基于 Transformer 模型与注意力机制的差分密码分析

肖超恩，李子凡，张磊，王建新，钱思源
（北京电子科技学院电子与通信工程系，北京 100071）

摘　要： 基于差分分析的密码攻击中，通常使用贝叶斯优化方法验证部分解密的数据是否具有差分特性。目前，主要采用基于深度学习的方式训练 1 个差分区分器，但随着加密轮数的增加，差分特征的精确度会呈现线性降低的趋势。为此，结合注意力机制和侧信道分析，提出了一种新的差分特性判别方法。根据多轮密文间的差分关系，基于 Transformer 训练了 1 个针对 SPECK32/64 算法的差分区分器。在密钥恢复攻击中，借助前一轮的密文对待区分密文影响最大特性，设计了新的密钥恢复攻击方案。在 SPECK32/64 算法的密钥恢复攻击中，采用 2^6 个选择明密文对，并借助第 20 轮密文对将第 22 轮 65 536 个候选密钥范围缩小至 17 个以内，完成对最后两轮子密钥的恢复攻击。实验结果表明，该方法的攻击成功率达 90%，可以有效应对加密轮数增多造成的密文差分特征难以识别的问题。

关键词： Transformer 模型；注意力机制；差分区分器；SPECK32/64 算法；密钥恢复攻击
中图分类号： TP391　　**文献标志码：** A　　**DOI：** 10.19678/j.issn.1000-3428.0068486

Differential Cryptanalysis Based on Transformer Model and Attention Mechanism

XIAO Chaoen, LI Zifan, ZHANG Lei, WANG Jianxin, QIAN Siyuan
(Department of Electronics and Communications Engineering, Beijing Electronic Science and Technology Institute, Beijing 100071, China)

【Abstract】 In differential analysis-based cryptographic attacks, Bayesian optimization is typically used to verify whether the partially decrypted data exhibit differential characteristics. Currently, the primary approach involves training a differential distinguisher using deep learning techniques. However, this method has a notable limitation in that, as the number of encryption rounds increases, the accuracy of the differential characteristics decreases linearly. Therefore, a new differential characteristic discrimination method is proposed based on the attention mechanism and side-channel analysis. Using the difference relationship between multiple rounds of the ciphertext, a difference partition for the SPECK32/64 algorithm is trained based on the transformer. In a key recovery attack, a novel scheme is designed based on the previous ciphertext treatment to distinguish the most influential features of the ciphertext. In the key recovery attack of the SPECK32/64 algorithm, 2^6 selected ciphertext pairs are used. Using the 20th round ciphertext pairs, the 65 536 candidate keys of the 22nd round can be screened within 17 on average, and the key recovery attack of the last two wheels can be completed. The experimental results show that this method achieves a success rate of 90%, effectively addressing the challenge of recognizing ciphertext differential features caused by an increase in the number of encryption rounds.

【Key words】 Transformer model; attention mechanism; differential distinguisher; SPECK32/64 algorithm; key recovery attack

0　引言

差分密码分析最初由 BIHAM 和 SHAMIR 提出[1]，用于研究 DES 算法[2]的安全性。该方法的核心思想是分析密码算法的输入差异和输出差异之间的关系，以尽可能地寻找高概率的差分特征。通过识别这些差分特征，攻击者可以将密码算法与随机置换等特性区分开来，从而获取有关密钥的信息。攻击者通过猜测密钥来验证部分解密后的差分值是否满足差分区分器的输出。攻击者的目标是找到一组明文对（输入差异），通过猜测密钥对这些明文进行加密后，得到的密文对（输出差异）满足差分区分器的要求。2011 年，BLONDEAU 等[3]提出了多差分攻击方法，利用多个输入差分和输出差分的差分特性进行密码分析，但都存在占用资源过多、复杂度较高的问题，而深度学习在这些方面都具有明显优势。

深度学习是一种机器学习技术，通过构建和训练多层神经网络来学习和提取数据的特征，并模拟人类思维进行开发。这种技术在计算机视

收稿日期： 2023-09-28　　**修回日期：** 2024-01-16
基金项目： 中央高校基本科研业务费资金资助（3282024009）。
通信作者 E-mail： xce@besti.edu.cn

觉[4]、自然语言处理[5]等领域已经取得了巨大的成功。GOHR[6]创造性地将深度学习与差分密码分析相结合，针对 SPECK32/64[7]提出了基于深度残差神经网络的差分区分器，结合贝叶斯优化搜索方法，进一步提出了基于深度学习的密钥恢复攻击，并成功恢复出第 11 轮的子密钥，说明使用深度学习进行密钥恢复攻击的复杂度要远低于传统密钥恢复攻击的复杂度。BENAMIRA 等[8]对 GOHR[6]提出的神经网络进行分析，发表了对神经区分器内部机理的研究报告，他们发现神经区分器的区分是基于倒数第二轮和倒数前二轮的密文对差异和内部状态差异。同期，宿恒川等[9]运用 PU 学习方法对减轮 SPECK 算法展开了研究，并建立了包括 5 轮和 6 轮的区分器，其准确度明显超过了 GOHR[6]构建的对应模型。CHEN 等[10]在文献[8]的基础上进一步引入了扩展差分线性连接表(EDCLT)，并利用该表构建了多种基于深度学习的差分区分器，最终使用 EDCLT 成功解释了与神经区分器相关的现象。SU 等[11]基于多面体差分与神经网络，针对 Simon32/64 算法[7]设计了准确率达 92%的神经网络区分器。付超辉等[12]将多面体差分应用在 Simeck32/64 算法[13]，得到了准确率高达 96.7%的区分器。SO 等[14]设计了基于深度学习的多密码分析模型，通过攻击简化版多种轻量级分组密码，验证了基于深度学习的密码分析可行性。YADAV 等[15]研究不同密码算法的结构特性，设计一种基于 Feistel、SPN 和 ARX 结构的通用机器学习差分区分器。BAO 等[16]设计了实用的 13 轮和改进的 12 轮基于神经网络差分区分器的密钥恢复攻击，深入探索了更广义的差分中性比特位，并且使用 DenseNet 和 SENet 得到 11 轮 Simon32/64 的神经网络差分区分器，从而实现了 16 轮密钥恢复攻击。BAKSI 等[17]将区分器的构建问题转变为神经网络的分类问题，使神经网络能够有效管理数据，并成功将其应用于流密码，扩大了深度学习技术在密码学领域的适用范围。JAIN 等[18]对 BAKSI 等[17]所提的模型进行了应用。杨小雪等[19]针对 Speckey 和 LAX32 密码算法，分析了线性运算模块对神经网络区分器的影响，并研究了输入数据信息对 Simon32/64 神经网络区分器造成的影响。陈怡等[20]针对减轮 SPECK 大状态分组密码，提出了深度学习辅助密钥恢复框架。

目前，这些主流方法是借助 ResNet 的特征提取优势，将训练过程看作黑盒，研究输入与输出之间的差分特征，达到判断密文是否具有差分特性的目的。但是随着加密轮数的增加，密钥和明密文之间的关系更加复杂，差分扩散更加强烈，使得差分特征难以区分。GOHR[6]提出的区分器分辨准确率直线下降，无法对更高轮数进行密钥恢复攻击。因此，基于特征提取方式的深度学习技术很难更深层次地发现差分特性，加密过程是通过轮函数进行一轮一轮的计算，而差分特性不仅与初始的输入有关，还与相关的各轮计算有关，即差分特性具有较强的上下文特征，而基于 ResNet 的特征提取网络无法考虑此特性。与其他神经网络相比，VASWANI 等[21]提出以 Self-Attention 为基本单元的 Transformer 模型，使注意力机制得到真正的运用，该模型与传统卷积神经网络(CNN)、ResNet 模型结构不同，其特有的注意力机制可兼顾上下文特征与局部细微特征，即具有出色的序列细粒度特征提取能力，又避免了文本上下文语义特征[22]的缺失。孙晓丽等[23]提出一种基于 seq2seq 模型的密码破译方法，将明密文之间的映射关系看作机器翻译问题，在满足模型输入格式的同时保留序列之间的相关性。

本文提出基于 Transformer 的密钥恢复框架，利用 Transformer 模型的注意力机制，研究密文结构中间状态的差异，设计了基于 Transformer 的分组密码差分区分器，此差分区分器不会随着加密轮数的增加影响模型评估效果。与其他方法相比，本文所提方法不需要差分中性比特位，不会随着加密轮数的增加影响密钥恢复攻击的成功率，使用前一轮的密文对来区分当前轮次密文对是否带有差分特性，以此推断该轮正确的候选密钥。

1 密钥恢复攻击方法

1.1 传统密钥恢复方法

在传统差分密码分析中，如果找到 1 条概率大于 2^{-n} 的 $n-1$ 轮差分特征，就能利用这条差分特征进行密钥恢复攻击，可恢复 n 轮密码算法的第 n 轮密钥。具体密钥恢复攻击流程如下：

1)寻找 1 条在 $n-1$ 轮密码算法中具有高概率的差分特征，其概率为 q。

2) 计算此差分特征密文对的输出差分，确定恢复的第 n 轮密钥的 l 个 bit。

3) 为第 n 轮所有候选密钥设置计数器，共有 2^l 个，并初始化为 0。

4) 均匀随机选取 m 个明文 P_0，计算 $P_1=P_0 \oplus d_{\text{diff}}$，其中 d_{diff} 是差分特征的输入差分。

5) 加密明文对 P_0 和 P_1，针对每个候选密钥，分别将密文对解密一轮，验证伪密文对的输出差分是否等于差分特征的输出差分，相等则对应计数器加 1。

6) 取计数器值最大的候选密钥为猜测密钥，返回结果。

1.2 基于深度学习的密钥恢复方法

2019 年，GOHR[6]首次提出了基于深度学习的密钥恢复攻击，采用 2 种新技术 UCB(Upper Confidence Bounds)[24]算法和贝叶斯优化，提出了加速的密钥恢复攻击，并应用于 11 轮、12 轮 SPECK32/64 算法。为了恢复 11 轮 SPECK32/64 子密钥，GOHR[6]在神经网络差分区分器前加了 1 个通过概率为 2^{-6} 的 2 轮差分(0x211,0xa04)→(0x40,0x0)。具体密钥恢复攻击流程如下：

1) 随机生成 1 个明密文对 P_0 和 P_1，$P_1=P_0\oplus d_{\text{diff}}$，$d_{\text{diff}}=$(0x0211,0x0a04)。

2) 使用该伪明文对 P_0、P_1 配合 k 个中性比特位，即明文中不影响差分传播的比特值，生成 1 个伪明文结构，在加密过程中先用 0 作为轮密钥解密，无损失扩展一轮，再加密 11 轮，得到所需的密文结构。

3) 使用第 11 轮的候选轮密钥 k_{11} 对步骤 2)中得到的密文结构进行一轮解密尝试，将解密后的伪密文结构输入到 7 轮神经网络差分区分器，得到每个候选密钥对应输出 S_i，$i\in[1,2^k]$。每个候选密钥得分的计算式如下：

$$V=\sum_{i=1}^{2^k}\text{lb}\left(\frac{S_i}{1-S_i}\right) \tag{1}$$

若 V 超过设定阈值 C_1，则遍历第 10 轮轮密钥所有候选密钥 k_{10}，用 k_{10} 对步骤 3)得到的伪密文结构继续进行一轮解密尝试，解密的伪密文对输入到 6 轮神经网络差分区分器，得到每个候选密钥对应输出并用式(1)计算 k_{10} 得分，当得分超过设定阈值 C_2 时，将(k_{11},k_{10})作为猜测密钥。

重复上述过程直到得到 1 个猜测密钥。

2 改进 SPECK32/64 密钥恢复攻击方案与模型构造

2.1 基于 Transformer 的区分器模型

本文基于 Transformer 网络结构，设计了针对 SPECK32/64 算法的差分区分器模型，该模型由 Transformer 编码模块和序列输出模块组成，网络结构如图 1 所示。

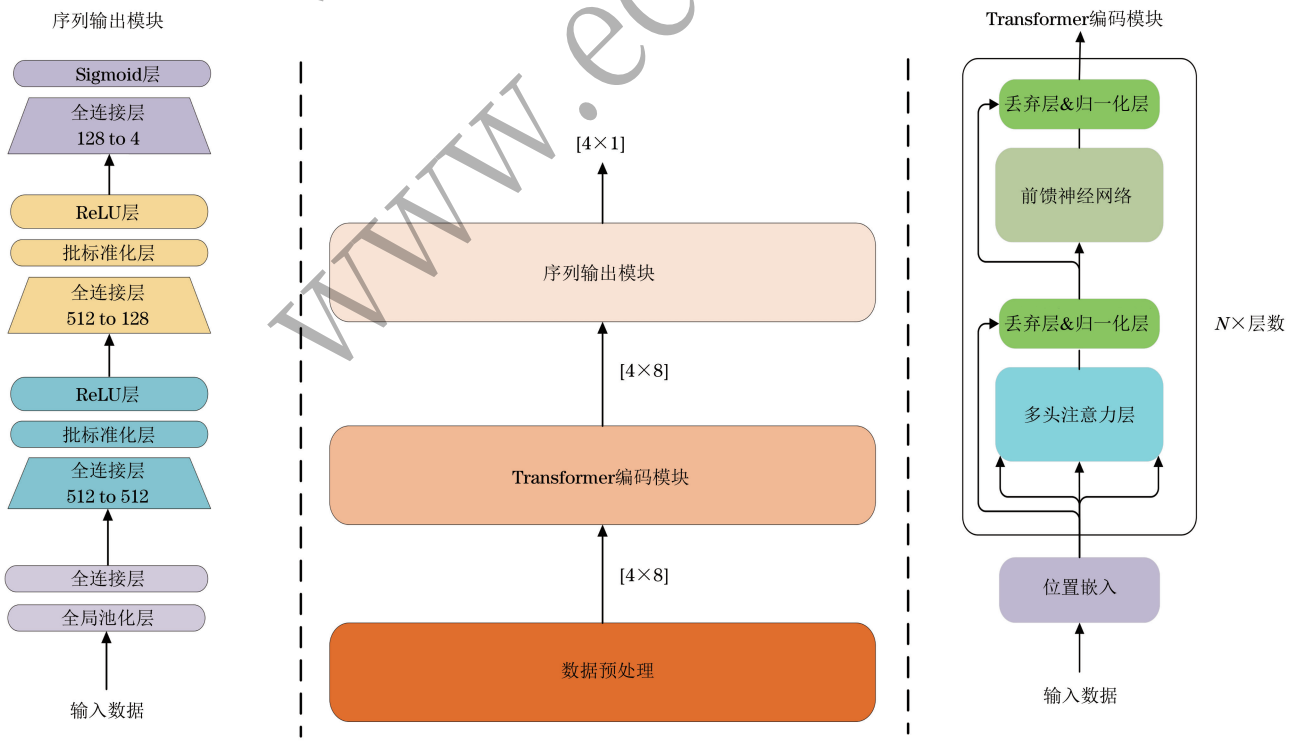

图 1 Transformer 差分区分器模型

Fig. 1 Transformer differential discriminator model

1)Transformer 编码模块。

编码部分包括输入层、位置嵌入层以及编码层。在输入层，输入数据包括加密的轮数和每一轮加密的密文对。之后，数据被传递至位置嵌入层，以便该层能够解析并理解密文对的顺序信息，最后输入到编码层。编码层由 1 个多头注意力机制、1 个前馈

神经网络、2 个归一化层和 2 个丢弃层组成，其中多头注意力层的注意力头数与加密轮数相对应，能够提取数据的内部特征。

2)序列输出模块。

序列输出模块包括 1 个全局池化层、2 个全连接网络和 1 个输出单元，其中，2 个全连接网络和 1 个输出单元都属于感知机结构。感知机是基础的线性二分类模型，即输出为 2 个状态，由 2 层神经元组成，对编码模块输出的结果进行序列化。全局池化得到的数据格式为$[C_{\text{ciphertext rounds}}\times 512]$的张量，处理后显著提高模型的泛化能力，输入到第 1 层全连接网络(512 to 512)，再输入到第 2 层全连接网络(512 to 128)。每层全连接层都经过批量归一化和 ReLU 激活函数。最后得到的张量被 1 个输入为 128，输出为$[C_{\text{ciphertext rounds}}\times 1]$的感知机接收并进行预测。输出经过 Sigmoid 激活函数之后得到(0,1)区间内的分数，将分数大于 0.5 的认定为真实密文对，小于等于 0.5 的认定为随机数据。

2.2 改进 SPECK32/64 密钥恢复攻击方案

在第 1.2 节介绍的攻击方案中，当所有密文全部通过 2 轮差分并且第 11 轮子密钥猜测正确时，密钥排名分数的差别会很大，通过这种方式使得密钥恢复攻击。但随着加密轮数的增加，神经网络差分区分器的区分准确率直线下降，严重影响了攻击成功率。为了有效攻击更高轮数的密文结构，本文改进了密钥恢复攻击方案，根据 Transformer 中注意力机制能够兼顾上下轮密文特征的优势，通过提取中间密文结构进行子密钥恢复。候选密钥筛选流程如下。

1)选择明密文对。均匀随机选择明文对 P_1，$P_2,\cdots,P_i$，确定输入差分 $d_{\text{diff}}=(0\text{x}0040/0\text{x}0000)$，令 $P'_i=P_i\oplus d_{\text{diff}}$，在同一密钥 k 下加密，获得第 $n-2$ 轮和第 n 轮相应密文对 $C_{i(n-2)}$、$C'_{i(n-2)}$、C_{in}、C'_{in}。

2)对第 n 轮密文对解密。由于每轮子密钥为 16 bit，因此候选密钥的数量为 2^{16}。利用候选密钥对第 n 轮密文对 C_{in}、C'_{in} 进行一轮解密，得到其 $n-1$ 轮密文对 $C_{i(n-1)}$、$C'_{i(n-1)}$。

3)将第 $n-2$ 轮密文对 $C_{i(n-2)}$、$C'_{i(n-2)}$ 与解密后的 $n-1$ 轮密文对 $C_{i(n-1)}$、$C'_{i(n-1)}$ 一起输入到 6 轮 Transformer 神经网络区分器进行评估(其他位置用 0 进行填充)，获得部分解密密文对的输出 $S_{i,k}$。

4)计算每个候选密钥解密后的得分：

$$V_k=\sum_{k=1}^{i}\ln\frac{S_{i,k}}{1-S_{i,k}} \tag{2}$$

对 V_k 降序排列，得到所有候选密钥的排名。基于上述候选密钥筛选过程，进行第 n、$n-1$ 轮 SPECK32/64 密钥恢复攻击，生成差分 $d_{\text{diff}}=(0\text{x}0040,0\text{x}0000)$ 的 N_{cts} 个随机明文对 P 和 P'。将明文对进行 n 轮加密，通过 UCB 算法[24]选择 1 组密文对，通过贝叶斯优化选择 N_{cand1} 个候选密钥 k_n 对第 n 轮密文进行 1 轮解密，将解密后的伪密文和第 $n-1$ 轮密文用 0 填充后送入 Transformer 神经网络差分区分器，获得部分解密密文对的输出 $S_{i,k}$，重复此操作 5 次。

使用式(2)进行候选密钥排名。当 V_k 超过阈值 C_1，遍历第 $n-1$ 轮 N_{cand2} 个候选密钥 k_{n-1}，重复步骤 3)和步骤 4)，当得分超过另一个阈值 C_2，得到(k_n,k_{n-1})作为最佳候选密钥。设置最大迭代次数 N_{it}，重复上述操作直到得到 1 个猜测密钥。

2.3 数据集构造

本文模型的关键是构造 1 个序列到序列的数据集，使得基于 Transformer 的差分区分器能识别给定密文对是否具有差分特征。本文构造的输入序列为 $X=[x_1,x_2,\cdots,x_i]$，其中 $x_i=\{x_i^1,x_i^2\}$。输出序列为 $Y=[y_1,y_2,\cdots,y_i]$，其中 $y_i\in\{0,1\}$。在输入输出序列中 x_i 与 y_i 一一对应。当 $y_i=1$ 时，表示输入序列 x_i 中 x_i^1 与 x_i^2 为一对具有差分特征的密文。当 $y_i=0$ 时，表示输入序列 x_i 中 x_i^1 与 x_i^2 为一对不具有差分特征的随机数。

假定输入差分记为 diff，差分区分器识别的密文轮数记为 i，则模型训练的数据集构造流程如图 2 所示(彩色效果见《计算机工程》官网 HTML 版，下同)。

1)随机生成包含元素个数为 n 的集合 P，并将集合 P 中元素与差分 diff 异或得到集合 P'；同时，随机生成包含元素个数为 n 的集合 K 作为密钥。将集合 P 中的元素作为明文，与密钥集合 K 中元素一一对应进行加密，得到密文集合 C。其中，C 集合的每个元素是由 P 集合中的元素前 i 轮密文构成。同理得到密文集合 C'。

2)将集合 C 与 C' 合并构成新的集合密文对集合 X'，其中，X' 集合中每个元素由 C 和 C' 中对应的元素组合而成。即每轮的密文构成一组密文对，i 轮密文对构成一个元素，并生成集合 Y'。Y' 中每个元素长度为 i bit 的序列，且序列中的每比特均设置为 1。

3)对 X'集合中每个元素,随机选取 i 轮,将选择的 i 轮密文对用随机数对进行替换,得到训练数据集 X,并将 Y'中对应元素对应比特设置为 0,得到标签数据集 Y。

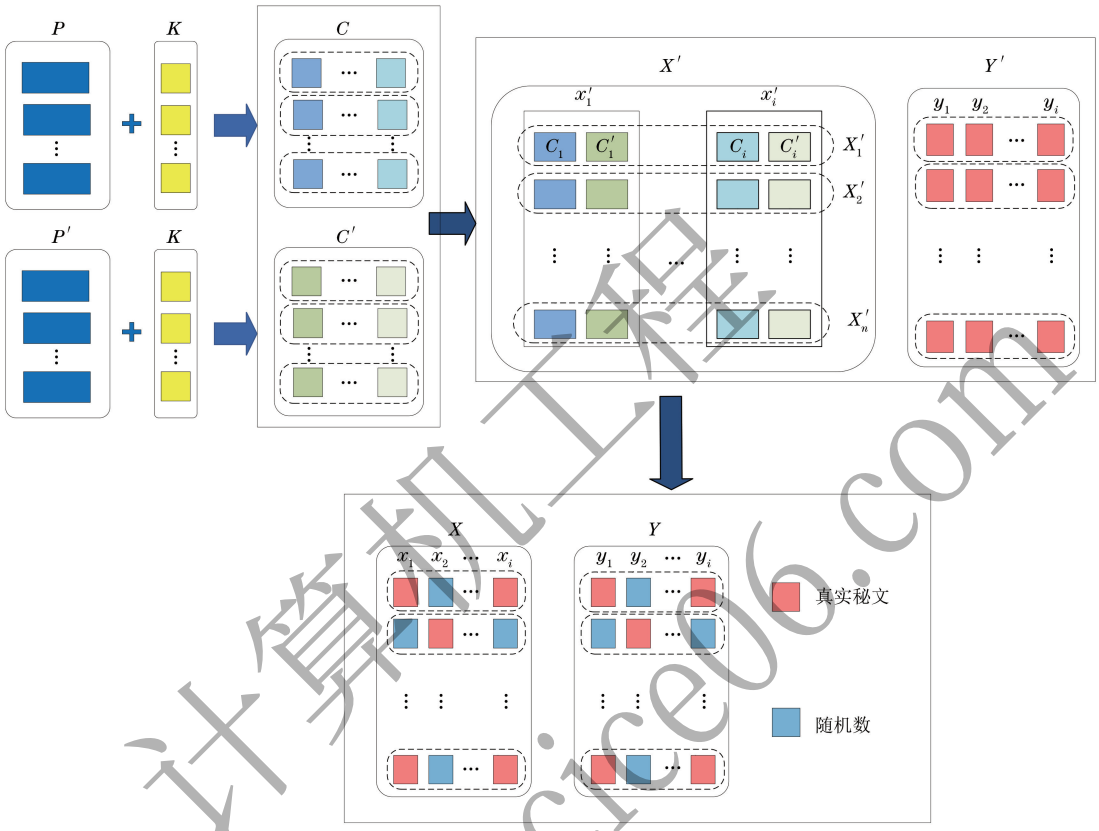

图 2 数据集构造流程

Fig. 2 Procedure of dataset construction

3 实验与分析

3.1 Transformer 模型设置与实验

本节探究了在 SPECK32/64 算法中基于注意力机制模型的差分区分器构造,介绍了模型的参数设置以及训练过程。对于 SPECK32/64 算法,在输入信息为 6 轮密文对,标签空间数为 2^3 个时,基于注意力机制模型的差分区分器效果最优。

3.1.1 Transformer 模型设置

针对 SPECK32/64 算法,选择输入差分为 $d_{\text{diff}}=$0x0040/0x0000,加密生成 10^7 个训练集、10^6 个验证集和 10^6 个测试集。为了使基于 Transformer 的区分器能够区分真实密文对和随机数据,快速收敛,本文对部分超参数进行设置。

损失函数设置为 binary_crossentropy,t 是二元标签 0 或 1,$p(t)$是输出等于标签 y 的概率。

$$L=-\frac{1}{N}\sum_{i=1}^{N}t_i\cdot\log_a(p(t_i))+(1-t_i)\cdot\log_a(1-p(t_i)) \tag{3}$$

批次大小:训练数据的 Batch_Size 设置为 2 000。

优化器:采用 Adam 算法。

周期:loss 随训练轮数不断降低,经过 20 个回合后不再降低,因此,训练周期设置为 20 个 Epoch。

学习率:采用循环学习率,其中,$l_{\text{low\_lr}}=1\times10^{-4}$,$h_{\text{high\_lr}}=9\times10^{-4}$ 和 $n_{\text{num\_Epochs}}=10$。

学习率的计算方式如下:

$$l_{\text{lr}_i}=l_{\text{low\_lr}}+\frac{((nm-1)-i)\bmod(nm)}{nm-1}\cdot(h_{\text{high\_lr}}-l_{\text{low\_lr}}) \tag{4}$$

3.1.2 Transformer 差分区分器训练与测试

训练集样本大小为 10^7 个,验证集样本大小为 10^6 个,正负样本数量各占其中的 1/2。当标签空间数为 2^3 个时,输入信息为 6 轮密文的 Transformer 神经网络差分区分器的准确率和损失函数随 Epoch 变化曲线如图 3 和图 4 所示。

6 轮神经网络差分区分器的标签设置如下:

$$Y=[1,*,1,*,1,*],\ *\in\{0,1\} \tag{5}$$

在 $Y=[1,*,1,*,1,*]$ 标签模式下,对于输入数据 x2,x4,x6 而言,即第 2、4、6 轮具有差分特性的密文对被随机数替换。因此,此时的标签空间数

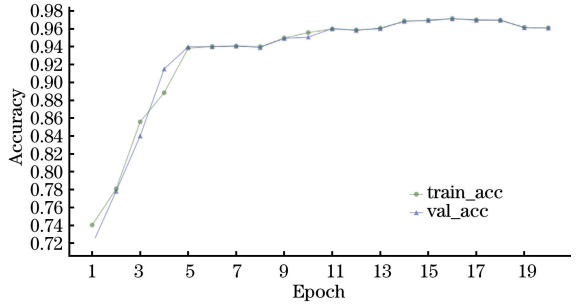

图 3 准确率随 Epoch 的变化曲线

Fig. 3 The change curves of accuracy with Epoch

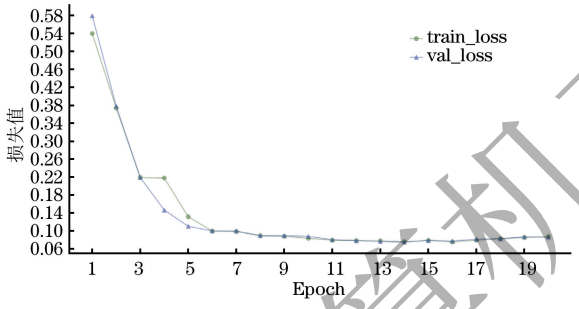

图 4 损失值随 Epoch 的变化曲线

Fig. 4 The change curves of loss values with Epoch

为 2^3 个。当标签空间数分别为 2^5(标签 $Y=[1,*,*,*,*,*]$)、2^4(标签 $Y=[1,*,1,*,*,*]$)、2^3(标签 $Y=[1,*1,*,1,*]$)个时,6 轮神经网络区分器测试结果如表 1 所示。其中,测试结果由区分准确率(ACC)、真阳性率(TPR)、真阴性率(TNR)、均方误差(MSE)来衡量。

表 1 6 轮神经网络区分器效果

Table 1 Effect of 6-round neural network discriminator

标签空间数/个	ACC	TPR	TNR	MSE
2^3	0.972 3	0.999 7	0.945 0	0.026 1
2^4	0.950 7	0.974 4	0.926 9	0.063 6
2^5	0.746 5	0.730 9	0.762 1	0.177 8

当标签空间数为 2^3 个时,基于 Transformer 的差分区分器对 SPECK32/64 的区分效果最好。在多头注意力机制训练过程中为每轮密文结构分配权重,能够充分提取密文结构间的差分性质,并对下一轮密文结构进行预测,区分真实密文和随机结构。

3.2 密钥恢复攻击实验

本节改进 GOHR[6] 提出的密钥恢复机制,应用于 SPECK32/64 的新密钥恢复攻击。讨论了当密钥恢复时,输入前多少轮相关密文对,能有效地区分当前轮次密文对的差分特性。实验结果表明,由于注意力机制能兼顾上下轮密文的差分特征,因此不仅能区分最初 1~6 轮密文对是否具有差分特性,还能分析后续任意密文对是否具有差分特性,利用区分器优秀的泛化能力只需提供前一轮的密文对,就能有效判断本轮的密文对是否具有差分特性。基于此方法,本文对 SPECK32/64 算法的最后 2 轮子密钥进行了恢复攻击。

3.2.1 差分区分器泛化能力研究

当选择 6 轮密文对和 2^3 个标签空间数时,能得到 1 个最佳的 Transformer 差分区分器。因此,在假设需要区分第 n 轮密文对是否具备差分特性的情况下,应当尽可能减少所需的信息量。因此,假设需要区分第 n 轮的密文对是否具有差分特性,本文采取了如下 3 种标签方式进行测试:$Y=[0,0,0,0,1,*]$,$Y=[0,0,0,0,*,*]$,$Y=[0,0,0,0,0,*]$。

对于标签 $Y=[0,0,0,0,1,*]$,输入数据 x1,x2,x3,x4 为随机数对,x5 为第 $n-1$ 轮的密文对,x6 进行随机设置,即有可能是第 n 轮的密文对,也可能为随机数对。

对于标签 $Y=[0,0,0,0,*,*]$,输入数据 x1,x2,x3,x4 为随机数对,x5,x6 进行随机设置,即有可能是第 $n-1$、n 轮的密文对,也可能为随机数对。

对于标签 $Y=[0,0,0,0,0,*]$,输入数据 x1,x2,x3,x4,x5 为随机数对,而 x6 进行随机设置,即有可能是第 n 轮的密文对,也可能为随机数对。

基于上述标签进行测试,前面位置对最后待区分位置准确率的影响如表 2 所示。

表 2 前一位置输入信息对最后位置评价指标的影响

Table 2 The influence of the input information of the previous position on the evaluation indicators of the final position

不同位置输入信息	ACC	TPR	TNR	MSE
$Y=[0,0,0,1,1,*]$	0.972 7	0.999 6	0.945 8	0.025 9
$Y=[0,0,0,0,1,*]$	0.972 3	0.999 7	0.945 0	0.026 1
$Y=[0,0,0,0,*,*]$	0.735 7	0.626 3	0.845 2	0.262 1
$Y=[0,0,0,0,0,0]$	0.501 2	0.057 5	0.944 9	0.487 2

其他位置的输入信息为随机数据,前 2 个位置的输入信息为第 $n-2$、$n-1$ 轮真实密文对,对第 n 轮密文对的区分准确率为 97.27%。当其他位置输入信息为随机数据,前一位置输入信息为第 $n-1$ 轮真实密文对时,对第 n 轮密文对的区分准确率为 97.23%。当其他位置输入信息为随机数据,前一位置输入信息为真实第 $n-1$ 轮密文对和随机数据混合时,第 n 轮密文对的区分准确率为 73.57%。当其他位置输入信息都为随机数据时,第 n 轮密文对的区分准确率为 50.12%。实验结果表明,只有前一轮密文对测试结果影响最大,因此在进行 n 轮部分子密钥恢复时只需要用到第 $n-2$ 轮密文和第 n 轮密文共 2 轮密文,其他位置用 0 填充即可。

传统方案、现有方案与本文方案的神经网络区分器效果对比如表 3 所示。文献[6,8,25]所提的方案均基于深度学习技术实现了 SPECK32/64 算法的差分区分器。传统差分区分器在第 5 轮时的准确率为 91.13%。文献[6]构建的神经网络区分器在第 7 轮和第 8 轮的准确率分别为 61.6%和 51.4%。文献[8]构建的神经网络区分器在第 6 轮和第 7 轮的准确率分别为 100%和 99.7%。文献[25]构建的神经网络区分器在第 8 轮时的准确率为 56.49%。而本文提出的方案在第 6 轮时的准确率为 97.23%,与文献[8]所提的方案相比降低。但是本文所提的方案可用于任何轮次到第 22 轮的差异特征区分,而文献[8]所提的方案仅用于第 6 轮和第 7 轮。实验结果表明,Transformer 模型的多头注意力机制为每轮密文赋予不同的权重,用于表示其与输入序列中其他轮密文的关联强度,从而捕捉到上下轮密文之间的差分关系,避免了每轮密文缺乏差分特征的问题,因此不管加密多少轮,只要有前一轮的密文特征,都不影响模型区分的准确率。本文基于 Transformer 差分区分器优秀的泛化能力,设计了新的密钥恢复方案。

表 3 神经网络区分器效果对比

Table 3 Comparison of the effects of neural network discriminators

方案	训练轮数/轮	区分轮数/轮	准确率/%
传统方案	5	5	91.13
文献[6]	7	7	61.60
	8	8	51.40
文献[8]	6	6	100.00
	7	7	99.70
文献[25]	7	7	88.19
	8	8	56.49
本文方案	6	6	97.23

3.2.2 改进 SPECK32/64 密钥恢复攻击的实现

基于第 2.1 节中改进 SPECK32/64 密钥恢复攻击方案进行第 21、22 轮子密钥恢复。实验参数设置如下:$N_{cts}=100$;$N_{it}=500$;$N_{cand1}=32$;$N_{cand2}=32$;$C_1=260$;$C_2=260$。

由于训练集包含 10^7 个样本,验证集包含 10^6 个样本,因此区分器的复杂度为 $10^7+10^6=10^{7.042}$。使用上述候选密钥筛选方案对第 22 轮 SPECK32/64 的 65 536 个候选密钥进行筛选,采用不同数据量的选择密文对进行 50 次实验。表 4 所示为以真实密钥为标准,大于真实密钥得分的候选密钥数量。筛选后的候选密钥数量越少,区分效果越好。当选择密文对数为 2^6 对时,候选密钥得分的平均取值范围−975.017～278.828 之间,平均能将 65 536 个密钥筛选到 17 个。

表 4 候选密钥测试结果对比

Table 4 Comparison of candidate key test results

选择密文对数/对	候选密钥得分	候选密钥数量/个	筛选后密钥数量/个
2^4	−253.414～−67.235	2^{16}	$2^{8.165}$
2^5	−489.997～137.798	2^{16}	$2^{5.248}$
2^6	−975.017～278.828	2^{16}	$2^{4.087}$

当筛选后最佳子密钥与真实子密钥之间差值的汉明权重和不超过 2,则攻击成功[6],如果在 500 次迭代内没有得到排名分数很高的猜测密钥,则视为攻击失败。总共进行了 50 次实验,其中 45 次实验成功,因此成功率为 90%。可见,Transformer 差分区分器有很好的实用性,区分效果较好,不需要浪费资源存储差分分布表。

文献[6]给出了 SPECK32/64 的最佳差分分布表需要 35 GB 的读写空间,而本文所提出模型权重只需要大约 10 MB。GOHR[6]提出的密钥恢复攻击只能对 SPECK32/64 进行 11、12 轮子密钥恢复。实验结果表明,基于 Transformer 的神经网络差分区分器,不仅有效攻击轮数更多,很好地解决了差分扩散问题,而且消耗的存储空间更少。

4 结束语

本文鉴于密码差分路径具有上下文的特征,将密码差分分析视为一种自然语言处理过程。利用 Transformer 模型注意力机制,提出一种全新的分组密码差分区分器模型,基于此模型设计了新的密钥恢复攻击方案。该攻击方案能有效地解决随加密轮数增加差分特征的识别准确率直线下降而无法进行子密钥恢复的问题。实验结果表明,以 SPECK32/64 密码算法为分析对象,在输入信息为 6 轮密文对,标签空间数为 2^3 个时,差分特征的识别准确率达到 97%。同时,在密钥恢复攻击中,在输入信息中前一位置密文信息的真实性对最后待区分密文影响最大。基于此特征,基于 6 轮的差分区分器,借助第 20 轮密文,对 SPECK32/64 最后两轮子密钥进行了恢复攻击,成功率为 90%。基于 Transformer 差分区分器的密钥恢复攻击方案增加了有效攻击轮数,成功缩减了候选密钥的范围。后续将此方法应用于其他密码分析中,如线性密码分析。

参考文献

[1] BIHAM E, SHAMIR A. Differential cryptanalysis of DES-like cryptosystems[EB/OL]. [2023-08-22]. https://link.springer.com/content/pdf/10.1007/BF00630563.pdf.

[2] Nation Bureau of Standards. Data encryption standard—Federal information processing standards publication[EB/OL]. [2023-08-22]. http://bitsavers.trailing-edge.com/pdf/nbs/fips/FIPS_46-1_Data_Encryption_Standard_Jan88.pdf.

[3] BLONDEAU C, GÉRARD B. Multiple differential cryptanalysis: theory and practice[EB/OL]. [2023-08-22]. https://link.springer.com/content/pdf/10.1007/978-3-642-21702-9_3.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[5] YOUNG T, HAZARIKA D, PORIA S, et al. Recent trends in deep learning based natural language processing[J]. IEEE Computational Intelligence Magazine, 2018, 13(3): 55-75.

[6] GOHR A. Improving attacks on round-reduced SPECK32/64 using deep learning[C]// Proceedings of the 39th Annual International Cryptology Conference. Berlin, Germany: Springer, 2019: 150-179.

[7] BEAULIEU R, SHORS D, SMITH J, et al. The SIMON and SPECK lightweight block ciphers[C]//Proceedings of the 52nd Annual Design Automation Conference. New York, USA: ACM Press, 2015: 1-6.

[8] BENAMIRA A, GERAULT D, PEYRIN T, et al. A deeper look at machine learning-based cryptanalysis[C]//Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques. Berlin, Germany: Springer, 2021: 805-835.

[9] 宿恒川, 朱宣勇, 段明. 基于 PU 分类的差分区分器及其应用[J]. 密码学报, 2021, 8(2): 330-337.
SU H C, ZHU X Y, DUAN M. Differential distinguisher based on PU learning and its application[J]. Journal of Cryptologic Research, 2021, 8(2): 330-337. (in Chinese)

[10] CHEN Y, YU H B. Bridging machine learning and cryptanalysis via EDLCT[EB/OL]. [2023-08-22]. https://eprint.iacr.org/2021/705.

[11] SU H C, ZHU X Y, MING D. Polytopic attack on round-reduced Simon32/64 using deep learning[M]// WU Y D, YUNG M. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2021: 3-20.

[12] 付超辉, 段明, 魏强, 等. 基于深度学习的多面体差分攻击及其应用[J]. 密码学报, 2020, 8(4): 591-600.
FU C H, DUAN M, WEI Q, et al. Polytopic differential attack based on deep learning and its application[J]. Journal of Cryptologic Research, 2020, 8(4): 591-600. (in Chinese)

[13] YANG G Q, ZHU B, SUDER V, et al. The Simeck family of lightweight block ciphers[C]//Proceedings of Conference on Cryptographic Hardware and Embedded Systems. Berlin, Germany: Springer, 2015: 307-329.

[14] SO J, KHOKHAR U M. Deep learning-based cryptanalysis of lightweight block ciphers[J]. Security and Communication Networks, 2020, 32: 3701067.

[15] YADAV T, KUMAR M. Differential-ML distinguisher: machine learning based generic extension for differential cryptanalysis[C]//Proceedings of International Conference on Cryptology and Information Security in Latin America. Berlin, Germany: Springer, 2021: 191-212.

[16] BAO Z, GUO J, LIU M, et al. Enhancing differential-neural cryptanalysis[C]//Proceedings of International Conference on the Theory and Application of Cryptology and Information Security. Berlin, Germany: Springer, 2020: 561-570.

[17] BAKSI A. Machine learning-assisted differential distinguishers for Lightweight ciphers[M]// Computer Architecture and Design Methodologies. Berlin, Germany: Springer, 2022: 141-162.

[18] JAIN A, KOHLI V, MISHRA G. Machine Learning Assisted Differential Distinguishers for Lightweight Ciphers[M]// Classical and Physical Security of Symmetric Key Cryptographic Algorithms. Berlin, Germany: Springer, 2022: 141-162.

[19] 杨小雪, 陈杰, 韩立东. 深度学习在 ARX 分组密码差分分析的应用[J]. 密码学报, 2022, 9(5): 923-935.
YANG X X, CHEN J, HAN L D. Application of deep learning in differential cryptanalysis of ARX block ciphers[J]. Journal of Cryptologic Research, 2022, 9(5): 923-935. (in Chinese)

[20] 陈怡, 包珍珍, 申焱天, 等. 用于大状态分组密码的深度学习辅助密钥恢复框架[J]. 中国科学: 信息科学, 2023, 53(7): 1348-1367.
CHEN Y, BAO Z Z, SHEN Y T, et al. A deep learning-aided key recovery framework for large-state block ciphers[J]. Scientia Sinica(Informationis), 2023, 53(7): 1348-1367. (in Chinese)

[21] VASWANI A, SHAZEER N M, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 6000-6010.

[22] YU H S, YANG Z G, TAN L, et al. Methods and datasets on semantic segmentation: a review[J]. Neurocomputing, 2018, 304: 82-103.

[23] 孙晓丽, 郭艳, 李宁, 等. 基于 seq2seq 模型的深度学习密码破译方法[J]. 通信技术, 2019, 52(9): 2217-2222.
SUN X L, GUO Y, LI N, et al. Deep learning password deciphering method based on seq2seq model[J]. Communications Technology, 2019, 52(9): 2217-2222. (in Chinese)

[24] AUER P. Using upper confidence bounds for online learning[C]//Proceedings of the 41st Annual Symposium on Foundations of Computer Science. Washington D. C., USA: IEEE Press, 2000: 56-64.

[25] HOU Z Z, REN J J, CHEN S Z. Improve neural distinguishers of SIMON and SPECK[J]. Security and Communication Networks, 2021(404): 1-1178.

编辑 薛晋栋