基于深度学习的侧信道分析综述

肖 冲 唐 明

(武汉大学国家网络安全学院空天信息安全与可信计算教育部重点实验室 武汉 430072)

摘 要 侧信道分析(SCA)已成为威胁现代密码系统安全的重大隐患,驱使相关防护对策与泄漏检测技术不断完善。受益于神经网络的快速普及与发展,基于深度学习的侧信道分析(DL-SCA)进入快速发展的阶段。深度学习技术的引入在放大侧信道攻击的潜在威胁的同时也降低了其攻击门槛,进而推动侧信道防护与检测技术的革新。本文将从攻击、防护和检测三方面入手,详细介绍基于深度学习技术的侧信道领域创新,总结当前研究趋势,并展望未来的发展潜力。依据逐年的统计数据,本文总结了当前的热点研究方向、分析未来发展趋势并为薄弱研究方向提供可行的技术路线。

关键词 侧信道分析;深度学习;建模侧信道分析;非建模侧信道分析;侧信道防护;泄漏检测

中图法分类号 TP309 **DOI 号** 10.11897/SP.J.1016.2025.00694

A Survey on Deep Learning-Based Side-Channel Analysis

XIAO Chong TANG Ming

(Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072)

Abstract Side-Channel Analysis (SCA) poses a significant threat to modern cryptographic systems, driving continuous advancements in defense mechanisms and leakage detection techniques. In recent years, Deep Learning-based Side-Channel Analysis (DL-SCA) has gained considerable attention, leveraging the self-learning capabilities of neural networks to enhance the effectiveness of these analyses. This paper aims to provide researchers with a comprehensive overview of the integration of deep learning techniques within the context of side-channel analysis. We survey DL-SCA across three key dimensions: attack, defense, and detection. The advancements in these dimensions reflect the evolution of classical side-channel techniques and offer valuable insights into their future trajectories. A primary focus of current research is on profiling attacks, which are particularly well-aligned with the deep learning paradigm due to their inherent characteristics. As a result, deep learning-based profiling side-channel analysis (DL-PSCA) has emerged as a prominent area of study and is widely recognized as one of the most powerful attack methodologies. This work classifies existing DL-PSCA research into four categories based on their research intents: adaptability, generalizability, explainability, and cost-effectiveness. Within the adaptability category, various neural network architectures, loss functions, and evaluation metrics are proposed to enhance the performance of DL-PSCA. The generalizability aspect addresses the critical challenge of bridging the gap between profiling devices and target devices, which has become a trending area of investigation in the field. Explainability is crucial, as it aids classical SCA researchers in understanding the underlying principles of deep learning-based techniques, fostering better collaboration between traditional and modern methodologies. The cost-effectiveness of DL-PSCA aims to minimize the

收稿日期:2023-10-12;在线发布日期:2024-11-01。本课题得到国家重点研发计划(2022YFB3103800)资助。**肖 冲**,博士研究生,主要研究方向为侧信道分析和密码分析. E-mail: xiaoch@whu.edu.cn。**唐 明**(通信作者),博士,教授,主要研究领域为密码学、安全密码芯片设计、抗侧信道轻量防护以及侧信道分析的系统性研究。E-mail: m.tang@whu.edu.cn。

resource overhead associated with model deployment. Beyond DL-PSCA, non-profiling side-channel analysis (NPSCA) also benefits from deep learning techniques. Within the context of NPSCA, there are two primary approaches: DL-driven NPSCA and DL-assisted NPSCA. The former utilizes deep learning techniques to enhance classical NPSCA methods, while the latter focuses on improving NPSCA performance by adapting deep learning techniques for feature engineering purposes. Following the exploration of attack techniques, we turn our attention to the evolution of defense mechanisms. The introduction of deep learning techniques has transformed the landscape of attacks, presenting significant challenges for defense strategies. We present relevant research efforts in two primary areas: defense against DL-PSCA and the design of defense mechanisms that leverage deep learning techniques. Moreover, we discuss leakage detection techniques, encompassing both leakage location and assessment, which have also integrated deep learning methodologies. Our comprehensive analysis provides suggestions for each category by comparing similar works An urgent phenomenon emerged: the studies focusing on attacks, defenses, and detections are markedly unbalanced, with attack research dominating this nascent field. Conversely, this imbalance suggests that defense and detection areas hold substantial potential for future exploration and development. To identify the defining characteristics of this emerging field of SCA, we conducted a chronological analysis of relevant research. Our findings indicate that research on adaptability, generalizability, and DL-NPSCA represent the top three trending areas of investigation. Furthermore, we predict future developments will address three critical aspects: tackling the challenges posed by realistic attack scenarios, demystifying the black-box nature of DL-SCA, and advancing defense and detection techniques. Furthermore, to overcome the technical barriers associated with the underdeveloped defense and detection areas, this work proposes two potential roadmaps from the perspectives of classical side-channel researchers and deep learning researchers, respectively.

Keywords　Side-Channel Analysis (SCA); deep learning; profiling SCA; non-profiling SCA; SCA countermeasure; leakage detection

1　引　言

侧信道分析(Side-Channel Analysis, SCA)兴起于 20 世纪 90 年代中期。当时的密码学家发现，基于硬件或软件的加密系统在运行期间，产生的物理信号(包含功耗、电磁辐射等)可以被攻击者捕获，用于推断加密过程中使用的密钥[1]。学术界普遍认可的侧信道攻击的分类为：非建模侧信道分析(Non-Profiling SCA, NPSCA)和建模侧信道分析(Profiling SCA, PSCA)。两者的主要区别在于：后者在攻击前需要利用部分侧信道样本来构造一个模型或模板。

在侧信道分析技术的发展过程中，经典的机器学习方法，特别是支持向量机，被大量用于建模侧信道分析[3-7]。文献[7]指出，基于机器学习的攻击方法在针对低维数据的攻击中表现优异；而在高维度数据的场景下，该类攻击的有效性会降低。换句话说，基于机器学习的建模侧信道分析(Machine Learning-based Profiling Side-Channel Analysis, ML-PSCA)并未取代传统的建模侧信道攻击方法，攻击者需要依据具体的应用场景进行选择。

随着深度学习在各行各业的迅速普及，侧信道分析领域也迎来了快速发展迭代的阶段。相较传统的基于统计分析或机器学习的侧信道分析技术，基于深度学习的侧信道分析(Deep Learning-based Side-Channel Analysis, DL-SCA)在处理各种复杂场景方面具有更大的潜力，表现出更强的适应性，且对攻击者能力的要求较低。基于这些优势，DL-SCA 已经成为目前最具威胁性的侧信道分析方法。

文献[8]详细总结了基于传统机器学习的侧信道分析工作，但只涵盖了少量基于深度学习的早期工作。之后，文献[9]侧重于基于深度学习的侧信道分析工作，将 DL-SCA 划分了多个攻击阶段，对每

个阶段既有的相关工作进行了详尽描述。区别现有综述,本文将从传统的侧信道领域的三个方向(攻击、防护和检测)切入,探讨深度学习技术在侧信道领域的全面融合及发展前景。此外,依据各个方向的统计数据,本文总结 DL-SCA 的研究热点、预测未来的发展趋势并为薄弱研究提供可行技术路线。

基于深度学习的侧信道攻击的实际应用非常广泛,除了典型的能量消耗和电磁辐射,还可以利用诸如光学信号、声学信号和微架构状态等侧信道信息来推测一些用户隐私,如用户的按键[10]、运行的应用[11]等。本文的研究范围未包括这些广义上的侧信道攻击。更具体而言,本文调查了基于深度学习对嵌入式设备密码系统进行侧信道分析的研究进展。

本文第 2 节为侧信道分析和深度学习相关的背景知识,按照攻击、防护和检测三个维度;第 3 节给出 DL-SCA 的相关工作分类;第 4 节和第 5 节分别为基于深度学习的建模侧信道分析和非建模侧信道分析的现有工作介绍与分析;第 6 节和第 7 节分别介绍深度学习在侧信道防护设计和泄漏检测中的研究现状;第 8 节基于 DL-SCA 的数据统计,总结当前的研究热点,发展趋势并指出薄弱研究的可行路线;第 9 节对全文进行总结;此外,附录 A 列出本文使用的英文缩写的对照表。

2 背 景

2.1 侧信道分析

侧信道分析是一种利用设备在处理数据的过程中产生的物理信号(如能耗、电磁辐射等)来还原相关秘密信息的攻击手段。下面的小节将介绍经典的侧信道攻击技术、防护策略以及侧信道泄漏的检测方法。

2.1.1 攻 击

一般而言,侧信道攻击可以分为两类:建模侧信道分析和非建模侧信道分析。

建模侧信道分析包括两个阶段:建模阶段(算法 1 的第 1、2 行)和攻击阶段(算法 1 的第 3、4 行)。在建模阶段,攻击者利用从建模设备上采集的时序侧信道信号(又称侧信道曲线)构建模板。之后,构建的模板在攻击阶段被用于匹配从目标设备上采集的曲线,以此推理出目标设备的秘密数据。典型的建模侧信道分析有模板攻击(Template Attack)[2]和随机攻击(Stochastic Attack)[12]。其中,模板攻击构建基于多元高斯分布的模板,从信息论角度被认为是最有效的侧信道攻击[11,13]。需要指出,传统的建模侧信道分析基于一个严格的攻击者能力假设:建模设备与目标设备完全一致,且受控于攻击者。

算法 1. 建模侧信道攻击流程

输入:建模侧信道曲线 X_p、明文 P_p 和密钥 K_p,攻击侧信道曲线 X_a 和明文 P_a

输出:攻击还原的密钥 K_a

1. 计算建模曲线对应的中间变量 $Y_p=f(K_p,P_p)$
2. 构建模型 $M()$:$X_p\rightarrow Y_p$
3. 将模型应用于攻击曲线 $Y_a=M(X_a)$
4. 还原密钥 $K_a\leftarrow f^{-1}(Y_a,P_a)$

不同于 PSCA 的二阶段攻击模式,非建模侧信道分析直接利用依赖密钥的敏感中间变量与侧信道泄漏之间的关系。如算法 2 所示,该类攻击利用特定区分器对所有密钥假设进行筛选,最高得分的密钥假设即认为是正确的密钥。常用的 NPSCA 包括差分功耗分析(Differential Power Analysis,DPA)[14]、相关功耗分析(Correlation Power Analysis,CPA)[15]和互信息分析(Mutual Information Analysis,MIA)[16]。DPA 分析目标设备的功耗模式是否依赖密钥假设,CPA 和 MIA 分别通过计算侧信道泄漏和猜测中间值之间的相关性或互信息区分正确密钥假设。传统的非建模侧信道分析存在局限性,例如:CPA 仅适用于侧信道泄漏和泄漏模型之间存在线性关系的情形[15],而 MIA 预估互信息的计算代价较高。

算法 2. 非建模侧信道攻击流程

输入:侧信道曲线 X,对应的加密明文 P

输出:攻击还原的密钥 k^*

1. FOR $k\in\mathcal{K}$
2. 计算中间变量 $S_k=f(k,P)$
3. 计算区分器得分 $D_k=\mathrm{Dist}(X,S_K)$
4. $k^*=\arg\max_k(D_k)$
5. RETURN k^*

2.1.2 防 护

为抵御传统侧信道攻击,当前的主流防护方案包含隐藏(Hiding)和掩码(Masking)两类[17]。其中,隐藏方法旨在降低物理信号的信噪比(Signal-to-Noise Ratio,SNR),使攻击者难以从侧信道曲线中提取有用的信息。换言之,配备了隐藏技术的加密设备仍会执行与无防护设备相同的运算操作、

产生相同的中间结果，但攻击者更难获得有用的信息。常用的隐藏技术包括伪操作插入、随机延迟(Random Delay)和乱序(Shuffling)等。

掩码防护[18-19]通过引入随机数 r 将敏感中间变量 x 拆分为多个分量，在确保运算结果正确的前提下让拆分的分量分别参与运算，使得产生的侧信道信号与原始中间变量相互独立。最常用的掩码为基于模加(异或)运算⊕的布尔掩码，将中间变量 x 拆分为被掩值 $x\oplus r$ 和掩码 r。此外，还有基于模乘运算的乘法掩码以及混合两种运算的算术掩码。

在对带掩码防护的密码实现进行侧信道分析时，攻击者需要进行一个额外的泄漏组合操作[20-21]，即将所有分量对应的侧信道泄漏进行组合，得到与原始中间变量存在关联的高阶泄漏。最为广泛应用的组合函数有乘积组合[20]和绝对差值组合[21]。

2.1.3　检　测

侧信道的泄漏检测有两个角度：泄漏定位和泄漏评估。其中，泄漏定位，又称 PoI(Point of Interest)定位，主要作为攻击前的特征点选择。传统的定位方法侧重于数据分布的比较，例如 SOSD[22]和 NICV[23]，依据泄漏变量的值将样本点分为数类，计算类间的差异度量来衡量该样本点是否为泄漏点。

泄漏评估是指一种评估加密设备泄漏量的方法论，通常依赖于信息论或统计分析。基于信息论的方法(如熵和互信息)用于对泄漏进行定量分析[24-25]，而基于统计分析的方法则基于某些假设和模型来估计泄漏量。例如，工业界最常用的泄漏评估方法 TVLA(Test Vector Leakage Assessment)[26]建立于 Welch's t-test 统计分析之上，而进行 t-test 的前提是样本服从正态分布。类似地，卡方检测也基于正态分布假设。

2.2　深度学习

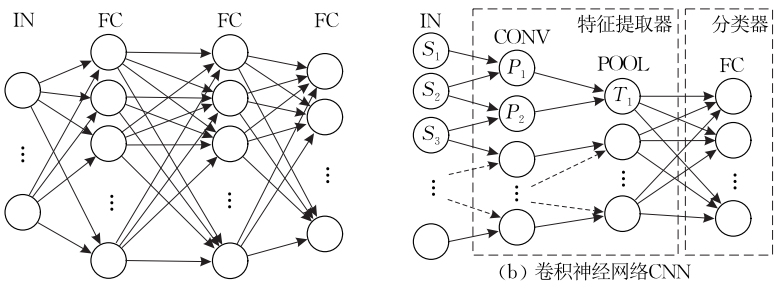
机器学习是一种采用计算机算法和统计模型，从数据中自主学习和提高的方法。经典的机器学习方法，如决策树[27]、随机森林[28]和支持向量机[29]在实际应用中非常普遍。然而，随着深度学习技术的发展，这些经典的机器学习方法正逐渐被取代。

深度学习是一种致力于大规模数据的自动学习和特征提取的机器学习技术[30]。与经典的机器学习算法的主要差异在于：深度学习依靠深度神经网络。深度学习已经在众多领域取得了重大进展，包括图像识别[31]、语音识别[32]和自然语言处理[33]等。事实上，深度学习也是一种机器学习，但本文进行了概念上的切割，将机器学习特指经典的机器学习算法。

2.2.1　神经网络的构成

神经网络通常由多个层次组成，每一层包含数个神经元，这些层次之间的连接具有权重，而整个网络的结构和参数由这些层次和连接组成。几种常用的网络层有：(1) 全连接(Fully-Connected，FC)层：该层每个节点都与前一层的每个节点相连；(2) 卷积(Convolutional，CONV)层：使用一个或多个卷积核来对输入向量以滑动窗口的模式进行卷积运算；(3) 池化(POOL)层：用于降低数据维度；(4) 激活函数(又称激活层)：对其他网络层的输出进行非线性映射。常见的激活函数包括 SELU[34]和 Softmax 等。

在基于深度学习的侧信道分析领域中，两类广泛使用的网络结构为多层感知器(Multi-Layer Perceptron，MLP)和卷积神经网络(Convolutional Neural Network，CNN)，其结构示例如图 1。MLP 由数个全连接层和激活函数叠加而成，结构简单但可训练参数较多。然而，卷积层的可训练参数取决于卷积核的尺寸和大小，当输入数据的维度较大时，效益更高，因此卷积神经网络较 MLP 更通用。典型的 CNN 由两部分构成：特征提取器和分类器，其中提取器由卷积层、池化层和激活层构成，而分类器一般由多个全连接层构成。



(a) 多层感知器MLP　　(b) 卷积神经网络CNN

(CONV和POOL层的核大小/步长：2/1，2/2)

图 1　MLP 和 CNN 的网络结构示例

2.2.2 神经网络的训练

根据神经网络训练期间使用的数据是否有标签,深度学习可以分为有监督的和无监督两类。无论监督与否,深度学习的核心在于通过反向传播[35]计算梯度,并使用优化算法更新网络参数,从而不断降低损失度量。常见神经网络优化算法有随机梯度下降 SDG[36]和自适应时刻估计方法 Adam[37];常用的损失函数有交叉熵(Cross Entropy,CE)和均方误差(Mean Square Error,MSE)。

(1)有监督学习

有监督学习的训练目标是使网络的输出逼近真实标签,因此,其损失函数用于衡量网络预测输出与真实标签之间的差异。在侧信道领域中,泄漏数据的推测通常被视为分类任务,因此交叉熵这一多分类任务的损失函数被广泛采用。

有一种特殊的有监督学习方法叫作对比学习,它在训练中利用标签有选择性地抽取样本,其目的是通过比较不同样本之间的相似性来学习有用的特征表示。主要思想是靠近正样本(同类别样本)的特征表示的同时远离负样本(不同类别样本)的特征表示,以便更好地捕获数据的结构和特征。对比损失用于衡量特征表示间的差异,常采用欧氏距离或余弦距离。经典的对比学习网络包括孪生网络[38]和三重网络[39],如图 2(a)和(b)所示。调研结果表明,尽管孪生网络和三重网络均已应用于 DL-SCA 领域,但相关研究仅各有一篇。

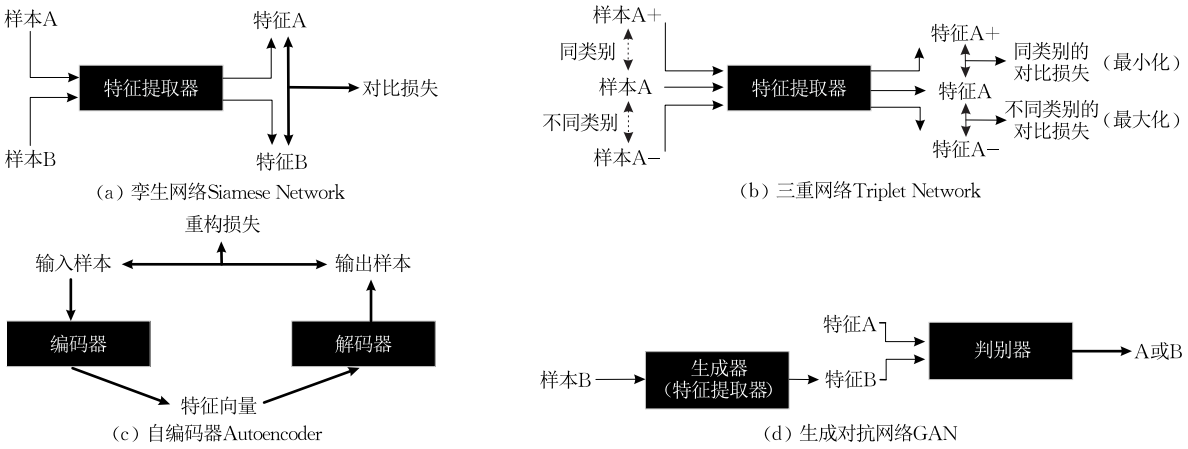


图 2 应用于 DL-SCA 的四类特殊网络

(2)无监督学习

无监督学习的目标是利用未标记的数据训练模型。目前已应用于 SCA 领域的无监督学习方法有重构任务和生成任务。

在重构任务中,模型被要求从输入数据中生成输出,然后与原始输入进行比较,以衡量生成输出的质量,通常使用 MSE 损失函数。典型的基于重构任务的网络结构有自编码器 Autoencoder[40],如图 2(c)所示。

生成任务旨在生成新的数据样本,这些样本在分布或特征上与训练数据相似。典型的生成模型有生成对抗网络(Generative Adversarial Network,GAN)[41],如图 2(d)。GAN 的训练是一场动态博弈:判别器需要区分特征是源自 A 还是 B,而生成器则试图产生与特征 A 尽可能类似的特征 B,使判别器无法区分特征的来源。目前,GAN 已被用于数据增强和迁移性研究。

(3)强化学习

强化学习[42]是一种通过与环境交互,依据奖励或惩罚来调整行为的学习框架。与监督学习提供明确答案和无监督学习自我发现规律不同,强化学习旨在最大化长期奖励,因此在复杂、动态环境中具有较高的灵活性和适应性。在侧信道领域内,强化学习的用途较为广泛,例如超参数的调节、泄漏定位以及防护设计等。

3 深度学习和侧信道分析的结合

目前,深度学习正逐渐融入侧信道领域的各个方面,涵盖了攻击、防护和检测三个方向。本节对 DL-SCA 的相关工作进行类别的划分,并介绍相关的数据集、预处理方法以及评估指标。

3.1 基于深度学习的侧信道分析 DL-SCA

深度学习技术最先应用于建模侧信道分析,后来扩展到非建模侧信道分析,此后,基于深度学习的侧信道防护和检测技术也开启了技术的迭代。本文整合了基于深度学习的侧信道分析领域,从攻击、防护和检测三个方向总结现有工作,列出框架图,如图 3 所示。

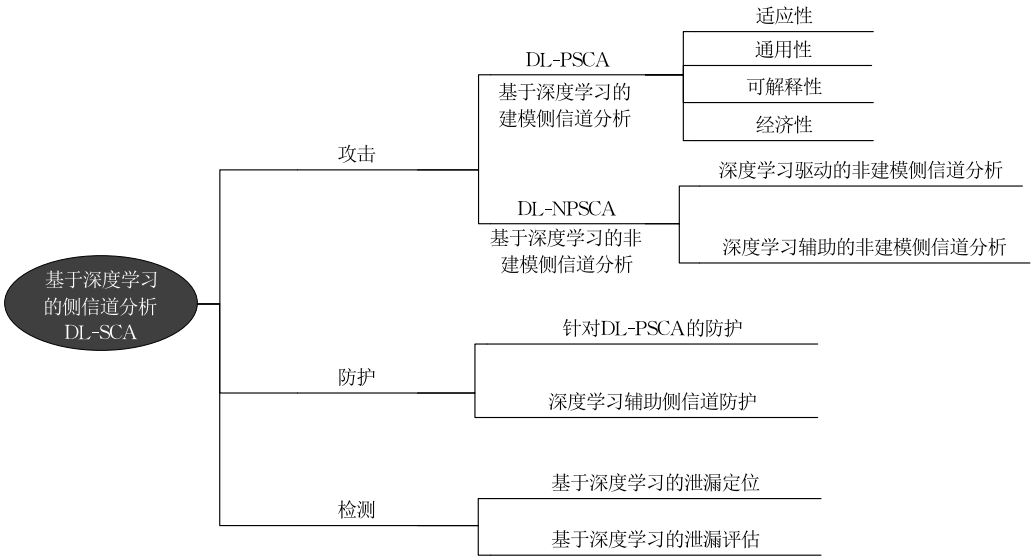


图 3　现有 DL-SCA 研究的分类框架

3.1.1　攻　击

类似传统 SCA，DL-SCA 也划分为：基于深度学习的建模侧信道分析（DL-PSCA）和基于深度学习的非建模侧信道分析（DL-NPSCA）。本文对 DL-PSCA 和 DL-NPSCA 研究采用不同的细分标准。由于 DL-PSCA 在 DL-SCA 领域的研究中占据主导地位，本文依据文献的研究目标，将相关工作细分为以下四个部分。

（1）适应性。指为将深度学习与侧信道领域深度融合而开展的工作。

（2）通用性。指为使训练好的模型在非建模设备上拥有较好的泛化能力而进行的尝试。

（3）可解释性。涉及剖析基于深度学习的侧信道分析的实际原理。

（4）经济性。指以最小资源投入获得最佳性能，通过有效利用计算资源、存储和时间，实现成本效益最大化。

经典的基于深度学习的建模侧信道分析流程类似于算法 1，不同之处在于构建的模型为训练的深度神经网络。此外，深度神经网络也被用于对侧信道曲线进行特征工程来辅助传统建模侧信道分析。

现有的关于 DL-NPSCA 的研究较为有限。根据深度学习技术的应用方式，可将其粗略分为两类：深度学习驱动的非建模侧信道分析和深度学习辅助的非建模侧信道分析。

3.1.2　防　护

自深度学习技术应用于侧信道分析以来，研究者开始反思现有防护方案的安全性，这也进一步推动了侧信道防护设计的进步。伴随 DL-SCA 的兴起，现有防护方案能否有效应对以及如何设计针对性的防御措施成为亟待解决的关键问题。本文将相关防护研究归为两类：针对 DL-PSCA 的防护和深度学习辅助侧信道防护。前者针对 DL-PSCA 提出新型防护，后者借用深度学习辅助设计侧信道防护。

3.1.3　检　测

与传统的侧信道泄漏检测类似，基于深度学习的侧信道泄漏检测包含：基于深度学习的泄漏定位和基于深度学习的泄漏评估。

3.2　数据集

区别于传统的侧信道分析依据需求采集曲线或使用模拟曲线，当前绝大多数 DL-SCA 工作的评估实验都基于公开数据集展开。一方面，具有高质量数据的公共数据集可以提供可靠的实验结果，提高研究成果的稳健性。此外，使用公开数据集，有助于验证和复制以前的研究结果，提高后续研究的可信度。

常见的用于 DL-SCA 的公开数据集有 ASCADf（采自于受一阶掩码保护的 AES-128 实现）[43] 和 DPAv4.1①（虽然该数据集对应有掩码防护的实现方案，但通常被当作已知掩码假设下的无防护实现）。文献[9]提供了一个用于侧信道分析的公开数据集清单（见文献[9]表 1）。在此不再赘述，仅补充三个用于研究模型可迁移性的数据集[44]。

（1）XMEGA。该数据集采集了 8 个 Atmel XMEGA 128A1U 的微控制器的功耗曲线，算法实现为无防护的 AES-128。8 个设备的加密密钥的第一个字节分别被设置为 0x01、0x02、0x03、0x04、

① https://dpacontest.telecom-paris.fr/v4/42_doc.php

0x05、0x06、0x07 和 0x08。每个设备采集了 30 000 条曲线，其中 25 000 条用于训练，5000 条用于攻击，每条曲线有 500 个特征点。

(2) XMEGA_EM。该数据集采集了 8 个 Atmel XMEGA 128A1U 的微控制器的电磁辐射曲线，探针位置存在人为误差，算法实现为无防护的 AES-128。8 个设备的加密密钥的第一个字节分别被为 0x01、0x02、0x03、0x04、0x05、0x06、0x07 和 0x08。对于每个设备的子数据集，有 25 000 条用于训练，5000 条用于攻击，每条曲线有 1500 个特征点。

(3) SAKURA_AES。样本采自无防护的 AES-128 硬件实现。侧信道曲线是从 3 个 SAKURA-G 评估板上获得的，芯片型号为 Xilinx Spartan-6。3 个设备的最后一轮密钥(第二字节)分别是 0x21、0xCD 和 0x8F。对于每个子数据集，有 90 000 条曲线用于训练，10 000 条用于攻击，每条曲线有 1000 个特征点。

3.3 预处理

数据预处理指对采样的生数据进行清洗和整理的过程。在基于深度学习的侧信道分析的背景下，公开数据集已经过初步清洗和整理。在此基础上，两类最常用的数据预处理方法有缩放(scaling)和标准化(standardization)。前者指将数据的每个维度的值进行缩放，使其落在特定区间内，推荐的范围有[0,1](又称归一化)和[−1,1]。后者的目标是使数据集中每个维度的特征服从标准正态分布。一般的方法是估计均值和标准差，然后从每个特征中减去均值，并将结果除以其标准差。区别于这两种在特征维度进行的垂直预处理，文献[45]提出在样本维度进行水平预处理，即对每条曲线实例进行缩放或标准化。

3.4 评估指标

在人工智能领域，分类任务的常见评价指标(如准确率和召回率)并不适用于侧信道分析。这是因为侧信道攻击的效果不是通过单次攻击的准确性来评估的，而是通过成功恢复密钥所需的曲线条数来衡量的。文献[43]提出了多个评估指标，其中，猜测熵(Guessing Entropy，GE)是目前最广泛认可的 DL-PSCA 性能指标。具体来说，GE 衡量攻击者根据从 n 条攻击曲线$\{x_i\}_{1\le i\le n}$获得的信息来识别正确密钥的能力。

攻击者使用一个神经网络模型 M 来评估曲线 $x\in\{x_i\}_{1\le i\le n}$，得到候选密钥的概率向量 $\boldsymbol{d}=[d_0,d_1,\cdots,d_{\|\mathcal{K}\|}]$，如式(1)所示。$k$ 是要恢复的密钥的变量，而$\|\mathcal{K}\|$是密钥空间的大小。GE 表示正确密钥在所有密钥假设中的期望排序，可以通过计算正确密钥 k^* 在所有猜测密钥中的平均位置来得到，如式(2)。

$$d_i=M(x)[i] \tag{1}$$

$$\mathrm{GE}=\frac{1}{n}\sum^{x_n}\sum_{k}^{\|\mathcal{K}\|}(d_{k^*}>d_k) \tag{2}$$

除开 GE，另一个在 DL-SCA 中常用的指标为网络中可训练参数的数量，它表示神经网络模型的复杂度，在一定程度上表示训练代价。

4 基于深度学习的建模侧信道分析 DL-PSCA

建模侧信道分析与深度学习的结合，最显著的体现是攻击方法的创新。有监督学习中的模型训练和模型预测，分别对应于传统建模侧信道分析的建模和攻击阶段。现有研究表明，基于深度学习的建模侧信道分析优于传统的建模侧信道分析方法[46-47]。Maghrebi 等人[46]选择无防护和有防护的 AES 硬件与软件实现作为攻击目标，对比了经典模板攻击和多种基于深度学习的建模侧信道分析方法的攻击性能，证实了 DL-PSCA 的显著优势。

为进一步推进深度学习在建模侧信道分析领域的应用，研究人员开展了四类面向不同目的的研究工作，接下来将展开介绍。

4.1 面向适应性

在侧信道分析领域应用深度学习技术时，应考虑侧信道领域专有的特性，进行细节或方向的调整，以更好地适应侧信道攻击这一任务，而非简单地将其视作一般的分类任务。适应性的调整覆盖了网络架构、损失函数和评估指标、超参数以及训练数据四个方面。

4.1.1 网络架构的选择与设计

无论是基于深度学习的图像领域，还是文本领域，设计高效的神经网络架构是一个长期目标。显然，在利用深度学习进行侧信道分析时也需因地制宜，挑选适合的网络架构。从简单的多层感知器 MLP，到目前最广泛应用的卷积神经网络 CNN，再到 Transformer，关于神经网络架构的选择与设计一直是 DL-PSCA 的重点。

(1) MLP 模型在侧信道领域的应用

① MLP 攻击无侧信道防护的实现：Martinasek

等人[48]首次使用 MLP 恢复无防护 AES 实现的加密密钥。之后，研究人员试图结合不同的特征处理技术来提高攻击性能，例如：平均曲线[49]、小波变换[50]和主成分分析[51]。然而，文献[46]指出：当使用原始曲线作为输入时，可以取得更好的性能。尽管如此，预处理或特征工程技术对 DL-PSCA 性能的影响依然不可忽视，攻击者应在不同场景下自主选择合适的方法。

②MLP 分析带掩码防护的实现：早期的研究基于一个假设：攻击者在建模时已知随机掩码。Gilmore 等人[51]使用 MLP 分别还原掩码和被掩值。随后，Martinasek 等人[52-53]也提出了类似的方法，使用 MLP 攻击公开数据集 DPA contest v4.1 和 v4.2(见本文第 6 页脚注①)——这两个数据集的密码实现含有轻量掩码防护 RSM[54]。这类工作的缺点显而易见，已知随机掩码的建模攻击等同于分别攻击两个无防护的中间变量。相反，基于未知掩码假设，即直接还原 Sbox 的原始输出，从而推测加密密钥的黑盒攻击具有更大的威胁性[46]。

③基于 MLP 结构的创新：为减少网络层数并加快学习阶段，Pfeifer 等人[55]提出了一个名为“spread”的网络层。其基本思想是将神经元的输出转化为空间编码的信息，代价是维度的增加。

(2)CNN 模型在侧信道领域的应用

近年来，CNN 正逐渐取代 MLP，成为 DL-PSCA 最常采用的网络模型，其原因主要有以下两点：

①卷积层的可训练参数较少，网络更轻量。

②CNN 具有移位不变(shift-invariant)的特性，使之能够更好地应对没有对齐的侧信道曲线。

(i)CNN 在 SCA 中的基本发展脉络。Maghrebi 等人[46]首次将CNN架构引入SCA领域。随后，Cagli 等人[56]提出了一个深度 CNN 架构，结合数据增强，利用 CNN 的移位不变性成功攻击带有隐藏防护措施的密码实现。为更好地设计网络架构以适应不同的实现，Zaid 等人[57]系统地研究了多个公开数据集的网络结构，提供了构建高效 CNN 架构的方法论。在这一基础上，Wouters 等人[45]继续该研究，纠正此前的错误观点，并在保持网络的性能水平的前提下进一步压缩网络复杂度。

(ii)二维 CNN。由于侧信道曲线是时间序列，前面提到的 CNN 研究都是针对一维 CNN 展开的，部分研究者试图跳出当前的思路，转向二维 CNN 网络。Yang 等人[58]提出将一维曲线转换为时频谱图。Hettwer 等人[59]研究了多种转化为二维图像的编码方法，并指出使用二维 CNN 能大幅度减少获得正确猜测密钥所需曲线条数。然而，就 ASCADf 数据集而言，当时显著的性能现在看来并不突出(需要大约 275 条攻击曲线使 GE<2)，这一结论同样适用于其他数据集。与此同时，二维 CNN 带来了指数级增长的计算代价。本文认为，虽然对曲线进行升维的方法具有新颖性，但并非主流，其成本效益相对较低。

(iii)基于建模侧信道分析特性进行的攻击优化。加密过程中的领域知识 DK(明文或密文)，也可以用于提升 DL-PSCA 的性能。文献[60]和[61]建议使用这些领域知识来辅助分类器执行分类决策，如图 4 所示。在先前的研究中，通常将一个字节大小的中间变量作为整体攻击目标。然而，还可以同时且独立地攻击字节中的每个比特。基于此，Zhang 等人[62]将泄漏的中间值字节编码为比特标签，等价于将 256 分类任务转化成 8 个独立的二分类任务。应用比特编码最直接的优势在于减少了输出层的维度从而降低了网络复杂度，另外，这种比特编码能够避免数据类别的不平衡，具体可以参考 4.1.4 节。为进一步提高 DL-PSCA 的性能，Won 等人[63]提出将多种 SCA 特征提取方法有机结合，并引入多尺度卷积神经网络 MCNN。该方法使用多种技术对输入曲线独立进行特征提取，然后将提取得到的特征进行拼接，以供 CNN 进行分类。结合前文给出的文献[46]对特征提取的实验结论，本文再次强调特征工程在 DL-PSCA 中的作用需要根据实际的攻击场景进行评估，而不能一概而论。此外，在先前对具有掩码保护的曲线进行攻击的研究中，并没有明确地考虑到多个泄漏点的组合，而是将组合方式交由神经网络自主学习。然而，这种方法在训练数据较少的情况下很难学习到有效的泄漏组

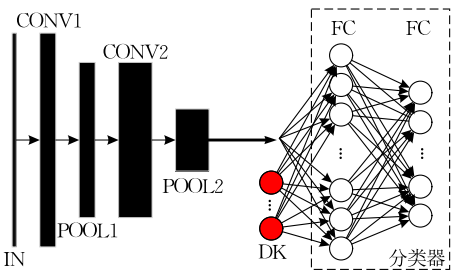


图 4　文献[60]中提出的网络结构

合表示,甚至可能导致网络完全不收敛。为解决这一局限,Cao 等人[64]建议在特征提取器和分类器之间增加一个联立层,即对特征向量求外积,以辅助网络快速建立高阶泄漏和标签之间的联系。结果表明,联立层对于小型网络和训练集较少的情况下表现优异,能将恢复 ASCADf 密钥的所需的攻击曲线降至 100 以内。

(3)新型网络结构的探索

虽然 CNN 在 DL-PSCA 任务中十分有效,其他网络结构的探索也从未停止,比如 LSTM[46]和残差网络(ResNet)[65]。最近,Transformer[66]这一革新了自然语言处理领域的神经网络架构,也被引入了侧信道领域。Hajra 等人[67]指出,Transformer 擅长捕捉侧信道曲线中相距较远的泄漏点之间的依赖关系,在攻击有掩码防护且去同步化①的数据集时有显著优势。为充分利用深度学习技术,执行端到端的建模侧信道攻击,即直接使用未筛选泄漏区域的原始采样曲线进行建模和攻击,Lu 等人[68]设计了一个由多级编码器、注意力机制和分类器组成的自定义网络架构。这些新型网络架构旨在应对特定的攻击场景,鉴于当前广泛采用的 CNN 在许多侧信道任务中均表现出色,这些新型网络并未受到广泛关注。

回顾 DL-PSCA 的网络结构选择与设计,其重心逐渐从简单实用向轻量高效倾斜,而近年来开始朝着定制化方向发展。众多文献的实验表明,基础的 MLP 和 CNN 已经能够很好地满足当前建模侧信道分析的需求。因而,未来网络选择的整体基调不会有大的变动,而面向具体场景的定制化网络结构或架构会不断地被提出。

4.1.2 定制损失函数/指标

(1)定制损失函数。当前,DL-PSCA 被视作分类任务,训练时最常使用交叉熵损失函数。由于侧信道分析与一般分类任务存在差异,研究人员提出了一些适用于不同场景的自定义损失函数,见表 1。Zaid 等人[69]提出了与侧信道分析的目标直接相关的排名损失 RrL。针对数据不平衡的场景(见 4.1.4 节),研究者引入了基于交叉熵比 CER[70]和焦点损失比 FLR[71]的损失函数。文献[72]的实验表明,针对 SCA 的损失函数 CER 表现非常好,在大多数评估设置中都优于其他损失函数(FLR 未列入评估)。此后,Zaid 等人[73]提出了集成损失函数,通过整合多个模型的独立预测结果来弥补单一模型局限性(即集成学习)来提升侧信道攻击的性能。值得一提的是,即便研究者提出了多种针对侧信道分析的损失函数并通过实验证实其相较 CE 的优势,但当前最广泛使用的仍为 CE 损失函数。一方面,研究者普遍认为:交叉熵损失函数是最为保险的选择。另一方面,文献[74]从信息论的角度证明选择交叉熵损失函数来驱动训练的合理性。

表 1 DL-PSCA 损失函数

损失函数	应用场景	说明
交叉熵 CE Cross Entropy	分类任务	分类任务最常用的损失函数。
排名损失 RrL[69] Ranking Loss	DL-PSCA	最小化正确密钥在所有密钥假设中的排名,使攻击成功率最大化。
交叉熵比 CER[70] Cross Entropy Ratio	数据 不平衡	正确密钥的交叉熵和其余密钥假设的交叉熵的期望的比。通过引入错误密钥的交叉熵以应对训练数据不平衡的场景。
焦点损失比 FLR[71] Focal Loss Ratio	数据 不平衡	基于 CER 的改进,对更难学习的样本分配更高的焦点比重。
集成损失 EL[73] Ensembing Loss	集成 学习	用于集成学习的自定义损失。

(2)定制评估指标。除了训练本身使用的损失函数之外,在模型训练的过程中,还需要时刻关注模型的性能,防止模型因过度拟合训练集而对未参与训练的攻击数据的分类效果下降。因此,在训练过程中,选择适当的性能指标来评估当前模型是必要的。当观察到模型开始表现出过拟合迹象时,提前停止训练是一种有效的策略[75]。然而,深度学习指标并不能全面地反映侧信道攻击的性能[76]。对此,研究者进行了深入研究,试图提出或改进现有的评估指标。文献[77]通过测量转移到输出层的互信息量,来确定网络在哪个训练节点的模型泛化能力最好,在此停止训练。除互信息外,文献[78]提出了一个由成功率[79]衍生的在线评估指标,也可以用来指导训练的提前停止。显然,衡量 DL-PSCA 性能的公认指标 GE 是评估模型性能的最佳指标,但它的计算代价较高。为了解决这个局限,Perin 团队[80]提供了一个快速的计算猜测熵的方法。同损失函数一样,即便性能指标也在不断优化,但大多数研究者在评估时仍直接应用 GE。GE 通过计算多次攻击中正确密钥的排名的平均值来估计真实值,而均值会受到一些奇异值影响导致较大误差。为更好地评估攻击性能,本文建议按照文献[81]的方法来计算猜测熵:用中位数代替算术平均值,以此减少因随机

① 去同步化由文献[36]提出,通过将对齐的曲线左移随机相位来模拟采样抖动(jittering),事实上也能用于模拟隐藏防护随机扰动(Random Delay)。

性造成的指标不稳定。

4.1.3　超参数调优

除了损失函数以外，其他的网络训练超参数对深度学习的贡献也不可忽视。超参数的调优是一项工作量繁重的任务[82-83]，在 DL-PSCA 中也不例外，目前的研究主要体现在两个方面。

(1) 调优方法的研究。Perin 团队[84]在预先定义的范围内进行了随机搜索，以构建多个服务于集成学习的深度学习模型。Rijsdijk 等人[85]提出使用强化学习来调整卷积神经网络超参数。Wu 等人[86]研究了基于贝叶斯优化的自动超参数调整。

(2) 特定超参数的定量分析，为超参数的调整提供了指导。Weissbart[87]分析了 MLP 的层数、神经元数目和激活函数，而文献[88]和[89]则分别研究了权重初始化和优化算法在 DL-PSCA 中的影响。文献[89]得出结论：Adam 和 RMSprop 优化算法对短训练阶段效果最好，而 Adagrade 对长训练阶段或较大的模型效果最好。

在 DL-PSCA 中，超参数调优并非研究的重点。一方面，大多数学者并未专注于寻找最优的参数配置。另一方面，当前所使用的网络架构相对简单，使用随机搜索方法足以满足大多数情况。因此，超参数调优在当前以及未来的 DL-PSCA 研究中并非热点问题。

4.1.4　数据平衡

理想情况下，训练集的各个类别的样本数目应该相等或相近。若不同类别的样本数存在显著差异，则称为数据不平衡。在 DL-PSCA 中，类别标签可以是选定的中间变量的数值本身(Identity，ID)，也可以是对应的汉明重量(Hamming Weight，HW)或汉明距离(Hamming Distance，HD)。此外，还可以使用文献[62]提出的多分类攻击框架，将 ID 标签编码成比特标签。

当加密明文随机时，对采集的曲线使用 ID 标签或比特标签进行类别划分不会导致数据失衡，而 HW 或 HD 标签对应的样本划分则会使类间样本数目呈现显著差异，这一问题最早由 Cagli 等人[56]指出。为解决这一问题，数据平衡技术 SMOTE[90]被引入 DL-PSCA[76,91]。SMOTE 通过合成新的少数类样本以增加其数量，改善模型在少数类别上的性能。文献[91]指出：平衡训练数据是进行 DL-PSCA 的必要步骤。

尽管 DL-PSCA 在使用 HW 或 HD 标签时，客观存在数据不平衡这一问题，但大多数工作并未进行数据平衡处理，得到的模型依旧健壮。我们将这一现象归因于两点：(1) 训练样本的特征显著，网络能够轻易收敛；(2) 样本总量相对较多。本文建议，当侧信道曲线的尺寸过大或样本量较少时，可以将数据平衡纳入训练数据的准备步骤。

4.2　面向通用性

通用性是指一个模型在未参与训练的数据上有效进行预测的能力。模型通用性在 DL-PSCA 中至关重要，因为攻击阶段使用的曲线与建模曲线客观上存在多方面的差异，例如采样探针位置、设备型号以及算法实现差异等。训练通用模型能够实现更高的性能，避免重头训练模型。为提高模型的通用性，本文从现有研究中归纳整理了五种可行的技术手段。

4.2.1　数据增强

数据增强(Data Augmentation，DA)指对现有数据样本进行各种形式的转换或变换以此构造新的样本，从而人为地增加数据集的规模，提高多样性。DA 能够降低模型过拟合风险并提高模型在未见过的数据上的良好表现，因此提高模型的性能和通用性。

(1) 为侧信道曲线设计的数据增强方法。这类数据增强方法或模拟数据采样阶段的随机性，或模拟隐藏防护，如随机延迟等，以此来丰富训练样本集的多样性。例如，随机水平移动曲线[56,92]，随机插入或移除一定数量的样本点[56]以及添加高斯噪声[93]。

(2) 迁移自其他领域的数据增强方法。Luo 等人[94]将一种名为 Mixup[95]的数据增强技术移植到 DL-PSCA 中。其基本原理是将多个不同的样本线性叠加，以生成一个新的样本，新样本的标签为原样本标签的线性组合。另外，数据平衡技术(4.1.4 节) SMOTE 可视作 Mixup 的一个特例。Won 等人[96]在多个数据集上进行了对比实验，指出 SMOTE 对模型性能的提升要优于文献[56]提出的数据增强技术。本文持谨慎立场，认为这个结论可能存在一定限制。由于 DA 方法引入了随机性，而 SMOTE 则是通过线性叠加已有样本。因此，在攻击曲线不够多样的情形下，DA 的优势无法充分体现。

(3) 基于生成式网络的数据增强方法。Wang 等人[97]借助条件生成对抗网络(Conditional GAN)来扩充建模训练集的规模。

当前，在图像、音频等领域，数据增强方法的应

用非常普遍,但在基于深度学习的侧信道领域使用较少。这是因为目前公开数据集的训练集和攻击集之间的差异非常小,有些甚至是由同一批采集的数据划分而来,因此无法充分展现数据增强的优势。然而,这并不意味着数据增强在 DL-SCA 中不重要。相反地,在应对非实验环境下的侧信道攻击(例如 4.2.3 节中的跨设备场景)时,数据增强具有显著的实际意义。本文认为,随着 DL-SCA 朝着实际攻击场景的发展,数据增强的潜力将得到充分挖掘。因此,尽管目前在 DL-SCA 中数据增强的应用相对较少,但随着对实际攻击场景需求的增加,数据增强的重要性将会进一步凸显,并在未来的研究中得到更多关注和应用。

4.2.2 特征工程

特征工程是指通过选择、转换、创建和优化数据特征,以提高后续建模性能的过程。区别于关注数据的清理和标准化的预处理,特征工程侧重于提取信息和优化特征。DL-PSCA 具有一大优势:无须繁杂的特征工程即可实现攻击目标。然而,利用深度学习技术对数据进行特征特提取,不仅能降低后续 DL-PSCA 的模型的复杂度,也可以提高传统侧信道分析的性能。

(1)基于自编码器的特征工程。自编码器在侧信道领域的用途广泛,主要有三点:① 去噪,若将一些隐藏防护对策视作噪声,则可以通过自编码器移除隐藏防护[98];② 对输入曲线降维[63,99],降低模型的复杂度;③ 处理异质数据集[100],使用自编码器将不同数据集的样本映射到一个公共的特征嵌入空间,从而能够使用一个神经网络模型攻击所有的数据集。

(2)基于对比学习的特征工程。借助三重网络,可以在最大化不同类别曲线的特征向量之间的距离的同时最小化同类别曲线的特征向量之间的距离,以此得到输入曲线的最佳特征嵌入表示,如图 2(b)。Wu 等人[101]将该特征表示用于模板攻击,得出一个与以往研究截然不同结论:模板攻击的性能可以与 DL-PSCA 相媲美,甚至在某些情况下表现得更优。

综上,特征工程能大幅度提高模型的通用性。在 DL-PSCA 面向实际攻击场景时,还有众多问题亟待解决,毫无疑问,特征工程将是一条可行的道路。同数据增强技术一样,基于深度学习的特征工程的发展潜力巨大。

4.2.3 迁移性

在实际的建模侧信道攻击中,攻击曲线的来源与建模时的训练曲线来源是不同的。因此,研究 DL-PSCA 模型的迁移性变得至关重要。根据建模和攻击样本的来源,目前存在 3 种 DL-PSCA 的迁移场景:

(1)跨设备:用于建模和攻击阶段收集的数据来自不同的设备。根据它们之间的差异程度,进一步分为 3 类[12]:

① 同型号设备:建模设备和目标设备的芯片是同一模型的两个物理副本,具有相同的设计和配置。3.2 节中介绍的 XMEAG 和 SAKURA_AES 均为同型号设备的跨设备数据集。

② 同构设备:建模设备和目标设备的芯片来自同一制造商,但型号和结构不同。

③ 异构设备:两种设备的核心芯片来自不同的制造商,因此在各方面都有所不同。这种程度的差异认为是最难解决而又实际存在的问题。

(2)跨实现:与建模设备相比,目标设备配备了额外的隐藏对策:噪声或抖动(jitter)。调研暂未发现实现层面上的更大跨度,例如:无掩码实现跨有掩码实现。

(3)跨信道:特指建模和攻击的曲线来自不同的信息渠道,例如,分别来自能耗和电磁,或来自不同探测位置的电磁。这是一个特殊的类别,目前为止没有引起太多的关注。3.2 节中介绍的 XMEAG_EM 即为跨信道的数据集。

图 5 则按时间线列出了现有的迁移性工作(包括发表会议或期刊),表 2 按照攻击的三个阶段,总结梳理了当前的迁移性工作:

(1)训练前:训练数据是否进行了特征工程?

(2)预训练:训练数据是否来自多个源?

(3)再训练:在攻击前是否对训练好的模型进行微调?若有,那么微调时的目标数据是否带标签,即是否是有监督的微调?

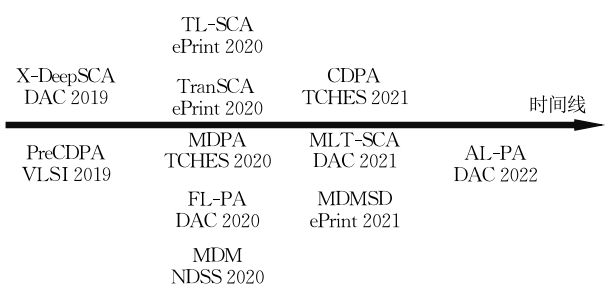


图 5 迁移性工作时间线(会议或期刊的缩写全称可参考附录 A 的缩写对照表)

表 2　DL-PSCA 迁移性工作对比

	训练前	预训练		再训练	跨设备	跨实现	跨信道
	特征工程	单源	多源	监督			
FL-PA[102]	FFT	√		N/A	异构	√	
PreCDPA[103]	PCA DTW	√	√	N/A	同型号		
X-DeepSCA[104]			√	N/A	同型号		
MDPA[105]	SOST	√	√	N/A	同型号		
MDMSD[106]		√		N/A	同型号		
MDM[107]		√	√	N/A	同型号		
TranSCA[108]		√		√		√	
TL-SCA[109]		√	√	√	同型号		√
MTL-SCA[110]		√		√	异构		√
CDPA[44]		√		×	同型号	√	√
AL-PA[111]		√	√	×	同型号	√	

注：(1) N/A 表示不适用，即无再训练阶段；(2) 最后三列为对应方案的实验覆盖面，而非该方案的应用范围。

(1) 训练前进行特征工程

当建模和攻击样本的来源存在较大差异时，通过进行特征工程可以提高模型的通用性和迁移性。例如，基于快速傅里叶变换(Fast Fourier Transform，FFT)的 FL-PA[102]，结合主成分分析(Principal Component Analysis，PCA)和动态时间规整(Dynamic Time Warping，DTW)的 PreCDPA[103] 以及使用 SOST 进行特征筛选的 MDPA[105]。其中，以 FL-PA 的效果最为突出，作者利用 FFT 将侧信道曲线从时域转换到频域，然后选择七个幅度最大的频率作为 MLP 模型的输入特征，成功地克服了异构设备的差异。据我们所知，这是目前成功跨越的最大差异。另外，其实验同时展示了跨实现①的迁移攻击。

(2) 训练时使用多源数据

训练时使用多源的训练数据，同样能够提高模型的迁移性。其原理类似数据增强：增加训练集的多样性。因此本文认为，数据增强能作为多源训练的替代方法。文献[102-105，107]均采用了多源数据训练，其中文献[105]指出，如果只使用单一设备进行建模，则在实践中难以成功进行跨设备攻击。Wu 等人[106]反其道而行之，提出了来自单一设备的多设备模型(MDMSD)，其目标是消除或减轻多个设备的攻击假设，并实现与多源训练相同或相近的性能水平。尽管 MDMSD 的技术路线是预训练移除部分模型使用带噪声的数据再训练，但整个过程仅使用了训练集，因此，本文将其归纳为不涉及微调的方案。

(3) 再训练微调模型

迁移学习是实现迁移攻击的一种直接方式[112]，指通过微调预训练模型，将在源域获得的知识应用于目标域。其中，建模数据所在域为源域，被迁移的目标数据所在域为目标域。依据微调模型时是否需要带标签的目标域数据，分为有监督和无监督的微调。

有监督的微调　TranSCA[108] 使用与目标设备相同型号的设备的少量带标签曲线，对除分类器外的其余层进行微调。尽管作者声称 TranSCA 能够跨越不同的 FPGA 系列，但论文实验仅使用了噪声水平不同的模拟曲线。根据本文的定义，将 TranSCA 归纳为跨实现的攻击。同一时期，TL-SCA[109] 也利用了迁移学习来避免为目标设备从头训练模型，该工作考虑了跨信道和跨设备(同型号设备)的情况。文献[110]结合迁移学习与元学习，提出了能够应对异构的目标设备的 MTL-SCA。元学习指：通过在训练时暴露模型于多个任务，并使其能够从这些任务中快速学到适应新任务的能力。需要强调的是，本文认为这类基于有监督微调的迁移攻击的应用场景受限：要求攻击者能够获取目标域的少量带标签数据是一个极为苛刻的假设。一个可行的应用场景为：已知一个在一段时间后失效的会话密钥，攻击者可以采集少量有标签的侧信道曲线进行有监督的微调，从而还原新的会话密钥。

无监督的微调　相较于有监督的微调，无监督的微调无需增强攻击者能力假设，符合通用的攻击场景。然而在不结合强大特征工程的前提下，目前的无监督微调仅限于跨同型号设备、跨实现和跨信道这种跨度较小的情形。CDPA[44] 使用最大平均差异(Maximum Mean Discrepancy，MMD)[113] 来衡量源域和目标域之间的差距。在微调过程中，将 MMD 作为原始损失函数的一个附加正则项，使模型在不降低源域性能的前提下适配目标域。另外，

① 文献实验部分的 Case 2：训练集为无防护的 AES-128 的实采数据，而目标数据集为 DPA v4.1，如 3.2 节中所述，DPA v4.1 被普遍视作已知掩码下的无防护数据集。因此，该跨实现实验并未跨越有无掩码的差异。

AL-PA[111]使用 GAN 网络来最小化领域差异。微调过程如图 2(d)所示,区分器试图区分输入是来自源域还是目标域,而编码器试图将目标域的曲线提取出的特征转化为源域特征表示。因此,微调的过程是区分器和编码器的动态博弈。

综上所述,特征工程、多源训练和微调技术共同影响模型的迁移性跨度。目前的研究现状显示,有监督微调的应用场景受到限制,而无监督微调仅在同型号跨设备等较小范围的情境中表现优异。相比之下,特征工程和多源训练对于迁移性跨度的影响值得深入探索。为了克服目标域样本与源域样本之间的较大跨度,可以考虑结合三个方面。

4.2.4 聚 合

为追求模型通用性,可以在建模时聚合多方的努力。根据多方合作的层面,有四种级别的聚合,如图 6 所示。

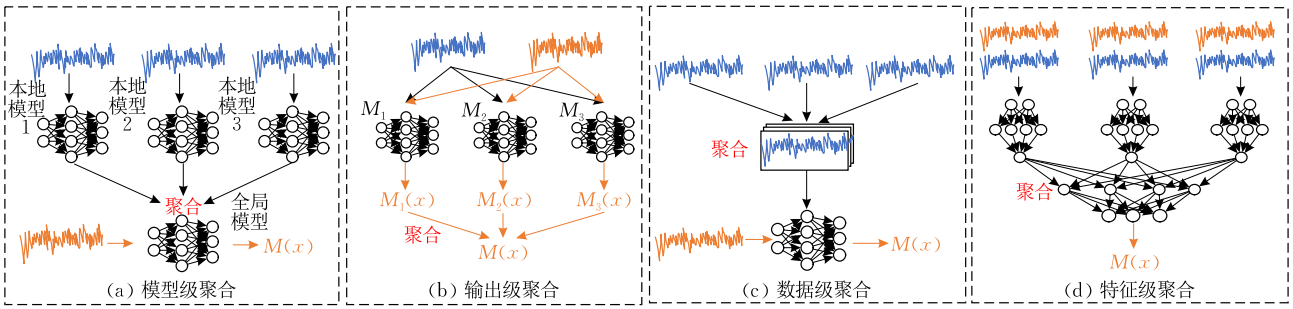


图 6 四种不同级别的聚合方式(图中,蓝色/黄色分别表示训练/攻击曲线)

(1)模型级聚合

图 6(a)为参考了水平联邦学习框架的模型级聚合。联邦学习[114]是为了在整合多方的计算资源的前提下,保证各方数据的保密性而提出的。每个设备利用自己的本地数据训练一个本地模型,将本地模型在中央服务器进行聚合生成一个全局的模型。调研发现,文献[115]是第一个,也是唯一一个在侧信道背景下应用联邦学习的研究。

(2)输出级聚合

图 6(b)为受 bagging[116]元算法启发的输出级聚合方法,该算法是一种集成学习[28]。bagging 方法指将所有单一模型的预测进行平均或线性组合(或加权和)得到最终的预测。一方面,集成方法能减少由权重初始化和超参数选择带来的随机误差[84],然而超参数调优对模型性能的贡献仍不可忽视。另一方面。集成模型的性能可以通过增加模型成员的多样性来提高[73]。

(3)数据级聚合

如图 6(c)所示,数据级聚合指的是收集来自多个源的数据用于模型训练。该聚合方法即 4.2.3 节中介绍的多源训练[103-105,107,109,111],在此不再赘述。另外,研究表明:建模设备和目标设备的加密密钥之间的差异也会对迁移攻击的性能产生影响[107,117],因此,确保多源数据包含变化的密钥对模型通用性存在增益。

(4)特征级聚合

如图 6(d),特征级聚合指独立提取不同的源的样本的特征,然后将特征进行聚合联立服务于下游任务。在侧信道领域的应用主要体现于多信道攻击[117](4.2.5 节)。另外,特征级聚合与 4.1.1 节中介绍的 MCNN 有所不同。MCNN 针对相同的数据采用不同的特征提取技术,旨在最大程度地提取有用信息,而非对不同信道的特征进行聚合。

表 3 对比了四种聚合方式。其中,基于联邦学习的模型级聚合场景在侧信道领域并不适用:因为当前研究者的关注点主要集中在其他方面,而非侧信道曲线的保密性。输出级聚合可以视作一种以空间换时间的方法。然而,训练多个模型进行联合决策来替代单个模型的调优,对整个领域的发展并没有指导性的影响。相比之下,数据级聚合和特征级聚合更具有较高的研究和应用价值。

表 3 四种聚合方式的对比

类型	技术路线	优势
模型级聚合	联邦学习	本地数据的保密性。
输出级聚合	集成学习	减少因参数选择造成的模型性能偏差。
数据级聚合	多源训练	多样性数据,增强模型通用性。
特征级聚合	多信道攻击	充分利用不同信道的特征。

4.2.5 多信道

经典的侧信道分析利用从单一侧信道采集的信息进行分析,Agrawal 等人[118]于 2003 年首次提出基于多信道的攻击。多信道特指单个源同时产生多个信道的泄漏,如一个密码设备加密时产生的能耗泄漏和电磁泄漏。随着 DL-SCA 的出现,多信道攻击重新引起了关注。需要强调的是:多信道攻击并

不等同于特征级聚合，尽管它们的范围存在一定的交叉。前者注重于训练数据的来源，而后者侧重于聚合的阶段在特征提取之后。

多信道信息的利用方式多种多样，取决于攻击者的目标。Yu 等人[119]通过对功耗噪声和电磁噪声之间的关系进行深度学习建模，以使用电磁噪声过滤功率噪声。Hettwer 等人[120]使用多头的网络结构(图 7)进行多信道攻击，并通过实验表明使用 Add 操作进行特征合并要优于其他 6 种操作。

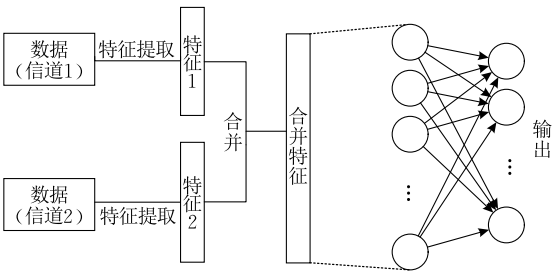


图 7　文献[120]中多信道攻击使用的网络结构

目前，深层次的多信道攻击研究并不突出。在多信道攻击中，信道间的差异越大，表示含有的专属特征越多，从这个角度来看，多信道攻击有望为迁移性攻击构建通用模型提供助力。

4.3　面向可解释性

经典侧信道分析以统计学和信息论为支撑，而 DL-SCA 是一种黑盒的攻击，人类无法直接观测到网络学习的知识。为此，将人工智能可解释性(eXplainable AI，XAI)扩展到 DL-SCA 对理解其原理和设计高效的网络具有重要意义。XAI 对深度学习侧信道攻击(DL-SCA)领域的贡献主要体现在三个方面。在这里需要指出，XAI 技术的应用范围并不受其在特定文献中的应用限制。换句话说，即便 XAI 技术在某篇论文中被用于特定目的，作为一种工具本身仍然可以用于其他方面。

(1) 超参数可解释性

超参数可解释性指通过了解网络学到的内容以服务于超参数调优的目标。van der Valk 等人[121]首先提出了 Guessing Entropy Bias-Variance Decomposition 工具，用于研究超参数设置对性能的影响。随后，该团队与 Bhasin 合作[122]，借助 SVCCA 工具[123]比较和分析神经网络学到的内部表示。值得一提的是，这项研究还为对抗迁移性提供了有益的启示。

(2) 防护可解释性

与传统侧信道不同，基于深度学习的攻击在面对带有掩码防护的实现时无需攻击者付出额外的努力。由此引发了侧信道领域研究人员的兴趣：深度学习网络在处理经过防护的样本时如何解读和利用信息，这也就是所谓的防护可解释性问题。一些研究通过信息论指标进行了深入分析。Wu 等人[106]通过比较各层进行“切除-再训练”前后的猜测熵差异，深入剖析了防护的处理过程，而 Perin 等人[124]从感知信息的角度对深度神经网络层的内部活动，特别是对掩码的处理，进行了解释。此外，还有一些研究运用 XAI 工具进行深入剖析。文献[125]采用了 XAI 工具 SHAP[126]，通过掩盖部分输入样本来揭示网络是否能够提取掩码泄漏。此外，作者还将梯度可视化用于掩码防护可解释性，而这一工具最初是用来定位 PoI[127]。

(3) 泄漏定位

XAI 还可用于事后的泄漏定位，即攻击成功后反过来进行泄漏定位。为服务于攻击-防御-检测框架，有关泄漏定位的现有工作将在 7.1 节中进行介绍。

毋庸置疑，作为弥补传统侧信道分析和基于深度学习的侧信道分析之间认知差异的可解释性研究将持续作为相关领域的重心，其发展潜力巨大。考虑到当前关于 DL-PSCA 的定量分析亟待攻克，可解释性研究或成为填补其空白的一块重要拼图。

4.4　面向经济性

DL-PSCA 在实验场景下表现优异，但当神经网络模型的规模过大时，需要耗费大量的计算和存储资源。降低模型的资源需求以适应资源受限的终端部署设备能够获得更高的经济效益。为实现该目标，可以采用模型压缩的技术手段。

模型压缩有多种方法，一个潜在解决方案是通过知识蒸馏[128]来训练一个小的模型[129]。具体来说，通过将深层网络的输出作为小型网络的合成标签，指导小型网络的输出向大型网络逼近。除了训练一个较小的网络外，另一个选择是应用剪枝算法对深度神经网络进行精简[130-131]。Perin 等人[131]依靠彩票假设，剪除了 90%以上的训练参数。这里的彩票假设指的是：一个充分过度参数化且带有随机初始化权重的神经网络中，存在一个子网络，无需额外训练即可与完整网络相媲美[132]。

模型压缩的研究与侧信道领域无直接关系，且其出发点相对受限：大多数情况，侧信道攻击模型不需要部署到资源受限的终端设备。因此，本文

不提倡 DL-SCA 研究者对模型压缩方法开展进一步研究。

5 基于深度学习的非建模侧信道分析 DL-NPSCA

非建模侧信道分析(NPSCA)依靠统计分析的差异来区分正确密钥与其他密钥,而不需要预先创建模板。正因如此,深度学习的"训练预测"模式无法直接套用到 NPSCA。然而,研究人员另辟蹊径,成功将深度学习技术应用于 NPSCA。

5.1 深度学习驱动的非建模侧信道分析

在 CHES 2019 会议中,Timon[133] 提出了首个基于深度学习的非建模侧信道分析:差分深度学习分析(Differential Deep Learning Analysis,DDLA)。如算法 3,DDLA 是简单且直接的,可以视作是 DPA 的神经网络版本。对于每一个密钥假设,攻击者使用相应的假设标签训练一个神经网络模型,而在正确的密钥假设下,训练模型的度量应当优于其余密钥假设下的训练度量。

算法 3. 差分深度学习分析(Differential Deep Learning Analysis,DDLA)

输入:训练曲线 X,对应的加密明文 P,神经网络 *net* 和训练轮数 n_e

输出:密钥 K^*

1. FOR $K \in \mathcal{K}$
2. 初始化 *net* 的可训练参数
3. 计算所有密钥假设下的候选变量 $H=f(P,K)$
4. 将候选变量编码为训练标签 $Y_K=g(H)$
5. 进行深度学习训练:$DL(net,X,Y_K,n_e)$
6. RETURN 获得最佳训练度量的密钥候选 K^*

DDLA 的提出奠定了深度学习驱动的非建模侧信道分析的基础,此后的大部分工作均为基于此的改进。Kuroda 等人[125] 使用 DDLA 成功攻击了含 RSM 掩码防护的 AES-128 软件实现以及 ASCADf 公开数据集。为提高 DDLA 性能,Won 等人[134] 建议将曲线转化为二维图片,与前文的二维曲线转换[58-59] 的不同点在于:这里的方案是对曲线的形状的编码,而前面的方案则是通过数值转换来完成。初始的 DDLA 使用的训练度量指标是基于敏感度分析(Sensitivity Analysis, SA)[135],Lu 等人[136] 提出了一个基于注意力机制的替代指标。此外,也有研究人员建议使用最小平均误差(MSE)作为 DDLA 的训练度量[137]。为解决 DDLA 训练成本较高的难题——需要训练与密钥假设数量一样多的模型,研究者提出了加速方案[138-139]。其中,文献[139]提出一种多输出多损失神经网络,可以在短时间内同时预测所有可能的假设密钥。

除了 DDLA,Zhang 等人[140] 也提出了一种深度学习驱动的 NPSCA,命名为神经互信息分析(Neural Mutual Information Analysis, NMIA)。NMIA 使用神经网络估计互信息值,作为区分器来执行非建模侧信道分析。

第三种深度学习驱动的 NPSCA 由 Ramezanpour 等人[141] 提出,他们没有试图在传统非建模侧信道分析的框架上进行技术迭代,而是基于强化学习提出了一个新的框架 SCARL。简单来说:一个自动编码器负责压缩特征,接着利用强化学习对特征进行聚类。

第四种深度学习驱动的 NPSCA 聚焦于非对称密码实现。通常情况下,对非对称密码系统的分析需要很长时间,Lee 等人[142] 提出利用孪生网络(如图 2(a))在合理时间内恢复全部密钥的方法。然而,该方法的一个限制在于,进行预测时需要一个支持集,即一小部分带标签的数据,作为参考数据集以便进行一次性学习(One-shot Learning)[143]。具体而言,孪生网络的输入包括攻击曲线和参考曲线,通过比较这两组样本的相似度来判断是否属于同一类,从而获取攻击曲线的预测密钥。由此,该方法并不严格遵守非建模侧信道分析的场景。

近年来,深度学习驱动的建模侧信道分析技术并未取得显著的发展。值得注意的是,侧信道领域对无监督学习的应用相对滞后。对于无监督学习或强化学习与非建模场景的潜在深度结合,本文持审慎态度。

5.2 深度学习辅助的非建模侧信道分析

深度学习技术在 NPSCA 中的另一个重要用途是辅助。例如:使用 MLP 拟合功耗模型的非线性映射,以此优化 CPA 和 MIA 攻击的性能[144];迭代修正水平攻击得到的私钥的比特位[145]。与此同时,当深度学习技术作为辅助手段时,并不限定其后续攻击是建模还是非建模场景。因此,先前介绍过的应用于建模场景的深度学习辅助技术在非建模场景中仍然具有适用性,例如:使用自编码器学习有效特征表示辅助 NPSCA[146-147]。

在展望深度学习对侧信道领域的辅助作用时,不应单独着眼于建模或非建模类型,而应统一地看

待。正如第 4 节所述，随着 DL-SCA 逐渐贴近实际攻击场景，特征工程等辅助技术的优势也将逐步显现。

6 防　护

针对传统 SCA 设计的防护对策在面对 DL-SCA 时是否仍然有效？前文介绍的攻击已经给出了答案：DL-SCA 能够有效的应对包括隐藏[56,98,148]和掩码[51-53,57,64,67]在内的防护策略。相较基于深度学习的侧信道攻击，利用深度学习设计反制措施的研究工作要相对贫瘠。

6.1 针对 DL-PSCA 的防护

随着 DL-PSCA 的兴起，针对这类新型攻击的反制措施理应同步发展。目前的防护措施主要有两种。

(1) 基于对抗样本的防护。鉴于 DL-PSCA 通常被视作分类任务，最直观的对抗手段就是误导分类器做出错误的预测。Picek 等人[149]实践了这种设想，将对抗样本这一针对神经网络的攻击作为 DL-PSCA 的防护对策。首先，攻击者使用一个代用模型来生成具有误导作用的噪声扰动。需要指出，为使扰动能够影响目标模型的预测输出，生成的噪声扰动具备迁移属性。之后，将攻击样本叠加上噪声扰动，就能导致目标模型输出错误的预测，如图 8 所示。目前，这项工作展示了非针对性的误导，即引导模型产生错误分类而非精确地将其误导为特定类别。这表明基于对抗样本的防护方案有待挖掘的潜力。此外，Bertrand 等人[150]也展示了对抗样本对 DL-PSCA 的误导，但他们的动机是质疑将机器学习和深度学习作为黑盒评估工具的可行性。

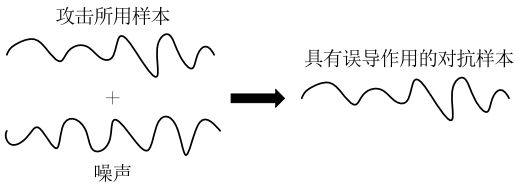


图 8　对抗样本作为侧信道防护

尽管这种对策理论可行，但其具体实现受到局限——怎样在攻击样本中添加噪声扰动？噪声要么伴随侧信道信号实时产生，要么在攻击者采集后人工添加。前者存在将噪声映射为硬件或软件实现的技术壁垒，后者的防护方不存在“可乘之机”。

(2) 基于现有防护的组合。除了基于对抗样本的防护，Rijsdijk 等人[151]认为可以通过叠加多种隐藏对策来抵御 DL-PSCA，他们使用强化学习指导隐藏策略的叠加，在附加最小代价的同时使得之前训练的模型失效。然而，其缺点显而易见：防护是后手。换句话说，该防御措施是建立在攻击模型可被防护者访问的假设之上的，且得到的组合对策仅适用这个特定的模型。尽管如此，强化已有防护，以抵御 DL-PSCA 不失为一个新的思路。

综上，目前提出的两种防护手段均存在很大的局限性，无法满足实际的防护需求。相较攻击，防护领域还有很大的空白需要填补，该领域理应得到重点关注。

6.2 深度学习辅助防护设计

深度学习的全面应用为 SCA 防护的辅助设计提供了广阔的前景，但目前的应用却相对较少。Krautter 等人[152]开展了一项有趣的研究，他们采用神经网络拟合并替代 AES 的 Sbox，以消除控制流中的数据依赖，从而提高对 CPA 攻击的抵抗。尽管该防护在资源开销方面仍在可接受的范围内，但时间开销却是标准 Sbox 的 50 倍。这凸显了在深度学习应用于 SCA 防护领域时，需要权衡资源和时间开销的挑战。

防护的设计在总体上滞后于攻击，也导致了深度学习辅助防护设计的匮乏。研究数据的缺少使我们无法预知其未来发展趋势，但该领域的发展潜力不可忽视。基于深度学习的侧信道防护仍是待攻克的难题，8.3 节将从传统侧信道研究者和深度学习研究者两个角度提供可行的技术路线。

7 检　测

除了攻击和防护，深度学习技术也被用于侧信道的泄漏检测，具体体现在：基于深度学习的泄漏定位和基于深度学习的泄漏评估。

7.1 基于深度学习的泄漏定位

本节将基于深度学习的泄漏定位归纳为事后和事前两类。

(1) 基于 XAI 的事后泄漏定位。在基于深度学习的建模侧信道分析中，神经网络通过自主学习泄漏与敏感变量的映射关系。因此，可以利用 XAI 工具对训练后的神经网络进行敏感性分析，以定位相关的泄漏。敏感性分析是一种用于确定系统或模型对其输入参数变化的输出有多敏感的技术。在文献[133]中，敏感性分析首次被引入侧信道分析：通过

累积 MLP 的第一个网络层的权重梯度的绝对值，攻击者可以确定泄漏发生的位置。类似地，名为梯度可视化(Gradient Visualization)[127]的工具通过计算模型的与输入有关的梯度，可以有效地定位 PoI。此外，Hettwer 及其团队[153]引入图像分类任务中的三种特征可视化方法 Occlusion[154]、Layer-wise Relevance Propagation[155] 和 Saliency Maps[156] 来研究 PoI 的泄漏问题。另外，Cao 等人[64]提出逐层相关性(Layer-Wise Correlation，LWC)，指代特定网络层的输出和原始输入之间的相关性，也可视作敏感性分析的方法。

(2)基于强化学习的事前泄漏定位。文献[157]提出可以进行自动定位的 AutoPoI 框架，其基本流程为：(a)选择 PoI；(b)进行模板攻击；(c)根据攻击结果进行奖励或者惩罚；(d)重复这个过程。通过反馈式的强化学习，能够定位到相关泄漏区域。尽管这并不是对模型进行事后解剖的定位方法，但这种定位方式并未完全脱离攻击，而是伴随着攻击进行的。

对于有掩码防护的样本曲线，传统的攻击方法无法在未知掩码随机数的情况下进行泄漏定位，但基于深度学习的泄漏定位能够做到。然而，其定位模式：攻击模型收敛之后的反向事后定位，导致基于深度学习的泄漏定位缺乏实用性。由此，基于深度学习的事前泄漏定位是个更有趣的方向，但目前有更简洁的替代方法——基于自编码器的特征工程。鉴于此，本文认为未来的研究重点不会集中在基于深度学习的泄漏定位上。

7.2 基于深度学习的泄漏评估

泄漏评估或基于信息论，或基于统计分析，均可以与深度学习相结合。

(1)基于信息论的泄漏评估。最直观的方法就是用神经网络来估计信息论指标[158-159]。例如，文献[158]提出了一种新的基于神经网络的互信息估计方式 MINE。之后，Cristiani 等人[160]使用 MINE 来评估设备的侧信道泄漏。

(2)基于统计分析的泄漏评估。传统的泄漏评估方法 TVLA 将采集的样本根据输入的明文被分为两组，通过考查两组是否存在显著的分布差异来评估泄漏。首个基于深度学习的泄漏评估方法 DL-LA[161]的核心思想是训练一个神经网络作为评估两组样本之间差异性的区分器。如果验证集中的两组样本能以高概率区分开来，则认为存在侧信道泄漏。DL-LA 和传统 TVLA 的对比如表 4 所示。

表 4 TVLA 和 DL-LA 的对比

	TVLA	DL-LA
数据采集	获得两组基于不同固定明文(fixed-vs-fixed)或一组固定明文一组随机明文(fixed-vs-random)加密过程中采集的侧信道样本，记作 I_A，I_B。	
样本准备	取样本 I_A，I_B 上一个时刻的样本点 Q_A，Q_B。	取样本 I_A，I_B 上一个时刻的样本点 Q_A，Q_B，分别拆成 T_A，V_A 和 T_B，V_B，构造训练集 $T=\{T_A,T_B\}$ 和验证集 $V=\{V_A,V_B\}$。使用训练集 T 训练一个二分类网络，该网络能够准确区分样本点的来源 A 或 B。
检验	对 Q_A，Q_B 的分布进行 t 检验。当 $\lvert t\rvert>4.5$ 时，认为该样本点存在泄漏	使用验证集 V 进行检验，若分类准确率大于门限(门限设定大于 0.5，因为 0.5 表示二分类中随机猜测的概率)，则认为存在泄漏。

目前，DL-LA 的可靠性受到质疑。一方面，神经网络的选择能直接影响甚至决定泄漏评估的结果。另一方面，网络模型易受误导，导致其评估结果不可信[150]。本文认为，将 DL-LA“开盒”是其向可靠评估方法靠近的必经之路——结果导向的黑盒评估方案无法提供合理的安全边界。由此可以预见：基于深度学习的泄漏评估需要依托于 XAI 的发展进步。

8 讨 论

从宏观角度来看，深度学习在侧信道领域的应用十分不平衡，绝大部分工作集中在攻击相关领域，而面向防护或检测的工作非常有限。为更好地分析近年的研究热点与趋势，本文绘制了 DL-SCA 领域的文献的统计数据图。如图 9 所示，多级分类使用不同的标记表示。例如，反斜杠标记的条形块表示对应年份的 DL-PSCA 相关的研究论文数目，蓝色的条形块则表示面向适应性的工作，而最浅的蓝色条形块代表定制损失函数/指标的论文数目。图 9 展示了一段时间内的研究侧重点，结合具体的研究现状，本节总结当前的 DL-SCA 的热点研究方向，并结合前文对 DL-SCA 的研究趋势进行展望。另外，我们还对当前的研究困境——防护和检测，分别从传统侧信道和深度学习的研究者的角度提供可行的技术路线。

8.1 研究热点

8.1.1 适应性研究

从文献的数据上看，深度学习与侧信道分析的深层次融合，即适应性工作，无疑是 DL-SCA 领域的研究热点。一方面，这是深度学习技术在侧信道分析领域最直观的应用研究。另一方面，DL-PSCA 在针对隐藏或掩码等防护措施时表现出的巨大优势

得到了众多侧信道领域研究人员的关注。从图 9 可以看到，对应于 DL-PSCA 的适应性工作的蓝色区域占据了近一半的比例，其特点是起步早且仍保持稳定发展速度。其中，网络架构的设计（最深的蓝色块）在 2020 年前后蓬勃发展，而近年来，架构的创新集中在针对特殊的场景定制网络结构。对有防护的算法实现考虑一些特殊的网络结构设计，针对性地面向掩码防护或随机扰动等隐藏防护。在定制损失函数或指标方面，尽管已经有大量研究致力于改进当前存在的不足，但创新性的应用并没有得到广泛采纳。研究者们普遍倾向于使用交叉熵损失函数和猜测熵评估指标。

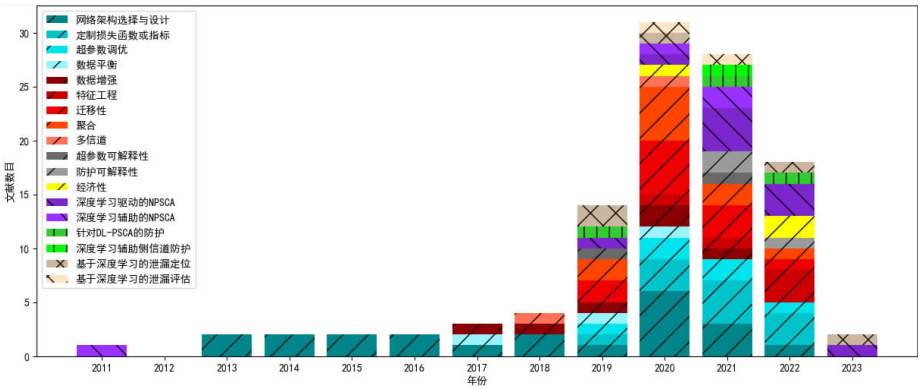


图 9　DL-SCA 领域研究工作的统计分析数据（条形样式用于区分不同类别：标有反斜线、斜线、竖线和十字的条形图分别代表 DL-PSCA、DL-NPSCA、防护和检测。相似颜色条类属同一子类别，而不同颜色条表示最细粒度的类别（请注意：截至这篇论文的写作时间，收集的 2023 年的工作有限，因此 2023 年的数据不具有代表性））

8.1.2　通用性研究

在图 9 中，通用性研究占比近三成（红色区域）。近年来，研究者们致力于改进深度学习预训练模型的通用性能力，数据增强和特征工程等提高模型通用性的方法不断涌现。与此同时，借助通用模型发起的迁移攻击正进一步扩大 DL-PSCA 的实际威胁。如何将训练好的模型用于非建模设备是建模侧信道攻击需要解决的问题。传统建模侧信道分析通过约束建模设备和目标设备之间的差异，尽可能避免这个问题。相比之下，基于深度学习的方法放宽了对攻击者能力的假设，允许建模设备与目标设备之间存在一定的差异。因此，弥补跨度差异的迁移性研究成为近年来的热门问题，相关的研究自 2019 年已经开始，虽然近年来进展缓慢，但仍是广大研究者重点关注的话题。4.2.3 节介绍的迁移性研究现状表明：目前的迁移跨度十分有限，大部分的工作都集中在对同型号设备的跨设备攻击。为扩展迁移性跨度，除了使用特征工程、多源训练以及微调提高模型通用性之外，还可以考虑构建一个通用的骨干模型（8.3.3 节）。

8.1.3　基于深度学习的非建模侧信道技术

从数量上看，基于深度学习的非建模侧信道分析（图 9 紫色区域）正经历快速发展阶段，已成为近年的热点研究问题。受益于众多无监督深度学习和强化学习的引入，这些技术在辅助非建模侧信道分析的攻击方面发挥了重要作用。与此同时，深度学习驱动的非建模侧信道分析也在不断优化改进，然而，革新式的方案却寥寥可数。

8.2　研究趋势

8.2.1　面向实际攻击场景

早期的 DL-PSCA 工作均立足于公开数据集，全然忽视这些数据集存在的问题：（1）样本单一；（2）经过样本对齐与切割等初步预处理；（3）训练集和攻击集是由同一批数据中划分得到的。由此导致此前考虑的场景都过于理想，高估了攻击者能力，而现实攻击场景更为复杂。解决这些问题是 DL-PSCA 从实验场景迈向实际攻击场景的必经之路。例如：4.1.1 节中，Lu 等人[68]提出一个自定义网络来对未切割的生数据进行端到端的建模侧信道分析；4.2 节为提高模型通用性而开展的多方面的研究。迄今为止，面向实际攻击场景的 DL-PSCA 仍存在众多问题，基于此的研究探索仍将继续推进。

8.2.2　DL-SCA 具象化

根源于传统侧信道坚实的理论基础，其每一步的操作行为都是有理可循的，因而传统侧信道分析被认为是具体可靠的。相反，深度神经网络被认为是黑盒模型，即我们可以观察到输入和输出，但很难理解网络内部的具体操作和决策过程，这种不可预知性和不可解释性对侧信道领域的研究者来说是难以接受的。对攻击、防护与检测都产生了负面影响：

由于训练过程中的随机性，基于深度学习的侧信道攻击有可能需要多次尝试才能得到收敛模型；不可预知性使防护的有效性受到挑战；不可解释性使基于深度学习的检测技术不可靠。研究者一直在积极探索技术和方法，具象化基于神经网络的侧信道技术的具体操作和决策过程，以应对不可预知性和随机性所带来的挑战。随着 XAI 在 DL-SCA 领域的进一步扩散，DL-SCA 将逐渐具象化，以此弥补神经网络的黑盒属性。

8.2.3 推动防护和检测技术

本文多次强调侧信道领域的三大分支与新技术的结合存在明显的失衡：基于深度学习的侧信道攻击技术遥遥领先，而另外两部分则鲜有引人注目的研究，推动缓慢。同时，目前关于防护和检测的研究工作的局限性已在相应章节中进行了讨论。纵观密码系统的安全研究进程，攻击与防护和检测是相互追赶与促进的，攻击技术的进步势必推动防护和检测技术的革新。由此可见，基于深度学习的防护和检测领域将成为未来研究的重点之一。

8.3 技术路线

目前，深度学习和侧信道分析的结合并不均衡，防护和检测领域存在巨大的空白待填补。考虑现有防护和检测技术的局限，本节中，作者将站在传统侧信道研究者和深度学习研究者的角度，提供能够解决这些问题的技术思路。

8.3.1 基于 XAI 的定量分析

站在传统侧信道研究者的角度来看，深度学习方法的不确定性和黑盒属性是阻碍其在防护与检测领域推进的核心问题。在传统的侧信道分析中，模型的训练与预测都是确定性的，但深度学习模型由于其复杂的结构和非线性特征，使得其内部操作变得不透明。这种不确定性与黑盒特性导致分析者难以预测模型在不同情况下的行为，从而难以准确评估其安全性和防护效果。

通过“开盒”分析，研究人员能够深入理解模型的内部机制，揭示其决策过程。这种方法的实现依赖于可解释人工智能（XAI）技术，XAI 通过提供模型的可解释性，使我们能够对其行为进行定量分析，并建立相应的安全基准。

在传统的侧信道分析中，密码实现的安全性可以通过基于探针模型[162]的可证明安全框架定义，也可以由基于信息论的度量来预估安全上界。同样地，在深度学习驱动的防护策略中，若能够定义类似的安全边界来评估其安全性，则能为研究人员提供清晰的安全评估标准。

基于深度学习的泄漏评估也需要在特定的基准下进行。DL-LA 方法受到网络架构选择或训练过程中的差异影响，从而导致误差和不准确的结果。因此，通过在统一的基准下进行评估，能够显著降低这些误差，确保评估结果的可靠性和一致性。这将有助于更准确地衡量防护措施的有效性，使其在实际应用中能够更好地抵御潜在的攻击。

总的来说，为解决深度学习方法的不确定性和黑盒属性问题，引入 XAI 来建立标准化的评估基准，对于推进基于深度学习的侧信道防护和检测至关重要。这些措施不仅能够增强我们对模型的理解，还将显著提升防护策略的实际效果。

8.3.2 通用的骨干模型

从深度学习的角度看，DL-SCA 的发展现状显得相对初级。它仍处于“一次一模型”的阶段，即每次攻击或评估都必须从头开始训练模型。而在深度学习的其他领域，使用预训练的骨干模型来处理下游任务已经成为主流。骨干模型，例如 VGG[163]和 GPT①（当前最热门的生成式人工智能工具 ChatGPT 正是基于该模型构建），通常指一个强大而通用的特征提取器，负责处理输入数据并提取重要特征，可以适用于各种下游任务，如物体检测或分类等。由于当前 DL-SCA 领域缺少通用的骨干模型，各项研究均基于独立训练的模型开展。一方面，各自训练的模型受超参数设置、参数初始化等随机因素的影响，导致模型性能不稳定性；另一方面，即便应用了当前的数据增强，单独训练的模型的通用性仍十分有限，阻碍了 DL-SCA 的进一步发展，尤其是迁移攻击和基于深度学习的泄漏评估两个方面。然而，训练 DL-SCA 的骨干模型需要海量且丰富的样本。参考计算机视觉领域的骨干模型的构建，利用了超大型公开数据集 ImageNet[164]，该数据集由超过 1400 万张有标注的图像组成，涵盖了 2 万多个物体类别。可以预见，一旦骨干模型建立成功，DL－SCA 将步入规范化阶段，攻击、防护和检测均会迎来飞跃性进步。

9 总 结

本文对基于深度学习的侧信道分析技术按三个

① Improving language understanding with unsupervised learning. OpenAI, 2018. https://openai.com/index/language-unsupervised/

维度：攻击、防护和检测，进行全面的整理与总结。结合当前的研究现状，展望各个方向的未来发展趋势，为薄弱方向的突破提供了思路。

神经网络与侧信道的结合主要集中在建模侧信道分析方面，旨在增强深度学习侧信道攻击的适应性、通用性、可解释性和经济性，从而提升攻击能力和实际威胁。与此同时，基于深度学习的非建模侧信道攻击的发展相对较为缓慢，缺乏突破性进展。目前，攻击研究的趋势明显偏向于实际应用，通用性研究的数量（如特征工程、迁移性研究和数据增强）呈现出显著的增长态势。另一方面，深度学习在防护和检测方面的研究相对受限且饱受质疑。本文认为，从侧信道角度出发，防护和检测领域的突破依赖于 XAI 技术的发展。参考其他深度学习领域，另一条可行的路径是建立骨干模型，推动深度学习侧信道分析领域的规范化。

参　考　文　献

[1] Kocher P, Jaffe J, Jun B, et al. Introduction to differential power analysis and related attacks. San Francisco, USA: Cryptography Research, Technical Report, 1998

[2] Chari S, Rao J R, Rohatgi P. Template attacks//Proceedings of the 4th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2002). Redwood Shores, USA, 2003: 13-28

[3] Hospodar G, Gierlichs B, De Mulder E, et al. Machine learning in side-channel analysis: A first study. Journal of Cryptographic Engineering, 2011, 1(4): 293-302

[4] Bartkewitz T, Lemke-Rust K. Efficient template attacks based on probabilistic multi-class support vector machines//Proceedings of the 11th International Conference on Smart Card Research and Advanced Applications. Graz, Austria, 2013: 263-276

[5] Heuser A, Zohner M. Intelligent machine homicide: Breaking cryptographic devices using support vector machines//Proceedings of the 3rd International Workshop on Constructive Side-Channel Analysis and Secure Design. Darmstadt, Germany, 2012: 249-264

[6] Lerman L, Bontempi G, Markowitch O. Power analysis attack: An approach based on machine learning. International Journal of Applied Cryptography, 2014, 3(2): 97-115

[7] Lerman L, Poussier R, Bontempi G, et al. Template attacks vs. machine learning revisited (and the curse of dimensionality in side-channel analysis)//Proceedings of the 6th International Workshop on Constructive Side-Channel Analysis and Secure Design. Berlin, Germany, 2015: 20-33

[8] Hettwer B, Gehrer S, Güneysu T. Applications of machine learning techniques in side-channel attacks: A survey. Journal of Cryptographic Engineering, 2020, 10(2): 135-162

[9] Picek S, Perin G, Mariot L, et al. SoK: Deep learning-based physical side-channel analysis. ACM Computing Surveys, 2023, 55(11): 1-35

[10] Harrison J, Toreini E, Mehrnezhad M. A practical deep learning-based acoustic side channel attack on keyboards//Proceedings of the 8th IEEE European Symposium on Security and Privacy Workshops. Delft, The Netherlands, 2023: 270-280

[11] Yang L, Chen Y C, Pan H, et al. MagPrint: Deep learning based user fingerprinting using electromagnetic signals//Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2020). Toronto, Canada, 2020: 696-705

[12] Schindler W, Lemke K, Paar C. A stochastic model for differential side channel cryptanalysis//Proceedings of the 7th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2005). Edinburgh, UK, 2005: 30-46

[13] Heuser A, Rioul O, Guilley S. Good is not good enough: Deriving optimal distinguishers from communication theory//Proceedings of the 16th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2014). Busan, Republic of Korea, 2014: 55-74

[14] Kocher P, Jaffe J, Jun B. Differential power analysis//Proceedings of the 19th Annual International Cryptology Conference (CRYPTO'99). Santa Barbara, USA, 1999: 388-397

[15] Brier E, Clavier C, Olivier F. Correlation power analysis with a leakage model//Proceedings of the 6th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2004). Cambridge, USA, 2004: 16-29

[16] Gierlichs B, Batina L, Tuyls P, et al. Mutual information analysis: A generic side-channel distinguisher//Proceedings of the 10th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2008). Washington, USA, 2008: 426-442

[17] Mangard S, Oswald E, Popp T. Power Analysis Attacks: Revealing the Secrets of Smart Cards: Volume 31. New York, USA: Springer Science & Business Media, 2008

[18] Chari S, Jutla C S, Rao J R, et al. Towards sound approaches to counteract power-analysis attacks//Proceedings of the 19th Annual International Cryptology Conference (CRYPTO'99). Santa Barbara, USA, 1999: 398-412

[19] Goubin L, Patarin J. DES and differential power analysis the "Duplication" method//Proceedings of the 1st International Workshop on Cryptographic Hardware and Embedded Systems. Worcester, USA, 1999: 158-172

[20] Schramm K, Paar C. Higher order masking of the AES//Proceedings of the Cryptographers' Track at the RSA Conference 2006. San Jose, USA, 2006: 208-225

[21] Messerges T S. Using second-order power analysis to attack DPA resistant software//Proceedings of the 2nd International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2000). Worcester, USA, 2000: 238-251

[22] Gierlichs B, Lemke-Rust K, Paar C. Templates vs. stochastic methods: A performance analysis for side channel cryptanalysis //Proceedings of the 8th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2006). Yokohama, Japan, 2006: 15-29

[23] Bhasin S, Danger J L, Guilley S, et al. NICV: Normalized inter-class variance for detection of side-channel leakage//Proceedings of the International Symposium on Electromagnetic Compatibility. Tokyo, Japan, 2014: 310-313

[24] Chatzikokolakis K, Chothia T, Guha A. Statistical measurement of information leakage//Proceedings of the 16th International Conference on Tools and Algorithms for the Construction and Analysis of Systems. Paphos, Cyprus, 2010: 390-404

[25] Chothia T, Guha A. A statistical test for information leaks using continuous mutual information//Proceedings of the 24th IEEE Computer Security Foundations Symposium. Cernay-la-Ville, France, 2011: 177-190

[26] Gilbert Goodwill B J, Jaffe J, Rohatgi P. A testing methodology for side-channel resistance validation//Proceedings of the Workshop on NIST Non-invasive Attack Testing. Nara, Japan, 2011, (7): 115-136

[27] Quinlan J R. Induction of decision trees. Machine Learning, 1986, 1: 81-106

[28] Breiman L. Random forests. Machine Learning, 2001, 45: 5-32

[29] Cortes C, Vapnik V. Support-vector networks. Machine Learning, 1995, 20: 273-297

[30] Zhang R, Li W, Mo T. Review of deep learning. arXiv preprint arXiv:1804.01653, 2018

[31] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778

[32] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 2012, 29(6): 82-97

[33] Deng L. Deep Learning in Natural Language Processing. Singapore: Springer, 2018

[34] Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. Advances in Neural Information Processing Systems, 2017, 30: 971-980

[35] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533-536

[36] Bottou L, Curtis F E, Nocedal J. Optimization methods for large-scale machine learning. SIAM Review, 2018, 60(2): 223-311

[37] Kingma D P, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014

[38] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 1735-1742

[39] Wang J, Song Y, Leung T, et al. Learning fine-grained image similarity with deep ranking//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1386-1393

[40] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786): 504-507

[41] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Advances in Neural Information Processing Systems, 2014, 27: 2672-2680

[42] Sutton R S, Barto A G. Reinforcement Learning: An Introduction. 2nd Edition. Cambridge, USA: MIT Press, 2018

[43] Benadjila R, Prouff E, Strullu R, et al. Deep learning for side-channel analysis and introduction to ASCAD database. Journal of Cryptographic Engineering, 2020, 10(2): 163-188

[44] Cao P, Zhang C, Lu X, Gu D. Cross-device profiled side-channel attack with unsupervised domain adaptation. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(4): 27-56

[45] Wouters L, Arribas V, Gierlichs B, Preneel B. Revisiting a methodology for efficient CNN architectures in profiling attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(3): 147-168

[46] Maghrebi H, Portigliatti T, Prouff E. Breaking cryptographic implementations using deep learning techniques//Proceedings of the International Conference on Security, Privacy, and Applied Cryptography Engineering. Hyderabad, India, 2016: 3-26

[47] Martinasek Z, Malina L. Comparison of profiling power analysis attacks using templates and multi-layer perceptron network//Proceedings of the International Conference on Mathematical Methods in Science and Engineering. Czech Republic, 2014: 134-139

[48] Martinasek Z, Zeman V. Innovative method of the power analysis. Radioengineering, 2013, 22(2): 586-594

[49] Martinasek Z, Hajny J, Malina L. Optimization of power analysis using neural network//Proceedings of the International Conference on Smart Card Research and Advanced Applications. Berlin, Germany, 2013: 94-107

[50] Saravanan P, Kalpana P, Preethisri V, Sneha V. Power analysis attack using neural networks with wavelet transform as pre-processor//Proceedings of the 18th International Symposium on VLSI Design and Test. Coimbatore, India, 2014: 1-6

[51] Gilmore R, Hanley N, O'Neill M. Neural network based attack on a masked implementation of AES//Proceedings of the IEEE International Symposium on Hardware Oriented Security and Trust. Washington, USA, 2015: 106-111

[52] Martinasek Z, Zapletal O, Vrba K, Trasy K. Power analysis attack based on the MLP in DPA contest v4//Proceedings of the 38th International Conference on Telecommunications and Signal Processing. Prague, Czech Republic, 2015: 154-158

[53] Martinasek Z, Dzurenda P, Malina L. Profiling power analysis attack based on MLP in DPA contest v4. 2//Proceedings of the 39th International Conference on Telecommunications and Signal Processing. Vienna, Austria, 2016: 223-226

[54] Hou S, Zhou Y, Liu H, Zhu N. Improved DPA attack on rotating S-boxes masking scheme//Proceedings of the IEEE 9th International Conference on Communication Software and Networks. Guangzhou, China, 2017: 1111-1116

[55] Pfeifer C, Haddad P. Spread: A new layer for profiled deep-learning side-channel attacks. Cryptology ePrint Archive, 2018: 880

[56] Cagli E, Dumas C, Prouff E. Convolutional neural networks with data augmentation against jitter-based countermeasures: Profiling attacks without pre-processing//Proceedings of the 19th International Conference on Cryptographic Hardware and Embedded Systems (CHES 2017). Taipei, China, 2017: 45-68

[57] Zaid G, Bossuet L, Habrard A, Venelli A. Methodology for efficient CNN architectures in profiling attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(1): 1-36

[58] Yang G, Li H, Ming J, Zhou Y. Convolutional neural network based side-channel attacks in time-frequency representations//Proceedings of the International Conference on Smart Card Research and Advanced Applications. Montpellier, France, 2018: 1-17

[59] Hettwer B, Horn T, Gehrer S, et al. Encoding power traces as images for efficient side-channel analysis//Proceedings of the IEEE International Symposium on Hardware Oriented Security and Trust. San Jose, USA, 2020: 46-56

[60] Hettwer B, Gehrer S, Güneysu T. Profiled power analysis attacks using convolutional neural networks with domain knowledge//Proceedings of the International Conference on Selected Areas in Cryptography. Calgary, Canada, 2018: 479-498

[61] Hoang A T, Hanley N, O'Neill M. Plaintext: A missing feature for enhancing the power of deep learning in side-channel analysis? Breaking multiple layers of side-channel countermeasures. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, (1): 49-85

[62] Zhang L, Xing X, Fan J, et al. Multilabel deep learning based side-channel attack. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2020, 40(6): 1207-1216

[63] Won Y S, Hou X, Jap D, et al. Back to the basics: Seamless integration of side-channel pre-processing in deep neural networks. IEEE Transactions on Information Forensics and Security, 2021, 16: 3215-3227

[64] Cao P, Zhang C, Lu X, et al. Improving deep learning based second-order side-channel analysis with bilinear CNN. IEEE Transactions on Information Forensics and Security, 2022, 17: 3863-3876

[65] Zhou Y, Standaert F X. Deep learning mitigates but does not annihilate the need of aligned traces and a generalized ResNet model for side channel attacks. Journal of Cryptographic Engineering, 2020, 10(1): 85-95

[66] Vaswani A. Attention is all you need. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008

[67] Hajra S, Saha S, Alam M, et al. TransNet: Shift invariant Transformer network for power attack. Cryptology ePrint Archive, 2021: 827

[68] Lu X, Zhang C, Cao P, et al. Pay attention to raw traces: A deep learning architecture for end-to-end profiling attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(3): 235-274

[69] Zaid G, Bossuet L, Dassance F, et al. Ranking loss: Maximizing the success rate in deep learning side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(1): 25-55

[70] Zhang J, Zheng M, Nan J, et al. A novel evaluation metric for deep learning-based side channel analysis and its extended application to imbalanced data. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(3): 73-96

[71] Kerkhof M, Wu L, Perin G, et al. Focus is key to success: A focal loss function for deep learning-based side-channel analysis//Proceedings of the 13th International Workshop on Constructive Side-Channel Analysis and Secure Design. Leuven, Belgium, 2022: 29-48

[72] Kerkhof M, Wu L, Perin G, et al. No (good) loss no gain: Systematic evaluation of loss functions in deep learning-based side-channel analysis. Journal of Cryptographic Engineering, 2023, 13(3): 311-324

[73] Zaid G, Bossuet L, Habrard A, et al. Efficiency through diversity in ensemble models applied to side-channel attacks: A case study on public-key algorithms. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(3): 60-96

[74] Masure L, Dumas C, Prouff E. A comprehensive study of deep learning for side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(1): 348-375

[75] Prechelt L. Early stopping—But when? Neural Networks: Tricks of the Trade, 1996, 1524: 55-69

[76] Picek S, Heuser A, Jovic A, et al. The curse of class imbalance and conflicting metrics with machine learning for side-channel evaluations. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2019, 2019(1): 1-29

[77] Perin G, Buhan I, Picek S. Learning when to stop: A mutual information approach to fight overfitting in profiled side-channel analysis. Cryptology ePrint Archive, 2020: 58

[78] Robissout D, Zaid G, Colombier B, et al. Online performance evaluation of deep learning networks for profiled side channel analysis//Proceedings of the 11th International Workshop on Constructive Side-Channel Analysis and Secure Design. Lugano, Switzerland, 2021: 200-218

[79] Standaert F X, Malkin T G, Yung M. A unified framework for the analysis of side-channel key recovery attacks//Proceedings of the 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2009). Cologne, Germany, 2009: 443-461

[80] Perin G, Wu L, Picek S. The need for speed: A fast guessing entropy calculation for deep learning-based SCA. Cryptology ePrint Archive, 2021: 1592

[81] Wu L, Perin G, Picek S. On the evaluation of deep learning based side-channel analysis//Proceedings of the 13th International Workshop on Constructive Side-Channel Analysis and Secure Design. Leuven, Belgium, 2022: 49-71

[82] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research, 2012, 13(2): 281-305

[83] Ruder S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016

[84] Perin G, Chmielewski Ł, Picek S. Strength in numbers: Improving generalization with ensembles in machine learning-based profiled side channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(4): 337-364

[85] Rijsdijk J, Wu L, Perin G, et al. Reinforcement learning for hyperparameter tuning in deep learning-based side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(3): 677-707

[86] Wu L, Perin G, Picek S. I choose you: Automated hyperparameter tuning for deep learning-based side-channel analysis. IEEE Transactions on Emerging Topics in Computing, 2022, 12(2): 546-557

[87] Weissbart L. On the performance of multilayer perceptron in profiling side-channel analysis. Cryptology ePrint Archive, 2019: 1476

[88] Li H, Krček M, Perin G. A comparison of weight initializers in deep learning-based side-channel analysis//Proceedings of the Workshops on Applied Cryptography and Network Security. Rome, Italy, 2020: 126-143

[89] Perin G, Picek S. On the influence of optimizers in deep learning based side-channel analysis//Proceedings of the 27th International Conference on Selected Areas in Cryptography. Halifax, Canada, 2021: 615-636

[90] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002, 16: 321-357

[91] Mukhtar N, Fournaris A P, Khan T M, et al. Improved hybrid approach for side-channel analysis using efficient convolutional neural network and dimensionality reduction. IEEE Access, 2020, 8: 184298-184311

[92] Pu S, Yu Y, Wang W, et al. Trace augmentation: What can be done even before preprocessing in a profiled SCA?//Proceedings of the 16th International Conference on Smart Card Research and Advanced Applications. Lugano, Switzerland, 2018: 232-247

[93] Kim J, Picek S, Heuser A, et al. Make some noise. Unleashing the power of convolutional neural networks for profiled side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2019, 2019(3): 148-179

[94] Luo Z, Zheng M, Wang P, et al. Towards strengthening deep learning-based side channel attacks with mixup//Proceedings of the IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications. Shenyang, China, 2021: 791-801

[95] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017

[96] Won Y S, Jap D, Bhasin S. Push for more: On comparison of data augmentation and smote with optimised deep learning architecture for side-channel//Proceedings of the 21st International Conference on Information Security Applications. Jeju Island, Republic of Korea, 2020: 227-241

[97] Wang P, Chen P, Luo Z, et al. Enhancing the performance of practical profiling side-channel attacks using conditional generative adversarial networks. arXiv preprint arXiv:2007.05285, 2020

[98] Wu L, Picek S. Remove some noise: On pre-processing of side channel measurements with autoencoders. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(4): 389-415

[99] Lei Q, Yang Z, Wang Q, et al. Autoencoder assist: An efficient profiling attack on high-dimensional datasets//Proceedings of the International Conference on Information and Communications Security. Canterbury, UK, 2022: 324-341

[100] Botocan C A. One network to rule them all. An autoencoder approach to encode datasets. Cryptology ePrint Archive, 2022: 890

[101] Wu L, Perin G, Picek S. The best of two worlds: Deep learning-assisted template attack. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2022, 2022(3): 413-437

[102] Zhang F, Shao B, Xu G, et al. From homogeneous to heterogeneous: Leveraging deep learning based power analysis across devices//Proceedings of the 57th ACM/IEEE Design Automation Conference. San Francisco, USA, 2020: 1-6

[103] Golder A, Das D, Danial J, et al. Practical approaches toward deep-learning-based cross-device power side-channel attack. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2019, 27(12): 2720-2733

[104] Das D, Golder A, Danial J, et al. X-DeepSCA: Cross-device deep learning side channel attack//Proceedings of the 56th Annual Design Automation Conference. Las Vegas, USA, 2019: 1-6

[105] Wouters L, Van den Herrewegen J, D Garcia F, et al. Dismantling DST80-based immobiliser systems. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2020, 2020(2): 99-127

[106] Wu L, Won Y S, Jap D, et al. Explain some noise: Ablation analysis for deep learning-based physical side-channel analysis. Cryptology ePrint Archive, 2021: 717

[107] Bhasin S, Chattopadhyay A, Heuser A, et al. Mind the portability: A warriors guide through realistic profiled side-channel analysis//Proceedings of the International Symposium on Network and Distributed System Security (NDSS 2020). San Diego, USA, 2020: 1-14

[108] Thapar D, Alam M, Mukhopadhyay D. TranSCA: Cross-family profiled side-channel attacks using transfer learning on deep neural networks. Cryptology ePrint Archive, 2020: 1258

[109] Genevey-Metat C, Gerard B, Heuser A. On what to learn: Train or adapt a deeply learned profile?. Cryptology ePrint Archive, 2020: 952

[110] Yu H, Shan H, Panoff M, Jin Y. Cross-device profiled side-channel attacks using meta-transfer learning//Proceedings of the 58th ACM/IEEE Design Automation Conference. San Francisco, USA, 2021: 703-708

[111] Cao P, Zhang H, Gu D, et al. AL-PA: Cross-device profiled side-channel attack using adversarial learning//Proceedings of the 59th ACM/IEEE Design Automation Conference. San Francisco, USA, 2022: 691-696

[112] Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. arXiv preprint arXiv:1911.02685, 2019

[113] Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test. The Journal of Machine Learning Research, 2012, 13(1): 723-773

[114] Bonawitz K, Ivanov V, Kreuter B, et al. Practical secure aggregation for federated learning on user-held data. arXiv preprint arXiv:1611.04482, 2016

[115] Wang H, Dubrova E. Federated learning in side-channel analysis//Proceedings of the International Conference on Information Security and Cryptology. Seoul, Republic of Korea, 2021: 257-272

[116] Breiman L. Bagging predictors. Machine Learning, 1996, 24: 123-140

[117] Egger M, Schamberger T, Tebelmann L, et al. A second look at the ASCAD databases//Proceedings of the 13th International Workshop on Constructive Side-Channel Analysis and Secure Design. Leuven, Belgium, 2022: 75-99

[118] Agrawal D, Rao J R, Rohatgi P. Multi-channel attacks//Proceedings of the 5th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2003). Cologne, Germany, 2003: 2-16

[119] Yu W, Chen J. Deep learning-assisted and combined attack: A novel side-channel attack. Electronics Letters, 2018, 54(19): 1114-1116

[120] Hettwer B, Fennes D, Leger S, et al. Deep learning multi-channel fusion attack against side-channel protected hardware //Proceedings of the 57th ACM/IEEE Design Automation Conference. San Francisco, USA, 2020: 1-6

[121] van der Valk D, Picek S. Bias-variance decomposition in machine learning-based side-channel analysis. Cryptology ePrint Archive, 2019: 570

[122] van der Valk D, Picek S, Bhasin S. Kilroy was here: The first step towards explainability of neural networks in profiled side-channel analysis//Proceedings of 11th International Workshop on the Constructive Side-Channel Analysis and Secure Design. Lugano, Switzerland, 2021: 175-199

[123] Raghu M, Gilmer J, Yosinski J, et al. SVCCA: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. Advances in Neural Information Processing Systems, 2017, 30: 6076-6085

[124] Perin G, Wu L, Picek S. I know what your layers did: Layer-wise explainability of deep learning side-channel analysis. Cryptology ePrint Archive, 2022: 1087

[125] Kuroda K, Fukuda Y, Yoshida K, et al. Practical aspects on non-profiled deep-learning side-channel attacks against AES software implementation with two types of masking countermeasures including RSM//Proceedings of the 5th Workshop on Attacks and Solutions in Hardware Security. Virtual, Republic of Korea, 2021: 29-40

[126] Lundberg S M, Lee S I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 2017, 30: 4765-4774

[127] Masure L, Dumas C, Prouff E. Gradient visualization for general characterization in profiling attacks//Proceedings of the 10th International Workshop on Constructive Side-Channel Analysis and Secure Design. Darmstadt, Germany, 2019: 145-167

[128] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015

[129] van der Valk D, Krcek M, Picek S, Bhasin S. Learning from a big brother-mimicking neural networks in profiled side-channel analysis//Proceedings of the 57th ACM/IEEE Design Automation Conference. San Francisco, USA, 2020: 1-6

[130] Lellis R, Soares R, Perin G. Pruning-based neural network reduction for faster profiling side-channel attacks//Proceedings of the 29th IEEE International Conference on Electronics, Circuits and Systems. Glasgow, UK, 2022: 1-4

[131] Perin G, Wu L, Picek S. Gambling for success: The lottery ticket hypothesis in deep learning-based side-channel analysis. Artificial Intelligence for Cybersecurity. Cham, Switzerland: Springer, 2022: 217-241

[132] Frankle J, Carbin M. The lottery ticket hypothesis: Finding sparse, trainable neural networks. arXiv preprint arXiv: 1803.03635, 2018

[133] Timon B. Non-profiled deep learning-based side-channel attacks with sensitivity analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2019, 2019(2): 107-131

[134] Won Y S, Han D G, Jap D, et al. Non-profiled side channel attack based on deep learning using picture trace. IEEE Access, 2021, 9: 22480-22492

[135] Saltelli A, Ratto M, Andres T, et al. Global Sensitivity Analysis: The Primer. West Sussex, England: John Wiley & Sons, 2008

[136] Lu X, Zhang C, Gu D. Attention-based non-profiled side channel attack//Proceedings of the International Symposium on Asian Hardware Oriented Security and Trust. Shanghai, China, 2021: 1-6

[137] Vijayakanthi G, Mohanty J P, Swain A K, et al. Differential metric based deep learning methodology for non-profiled Side Channel Analysis//Proceedings of the IEEE International Symposium on Smart Electronic Systems. Jaipur, India, 2021: 200-203

[138] Kwon D, Hong S, Kim H. Optimizing implementations of non-profiled deep learning-based side-channel attacks. IEEE Access, 2022, 10: 5957-5967

[139] Do N T, Le P C, Hoang V P, et al. MO-DLSCA: Deep learning based non-profiled side channel analysis using multi-output neural networks//Proceedings of the International Conference on Advanced Technologies for Communications. Hanoi Capital, Vietnam, 2022: 245- 250

[140] Zhang C, Lu X, Cao P, et al. A non-profiled side channel analysis based on variational lower bound related to mutual information. Science China Information Sciences, 2023, 66(1): 119

[141] Ramezanpour K, Ampadu P, Diehl W. SCARL: Side-channel analysis with reinforcement learning on the ascon authenticated cipher. arXiv preprint arXiv: 2006.03995, 2020

[142] Lee N, Hong S, Kim H. Single-trace attack using one-shot learning with Siamese network in non-profiled setting. IEEE Access, 2022, 10: 60778-60789

[143] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning. Advances in Neural Information Processing Systems, 2016, 29: 3630-3638

[144] Yang S, Zhou Y, Liu J, et al. Back propagation neural network based leakage characterization for practical security analysis of cryptographic implementations//Proceedings of the International Conference on Information Security and Cryptology. Seoul, Republic of Korea, 2011: 169-185

[145] Perin G, Chmielewski Ł, Batina L, et al. Keep it unsupervised: Horizontal attacks meet deep learning. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(1): 343-372

[146] Ramezanpour K, Ampadu P, Diehl W. SCAUL: Power side-channel analysis with unsupervised learning. IEEE Transactions on Computers, 2020, 69(11): 1626-1638

[147] Kwon D, Kim H, Hong S. Non-profiled deep learning-based side-channel preprocessing with autoencoders. IEEE Access, 2021, 9: 57692-57703

[148] Nomikos K, Papadimitriou A, Psarakis M, et al. Evaluation of hiding-based countermeasures against deep learning side channel attacks with pre-trained networks//Proceedings of the IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems. Austin, USA, 2022: 1-6

[149] Picek S, Jap D, Bhasin S. Poster: When adversary becomes the guardian—Towards side-channel security with adversarial attacks//Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. London, UK, 2019: 2673-2675

[150] Bertrand van Ouytsel C H, Bronchain O, Cassiers G, et al. How to fool a black box machine learning based side-channel security evaluation. Cryptography and Communications, 2021, 13(4): 573-585

[151] Rijsdijk J, Wu L, Perin G. Reinforcement learning-based design of side-channel countermeasures//Proceedings of the 11th International Conference on Security, Privacy, and Applied Cryptography Engineering. Kolkata, India, 2022: 168-187

[152] Krautter J, Tahoori M B. Neural networks as a side-channel countermeasure: Challenges and opportunities//Proceedings of the IEEE Computer Society Annual Symposium on VLSI. Tampa, USA, 2021: 272-277

[153] Hettwer B, Gehrer S, Güneysu T. Deep neural network attribution methods for leakage analysis and symmetric key recovery//Proceedings of the 26th International Conference on Selected Areas in Cryptography. Waterloo, Canada, 2020: 645-666

[154] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks//Proceedings of the 13th European

Conference on Computer Vision. Zurich, Switzerland, 2014: 818-833

[155] Bach S, Binder A, Montavon G, et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One, 2015, 10(7): e0130140

[156] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034, 2013

[157] Remmerswaal M G, Wu L, Tiran S, et al. AutoPOI: Automated points of interest selection for side-channel analysis. Cryptology ePrint Archive, 2023: 8

[158] Belghazi M I, Baratin A, Rajeshwar S, et al. Mutual information neural estimation//Proceedings of the 35th International Conference on Machine Learning. Stockholmsmassan, Sweden, 2018: 531-540

[159] Gabrie M, Manoel A, Luneau C, et al. Entropy and mutual information in models of deep neural networks. Advances in Neural Information Processing Systems, 2018, 31: 1826-1836

[160] Cristiani V, Lecomte M, Maurine P. Leakage assessment through neural estimation of the mutual information//Proceedings of the Workshops on Applied Cryptography and Network Security. Rome, Italy, 2020: 144-162

[161] Moos T, Wegener F, Moradi A. DL-LA: Deep learning leakage assessment: A modern roadmap for SCA evaluations. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(3): 552-598

[162] Ishai Y, Sahai A, Wagner D. Private circuits: Securing hardware against probing attacks//Proceedings of the 23rd Annual International Cryptology Conference (CRYPTO 2003). Santa Barbara, USA, 2003: 463-481

[163] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014

[164] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 248-255

附录 A.

鉴于本文使用了较多英文缩写，在此列出英文缩写的对照，如下表所示.

表 5　英文缩写对照表

缩写	全称或解释
SCA	Side-Channel Analysis，侧信道分析
PSCA	Profiling Side-Channel Analysis，建模侧信道分析
NPSCA	Non-Profiling Side-Channel Analysis，非建模侧信道分析
ML-SCA	Machine Learning-based Side-Channel Analysis，基于机器学习的侧信道分析
DL-SCA	Deep Learning-based Side-Channel Analysis，基于深度学习的侧信道分析
DPA	Differential Power Analysis，差分功耗分析
CPA	Correlation Power Analysis，相关功耗分析
MIA	Mutual Information Analysis，互信息分析
SNR	Signal-to-Noise Ratio，信噪比
PoI	Point of Interest，在 SCA 中通常指选择的特征点
NICV	Normalized Inter-Class Variance，标准化类间方差
SOSD	Sum Of Squared pairwise Differences，平方对差之和
TVLA	Test Vector Leakage Assessment，测试向量泄漏评估
MLP	Multi-Layer Perceptron，多层感知器
CNN	Convolutional Neural Network，卷积神经网络
CE	Cross Entropy，交叉熵
MSE	Mean Square Error，均方误差
GAN	Generative Adversarial Network，生成对抗网络

（续　表）

缩写	全称或解释
DL-PSCA	Deep Learning-based Profiling Side-Channel Analysis，基于深度学习的建模侧信道分析
DLNPSCA	Deep Learning-based Non-Profiling Side-Channel Analysis，基于深度学习的非建模侧信道分析
GE	Guessing Entropy，猜测熵
MCNN	Multi-scale Convolutional Neural Network，多尺度卷积神经网络
ID/HW/HD	Identity/Hamming Weight/Hamming Distance，三种标签方法
DA	Data Augmentation，数据增强
DAC	Design Automation Conference，设计自动化大会
VLSI	Very-Large-Scale Integration，超大规模集成技术会议
CHES	Cryptographic Hardware and Embedded Systems，密码学硬件与嵌入式系统会议
TCHES	IACR Transactions on Cryptographic Hardware and Embedded Systems，CHES 论文发表期刊
NDSS	Network and Distributed System Security Symposium，网络与分布式系统安全研讨会
FFT	Fast Fourier Transform，快速傅里叶变换
PCA	Principal Component Analysis，主成分分析
DTW	Dynamic Time Warping，动态时间规划
XAI	Explainable Artificial Intelligence，人工智能可解释性
DDLA	Differential Deep Learning Analysis，差分深度学习分析
SA	Sensitivity Analysis，敏感性分析
NMIA	Neural Mutual Information Analysis，神经互信息分析
LWC	Layerwise Correlation，逐层相关性

XIAO Chong, Ph. D. candidate. His current research interests include side-channel analysis and cryptanalysis.

TANG Ming, Ph. D., professor. Her current research interests include cryptography, secure design of cryptography chips, lightweight countermeasures against side-channel analysis, and systematic research on side-channel analysis.

Background

This paper presents a comprehensive survey on the development of Deep Learning-based Side-Channel Analysis (DL-SCA). Ever since deep learning techniques were introduced to the field of side-channel analysis, it has been proven a powerful and attacker-friendly type of attack. DL-SCA outperforms traditional side-channel analysis in various ways and it has been acknowledged as the most powerful side-channel attack among researchers. Deep learning techniques are renovating all aspects of the side-channel domain, from basic attacking methods to countermeasures, and leakage detection. Although there is an internationally recognized Systematization of Knowledge (SoK) paper, as citation [9] in this article we intend to provide a distinct perspective on the advancement of DL-SCA. Differing from their perspective of presenting research works at different stages of DL-SCA, our work follows the roadmap of traditional SCA in attacking, defense, and detection areas and their integration with deep learning techniques.

Our classification refers to the traditional side-channel research and then subclassifies related works according to their intents of research. In general, the intersection of deep learning techniques and the side-channel domain is mainly focused on profiled side-channel analysis, in which a large amount of works revolve around adaptability, generalizability, explainability, and cost-effectiveness for deep learning to better cope with side-channel analysis. Accordingly, we present comparisons among similar works and analyze the prospect. In recent years, deep learning-based non-profiled side-channel analysis has also started to evolve, but progress has been slow overall. Some of the interesting work focuses on using deep learning as an auxiliary tool to boost the performance of side-channel analysis. In addition, there is a relative lack of attention on the countermeasure of side-channel analysis as well as leakage detection. Finally, we statistically analyze the relevant papers by year, summarize the trending topics, and give insights and suggestions for future dedication in the under-development domain.

This work was supported by the National Key R&D Program of China (No. 2022YFB3103800).