds respectively.

In summary, not only the number but also coefficients of the constraints affect the overall efficiency. It is interesting to study the characterizations of optimal constraints in future.

## E.3 Comparison of Alternative Extended Bit-wise Modeling Methods

In our extended bit-wise approach, there are also three alternatives for modeling S-boxes, namely, $\Psi^\dagger_{\mathcal{S}} \cup \Psi^{\mathrm{bn}}_{\mathcal{S}}$, $\Psi^\dagger_{\mathcal{S}}$ and $\Psi^4_{\mathcal{S}}$. Note that $\Psi^\dagger_{\mathcal{S}} = \Psi^1_{\mathcal{S}} \cup \Psi^2_{\mathcal{S}} \cup \Psi^4_{\mathcal{S}}$ if the S-box $\mathcal{S}$ is injective, otherwise $\Psi^\dagger_{\mathcal{S}} = \Psi^1_{\mathcal{S}} \cup \Psi^4_{\mathcal{S}}$, where $\Psi^1_{\mathcal{S}}$, $\Psi^2_{\mathcal{S}}$ and $\Psi^{\mathrm{bn}}_{\mathcal{S}}$ are redundant.

**Results of $\Psi^\dagger_{\mathcal{S}} \cup \Psi^{\mathrm{bn}}_{\mathcal{S}}$ vs. $\Psi^\dagger_{\mathcal{S}}$ vs. $\Psi^4_{\mathcal{S}}$.** We conduct experiments on PRESENT (up to 11 rounds), TWINE (up to 12 rounds) and SKINNY (up to 7 rounds), where $\Psi^4_{\mathcal{S}}$ is constructed using the technique from [Sun et al. 2014b]. The results are given in Figure 14a, Figure 14b and Figure 14c, where timeout is set the same as above, i.e., 1.5 times of the best one for the largest round number. We can observe that $\Psi^\dagger_{\mathcal{S}} \cup \Psi^{\mathrm{bn}}_{\mathcal{S}}$ achieves the best performance for all rounds of PRESENT and TWINE, but $\Psi^4_{\mathcal{S}}$ performs better than $\Psi^\dagger_{\mathcal{S}} \cup \Psi^{\mathrm{bn}}_{\mathcal{S}}$ on SKINNY. Thus, overall, $\Psi^\dagger_{\mathcal{S}} \cup \Psi^{\mathrm{bn}}_{\mathcal{S}}$ achieves the best performance.



(a) GIFT-COFB

(b) GIFT

Fig. 15. Comparison of techniques for constructing $\Psi^4_{\mathcal{S}}$ in the overall security analysis.

**Results of techniques for constructing $\Psi^4_{\mathcal{S}}$** We compare the techniques $\mathbf{T}_i$, $1 \leq i \leq 8$, for constructing $\Psi^4_{\mathcal{S}}$ using GIFT-COFB and GIFT. Note that GIFT-COFB and GIFT use the same S-box.

The comparison of the techniques $\mathbf{T}_i$, $1 \leq i \leq 8$ for constructing $\Psi^4_{\mathcal{S}}$ of GIFT-COFB (GIFT) in terms of number of constraints and execution time are given in Table 13, where timeout (TO) is set to 1 hour. We can observe that

- $\mathbf{T}_2$ produces the fewest constraints,
- $\mathbf{T}_3$ is the most efficient,
- $\mathbf{T}_4$, $\mathbf{T}_5$, $\mathbf{T}_6$ and $\mathbf{T}_7$ all run out of time,
- $\mathbf{T}_8$ fails due to bugs.

Note that we use the open-source implementations of $\mathbf{T}_7$ and $\mathbf{T}_8$ provided by their authors and the bugs have been reported to the author of $\mathbf{T}_8$.

The comparison of the techniques for the overall security analysis in terms of execution time are reported in Figure 15a and Figure 15b respectively, using $\Psi^{\dagger}_{\mathcal{S}} \cup \Psi^{\mathsf{bn}}_{\mathcal{S}}$. We found that $\mathbf{T}_1$ and $\mathbf{T}_2$ are comparable, and performs better than $\mathbf{T}_3$, probably because $\mathbf{T}_3$ produces too many constraints.

### E.4 Bounding Condition

Inspired by [Zhang et al. 2018], we formalize the bounding condition of [Matsui 1994] for the bit-wise approach, where the formalizations for the word-wise approach and the extended bit-wise approach can be defined similarly.

Suppose we are going to determine the lower bound of the minimum number of active S-boxes of $r$-round differential characteristics. We will add the following additional constraints into the set of IL constraints $\Phi \cup \{\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1\}$ that are constructed by our bit-wise approach (cf. Section 6):

(1) for every $1 \leq i \leq r$, $\sum_{j=0}^{s\cdot n-1} \gamma(x_i, j) \geq 1$ is added, where $x_i$ denotes the variable assigned by the return of the $i$-th round function, and its type is `uints`$[n]$. Intuitively, the output difference $\gamma(x_i, j)$ of each round function should be nonzero when the input difference of the cipher is nonzero which is ensured by $\sum_{i=0}^{s\cdot n-1} \gamma(txt, i) \geq 1$.

(2) for every $1 \leq i < r$, $\sum_{j=1}^{i} \sum_{b\in\Theta_j} b \geq N_i$ is added, where $\Theta_j$ denotes the set of Boolean variables modeling the activeness of the S-boxes in the $j$-th round and $N_i$ is a lower bound of the number of active S-boxes of $i$-round differential characteristics. Intuitively, the number of active S-boxes of $r$-round differential characteristics that occurs in the first $i$-rounds should be no less than $N_i$. Furthermore, $\sum_{j=r-i}^{r} \sum_{b\in\Theta_j} b \geq N_i$ is also added if the first $i$-rounds and the $r-1$ to $r$ rounds are the same.

Table 14. Results of the bit-wise approach without the bounding condition, Timeout is 24 hours.

| **Rounds** | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ASCON (12) | **#AS** | 1 | 4 | 15 | N/A | | | | | | | | | | | |
| | **Time** | 1s | 31s | 6913s | Timeout | | | | | | | | | | | |
| ELEPHANT (80) | **#AS** | 1 | 2 | 4 | 6 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | N/A | | | |
| | **Time** | 0s | 0s | 2s | 12s | 66s | 428s | 761s | 2040s | 3834s | 7557s | 22957s | Timeout | | | |
| GIFT-COFB (40) | **#AS** | 1 | 2 | 3 | 5 | 7 | 10 | 13 | 17 | 19 | 21 | 23 | N/A | | | |
| | **Time** | 0s | 0s | 1s | 6s | 26s | 49s | 1074s | 3578s | 5744s | 42615s | 66725s | Timeout | | | |
| GIFT (28) | **#AS** | 1 | 2 | 3 | 5 | 7 | 10 | 13 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 |
| | **Time** | 0s | 0s | 0s | 2s | 4s | 6s | 38s | 240s | 276s | 308s | 6412s | 3450s | 9672s | 13890s | 51175s |
| PRESENT (31) | **#AS** | 1 | 2 | 4 | 6 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | N/A | |
| | **Time** | 0s | 0s | 0s | 2s | 5s | 11s | 54s | 27s | 277s | 745s | 3515s | 7750s | 33215s | Timeout | |
| RECTANGLE (25) | **#AS** | 1 | 2 | 3 | 4 | 6 | 8 | 11 | 13 | 15 | 17 | 19 | 21 | 23 | N/A | |
| | **Time** | 0s | 0s | 1s | 1s | 3s | 7s | 20s | 27s | 115s | 292s | 1683s | 2322s | 28772s | Timeout | |
| SKINNY (36) | **#AS** | 1 | 2 | 5 | 8 | 12 | 16 | 26 | 36 | 41 | 46 | 51 | 55 | N/A | | |
| | **Time** | 0s | 0s | 1s | 2s | 6s | 12s | 88s | 291s | 1672s | 3337s | 7391s | 25231s | Timeout | | |
| SPARKLE (7) | **#AS** | 1 | 2 | 5 | 6 | 7 | N/A | | | | | | | | | |
| | **Time** | 2s | 44s | 1624s | 4416s | 17592s | Timeout | | | | | | | | | |

### E.5 Results of the Bit-wise Approach without Bounding Condition

The results of our bit-wise approach without adding the bounding condition are shown in Table 14. By comparing with the results given in Table 6, we can observe that on ASCON, ELEPHANT, PRESENT, RECTANGLE and SPARKLE, the bit-wise approach with the bounding condition achieves a better performance than the one without the bounding condition. in particular, the improvement brought by the bounding condition on RECTANGLE is more than 10 times for large round numbers. However, the bit-wise approach with bounding condition performs worse than the one without the bounding condition on GIFT-COFB and GIFT. This indicates that the bounding condition does not necessarily improve MILP solving efficiency, but overall it is recommended to be used.

## F SUMMARY OF RELATED WORK

To facilitate the comparison with the related work, we give a summary of the related work in Table 15, which lists the existing MILP/SAT/SMT-based approaches for the security analysis of cryptographic primitives against differential cryptanalysis.

In Table 15, the second column shows the supported operations in the respective work; the third column indicates whether its modeling is word-wise, bit-wise or extended bit-wise (Note that the extended bit-wise here means that the probabilities of bit-level input and output differences of operations are encoded in constraints); the fourth column gives the techniques used for S-box modeling (in bit-wise or extended bit-wise approach) and the last column gives the main strategy for the security analysis, recall that #AS denotes the lower bound of the minimum number of active S-boxes and Pr denotes the upper bound of the probability of optimal differential characteristics.

In the techniques shown in the fourth column, the H-representation (H-Rep.) generation of convex hull, the logic condition (Log.) method, the conjunctive normal form (CNF) of Boolean functions, the methods in [Boura and Coggia 2020] (Alg.1, Alg.2, Alg.3 and Proposition 3), the SuperBall method and the method in [Udovenko 2021] applying monotone Boolean functions are the methods for generating linear inequalities for modeling the possible differential propagation in S-boxes; the greedy algorithm (greedy Alg.) and the MILP-based algorithm in [Sasaki and Todo 2017] (MILP Alg.) are the methods for selecting a subset of the generated inequalities.

Received 2023-07-11; accepted 2023-11-07

Table 15. Summary of existing MILP/SAT/SMT-based approaches, where H-Rep. denotes H-representation and Log. denotes logical condition modeling

| Ref | Supported operations | Word-/(Extended) Bit-wise | $\Psi_{\mathcal{S}}$ for S-box | Evaluation Strategy |
|---|---|---|---|---|
| [Mouha et al. 2011] | XOR, S-box, Linear transformation | Word-wise | N/A | MILP+#AS |
| [Sun et al. 2013] | XOR, S-box, Linear transformation | Bit-wise | None | MILP+#AS |
| [Sun et al. 2014a] | XOR, AND, S-box, Linear transformation | Bit-wise | H-Rep.+gready Alg. | MILP+#AS |
| [Sun et al. 2014b] | XOR, AND, S-box, Linear transformation | (Extended) Bit-wise | H-Rep.+gready Alg. | MILP+#AS/Pr |
| [Sasaki and Todo 2017] | XOR, S-box | Bit-wise | H-Rep./Log.+MILP Alg. | N/A |
| [Abdelkhalek et al. 2017] | XOR, S-box | Word-wise<br>(Extended) Bit-wise | CNF+QM Alg. | MILP+#AS/Pr |
| [Boura and Coggia 2020] | XOR, S-box, Linear transformation | Bit-wise | Alg. 1, Alg. 2, Alg. 3 or Proposition 3 + MILP Alg. | N/A |
| [Sun 2021] | S-box | Bit-wise | SuperBall + MILP Alg. | N/A |
| [Li and Sun 2022] | S-box | Bit-wise | SuperBall + MILP Alg. | MILP+#AS |
| [Udovenko 2021] | S-box | Bit-wise | Monotone Boolean functions + MILP Alg. | N/A |
| [Cui et al. 2016] | XOR, S-box, modular addition | Bit-wise | [Sun et al. 2014b,a] | N/A |
| [Li et al. 2019] | XOR, S-box | Extended Bit-wise | [Sun et al. 2014b,a] | MILP+Pr |
| [Yin et al. 2017] | XOR, modular addition | Extended Bit-wise | None | MILP+Pr |
| [Ilter and Selçuk 2021] | XOR, S-box, Linear transformation | (Extended) Bit-wise | [Sun et al. 2014b]<br>[Sasaki and Todo 2017] | MILP+#AS/Pr |
| [Zhang and Zhang 2018] | XOR, S-box, Linear transformation | Bit-wise | [Sun et al. 2014a] | MILP+#AS |
| [Fu et al. 2016] | XOR, modular addition | Extended Bit-wise | None | MILP+Pr |
| [Wang et al. 2018] | XOR, AND | Extended Bit-wise | None | MILP+Pr |
| [Zhang et al. 2018] | XOR, S-box modular addition | Extended Bit-wise | [Sun et al. 2014b,a] | MILP+Pr |
| [Zhou et al. 2019] | XOR, S-box, Linear transformation | Word-wise<br>(Extended) Bit-wise | [Sasaki and Todo 2017]<br>[Sun et al. 2014b,a] | MILP+#AS/Pr |
| [Mouha and Preneel 2013] | XOR, modular addition | Extended Bit-wise | None | SAT+Pr |
| [Aumasson et al. 2014] | XOR, AND | Extended Bit-wise | None | SAT/SMT+Pr |
| [Kölbl et al. 2015] | XOR, AND | Extended Bit-wise | None | SAT/SMT+Pr |
| [Song et al. 2016] | XOR, modular addition | Extended Bit-wise | None | SAT+Pr |
| [Liu et al. 2021] | XOR, S-box | (Extended) Bit-wise | None | N/A |
| [Sun et al. 2018] | XOR, S-box, Linear transformation | Extended Bit-wise | None | SAT+Pr |
| [Azimi et al. 2022] | XOR, modular addition | Extended Bit-wise | None | SMT+Pr |
| [Makarim and Rohit 2022] | XOR, S-box | (Extended) Bit-wise | [Sun et al. 2014a]<br>[Abdelkhalek et al. 2017] | MILP/SMT+#AS/Pr |
| [Sun et al. 2021] | XOR, AND, S-box, modular addition | (Extended) Bit-wise | None | SAT+#AS/Pr |