

引用格式: 谢欣宜, 刘帅, 王立辉, 等. 基于卷积神经网络可视化的侧信道泄露区间定位[J]. 微电子学与计算机, 2021, 38(3): 1-7. [XIE X Y, LIU S, WANG L H, et al. Revealing leakage location in side-channel attack based on convolutional neural network visualizing[J]. Microelectronics & Computer, 2021, 38(3): 1-7.]

基于卷积神经网络可视化的侧信道泄露区间定位

谢欣宜¹, 刘 帅¹, 王立辉¹, 李 清^{1,2}, 俞 军^{1,2}

(1 上海复旦微电子集团股份有限公司, 上海 200433;

2 复旦大学 专用集成电路与系统国家重点实验室, 上海 201203)

摘 要: 侧信道攻击中, 泄露区间的定位对进一步提升攻击效果和芯片防护产生至关重要的作用. 对于基于神经网络的模板攻击, 由于神经网络是黑盒模型, 使得泄露区间定位成为难题. 本文通过改进加权梯度类激活映射(Grad-CAM), 提出一种新的可以定位功耗或电磁曲线泄露位置的方法 WAvg-Grad-CAM. 该方法利用梯度计算进行区间定位, 计算输入曲线的不同位置对攻击成功的影响程度, 从而可以准确地定位输入曲线的泄露区间. 通过实验对比其他常见的可视化方法在模板攻击中的应用效果, 发现其拥有曲线平滑和噪音小的优势. 实验证明在泄露情况未知且不进行二阶曲线处理的情况下, 即使网络成功率较低, 也可以直接对带掩码实现的曲线进行精准定位.

关键词: 模板攻击; 神经网络; 可视化

中图分类号: TP183; TP309.2

文献标识码: A

文章编号: 1000-7180(2021)03-0001-07

Revealing leakage location in side-channel attack based on convolutional neural network visualizing

XIE Xin-yi¹, LIU Shuai¹, WANG Li-hui¹, LI Qing^{1,2}, YU Jun^{1,2}

(1 Shanghai Fudan Microelectronics Group Company Limited, Shanghai 200433, China;

2 State Key Laboratory of ASIC and System, Fudan University, Shanghai 201203, China)

Abstract: Locating leakage plays a vital role in further improving the attack success rate and countermeasure in side-channel attacks. For template attacks based on convolutional neural network, it is difficult to locate the leakage due to the invisibility of the network. This paper proposed a new method named WAvg-Grad-CAM to locate the leakage spot of side-channel traces, based on the ideal of Gradient-weighted Class Activation Mapping (Grad-CAM). This method applies gradient mapping to compute the attacking effectiveness of different positions on the input trace. According to the attacking effectiveness, the leakage position of the input trace can be precisely located. The experiment shows that this new method has advantages of smooth result and lower noise, compared with other common visualization methods. The result shows that this attack method can be used to localize the masked implementations precisely, without leakages combination processing and without knowledge of the implemented protections, even on the condition of low successful rate.

Key words: template attack; neural network; visualization

1 引言

智能卡作为现代常用的密码设备,因其存放着密钥等敏感信息和人们对隐私的逐渐关注,使得智能卡安全风险问题得到广泛的研究. 侧信道攻击(Side Channel Attack, SCA)主要包括计时攻击、能量分析攻击和电磁攻击,它的出现给芯片安全带来了极大的挑战. 其攻击本质是利用密码设备在运行过程中产生依赖于密钥的旁路信息来实施密钥恢复. 由于侧信道攻击不会破坏原有芯片或改变其正常工作状态,实施起来相对容易,使得密码设备的安全性遭到了严重威胁. 模板攻击更因其效率高成本低,得到了广泛的研究和应用.

传统模板攻击中,攻击者通过能量采集设备来采集芯片进行相关密码操作的能耗或电磁信号曲线,加密和解密操作需要一定时间,并且由于高采样率的设置,使得采样点多曲线长. 而模板攻击在建模中存在特征点数量选取和稀疏矩阵求逆的问题,往往需要进行泄露位置的定位,传统的泄露定位方法包括相关性分析、t 检验和标准类内方差等.

随着深度学习的广泛应用,基于卷积神经网络(Convolutional Neural Network, CNN)的模板攻击也成为主流攻击手段,相较于传统基于多元高斯分布的模板攻击,它拥有更好的特征提取和拟合能力,能提高模板攻击的效率和正确率,对曲线的预处理和特征点数量的选择方面也不再受到那么严格的限制. 但 CNN 属于黑盒模型,攻击者无法了解输入曲线的各个点对攻击结果的影响. 对曲线泄露区间的定位可以降低攻击难度,当网络输入维度过高时,网络收敛困难,超参数调整难度高. 对具有一定成功度的网络进行可视化分析,根据结果调整网络参数,通过减少输入维度和对超参数的调整,攻击速度会加快,网络也可以进一步被优化提高其正确率. 此外定位泄露位置可以反向分析芯片在当前采样时刻的运算操作,从而分析过程中出现信息泄露的原因,在芯片的迭代设计中可以针对该位置进行防护.

针对神经网络可视化这个难题,在深度学习图像识别领域,研究者通过提出了反卷积、导向反向传播和加权梯度类激活映射(Gradient-weighted Class Activation Mapping, Grad-CAM)等方法来分析神经网络输出对输入图片的关注度权重分布,从而达到目标检测的目的.

本文将图像处理和识别领域的可视化方法应用到模板攻击中,提出 WAvg-Grad-CAM 一种新的基

于神经网络的泄露区间定位方法,较图像处理和识别领域的可视化方法而言,该方法更适合于侧信道攻击,其可视化分析曲线平滑噪声小. 相较于传统的泄露区间定位方法而言,该方法可以在对掩码实现未知,且没有进行泄露点组合的曲线预处理情况下,直接进行泄露定位.

2 概念描述

模板攻击是一种结合概率统计的侧信道攻击方法,其本质上是一个分类问题,它分为训练和攻击两个阶段,在第一步的训练过程中,攻击者利用已经完全控制的设备建立分类模型;然后在第二步攻击过程中,利用已知的分类模型来得到被攻击设备的秘密信息. 用数学符号表示这个过程如下:传统的模板攻击过程相当于函数映射 F_{w_i} , w_i 代表函数中的参数,例如对于正态函数族来说,该参数是均值 μ 和协方差矩阵 Σ ,即 $w_i = (\mu_i, \Sigma_i)$. 函数 F_{w_i} 是从侧信道信息 x 到秘密信息 k 的映射,即 $k = F_{w_i}(x)$. 在训练过程中,攻击者采集大量的侧信道信息 x 及其所对应的秘密信息 k ,组成很多个数据对,即 (x_i, k_i) ,根据秘密信息 k 建立 p 个模板, $i \in [1, p]$,通过对应 k_i 的所有数据对来计算模板函数中的参数 w_i . 训练完成后,在攻击阶段采集被攻击芯片的侧信道信息 x ,结合模板函数 $F_{w_i}(x)$,可以得到被攻击芯片的秘密信息 k . 不同模板攻击方法的区别在于他们的函数映射 $F_{w_i}(x)$ 的不同,传统模板攻击采用的函数是多元正态分布函数族,参数 $w = (\mu_i, \Sigma_i)$, $i \in [1, p]$:

$$F_w = \underset{i}{\operatorname{argmax}} \left[\frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} e^{-\frac{1}{2} \cdot \langle x - \mu_i \rangle^T \Sigma_i^{-1} \langle x - \mu_i \rangle} \right] \quad (1)$$

近年来,神经网络得到很大程度发展,逐渐应用于侧信道分析中,成为主流的攻击方法之一. 采用卷积神经网络的模板攻击方法和传统方法主要有两点不同:第一,模板函数 $F_{w_i}(x)$ 采用卷积神经网络,根据网络的不同结构,模板函数的参数 w 可以多达百万个,从而比多元正态函数有了更强的拟合能力;第二,传统方法在训练过程中,确定模板函数的参数 w 采用了最大似然估计的方法,而基于卷积神经网络的模板攻击则是采用反向传播来确定网络中的参数. 利用卷积神经网络进行模板攻击,可以对未对齐曲线或带掩码的曲线攻击成功^[1],使得卷积神经网络的攻击相对于传统模板攻击拥有一定优势.

利用神经网络进行攻击,首先需要采集数据, CNN 网络总共需要三个数据集,包括训练集、验证

集和测试集,其中训练集和验证集是由攻击者控制的芯片所采集的已知秘密信息的能量迹,而测试集则是采集被攻击芯片未知密钥所得的能量迹。根据上述攻击过程,第一步在训练阶段,攻击者通过训练集和验证集数据,通过反向传播来确定神经网络的参数。网络收敛之后,在第二步攻击过程中,攻击者可以用神经网络来获取能量迹的秘密信息。卷积神经网络由输入层,隐含层和输出层组成,其中隐含层包括卷积、激活、池化等操作。用于模板攻击的卷积神经网络的输入是一条能量迹,通过隐含层,最后输出这条曲线所对应的标签,即秘密信息 k 。

3 泄露区间定位方法

3.1 经典泄露点定位方法

侧信道攻击中,密码设备的能量消耗往往依赖于该设备所执行的操作和操作数,所以攻击者采集的能耗曲线的功耗值往往和芯片正在处理的数据有关。在设备进行加密等有关敏感信息操作时,攻击者测量并记录芯片的能量消耗,即在加密不同时刻 t 下的侧信道信息例如功耗值 $x(t)$,同时记录对应的明密文或者秘密信息 k 等信息。攻击者往往利用密钥和能耗的相关性来进行曲线泄露区间的定位:

$$r(k, x) = \frac{\text{Cov}(k, x)}{\sqrt{\text{Var}(k) \cdot \text{Var}(x)}} \quad (2)$$

式中, Cov 和 Var 分别表示协方差和方差,相关性越高,代表该功耗点泄露的密钥相关信息越多,相关性曲线的峰值所对应的区域即被认为是该能量曲线的泄露位置。

由于泄露点的功耗值会依赖于密钥值,通过对不同类别的秘密信息进行建模,分析不同类的侧信道信息的变化来进行泄露点提取,假设秘密信息属于类别 i ,对应该秘密信息 k_i 的侧信道曲线的均值记为 $\mathbb{E}(x | k_i)$,泄露时刻出现在 $\sum_{i,j} (\mathbb{E}(x | k_i) - \mathbb{E}(x | k_j))$ 的极大值上^[2]。文献[3] 对该方法进行了优化,利用平方差和(Sum of Squared pairwise difference, SOSD) $\sum_{i,j} (\mathbb{E}(x | k_i) - \mathbb{E}(x | k_j))^2$ 来进行定位。通过归一化对 SOSD 进一步优化,即 SOST(Sum of Squared pairwise T-difference, T-difference 就是 t 检验),对应秘密信息 k_i 的侧信道曲线数记为 n_i ,则可以通过公式(3) 进行计算:

$$\sum_{i,j=1} \left[\frac{\mathbb{E}(x | k_i) - \mathbb{E}(x | k_j)}{\sqrt{\frac{\text{Var}(x_i)}{n_i} + \frac{\text{Var}(x_j)}{n_j}}} \right]^2 \quad (3)$$

文献[4] 中提出标准类内方差(Normalized Inter-Class Variance, NICV) 的方法,通过公式(4) 进行分析。

$$\frac{\text{Var}[\mathbb{E}(x | y)]}{\text{Var}[x]} \quad (4)$$

式中, y 代表明文或密文,该方法省去了建模的时间,所以泄露点分析的过程可以被大大加快,且只需要通过公共信息例如明密文就可以进行定位,较之前的分析方法拥有明显优势。

另一种常见的降维方法是主成分分析法^[5] (Principal Components Analysis, PCA),在尽量保持数据方差一致的前提下,通过线性变化将高维数据投影到低维空间中,该方法首先将坐标系旋转到数据方差最大的方向上重新建立新的坐标系,丢弃次要方向并保留最重要的维度从而达到降维的效果。

这些传统泄露点提取方法都拥有数据量大,计算资源耗费高的问题。此外,他们都需要对曲线进行对齐等预处理,对齐程度对分析结果会起到决定性作用。此外,对于利用掩码进行防护的曲线,利用这些经典方法进行泄露区间定位往往会失败。

3.2 反向传播

在图像处理和识别领域,通过反卷积和导向反向传播可以实现卷积神经网络的可视化,该方法均基于反向传播,但是和训练过程中的方向传播相反,可视化方法是网络输出对网络输入或者网络中间层进行求导(Input-Grad),通过导数来表示输入的每个像素点对结果的影响,导数值越大,代表该像素点所含有效信息越多,对网络分类的影响也越大,该方法可用于能量迹中秘密信息泄露区间定位。针对 N 条训练能量迹 $x = (x_i)_{1 \leq i \leq N}$,每条曲线含有 m 个采样点,计算神经网络损失对输入的导数值:

$$\nabla T_{i,j} = \frac{\partial L_{T_i}}{\partial x_j}, \quad 1 \leq i \leq N, 1 \leq j \leq m \quad (5)$$

这种方法对于神经网络的网络结构没有限制,实现方法简单。文献[6] 中,此类方法被应用于侧信道攻击中,并通过实验证明可以利用导数进行泄露区间的定位。

3.3 加权梯度类激活映射 Grad-CAM

图像处理和识别领域中,为了克服对类别不敏感的缺陷,研究者提出了 Grad-CAM 的方法^[7],该方法可以生成对应标签的热力图,从而完成目标检测。文献[8] 中阐释了高层卷积层输出的特征图包含了输入的具象特征,Grad-CAM 则选取 CNN 最后

一层卷积层输出的若干张特征图进行分析处理从而定位输入的特征区间. Grad-CAM 的思想是首先得到每张特征图所对应的权重, 然后对特征图求加权和来得到热力图. 权重通过特征图的梯度的全局平均来计算:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (6)$$

式中, α_k^c 代表针对类 C , 第 k 张特征图的权重, z 代表特征图的大小, y^c 代表未经过 softmax 之前时类 C 的得分, A_{ij}^k 代表第 k 张特征图中位置为 (i, j) 的像素点的激活值. 由此类 C 的 Grad-CAM 为:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}(\sum_k \alpha_k^c A^k) \quad (7)$$

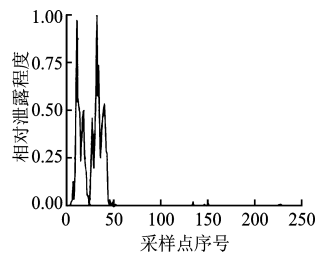
式中, ReLU 为线性整流函数. 这种可视化方法的优势在于不需要对卷积神经网络的架构进行修改, 使得其可以应用于几乎任何基于 CNN 的网络架构.

4 优化方法——WAvg-Grad-CAM

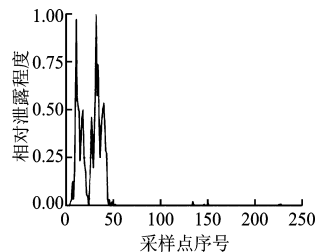
在图像识别中, 底层的卷积层可以捕获图像的边缘角点以及颜色的抽象特征, 较深的卷积层输出的特征图会呈现更具象的特征, Grad-CAM 可视化方法利用这一特性, 对最后一层卷积层输出的整张特征图求取平均值作为该特征图的权重值, 以此来挑选出对该类贡献较大的几张特征图, 从而完成对物体的定位.

模板攻击区别于图像识别主要因为输入的数据集的特征不同, 在图像识别中, 目标物体的特征较为明显, 人眼可分别, 但对于模板攻击所用的能耗或电磁曲线来说, 特征微弱, 对于特定采样时刻点而言, 不同密钥在该采样点功耗值的概率分布会有一部分重叠, 这导致并不是几张特定的特征图就可以实现区间的定位. 而 Grad-CAM 更多地关注于每张特征图的整体贡献, 通过特征图的叠加来进行物体定位. 对于泄露区间定位而言, 当特征不再明显时, 一张特征图所包含具象特征的参考意义也随之减弱, 泄露区间的定位则需要更多地关注在特征图上不同位置的点对结果的贡献大小. 为了说明特征图和权重分布在模板攻击中的对定位的贡献, 采用权重乘以特征图求平均 $\text{avg}(\alpha_k^c A^k)$ 的方式和直接对权重值求平均 $\text{avg}(\alpha_k^c)$ 的方式进行对比, 得到效果如图 1 所示.

可以发现, 去掉特征图直接用导数进行定位和 Grad-CAM 原有方法的定位区间基本相同, 说明位置信息更多的保留在权重中而非特征图中. 针对这个结论, 我们对 Grad-CAM 做了方法改进, 在卷积



(a) 权重与特征图乘积的平均结果图



(b) 权重平均结果图

图 1 梯度乘以特征图平均、梯度平均对比

神经网络训练到具有一定成功率时, 对于一条标签为 C 的能量迹曲线, 卷积层输出 k 张特征图, 其中 k 为该层卷积层所用卷积核的个数, 用未经过 softmax 层的对应标签为 C 的得分对特征图每一个点求取偏导值的绝对值, 先按位置与特征图的激活值相乘来保留每一个位置像素点的权重值的影响大小, 然后通过 ReLU 函数保留对结果有正激励的特征, 最后在卷积核个数的维度上求取平均, 得到每个点的平均权重值, 计算过程如公式(8)所示.

$$\omega_i = \frac{1}{k} \sum_k \text{ReLU}\left(\left|\frac{\partial y^c}{\partial A_i^k}\right| \cdot A_i^k\right) \quad (8)$$

式中, ω_i 是第 i 个点的权重, 表示该时刻的功耗对分类结果的影响大小, y^c 代表未经过 softmax 之前时类 C 的得分, A_i^k 代表第 k 张特征图第 i 个像素点的激活值. WAvg-Grad-CAM 方法可以用图 2 来表示整个可视化过程.

5 实验测试

5.1 定位评估

本文采用一款无安全防护的芯片和能耗曲线采集设备进行相关测试, 向芯片发送传输 8 字节密钥的指令, 利用采集平台捕获该芯片在执行此命令时的功耗曲线.

对于训练集的数据集, 密钥随机生成, 共采取 101 000 条数据, 其中 2 万条作为验证集使用, 测试集随机选取 9 组不同的固定密钥值, 每一组采集 1 000 条数据, 共 11 万条信号曲线用于模板攻击, 来

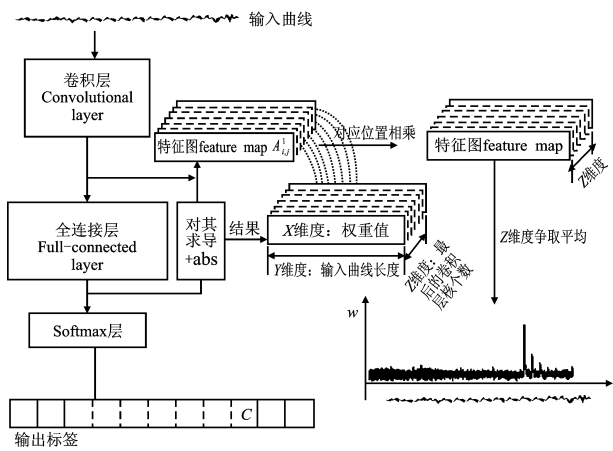
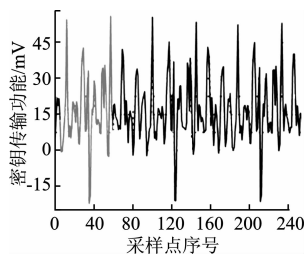
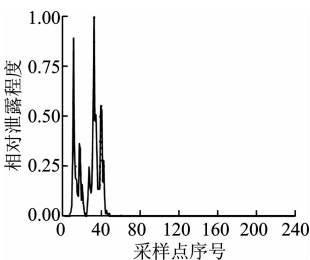


图 2 WAvG-Grad-CAM 方法的基本流程

攻击传输的第一字节的密钥值. 采用 CNN 的网络结构为三层卷积层, 两层全连接层, 输出 256 种标签. 为了方便对可视化准确性地评估, 选择密钥分别传输前 3 字节的执行命令时间段作为攻击数据区间, 共 256 个采样点, 如图 3(a)所示, 其中灰色表示第一字节的传输.



(a) 前三字节密钥传输曲线



(b) WAvG-Grad-CAM 可视化效果图

图 3 改进方法定位准确度分析结果

对采集曲线的训练集进行训练, 待网络收敛且未出现过拟合时停止训练. 利用训练好的网络对测试集数据进行 WAvG-Grad-CAM 分析, 可视化结果如图 3 (b)所示. 可以发现改进后方法 WAvG-Grad-CAM 所做的可视化效果图, 其纵坐标代表该点对结果成功分类的影响大小, 峰值区域定位在了原始曲线传输第一字节的范围中, 即图 3(a)中的灰区域部分, 与期望区间位置一致, 说明利用改进后的方法

可以准确地进行泄露区间的定位.

5.2 方法对比

5.2.1 一阶泄露曲线对比实验

为了说明改进后 WAvG-Grad-CAM 方法相对于 Grad-CAM 等其他常见可视化方法的优势, 下面进一步对相关性分析、反向传播、Grad-CAM 和其改进方法进行对比试验, 探讨这几种可视化方法的效果. 由于测试集采用了固定密钥, 故选取了验证集进行对比试验, 随机两千条预测成功的曲线数据进行相关性分析和可视化分析, 为了方便与相关性分析进行对比, 其他三类方法均被归一化到单位区间, 得到的结果如图 4 所示.

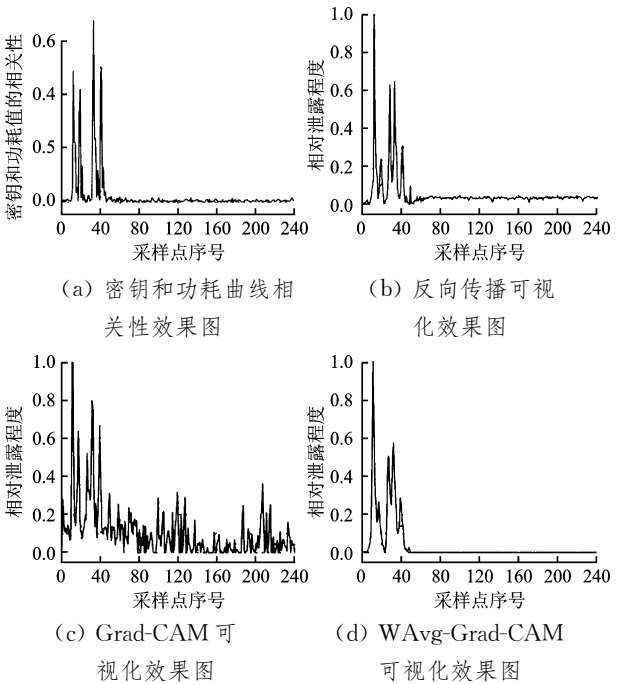


图 4 2 000 条数据可视化效果对比

三种方法定位的区间与相关性定位区间一致, 利用 Grad-CAM 进行可视化得到的泄露分析曲线的噪声较大. 利用反向传播进行可视化和改进的方法定位效果基本一致, WAvG-Grad-CAM 在非泄露区间的曲线相比更为平滑, 噪声较小.

经过上述对比, 对于一阶泄露曲线, 我们可以发现这四种可视化方法都可以对能量迹曲线进行泄露区间定位, 但相关性分析、反向传播可视化方法和 WAvG-Grad-CAM 更能表现网络分类成功所关注的泄露位置, 且 WAvG-Grad-CAM 在非泄露区间相较于其他方法而言, 拥有着噪音小, 曲线平滑的优势. 这使得在可视化分析中, 更容易定位到泄露时刻, 便于进一步的攻击或防护.

5.2.2 二阶泄露曲线对比实验

相较于传统的泄露区间定位方法而言, WAvG-

Grad-CAM 对于高阶泄露也有着优秀的表现. 加密操作的常见的一种防护手段是增加掩码, 掩码的加入可以隐藏加密操作中间敏感值的泄露, 例如在 AES 加密过程中, S 盒的输出往往不用明文进行表示, 而是使用 $S = S_{\text{box}}(d \oplus k) \oplus m_1 \oplus \dots \oplus m_s$, 其中 $m_1, \dots, m_s$ 称为掩码值, 它是每一次执行 AES 加密算法时生成的随机数, 而 S 表示带掩码的 S 盒输出值. 对于高阶泄露曲线需要通过高阶攻击^[9], 例如高阶 CPA 才能进行破解. 而高阶攻击需要在攻击阶段前进行数据准备, 将 mask 的泄露和掩码的 S 盒输出值的泄露进行联合处理, 如求差等操作, 再进行攻击. 而对于基于神经网络的 WAvG-Grad-CAM 方法, 网络可以在掩码未知的情况下, 自动捕获到掩码的位置和带掩码 S 盒输出的泄露位置, 这使得该方法较传统泄露定位方法而言具有明显优势.

以二阶泄露为例, 仅使用掩码 m_1 来进行防护, 带掩码的 S 盒输出值为 $S_{\text{box}}(d \oplus k) \oplus m_1$, 对于传统定位方法而言, 需要在数据准备阶段, 攻击者拿到掩码值来分析相关性, 才能确定掩码的泄露位置和带掩码的 S 盒输出值的泄露位置. 如果这两者的泄露位置未知, 则需要将所有可能点组合进行曲线预处理, 假设曲线共有 n 个采样点, 则联合处理后的曲线则有 $n(n-1)/2$ 个采样点, 当 n 足够大时, 这种联合分析就极其耗时. 对于 WAvG-Grad-CAM 而言, 在实现掩码未知的情况下, 我们可以使用 $S_{\text{box}}(d \oplus k)$ 直接作为标签进行网络训练, 在网络具有一定成功率的时候进行可视化分析, 就可以对两个泄露位置进行精准定位.

为了模拟一阶掩码防护的场景, 我们生成了一万条每条曲线有 50 个采样点的数据, 在第 5 个采样点模拟掩码加载的操作, 针对每一条曲线都随机产生一个 0 到 255 的随机值作为掩码值 m_1 , 即该采样点的值为 m_1 . 在第 25 个采样点模拟带掩码的 S 盒操作, 即该采样点的值为 $S_{\text{box}}(d \oplus k) \oplus m_1 + N(0, 1)$, 每一条曲线的 d 为 0 到 255 的随机值, k 为一字节的固定值, $N(0, 1)$ 为随机噪声. 该曲线其他采样点均为 0 到 255 的随机值, 用这些数据来模拟功耗值, 图 5 展示了其中一条仿真的二阶泄露曲线.

该仿真数据集随机选取其中的 8 000 条作为神经网络的训练集, 1 000 条作为验证集, 1 000 条作为测试集进行训练. 用训练集进行网络训练, 在网络攻击验证集成功率达到 1.26% 时, 对测试集数据进行可视化分析, 结果如图 6 所示.

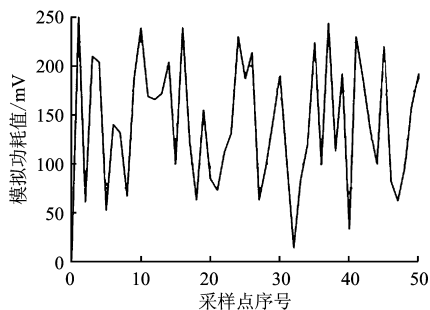
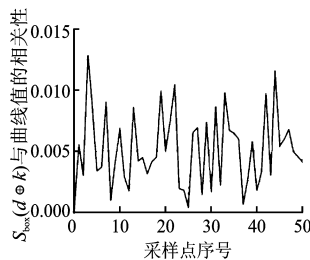
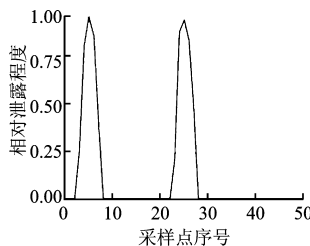


图 5 仿真二阶泄露功耗曲线



(a) 相关性分析效果图



(b) WAvG-Grad-CAM 可视化效果图

图 6 二阶泄露曲线定位能力分析

行可视化的效果图, 可以看到在第 5 个采样点和第 25 个采样点有明显尖峰, 和实际泄露位置一致, 说明该方法对于二阶泄露的曲线也可以精准定位. 这说明在掩码未知的条件下, 即使在成功率较低的网络, WAvG-Grad-CAM 仍然可以精准定位, 且相对泄露程度曲线仍具有光滑噪声小的优点. 对原有 50 个数据进行继续训练, 最后可以达到验证集的成功率为 7.65% 和测试集的成功率为 7.26%, 针对泄露区域进行挑选, 共提出 6 个兴趣点, 采样点序号分别为 {4, 5, 6, 24, 25, 26} 的数据点进行进一步训练, 成功率可以提升到验证集 14.53% 和测试集 11.23%. 可以发现对泄露位置定位, 不仅可以定位泄露的位置, 帮助芯片设计人员迭代设计时进行针对性的防护, 而且可以在攻击阶段降低神经网络的输入维度, 进一步的提升成功率.

6 结束语

本文提出的泄露区间定位方法 WAvG-Grad-

图 6(b)展示了 WAvG-Grad-CAM 对测试集进

CAM,是基于神经网络对侧信道信息的处理,通过对卷积神经网络的可视化,在导数和特征图信息结合时,侧重于保留导数获取的位置信息,从而达到对泄露位置进行精确定位。相较于图像处理和识别领域的可视化方法,该方法拥有噪声小曲线平滑的优势。相较于传统方法,它在泄露情况未知和不进行联合预处理的前提下,可以直接对带掩码实现进行准确定位。该方法在网络成功率较低的情况下,仍然可以定位成功,对该区域进行提取和处理,可以降低网络的输入维度,在网络训练中,使网络训练速度提升成功率提高。此外,对侧信道曲线进行泄露区间定位,也可以给芯片的后续防护提供依据。

参考文献:

- [1] CAGLI E, DUMAS C, PROUFF E. Convolutional neural networks with data augmentation against jitter-based countermeasures: profiling attacks without preprocessing[C]//Proceedings of the 19th International Conference on Cryptographic Hardware and Embedded Systems, Taipei, China; Springer, 2017: 45-68. DOI:10.1007/978-3-319-66787-4_3.
- [2] CHARI S, RAO J R, ROHATGI P. Template attacks[C]//Proceedings of the 4th International Workshop Redwood Shores on Cryptographic Hardware and Embedded Systems, Redwood Shores, CA, USA; Springer, 2003: 13-28. DOI:10.1007/3-540-36400-5_3.
- [3] GIERLICH B, LEMKE-RUST K, PAAR C. Templates vs. stochastic methods: a performance analysis for side channel cryptanalysis[C]//Proceedings of the 8th International Workshop on Cryptographic Hardware and Embedded Systems, Yokohama, Japan; Springer, 2006: 15-29. DOI:10.1007/11894063_2.
- [4] BHASIN S, DANGER J L, GUILLEY S, et al. NICV: normalized inter-class variance for detection of side-channel leakage [C]//Proceedings of the 2014 International Symposium on Electromagnetic Compatibility, Tokyo, Japan; IEEE, 2014: 310-313.
- [5] ARCHAMBEAU C, PEETERS E, STANDAERT F X, et al. Template attacks in principal subspaces [C]//Proceedings of the 8th International Workshop on Cryptographic Hardware and Embedded Systems, Yokohama, Japan; Springer, 2006: 1-14. DOI:10.1007/11894063_1.
- [6] TIMON B. Non-profiled deep learning-based side-channel attacks with sensitivity analysis [J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2019, 2019 (2): 107-131. DOI:10.13154/tches.v2019.i2.107-131.
- [7] SELVARAJU R R, COGSWELL M, DAS A, et al. GRAD-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy; IEEE, 2017: 618-626. DOI:10.1109/ICCV.2017.74.
- [8] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland; Springer, 2014: 818-833. DOI:10.1007/978-3-319-10590-1_53.
- [9] MESSERGES T S. Using second-order power analysis to attack DPA resistant software[C]//Proceedings of the Second International Workshop on Cryptographic Hardware and Embedded Systems, Worcester, MA, USA; Springer, 2000: 238-251. DOI:10.1007/3-540-44499-8_19.

作者简介:

谢欣宜 女,(1993-),硕士,工程师.研究方向为信息安全.

刘 帅 男,(1989-),硕士,工程师.研究方向为信息安全.

王立辉 男,(1982-),博士,高级工程师.研究方向为密码芯片安全.

李 清(通讯作者) 女,(1968-),硕士,教授级高级工程师.研究方向为集成电路设计开发.

E-mail: liqing@fmsh.com.cn.

俞 军 男,(1968-),硕士,教授级高级工程师.研究方向为集成电路设计开发.