ARX 结构分组密码积分区分器的自动化搜索

韩亚[1,2]，王明生[1,2]

（1. 中国科学院信息工程研究所信息安全国家重点实验室，北京 100093；
2. 中国科学院大学网络空间安全学院，北京 100049）

摘 要：首先，基于三子集传播的积分可分性质，分别构造 ARX 结构分组密码积分的 *K* 集和 *L* 集传播方程，其中，经过分组密码轮函数异或操作时，*L* 集所有向量影响 *K* 集向量传播；然后，利用 SAT/SMT 求解器，建立 ARX 结构分组密码积分传播方程；最后，遍历满足一定数据复杂度的积分输入，自动化搜索缩减轮数的 ARX 结构分组密码积分区分器。利用该方法能高效地自动化搜索 ARX 结构，包括类 SIMON 簇、HIGHT、SPECK 簇和 LEA 等分组密码算法的积分区分器。

关键词：ARX；三子集；积分区分器；SAT/SMT

中图分类号：TP391

文献标识码：A

doi: 10.11959/j.issn.1000-436x.2018081

Automatic method for searching integral distinguishers of ARX block ciphers

HAN Ya[1,2], WANG Mingsheng[1,2]

1. State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Firstly, based on three subsets division property propagation technique, the propagation function of the *K*-set and *L*-set of ARX block ciphers was constructed respectively. All vectors in *L*-set affected the propagation of *K*-set when propagate through xored round key operation. With SAT/SMT solver, round reduced integral propagation functions of ARX block ciphers could be established. Finally, by exhausting all possible input integral characteristics with proper data complexity, round reduced integral distinguishers of ARX block ciphers could be found. The proposed method can be used for searching integral distinguishers of ARX block ciphers including SIMON-like family block ciphers, HIGHT, SPECK family block ciphers and LEA effectively.

Key words: ARX, three subsets, integral distinguishers, SAT/SMT

1 引言

2015 年，Todo[1]在欧密会首次提出积分可分性质，并利用该性质搜索分组密码积分区分器。同年，Todo[2]在美密会提出了分组密码算法 MISTY1 的 6 轮积分区分器，并对全轮 MISTY1 实施了理论的积分分析。根据可分性质积分传播规则，Todo 给出 10 轮 SIMON32 的积分区分器，比实际搜索结果少了 5 轮。Wang 等[3]提出了 SIMON 在原积分区分器的基础上增加一轮而不增加数据复杂度的方法。随后，Todo 等[4]根据类 SIMON 簇分组密码的结构特征，提出基于比特的可分性质，并搜索得到 14 轮 SIMON32 积分区分器。同时，他们通过改进基于比特的可分性质，提出利用三子集描述积分传播过程。利用三子集积分传播特征，他们搜索得到 15 轮 SIMON32 积分区分器。针对其他 SIMON 簇分组密码算法[5]

收稿日期：2017-12-26；修回日期：2018-04-20
基金项目：国家自然科学基金资助项目（No.61772516）
Foundation Item: The National Natural Science Foundation of China (No.61772516)

SIMON48/64/96/128，Todo 等仅给出理论的积分区分器上界。

2016 年，Xiang 等[6]在亚密会上通过学习基于比特的可分性质，提出了一种基于混合线性规划（MILP, mixed integer linear programming）方法的积分区分器自动化搜索算法。该算法可用于搜索类 SIMON 簇及基于 S 盒的分组密码算法的积分区分器。针对 SIMON 簇分组密码算法 SIMON32/48/64/96/128，他们分别给出了 14、16、18、22 和 26 轮积分区分器。但是他们无法给出与实际相符的 15 轮 SIMON32 积分区分器。随后，Sun 等[7]通过改进 Xiang 等[6]的方法，提出搜索 ARX 结构分组密码积分区分器算法。基于比特可分性质，Sun 等[7]通过建立积分传播模型并利用混合线性规划求解得到 18 轮 HIGHT[8]积分区分器以及 7 轮 LEA[9]积分区分器。同时，Sun 等[7]还第一次给出了 TEA[10]、XTEA、KATAN 和 KTANTAN 等 ARX 结构分组密码算法的积分区分器。

针对一些小状态 ARX 结构分组密码，混合线性规划方法能有效地搜索积分区分器。但是对于大状态 ARX 结构分组密码算法如 SPECK128，混合线性规划方法并不能给出有效的解决方案。通过学习基于比特的可分性质，利用三子集传播方程，本文提出一种基于 SAT/SMT 求解器自动化搜索 ARX 结构分组密码积分区分器的方法。SAT/SMT 求解器可通过刻画比特向量可分性质，等价描述比特可分性质传播过程，并简化积分模型建立，加速模型求解过程。

2 比特可分性质传播模型

2.1 可分性质

积分可分性质由 Todo[1]首次提出并用于搜索分组密码积分区分器。集合空间的可分性质可通过点积运算描述。取输入变量 $x\in F_2^n$，点积运算 $\pi_{\boldsymbol{u}},\boldsymbol{u}\in F_2^n$ 满足

$$\pi_{\boldsymbol{u}}(x)\coloneqq\prod_{i=0}^{n-1}x[i]^{\boldsymbol{u}[i]} \tag{1}$$

任意 $\boldsymbol{u}=(u_0,u_1,\cdots,u_{m-1})\in(F_2^{n_0}\times F_2^{n_1}\times\cdots F_2^{n_{m-1}})$，输入变量 $x=(x_0,x_1,\cdots,x_{m-1})$，点积运算满足

$$\pi_{\boldsymbol{u}}(x)\coloneqq\prod_{i=0}^{m-1}\pi_{\boldsymbol{u}_i}(x_i) \tag{2}$$

对于集合 X，任意 $x\in X$ 属于有限域空间 $(F_2^{n_0}\times F_2^{n_1}\times\cdots F_2^{n_{m-1}})$，如果集合 X 满足可分性质 $D_K^{n_0,n_1,\cdots,n_{m-1}}$，其中，$\boldsymbol{k}\in K$ 为 m 维向量且第 i 个元素满足 $0\leqslant\boldsymbol{k}_i\leqslant n_{i-1}$，对所有 $x\in X$ 满足

$$\oplus\pi_{\boldsymbol{u}}(x)=\begin{cases}U,\boldsymbol{k}\in K\text{且}Wt(\boldsymbol{u})\geqslant\boldsymbol{k}\\0,\text{其他}\end{cases} \tag{3}$$

其中，U 表示不确定取值，$Wt(\boldsymbol{a})$ 表示向量 $\boldsymbol{a}$ 的向量汉明重量，且 $Wt(\boldsymbol{a})=\left(wt(\boldsymbol{a}_0),wt(\boldsymbol{a}_1),\cdots,wt(\boldsymbol{a}_{m-1})\right)$。向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的所有分量 $\boldsymbol{a}_i$、$\boldsymbol{b}_i$ 均满足 $\boldsymbol{b}_i\leqslant\boldsymbol{a}_i$，则 $\boldsymbol{b}\leqslant\boldsymbol{a}$。

2.2 比特三子集积分可分性质

Todo 等[4]将基于字的积分可分性质转换为比特状态，给出比特状态下的可分性质传播规则，并利用三子集更精确地描述积分可分性质的传播过程。

对于任意集合 X，其元素属于 $(F_2)^m$，如果集合 X 满足三子集积分可分性质 $D_{K,L}^{1^m}$，其中，$\boldsymbol{k}\in K$ 为 m 维比特向量且第 i 个元素取值为 0 或 1，所有 $x\in X$ 满足

$$\oplus\pi_{\boldsymbol{u}}(x)=\begin{cases}U,\boldsymbol{k}\in K\text{且}Wt(\boldsymbol{u})\geqslant\boldsymbol{k}\\1,\boldsymbol{l}\in L\text{且}Wt(\boldsymbol{u})=\boldsymbol{l}\\0,\text{其他}\end{cases} \tag{4}$$

三子集积分可分性质比传统积分可分性质在传播过程中多了 L 集传播，能更加精确地描述积分可分性质的传播过程。

2.3 比特三子集传播规则

基于比特的积分可分性质，K 集和 L 集在经过 ARX 结构分组密码轮函数中“分支”“异或”“与”操作时相互独立。

定义 1 F 为分支操作，其输入 $(x_0,x_1,\cdots,x_{m-1})\in(F_2)^m$。任意比特 $x_i,0\leqslant i\leqslant m-1$ 经过分支操作 F 的 K 集传播 $\boldsymbol{k}_{x_i}\to(\boldsymbol{k}_{y_i},\boldsymbol{k}_{z_i})$ 满足

$$(\boldsymbol{k}_{y_i},\boldsymbol{k}_{z_i})=\begin{cases}(0,0),\boldsymbol{k}_{x_i}=0\\(1,0),(0,1),\boldsymbol{k}_{x_i}=1\end{cases} \tag{5}$$

L 集传播 $\boldsymbol{l}_{x_i}\to(\boldsymbol{l}_{y_i},\boldsymbol{l}_{z_i})$ 满足

$$(\boldsymbol{l}_{y_i},\boldsymbol{l}_{z_i})=\begin{cases}(0,0),\boldsymbol{l}_{x_i}=0\\(0,1),(1,0),(1,1),\boldsymbol{l}_{x_i}=1\end{cases} \tag{6}$$

定义 2 F 为异或操作，其输入 $(x_0,x_1,\cdots,x_{m-1})\in(F_2)^m$。任意比特 $x_i,x_j,0\leqslant i\neq j\leqslant m-1$ 经过异或操作 F 的 K 集传播 $(\boldsymbol{k}_{x_i},\boldsymbol{k}_{x_j})\to\boldsymbol{k}_{y_i}$ 满足

$$\boldsymbol{k}_{y_i}=\begin{cases}0,(\boldsymbol{k}_{x_i},\boldsymbol{k}_{x_j})=(0,0)\\1,(\boldsymbol{k}_{x_i},\boldsymbol{k}_{x_j})=(1,0),(0,1)\end{cases}\tag{7}$$

L 集传播 $(\boldsymbol{l}_{x_i},\boldsymbol{l}_{x_j})\to\boldsymbol{l}_{y_i}$ 满足

$$\boldsymbol{l}_{y_i}=\begin{cases}0,(\boldsymbol{l}_{x_i},\boldsymbol{l}_{x_j})=(0,0)\\1,(\boldsymbol{l}_{x_i},\boldsymbol{l}_{x_j})=(1,0),(0,1)\end{cases}\tag{8}$$

定义 3　F 为与操作，其输入 $(x_0,x_1,\cdots,x_{m-1})\in(F_2)^m$。任意比特 $x_i,x_j,0\leqslant i\neq j\leqslant m-1$ 经过与操作 F 的 K 集传播 $(\boldsymbol{k}_{x_i},\boldsymbol{k}_{x_j})\to\boldsymbol{k}_{y_i}$ 满足

$$\boldsymbol{k}_{y_i}=\begin{cases}0,(\boldsymbol{k}_{x_i},\boldsymbol{k}_{x_j})=(0,0)\\1,(\boldsymbol{k}_{x_i},\boldsymbol{k}_{x_j})=(1,0),(0,1),(1,1)\end{cases}\tag{9}$$

L 集传播 $(\boldsymbol{l}_{x_i},\boldsymbol{l}_{x_j})\to\boldsymbol{l}_{y_i}$ 满足

$$\boldsymbol{l}_{y_i}=\begin{cases}0,(\boldsymbol{l}_{x_i},\boldsymbol{l}_{x_j})=(0,0)\\1,(\boldsymbol{l}_{x_i},\boldsymbol{l}_{x_j})=(1,1)\end{cases}\tag{10}$$

定义 4　F 为异或轮密钥操作，其输入 $(x_0,x_1,\cdots,x_{m-1})\in(F_2)^m$。任意 L 集向量 $(\boldsymbol{l}_{x_0},\boldsymbol{l}_{x_1},\cdots,\boldsymbol{l}_{x_{m-1}})$ 经过异或轮密钥操作 F 后，如果满足 $\boldsymbol{l}_{x_i}=0$，需向 K 集中添加向量 $(\boldsymbol{l}_{x_0},\boldsymbol{l}_{x_1},\cdots,\boldsymbol{l}_{x_i}\vee 1,\cdots,\boldsymbol{l}_{x_{m-1}})$。

例如，取 $m=4$，异或轮密钥操作输入 L 集向量集合为 $((0,0,0,1),(1,0,0,0),(1,1,0,0))$。其中，向量 $(0,0,0,1)$ 经过异或轮密钥操作后 K 集增加 3 个额外向量 $((1,0,0,1),(0,1,0,1),(0,0,1,1))$；$(1,0,0,0)$ 经过异或轮密钥操作后 K 集增加 3 个额外向量 $((1,1,0,0),(1,0,1,0),(1,0,0,1))$；$(1,1,0,0)$ 经过异或轮密钥操作后 K 集增加 2 个额外向量 $((1,1,1,0),(1,1,0,1))$，其中，$(1,1,1,0)\geqslant(1,1,0,0)$，$(1,1,0,1)\geqslant(1,1,0,0)$ 被筛除。K 集共增加 5 个向量 $((1,0,0,1),(0,1,0,1),(0,0,1,1),(1,1,0,0),(1,0,1,0))$。

定义 5　F_r 为分组密码轮函数，假设初始三子集积分可分性质为 $D_{k,l}^{1^m}$。经过第 i 轮函数的输出，三子集积分可分性质满足 $D_{K_i,L_i}^{1^m}$，则 r 轮三子集积分可分性质传播路径表示为

$$(\boldsymbol{k},\boldsymbol{l})=(K_0,L_0)\to(K_1,L_1)\to\cdots\to(K_r,L_r)$$

2.4　向量三子集传播规则

基于向量的积分可分性质，K 集和 L 集在经过 ARX 结构分组密码轮函数“分支”“异或”“与”操作时相互独立。

定义 6　二分支操作 $x\to(y,z)$，输入 $x=(x_0,x_1,\cdots,x_{m-1})\in(F_2)^m$。经过二分支操作的 K 集传播 $kp_{2\text{copy}}(\boldsymbol{k}_x\to(\boldsymbol{k}_y,\boldsymbol{k}_z))$ 满足

$$((\sim\boldsymbol{k}_x)\wedge(\sim\boldsymbol{k}_y)\wedge(\sim\boldsymbol{k}_z))\oplus(\boldsymbol{k}_x\wedge(\boldsymbol{k}_y\oplus\boldsymbol{k}_z))=1\tag{11}$$

L 集传播 $lp_{2\text{copy}}(\boldsymbol{l}_x\to(\boldsymbol{l}_y,\boldsymbol{l}_z))$ 满足

$$\boldsymbol{l}_x\oplus((\sim\boldsymbol{l}_y)\wedge(\sim\boldsymbol{l}_z))=1\tag{12}$$

定义 7　三分支操作 $x\to(y,z,t)$，输入 $x=(x_0,x_1,\cdots,x_{m-1})\in(F_2)^m$。经过三分支操作的 K 集传播 $kp_{3\text{copy}}(\boldsymbol{k}_x\to(\boldsymbol{k}_y,\boldsymbol{k}_z,\boldsymbol{k}_t))$ 满足

$$((\sim\boldsymbol{k}_z)\wedge(\sim\boldsymbol{k}_t)\wedge(\boldsymbol{k}_y\oplus(\sim\boldsymbol{k}_x)))\oplus((\sim\boldsymbol{k}_y)\wedge\boldsymbol{k}_x\wedge(\boldsymbol{k}_z\oplus\boldsymbol{k}_t))=1\tag{13}$$

L 集传播 $lp_{3\text{copy}}(\boldsymbol{l}_x\to(\boldsymbol{l}_y,\boldsymbol{l}_z,\boldsymbol{l}_t))$ 满足

$$\boldsymbol{l}_x\oplus((\sim\boldsymbol{l}_y)\wedge(\sim\boldsymbol{l}_z)\wedge(\sim\boldsymbol{l}_t))=1\tag{14}$$

定义 8　二异或操作 $x\oplus y=z$，输入 $x=(x_0,x_1,\cdots,x_{m-1}),y=(y_0,y_1,\cdots,y_{m-1})\in(F_2)^m$。经过二异或操作的 K 集传播 $kp_{2\text{xor}}((\boldsymbol{k}_x,\boldsymbol{k}_y)\to\boldsymbol{k}_z)$ 满足

$$((\sim\boldsymbol{k}_x)\wedge(\sim\boldsymbol{k}_y)\wedge(\sim\boldsymbol{k}_z))\oplus(\boldsymbol{k}_z\wedge(\boldsymbol{k}_y\oplus\boldsymbol{k}_x))=1\tag{15}$$

L 集传播 $lp_{2\text{xor}}((\boldsymbol{l}_x,\boldsymbol{l}_y)\to\boldsymbol{l}_z)$ 满足

$$((\sim\boldsymbol{l}_x)\wedge(\sim\boldsymbol{l}_y)\wedge(\sim\boldsymbol{l}_z))\oplus(\boldsymbol{l}_z\wedge(\boldsymbol{l}_y\oplus\boldsymbol{l}_x))=1\tag{16}$$

定义 9　三异或操作 $x\oplus y\oplus z=t$，输入 $x,y,z\in(F_2)^m$。经过三异或操作的 K 集传播 $kp_{3\text{xor}}((\boldsymbol{k}_x,\boldsymbol{k}_y,\boldsymbol{k}_z)\to\boldsymbol{k}_t)$ 满足

$$((\sim\boldsymbol{k}_z)\wedge(\sim\boldsymbol{k}_x)\wedge(\boldsymbol{k}_y\oplus(\sim\boldsymbol{k}_t)))\oplus((\sim\boldsymbol{k}_y)\wedge\boldsymbol{k}_t\wedge(\boldsymbol{k}_z\oplus\boldsymbol{k}_x))=1\tag{17}$$

L 集传播 $lp_{3\text{xor}}((\boldsymbol{l}_x,\boldsymbol{l}_y,\boldsymbol{l}_z)\to\boldsymbol{l}_t)$ 满足

$$((\sim\boldsymbol{l}_z)\wedge(\sim\boldsymbol{l}_x)\wedge(\boldsymbol{l}_y\oplus(\sim\boldsymbol{l}_t)))\oplus((\sim\boldsymbol{l}_y)\wedge\boldsymbol{l}_t\wedge(\boldsymbol{l}_z\oplus\boldsymbol{l}_x))=1\tag{18}$$

定义 10　与操作 $x\wedge y=z$，输入 $x=(x_0,x_1,\cdots,x_{m-1}),\ y=(y_0,y_1,\cdots,y_{m-1})\in(F_2)^m$。经过与操作的 K 集传播 $kp_{\text{and}}((\boldsymbol{k}_x,\boldsymbol{k}_y)\to\boldsymbol{k}_z)$ 满足

$$\boldsymbol{k}_z\oplus((\sim\boldsymbol{k}_x)\wedge(\sim\boldsymbol{k}_y))=1\tag{19}$$

L 集传播 $lp_{\text{and}}((\boldsymbol{l}_x,\boldsymbol{l}_y)\to\boldsymbol{l}_z)$ 满足

$$((\sim\boldsymbol{l}_x)\wedge(\sim\boldsymbol{l}_y)\wedge(\sim\boldsymbol{l}_z))\oplus(\boldsymbol{l}_x\wedge\boldsymbol{l}_y\wedge\boldsymbol{l}_z)=1 \quad (20)$$

3 ARX 结构可分性质模型

模加运算是 ARX 结构分组密码算法中唯一的非线性部件。常规模加运算输入 $x,y\in(F_2)^m$，模加运算输出 $z\in(F_2)^m$，满足 $z=(x+y)\bmod 2^m$。在某些 ARX 结构分组密码算法（HIGHT、TEA、XTEA 等）中，存在一种常数模加运算，其输入变量 y 为常数（轮密钥），满足 $z=(x+rk)\bmod 2^m$。

3.1 常规模加运算模型

常规模加运算输入 $x,y\in(F_2)^m$，输出 $z\in(F_2)^m$ 满足

$$\begin{cases} z_{m-1}=x_{m-1}\oplus y_{m-1} \\ z_{m-2}=x_{m-2}\oplus y_{m-2}\oplus c_{m-2}, c_{m-2}=x_{m-1}\wedge y_{m-1} \\ z_{m-3}=x_{m-3}\oplus y_{m-3}\oplus c_{m-3}, c_{m-3} \\ \quad =x_{m-2}\wedge y_{m-2}\oplus(x_{m-2}\oplus y_{m-2})\wedge c_{m-2} \\ \quad\quad\vdots \\ z_0=x_0\oplus y_0\oplus c_0, c_0=x_1\wedge y_1\oplus(x_1\oplus y_1)\wedge c_1 \end{cases} \quad (21)$$

其中，$c=(c_0,c_1,\cdots,c_{m-1})\in(F_2)^m$ 是模加运算的进位。常规模加运算 K 集传播过程如表 1 所示。

增加 12 个变量 $(kx^*,kx^{**},kx^{***},ky^*,ky^{**},ky^{***},kc^*,kc^{**},kc^{***},ku,kv,kw)$ 辅助表示模加运算 K 集传播，满足

$$\begin{cases} kz=kx^*\oplus ky^*\oplus kc^* \\ kx\to(kx^*,kx^{**},kx^{***}) \\ ky\to(ky^*,ky^{**},ky^{***}) \\ kc^{**}\to(kc^*,kc^{***}) \\ ku=kx^{**}\wedge ky^{**} \\ kv=kx^{***}\oplus ky^{***} \\ kw=kv\wedge kc^{***} \\ kc^{**}=(ku\oplus kw)<<1 \end{cases} \quad (22)$$

其中，8 个比特 $(kx_0^{**},ky_0^{**},kx_0^{***},ky_0^{***},kx_{m-1}^{***},ky_{m-1}^{***},kc_0^{***},kc_{m-1}^{***})$ 为 0 的分支状态。因此，在 K 集传播过程中该 8 bit 限制为 0。常规模加运算 K 集传播系统 $S_{k+}((\boldsymbol{k}_x,\boldsymbol{k}_y)\to\boldsymbol{k}_z$ 满足

$$\begin{cases} kp_{3\text{xor}}((kx^*,ky^*,kc^*)\to kz)=1 \\ kp_{3\text{copy}}(kx\to(kx^*,kx^{**},kx^{***}))=1 \\ kp_{3\text{copy}}(ky\to(ky^*,ky^{**},ky^{***}))=1 \\ kp_{2\text{copy}}(kc^{**}\to(kc^*,kc^{***}))=1 \\ kp_{\text{and}}((kx^{**},ky^{**})\to ku)=1 \\ kp_{2\text{xor}}((kx^{***},ky^{***})\to kv)=1 \\ kp_{\text{and}}((kv,kc^{***})\to kw)=1 \\ kp_{2\text{xor}}((ku<<1,kw<<1)\to kc^{**})=1 \end{cases} \quad (23)$$

根据 K 集传播过程，可得到常规模加运算 L 集传播系统 $S_{l+}((\boldsymbol{l}_x,\boldsymbol{l}_y)\to\boldsymbol{l}_z$ 满足

表 1 常规模加运算 K 集传播

输出状态	K 集传播	进位状态	K 集传播
kz_{m-1}	$\underbrace{kx_{m-1}}_{kx^*_{m-1}}\oplus\underbrace{ky_{m-1}}_{ky^*_{m-1}}\oplus\underbrace{0}_{kc^*_{m-1}}$	$\underbrace{kc_{m-1}}_{kc^{**}_{m-1}}$	$\overbrace{\underbrace{0}_{kx_0^{**}}\wedge\underbrace{0}_{ky_0^{**}}\oplus(\underbrace{0}_{kx_0^{***}}\wedge\underbrace{0}_{ky_0^{***}})\wedge\overbrace{0}^{kc_0^{***}}}^{kw_0}}^{ku_0}$
kz_{m-2}	$\underbrace{kx_{m-2}}_{kx^*_{m-2}}\oplus\underbrace{ky_{m-2}}_{ky^*_{m-2}}\oplus\underbrace{kc_{m-2}}_{kc^*_{m-2}}$	$\underbrace{kc_{m-2}}_{kc^{**}_{m-2}}$	$\overbrace{\underbrace{kx_{m-1}}_{kx^{**}_{m-1}}\wedge\underbrace{ky_{m-1}}_{ky^{**}_{m-1}}\oplus(\underbrace{0}_{kx^{***}_{m-1}}\wedge\underbrace{0}_{ky^{***}_{m-1}})\wedge\overbrace{0}^{kc^{***}_{m-1}}}^{kw_{m-1}}}^{ku_{m-1}}$
kz_{m-3}	$\underbrace{kx_{m-3}}_{kx^*_{m-3}}\oplus\underbrace{ky_{m-3}}_{ky^*_{m-3}}\oplus\underbrace{kc_{m-3}}_{kc^*_{m-3}}$	$\underbrace{kc_{m-3}}_{kc^{**}_{m-3}}$	$\overbrace{\underbrace{kx_{m-2}}_{kx^{**}_{m-2}}\wedge\underbrace{ky_{m-2}}_{ky^{**}_{m-2}}\oplus(\underbrace{kx_{m-2}}_{kx^{***}_{m-2}}\wedge\underbrace{ky_{m-2}}_{ky^{***}_{m-2}})\wedge\overbrace{kc_{m-2}}^{kc^{***}_{m-2}}}^{kw_{m-2}}}^{ku_{m-2}}$
⋮	⋮	⋮	⋮
kz_0	$\underbrace{kx_0}_{kx_0^*}\oplus\underbrace{ky_0}_{ky_0^*}\oplus\underbrace{kc_0}_{kc_0^*}$	$\underbrace{kc_0}_{kc_0^{**}}$	$\overbrace{\underbrace{kx_1}_{kx_1^{**}}\wedge\underbrace{ky_1}_{ky_1^{**}}\oplus(\underbrace{kx_1}_{kx_1^{***}}\wedge\underbrace{ky_1}_{ky_1^{***}})\wedge\overbrace{kc_1}^{kc_1^{***}}}^{kw_1}}^{ku_1}$

$$\begin{cases} lp_{3\text{xor}}((lx^{*},ly^{*},lc^{*})\to lz)=1 \\ lp_{3\text{copy}}(lx\to(lx^{*},lx^{**},lx^{***}))=1 \\ lp_{3\text{copy}}(ly\to(ly^{*},ly^{**},ly^{***}))=1 \\ lp_{2\text{copy}}(lc^{**}\to(lc^{*},lc^{***}))=1 \\ lp_{\text{and}}((lx^{**},ly^{**})\to lu)=1 \\ lp_{2\text{xor}}((lx^{***},ly^{***})\to lv)=1 \\ lp_{\text{and}}((lv,lc^{***})\to lw)=1 \\ lp_{2\text{xor}}((lu<<1,lw<<1)\to lc^{**})=1 \end{cases} \tag{24}$$

3.2　常数模加运算模型

常数模加运算输入 $x,rk\in(F_2)^m$，输出 $z\in(F_2)^m$ 满足

$$\begin{cases} z_{m-1}=x_{m-1}\oplus rk_{m-1} \\ z_{m-2}=x_{m-2}\oplus rk_{m-2}\oplus c_{m-2},c_{m-2}=x_{m-1}\wedge rk_{m-1} \\ z_{m-3}=x_{m-3}\oplus rk_{m-3}\oplus c_{m-3},c_{m-3} \\ \quad =x_{m-2}\wedge rk_{m-2}\oplus(x_{m-2}\oplus rk_{m-2})\wedge c_{m-2} \\ \quad\vdots \\ z_0=x_0\oplus rk_0\oplus c_0,c_0=x_1\wedge rk_1\oplus(x_1\oplus rk_1)\wedge c_1 \end{cases} \tag{25}$$

其中，$c=(c_0,c_1,\cdots,c_{m-1})\in(F_2)^m$ 是模加运算的进位。由于轮密钥为常数，则轮密钥的三子集积分可分性质满足 $D^{1^m}_{K,L}=(0,0)$。常数模加运算 K 集传播过程如表 2 所示。

表 2　　**常数模加运算 K 集传播**

输出状态	K 集传播	进位状态	K 集传播
kz_{m-1}	$\underbrace{kx_{m-1}}_{kx^{*}_{m-1}}\oplus\underbrace{0}_{kc^{*}_{m-1}}$	$\underbrace{kc_{m-1}}_{kc^{**}_{m-1}}$	$\underbrace{0}_{kx^{**}_0}\oplus\overbrace{\underbrace{0}_{kx^{***}_0}\wedge\underbrace{0}_{kc^{***}_0}}^{ku_0}$
kz_{m-2}	$\underbrace{kx_{m-2}}_{kx^{*}_{m-2}}\oplus\underbrace{kc_{m-2}}_{kc^{*}_{m-2}}$	$\underbrace{kc_{m-2}}_{kc^{**}_{m-2}}$	$\underbrace{kx_{m-1}}_{kx^{**}_{m-1}}\oplus\overbrace{\underbrace{0}_{kx^{***}_{m-1}}\wedge\underbrace{0}_{kc^{***}_{m-1}}}^{ku_{m-1}}$
kz_{m-3}	$\underbrace{kx_{m-3}}_{kx^{*}_{m-3}}\oplus\underbrace{kc_{m-3}}_{kc^{*}_{m-3}}$	$\underbrace{kc_{m-3}}_{kc^{**}_{m-3}}$	$\underbrace{kx_{m-2}}_{kx^{**}_{m-2}}\oplus\overbrace{\underbrace{kx_{m-2}}_{kx^{***}_{m-2}}\wedge\underbrace{kc_{m-2}}_{kc^{***}_{m-2}}}^{ku_{m-2}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
kz_0	$\underbrace{kx_0}_{kx^{*}_0}\oplus\underbrace{kc_0}_{kc^{*}_0}$	$\underbrace{kc_0}_{kc^{**}_0}$	$\underbrace{kx_1}_{kx^{**}_1}\oplus\overbrace{\underbrace{kx_1}_{kx^{***}_1}\wedge\underbrace{kc_1}_{kc^{***}_1}}^{ku_1}$

增加 7 个变量 $(kx^{*},kx^{**},kx^{***},kc^{*},kc^{**},kc^{***},ku)$ 辅助表示常数模加运算 K 集传播，满足

$$\begin{cases} kz=kx^{*}\oplus kc^{*} \\ kx\to(kx^{*},kx^{**},kx^{***}) \\ kc^{**}\to(kc^{*},kc^{***}) \\ ku=kx^{***}\wedge kc^{***} \\ kc^{**}=(kx^{**}\oplus ku)<<1 \end{cases} \tag{26}$$

其中，5 个比特 $(kx^{**}_0,kx^{***}_0,kx^{***}_{m-1},kc^{***}_0,kc^{***}_{m-1})$ 为 0 的分支状态。因此，在 K 集传播过程中该 5 bit 限制为 0。常数模加运算 K 集传播系统 $S_{k+c}((\boldsymbol{k}_x,\boldsymbol{k}_{rk})\to\boldsymbol{k}_z$ 满足

$$\begin{cases} kp_{2\text{xor}}((kx^{*},kc^{*})\to kz)=1 \\ kp_{3\text{copy}}(kx\to(kx^{*},kx^{**},kx^{***}))=1 \\ kp_{2\text{copy}}(kc^{**}\to(kc^{*},kc^{***}))=1 \\ kp_{\text{and}}((kx^{***},kc^{***})\to ku)=1 \\ kp_{2\text{xor}}((ku<<1,kx^{**}<<1)\to kc^{**})=1 \end{cases} \tag{27}$$

根据 K 集传播过程，可得到常数模加运算 L 集传播系统 $S_{l+c}((\boldsymbol{l}_x,\boldsymbol{l}_{rk})\to\boldsymbol{l}_z$ 满足

$$\begin{cases} lp_{2\text{xor}}((lx^{*},lc^{*})\to lz)=1 \\ lp_{3\text{copy}}(lx\to(lx^{*},lx^{**},lx^{***}))=1 \\ lp_{2\text{copy}}(lc^{**}\to(lc^{*},lc^{***}))=1 \\ lp_{\text{and}}((lx^{***},lc^{***})\to lu)=1 \\ lp_{2\text{xor}}((lu<<1,lx^{**}<<1)\to lc^{**})=1 \end{cases} \tag{28}$$

3.3　搜索算法

根据三子集可分性质经过“分支”“异或”“与”及“模加”操作的传播规则，SAT/SMT 求解器可以建立 ARX 结构分组密码轮函数的积分传播模型。由三子集可分性质经过 ARX 结构轮函数传播路径，可建立 r 轮积分传播系统。K 集和 L 集传播过程中满足一定传播条件，可通过 L 集过滤算法减少 L 集中的冗余向量从而降低搜索时间复杂度。给定输入集合 X 满足三子集积分 $(\boldsymbol{k},\boldsymbol{l})=(K_0,L_0)$，对于 r 轮输出集合 Y 满足三子集可分性质 $D^{1^m}_{K_r,L_r}$，如果 K_r 集中包含 m 个相异的单位向量，则不存在以 $(\boldsymbol{k},\boldsymbol{l})$ 积分特征为输入的 r 轮积分区分器。如果 K_r 集中不存在某单位向量 $\boldsymbol{e}_i$，对任意 $\boldsymbol{k}\in K_r$ 满足 $\oplus_{y\in Y}\pi_{\boldsymbol{e}_i}(y)=0$，则 r 轮输出的第 i bit 为平衡比特。积分自动化搜索算法的具体描述如下。

1) 初始化参数。给定三子集输入积分 $(\boldsymbol{k},\boldsymbol{l})=(K_0,L_0)$，初始化平衡集合 *Bset* 为空集，不确定集合 *Uset* 为空集。

2) 建立模型。根据三子集可分性质，通过 ARX 结构分组密码轮函数 F_r 传播规则，利用 SAT/SMT 求解器建立满足积分传播路径的 r 轮积分传播系统。

3) 求解。利用 SAT/SMT 求解器求解 r 轮积分传播系统。如果存在一组解满足步骤 2)的传播条件，添加集合 $\{i\,|\,i\in[0,m-1]\}$ 到 *Uset*，并判定不存在以 $(\boldsymbol{k},\boldsymbol{l})$ 积分特征为输入的 r 轮积分区分器。否则对所

有 $i\in[0,m-1]$ 测试 $\boldsymbol{e}_i$ 是否满足 $\boldsymbol{e}_i\in K_r$，如果满足则添加 i 到 $Uset$ 中并继续测试，否则，添加 i 到 $Bset$ 中并继续测试。最后添加 $\{D^{1^m}_{K_0,L_0}, Bset\bigcup Uset\}$ 到积分区分器集合 ID 中。

4) 搜索所有可能的积分区分器。遍历所有可能的三子集输入积分 $(\boldsymbol{k},\boldsymbol{l})$，满足 $wt(\boldsymbol{l})=m-1$, $wt(\boldsymbol{k})=m$，执行步骤 1)~步骤 3)，最终输出积分区分器集合 ID。

在建立模型的过程中，会有大量的冗余 L 集向量产生，但冗余 L 集向量并不会影响 K 集向量传播，因此，冗余的 L 集向量可通过筛除算法筛除或不做任何处理。经过模加轮密钥操作时，L 集所有向量均影响 K 集的向量传播，利用 L 集扩展算法扩展 L 集向量并添加到 K 集中。该算法的时间复杂度为 $O(m^2)$，数据复杂度 $O(1)$，空间复杂度为 $O(1)$。其中，L 集向量筛除算法 *SieveL* 的具体描述如下。

1) 初始化参数：三子集积分变量 (K,L)，积分向量元素长度 m。

2) L 集向量筛除。SAT/SMT 限制语言描述如下。

①初始化空命令字串

②向空字串添加否定断言

③for i in (0,m)

④ if i==m−1

⑤ 向命令字串添加命令“BVGE(L[i:i], L[i:i])”

⑥ 返回

⑦ else

⑧ 向命令字串添加命令“BVGE(L[i:i], L[i:i]) AND”

L 集扩展算法 *ExtendL* 的具体描述如下。

1) 初始化参数：L 集积分变量，临时变量 k，积分向量元素长度 m。

2) L 集扩展。SAT/SMT 限制语言描述如下。

①初始化空命令字串

②向空字串添加赋值断言

③for i in (0, m−1)

④向命令字串添加命令“(if L[i:i] == 0b1 then 0 else BVXOR(L,1<<i) end if) or k=”

⑤向命令字串添加命令“(if L[m−1:m−1] == 0b1 then 0 else BVXOR(L,1<<(m−1)) end if))”

4 应用

本文基于 STP[11]求解器实现积分区分器自动化搜索，平台搭载 Intel(R) Core(TM) CPU i5-4210M (2.6 GHz, 1 GB RAM, Ubuntu14.04.1)。

4.1 SIMON

SIMON 簇分组密码算法是由 NSA 提出的一种轻量级分组密码算法。SIMON 采用类 Feistel 结构，分组长度为 $2n$，表示为 SIMON- $2n$，字节长度 $n\in\{16,24,32,48,64\}$。SIMON 分组密码轮函数表示为

$$R(x_i,y_i)=(((x_{i-1}<<<\alpha)\wedge(x_{i-1}<<<\beta))\oplus (x_{i-1}<<<\gamma)\oplus y_{i-1}\oplus rk_i, x_{i-1})$$

其中，循环移位常数 $\alpha=8$、$\beta=1$、$\gamma=2$。SIMON 簇分组密码算法轮函数如图 1 所示。

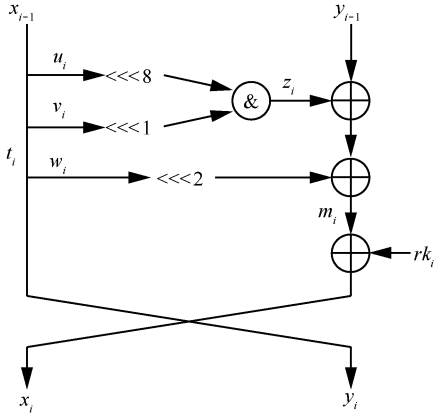

图 1 SIMON 簇分组密码算法轮函数

SIMON 轮函数 $(x_{i-1},y_{i-1})\mapsto(x_i,y_i)$，输入集合 X 满足三子集可分性质 $D^{1^{2n}}_{(kx_{i-1},ky_{i-1}),(lx_{i-1},ly_{i-1})}$。根据三子集传播规则，$K$ 集传播满足

$$\begin{cases} kp_{3\text{copy}}(kx_{i-1}\to(ku_i,kv_i,kt_i))=1\\ kp_{2\text{copy}}(kt_i,\to(kw_i,ky_i))=1\\ kp_{\text{and}}((ku_i<<<8,kv_i<<<1)\to kz_i)=1\\ kp_{3\text{xor}}((kw_i<<<2,kz_i,ky_{i-1})\to km_i)=1 \end{cases} \tag{29}$$

L 集传播满足

$$\begin{cases} lp_{3\text{copy}}(lx_{i-1}\to(lu_i,lv_i,lt_i))=1\\ lp_{2\text{copy}}(lt_i,\to(lw_i,ly_i))=1\\ lp_{\text{and}}((lu_i<<<8,lv_i<<<1)\to lz_i)=1\\ lp_{3\text{xor}}((lw_i<<<2,lz_i,ly_{i-1})\to lm_i)=1 \end{cases} \tag{30}$$

经过异或轮密钥操作时，对 L 集变量 lx_i 扩展 $ExtendL(lx_i,ktemp_i,n)$ 为辅助变量 $ktemp_i$，并将变量 km_i 和 $ktemp_i$ 赋值给 kx_i。轮函数输出三字集满足 $SieveL(lx_i\,|\,ly_i,kx_i\,|\,ky_i,2n)$。利用该搜索算法，

SIMON 分组密码积分区分器如表 3 所示。

表 3　SIMON 分组密码积分区分器

分组密码	轮数	数据量	平衡比特	条数
SIMON32	14	2^{31}	3	32
SIMON48	15	2^{47}	24	48
SIMON64	17	2^{63}	22	64
SIMON96	21	2^{95}	5	96
SIMON128	25	2^{127}	3	128

4.2　HIGHT

HIGHT 分组密码算法是由 Deukio 等[8]在 CHES2006 中提出的一种轻量级分组密码算法。HIGHT 分组密码长度为 64 bit，密钥长度为 128 bit。HIGHT 分组密码算法轮函数如图 2 所示。

图 2 中的 $F_t, t\in[0,1]$ 函数定义为

$$F_t(x)=(x<<<\alpha)\oplus(x<<<\beta)\oplus(x<<<\gamma) \quad (31)$$

HIGHT 轮函数 K 集传播变量 $m_{i,j}, j\in[0,3]$ 传播经过函数 F_t 到 $n_{i,j}, j\in[0,3]$，当 $t=0$ 时满足 $(\alpha,\beta,\gamma)=(1,2,7)$，当 $t=1$ 时满足 $(\alpha,\beta,\gamma)=(3,4,6)$。需要增加额外的 6 个变量 $m_{i,j,1}, m_{i,j,2}, m_{i,j,3}$，$n_{i,j,1}, n_{i,j,2}, n_{i,j,3}$ 来描述 K 集传播，满足传播系统 $S_{F_t}(\boldsymbol{k}_m\to\boldsymbol{k}_n)$

$$\begin{cases} kp_{3\text{copy}}(m_{i,j}\to(m_{i,j,1},m_{i,j,2},m_{i,j,3}))=1 \\ kp_{3\text{xor}}((n_{i,j,1},n_{i,j,2},n_{i,j,3})\to n_{i,j})=1 \\ n_{i,j,1}=m_{i,j,1}<<<\alpha \\ n_{i,j,2}=m_{i,j,2}<<<\beta \\ n_{i,j,3}=m_{i,j,3}<<<\gamma \end{cases} \quad (32)$$

由于模加运算输入变量 $n_{i,j}, j\in[0,3]$ 需要额外增加 3 个变量描述模加积分传播，且满足

$$kp_{3\text{xor}}((n_{i,j,1},n_{i,j,2},n_{i,j,3})\to n_{i,j})=1 \quad (33)$$

则 K 集传播时可以省略 3 个额外变量以降低搜索时间复杂度，HIGHT 轮函数 K 集传播满足传播方程

$$\begin{cases} kp_{2\text{xor}}((kx_{i-1,7},kz_{i,0})\to kx_{i,0})=1 \\ kp_{2\text{copy}}(kx_{i-1,6}\to(km_{i,0},kx_{i,7}))=1 \\ kp_{2\text{copy}}(kx_{i-1,4}\to(km_{i,1},kx_{i,5}))=1 \\ kp_{2\text{xor}}((kx_{i-1,3},kz_{i,2})\to kx_{i,4})=1 \\ kp_{2\text{copy}}(kx_{i-1,2}\to(km_{i,2},kx_{i,3}))=1 \\ kp_{2\text{copy}}(kx_{i-1,0}\to(km_{i,3},kx_{i,1}))=1 \\ S_{k+c}(kn_{i,0},krk_{4i-1})\to kz_{i,0} \\ S_{k+}(kz_{i,1},kx_{i-1,5})\to kx_{i,6} \\ S_{k+c}(kn_{i,2},krk_{4i-3})\to kz_{i,2} \\ S_{k+}(kz_{i,3},kx_{i-1,1})\to kx_{i,2} \\ S_{F_t}(km_{i,j}\to kn_{i,j}) \end{cases} \quad (34)$$

L 集传播过程中 3xor 与 3copy 操作时不等价，经过 F_t 函数后 3 个额外变量不能省略。HIGHT 轮函数 L 集传播同 K 集传播方程。经过异或轮密钥操作时，对 L 集变量 $ln_{i,1}$、$ln_{i,3}$ 扩展为辅助变量 $ktemp_{i,1}$、$ktemp_{i,3}$，并将辅助变量 $ktemp_{i,1}$、$ktemp_{i,3}$ 分别赋值给 $kn_{i,1}$、$kn_{i,3}$。轮函数输出三字集满足 L 集向量筛除规则。利用该搜索算法，HIGHT 积分区分器如表 4 所示。

表 4　HIGHT 积分区分器

分组密码	轮数	数据量	平衡比特	条数
HIGHT	18	2^{63}	1	4
	18	2^{63}	2	4

4.3　其他算法

利用该算法搜索得到，SIMECK 簇分组密码算法[12]SIMECK32/48/64 分别存在 14、17 和 20 轮积分区分器；SPECK 簇分组密码算法 SPECK32/48/64/

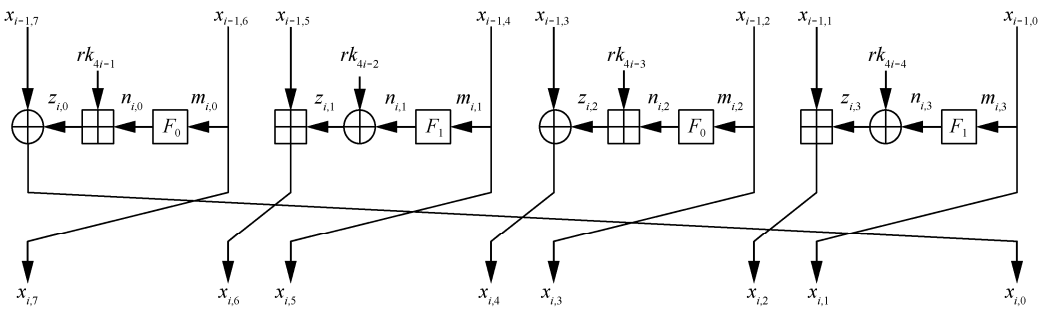

图 2　HIGHT 分组密码算法轮函数

96/128 存在 6 轮积分区分器；LEA 分组密码算法存在 8 轮积分区分器。

5 结束语

本文提出一种基于 SAT/SMT 求解器自动化搜索 ARX 结构分组密码积分区分器的方法。利用三子集积分可分技术，通过建立缩减轮数的 ARX 结构分组密码算法积分传播系统，求解得到缩减轮数的积分区分器，并对所有版本的 SIMON32/48/64/96/128 算法进行三子集积分区分器搜索，分别得到 14、15、17、21 和 25 轮积分区分器，进一步精确了 Todo 提出的 SIMON 积分界。利用该算法，搜索得到 8 条 18 轮 HIGHT 积分区分器。

SAT/SMT 求解器同样能够应用到 SPN 结构及 Feistel 结构分组密码算法中，但不能有效完整地描述大状态 S 盒积分传播规则。如何自动化搜索大状态 S 盒积分传播，是未来需要研究的工作。

参考文献：

[1] TODO Y. Structural evaluation by generalized integral property[C]//EUROCRYPT. 2015: 287-314.

[2] TODO Y. Integral cryptanalysis on full MISTY1[C]//CRYPTO. 2015: 413-432.

[3] WANG Q J, LIU Z Q, KEREM V, et al. Cryptanalysis of reduced-round SIMON32 and SIMON48[C]//INDOCRYPT. 2014: 143-160.

[4] TODO Y, MORII M. Bit-based division property and application to simon family[C]//Fast Software Encryption. 2016: 357-377.

[5] ALEX B, ARNAB R, VESSELIN V. Differential analysis of block ciphers SIMON and SPECK[C]//Fast Software Encryption. 2014: 546-570.

[6] XIANG Z J, ZHANG W T, BAO Z Z, et al. Applying MILP method to searching integral distinguishers based on division property for 6 lightweight block ciphers[C]// ASIACRYPT. 2016: 648-678.

[7] SUN L, WANG W, LIU R, et al. Milp-aided bit-based division property for arx-based block cipher[M]. IACR Cryptology ePrint Archive, 2016.

[8] DEUKIO H, JAECHUL S, SEOKHIE H, et al. HIGHT: a new block cipher suitablefor low-resource device[C]//Cryptographic Hardware and Embedded Systems. 2006: 46-59.

[9] DEUKIO H, JUNG K L, DONG C K, et al. LEA: a 128-bit block cipher for fast encryption on common processors[C]//WISA. 2013: 3-27.

[10] DAVID J, WHEELER, ROGER M. Tea, a tiny encryption algorithm[C]//Fast Software Encryption. 1994: 363-366.

[11] YAO J T, LIU W N. The STP model for solving imprecise problems[C]//GrC. 2006: 683-687.

[12] YANG G Q, ZHU B, VALENTIN S, et al. The simeck family of lightweight block ciphers[C]//Cryptographic Hardware and Embedded Systems. 2015: 307-329.

[作者简介]

韩亚（1989-），男，河南商丘人，中国科学院信息工程研究所博士生，主要研究方向为信息安全、密码学等。

王明生（1967-），男，四川遂宁人，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为轻量密码学、大数据密码和密码相关的困难问题等。