

基于深度学习的CRAFT算法差分分析

沈璇¹, 沈霞民², 卜予彤², 代威¹, 黄金杰¹, 郭连城¹

(1. 信息支援部队工程大学, 湖北 武汉 430010; 2. 国防科技大学, 湖南 长沙 410073)

摘要: 基于深度学习的分组密码差分分析已成为当前密码学领域研究热点, 为研究不同数据格式对神经网络差分区分器性能的影响, 结合单差分、多差分区分器数据格式特点, 利用CRAFT算法S盒组件具有差分传播不确定性的特性, 设计提出“单差分+S盒中间状态”区分器数据格式, 通过残差神经网络模型分别训练获得3类数据格式对应的CRAFT算法5~7轮神经网络差分区分器。对于5轮神经网络差分区分器, 3类区分器准确率均在97%以上; 对于6轮神经网络差分区分器, “单差分+S盒中间状态”区分器准确率为96.14%, 明显优于单差分区分器的61.39%和多差分区分器的72.78%; 对于7轮神经网络差分区分器, 3类区分器准确率略高于50%。其中, 单差分和“单差分+S盒中间状态”区分器对应模型训练数据复杂度为 $2^{21.07}$, 多差分区分器对应模型训练数据复杂度为 $2^{21.66}$ 。进一步利用6轮“单差分+S盒中间状态”区分器对7轮CRAFT算法进行部分密钥恢复攻击, 当选择明文对数量为10 000个时, 在100次攻击实验中成功率可达93%, 一次完整的密码分析攻击时间复杂度为 $2^{22.01}$ 。

关键词: 深度学习; 残差神经网络; 轻量级分组密码; CRAFT; 神经网络差分区分器; 密钥恢复攻击

中图分类号: TN918

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.AQ25290

Differential analysis of the CRAFT algorithm based on deep learning

Shen Xuan¹, Shen Xiamin², Bu Yutong², Dai Wei¹, Huang Jinjie¹, Guo Liancheng¹

1. Information Support Force Engineering University, Wuhan 430010, China

2. National University of Defense Technology, Changsha 410073, China

Abstract: The differential analysis of block ciphers based on deep learning has become a hot research topic in the field of cryptography. To investigate the impact of different data formats on the performance of neural network differential distinguisher, by combining the characteristics of single-differential and multi-differential distinguisher data formats, and taking advantage of the differential propagation uncertainty of the S-box component in the CRAFT, a “single-differential + S-box intermediate state” distinguisher data format was designed and proposed. Through residual neural network (ResNet) models, three types of data formats corresponding to the 5-7 rounds of neural network differential distinguisher of the CRAFT were obtained. For the 5-round neural network differential distinguisher, the accuracy of all three discriminators was above 97%. For the 6-round neural network differential distinguisher, the accuracy of the “single-differential + S-box intermediate state” distinguisher was 96.14%, significantly better than the 61.39% of the single-differential distinguisher and the 72.78% of the multi-differential distinguisher. For the 7-round neural network

收稿日期: 2025-12-07; 修回日期: 2026-05-18

通信作者: 沈霞民, sxmc042012@163.com

基金项目: 国家自然科学基金资助项目 (No.62272470)

Foundation Item: The National Natural Science Foundation of China (No.62272470)

differential distinguisher, the accuracy of the three distinguishers was slightly above 50%. Among them, the training data complexity of the model corresponding to the single-differential distinguisher was $2^{21.07}$, while the training data complexity of the multi-differential and “single-differential + S-box intermediate state” distinguisher was $2^{21.66}$. Further, by using a 6-round “single difference + S-box intermediate state” distinguisher to conduct partial key recovery attacks on the 7-round CRAFT, when the number of plaintext-ciphertext pairs selected was 10 000, the success rate can reach 93% in 100 attack experiments, the time complexity of a complete cryptanalysis attack was $2^{22.01}$.

Key words: deep learning, residual neural network, lightweight block cipher, CRAFT, neural network differential distinguisher, key recovery attack

0 引言

密码算法是确保数据安全的重要保证, 轻量级分组密码^[1]具有计算效率高、存储需求小、功耗低等特点, 被广泛运用于资源受限的信息化设备中, 如 SKINNY^[2]、CRAFT^[3]、SPECK^[4]、CLEFIA^[5-6]等。针对轻量级分组密码分析, 2019年 Gohr^[7]首次提出基于深度学习的轻量级分组密码安全性分析, 采用残差神经网络构建了减轮 SPECK32/64 神经网络差分区分器, 可有效区分固定明文差分加密的密文对和随机数据, 证明了神经网络差分区分器的有效性, 为深度学习辅助密码分析开辟了新的研究方向。

近年来, 密码学界陆续展开深入研究, 研究表明借助深度学习技术的安全性分析相比于传统的攻击手段能够降低攻击复杂度、减少计算成本, 而提高神经网络差分区分器准确率是密码安全分析的关键, 主要通过完善优化神经网络模型和调整改变数据集格式来实现, 具体可细分为以下4类研究方向。第一类调整数据集格式: 2021年, Chen等^[8]在 Gohr 模型的基础上进行了优化, 原模型神经网络输入为单个密文对, 优化后的模型将神经网络输入调整为多个密文对, 新模型得到的结果明显优于原模型。2024年, 杨小雪等^[9]以 PRESENT 和 KLEIN 为研究对象, 通过改变输入矩阵宽度选取合适的输入矩阵格式, 可有效地提高神经网络差分区分器的准确率。第二类引入不同固定输入差分: 2021年, 段明等^[10]提出了一种多差分残差神经网络区分器, 该区分器使用多个输入差分得到多个密文对作为分析对象, 通过训练得到了 RC5 算法 5 轮、6 轮神经网络差分区分器。2023年, 蒋若怡等^[11]设计提出多差分神经网络区分器通用模型, 该模型输入被设置为多个明文差分、相应密文及密文差分, 将其应用

于分析 SPECK32/64 算法, 结果表明 5~7 轮神经网络差分区分器准确率均有显著提升。2023年, 王慧娇等^[12]研究验证了 SM4 算法不同输入差分的汉明权重对神经网络区分器性能存在一定程度的影响。第三类优化神经网络模型: 2022年, 侯泽洲等^[13]基于卷积神经网络和残差神经网络, 通过优化模型参数训练得到了 SIMON32/64 算法 7 轮、8 轮和 SPECK32/64 算法 6 轮、7 轮神经网络差分区分器。第四类综合神经网络模型优化以及数据集格式改进: 2024年, 栗琳珂等^[14]针对 SIMON、SIMECK 和 SPECK 3 类密码算法, 基于引入注意力机制的深度残差网络构造了精度更高的神经网络差分区分器, 并研究了输入数据中只包含密文对和包含密文对、倒数第二轮差分信息两种数据格式对神经网络差分区分器精度的影响。

基于已有研究成果, 本文研究重点为通过改进丰富数据集特征信息, 提升神经网络差分器的准确率, 以 CRAFT 算法为研究对象, 利用残差神经网络模型, 研究不同数据集格式对 CRAFT 算法神经网络差分区分器准确率的影响。CRAFT 算法是一种可调轻量级分组密码算法, 设计者对其进行了差分分析、线性分析、中间相遇攻击等安全性分析, Hadipour 等^[15]找到了该算法概率为 $2^{-40.20}$ 的 9 轮差分区分器, Liu 等^[16]基于布尔可满足性问题 (SAT) 搜索方法得到了该算法概率为 $2^{-40.68} + 2^{-48.60}$ 的 9 轮差分区分器。当前, 现有文献暂无针对 CRAFT 算法的神经网络差分区分器分析研究, 类比于 CRAFT 算法同类型的分组密码相关研究, 2022年, 蒋若怡^[17]基于深度学习对 SKINNY-64 进行了差分分析, 在单差分和多差分两种模型中分别捕获了 5 轮和 6 轮 SKINNY-64 差分神经网络区分器。鉴于 CRAFT 和 SKINNY 同属轻量级可调分组密码, 均采用代换-置换网络

(SPN) 密码结构, 组件设计思路相似 (如 4 bit S 盒、列混淆、轮密钥加等), 在差分分析、截断差分分析、矩阵分析等攻击中, 它们总是被作为一组具有代表性的目标算法共同研究和比较。例如, 2020 年, Ebrahimi 等^[18]同时分析了 SKINNY 和 CRAFT 的截断差分特性, 并改进了针对这两种算法的分析结果。2021 年, Hadipour 等^[19]通过自动化搜索工具, 针对两种轻量级密码算法 SKINNY 和 CRAFT, 找到了更长的矩形攻击区分器。2023 年, Hadipour 等^[20]提出了一种新的自动化搜索方法, 并同时应用于 SKINNY 和 CRAFT, 显著改进了针对这两种算法的多种攻击结果。基于此, 本文进一步开展了 CRAFT 神经网络差分区分器构造研究, 与传统的差分分析相同, 训练神经网络差分区分器同样需要合适的输入差分, 选取文献[15-16]研究结果中的 9 轮差分路径的输入差分作为神经网络差分模型的固定输入差分。为探究不同数据格式对神经网络差分区分器的影响, 设计了单差分、多差分 and “单差分+S 盒中间状态” 3 类区分器数据集格式, 分别训练得到了 CRAFT 算法 5~7 轮神经网络差分区分器, 如表 1 所示。其中, 在开展 CRAFT 算法 r 轮神经网络差分区分器研究时, C_i 表示 r 轮加密后的密文结果; C_i^{-s} 表示 C_i 作 S 盒逆运算后的值, 即第 r 轮加密过程中, 经 S 盒运算前的中间状态值; Y 表示数据样本标签。对于 5 轮神经网络差分区分器, 3 种情形准确率均在 97% 以上; 对于 6 轮神经网络差分区分器, 准确率由高到低为“单差分+S 盒中间状态”、多差分、单差分, “单差分+S 盒中间状态” 区分器准确率为 96.14%, 明显优于其他两种情形; 对于 7 轮神经网络差分区分器, 3 种情形准确率略高于 50%。其中, 单差分和“单差分+S 盒中间状态” 区分器对应模型训练数据复杂度为 $2^{21.07}$, 多差分区分器对应模型训练数据复杂度为 $2^{21.66}$ 。为进一步验证“单差分+S 盒中间状态” 这一最优数据格式是否具备真实密码攻击场景下的可获得性, 使用 6 轮“单差分+S 盒中间状态” 区分器对 7 轮 CRAFT 算法进行部分密钥恢复攻击, 当选择 10 000 组明密文对时, 在 100 次攻击中密钥恢复成功率可达 93%, 一次完整的密码分析攻击时间复杂度为 $2^{22.01}$ 。

表 1 CRAFT 算法 5~7 轮神经网络差分区分器评估结果

训练数据格式	5 轮	6 轮	7 轮
$\{C_1, C_2, C_1 \oplus C_2; Y\}$	97.78%	61.39%	50.18%
$\{C_1, C_2, C_3, C_1 \oplus C_2, C_1 \oplus C_3, C_2 \oplus C_3; Y\}$	99.83%	72.78%	50.26%
$\{C_1, C_2, C_1^{-s}, C_2^{-s}, C_1 \oplus C_2, C_1^{-s} \oplus C_2^{-s}; Y\}$	99.60%	96.14%	50.46%

综上所述, 本文主要的研究工作如下。

(1) 在单差分和多差分区分器基础上, 利用 CRAFT 算法 S 盒组件具有差分传播不确定性的特点, 提出了“单差分+S 盒中间状态” 区分器数据集格式, 以此获取更多数据信息特征。

(2) 基于残差神经网络, 采用差分分析思想, 构建了 CRAFT 算法残差神经网络模型, 通过训练调试优化了模型超参数的设置。

(3) 利用 CRAFT 算法残差神经网络模型, 输入 3 类不同的区分器数据集格式, 分别训练得到了 CRAFT 算法 5~7 轮神经网络差分区分器, 并对比分析不同数据格式对区分器精度的影响, 得出“单差分+S 盒中间状态” 区分器可有效提升准确率, 且相比多差分区分器对应模型训练数据复杂度更低的结论。

(4) 实验评估结果表明, 该攻击的有效性在 100 次独立重复实验中得以验证, 平均密钥恢复成功率为 93%。

1 预备知识

1.1 CRAFT 算法

轻量级可调整分组密码 CRAFT 算法分组长度为 64 bit、密钥长度为 128 bit、调柄值为 64 bit, CRAFT 算法的内部状态 Stt_i 可以表示为 4×4 半字节数组。

$$\text{Stt}_i = \begin{Bmatrix} m_0 & m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 & m_7 \\ m_8 & m_9 & m_{10} & m_{11} \\ m_{12} & m_{13} & m_{14} & m_{15} \end{Bmatrix} \quad (1)$$

内部状态经过 32 轮加密运算后, 生成 64 bit 密文。前 31 轮中的每一轮都包含 5 个函数, 分别为列混淆 (MC)、常数加 (RC)、轮密钥加 (RT)、置换 (PN) 和 S 盒替换 (SB), 最后一轮仅包含列混淆、常数加和轮密钥加, 如图 1 所示,

其中R_shift代表右循环移位, shuffle代表单元置换, L_shift代表左循环移位。

(1) 列混淆。内部状态的每一列都乘以下面的二进制矩阵 M 。

$$M = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

(2) 常数加。将状态矩阵异或一个轮常量, 轮常量如表2所示。在第 i 轮加密过程中, 中间状态 Stt_4 和 Stt_5 分别与 RC_4 、 RC_3 异或。

表2 CRAFT算法轮常量

轮数 i	$RC_i = (a, b)$
0~15	11, 84, 42, 25, 96, c7, 63, b1, 54, a2, d5, e6, f7, 73, 31, 14
16~31	82, 45, 26, 97, c3, 61, b4, 52, a5, d6, e7, f3, 71, 34, 12, 85

$$RC_4 = [1, 8, 4, 2, 9, c, 6, b, 5, a, d, e, f, 7, 3, 1, 8, 4, 2, 9, c, 6, b, 5, a, d, e, f, 7, 3, 1, 8] \quad (3)$$

$$RC_3 = [1, 4, 2, 5, 6, 7, 3, 1, 4, 2, 5, 6, 7, 3, 1, 4, 2, 5, 6, 7, 3, 1, 4, 2, 5, 6, 7, 3, 1, 4, 2, 5] \quad (4)$$

(3) 轮密钥加。内部状态的每半字节与轮密钥 TK_i 的每半字节进行异或。轮密钥 TK_i 编排方案如下: 128 bit主密钥 K 被分成两个64 bit密钥 K_0 和 K_1 。从 K_0 、 K_1 和相关的调柄值 T 派生出4个64 bit密钥 TK_0 、 TK_1 、 TK_2 和 TK_3 , 最终生成轮密钥 TK_i 。

$$\begin{cases} TK_0 = K_0 \oplus T \\ TK_1 = K_1 \oplus T \\ TK_2 = K_0 \oplus Q(T) \\ TK_3 = K_1 \oplus Q(T) \end{cases} \Rightarrow TK_i = TK_{i \bmod 4} \quad (0 \leq i \leq 31) \quad (5)$$

$Q(T) = [12, 10, 15, 5, 14, 8, 9, 2, 11, 3, 7, 4, 6, 0, 1, 13]$ 是在应用调柄值 T 的半字节上的置换。

(4) 置换。将置换 P 应用于内部状态进行重新排序, $P = [15, 12, 13, 14, 10, 9, 8, 11, 6, 5, 4, 7, 1, 2, 3, 0]$ 。

(5) S盒替换。使用表3所示的S盒并行地应用于内部状态的每个半字节, a、b、c、d、e、f代表16进制数。

表3 CRAFT算法的S盒函数

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$S(x)$	c	a	d	3	e	b	f	7	8	9	1	5	0	2	4	6

1.2 差分分析

差分分析在分组密码分析领域是一种重要的研究方法, 针对分组密码差分分析的重点在于找到一个轮数或者概率更大的差分区分器。

引理^[21] 对 $\{0, 1\}^n$ 上的随机置换 $\mathcal{R}$, 任意给定差分 $\{\alpha, \beta\}$, 则 $DP(\alpha, \beta) = \frac{1}{2^n - 1} \approx \frac{1}{2^n}$, ($\alpha \neq 0$, $\beta \neq 0$)。如果找到了一条 r 轮差分, 其概率大于 $\frac{1}{2^n}$, 就可以将 r 轮的加密算法与随机置换区分开, 利用该差分区分器对分组密码进行密钥恢复攻击。

差分区分器对明文或者密文的区分, 从本质上说就是对明文数据或者密文数据的分类, 而深度学习技术在图像语音识别^[22]、异常流量检测^[23]、分组密码结构识别^[24]等领域被广泛应用于分类问题, 这与差分区分器的功能是相似的。与传统差分区分器效果相同, 神经网络差分区分器是在输入差分确定的情况下, 对大量已知密文数据进行学习, 挖掘并提取大量有效密文数据特征, 从而实现对未知密文数据的分类。

1.3 残差神经网络

残差神经网络(ResNet)由He等^[25]在2016年提出, 是一种高效且适用范围广的卷积神经网络。与卷积神经网络不同, 残差网络加入了捷径连接来形成残差单元, 它可将其中一层的激活值绕道传给更深层的网络, 在大幅提高网络结构的

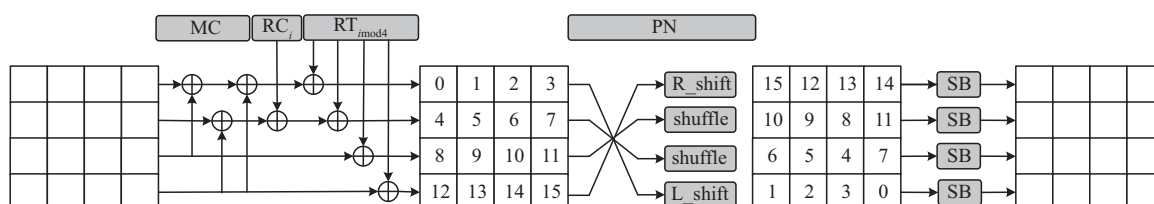


图1 CRAFT算法的加密过程

层级时, 提高网络训练的准确率, 同时有效避免梯度消失问题。ResNet 的核心思想是引入“跳跃链接”, 即允许前面某一层的输出直接跨过一层或多层, 叠加到后面数据层的输入部分, 这种结构被称为残差块, 该残差块分为两个部分: 残差映射 $F(x)$ 和恒等映射 x , 如图 2 所示, ReLU 为激活函数。

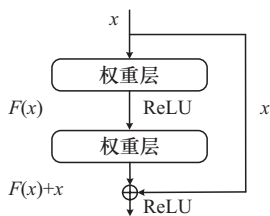


图2 残差块

2 神经网络差分区器构造

神经网络差分区器构造的关键在于数据集构造和深度学习神经网络模型设计, 本文主要研究如何调整改进数据形式来提升神经网络差分区器的准确率。

2.1 高概率差分选取

找到一个轮数尽可能长、概率尽可能大的差分区器, 离不开良好的输入差分, 与传统的差分区器相同, 神经网络差分区器也需要给定输入差分。为了保证训练样本有效性, 根据文献[15-16]关于 CRAFT 算法的差分分析结果, 选取 $\Delta x_1 = \text{aa0a, aa00, 0000, aa00}$ 和 $\Delta x_2 = \text{aaa0, aa00, 0000, aa00}$ 作为神经网络差分区器的固定输入差分。

2.2 神经网络差分区器数据结构

2021 年, Benamira 等^[26]深入研究 Gohr^[7]提出的神经网络差分区器, 表明神经网络差分区器性能主要依赖于密文对的差分分布, 但也依赖于倒数第二轮的差分分布。2023 年, Bao 等^[27]指出 Gohr 提出的神经网络差分区器除了利用纯密文差分信息外, 还利用了密文的偏值、密文对的差分 and 中间状态的差分之间的相关性。基于上述思想, 仅以最终轮密文作为输入, 难以有效捕获算法内部中间状态的非线性传播规律, 为进一步提取更加丰富的数据特征信息, 在构造缩减轮为 r 轮的神经网络差分区器时, 记录获取第 r 轮加密过程中某一部件前后的中间状态值作为数据集的补充部分, 具体分析对比以下 3 类数据格式。

(1) 单差分区器。固定输入差分数量为 1, 选取随机明文和固定输入差分, 通过 CRAFT 算法加密生成对应密文对, 然后异或得到密文差分, 将密文对、密文差分以及样本标签结合起来构成单个数据样本。

(2) 多差分区器。本文考察固定输入差分数量为 2 个的情形, 选取随机明文和 2 个固定输入差分, 通过 CRAFT 算法加密生成对应的 3 个密文, 然后两两异或得到 3 个密文差分, 将密文、密文差分以及样本标签结合起来构成单个数据样本。

(3) “单差分+S 盒中间状态”区分器。固定输入差分数量为 1 个, 选取随机明文和固定输入差分, 通过缩减轮为 r 轮的 CRAFT 算法加密生成密文对以及在第 r 轮加密过程中进入 S 盒运算前的中间状态值, 然后将密文和中间状态值分别异或, 附上样本标签结合起来构成单个数据样本。

2.3 数据生成

步骤 1 生成正标签密文样本。随机选择明文 P_1 , 使得 $P_2 = \Delta x_1 \oplus P_1$, $P_3 = \Delta x_2 \oplus P_1$, 组成明文对 (P_1, P_2) 或 (P_1, P_2, P_3) , 在密钥 K_i 下 (为了防止神经网络出现单一的密钥依赖, 每一组密文都选择由不同的密钥加密而成), 经 r 轮加密, 得到密文对 (C_1, C_2) 、 (C_1, C_2, C_3) 以及第 r 轮加密过程中进入 S 盒运算前的中间状态值 (C_1^{-s}, C_2^{-s}) , 记为正标签密文样本, 标记 $Y = 1$ 。

步骤 2 生成负标签密文样本。同步步骤 1, 将上述的固定输入差分替换为随机输入差分, 随机输入差分生成规则为“循环调用 Python 的 random.choice() 函数 16 次, 从 0x0 到 0xf 的十六进制数中随机选取元素, 得到一个包含 16 个半字节的随机数组, 数据具有离散均匀分布特性”, 通过该生成规则产出相同数量的负标签密文样本, 标记 $Y = 0$ 。

步骤 3 拼接构造训练数据集。①单差分区器: 数据格式为 $\{C_1, C_2, C_1 \oplus C_2; Y\}$, 单个样本共 193 bit, 含 1 bit 样本标签; ②多差分区器: 数据格式为 $\{C_1, C_2, C_3, C_1 \oplus C_2, C_1 \oplus C_3, C_2 \oplus C_3; Y\}$, 单个样本共 385 bit, 含 1 bit 样本标签; ③“单差分+S 盒中间状态”区分器: 数据格式为 $\{C_1, C_2, C_1^{-s}, C_2^{-s}, C_1 \oplus C_2, C_1^{-s} \oplus C_2^{-s}; Y\}$, 单个样本共 385 bit, 含 1 bit 样本标签。

综上, 训练样本、验证样本数据集生成算法如算法1所示。

算法 数据集生成算法

输入 数据样本数量 n , 固定输入差分 Δx_1 、 Δx_2

输出 数据集 X , 标签 Y

(1) 随机分别生成 $\frac{n}{2}$ 个 64 bit 正、负标签明文
for i in range $\left(\frac{n}{2}\right)$:

$(P_i, T, Y_i) \leftarrow (\text{data_random_64}(), 1)$

$(P_i, F, Y_i) \leftarrow (\text{data_random_64}(), 0)$

(2) 随机生成 n 个 128 bit 密钥和 64 bit 输入差分 for i in range (n) :

$K_i \leftarrow \text{data_random_128}()$

$\Delta x \leftarrow \text{data_random_64}()$

(3) 根据标签值, 组成对应明文对

if $Y_i = 1$

$(P_1, P_2, P_3) \leftarrow (P_i, T, P_i \oplus T \oplus \Delta x_1, P_i \oplus T \oplus \Delta x_2)$ else

$(P_1, P_2, P_3) \leftarrow (P_i, F, P_i \oplus F \oplus \Delta x_1, P_i \oplus F \oplus \Delta x_2)$

(4) 通过CRAFT算法经密钥 K_i 进行减轮加密

$(C_1, C_1^{-s}) \leftarrow (\text{Enc}(P_1, K_i), \text{Enc}_S(P_1, K_i))$

$(C_2, C_2^{-s}) \leftarrow (\text{Enc}(P_2, K_i), \text{Enc}_S(P_2, K_i))$

$C_3 \leftarrow \text{Enc}(P_3, K_i)$

(5) 分别构造单差分、多差分 and “单差分+S盒中间状态”数据集格式

$X \leftarrow (C_1, C_2, C_1 \oplus C_2)$
 $X \leftarrow (C_1, C_2, C_3, C_1 \oplus C_2, C_1 \oplus C_3, C_2 \oplus C_3)$
 $X \leftarrow (C_1, C_2, C_1^{-s}, C_2^{-s}, C_1 \oplus C_2, C_1^{-s} \oplus C_2^{-s})$

(6) 返回神经网络模型所需数据集

return (X, Y)

2.4 神经网络模型设置

基于ResNet训练构造SPECK、SIMON算法的神经网络差分区分器时, 神经网络模型主要分为5个部分, 分别是输入层、一维卷积层、残差网络层、全连接层和输出层, 且研究表明残差网络层中残差块个数对模型性能影响存在以下特点: 蒋若怡等^[11]得出, 当残差块的数量达到一定值后, 模型在验证集上的准确率无明显变化, 但是会随着深度增加, 运行时间逐渐增加; 侯泽洲等^[13]得出, 5个残差块的ResNet5模型相比10个残差块的ResNet10模型, 在验证集上的准确率不升反降, 且时间开销增加1倍。基于上述研究结果, CRAFT算法的神经网络模型结构如图3所示。其中, 输入模块是网络的输入层, 将分析对象转换成神经网络所需的格式, 对于单差分区分器, 输入数据格式为64×3的矩阵, 对于多差分 and “单差分+S盒中间状态”区分器, 输入数据格式为32×12的矩阵; 模块1是一个简单的一维卷积层, 模块2共循环5次, 并引入了残差块, 这两个模块神经元个数为64个, 还包括正则化函数以及激活函数ReLU, 主要功能是提取密文

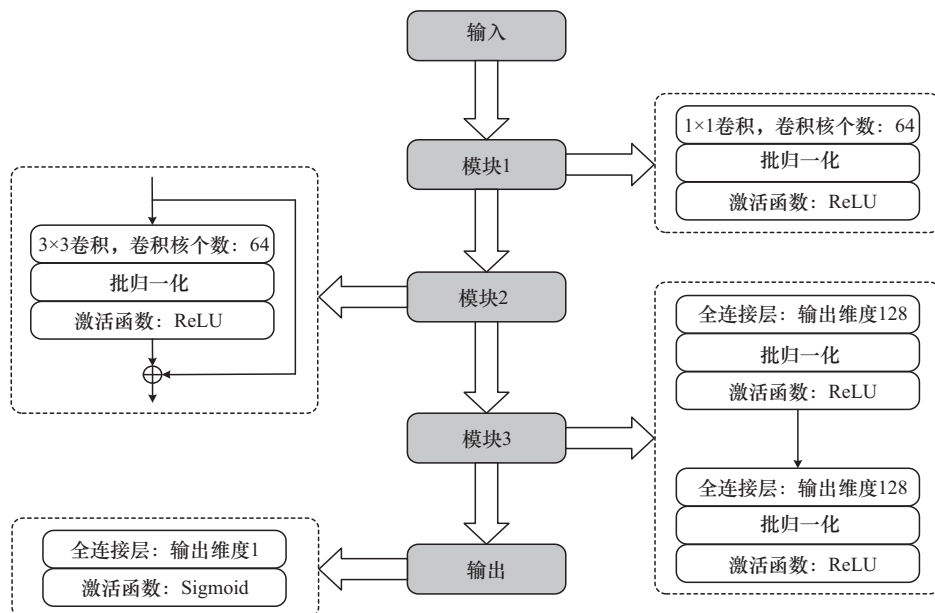


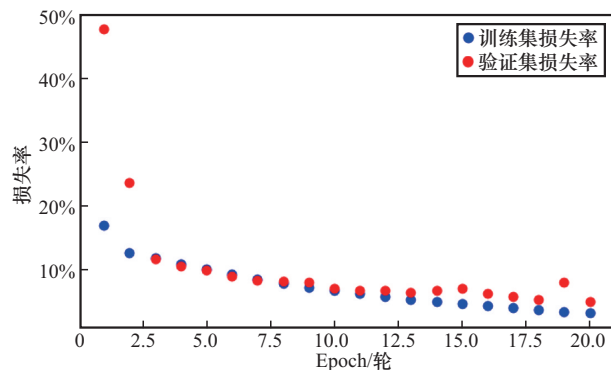
图3 CRAFT算法的神经网络模型结构

对、密文差分对以及中间状态值之间的特征；模块3由两个全连接层构成，神经元个数为128个；输出模块是输出层，神经元个数为1个，激活函数为Sigmoid，主要对数据进行分类，判别其是真实密文对或随机数据。

文献[10-11,14]基于ResNet训练构造了神经网络差分区器，超参数通常设置为批次大小 (Batch Size) =5 000、损失函数 (Loss Function) 为均方误差 (MSE) 函数、优化器 (Optimizer) 为Adam、学习速率 (Lr) =0.000 1或0.000 01、训练轮次 (Epoch) =20~200 轮、卷积核大小 (kernel_size) =3 (残差网络层)、激活函数 (Activation) =ReLU (卷积层) 和 Sigmoid (输出层)。在此基础上，CRAFT算法神经网络差分区器模型部分超参数除取上述文献通用值外，其他参数经多次训练调试后取最优值，其中，Epoch接近20轮时，在验证集上的准确率趋于平稳，再增加时没有太大变化，反而增大了训练时间；单差分区器对应数据为192 bit，输入矩阵设置为64×3优于32×6，多差分“单差分+S盒中间状态”区分器对应数据为384 bit，输入矩阵设置为32×12优于64×6。具体参数如表4所示。

表4 神经网络差分区器模型参数

超参数	值
Epoch	20 轮
Batch Size	5 000
Loss Function	MSE
Optimizer	Adam
Lr	0.000 1
Activation	卷积层: ReLU; 输出层: Sigmoid
输入矩阵格式 (Shape)	单差分区器: 64×3; 多差分“单差分+S盒中间状态”区分器: 32×12

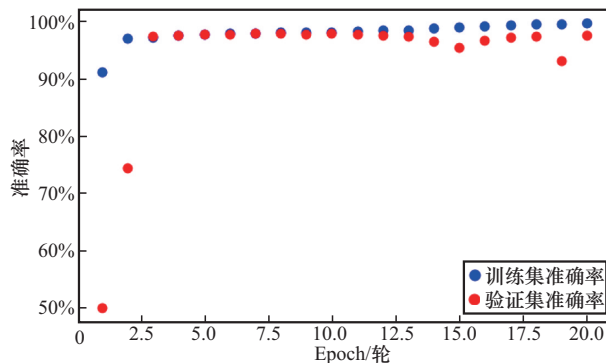


(a) 训练集和验证集损失率

3 模型训练

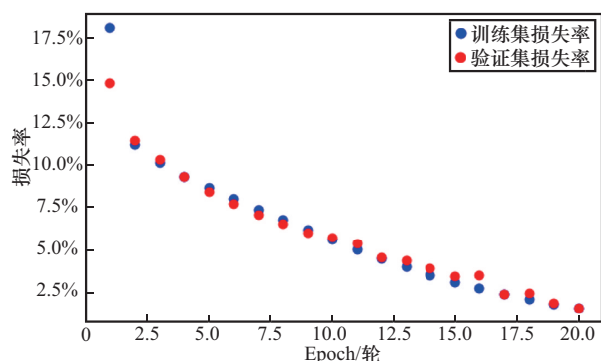
按照2.4节介绍的ResNet模型，选取 $\Delta x_1 = aa0a, aa00, 0000, aa00$ 和 $\Delta x_2 = aaa0, aa00, 0000, aa00$ 作为固定输入差分，采用2.3节介绍的数据集生成算法，分别加密生成5~7轮的 10^6 个训练样本数据集、 10^5 个验证样本数据集以此训练构造CRAFT算法5~7轮神经网络差分区器，单差分神经网络差分区器模型为ND₁、多差分神经网络差分区器模型为ND₂、“单差分+S盒中间状态”神经网络差分区器模型为ND₃。在相应的验证样本数据集上分别对ND₁、ND₂和ND₃3个模型进行评估，通过实验得到了5~7轮CRAFT算法神经网络差分区器的准确率 (Accuracy) 以及损失率 (Loss)。图4~图6分别代表5轮ND₁、ND₂、ND₃评估结果，图7~图9分别代表6轮ND₁、ND₂、ND₃评估结果，图10~图12分别代表7轮ND₁、ND₂、ND₃评估结果。

实验结果表明，本文设计的神经网络差分区器模型在CRAFT算法5轮加密后，3类差分区器的准确率分别达到了97.78%、99.83%和99.60%，模型判别准确率都较高；对于CRAFT算法6轮区分器，单差分区器准确率为61.39%、多差分区器准确率为72.78%、“单差分+S盒中间状态”区分器准确率为96.14%；对于CRAFT算法7轮区分器，3类差分区器对应的准确率略高于50%。通过研究分析，明文经过CRAFT算法5轮加密后，密文随机性较弱，3类神经网络差分区器模型都能较好地学习到密文信息特征，所以准确率都比较高、差距不大；经CRAFT算法6轮加密后，密文随机性扩大，准确

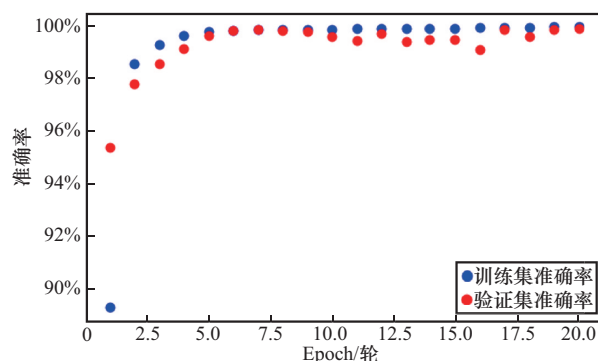


(b) 训练集和验证集准确率

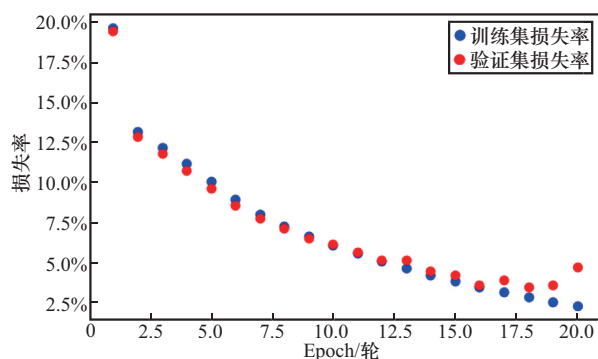
图4 5轮CRAFT算法ND₁评估结果



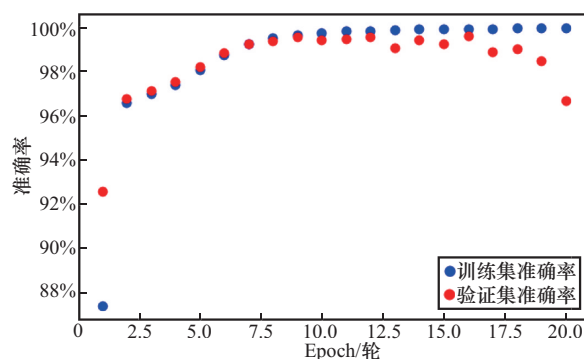
(a) 训练集和验证集损失率



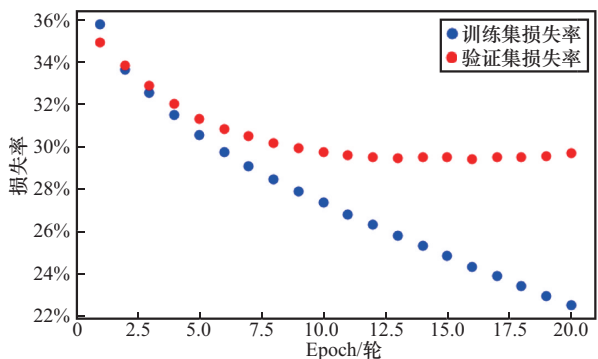
(b) 训练集和验证集准确率

图5 5轮CRAFT算法ND₂评估结果

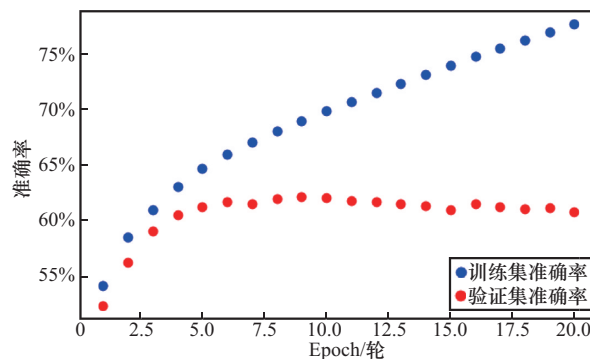
(a) 训练集和验证集损失率



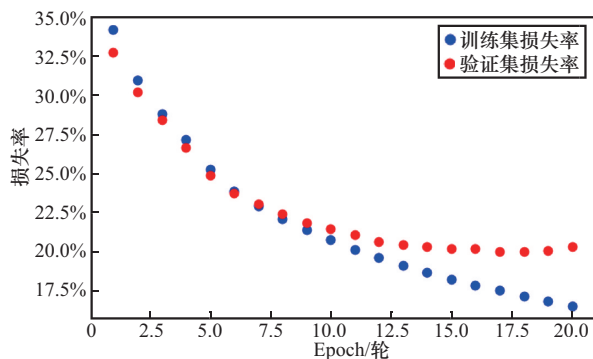
(b) 训练集和验证集准确率

图6 5轮CRAFT算法ND₃评估结果

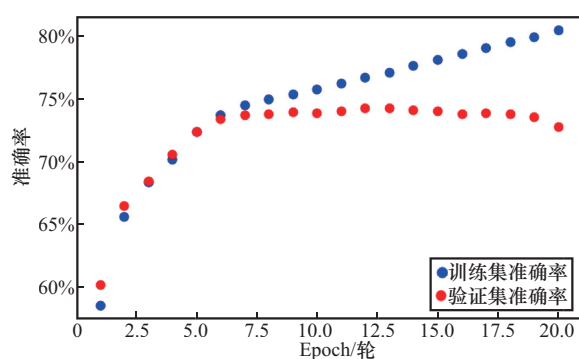
(a) 训练集和验证集损失率



(b) 训练集和验证集准确率

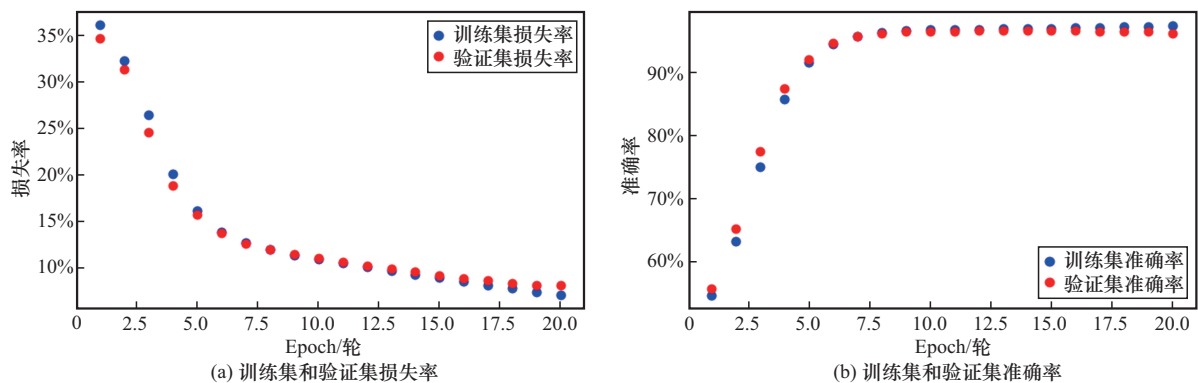
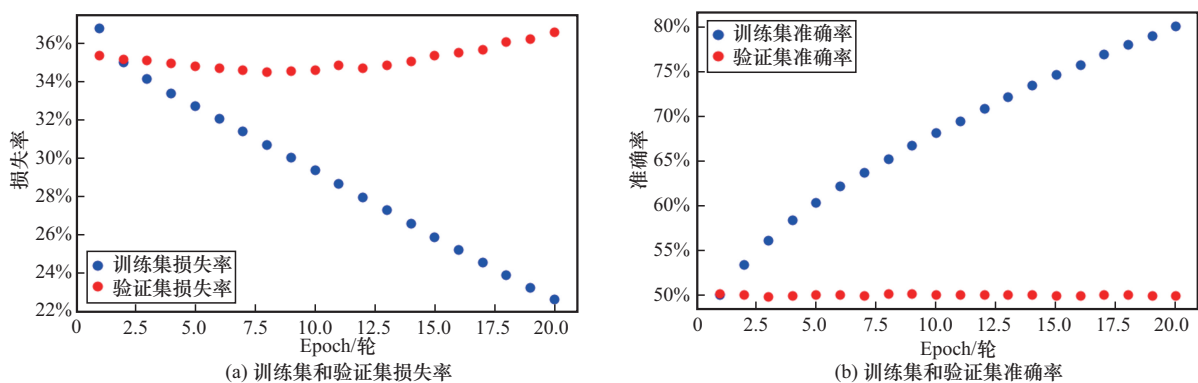
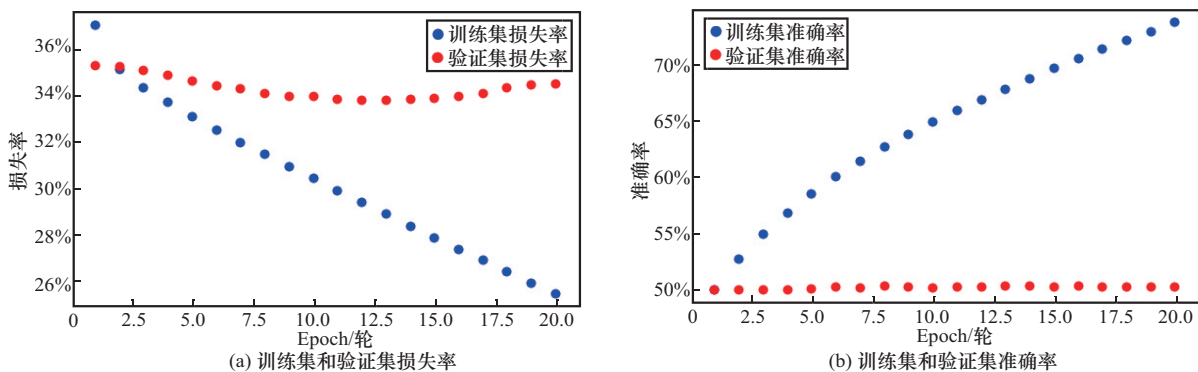
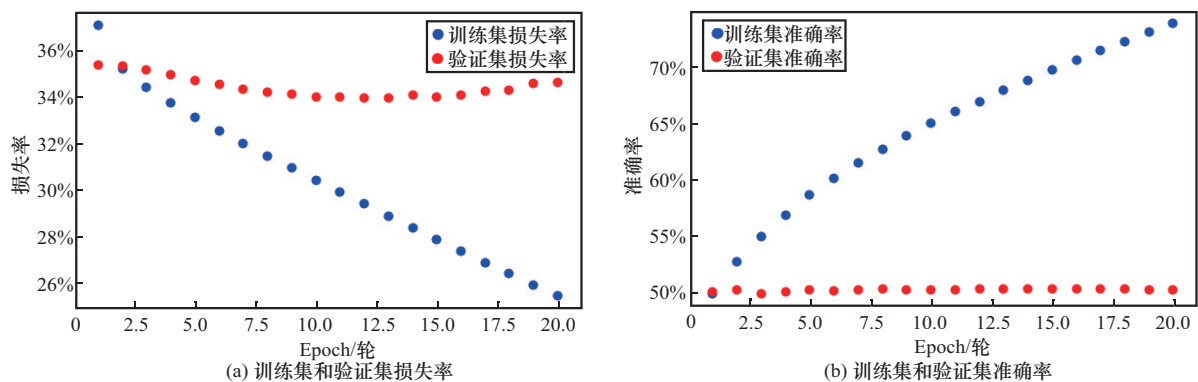
图7 6轮CRAFT算法ND₁评估结果

(a) 训练集和验证集损失率



(b) 训练集和验证集准确率

图8 6轮CRAFT算法ND₂评估结果

图9 6轮CRAFT算法ND₃评估结果图10 7轮CRAFT算法ND₁评估结果图11 7轮CRAFT算法ND₂评估结果图12 7轮CRAFT算法ND₃评估结果

率相比5轮的有所下降,对比3类差分区分器对应的数据集格式,多差分区分器数据样本比单差分区分器数据样本特征信息丰富,“单差分+S盒中间状态”区分器数据样本引入了S盒运算前的中间状态值,通过S盒的差分传播不确定性提取更多信息特征,进而获得更高的准确率,可见本文提出的“单差分+S盒中间状态”区分器准确率明显优于其他两种情形,可以更好地学习捕捉到数据特征,得到较高的预测精度;经CRAFT算法7轮加密后,密文随机性进一步扩大,密文对、密文差分、中间状态差分之间的特征都很难被学习到,故模型预测效果不佳。神经网络差分区分器训练数据复杂度,通常以破译所需明密文数计量,具体由神经网络模型训练样本与验证样本量影响决定,3类区分器样本数量一致,训练集为 10^6 个、验证集为 10^5 个。其中,单差分区分器格式为 $\{C_1, C_2, C_1 \oplus C_2; Y\}$,“单差分+S盒中间状态”区分器格式为 $\{C_1, C_2, C_1^{-s}, C_2^{-s}, C_1 \oplus C_2, C_1^{-s} \oplus C_2^{-s}; Y\}$,所需明密文数均为 $2 \times (10^6 + 10^5) \approx 2^{21.07}$;多差分区分器格式为 $\{C_1, C_2, C_3, C_1 \oplus C_2, C_1 \oplus C_3, C_2 \oplus C_3; Y\}$,所需明密文数为 $3 \times (10^6 + 10^5) \approx 2^{21.66}$ 。可见,“单差分+S盒中间状态”区分器相比多差分区分器而言,使用较少的数据量可以得到更高的准确率,表明“单差分+S盒中间状态”区分器在训练数据复杂度和准确率上具有优异性。

4 密钥恢复攻击

在上文基础上,进一步验证6轮“单差分+S盒中间状态”区分器对7轮CRAFT算法进行部分密钥恢复攻击的有效性。根据CRAFT算法密钥编排方案,搜索恢复第7轮加密对应的密钥,即轮密钥 TK_2 。

4.1 基本思路

对每个差分值为 $\Delta x_1 = aa0a, aa00, 0000, aa00$ 的明文对进行7轮加密,得到对应的CRAFT密文对,使用候选轮密钥对7轮CRAFT算法密文进行一轮解密,并经S盒逆运算获取中间状态值,拼接成“单差分+S盒中间状态”区分器对应数据样本格式,然后利用6轮“单差分+S盒中

间状态”区分器 ND_3 对每个数据样本进行预测,将预测的结果作为候选密钥的分值,最后按照分值大小对密钥进行降序排列。实验进行100次密钥恢复攻击,具体步骤如下。

步骤1 生成密文对。随机选择 $n = 10\,000$ 组明文对 (P_1, P_2) , $P_2 = \Delta x_1 \oplus P_1$,然后选定一个主密钥 K ,依次对明文进行7轮加密,得到密文对 $\{C_1, C_2\}$ 。

步骤2 选择猜测密钥集。每次实验使用不同的主密钥 K ,按照步骤1的方式进行加密。同时每次实验选择100个候选轮密钥 $k = \{k_1, k_2, \dots, k_{100}\}$,作为猜测密钥集合(考虑到待恢复轮密钥为64 bit,若是全空间遍历,计算复杂度为 2^{64} ,远超计算机性能,故缩小密钥空间,以验证区分器攻击有效性为主),集合中包含1个真实第7轮轮密钥 k_{th} ,根据CRAFT算法轮密钥编排规则,通过主密钥 K ,可计算出第7轮轮密钥 $k_{th} = TK_2$ 。

步骤3 解密一轮。依次使用猜测密钥集合中候选密钥对密文对 $\{C_1, C_2\}$ 解密一轮,得到 $\{C'_1, C'_2\}$,再经S盒逆运算,得到中间状态值 $\{C_1'^{-s}, C_2'^{-s}\}$ 。

步骤4 构造数据样本。按照“单差分+S盒中间状态”区分器对应数据格式,拼接数据样本 $C_i^k = \{C'_1, C'_2; C_1'^{-s}, C_2'^{-s}; C'_1 \oplus C'_2, C_1'^{-s} \oplus C_2'^{-s}\}$ 。

步骤5 计算候选密钥得分。利用6轮“单差分+S盒中间状态”区分器 ND_3 对每个数据样本 C_i^k 进行预测,得到相应分值 Z_i^k ,将猜测密钥集合中每个候选密钥 k_i 对应的所有 Z_i^k 通过式(6)计算得到该密钥的最终分值 V_k 。

$$V_k = \text{lb} \left(\frac{Z_i^k}{(1 - Z_i^k)} \right) \quad (6)$$

步骤6 保存高分候选密钥。将猜测密钥集合中所有候选密钥 k_i 按其得分值 V_k 进行降序排列,将分值最高的候选密钥与真实密钥进行比对,两者若相等,则判定为一次成功的有效攻击。

综上,密钥恢复攻击算法如算法2所示。

算法2 密钥恢复攻击算法

输入 数据样本数量 n ,固定输入差分 Δx_1 ,主密钥 K ,CRAFT算法加密操作 Enc_{7r} ,解密一轮操作 Dec_{1r} ,S盒逆运算 S^{-1} ,6轮“单差分+S

盒中间状态”区分器 ND_3

输出 候选密钥降序列表 L_k , 分值最高密钥 $k_{\max}$

(1) 生成明文对

for i in range(n):
 $P_i \leftarrow \text{data_random_64}()$
 $(P_1, P_2) \leftarrow (P_i, P_i \oplus \Delta x_1)$

(2) 选择猜测密钥集

for i in range(m):
 $K_i \leftarrow \text{data_random_128}()$
 $k_{th} \leftarrow TK_2 \leftarrow K_i$
 $k \leftarrow \{k_1, \dots, k_{th}, \dots, k_{100}\}$

(3) 生成密文对

$(C_1, C_2) \leftarrow (\text{Enc}_{7r}(P_1), \text{Enc}_{7r}(P_2))$

(4) 解密一轮

$\{C'_1, C'_2\} \leftarrow (\text{Dec}_{1r}(C_1), \text{Dec}_{1r}(C_2))$
 $\{C'^{-s}_1, C'^{-s}_2\} \leftarrow (S^{-1}(C'_1), S^{-1}(C'_2))$

(5) 构造“单差分+S盒中间状态”区分器样本

$X \leftarrow \{C'_1, C'_2; C'^{-s}_1, C'^{-s}_2; C'_1 \oplus C'_2, C'^{-s}_1 \oplus C'^{-s}_2\}$

(6) 计算候选密钥得分

for i in range(n):
 $Z_i^k \leftarrow ND_3(X_i)$
 $V_k = \text{lb}\left(\frac{Z_i^k}{(1 - Z_i^k)}\right)$

(7) 输出高分候选密钥

$L_k \leftarrow \text{append}(k, V_k)$
 return($k_{\max}$)

4.2 实验结果

实验过程中, 采用的硬件设备型号为华为 MateBook 14 笔记本电脑, 处理器为 AMD Ryzen 5 3500U with Radeon Vega Mobile GFX、操作系统为 Windows 10 家庭专业版、内存为 16 GB, 编程语言为 Python3.7, 编译工具为 Pycharm 2019。在 7 轮攻击实验中, 每次实验明密对数量为 10 000 个、候选轮密钥个数为 100 个, 共进行了 100 次实验, 恢复单个密钥平均用时 445 s。其中, 93 次实验中分值最高的候选轮密钥与真实轮密钥一致, 另外 7 次实验正确候选轮密钥得分值排名均为第 2, 密钥恢复成功率为 93%。

在差分分析中, 时间复杂度通常以加密次数衡量, 一次完整的密码分析包括区分器训练构造

和密钥恢复两个阶段。训练阶段明密文数为 $2 \times (10^6 + 10^5) \approx 2^{21.07}$, 训练区分器需加密 $2^{21.07}$ 次; 攻击阶段利用密钥恢复攻击算法, 每次对 7 轮 CRAFT 算法的 100 个猜测密钥进行筛选, 使用了 10 000 个选择明密文对, 密钥恢复需加密 $2 \times 10\,000 \times 100 = 2 \times 10^6$ 次 (7 轮加密), 故攻击时间复杂度为 $2^{21.07} + 2 \times 10\,000 \times 100 \approx 2^{22.01}$ 。

5 结束语

本文基于残差神经网络对 CRAFT 算法神经网络差分区器构造及其应用进行了深入研究, 神经网络模型、固定输入差分和数据集格式是决定神经网络差分区器性能的关键因素, 本文重点研究了 3 类不同数据集格式对神经网络差分区器的影响, 结合单差分和多差分区器数据集特点, 针对 CRAFT 算法结构设计提出了“单差分+S盒中间状态”区分器, 通过训练分别获得了 3 类数据集格式对应的 5~7 轮神经网络差分区器, 尤其在 6 轮神经网络差分区器中, 准确率明显高于单差分和多差分区器。实验结果证明, 7 轮加密后的密文侧解密一轮, 并经 S 盒逆运算推出中间状态值, 将构造好的数据样本经 6 轮“单差分+S盒中间状态”区分器预测后, 对 7 轮 CRAFT 算法可有效进行部分密钥恢复攻击。本文研究结果对于基于深度学习的神经网络差分区器分析探索具有重要意义, 可拓展应用于 SPN 结构的 SKINNY、PRESENT 等算法, 验证 S 盒中间状态值是否含有更多密文差异信息, 起到提升神经网络差分区器准确率的作用, 进而开展有效的密钥恢复攻击。后续, 可以从引入新的深度学习模型或者优化数据处理方式等多个维度, 进一步研究探索提高神经网络差分区器性能的方法路径。

参考文献:

- [1] 钟悦, 谷杰铭, 曹洪林. 轻量级分组密码算法综述[J]. 计算机科学, 2023, 50(9): 3-15.
- Zhong Y, Gu J M, Cao H L. Survey of lightweight block cipher[J]. Computer Science, 2023, 50(9): 3-15.
- [2] Beierle C, Jean J, Kölbl S, et al. The SKINNY family of block ciphers and its low-latency variant MANTIS[C]//Advances in Cryptology-CRYPTO 2016. Berlin: Springer, 2016: 123-153.
- [3] Beierle C, Leander G, Moradi A, et al. CRAFT: lightweight tweak-

- able block cipher with efficient protection against DFA attacks[J]. IACR Transactions on Symmetric Cryptology, 2019: 5-45.
- [4] Beaulieu R, Shors D, Smith J, et al. The SIMON and SPECK families of lightweight block ciphers[J]. IACR Cryptology EPrint Archive, 2013, 2013: 404.
- [5] 沈璇, 刘国强, 孙兵, 等. 两类动态密码结构抵抗不可能差分和相关线性能力评估[J]. 电子学报, 2024, 52(3): 709-718.
- Shen X, Liu G Q, Sun B, et al. Security evaluation against impossible differential cryptanalysis and zero correlation linear cryptanalysis for two dynamic cryptographic structures[J]. Acta Electronica Sinica, 2024, 52(3): 709-718.
- [6] 成磊, 沈璇, 任传伦. 广义类CLEFIA动态密码结构抵抗差分和线性密码分析的安全性评估[J]. 电子学报, 2024, 52(8): 2571-2580.
- Cheng L, Shen X, Ren C L. Security evaluation of generalized CLEFIA-like dynamic cipher structures against differential and linear cryptanalysis[J]. Acta Electronica Sinica, 2024, 52(8): 2571-2580.
- [7] Gohr A. Improving attacks on Speck32/64 using deep learning[J]. IACR Cryptology EPrint Archive, 2019, 2019: 37.
- [8] Chen Y, Shen Y T, Yu H B, et al. A new neural distinguisher considering features derived from multiple ciphertext pairs[J]. The Computer Journal, 2023, 66(6): 1419-1433.
- [9] 杨小雪, 陈杰. 几类密码算法的神经网络差分区分器的改进[J]. 西安电子科技大学学报(自然科学版), 2024, 51(1): 210-222.
- Yang X X, Chen J. Improvement of the neural distinguishers of several ciphers[J]. Journal of Xidian University (Natural Science), 2024, 51(1): 210-222.
- [10] 段明, 付超辉, 魏强, 等. 基于深度学习的多差分神经网络区分器及其应用[J]. 信息工程大学学报, 2021, 22(3): 347-350.
- Duan M, Fu C H, Wei Q, et al. Multi differential attack based on deep learning and its application[J]. Journal of Information Engineering University, 2021, 22(3): 347-350.
- [11] 蒋若怡, 韦永壮, 王慧娇. 基于深度学习的差分神经区分器求解方法[J]. 计算机工程与设计, 2023, 44(6): 1629-1634.
- Jiang R Y, Wei Y Z, Wang H J. Method for solving differential neural distinguishers based on deep learning[J]. Computer Engineering and Design, 2023, 44(6): 1629-1634.
- [12] 王慧娇, 张鑫, 韦永壮, 等. 基于深度学习的SM4密码算法新型区分器[J]. 通信学报, 2023, 44(7): 171-184.
- Wang H J, Zhang X, Wei Y Z, et al. Novel distinguisher for SM4 cipher algorithm based on deep learning[J]. Journal on Communications, 2023, 44(7): 171-184.
- [13] 侯泽洲, 陈少真, 任炯炯. 深度学习在分组密码差分区分器上的研究应用[J]. 软件学报, 2022, 33(5): 1893-1906.
- Hou Z Z, Chen S Z, Ren J J. Research and application of deep learning on differential distinguisher of block cipher[J]. Journal of Software, 2022, 33(5): 1893-1906.
- [14] 栗琳轲, 陈杰, 刘君. 神经网络差分区分器的改进方案与应用[J]. 西安电子科技大学学报(自然科学版), 2025, 52(1): 196-214.
- Li L K, Chen J, Liu J. Improved schemes and applications of the neural network differential distinguisher[J]. Journal of Xidian University (Natural Science), 2025, 52(1): 196-214.
- [15] Hadipour H, Sadeghi S, Niknam M M, et al. Comprehensive security analysis of CRAFT[J]. IACR Transactions on Symmetric Cryptology, 2020: 290-317.
- [16] Liu H M, Zhang W Y, Zhang J J, et al. Clustering of differentials in CRAFT with correlation matrices[J]. International Journal of Intelligent Systems, 2022, 37(12): 12113-12134.
- [17] 蒋若怡. 分组密码的差分类自动化分析方法研究[D]. 桂林: 桂林电子科技大学, 2022.
- Jiang R Y. Research on automatic analysis methods of differential classification in block ciphers[D]. Guilin: Guilin University of Electronic Technology, 2022.
- [18] Ebrahimi M A, Ahmadian Z. New automatic search method for truncated-differential characteristics application to midori, SKINNY and CRAFT[J]. The Computer Journal, 2020, 63(12): 1813-1825.
- [19] Hadipour H, Bagheri N, Song L. Improved rectangle attacks on SKINNY and CRAFT[J]. IACR Transactions on Symmetric Cryptology, 2021: 140-198.
- [20] Hadipour H, Sadeghi S, Eichlseder M. Finding the impossible: automated search for full impossible-differential, zero-correlation, and integral attacks[C]//Advances in Cryptology-EUROCRYPT 2023. Berlin: Springer, 2023: 128-157.
- [21] 李超, 孙兵, 李瑞林. 分组密码的攻击方法与实例分析[M]. 北京: 科学出版社, 2010.
- LI C, SUN B, LI R L. Attack methods and case analysis of block ciphers[M]. Beijing: Science Press, 2010.
- [22] 黄锦坤, 黄远航, 黄文敏, 等. 基于跨流注意力增强中心差分卷积网络的CG图像检测[J]. 网络与信息安全学报, 2024, 10(6): 96-108.
- Huang J K, Huang Y H, Huang W M, et al. Cross-stream attention enhanced central difference convolutional network for CG image detection[J]. Chinese Journal of Network and Information Security, 2024, 10(6): 96-108.
- [23] 刘子翔, 莫秀良, 宫良一, 等. 基于时空特征融合深度学习的NXN-SAttack流量识别方法[J]. 网络与信息安全学报, 2025, 11(4): 173-188.
- Liu Y X, Mo X L, Gong L Y, et al. NXNAttack traffic detection method based on spatio-temporal feature fusion deep learning[J]. Chinese Journal of Network and Information Security, 2025, 11(4): 173-188.
- [24] 夏锐琪, 李曼曼, 陈少真. 基于机器学习的分组密码结构识别[J]. 网络与信息安全学报, 2023, 9(3): 79-89.
- Xia R Q, Li M M, Chen S Z. Identification on the structures of block ciphers using machine learning[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 79-89.
- [25] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [26] Benamira A, Gerauld D, Peyrin T, et al. A deeper look at machine learning-based cryptanalysis[C]//Advances in Cryptology-EUROCRYPT 2021. Berlin: Springer, 2021: 805-835.
- [27] Bao Z Z, Lu J Y, Yao Y R, et al. More insight on deep learning-aided cryptanalysis[C]//Advances in Cryptology-ASIACRYPT 2023. Berlin: Springer, 2023: 436-467.

[作者简介]



沈璇 (1990-), 男, 博士, 信息支援部队工程大学副教授, 主要研究方向为对称密码的设计与分析。



代威 (1991-), 男, 硕士, 信息支援部队工程大学讲师, 主要研究方向为网络安全防护。



沈霞民 (1993-), 男, 国防科技大学硕士生, 主要研究方向为对称密码的设计与分析。



黄金杰 (1990-), 男, 信息支援部队工程大学讲师, 主要研究方向为信息安全。



卜予彤 (2001-), 女, 国防科技大学硕士生, 主要研究方向为对称密码的设计与分析。



郭连城 (1980-), 男, 信息支援部队工程大学副教授, 主要研究方向为软件漏洞分析、软件加解密。