Advances in Applied Mathematics 应用数学进展, 2023, 12(2), 679-689
Published Online February 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.122069

Hans汉斯

基于改进Transformer模型的模板攻击

彭　静，王　敏*，王　燚

成都信息工程大学网络安全学院，四川　成都

收稿日期：2023年1月21日；录用日期：2023年2月20日；发布日期：2023年2月27日

摘　要

模板攻击是最强的侧信道攻击方法，然而传统模板攻击在处理高维特征数据时，可能会遇到数值计算问题。掩码策略是抵抗侧信道攻击的常见策略之一，其主要思想是利用随机掩码使密码算法运行过程中的敏感信息泄露能耗随机化。针对传统模板攻击存在的问题和加掩抵抗策略，本文重点研究了在机器翻译领域取得了显著成果的Transformer网络模型，首次提出了一种基于Transformer网络模型的模板攻击新方法。为了使适用于机器翻译的神经网络适应侧信道一维数据特征，本文对网络模型结构进行了适当的调整。实验对加掩防护的AES128算法采集能耗曲线，选取第一轮第三个S盒的输出作为攻击点，分别采用了多层感知机、一维卷积神经网络和基于改进Transformer的神经网络模型建立模板。最终实验结果表明，改进Transformer模型的卷积层在训练过程中会结合能量迹的不同兴趣点进行学习，自注意力机制能够赋予大的权值给重要的特征来提取出对模型分类重要的兴趣点，由此基于改进Transformer模型的模板攻击能够成功实现对带掩防护数据集的攻击，且需要的能迹数少于多层感知机和一维卷积神经网络。

关键词

Transformer模型，注意力机制，模板攻击

Template Attack Based on Improved Transformer Model

Jing Peng, Min Wang*, Yi Wang

School of Cybersecurity, Chengdu University of Information Technology, Chengdu Sichuan

Received: Jan. 21st, 2023; accepted: Feb. 20th, 2023; published: Feb. 27th, 2023

Abstract

Template attack is the strongest method of side-channel attack. However, traditional template at-

*通讯作者。

文章引用: 彭静, 王敏, 王燚. 基于改进 Transformer 模型的模板攻击[J]. 应用数学进展, 2023, 12(2): 679-689.
DOI: 10.12677/aam.2023.122069

tack may encounter numerical problems when processing high-dimensional feature data. Mask strategy is one of the common strategies to resist side-channel attacks. Its main idea is to use random mask to randomize the energy consumption of sensitive information leakage during the operation of cryptographic algorithms. Aiming at the problems of traditional template attacks and masking resistance strategies, this paper focuses on the Transformer network model, which has achieved remarkable results in the field of machine translation, and proposes a new template attack method based on the Transformer network model for the first time. In order to adapt the neural network suitable for machine translation to the one-dimensional data characteristics of the side channel, the structure of the network model has been appropriately adjusted. The experiment collects the energy consumption curve of AES128 algorithm for masking protection, selects the output of the third S-box in the first round as the attack point, and uses multi-layer perceptron, one-dimensional convolutional neural network and neural network model based on improved Transformer to build the template. The final experimental results show that the convolution layer of the improved Transformer model will combine the different interest points of the energy trace to learn during the training process, and the self-attention mechanism can give large weights to important features to extract the important interest points of the model classification, so the template attack based on the improved Transformer model can successfully achieve the attack on the masking protection data set. The number of traces required is less than that of multi-layer perceptron and one-dimensional convolutional neural network.

Keywords

Transformer Model, Attention Mechanism, Template Attack

1. 引言

密码设备在进行加密操作时会产生热量消耗、功率消耗、电磁辐射等信息的泄露，密码分析工作者可以通过这些信息的泄露与相关密码算法的输入输出、S 盒操作和相关设计细节对硬件密码设备来进行攻击，这一攻击方式也称作侧信道攻击。侧信道攻击可分为有学习、无学习两类。无学习的侧信道攻击包括简单能量分析(SPA, simple power analysis) [1]和差分能量分析(DPA, differential power analysis) [2]。有学习的侧信道攻击包括模板攻击(TA, template attack) [3]、随机攻击(SA, stochastic attack) [4]等。模板攻击是有学习的侧信道攻击中最成功的方法。

模板攻击利用了加密设备在运行过程中产生的能量消耗依赖于正在处理的数据这一事实。传统模板攻击通常分为两个阶段：第一个阶段利用多元高斯分布刻画能量消耗特征，第二个阶段实施攻击。然而传统的模板攻击使用的多元高斯分布中，协方差矩阵的运算量随着输入点的增加呈现几何增长。当矩阵维度过高时，会出现奇异矩阵问题而导致攻击失败。随着机器学习的发展，神经网络[5]被应用到了模板攻击中，这克服了传统模板攻击的数值计算问题。近年来，深度学习领域发展迅速，许多较为优秀的网络模型被应用到了侧信道攻击中。

本文针对传统模板攻击存在的问题以及密码算法的加掩措施，提出了一种基于 Transformer [6]神经网络模型的模板攻击新方法。

2. 相关概念

2.1. AES-128 加密算法

AES 高级加密标准[7]是一种基于有限域运算的分组密码算法，加密是输入是 128 bit 的明文，密钥可以为 128、192 或 256 bit 三种长度，各个长度的变换轮数和分组长度如表 1 所示。

本研究中的数据集使用的加密算法是 AES-128，加密过程一共 10 轮。前九轮每轮都包含字节替换，行移位，列混淆和轮秘钥加这四个基本步骤，而第 10 轮不包含列混淆。每个步骤的具体操作如表 2 所示。

字节替换也叫 S 盒替换，是 AES 算法中唯一的非线性变换，其产生的功耗也比一般操作更明显。因此，为了提高侧信道分析效率，攻击时都选择 AES 算法的第一轮 S 盒替换或最后一轮 S 盒替换作为攻击点。

Table 1. Relationship between AES algorithm's grouping, key length and encryption rounds
表 1. AES 算法的分组、密钥长度和加密轮数的关系

标准	明文分组	密钥长度	加密轮数
AES-128	128 bit	128 bit	10 轮
AES-192	128 bit	192 bit	12 轮
AES-256	128 bit	256 bit	14 轮

Table 2. Introduction to relevant operations in AES encryption algorithm
表 2. AES 加密算法中的相关操作介绍

操作	主要功能
轮密钥加(AddRoundKey)	这一部分主要是将明文块以 128bit 分组然后与轮密钥(RoundKey)做异或(XOR)逻辑运算
字节替换(SubBytes)	通过查表的方式(非线性替换)将每个字节进行替换，也就是 Sbox 操作；以 AES-128 为例，它有 16 个非线性操作
行移位(ShiftRows)	矩阵每一行以字节为基础单位进行循环左移
列混合(MixColumns)	通过与固定矩阵相乘实现对矩阵的每一列的四个字节进行混合操作

2.2. 卷积神经网络

卷积神经网络[8]一般由卷积层、池化层、全连接层组成。卷积层通过一些过滤器从输入中提取信息，使用几个过滤器的原因是期望每个过滤器从输入中提取一种不同的特征。池化层有最大池化和平均池化两种，通过使用过滤器在其输入中滑动来实现数据的降维，具有平移不变性。全连接层融合卷积提取到的特征，得到一个依赖于整个输入的全局结果。如图 1 所示，在实际场景中，卷积层、池化层、全连接层的层数都可以根据实际效果和需求进行相关的调整。

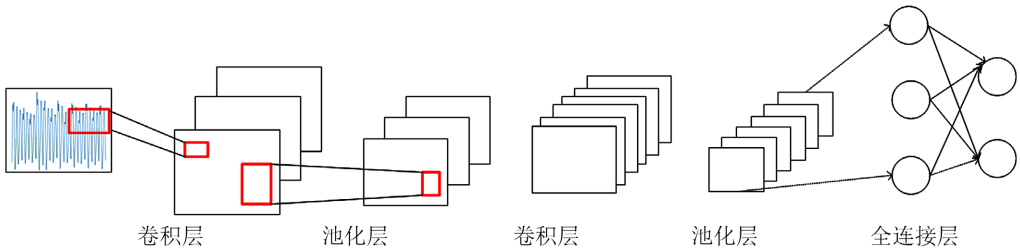

Figure 1. Convolution network structure
图 1. 卷积网络结构

2.3. 加掩

掩码技术是侧信道常见防护手段之一，使用一个或者多个随机掩码与可能发生泄露的中间值进行异或，从而使泄露的能耗随机化。本文使用加掩算法的伪代码的主要部分如下：

算法 1 AES128加掩算法

for $i = 1 \rightarrow 16$ **do**
    $state0[i] \leftarrow p[i] \oplus r[i]$
    $state1[i] \leftarrow r[i]$
    $key[i] \leftarrow mki$
end for
for $round = 1 \rightarrow 10$ **do**
    $key[1], ..., key[16] \leftarrow KeyScheduling(key[1], ..., key[16])$
end for
for $i = 1 \rightarrow 16$ **do**
    $state0[i] \leftarrow (state0[i] \oplus key[i] \oplus state1[i])$
    $state0[i] \leftarrow sbox[state0[i]]$
    $state0[i] \leftarrow (state0[i] \oplus state1[i]) \oplus rout$
end for

其中，p[i]是明文的第 i 个字节，r[i]是第 i 个明文字节对应的掩码，rin 和 rout 同样是掩码，分别用来保护 S 盒的输入和输出，mki 是每轮轮密钥的第 i 个字节，KeyScheduling 指密钥拓展。

3. Transformer 神经网络结构与调整

3.1. Transformer 神经网络结构

Vaswani 等人在 2017 年提出了 Transformer 神经网络，模型结构如图 2 所示。模型分为 Encoder 端 Decode 端。输入经过 Encoder 端的输出通过交互注意力与 Decoder 相联。Encoder 和 Decoder 的个数可根据实际问题进行调整。

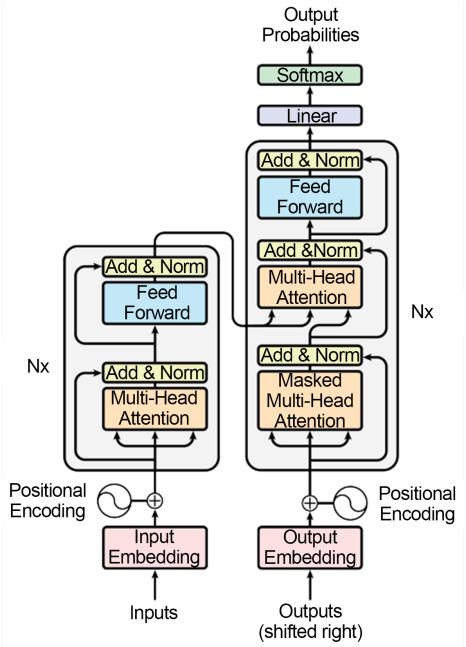

Figure 2. Transformer network model structure
图 2. Transformer 网络模型结构

随着 Transformer 模型在 NLP 领域的广泛应用，有研究逐渐将其应用到计算机视觉领域，例如图像分类、目标检测等[9]。当 Transformer 用于分类任务时只需要模型中的 Encoder 部分，例如用于文本分类的 BERT 模型[10]、图像分类的 ViT 模型[11]。本研究将其应用于侧信道攻击中，本质上也是用于分类任务，只需要其 Encoder 部分，所以本文只对 Encoder 部分进行介绍。

3.1.1. 嵌入层

Transformer 使用词嵌入层来将输入序列中每个元素的维度转换为 d_{model}。这与其他序列转换模型类似。原文中 d_{model} 为 512。如式(1)所示。

$$X_E = XW_E \tag{1}$$

其中 X 表示输入序列，W_E 表示可训练的权重矩阵。

3.1.2. 多头自注意力机制

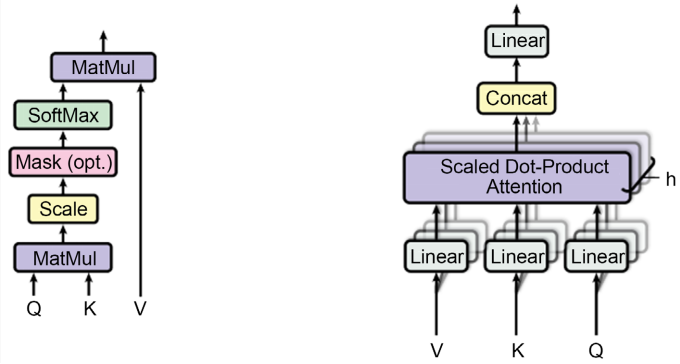

Figure 3. Self-attention mechanism and multi-head attention mechanism
图 3. 自注意力机制和多头注意力机制

Transformer 利用了自注意力机制，如图 3 所示。自注意力机制的输入、输出都是一个序列。对于输入序列中的每个元素，通常通过点积的方法让其学习与序列中所有元素的相关性。如式(2)所示。

$$Attention(Q,K,V) = softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{2}$$

其中，$Q = XW_q, K = XW_k, V = XW_v$，$X$ 是输入序列，W_q, W_k, W_v 是通过训练得到的权重参数，即 Q,K,V 是输入序列 X 的线性变换。$\frac{1}{\sqrt{d_k}}$ 是缩放因子。当点积在数量级上大幅增长，会将 softmax 函数推向具有极小梯度的区域。为了抵消这种影响，可以对点积扩展 $\frac{1}{\sqrt{d_k}}$ 倍。公式(2)实际上是获得输入序列中各元素之间的关联，利用关联作为一种权重再来加权输入的各元素。

多头自注意力机制可以理解为用多头模拟卷积的多输出通道。因为在实际问题中，我们往往希望关注到不同的相关性。给定输入向量和头数 h，输入向量首先被转换成三组不同的向量，即上述的 Q,K,V。每个 Q,K,V 都会被转换为一组有 h 个个向量的向量组，每个向量的维度为 dmodel/h。此时 Q,K,V 可被表示为：$\{Q_i\}^h_{i-1}, \{K_i\}^h_{i-1}, \{V_i\}^h_{i-1}$，计算过程如式(3)所示。

$$\{Q_i\}^h_{i-1} MultiHead(Q',K',V') = Concat(head_1,\cdots,head_h)W^o \tag{3}$$

其中，$head_i = Attetion\left(Q_i, K_i, V_i\right)$，$Q', K', V'$ 分别是 $\{Q_i\}^{h}_{i-1}, \{K_i\}^{h}_{i-1}, \{V_i\}^{h}_{i-1}$ 的整合。W^o 是线性投影矩阵。

对多头注意力机制可以简单理解为，如果多头的头数为 n，那么待解决问题中的输入有 n 种不同的相关性。

3.1.3. 多层感知器

多层感知器[12] [13]将自注意模块的输出会被传递到两层多层感知器(multi-layer perceptrons, MLP)中，两层 MLP 之间具有激活函数 ReLU，激活函数的非线性增加了分类函数的复杂性。如式(4)所示。

$$FFN\left(X_A\right) = F_2\left(ReLU\left(F_1\left(X_A\right)\right)\right) \tag{4}$$

其中 X_A 是第一层 MLP 的输入，F_1 和 F_2 可表示为 $Wx+b$，x 是多层感知器的输入。W 是线性变换层的参数矩阵，b 是偏置。自注意力层和前馈神经网络层交替使用，原结构中交替使用了 6 次。

3.1.4. 残差连接和层归一化

如图 2，在编码器(解码器)的每一个子层后都使用了残差连接[14]，这里的子层指多头注意力层和 MLP 网络层。残差连接就是把输入和输出连接起来得到一个新的向量，再把新的向量输入到后面的层。在 Transformer 结构中，残差连接之后会进行 LayerNormalization。如式(5)所示。

$$LayerNorm\left(x + Attention\left(x\right)\right) \tag{5}$$

其中，x 是自注意力层的输入。

3.2. 模型结构调整

为了将 Transformer 网络模型应用于侧信道数据分析，对模型作以下调整：只保留 Transformer 中的 Encoder 部分。由于将 Transformer 网络模型应用到侧信道攻击中本质上是用其处理分类问题，所以只需要 Encoder 部分，其中用卷积、池化层代替 Transformer 原结构中的词嵌入层。Transformer 最初用于自然语言处理，序列中的每个单元是单词，每个单词经过词嵌入后转换为向量，可以表达丰富的语法和语义信息。而侧信道的数据序列中单元是一个个的能耗值。本文使用的数据集是 AES128 算法加掩防护的能迹，卷积层在训练过程中会结合能量迹的不同兴趣点进行学习，能有效攻破具有掩码防护的加密算法。所以本文利用卷积提取特征来进行向量化处理，即每个样本通过卷积、池化层后变成了向量。采用卷积神经网络来作向量化处理是因为本文使用的数据集是加掩防护的，卷积神经网络在训练过程中，会结合能量迹中不同的兴趣点进行学习，可以有效的攻破具有掩码防护的加密算法，并且可以将整条能量迹作为模型的输入，在学习过程中会自动提取特征，不需要对能量迹进行预处理操作。具体调整后的网络结构如图 4 所示。

嵌入层中，每一个卷积层的卷积核个数依次为 4、8、16、32、64、128，前两个卷积层大小分别是 7 和 4，后面几个卷积核大小均为 3，卷积步长恒为 1。池化均为最大池化，池化窗口、步长都为 2。图中的全连接层 1、2 是 Transformer [15]注意力机制后面中的两个全连接层。最后的输出层就是分类层。

4. 实验设计与结果分析

4.1. 实验数据

实验是基于测量一个执行在智能卡上的 AES128 加掩算法实施的，实验设备由示波器、智能卡、读卡器、计算机等组成。其中计算机负责向读卡器下发命令控制智能卡运行 AES 加密算法，同时，通过 USB 线向智能卡下发明文信息。在智能卡工作的同时，给示波器触发信号，以进行能量迹采集。示波器

将采集到的能迹信息发送到计算机中进行存储。本文使用的加密算法在实现的时候，第一轮的第一个和第二个 S 盒操作是未加防护的，所以选择第一轮的第三个 S 盒的输出作为敏感中间值。为了识别与sbox(p[3]⊕k[3])相关的泄漏样本，分析了几个信噪比，分别是：sbox(p[3]⊕k[3])、sbox(p[3]⊕k[3])⊕rout、rout、sbox(p[3]⊕k[3])⊕r[3]、r[3]。原始能迹一共有 10 万个样本点，通过泄露分析发现其中[45400,46100]这 700 个样本点同时包含了上述几种中间值的泄露，可以综合其中的能耗，识别无掩的 S 盒输出。本实验采集的数据集总共包括 60000 条能量迹，其中 50000 条能迹作为训练集，其余的 10000 条能迹用来攻击。

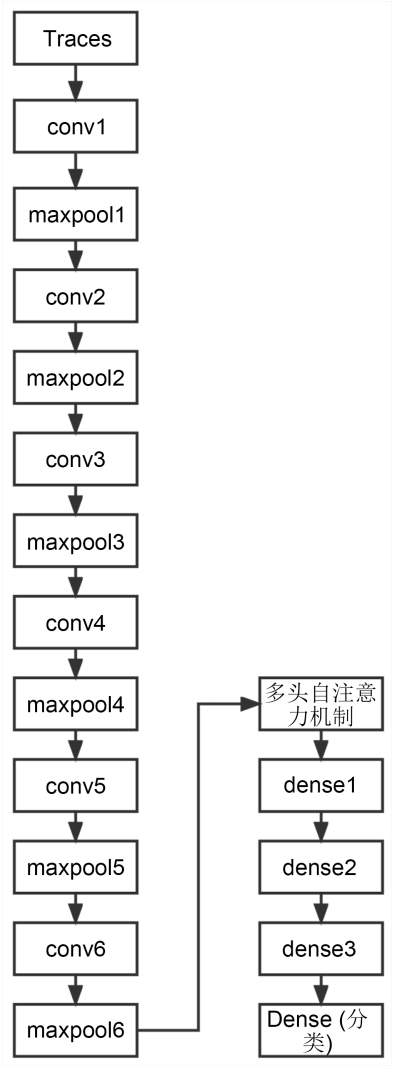

Figure 4. Transformer improved network structure
图 4. Transformer 改进网络结构图

4.2. 实验指标

本文用猜测熵作为描述攻击模型性能的指标。猜测熵是指在多次攻击中，正确密钥在攻击模型推荐的候选密钥队列中的平均排名。猜测熵越小意味着攻击成功时所需要的能迹数量越少。猜测熵如式(6)所示。

$$ge = \left| K \in k \mid P(K) > P\left(K^*\right) \right| \tag{6}$$

其中，$P(K)$表示密钥猜测的最终得分，$P\left(K^*\right)$表示获得正确密钥的得分。猜测熵最小是 0，此时正确密钥的得分大于其他猜测密钥的得分，代表攻击成功。

4.3. 实验对比

本文将基于 Transformer 网络模型的模板攻击与基于 MLP、一维卷积的模板攻击进行了对比。本文建立上述三个攻击模型后，用训练集对其进行训练。针对神经网络中的各种超参数的寻优，方法是首先通过随机寻优的方式：在模型的各超参数的实际取值范围内，均匀随机取值，使用这些随机超参数值配置模型，进行模型训练。然后是进行网格寻优：找到较好的模型后，将需要调整的超参数设定为离散值，对超参数的各种取值进行交叉组合，使用超参数的每个组合配置模型，然后进行模型训练。根据模型最终使用验证集取得的猜测熵，确定并保存最佳模型。利用三个网络的最佳模型进行对比。为确保实验结果的一般性，本文取 10 次攻击的平均结果，比较三个网络的平均攻击性能。本文旨在用最少的能迹恢复正确密钥。

攻击效果如图 5 所示。对于 MLP 网络，在使用 390 条能迹时猜测熵首次达到 0；对于 CNN 网络，在使用 320 条能迹时猜测熵首次达到 0；对于 Transformer 网络，在使用 100 条能迹时猜测熵就首次达到 0。MLP 网络在使用大于等于 550 条能迹之后猜测熵完全等于 0，CNN 在使用大于等于 510 条能迹之后猜测熵完全等于 0，而 Transformer 在使用大于等于 250 条能迹之后猜测熵就完全等于 0，此数量相比 MLP、CNN 网络减少了一半。由此可见，Transformer 网络的猜测熵随攻击能迹数的增加而减少的趋势也更加稳定。综上，多层感知器有自适应的学习能力和强大的非线性变换能力，可以直接作用于加掩的侧信道功耗数据，无需知道掩码即可对密钥进行还原。卷积神经网络在训练过程中能结合能量迹不同兴趣点学习，能有效攻破具有掩码防护的加密算法。本文的 Transformer 网络结构首先利用卷积层对能耗值进行向量化处理，当卷积层的输出经过自注意力层时，注意力机制可以有效抑制噪声的表达并提高重要特征的运算权重，发现各样本泄露的显著性，在一定程度可以起到降噪的作用，由此取得了更好的效果。

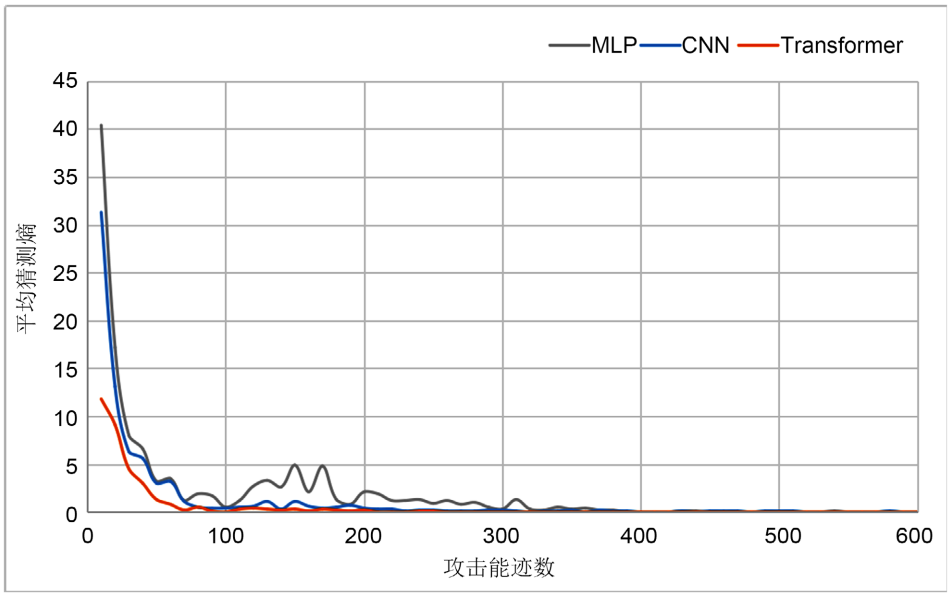

Figure 5. Average guess entropy of MLP network, CNN network and Transformer network
图 5. MLP 网络、CNN 网络、Transformer 网络的平均猜测熵

本文的嵌入层使用了卷积、池化层，由于卷积神经网络本身能够有效攻破带掩码防护的加密算法，所以下面将网络结构中的卷积、池化部分与改造后的 Transformer 结构进行对比，onlyCNN 代表只有卷积、池化部分。结果如图 6 所示，可见嵌入层使用卷积、池化层的改造后的 Transformer 网络结构能够取得更好的效果。

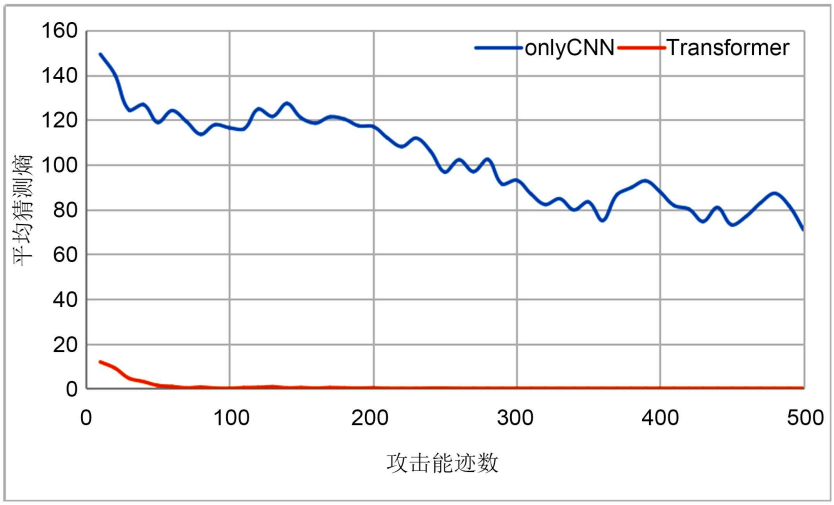

Figure 6. Average Guess Entropy of Transformer and onlyCNN
图 6. Transformer、onlyCNN 的平均猜测熵

在做泄露分析选取能迹范围时，发现 700 个样本点附近有一些微弱的泄露，所以扩大样本区域范围至 10000 个样本点，尝试利用这些微弱泄露。Transformer 结构的实验结果如图 7 所示，仅用 15 条能迹就能成功实施攻击，远少于 700 个样本点的能迹数。可见，无掩中间值的泄露特征是泄露非常微弱，但泄露区域广泛。在具有明显泄露的情况下，不需要利用这些微弱泄露进行训练和攻击。但在本质泄露非常低的情况下，有必要利用微弱的泄露。同时也对 MLP、CNN 进行实验，实验结果如图 8 所示，同样都取得了更好的效果，但所需能迹数仍然多于 Transformer。

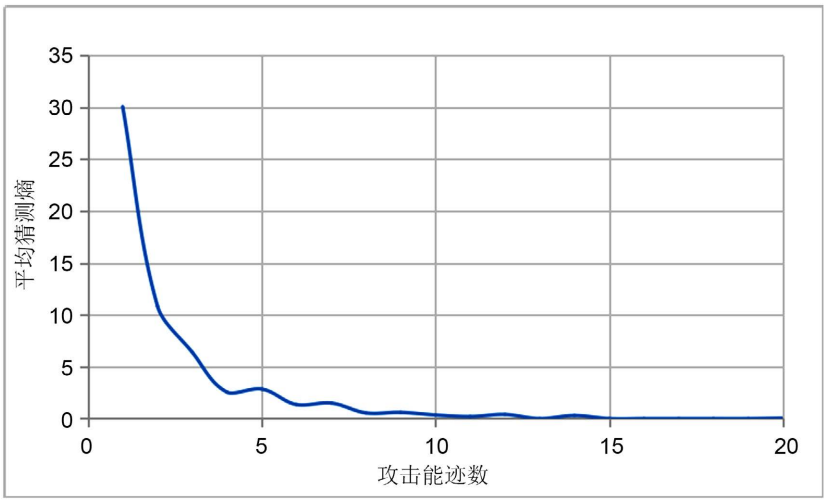

Figure 7. Average Guess Entropy of Transformer on Data Set with Increased Sample Range
图 7. Transformer 在增大样本范围数据集上的平均猜测熵

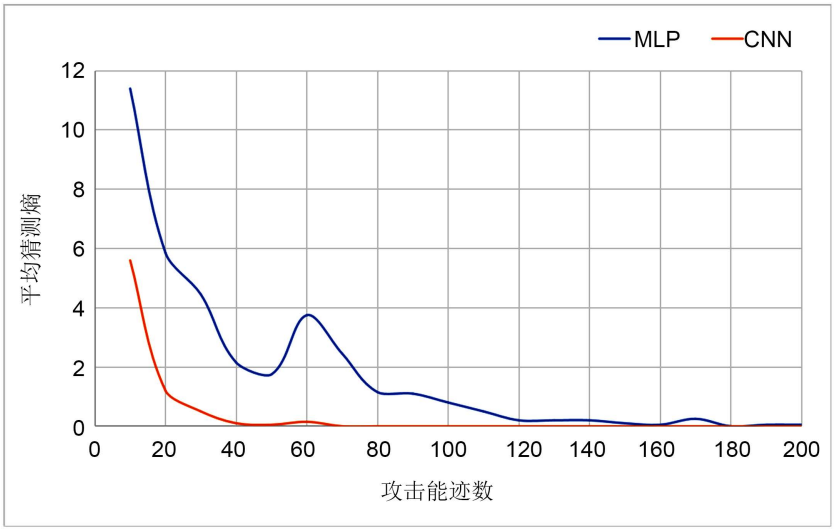

Figure 8. Average guess entropy of MLP and CNN network structure on the data set with increased sample range
图 8. MLP、CNN 网络结构在增大样本范围数据集上的平均猜测熵

5. 结束语

本文将改进后的 Transformer 神经网络用于侧信道模板攻击，与 MLP、CNN 网络模型进行了对比实验，证明了改进后的 Transformer 神经网络能使用更少的能迹数攻破具有掩码防护的加密算法。接着扩大了样本范围，利用一些微弱泄露，取得了更好的效果。本文的 Transformer 模型嵌入层使用的卷积、池化层，其平移不变性能够有效攻破具有抖动防御的数据集，接下来可以进行相关的实验研究；Transformer 中使用的自注意力模型，今后可以尝试将其中的自注意力模型改成其他注意力模型，也许可以达到更好的效果。

基金项目

四川省科技计划资助(项目号：2021ZYD0011)。

参考文献

[1] Chari, S., Rao, J.R. and Rohatgi, P. (2003) Template Attacks. In: Kaliski Jr., B.S., Koç, Ç.K. and Paar, C., Eds., *Cryptographic Hardware and Embedded Systems—CHES* 2002, *Lecture Notes in Computer Science*, Vol. 2523, Springer, Berlin, 13-28. https://doi.org/10.1007/3-540-36400-5_3

[2] Kocher, P. (1999) Differential Power Analysis and Related Attacks. *Annual International Cryptology Conference*, 388-397. https://doi.org/10.1007/3-540-48405-1_25

[3] 夏璇, 陈柏沩, 钟卫东. 侧信道攻击与防御概述[J]. 科学与信息化, 2020(35): 57.

[4] Schindler, W., Lemke, K. and Paar, C. (2005) A Stochastic Model for Differential Side Channel Cryptanalysis. In: Rao, J.R. and Sunar, B., Eds., *Cryptographic Hardware and Embedded Systems—CHES* 2005, *Lecture Notes in Computer Science*, Vol. 3659, Springer, Berlin, 30-46. https://doi.org/10.1007/11545262_3

[5] Siddiqui, S., Khan, M.S., Ferens, K. and Kinsner, W. (2016) Detecting Advanced Persistent Threats Using Fractal Dimension based Machine Learning Classification. *Proceedings of the* 2016 *ACM on International Workshop on Security and Privacy Analytics*, New Orleans, 11 March 2016, 64-69. https://doi.org/10.1145/2875475.2875484

[6] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. ArXiv: 1706.03762.

[7] Jain, R. (2001) Advanced Encryption Standard (AES). Washington University in St. Louis, Saint Louis.

[8] Friedberg, I., Skopik, F., Settanni, G. and Fiedler, R. (2015) Combating Advanced Persistent Threats: From Network Event Correlation to Incident Detection. *Computers & Security*, **48**, 35-57. https://doi.org/10.1016/j.cose.2014.09.006

[9] Tay, Y., Dehghani, M., Bahri, D. and Metzler, D. (2022) Efficient Transformers: A Survey. *ACM Computing Surveys*, **55**, Article No. 109. https://doi.org/10.1145/3530811

[10] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT* 2019, Minneapolis, 2-7 June 2019, 4171-4186.

[11] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2021) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.

[12] Martinasek, Z., Dzurenda, P. and Malina, L. (2016) Profiling Power Analysis Attack Based on MLP in DPA Contest V4.2. 2016 39*th International Conference on Telecommunications and Signal Processing*, Vienna, 27-29 June 2016, 223-226. https://doi.org/10.1109/TSP.2016.7760865

[13] Martinasek, Z., Hajny, J. and Malina, L. (2014) Optimization of Power Analysis Using Neural Network. In: Francillon, A. and Rohatgi, P., Eds., *Smart Card Research and Advanced Applications*, *CARDIS* 2013, *Lecture Notes in Computer Science*, Vol. 8419, Springer, Cham, 94-107. https://doi.org/10.1007/978-3-319-14123-7_7

[14] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[15] Jaderberg, M., Simonyan, K., Zisserman, A., *et al*. (2015) Spatial Transformer Networks. MIT Press, Cambridge, Massachusetts.