基于分治策略的 SAT 差分自动化搜索算法及其应用

胡斌[1]，谈潇[1]，王森鹏[1,2]
（1.信息工程大学密码工程学院，河南 郑州 450001；2. 密码科学技术国家重点实验室，北京 100878）

摘 要：为了提高自动化搜索效率，结合分治策略提出了一种基于 SAT 模型的最优差分特征搜索算法。利用任意部分连续轮的 Matsui 边界条件提供的信息，将搜索空间划分为互不相交的子集。通过分析 SAT 差分模型间的可满足性关系，提出一种降序分支搜索链模型。进一步地，在模型优化层面，减少了需搜索划分子集数量的方法；在算法实现层面，结合并行技术实现对模型搜索空间的约减。将加速算法应用于 ARX 类密码算法族 SPECK，获得了 20 轮、14 轮、11 轮 SPECK-48、SPECK-96、SPECK-128 的最优差分特征，较现有最好结果分别提高了 1 轮、4 轮、2 轮。

关键词：差分特征；分组密码；自动化搜索；分治策略
中图分类号： TP391
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2023082

SAT-based differential automatic search algorithm using divide-and-conquer strategy and its applications

HU Bin[1], TAN Xiao[1], WANG Senpeng[1,2]
1. Department of Cryptogram Engineering, Information Engineering University, Zhengzhou 450001, China
2. State Key Laboratory of Cryptology, Beijing 100878, China

Abstract: To improve the efficiency of automatic search, an algorithm for searching the optimal differential characteristics based on SAT model was proposed by combining the divide-and-conquer strategy. The search space was divided into disjoint subsets by using the information from Matsui boundary conditions of arbitrary continuous rounds. By analyzing the relationships between satisfiability of differential models based on SAT, a descending branch search chain model was proposed. Furthermore, at the model optimization level, the number of subsets that need to be searched and partitioned was decreased. At the level of algorithm implementation, the search space was reduced by utilizing the parallel technology. Finally, the accelerated algorithm was applied to SPECK family of ARX cryptographic algorithms. The 20, 14, 11-round optimal differential characteristics of SPECK-48, SPECK-96, SPECK-128 are obtained, which increase the previous best results by 1, 4, 2 rounds respectively.

Keywords: differential characteristic, block cipher, automatic search, divide-and-conquer strategy

0 引言

ARX（addition-rotation-xor）类密码算法为仅由模加、循环移位和异或运算构成的算法，具有软件实现简单、运行快速且易于替换等优点，常被视为基于 S 盒设计的密码算法，如数据加密标准（DES, data encryption standard）[1]和高级加密标准（AES, advanced encryption standard）[2]算法的替代者，在轻量级算法领域广受欢迎。许多高效的轻量级密码算法都是基于 ARX 的结构设计的，如 TEA（tiny encryption algorithm）[3]、Salsa20[4]、SipHash[5]、SPECK[6]、LEA[7]和 Chaskey[8]等。

收稿日期：2022–11–09；**修回日期：**2023–02–03
通信作者：王森鹏，wsp2110@126.com
基金项目：国家自然科学基金资助项目（No.62102448）
Foundation Item: The National Natural Science Foundation of China (No.62102448)

差分分析是对称密码算法最有效的分析方法之一，由 Biham 和 Shamir[9]提出，并用于攻击 DES。经过多年的发展，差分分析已经成为评估对称密码算法安全强度的重要准则，并通过搜索高概率的差分特征作为初步评估的依据。常用的搜索高概率的差分特征的方法有 3 种：Matsui 分支定界算法[10]、基于混合整数线性规划（MILP, mixed integer linear programming）[11]的自动化搜索和基于布尔可满足性（SAT）问题[12]的自动化搜索。

在 ARX 类密码算法的差分特征搜索方面，由于算法采用模加代替 S 盒，构造完整的差分分布表在实现上不可行，因此许多 ARX 类密码算法差分特征的搜索是通过构造满足一定阈值的部分差分分布表、限制差分特征的汉明重量或引入随机算法等方式实现的。例如，Biryukov 和 Velichkov[13]利用部分差分分布表技术和 Mutsui 分支定界算法，提出了第一个 ARX 类密码差分特征的自动化搜索算法，应用于(X)TEA 和 SPECK。Song 等[14]对模加结构进行分割，固定部分输入变量条件搜索高概率差分特征。这些方法得到的结果是启发式的，不能保证差分特征的全局最优性。

为了得到 ARX 类密码算法更高轮数的全局最优差分特征，学者提出了一些改进的搜索方法。例如，Biryukov 等[15]通过改进 Matsui 分支定界算法，提出 ARX 类密码最优差分特征和线性特征的搜索算法，该算法不是基于求解器的自动化搜索算法。Fu 等[16]利用 Lipmaa 等[17]给出的模加差分性质刻画结果，将基于 MILP 的自动化搜索算法扩展到 ARX 类密码算法中，用于寻找最佳差分/线性特征。Zhang 等[18]提出在 MILP 模型中添加 Matsui 边界条件的方法进行加速搜索，但对于 ARX 类密码算法 SPECK 来说并没有取得更好的结果。Sun 等[19]将 Matsui 边界条件加入 SAT 模型中，大幅提高了最优差分特征搜索的轮数。Wang 等[20]研究了 Matsui 边界条件提供的全部限制信息，给出了最简边界条件的模型，减少所刻画子句和变量的个数，进一步提高了全局搜索的效率。但是对于 ARX 类密码算法，如最简单的 SPECK 族分组密码算法，搜索轮数和效率仍然不高。为了解决这一问题，本文在上述工作的启发下，尝试将 SAT 搜索算法与分治策略相结合。本文工作的主要贡献如下。

1) 提出了一种划分子集的方法。从 Mutsui 边界条件提供的完整信息着手，选取其中任意连续轮，按照概率的所有可能取值对模型搜索空间进行划分，即子集划分。同时，给出了所划分的子集之间的关系，以及它们的可满足性的联系。

2) 构建新的有序搜索链模型。在满足求解的充要条件的前提下，结合本文子集划分方法、子集包含关系与状态的讨论以及求解器的特点，确定了搜索顺序以及搜索节点的选取策略，有效约减了搜索空间并提高了搜索效率。

3) 将加速搜索算法应用于 ARX 类密码算法 SPECK。在选择了具体划分条件后，应用加速搜索算法寻找 SPECK-48、SPECK-96、SPECK-128 的全局最优差分特征，提升了搜索效率，得到了更高轮数的最优差分特征。

1 预备知识

1.1 SAT 问题转化

SAT 问题即布尔可满足性问题，它考虑给定布尔函数表达式的可满足性。即给定二元变量集合 X 和 X 上的布尔函数表达式 $F(X)$，判断是否存在一个真值指派使 $F(X)=1$。布尔函数表达式属于命题逻辑公式，可以通过布尔代数运算律转化为等价的合取范式形式（CNF, conjunctive normal form），即用“与”运算连接一系列由“或”运算连接的子句的形式。

对于分组密码算法差分安全性评估，一般需要估计密码算法的最大差分概率。不考虑差分聚集效应时，往往通过搜索最大差分特征概率来进行初步评估。基于 SAT 问题的自动化搜索方法将差分和概率用二元变量表示，差分传播和概率约束用合取范式形式刻画，将差分安全性评估问题转化为可调用 SAT 求解器求解的数学模型。一个满足 SAT 模型的真值指派即对应一条可能的差分特征，则所有可能的差分特征对应 SAT 模型的解空间。

1.2 基本 SAT 模型刻画

ARX 类密码包含 3 种密码组件：异或、模加、循环移位。在刻画 ARX 类密码算法差分传播时，由于循环移位操作的输出差分变量可通过改变输入差分变量的索引直接刻画，因此这里只对异或和模加这 2 种密码组件的刻画模型进行介绍。

模型 1 异或[19]。令 $\boldsymbol{\gamma}=f(\boldsymbol{\alpha},\boldsymbol{\beta})$ 是 $2n$ bit 输入的异或操作，其中，$\boldsymbol{\alpha}=(\alpha_0,\alpha_1,\cdots,\alpha_{n-1})\in F_2^n$ 和 $\boldsymbol{\beta}=(\beta_0,\beta_1,\cdots,\beta_{n-1})\in F_2^n$ 为输入变量，$\boldsymbol{\gamma}=(\gamma_0,\gamma_1,\cdots,\gamma_{n-1})\in F_2^n$ 为输出变量，且满足 $\gamma_i=\alpha_i\oplus\beta_i$，$\forall i,0\leqslant i\leqslant n-1$。则逻辑表达式(1)可以刻画比特差分在异或运算中的传播规律。

$$\begin{aligned}
&\alpha_i \vee \beta_i \vee \overline{\gamma_i} &&=1\\
&\alpha_i \vee \overline{\beta_i} \vee \gamma_i &&=1\\
&\overline{\alpha_i} \vee \beta_i \vee \gamma_i &&=1\\
&\overline{\alpha_i} \vee \overline{\beta_i} \vee \overline{\gamma_i} &&=1
\end{aligned} \tag{1}$$

其中，$0 \leqslant i \leqslant n-1$。

模型 2　模加[19]。设 $\boldsymbol{\gamma}=f(\boldsymbol{\alpha},\boldsymbol{\beta})$ 是 n bit 模加操作，其中，$\boldsymbol{\alpha}=(\alpha_0,\alpha_1,\cdots,\alpha_{n-1})\in F_2^n$ 和 $\boldsymbol{\beta}=(\beta_0,\beta_1,\cdots,\beta_{n-1})\in F_2^n$ 为输入变量，$\boldsymbol{\gamma}=(\gamma_0,\gamma_1,\cdots,\gamma_{n-1})\in F_2^n$ 为输出变量。则 $(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\in F_2^{3\times n}$ 是一个可能的差分对应，当且仅当满足以下条件

$$\begin{aligned}
&\alpha_{n-1}\oplus\beta_{n-1}\oplus\gamma_{n-1}=0\\
&\alpha_i \vee \beta_i \vee \overline{\gamma_i} \vee \alpha_{i+1} \vee \beta_{i+1} \vee \gamma_{i+1} &&=1\\
&\alpha_i \vee \overline{\beta_i} \vee \gamma_i \vee \alpha_{i+1} \vee \beta_{i+1} \vee \gamma_{i+1} &&=1\\
&\overline{\alpha_i} \vee \beta_i \vee \gamma_i \vee \alpha_{i+1} \vee \beta_{i+1} \vee \gamma_{i+1} &&=1\\
&\overline{\alpha_i} \vee \overline{\beta_i} \vee \overline{\gamma_i} \vee \alpha_{i+1} \vee \beta_{i+1} \vee \gamma_{i+1} &&=1\\
&\alpha_i \vee \beta_i \vee \gamma_i \vee \overline{\alpha_{i+1}} \vee \overline{\beta_{i+1}} \vee \overline{\gamma_{i+1}} &&=1\\
&\alpha_i \vee \overline{\beta_i} \vee \overline{\gamma_i} \vee \overline{\alpha_{i+1}} \vee \overline{\beta_{i+1}} \vee \overline{\gamma_{i+1}} &&=1\\
&\overline{\alpha_i} \vee \beta_i \vee \overline{\gamma_i} \vee \overline{\alpha_{i+1}} \vee \overline{\beta_{i+1}} \vee \overline{\gamma_{i+1}} &&=1\\
&\overline{\alpha_i} \vee \overline{\beta_i} \vee \gamma_i \vee \overline{\alpha_{i+1}} \vee \overline{\beta_{i+1}} \vee \overline{\gamma_{i+1}} &&=1
\end{aligned} \tag{2}$$

其中，$0 \leqslant i \leqslant n-2$。为了刻画模加的差分转移概率，引入 $n-1$ 个二元变量 $w_0,w_1,\cdots,w_{n-2}$，满足以下条件

$$\begin{aligned}
&\alpha_{i+1} \vee \gamma_{i+1} \vee w_i &&=1\\
&\beta_{i+1} \vee \overline{\gamma_{i+1}} \vee w_i &&=1\\
&\alpha_{i+1} \vee \overline{\beta_{i+1}} \vee w_i &&=1\\
&\alpha_{i+1} \vee \beta_{i+1} \vee \gamma_{i+1} \vee \overline{w_i} &&=1\\
&\overline{\alpha_{i+1}} \vee \overline{\beta_{i+1}} \vee \overline{\gamma_{i+1}} \vee \overline{w_i} &&=1
\end{aligned} \tag{3}$$

其中，$0 \leqslant i \leqslant n-2$。则差分转移概率可由 $p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})=2^{-\sum_{i=0}^{n-2} w_i}$ 计算，其中，w_i 为重量变量。

除了刻画密码组件的模型之外，还需考虑 R 轮差分特征概率的约束条件。因为本文工作基于马尔可夫密码假设，考虑单密钥情况的最优差分特征概率。对于一个分组规模为 N bit 的 R 轮迭代分组密码算法 E，R 轮差分特征概率可用每轮差分转移概率的乘积来表示，其指数的绝对值称为 R 轮差分转移概率的重量。则基本的 SAT 模型还有以下两个必不可少的约束条件。

1) 目标约束条件。由于求解目标是获得最大概率差分特征，则 R 轮差分转移概率的重量应该被限制为一个特定的值，或不超过重量变量的总数。

2) 初始输入差分约束条件。为了保证结果的非平凡性，第一轮输入差分变量应不全为 0。

1.3　最简 Matsui 边界条件

Matsui 边界条件来源于 Matsui[10]提出的分支定界算法。可以体现其算法思想的一种简单的情况如下：前 r 轮实际差分转移概率+后 $R-r$ 轮最大差分转移概率$\leqslant R$ 轮初始差分转移概率。即对前 r 轮差分变量进行赋值，判断是否满足该不等式条件，若不满足，则前 r 轮的赋值不可能得到最优差分概率特征；若满足，则继续对下一轮赋值。直到 R 轮完全赋值为止，更新 R 轮最大差分转移概率的下界，再进行下一分支的搜索。引理 1 给出 Matsui 边界条件一般形式，并且进一步给出前 r 轮实际差分转移概率的上界表示。

引理 1　边界条件一般形式[20]。对于一个 R 轮的分组密码算法，令 $P_{\text{ini}}(R)$ 为差分概率边界的初始估计值，假设已知前 i 轮的最优差分概率边界 $P_{\text{opt}}(i),1\leqslant i\leqslant R-1$。则式(4)所示的边界条件可以利用由 $\left\{P_{\text{ini}}(R),P_{\text{opt}}(i),1\leqslant i\leqslant R-1\right\}$ 提供的所有信息。

$$\begin{aligned}
&\sum_{r}^{i=j}(-\text{lb}(p(\alpha^i\rightarrow\alpha^{i+1})))\leqslant \text{lb}P_{\text{opt}}(j)+\\
&\text{lb}P_{\text{opt}}(R-1-r)-\text{lb}P_{\text{ini}}(R)\\
&\sum_{r}^{i=j}(-\text{lb}(p(\alpha^i\rightarrow\alpha^{i+1})))\geqslant -\text{lb}P_{\text{opt}}(r+1-j)
\end{aligned} \tag{4}$$

其中，$0\leqslant j\leqslant r\leqslant R-1$。

通过对上述 Matsui 边界条件进行抽象，则所有的 Matsui 边界条件均可以表示为不等式约束形式，即

$$l_{e_1,e_2}\leqslant\sum_{i=e_1}^{e_2}w_i\leqslant m_{e_1,e_2} \tag{5}$$

模型 3　最简边界条件[20]。最简边界条件模型建立在 Sun 等[19]的序列编码模型的基础上，利用辅助 MILP 模型求解出引理 1 中式(5)的精确上下界，并由此给出重量变量的约束条件。该模型利用了 Matsui 边界条件提供的所有信息，并使整合的 Matsui 边界条件对应的重量变量个数最小，从而减少序列编码模型刻画所需 CNF 子句的数量[20]。

2　结合分治策略的加速搜索算法

2.1　差分搜索模型子集划分技术

针对 ARX 类密码长轮模型求解困难的问题，

本文受最简边界条件模型的启发，提出一种将大规模搜索空间划分为若干小规模子集的分治方法，通过解决小规模子集空间的搜索问题，从而完成大规模空间的搜索。

对于一个 R 轮分组密码算法，Matsui 边界条件刻画了部分连续轮的概率边界范围。记 $E(i_1,i_2)$ 为从第 i_1 轮到第 i_2 轮的部分密码算法，满足 $i_1 \leqslant i_2$。设 $W(i_1,i_2)$ 为从第 i_1 轮到第 i_2 轮密码算法对应的差分转移概率重量。当 $i_1=1, i_2=R$ 时，$E(1,R)=E$，简记 $W(1,R)$ 为 W。当 $i_1=i_2=i$ 时，$E(i,i)$ 表示密码算法的第 i 轮，简记为 E_i，$W(i,i)$ 表示第 i 轮差分转移概率重量，简记为 W_i。

令 $B_{\text{opt}}(i), i\in\{1,\cdots,R-1\}$ 表示已知的前 i 轮的最小重量，$B_{\text{ini}}(R)$ 表示 R 轮的初始重量，$B_{\text{target}}(R)$ 表示 R 轮待判断的目标重量。

定理 1 由引理 1 中式(4)可以求得 $W(i_1,i_2)$ 的取值范围为

$$B_{\text{opt}}(i_2-i_1+1) \leqslant W(i_1,i_2) \leqslant B_{\text{target}}(R)-B_{\text{opt}}(i_1-1)-B_{\text{opt}}(R-i_2) \tag{6}$$

其中，$W(i_1,i_2)=\sum_{i=i_1}^{i_2}(-\text{lb}(p(\alpha^i\to\alpha^{i+1})))$。

定义 1 对任意一个 R 轮密码的差分 SAT 模型，其 R 轮重量为 W。选择 R 轮中任意一段连续轮密码算法 $E(i_1,i_2)$，利用其重量 $W(i_1,i_2)$ 确定划分子集，用符号表示为 $M_{R,W,W(i_1,i_2)}$。对 $W(i_1,i_2)$ 的任意一个取值 ϖ，子集的约束条件表示为

$$W_i+W_{i+1}+\cdots+W_j=\varpi \tag{7}$$

且 $B_{\text{opt}}(i_2-i_1+1)\leqslant\varpi\leqslant W-B_{\text{opt}}(i_1-1)-B_{\text{opt}}(R-i_2)$ 成立。其中，任意一个子集的可满足性都存在以下 3 种状态：可满足、不可满足和未知。

上述子集划分方法从边界条件出发，其约束条件是仅与概率重量变量有关的线性等式，因此将该等式添加到辅助 MILP 模型中即可实现对子集的约束，详细描述如算法 1 所示。

算法 1 辅助 MILP 算法

输入 边界条件集合 D，概率重量变量部分和 $\sum_{i=e_1}^{e_2}w_i$，部分连续轮 $E(i_1,i_2)$ 的重量 ϖ

输出 $l_{e_1,e_2}\leqslant\sum_{i=e_1}^{e_2}w_i\leqslant m_{e_1,e_2}$

1) 创建空 MILP 模型 M

2) for $d\in D$ do

3) 添加当前 Matsui 边界条件 d

4) end for

5) 添加子集约束条件 $W_i+W_{i+1}+\cdots+W_j=\varpi$

6) 创建下界模型 $M_l=M$

7) 添加目标函数，求 $\sum_{i=e_1}^{e_2}w_i$ 最小值

8) 调用 MILP 求解器，得到精确下界 l_{e_1,e_2}

9) 创建上界模型 $M_m=M$

10) 添加目标函数，求 $\sum_{i=e_1}^{e_2}w_i$ 最大值

11) 调用 MILP 求解器，得到精确上界 m_{e_1,e_2}

引理 2 给出一般情况下 SAT 模型搜索空间的包含关系与模型可满足性之间的联系。进一步地，考虑本节中划分方法在同一部分连续轮 $E(i_1,i_2)$ 的选取下，存在的两类具有包含关系的模型搜索空间组合的可满足性。

引理 2 设 M_1、M_2 为任意 2 个 SAT 模型，若 M_1 的搜索空间包含在 M_2 的搜索空间中，记为 $M_1\subseteq M_2$。则 M_1、M_2 的可满足性满足下面 2 个规则。

1) M_2 为不可满足 $\Rightarrow M_1$ 为不可满足。

2) M_1 为可满足 $\Rightarrow M_2$ 为可满足。

命题 1 固定总轮数 R、部分连续轮 $E(i_1,i_2)$ 和对应重量 $W(i_1,i_2)$，设 W_1、W_2 为 W 的 2 个取值，满足 $W_1\leqslant W_2$。则子集 $M_{R,W_1,W(i_1,i_2)}$ 和 $M_{R,W_2,W(i_1,i_2)}$ 的搜索空间存在如下包含关系

$$M_{R,W_1,W(i_1,i_2)}\subseteq M_{R,W_2,W(i_1,i_2)} \tag{8}$$

由引理 2 可知，其模型可满足性满足以下关系。

1) $M_{R,W_2,W(i_1,i_2)}$ 为不可满足 $\Rightarrow M_{R,W_1,W(i_1,i_2)}$ 为不可满足。

2) $M_{R,W_1,W(i_1,i_2)}$ 为可满足 $\Rightarrow M_{R,W_2,W(i_1,i_2)}$ 为可满足。

命题 2 固定总轮数 R 和 R 轮重量上界 W，选取任意部分连续轮 $E(i_1,i_2)$ 和对应重量 $W(i_1,i_2)$，显然原集合 $M_{R,W}$ 和划分子集 $M_{R,W,W(i_1,i_2)}$ 之间的关系为

$$\begin{cases} M_{R,W}=\bigcup\limits_{W(i_1,i_2)} M_{R,W,W(i_1,i_2)} \\ M_{R,W,W(i_1,i_2)}M_{R,W,W'(i_1,i_2)}=\varnothing, \forall W(i_1,i_2)\neq W'(i_1,i_2) \end{cases} \tag{9}$$

由引理 2 可知，其模型可满足性满足以下关系。

1) $M_{R,W}$ 为不可满足 $\Leftrightarrow \exists M_{R,W,W(i_1,i_2)}$ 为不可满足。

2) $M_{R,W}$ 为可满足 $\Leftrightarrow \forall M_{R,W,W(i_1,i_2)}$ 为可满足。

2.2　差分搜索模型子集的搜索模式

本节研究差分搜索划分子集的搜索模式。首先给出判断最优差分概率的充要条件。

命题 3　固定总轮数 R 并选取一个部分连续轮 $E(i_1,i_2)$，则在子集维度上，使 $M_{R,W}$ 判断为可满足时，W 的取值恰好为 $B_{\text{opt}}(R)$ 的充要条件如下。

1) $\exists M_{R,W,W(i_1,i_2)}$ 为可满足。

2) $\forall M_{R,W-1,W(i_1,i_2)}$ 为不可满足。

2.2.1　构造模型子集的有序搜索链

例 1　原升序搜索链。在之前的 SAT 搜索算法中，设 R 轮差分特征重量的初始值为 $B_{\text{opt}}(R-1)$。然后，判断相应的 SAT 模型是否可满足，若不可满足则将目标重量值加 1，直到目标重量值取到 $B_{\text{opt}}(R)$，即对应 SAT 模型判断结果为可满足时，停止搜索。R 轮重量上界取值 W 的顺序为

$$B_{\text{opt}}(R-1)=B_{\text{ini}}(R)\to(B_{\text{ini}}(R)+1)\to\cdots\to B_{\text{end}}(R)=B_{\text{opt}}(R) \tag{10}$$

此时，模型 $M_{R,W}$ 的搜索顺序与 W 的取值顺序一致，具体表示为

$$M_{R,W}:M_{R,B_{\text{opt}}(R-1)}=M_{R,B_{\text{ini}}(R)}\to M_{R,(B_{\text{ini}}(R)+1)}\to\cdots\to M_{R,B_{\text{end}}(R)}=M_{R,B_{\text{opt}}(R)} \tag{11}$$

该升序搜索链在满足命题 3 的前提下，将未划分的模型集合按照目标重量增加的顺序进行搜索。结合前文对划分子集的研究，给出划分子集中类似的搜索链的描述。

特别地，本文倾向于选择与例 1 相反的搜索顺序。目前，SAT 求解器判定可满足性问题的速度往往比判定不可满足性问题的速度快。这一点满足实验测试经验，且 Liu 等[21]在搜索线性逼近时曾表述过相同的观点。因此本文希望需要判断不可满足的子集个数尽可能少，于是提出满足命题 3 的降序分支搜索链的概念。

定义 2　固定总轮数 R 和选取的部分连续轮 $E(i_1,i_2)$，W 的取值顺序为

$$B_{\text{ini}}(R)\to(B_{\text{ini}}(R)-1)\to\cdots\to B_{\text{end}}(R) \tag{12}$$

则对任意一个可能的 $W(i_1,i_2)$ 代表的全部子集可以按照式(13)搜索。

$$M_{R,W,W(i_1,i_2)}:M_{R,B_{\text{ini}}(R),W(i_1,i_2)}\to M_{R,B_{\text{ini}}(R)-1,W(i_1,i_2)}\to\cdots\to M_{R,B_{\text{end}}(R),W(i_1,i_2)} \tag{13}$$

其中，箭头方向表示子集搜索的顺序的同时，也符合由较大子集指向其包含的较小子集的事实，故将式(13)称为降序分支搜索链。

此外，在不添加额外的加速条件时，定义 2 中起点和终点应该覆盖增加一轮重量的所有可能情况，即应分别选取 $B_{\text{ini}}(R)=B_{\text{opt}}(R-1)+m-1$ 和 $B_{\text{end}}(R)=B_{\text{opt}}(R)-1$。

2.2.2　降序搜索策略的实现方法

为了高效实现上述降序分支搜索链，对搜索的起点、终点和中间节点的搜索技术进一步研究，以达到尽可能地约简搜索空间的目的。3 种搜索节点确定条件如下。

1) 起点：预计算上界。降序搜索策略需要选定差分概率重量的上界作为搜索链的起点，上界越靠近最优差分概率重量，则需要判断的搜索空间越小。

对于一个待求解的 R 轮搜索链，若已知前 R−1 轮的一些高概率差分特征，分别将它们的输出差分作为第 R 轮的输入差分，向后再计算一轮，则得到一些可能的 R 轮差分特征。取这些 R 轮差分特征中差分转移概率重量最小值作为初始的重量上界。此过程在求解 R−1 轮的差分特征时，是对 R 轮差分转移概率重量上界的预计算。

2) 终点：终止条件。由命题 1 可以得到分支搜索链终止搜索的判断条件。即在一条搜索链中，若出现某个子集判断为不可满足，则停止对后续子集的搜索；若子集的状态为可满足，则继续沿链的方向搜索下一子集。

3) 中间节点：跳过条件。由命题 2 可以给出链中间可跳过节点的判断条件。即对于不同链中满足相同的 R 轮概率重量的子集，如果已证明存在可满足的子集，可推出当前 R 轮概率重量对应的差分特征存在，跳过对剩余子集的搜索；否则继续搜索满足当前 R 轮概率重量的子集。

结合上述 3 种搜索节点的确定条件，给出改进的降序分支搜索链模型。为了便于算法描述，引入一个维数随搜索深度增加而增加的向量 **Flag**，其每个分量取值为 0 或 1。其中，第 k 个分量 **Flag**[k]取值为 0 表示搜索深度为 k 的所有子集可满足性为未知或不可满足；取值为 1 则表示存在搜索深度为 k 的子集被判断为可满足。搜索的详细步骤如算法 2 所示。

算法 2　降序分支搜索链模型

输入　密码算法 E，目标轮数 r，初始重量 $B_{\text{init}}(r)$，部分连续轮重量 $W(i_1,i_2)$ 的取值集合，公共内存向量 **Flag**

输出　r+1 轮重量上界 $B_{W(i_1,i_2),\text{extra}}(r)$

1) $k=0$
2) $B(r)\leftarrow B_{\text{init}}(r)-k$
3) $B_{W(i_1,i_2),\text{extra}}(r)\leftarrow B(r)+m$
4) if **Flag**[0] = 0，建立子集 $M_{r,B(r),W(i_1,i_2)}$ 的模型，调用求解器判断可满足性
5) else $M_{r,B(r),W(i_1,i_2)}$ 的可满足性未知
6) end if
7) while $M_{r,B(r),W(i_1,i_2)}$ 为可满足或未知 do
8) if 至少存在一个子集 $M_{r,B(r),W(i_1,i_2)}$ 判断为可满足，**Flag** 的末位增加一个 0 分量
9) end if
10) $k+=1$
11) $B(r)\leftarrow B_{\text{init}}(r)-k$
12) if **Flag**[k]==1，则跳过 $M_{r,B(r),W(i_1,i_2)}$ 模型建立与搜索，可满足性记为未知
13) elif **Flag**[k]==0，调用算法 1 求解最简边界条件中精确上下界，建立 $M_{r,B(r),W(i_1,i_2)}$ 的 SAT 模型并调用求解器求解
14) if $M_{r,B(r),W(i_1,i_2)}$ 是可满足的，则更新 **Flag**[k]=1。提取解文件中对应的输出差分值，向后推一轮，得到对应的 r+1 轮可满足重量。更新 $B_{W(i_1,i_2),\text{extra}}(r)$ 为当前 r+1 轮可满足重量最小值
15) else $M_{r,B(r),W(i_1,i_2)}$ 不可满足，终止循环
16) end if
17) end if
18) end while

2.3 加速搜索算法

本节按照由前 r−1 轮最优差分特征的概率重量归纳得到 r 轮最优差分特征的概率重量的总体思路，给出改进的最优差分特征自动化搜索算法。主要过程为以下两步。

1) 选取部分连续轮 $E(i_1,i_2)$ 利用前 $r-1$ 轮最优差分特征的概率重量，计算初始子集划分条件的取值范围，作为分支搜索链模型的选取依据。

2) 使用并行池技术，计算 r 轮的全部分支搜索链模型，得到 r 轮的最优差分特征的概率重量，以及 $r+1$ 轮最优差分特征概率重量的预计算上界。

详细过程如算法 3 所示。

算法 3 加速的 SAT 差分自动化搜索算法

输入 密码算法 E，搜索总轮数 R

输出 最优差分特征概率重量 $B_{\text{opt}}(1),B_{\text{opt}}(2),\cdots,B_{\text{opt}}(R)$

1) 初始化 $B_{\text{init}}(r)\leftarrow 0$
2) if $r=1$，推导或建模求解 $B_{\text{opt}}(1),B_{\text{opt}}(2),\cdots,B_{\text{opt}}(R)$ 并预计算两轮最优差分特征概率重量上界 $B_{\text{extra}}(1)$
3) end if
4) for r=2～R do
5) 选取部分连续轮 $E(i_1,i_2)$，计算 $W(i_1,i_2)$ 的取值范围
6) 设立公共内存向量 **Flag**，初始化其元素为全 0
7) 利用进程池并行实现算法 2，输入 E、$B_{\text{init}}(r)$、$W(i_1,i_2)$ 取值集合以及 **Flag**，计算 r 轮的全部降序分支搜索链模型
8) 返回 **Flag** 和第 $r+1$ 轮最优差分特征概率重量上界的集合
9) $B_{\text{opt}}(r)\leftarrow B_{\text{init}}(r)-\text{len}(\textbf{Flag})+1$
10) $B_{\text{init}}(r+1)\leftarrow \min(B_{W(i_1,i_2),\text{extra}}(r))$
11) end for

3 应用于SPECK-48、SPECK-96、SPECK-128

3.1 SPECK 族密码算法简介

SPECK 族密码算法是美国国家安全局提出的基于 ARX 原语的分组密码算法，根据不同的分组规模与密钥规模，它包含 10 种结构相同的变体。SPECK 族密码算法轮函数由 3 种运算构成：循环移位运算，循环右移 α bit 和循环左移 β bit 分别用 $S^{-\alpha}$ 和 S^{β} 表示；异或运算，用 $\oplus$ 表示；模加运算，用 $\boxplus$ 表示。其中，模加运算作为算法的非线性组件。SPECK 族密码算法的相关参数如表 1 所示，轮函数结构如图 1 所示。

表 1 SPECK 族密码算法的相关参数

分组规模/组	α	β	轮数/轮	密钥规模/bit
32	7	2	22	64
48	8	3	22	72
			23	96
64	8	3	26	96
			27	128
96	8	3	28	96
			29	144
128	8	3	32	128
			33	192
			34	256

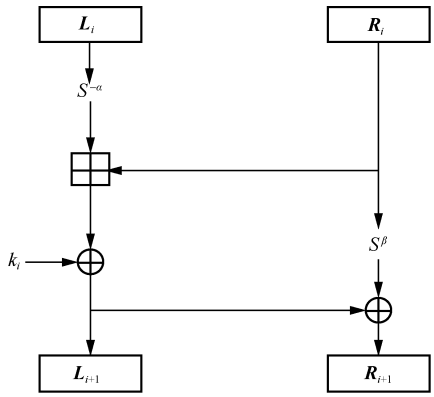

图 1　SPECK 族密码算法轮函数结构

3.2　SPECK-48、SPECK-96、SPECK-128 的差分安全边界

将加速算法应用于 SPECK 族密码算法，使用在实验中效果较好的部分连续轮选取方式 $E(i_1,i_2)=E\left(1,\left\lfloor\frac{R}{2}\right\rfloor\right)$，计算部分连续轮的概率重量 $W\left(1,\left\lfloor\frac{R}{2}\right\rfloor\right)$ 的取值范围，并将有序分支链按照取值范围的两端向中间交替排列的顺序作为并行池中进程函数输入 $W(i_1,i_2)$ 取值列表。SPECK-2n 的 SAT 子集模型 $M_{R,W,W(i_1,i_2)}$ 的描述如下。

设第 i 轮输入差分变量为 $(\boldsymbol{L}_i,\boldsymbol{R}_i)=(x_1^i,x_2^i,\cdots,x_n^i,x_{n+1}^i,x_{n+2}^i,\cdots,x_{2n}^i)$，输出差分变量为 $(\boldsymbol{L}_{i+1},\boldsymbol{R}_{i+1})=(x_1^{i+1},x_2^{i+1},\cdots,x_n^{i+1},x_{n+1}^{i+1},x_{n+2}^{i+1},\cdots,x_{2n}^{i+1})$，模加中概率重量变量为 $\boldsymbol{W}_i=(w_1^i,w_2^i,\cdots,w_{n-1}^i)$。

$\boldsymbol{L}_i$ 循环右移 α bit 运算可以描述为变量置换，即 $(x_1^i,x_2^i,\cdots,x_n^i)\to(x_{n+1-\alpha}^i,\cdots,x_n^i,x_1^i,\cdots,x_{n-\alpha}^i)$。

$\boldsymbol{R}_i$ 循环左移 β bit 运算可以描述为变量置换，即 $(x_{n+1}^i,x_{n+2}^i,\cdots,x_{2n}^i)\to(x_{n+1+\beta}^i,\cdots,x_{2n}^i,x_{n+1}^i,\cdots,x_{n+\beta}^i)$。

模加运算的输入变量为 $(x_{n+1-\alpha}^i,\cdots,x_n^i,x_1^i,\cdots,x_{n-\alpha}^i)$ 和 $(x_{n+1}^i,x_{n+2}^i,\cdots,x_{2n}^i)$，输出变量为 $(x_1^{i+1},x_2^{i+1},\cdots,x_n^{i+1})$，概率重量变量为 $(w_1^i,w_2^i,\cdots,w_{n-1}^i)$，可以利用模型 2 进行刻画。

异或运算的输入变量为 $(x_1^{i+1},x_2^{i+1},\cdots,x_n^{i+1})$ 和 $(x_{n+1+\beta}^i,\cdots,x_{2n}^i,x_{n+1}^i,\cdots,x_{n+\beta}^i)$，输出变量为 $(x_{n+1}^{i+1},x_{n+2}^{i+1},\cdots,x_{2n}^{i+1})$，可以利用模型 1 进行刻画。

初始输入非平凡约束为 $x_1^i\vee x_2^i\vee\cdots\vee x_n^i\vee x_{n+1}^i\vee x_{n+2}^i\vee\cdots\vee x_{2n}^i=1$。

目标约束为 $\sum_{i=1}^{R}\sum_{j=1}^{n-1}w_j^i\leqslant W$。

子集约束为 $\sum_{i=i_1}^{i_2}\sum_{j=1}^{n-1}w_j^i=W(i_1,i_2)$。

最简 Matsui 边界条件求解过程如下：向 MILP 模型中添加所有 Matsui 边界条件和子集约束条件，由算法 1 求得约简的不等式 $l_{e_1,e_2}\leqslant\sum_{i=e_1}^{e_2}w_i\leqslant m_{e_1,e_2}$。再用模型 3 进行刻画。

本文实验在处理器配置为 64 位 AMD 锐龙 R7 4800H @2.90 GHz 的 8 核 64 位笔记本电脑上进行，用 Python3.7 编程并调用求解器 CaDiCaL2020 和 Gurobi9.1.1 进行实现。在有限的时间内，SPECK-48、SPECK-96、SPECK-128 的全局最优差分特征的概率重量分别搜索到 20 轮、15 轮、11 轮，较现有最好结果分别增加了 1 轮、4 轮、2 轮，如表 2 所示。

表 2　SPECK-48、SPECK-96、SPECK-128 的全局最优差分特征概率

算法	轮数/轮	最优差分概率	时间	文献
SPECK-48	19	2^{-89}	1 736 050.9 s	文献[20]
			827 299.5 s	本文
	20	2^{-96}	>10 d	本文
SPECK-96	10	2^{-49}	1 323 894.2 s	文献[20]
			160 641.0 s	本文
	11	2^{-58}	998 381.2 s	本文
	12	2^{-62}	154 312.1 s	本文
	13	2^{-66}	63 628.4 s	本文
	14	2^{-72}	170 265.6 s	本文
SPECK-128	9	2^{-39}	247 510.6 s	文献[20]
			7 043.6 s	本文
	10	2^{-49}	145 132.6 s	本文
	11	2^{-58}	1 054 561.7 s	本文

4　结束语

本文在最简 Matsui 边界条件模型的启发下，结合分治思想提出了一种新的基于 SAT 的最优差分特征搜索算法。对于任意一个分组密码算法，本文将待搜索的可能存在差分特征的模型按照一部分连续轮的重量进行划分子集，并且构造了一种有序分支链模型对划分子集进行分类。为了进一步加快搜索，本文提出了不同的方法给出链的搜索方向以及起点、终点、中间节点的选择策略，尽可能地约减搜索空间，提高搜索效率。本文将加速搜索算法应用到 ARX 类密码算法 SPECK 上，得到了

SPECK-48、SPECK-96、SPECK-128 的更长轮的全局最优的差分特征，并且减少了求解时间。进一步地，由于改进的算法不涉及对密码组件刻画的修改，也可以用于搜索 ARX 类其他结构的密码或者带 S 盒的分组密码的最优差分特征。

参考文献：

[1] National Bureau of Standards. Data encryption standard[S]. 1977.

[2] DAEMEN J, RIJMEN V. AES proposal: Rijndael[R]. 1999.

[3] WHEELER D J, NEEDHAM R M. TEA, a tiny encryption algorithm[M]. Berlin: Springer, 1995.

[4] BERNSTEIN D J. The Salsa20 family of stream ciphers[M]. Berlin: Springer, 2008.

[5] AUMASSON J P, BERNSTEIN D J. SipHash: a fast short-input PRF[C]//Proceedings of International Conference on Cryptology in India. Berlin: Springer, 2012: 489-508.

[6] BEAULIEU R, TREATMAN-CLARK S, SHORS D, et al. The SIMON and SPECK lightweight block ciphers[C]//Proceedings of 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC). Piscataway: IEEE Press, 2015: 1-6.

[7] HONG D, LEE J K, KIM D C, et al. LEA: a 128-bit block cipher for fast encryption on common processors[C]//Proceedings of International Workshop on Information Security Applications. Berlin: Springer, 2014: 3-27.

[8] MOUHA N, MENNINK B, HERREWEGE A V, et al. Chaskey: an efficient MAC algorithm for 32-bit microcontrollers[C]//Proceedings of International Conference on Selected Areas in Cryptography. Berlin: Springer, 2014: 306-323.

[9] BIHAM E, SHAMIR A. Differential cryptanalysis of DES-like cryptosystems[C]//Advances in Cryptology-CRYPTO' 90. Berlin: Springer, 1991: 2-21.

[10] MATSUI M. Linear cryptanalysis method for DES cipher[M]. Berlin: Springer, 1994.

[11] MOUHA N, WANG Q J, GU D W, et al. Differential and linear cryptanalysis using mixed-integer linear programming[C]//Proceedings of International Conference on Information Security and Cryptology. Berlin: Springer, 2012: 57-76.

[12] MASSACCI F, MARRARO L. Logical cryptanalysis as a SAT problem[J]. Journal of Automated Reasoning, 2000, 24(1): 165-203.

[13] BIRYUKOV A, VELICHKOV V. Automatic search for differential trails in ARX ciphers[C]//Proceedings of Cryptographers' Track at the RSA Conference. Berlin: Springer, 2014: 227-250.

[14] SONG L, HUANG Z J, YANG Q Q. Automatic differential analysis of ARX block ciphers with application to SPECK and LEA[C]//Proceedings of Australasian Conference on Information Security and Privacy. Berlin: Springer, 2016: 379-394.

[15] BIRYUKOV A, VELICHKOV V, CORRE Y L. Automatic search for the best trails in ARX: application to block cipher speck[C]// Proceedings of International Conference on Fast Software Encryption. Berlin: Springer, 2016: 289-310.

[16] FU K, WANG M Q, GUO Y H, et al. MILP-based automatic search algorithms for differential and linear trails for SPECK[C]//Proceedings of International Conference on Fast Software Encryption. Berlin: Springer, 2016: 268-288.

[17] LIPMAA H, MORIAI S. Efficient algorithms for computing differential properties of addition[C]//Proceedings of Fast Software Encryption. Berlin: Springer, 2002: 336-350.

[18] ZHANG Y J, SUN S W, CAI J H, et al. Speeding up MILP aided differential characteristic search with Matsui's strategy[C]//Proceedings of International Conference on Information Security. Berlin: Springer, 2018: 101-115.

[19] SUN L, WANG W, WANG M. Accelerating the search of differential and linear characteristics with the SAT method[J]. IACR Transactions on Symmetric Cryptology, 2021(1): 269-315.

[20] WANG S P, FENG D, HU B, et al. The simplest SAT model of combining Matsui's bounding conditions with sequential encoding method[J]. IACR Cryptol ePrint Arch, 2022, 2022: 626.

[21] LIU Y W, WANG Q J, RIJMEN V. Automatic search of linear trails in ARX with applications to SPECK and Chaskey[C]//Proceedings of International Conference on Applied Cryptography and Network Security. Berlin: Springer, 2016: 485-499.

[作者简介]

胡斌（1971– ），男，河南信阳人，博士，信息工程大学教授、博士生导师，主要研究方向为密码学与信息安全。

谈潇（1998– ），女，湖北鄂州人，信息工程大学硕士生，主要研究方向为分组密码自动化分析。

王森鹏（1990– ），男，河南商丘人，博士，信息工程大学讲师，主要研究方向为对称密码算法的设计与分析。