对加掩加密算法的盲掩码模板攻击

王燚，吴震，蔺冰
（成都信息工程大学网络空间安全学院，四川 成都 610225）

摘 要：加掩是在加密算法的实现中使用随机掩码使敏感信息的泄露能耗随机化，从而防止差分能量攻击的技术手段。目前，对加掩防护加密算法的模板攻击的方法均要求攻击者在学习阶段了解使用的掩码。这一要求不仅提高了攻击的条件，同时也可能导致模板学习阶段使用的加密代码与实际设备的代码有所不同，进而导致对实际设备攻击效果较差。盲掩码模板攻击不需要了解训练能迹使用的掩码，直接学习无掩中间组合值的模板，以此攻击加掩加密设备。实验中分别采用传统的高斯分布和神经网络建立模板。实验结果证明这种方法是可行的，而且基于神经网络的盲掩码模板攻击对加掩加密设备的攻击成功率非常接近于传统模板攻击对无掩加密设备的攻击成功率。

关键词：侧信道攻击；模板攻击；盲掩码攻击；加掩防护；神经网络
中图分类号：TP309.1
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2019007

Blind mask template attacks on masked cryptographic algorithm

WANG Yi, WU Zhen, LIN Bing
College of Information Security Engineering, Chengdu University of Information Technology, Chengdu 610225, China

Abstract: Masking is a countermeasure against differential power analysis (DPA) attacks on cryptographic devices by using random masks to randomize the leaked power of sensitive information. Template attacks (TA) against cryptographic devices with masking countermeasure by far require attackers have knowledge of masks at the profiling phase. This requirement not only increase the prerequisite of template attacking, but also lead to some sort of difference between the experimental encryption codes of the profiling device and the codes of commercial cryptographic devices, which might degrade performance in real world attacking. Blind mask template attack directly learns templates for the combination of no mask intermediate values without the need of knowing the masks of training power traces, and then uses these templates to attack masked cryptographic devices. Both traditional Gaussian distribution and neural network were adopted as the templates in experiments. Experimental results verified the feasibility of this new approach. The success rate of neural network based blind mask template attacking against masked cryptographic devices is very close to that of traditional template attacks against cryptographic devices without masking countermeasure.

Key words: side channel attack, template attack, blind mask template attack, masking countermeasure, neural network

1 引言

能量分析是利用加密设备在加密过程中泄露能量与敏感信息的相关性来攻击其密钥的一种方法。能量分析攻击有各种各样的实现方法[1-4]，但大体可分为无学习和有学习这两类。无学习的能量攻

收稿日期：2018−03−13；**修回日期**：2018−08−03
通信作者：吴震，wuzheng@cuit.edu.cn
基金项目：“十三五”国家密码发展基金资助项目（No.MMJJ20180224）；国家重点研发计划基金资助项目（No.2018YFB0904900, No.2018YFB0904901）；四川省教育厅科研基金资助项目（No.17ZB0082）
Foundation Items: The 13th Five-Years National Cryptogram Development Fund (No.MMJJ20180224), The National Key Research and Development Program of China (No.2018YFB0904900, No.2018YFB0904901), Sichuan Provincial Education Department Scientific Research Projects (No.17ZB0082)

击包括简单能量分析（SPA, simple power analysis）[5]和差分能量分析（DPA, differential power analysis）[6]。DPA根据某种泄露模型，利用统计方式来识别猜测密钥的正确性，主要的泄露模型有汉明重量模型和汉明距离模型[7]。DPA采用的统计识别方法主要有分组能耗均值差[8]、Pearson相关系数[9]和互信息[10]。有学习的能量分析利用一台可控制的、与目标设备相似或相同的设备（称为实验设备[11]），精确地建立泄露信息的能耗模型（称为模板），以此对同类设备进行攻击。有学习的能量分析包括模板攻击（TA, template attack）[11]、随机攻击（SA, stochastic attack）[12]等。无学习的能量分析的优势是攻击者不需要拥有实验设备便直接对目标设备进行攻击，缺点是在能迹较少时攻击成功率比较低。有学习的能量分析由于建立了关于泄露信息的精确能耗模型，攻击时仅需较少的能迹（甚至只需要一条能迹）就能攻击成功，但缺点也很明显，即攻击者必须拥有一台可控制的实验设备。模板攻击是有学习能量分析中最成功、研究得最深入的一种方法，其中模板最基本的形式是多元高斯分布，其分布参数在模板训练阶段统计计算。为提高模板的攻击效果，部分研究着眼于改进多元高斯分布协方差矩阵，分别采用单位协方差、共享协方差、池化协方差等方法，在一定程度上降低计算成本、提高攻击效果[13-15]。也有一些研究着眼于充分利用多种类型泄露信息的联合能量攻击，以提高模板攻击的效率[16]。其他提高模板攻击效率的方法包括：利用相位相关性消除构架模板中的数据干扰，以构建高质量的模板[17]；采用主成分分析（PCA, principal component analysis）对能耗数据进行降维处理，有效降低协方差矩阵的计算复杂度并在一定程度上提高攻击效率[18-19]。模板攻击的另一种研究思路是采用机器学习中的分类算法替代传统的多元高斯分布模板，例如，将贝叶斯分类的算法应用在模板攻击中[20]，利用多类支持向量机作为模板实施模板攻击[21-22]，采用神经网络作为模板进行模板攻击[23-24]等。

为应对能量分析对加密设备造成的显著威胁，研究人员发展出很多对抗措施，如抖动防护、加掩防护[25]等。加掩防护的原理是使用一个或多个随机掩码与可能发生泄露的中间值进行异或，从而使其泄露的能耗随机化，达到防御DPA攻击的目的。而研究人员又发展出高阶DPA攻击[26-27]的方法来攻击具有加掩防护的加密设备。高阶DPA攻击利用多个中间值的组合值与它们对应泄露能耗的组合值具有一定的相关性这一事实实施攻击。高阶DPA比一阶DPA攻击需要更多的能迹，并且非常耗时。这是因为攻击者无法通过统计手段发现多个中间值的能耗泄露的准确位置，只能在猜测的泄露范围内对能耗进行交叉组合。这种方式极大地增加了攻击的计算量，降低了攻击的效率。虽然文献[28]中提出采用FFT（fast Fourier transform）加速能耗组合计算的方法，但仍不能从根本上解决高阶DPA攻击效率低的问题。

模板攻击也被应用于攻击具有加掩防护的加密设备（后文中简称为加掩设备）。一种途径是利用模板辅助对加掩设备实施DPA攻击[29]，包括几种具体方法。1) 在训练阶段根据已知掩码和明文，建立关于带掩中间值的模板，用于识别和计算攻击能迹的带掩中间组合值，攻击时以此代替能耗组合值实施高阶DPA攻击。2) 在训练阶段根据已知掩码建立关于掩码的模板，攻击时识别攻击能迹的掩码并过滤特定的掩码以造成掩码不平衡，从而可以实施一阶DPA攻击。3) 在训练时根据已知的明文和掩码，发现多个中间值的准确泄露位置，并计算其能耗组合值，建立无掩中间组合值关于能耗组合值的模板，用于在攻击时识别攻击能迹的无掩中间组合值，进而通过计算与猜测密钥对应的猜测中间组合值的相关系数实施DPA攻击。在不采用DPA而完全基于模板攻击的方法中，一种方法是采用<密钥，掩码>对的模板实施攻击[30]，这种方法需要建立大量的模板，并同时攻击掩码和密钥；另一种方法是使用支持向量机（SVM, support vector machine）和神经网络（NN, neural network）建立攻击掩码的模板以及攻击带掩中间值的模板，采用先攻击得到掩码，去掩后攻击带掩中间值的方式获取设备密钥[31-32]。根据了解，所有对加掩设备的模板攻击均需要攻击者在训练阶段（profiling phase）了解使用的掩码。这样，在训练阶段使用的加密代码中必须包含对掩码的输入或输出，这与真实加密设备的代码必然有所不同。这种差异可能导致习得的能耗模型与真实设备的能耗模板存在一定差异，进而可能导致对真实设备的攻击成功率不高。此外，对硬实现的加密算法，攻击者无法改变加

密的实现，无法获取设备加密时采用的随机掩码，因而以上方法均无法实施。

本研究的目标是攻击者使用一个仅知道密钥的加掩设备作为实验设备，不需要了解掩码，直接学习关于无掩中间组合值的模板，从而可以采用模板攻击的方法获取同类设备的密钥。这种方法称为盲掩码模板攻击。本文的主要贡献包括以下 3 个方面。

1) 提出了对加掩加密算法的盲掩码模板攻击方法，并从实验上证实了该方法的可行性。盲掩码模板攻击极大地降低了对加掩防护加密设备实施模板攻击条件和门槛。

2) 提出采用人工神经网络作为优化的盲掩码模板。通过大量的训练和攻击的实验，指出了在使用信噪比极低的能耗特征数据的神经网络训练中，必须采用全批而不能采用 mini-batch 的训练策略。

3) 针对实验中神经网络训练极易出现过拟合的问题，分析并指出了导致这种过拟合的原因，并提出了有针对性的部分特征 PCA 的数据预处理方法。该方法与 L2 权重正则化配合，可以有效地防止神经网络的训练发生过拟合，且不易出现欠拟合的现象。实验结果表明，基于神经网络的盲掩码模板攻击的成功率接近对无加掩防护设备的模板攻击。在对实验数据的攻击中，仅需 8 条攻击能迹就能够达到 100%的成功率，最少攻击能迹数仅为一条（成功率为 12%）。

2　对加掩设备的模板攻击研究现状

Oswald 等[29]提出了几种对加掩设备的模板攻击方法。第一种称为预处理前的模板，该方法要求攻击者了解训练能迹所使用的掩码 r，从而可以计算出带掩的 Sbox（substitution-box）的输入 u_r 和输出 v_r。训练时建立 u_r 和 v_r 的能耗模型 $\Pr(\boldsymbol{e}\,|\,u_r)$ 和 $\Pr(\boldsymbol{e}\,|\,v_r)$。攻击时使用这些能耗模型识别攻击能迹的 u_r 和 v_r，并计算带掩中间组合值 $\mathrm{pre}(u_r,v_r)$ 与根据猜测密钥 k 计算的猜测无掩中间组合值 $\mathrm{pre}(u,v)$ 的相关系数，从而实施 CPA（correlation power analysis）攻击。其中能迹预处理函数 $\mathrm{pre}(\bullet)$ 和中间值组合函数 $\mathrm{comb}(\bullet)$ 采用三角多项式实现组合，以提高两者的理论相关系数。第二种方法称为预处理中的模板，该方法同样要求攻击者了解训练能迹所使用的掩码。攻击者建立关于掩码的能耗模型 $\Pr(\boldsymbol{e}\,|\,r)$，用于识别攻击能迹的掩码 r。攻击时通过抛弃使用部分特定掩码的能迹，使攻击能迹中的掩码分布不平衡，导致中间值不能完全被掩码所掩盖，从而可以使用一阶 DPA 攻击设备密钥。第三种方法称为预处理后的模板，该方法利用已知的密钥、明文和掩码，在训练能迹中发现 Sbox 输入和输出的准确泄露位置，据此计算能耗组合值 pre，然后建立无掩中间组合值 comb 与 pre 的能耗模型 $\Pr(\mathrm{pre}\,|\,\mathrm{comb})$。攻击时利用该模型识别攻击能迹的无掩中间组合值 comb，并据此实施高阶 DPA 攻击。文献[29]中指出，由于计算能耗组合值 pre 时丢失了大量信息，这种攻击方法需要大量攻击能迹，不能体现模板攻击的优势。Lemke-Rust 和 Paar[30]提出建立<中间值 c，掩码 r>对的能耗模型 $\Pr(\boldsymbol{e}\,|\,c,r)$，攻击时需要同时攻击密钥和掩码对 $<\hat{k},\hat{r}>$。由于 $<k,r>$ 的组合数非常多，只能实施按位的攻击。而按位攻击的弱点是，当攻击第 i 位时，其余位的数据产生的能耗对第 i 位而言就是噪声，这使得攻击时数据的信噪比更低，需要更多的能迹才能攻击成功。Lerman 等[31]提出使用概率型支持向量机建立掩码的能耗模型 $\mathrm{SVM}(\boldsymbol{e}\,|\,r)$，用于攻击时识别攻击能迹的掩码 $\hat{r}$，一旦攻击能迹的掩码被获知，加掩防护也就失去了作用。Gilmore 等[32]采用的思路与上文基本相同，但使用神经网络建立掩码的能耗模型 $\mathrm{NN}(\boldsymbol{e}\,|\,r)$，获得了更高的掩码识别准确率。

上述对加掩设备的模板攻击均需要在学习阶段了解训练能迹所使用的随机掩码。因此攻击者必须能够完全控制加密设备的代码，以便输出每次使用的掩码，或从外部输入每次使用的掩码。这一要求不仅不易达到，而且其实验设备使用的代码由于需要输入或输出随机掩码，与真实加密设备的代码必然有所不同，对真实设备进行实际攻击时往往成功率比较低。

3　对加掩设备的盲掩码模板攻击原理

3.1　盲掩码模板攻击的基本思路

盲掩码模板攻击的思路来源于高阶 DPA 攻击。高阶 DPA 是对加掩设备的无学习攻击方法。高阶 DPA 攻击不需要了解设备的随机掩码，利用多个无掩中间值的组合值对能迹进行分组，各分组的组合能耗的均值呈现出差异[26]，从而可以采用组间差实施攻击。高阶 CPA 是高阶 DPA 的一种特殊方式。Coron 等[33]证明了尽管由于随机掩码的作用，无掩

中间值 c_i 与其对应操作 o_i 的泄露能耗 e_i 不再具有相关性，但特定的多个无掩中间值的组合值 $\mathrm{comb}(c_1,\cdots,c_d)$ 与相应的能耗组合值 $\mathrm{pre}(e_1,\cdots,e_d)$ 之间却存在一定程度的线性相关性，即它们的 Pearson 相关系数 $\rho(\mathrm{comb},\mathrm{pre})\neq 0$，从而可以实施高阶 CPA 攻击。

无掩中间组合值 comb 与能耗组合值 pre 有一定程度的相关性，意味着能够根据 pre 的值以一定的概率识别出正确的 comb 值，即存在概率分布 $\Pr(\mathrm{comb}\,|\,\mathrm{pre})$。在统计或训练该概率分布时，由于无掩中间组合值 comb 与掩码无关，因此可以消除训练时对掩码的依赖。如果能够在训练阶段得到该概率分布，就可以采用模板攻击的方法识别正确的猜测密钥，如式(1)所示。

$$\hat{K}=\underset{K}{\mathrm{argmax}}\prod_{i=1}^{m}\Pr(\mathrm{comb}(x_i,k)|\mathrm{pre}) \tag{1}$$

其中，$\mathrm{comb}(x_i,k)$ 表示根据明文 x_i 和猜测密钥 k 计算得到的无掩中间组合值。由于训练概率分布 $\Pr(\mathrm{comb}\,|\,\mathrm{pre})$ 时并不使用掩码，攻击时也不需要掩码，因此本文把这种模板攻击方式称为盲掩码模板攻击。

盲掩码模板攻击的关键在于找到能耗有一定程度泄露的无掩中间组合值。不同加密算法或同一加密算法的不同实现，无掩中间组合值 comb 的组合形式可能不同。由于训练设备的密钥已知，可以计算得到各种无掩中间组合值与能耗组合值的相关系数，并选择相关系数较大的无掩中间组合值的组合形式。

盲掩码模板攻击的首要目标是降低对加掩防护加密算法实施的模板攻击的门槛。根据上述的论述，通过对无掩中间值建模，可以消除创建模板时对掩码的依赖，从而实现了该目标。但之所以要对加掩防护加密算法实施模板攻击，就是希望在攻击阶段可以使用少量攻击能迹获取设备密钥，提高攻击的效率，要实现该目标，就需要高质量的模板。但由于掩码的存在，无掩中间组合值对应的能耗组合值中包含了大量掩码带来的噪声，且这类噪声的高斯性不强，可能导致传统的基于高斯分布的模板在攻击中需要大量能迹，而无法体现出模板攻击的优势。而神经网络不需要假设噪声的分布，并具有强大的非线性转换能力，可以用于创建质量更高的模板。

3.2 基于一元高斯分布的盲掩码模板攻击

如果能够找到泄露无掩中间组合值 $\mathrm{comb}(c_1,\cdots,c_d)$ 信息的能耗组合形式 $\mathrm{pre}(e_1,\cdots,e_d)$，其中 $e_1,\cdots,e_d$ 对应的样本位置为 $\boldsymbol{t}=\{t_1,\cdots,t_d\}$，则模板攻击中各训练能迹和攻击能迹可以被转换为一个标量的能耗组合值 pre。设无掩中间组合值 comb 有 n 个取值，其中 comb_i 对应的能迹分组为 $S_i, i\in[1,n]$，如果 S_i 分组内各能迹的能耗组合值 pre 呈高斯分布，则模型 $\Pr(\mathrm{pre}\,|\,\mathrm{comb})$ 可以用一元高斯分布来描述，如式(2)所示。

$$\Pr\left(\mathrm{pre}|\mathrm{comb}_i\right)=\frac{1}{\sqrt{2\pi\sigma_i^2}}\mathrm{e}^{-\frac{(\mathrm{pre}-\mu_i)^2}{2\sigma_i^2}} \tag{2}$$

其中，

$$\mu_i=\frac{1}{|S_i|}\sum_{\mathrm{pre}\in S_i}\mathrm{pre}$$

$$\sigma_i^2=\frac{1}{|S_i|-1}\sum_{\mathrm{pre}\in S_i}(\mathrm{pre}-\mu_i)^2$$

若攻击时使用 m 条攻击能迹，则最可能的密钥为

$$\hat{k}=\underset{k}{\mathrm{argmax}}\prod_{j=1}^{m}\Pr\left(\mathrm{pre}_j\,|\,\mathrm{comb}(x_j,k)\right) \tag{3}$$

基于一元高斯分布的盲掩码模板攻击实际上是首先对能迹进行预处理，将各能迹转换为一个组合能耗值 pre。在实际情况中，无掩中间组合值可能在多种能耗组合形式下发生泄露，而一元高斯分布的模板仅采用其中一种泄露，将丢失大量的泄露信息，对模板的质量和攻击效率不利。

3.3 基于多元高斯分布的盲掩码模板攻击

假设泄露无掩中间组合值 $\mathrm{comb}(c_1,\cdots,c_d)$ 信息的能耗组合形式 $\mathrm{pre}(e_1,\cdots,e_d)$ 有多种，它们包含的能耗样本位置的并集为 $\boldsymbol{t}=\{t_1,\cdots,t_n\}, n>d$。提取能迹 p 上位置集 $\boldsymbol{t}$ 的能耗作为能迹特征向量 $\boldsymbol{e}=p[\boldsymbol{t}]$，则 $\Pr(\boldsymbol{e}\,|\,\mathrm{comb})$ 可以用多元高斯分布来描述。

$$\Pr\left(\boldsymbol{e}\,|\,\mathrm{comb}_i\right)=\frac{\mathrm{e}^{-\frac{1}{2}(\boldsymbol{e}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{e}-\boldsymbol{\mu}_i)}}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_i|}} \tag{4}$$

其中，

$$\boldsymbol{\mu}_i=\frac{1}{|S_i|}\sum_{\boldsymbol{e}\in S_i}\boldsymbol{e}$$

$$\boldsymbol{\Sigma}_i=\frac{1}{|S_i|-1}(\boldsymbol{e}-\boldsymbol{\mu}_i)(\boldsymbol{e}-\boldsymbol{\mu}_i)^{\mathrm{T}},\boldsymbol{e}\in S_i$$

若攻击时使用 m 条攻击能迹，则最可能的密钥为

$$\hat{k} = \underset{k}{\operatorname{argmax}} \prod_{j=1}^{m} \Pr\left(\boldsymbol{e}_j \mid \operatorname{comb}(x_j, k)\right) \tag{5}$$

3.4 基于神经网络的盲掩码模板攻击

对加掩防护加密算法实施模板攻击的目的是以少量能迹获取设备密钥，提高攻击效率。实现该目标的前提是建立高质量的模板。而由于随机掩码的存在，能耗中包含了大量的数字噪声，且其噪声可能并不符合高斯分布。这将导致前两种基于高斯分布的模板质量低下，其攻击效率并不能体现出模板攻击的优势。因此，需要探索一种弱假设或无假设的创建模板的方法。

模板攻击的基本操作是根据能耗判断某个中间值（或中间组合值）的概率，其本质是机器学习的概率化分类识别。因此，可以用各种机器学习模型来代替高斯分布。目前，已有很多这方面的研究，例如使用支持向量机（SVM, support vector machine）[21-34]、随机森林（RF, random forest）[35-36]、人工神经网络（ANN, artificial neural network）[23-24]等作为概率型的分类模型。能耗的概率分布可能不符合或不完全符合高斯分布，这会导致基于高斯分布的模板攻击成功率不高。这些机器学习模型没有预先假设的概率分布，可以通过训练习得任何类型的概率分布，因此攻击成功率可以高于基于高斯分布的模板攻击。从已发表的实验结果看，当训练能迹足够时，随机森林树的攻击效果不如 SVM 和 ANN。这是因为树的构造中必须将能耗离散化处理，而导致了信息的丢失。SVM 与 ANN 的结果接近，但由于 SVM 本质上是二类分类器，多类 SVM 是一对一 SVM 或一对其余 SVM 的集合，训练多类 SVM 需要训练大量的二分类 SVM，因此训练过程非常漫长。同时，根据实验，对于信噪比很低的数据，SVM 的支持向量中包含了大部分训练向量，这意味着训练出现了严重的过拟合。因此，本研究采用神经网络代替高斯分布作为模板。

设训练阶段习得的分类器的概率模型为 $\mathrm{NN}(\boldsymbol{e} \mid \mathrm{comb})$，则攻击密钥的操作为

$$\hat{k} = \underset{k}{\operatorname{argmax}} \prod_{j=1,\cdots,m} \mathrm{NN}\left(\boldsymbol{e}_j \mid \operatorname{comb}(x_j, k)\right) \tag{6}$$

基于神经网络的盲掩码模板攻击中，神经网络的训练效果至关重要。由于是对无掩中间组合值建模，相应的能耗组合值中包含了大量掩码带来的数字噪声，其信噪比极低。在训练中容易出现 3 种现象：1) 神经网络在训练中无法收敛；2) 在训练精度很低时就出现过拟合；3) 同时出现这 2 种情况。关于极低信噪比数据的神经网络的训练方法还未见有专门的研究。通过大量实验，本文提出了一种数据预处理的方法——分段 PCA 处理，并探索出综合现有的防过拟合的措施的神经网络训练方法。

4 实验分析

4.1 实验数据

盲掩码模板攻击的实验采用公开的 DPA Contest V4 能迹集。DPA Contest[37]公开了一组标准的能迹集，供研究人员测试和比较各种侧信道攻击方法，其第 4 版（DPA Contest V4）提供了在 Atmel ATMega-163 智能卡上实现 AES-256（advanced encryption standard-256）加密算法，使用电磁探测获取的能迹集，共包括 10 万条能迹。DPA Contest V4 中采用了 RSM (rotating S-boxes making) [38]加掩防护措施。本文的实验中使用其中 5 万条能迹，其中 3 万条用于模板训练，1 万条用于训练的验证，1 万条用于攻击实验。

4.2 攻击目标

根据第 3 节的介绍，盲掩码模板攻击的目标中间组合值来源于高阶 CPA 攻击的目标中间组合值。因此，本文首先对 DPA Contest V4 实施了二阶 CPA 攻击。二阶 CPA 攻击的目标中间组合值 comb 是 Sbox 无掩输入和输出的汉明距离。以第一轮 Sbox0 为例，其对应的子密钥为 k=108。设 Sbox0 的无掩输入为 $u = x \oplus k$，Sbox0 的无掩输出为 $v = \mathrm{Sbox}\ (x \oplus k)$；汉明距离为 $\mathrm{comb} = \mathrm{hw}(u \oplus v)$；对应的能耗组合值为 $\mathrm{pre} = e_1 \times e_2$，其中 e_1、e_2 分别代表 Sbox0 输入、输出操作的泄露能耗；位置在攻击前是未知的。通过对 3 万条训练能迹的平均能迹的可视化观察，确定了 Sbox0 输入和输出的 2 个大致的样本范围。pre 是这 2 个范围中任意一对位置上能耗的乘积（能迹进行了中心化预处理[39]）。攻击方法为

$$\begin{aligned}\hat{k} &= \underset{k}{\operatorname{argmax}}\, \rho(\mathrm{comb}, \mathrm{pre}) \\ &= \underset{k}{\operatorname{argmax}}\, \rho(\mathrm{hw}(u(x,k) \oplus v(x,k)), e_1 \times e_2)\end{aligned} \tag{7}$$

二阶 CPA 攻击结果如图 1 所示。

key	corr	t1	t2
108	0.173459	50178	51587
108	0.172838	50178	51585
108	0.163745	50178	51584
108	0.163171	50178	51586
108	0.161561	50178	51588
108	0.155341	50178	51589
108	0.154832	50180	51587
108	0.150743	50179	51587
108	0.149912	50180	51588
108	0.149223	50177	51587
108	0.147163	50180	51585
108	0.145576	50177	51584

图 1 二阶 CPA 攻击部分结果

图 1 中，t_1 与 t_2 表示攻击时组合能耗的位置，corr 表示中间组合值与能耗组合值的 Pearson 相关系数，key 为最可能的密钥。图中所示的猜测密钥 108 与正确的密钥完全相同，说明攻击是成功的。

二阶 DPA 攻击成功，说明使用正确位置上的能耗组合值 $\text{pre}=e_1\times e_2$，能够以一定的概率识别出正确的中间组合值 $\text{comb}=\text{hw}(u\oplus v)$，即存在概率分布 $\Pr(\text{hw}(u\oplus v)|e_1\times e_2)$。这样，可以使用 $\text{comb}=\text{hw}(u\oplus v)$ 作为盲掩码模板攻击的目标。

此外，还可以选择以 $\text{comb}=u\oplus v$ 作为盲掩码模板攻击的目标，这是因为按照 $u\oplus v$ 值分组时，分组中的所有能迹的 $\text{hw}(u\oplus v)$ 是相同的，即有 $\Pr(u\oplus v|e_1\times e_2)=\Pr(\text{hw}(u\oplus v)|e_1\times e_2)$。也就是说，以 $u\oplus v$ 为攻击目标，理论上并不影响概率分布。更重要的是，如果使用 $\Pr(u\oplus v|e_1\times e_2)$ 识别出正确的 $u\oplus v$ 值，就可以根据已知明文 x 推导出唯一的 k 值；而使用 $\Pr(\text{hw}(u\oplus v)|e_1\times e_2)$ 识别出正确的 $\text{hw}(u\oplus v)$ 值，可能的 k 值不是唯一的。此外，$u\oplus v$ 的值实际上表达了所有的位组合形式。而物理电路上，不同的位的能耗不完全相同，以 $u\oplus v$ 为目标建模能够更精确地表达能耗模型，因此，以 $u\oplus v$ 为目标的攻击效率可能高于以 $\text{hw}(u\oplus v)$ 为目标。在后面的实验中，本文对这 2 种目标的模板攻击进行了比较，证实了这一推测。

4.3 特征选择

特征选择是模板攻击的重要步骤，准确选择信息泄露程度高的样本位置，对模板攻击能否成功以及成功率的高低有决定性的作用。最简单的特征选择是观察平均能迹的能耗变化规律，可视化地判断攻击目标操作的样本范围。但这种方法选择的特征区域包含大量与攻击目标信息无关的能耗，会对模板攻击造成非常不利的影响。目前基于统计的特征选择方法有平均能耗差（DOM，difference of means）[11]、能耗方差和（SOSD，sum of squared difference）[40]、能耗学生测试和（SOST，sum of student test）[40]、正则化类间差（NICV，normalized inter-class variance）[41]、基于子空间的方法[18, 42]等。这些方法都必须有足够的信息（如明文或密文、密钥、掩码等）来计算要攻击的目标值，并据此对训练能迹分组，然后统计各分组的平均能迹之间的差异，从而得到揭示目标值信息泄露的指标。

本研究的目标是在未知掩码的情况下实施模板攻击，这意味着在训练阶段并不知道各能迹的掩码，无法计算加掩的攻击目标值。因此上述方法均不适用。但 4.2 节中进行的二阶 CPA 攻击结果提供了 Sbox 输入和输出的泄露位置。从图 1 可以看出，二阶 CPA 在多个位置组合上都能够攻击成功。事实上，使用 3 万条能迹的二阶 CPA 在 2 302 个位置对上攻击成功，虽然其中包含了大量重复的位置，去重后，仍有 193 个泄露位置。这些泄露位置上信息泄露的程度并不相同，可以认为，二阶 DPA 攻击中相关系数较高的位置上，能耗的信息泄露更加明显。因此，仅选择部分泄露位置提取能耗特征，本文根据这些位置对所对应的相关系数，从高到低选择，并计算这些位置对中各位置的并集。

4.4 基于一元高斯分布的盲掩码模板攻击

基于高斯分布的模板攻击要求泄露能耗呈高斯分布。根据 3.2 节与 4.2 节所述，一元高斯分布盲掩码模板攻击中采用能耗组合值 pre，$\text{pre}=e_1\times e_2$，其中 e_1 与 e_2 分别是 Sbox0 输入和输出的能耗。首先检查 pre 是否符合高斯分布。设攻击的目标为 $\text{hw}(u\oplus v)$，在 $\text{hw}(u\oplus v)=4$ 的分组能迹上，取二阶 CPA 攻击中相关系数最高的一对位置的能耗作为 e_1 和 e_2。能耗组合值 pre 的正则化直方图与其理想高斯分布的对比如图 2 所示。

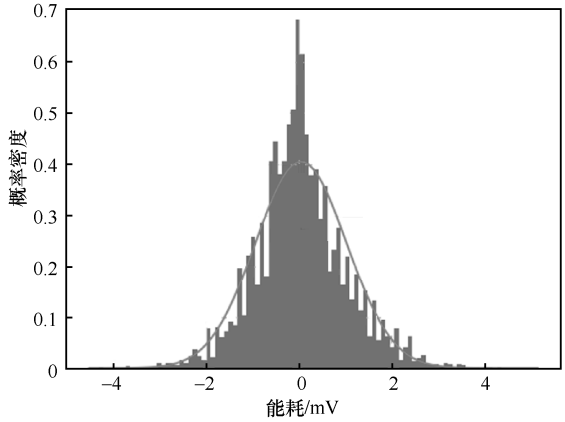

图 2 能耗组合值的高斯性定性分析

从图 2 中可以看出，$\mathrm{pre}=e_1\times e_2$ 的分布具有一定的高斯性，应该可以支持一元高斯分布的模板攻击。

据此，分别以 $\mathrm{comb}=\mathrm{hw}(u\oplus v)$ 和 $\mathrm{comb}=u\oplus v$ 为目标，训练一元高斯模板。 $\mathrm{comb}=\mathrm{hw}(u\oplus v)$ 时共有 9 个模板， $\mathrm{comb}=u\oplus v$ 时共有 163 个模板。使用测试能迹集进行不同攻击能迹数的攻击实验。图 3 中比较了这 2 种模板攻击以及二阶 CPA 攻击的成功率与攻击能迹数的关系。

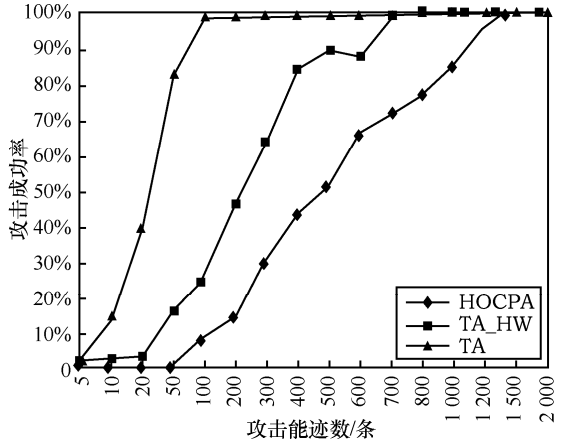

图 3　一元盲掩码模板攻击与二阶 DPA 攻击比较

图 3 中 HOCPA 表示二阶 CPA 攻击，TA_HW 表示以 $\mathrm{comb}=\mathrm{hw}(u\oplus v)$ 目标的盲掩码模板攻击，TA 表示以 $\mathrm{comb}=u\oplus v$ 为目标的盲掩码模板攻击。达到 100%攻击成功率时，HOCPA 需要的能迹数为 1 500 条，TA_HW 需要 700 条，TA 只需要 100 条。TA_HW 和 TA 这 2 种模板攻击成功率达到 100%所需的能迹数都比 HOCPA 所需的能迹数少，证明了盲掩码模板攻击的有效性，其中 TA 需要的能迹数比 TA_HW 少，说明以 $\mathrm{comb}=u\oplus v$ 为目标是更好的选择。因此在以后的实验中，均以 $\mathrm{comb}=u\oplus v$ 为攻击目标。

4.5　基于多元高斯分布的盲掩码模板攻击

为了提高模板攻击的成功率，应该尽可能利用能迹上的全部泄露信息。根据 4.3 节所述，二阶 CPA 攻击共得到 193 个泄露位置。这些位置都不同程度地泄露了 Sbox0 输入或输出的信息。要充分利用这些信息泄露，就需要采用多元高斯分布。实验比较了不同能耗特征数量与攻击成功率的关系，采用 50 条攻击能迹，实验结果如图 4 所示。图 4 中横坐标中的 1*表示一元高斯分布的盲掩码模板攻击，作为比较基准。

从图 4 可以看出，采用 2 个能耗特征的二元高斯盲掩码模板攻击与基准攻击成功率相同。当能耗特征达到 5 个时，成功率略有上升，但继续增加能耗特征数量，成功率反而出现大幅下降。产生该现象的一个原因是，单一的能耗特征的分布并不符合高斯分布。图 5 是 $u\oplus v$=4 的分组中，一个泄露位置的能耗分布的直方图与高斯分布的比较。从图 5 可以看出，单个位置的能耗分布明显与高斯分布不同。而在多元高斯分布中，各边际概率都应该符合高斯分布，因此多元高斯分布不能准确地反映能耗的实际分布情况，这是造成能耗特征增加、攻击效果不佳的原因之一。

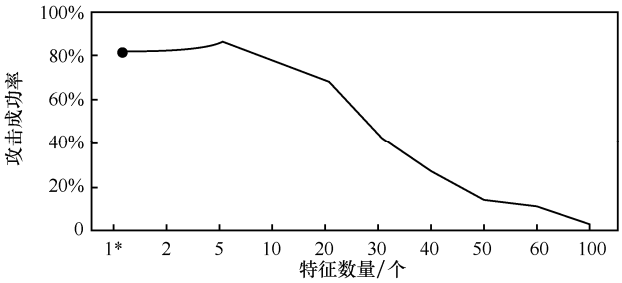

图 4　多元高斯分布盲掩码模板攻击中特征数与成功率的关系

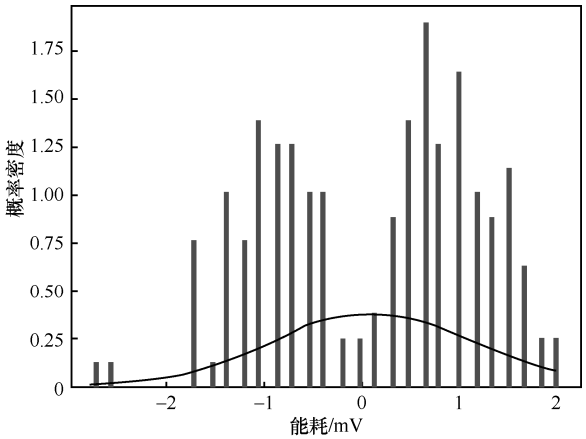

图 5　单个泄露位置能耗的分布情况

此外，还存在另一个造成能耗特征增加、攻击成功率下降的因素，即由于特征数过多而造成病态协方差矩阵，这个因素在本实验中表现得特别明显。为了尽可能地丢弃泄露信息，本文在选择特征时并未遵循文献[43]中提出的时间间隔原则。因此，有很多特征的样本位置是相邻的，这些特征之间相关性很高，更容易形成病态协方差矩阵。

对此，一个解决方案是采用 PCA 对能迹的特征向量进行预处理。PCA 在模板攻击中主要用于降维处理[42]，这里采用 PCA 的主要目的是防止产生病态协方差矩阵。图 6 是采用 PCA 预处理后的多元高斯分布的盲掩码模板攻击，其中，PCA0.4、

PCA0.6、PCA0.8 表示对能耗特征进行 PCA 处理，分别保留前 40%、60%、80%方差的 PCA 维作为特征向量。ref(*n* pois)表示不使用 PCA 预处理时，采用 *n* 个特征的攻击结果。从图 6 可以看出，全部能耗特征经 PCA 预处理后的攻击成功率均超过未经 PCA 预处理且采用 100 个特征的攻击成功率（ref(100 pois)），这证明了 PCA 防止产生病态协方差矩阵的能力。然而，所有采用 PCA 预处理的攻击成功率均未能超过未经 PAC 预处理且仅使用 5 个特征的攻击成功率（ref(5 pois)）。有以下 2 个原因：1) PCA 在降维的同时会造成信息丢失；2) PCA 维是各元素能耗特征的加权和，其概率分布将会产生很大变异。由于本实验中各能耗特征本身就不符合高斯分布，因此 PCA 维的高斯性更差。

由上述分析可知，对加掩设备的多元高斯分布的盲掩码模板攻击中，不宜采用 PCA 处理，而只能选择有限的特征数量。这一限制造成了很大的信息损失。

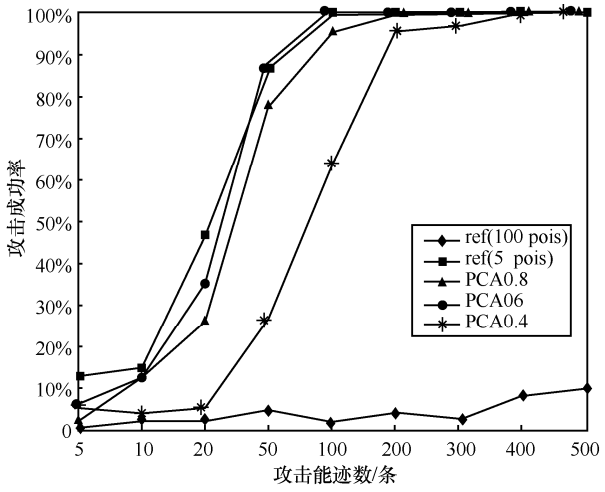

图 6 特征 PCA 处理的多元高斯模板攻击

4.6 基于神经网络的盲掩码模板攻击

4.6.1 神经网络结构与训练

在神经网络的结构选择上，已有实验证明，单个隐层的前向神经网络最符合模板攻击的需要。这是由于特征能耗中不存在像图像识别等拥有的高阶抽象特征，因而并不需要采用深度神经网络来提取高阶特征。神经网络在本研究中的作用是自动学习能耗的组合方式，并计算组合能耗相对于中间组合值的概率。在神经元的激活函数的选择方面，目前深度学习采用的 RELU（rectified linear unit）、SELU（sealed exponential linear unit）等激活函数也并不适用。这是由于 RELU 的非线性转换能力不如 tanh 函数或 sigmoid 函数。RELU 主要应用在多层卷积网络中，其主要作用是防止梯度消失，而梯度消失的问题在单个隐层的前向神经网络中并不严重。经反复实验，本文采用的网络结构为[输入层=特征数，隐层=100 个神经元，输出层=分类数]，输入层的能耗特征进行了正则化处理（即将输入能耗特征数据映射到平均能耗为 0，标准差为 1 的范围内）；隐层采用 tanh 函数；输出层每个神经元代表一个类别的得分，采用 softmax 转换为各类的概率。需要特别说明的是，由于 AES Sbox 转换的特点，无掩中间组合值 $u \oplus v$ 的取值范围并不是 0~255，而只是其中 163 个值，因此，输出层的神经元数量为 163。训练和预测前，需要将实际的 $u \oplus v$ 值映射到连续的[0,162]之间的值域。实验中使用的神经网络结构如图 7 所示。实验中神经网络在 tensorflow 中实现，使用一块 NVIDIA GTX980Ti 显卡（6 GB 内存）进行训练。

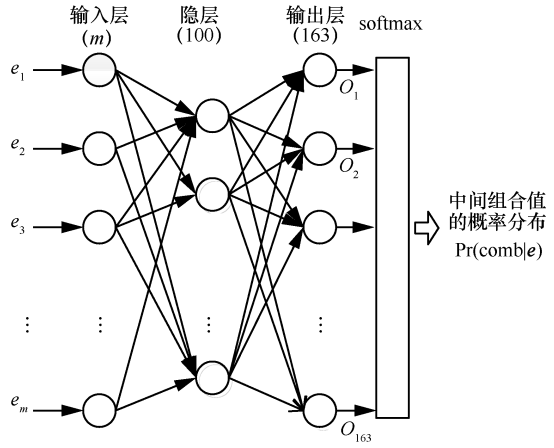

图 7 神经网络结构示意

神经网络采用交叉熵损失函数（cross-entropy），采用 Adam 优化器训练。经反复测试，学习率采用 0.01。每次训练迭代 1 万次，每次迭代都输入训练集的全部能耗特征向量，计算训练集的损失和精度，并由 Adam 优化器调整网络权重。每迭代 10 次输入验证集能耗特征向量，计算验证集的损失和精度，并保存验证集精度最高时的网络参数，作为后期攻击时的模型。

每次迭代采用训练集的全部能耗特征向量（以下简称为“全批训练”），这与目前神经网络训练通常建议采用 mini-batch 的方法不同。为了防止过拟合，也因为学习数据量可能太大，不能一次性加载到内存，训练神经网络时往往采用 mini-batch 的方

式学习，即每次学习迭代时仅采用部分数据。但实验证明这种策略不适用对无掩中间组合值的能耗特征模型的学习。其原因是掩码产生了大量数字噪声，实验中能耗的信噪比太低，采用 mini-batch 难以发现梯度调整的正确方向。使用 mini-batch 的实验中，学习损失和精度一直呈现大幅振荡，难以收敛，而且 mini-batch 越小，振荡越大，收敛越困难。只有当每次迭代输入全部能耗特征数据时，模型的收敛最好。在本文的实验中，由于显卡内存足够大，能耗特征值采用 32 bit 浮点数表示，全部数据可以一次性加载到 GPU 中运算。

4.6.2　初步的神经网络盲掩码模板攻击

首先测试能耗特征数量与神经网络盲掩码模板攻击成功率的关系，其结果如图 8 所示。

图 8 中，NN20 表示采用 20 条攻击能迹时基于神经网络的盲掩码攻击的成功率，TA20 表示采用 20 条攻击能迹时基于多元高斯分布的盲掩码模板攻击的成功率，在这里作为对比的基准。从图 8 中可以看出，当特征数小于或等于 5 时，多元高斯模板的成功率高于神经网络；但特征数超过 5 个之后，神经网络的成功率迅速上升，而高斯模板的则开始下降。这是因为神经网络不存在病态协方差的问题，而其强大的非线性转换能力，使其可以有效利用更多能耗特征中包含的信息。当特征数达到 40 个时，20 条能迹的神经网络攻击成功率达到 100%；但当特征数超过 60 个之后，神经网络的攻击成功率开始急速下跌。这是因为在神经网络的训练中产生了严重的过拟合。神经网络的学习曲线如图 9 所示，可以清楚地看到过拟合现象的发生。

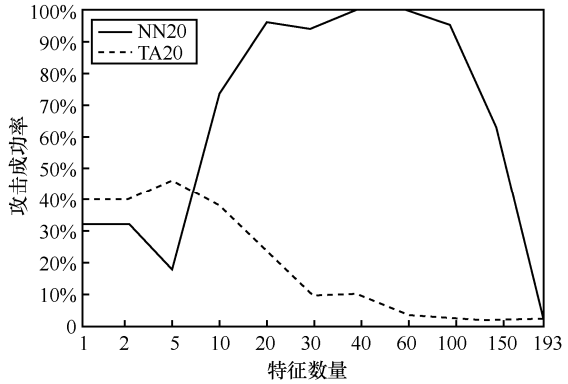

图 8　神经网络盲掩码模板攻击中特征数量与成功率的关系

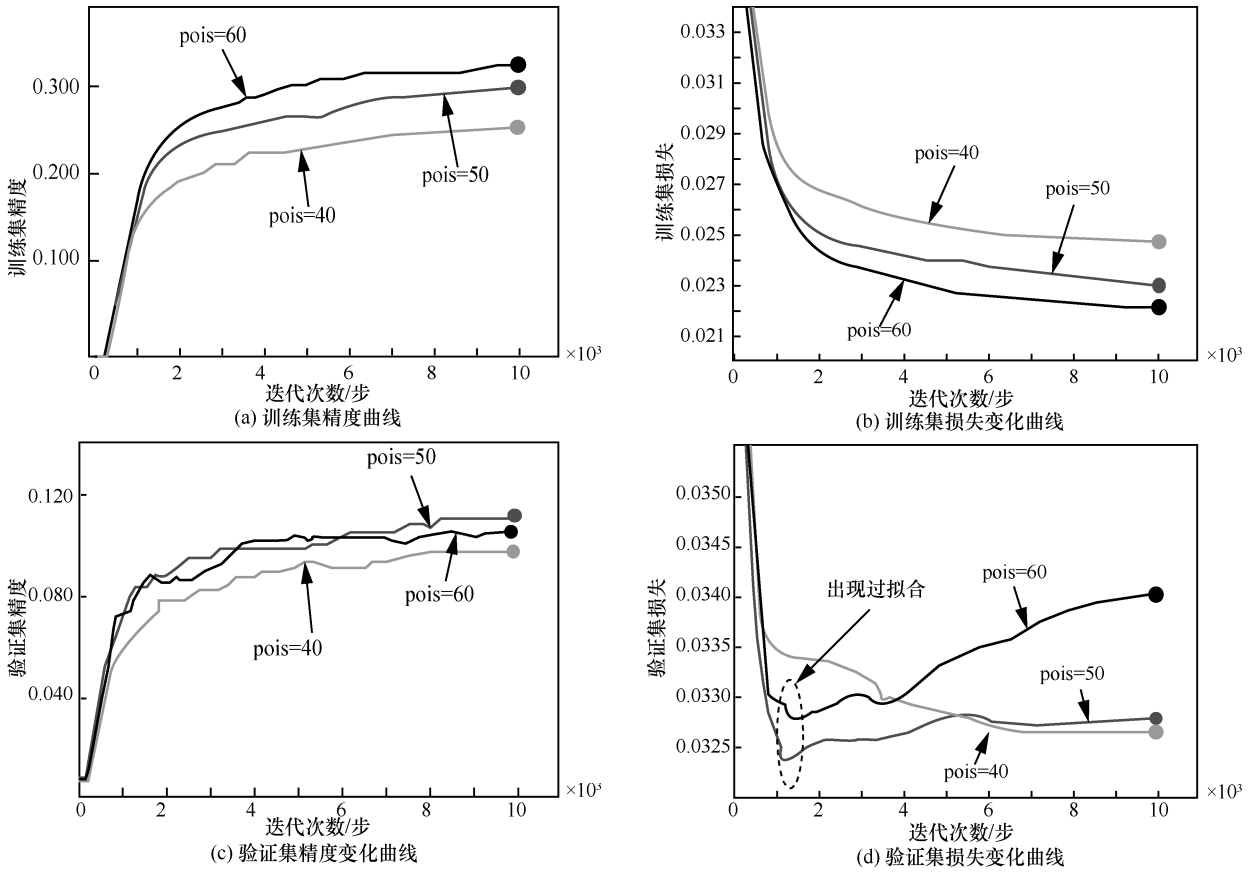

图 9　神经网络训练过程中的过拟合与特征数量的关系

在图 9 中，当仅使用 40 个能耗特征时（pois=40），验证集损失持续下降，没有出现过拟合。当使用 50 个能耗特征时（pois=50），验证集损失在 1 200 步之后开始上升，出现轻微的过拟合，但验证精度超过 pois=40 的精度。当使用 60 个能耗特征时（pois=60），过拟合变得比较严重，验证集精度也低于 pois=50 的精度。由于比例关系，图 9 中没有表示出使用更多特征时的训练情况。事实是，当特征更多时，过拟合更严重，精度下降更大。当使用全部 193 个特征时，验证集精度下降到只有 0.018 2。

此外，从图 9 可以看出，训练集的最小损失（即出现过拟合时的损失）随能耗特征的增加而减少，验证集中 pois=40 和 pois=50 也符合这一规律。这是因为随着特征数量的增加，网络能够更好地拟合到目标值。然而验证集中当 pois=60 时出现例外，其最小损失大于 pois=40 或 pois=50 的最小损失值。这一现象说明，导致过拟合的原因是增加能耗特征数量后，数据中的噪声大量增加，训练中利用了训练集中的噪声特征拟合目标值，导致模型泛化能力大幅下降，从而出现过拟合。

4.6.3 神经网络训练过拟合的处理

过拟合是机器学习中经常出现的现象。常见的防止过拟合的方法有 dropout、采用更简单的模型、使用更多的学习数据、权重正则化、在输入或权重或网络响应上增加噪声等。在对能耗数据的训练中，dropout 并不适用。根据 4.6.1 节的实验，能耗特征训练时必须使用全批训练才能收敛。dropout 丢弃部分数据，其效果等效于使用 mini-batch，会导致学习不收敛。使用更多的数据以防止过拟合的本质是增加训练数据的覆盖范围，避免训练时仅能访问部分特征。而能耗数据的训练中出现过拟合的原因是训练数据的噪声过大，因此使用更多的能迹并不能防止此类过拟合。最后，增加噪声的方法在这里显然更不可取，因为能耗数据本身信噪比就已经非常低了。因此，本文在防止过拟合的实验中采用 L2 规范化。

1) L2 规范化防止过拟合的实验分析

L2 规范化是在原损失函数（本文中为交叉熵损失函数）的基础上，增加连接权重的 L2 惩罚因子，从而防止权重的方差过大。一般而言，更小的权重方差意味着更简单的拟合模型，而简单的拟合模型不易出现过拟合现象。带有 L2 规范的训练损失函数形式为

$$\text{loss} = \text{crossentropy} + \lambda \frac{1}{N} \sum_{w} w^2 \tag{8}$$

本文使用全部 193 个能耗特征，分别采用 $\lambda = 0.01$、0.02、0.05、0.1、0.2、0.3 进行了防止过拟合的实验。图 10 显示了各种 λ 取值下神经网络的训练精度（用验证集精度表示）的对比情况。

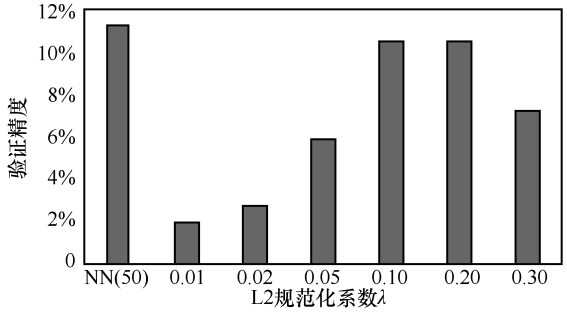

图 10 L2 规范化神经网络的训练结果对比

图 10 中 NN(50)表示采用 50 个特征的未经 L2 规范化的神经网络训练。NN(50)是 4.6.2 节特征数量与神经网络攻击实验中训练精度最高、攻击成功率最好的，在这里作为对比的基准。L2 规范化训练中，$\lambda = 0.01$、0.02、0.05 时网络训练均出现大幅过拟合，即未能阻止过拟合，在图 10 中对应的精度也很差；$\lambda = 0.1$、0.2、0.3 时网络训练均未出现过拟合，但网络训练精度仍然不如 NN(50)。这是因为在网络训练中通过 L2 规范化在减小权重变化方差的同时，也弱化了神经网络的拟合能力。由此看出，单纯采用 L2 规范化虽然可以防止过拟合，但并不能提高网络训练精度。

2) 能耗特征 PCA 预处理防止过拟合的实验分析

经分析认为，本实验中出现过拟合的原因在于大量的能耗特征中信息泄露并不明显。如 4.3 节所述，实验中采用的特征位置来源于二阶 CPA 攻击成功的样本位置对。这些样本位置对按相关系数降序排序，越靠后的特征，其包含的泄露信息越少，与泄露信息无关的噪声特征越多。在训练过程中，当采用更多能耗特征时，训练数据中包含了更多的噪声更大、信息泄露更不明显的能耗特征。这些无关的特征被用于训练集的拟合，虽然提高了训练集的拟合精度，却弱化了模型的泛化能力，从而导致过拟合。

对此，可以对数据进行 PCA 预处理，通过提取能耗特征中的主成分，消除部分噪声，从而达到防止过拟合的目的。但 PCA 处理不可避免会同时带来信

息的丢失，为了避免出现严重的信息丢失，本文提出一种分段 PCA 处理的方法。分段 PCA 是指根据泄露位置对应的 Pearson 相关系数（参见 4.3 节），将能耗特征分为高泄露和低泄露 2 个部分，分别采用不同的降维幅度进行 PCA 处理。高泄露部分降维幅度小，保留大部分特征；低泄露部分降维幅度大，以过滤更多的噪声。本文对采用全部特征进行 PCA 预处理（整体 PCA 预处理）和分段 PCA 预处理 2 种方案进行了实验对比，实验结果如图 11 所示。

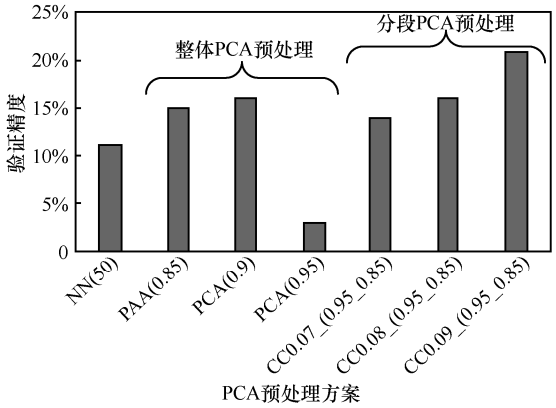

图 11　PCA 预处理的神经网络训练结果对比

图 11 中，NN(50)表示采用 50 个特征的无 PCA 预处理的神经网络攻击结果，在这里作为对比的基准；PCA（x）表示对全部能耗特征进行 PCA 预处理，x 表示保留方差比例为 x 的主成分作为特征向量；CCy_(a_b)表示对能耗特征进行分段 PCA 预处理，y 表示划分高、低泄露特征的相关系数（以下称为阈值相关系数）；分别采用了 0.07、0.08 和 0.09 3 种相关系数。a 和 b 分别表示高、低泄露特征的 PCA 处理中的保留方差比例。

从图 11 中可以看出，整体 PCA 处理中，除保留方差比例为 0.95 外，训练精度均超过基准。保留方差比例为 0.9 时，训练出现轻微过拟合。保留方差比例为 0.95 的训练精度与采用全部特征而不做 PCA 相近，仍然出现了大幅过拟合，原因是其中包含的噪声仍然过多。分段 PCA 预处理中，阈值相关系数为 0.09 时，训练达到了最高验证精度 20.79%。阈值相关系数为 0.07 和 0.08 时，训练均出现了不同程度的较弱的过拟合。

3) 结合 PCA 预处理和 L2 规范化防止过拟合实验

鉴于 PCA 处理中仍然存在一定的过拟合，因此使用 L2 规范化进一步优化网络训练。以下实验对图 11 中的几种 PCA 预处理方案均采用 L2 规范化重新训练网络，其中，L2 规范化的系数 λ 根据几种方案中出现的过拟合程度来取值，过拟合严重的取 0.1，比较严重的取 0.01，非常微弱的取 0.001，然后根据训练情况调整 λ。图 12 显示了各 PCA 预处理方案的最佳 λ 的 L2 规范化训练精度与未采用 L2 规范化的训练精度对比。

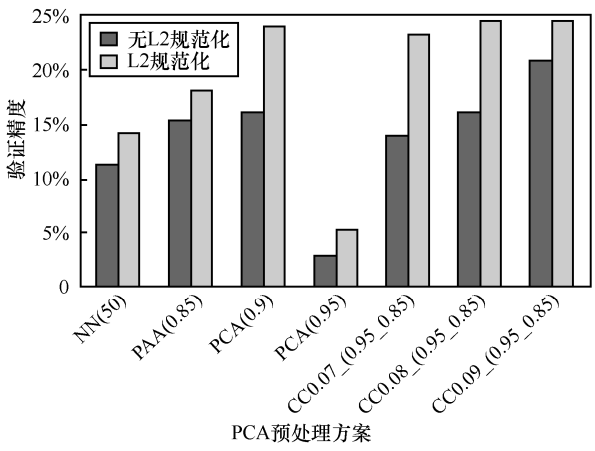

图 12　结合 PCA 预处理和 L2 规范化防过拟合的神经网络训练

从图 12 的实验结果可以看出，选择恰当的 λ 时，L2 规范化训练均可以提高训练精度。且过拟合越严重，训练精度提高越大。即使对于没有出现明显过拟合的情况，如 CC0.09_(0.95_0.85)，L2 规范化训练仍然能够提高精度。

4.6.4　基于神经网络的盲掩码攻击实验

本实验采用最佳神经网络训练结果作为模板，进行了盲掩码模板攻击，其结果与使用已知掩码的模板攻击的比对如图 13 所示。已知掩码的模板攻击是指在训练阶段和攻击阶段，能迹使用的掩码均是已知的。这对加掩防护的加密设备不是一种现实可行的攻击方法，但它模拟了对无掩码防护的加密设备的模板攻击，由于掩码在此并未起到防护作用，因此其攻击成功率应当相当高。本文选择这种攻击模式是为了验证基于神经网络的盲掩码攻击的真实效果。图 13 中，TA_WITH_MASK 表示使用已知掩码的模板攻击，NN_NOMASK 表示基于神经网络的盲掩码模板攻击。

从图 13 可以看出，正确训练的神经网络盲掩码攻击的成功率已接近对无掩码防护的模板攻击成功率。4 条攻击能迹时，攻击成功率已达到 90%以上；8 条攻击能迹已达到 100%的成功率；3 条能迹的猜测熵[44]（多次攻击中正确密钥的平均排名，

最小为 1）为 2.12，已在实际可攻击的范围内；最少一条能迹就能攻击成功，成功率为 18.4%。

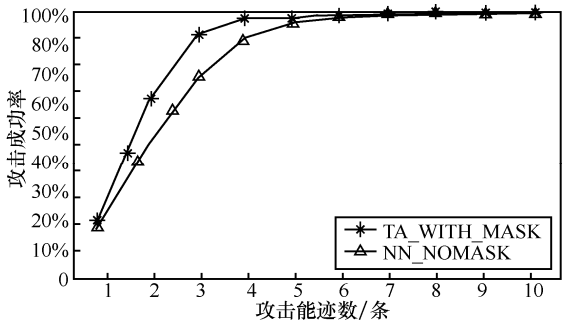

图 13 神经网络盲掩码攻击的成功率与攻击能迹数的关系

5 结束语

相比现有的针对加掩防护加密算法的模板攻击，盲掩码模板攻击可以有效地降低对加掩防护加密算法的模板攻击的条件和难度。现有的技术均要求攻击者能够了解设备在每次加密时使用的随机掩码。这意味着攻击者必须能够完全控制训练设备，可以修改设备上加密算法的实现代码，从而在每次加密时输出使用的随机掩码，而且，为了保证模板的对真实设备攻击时的有效性，修改代码时还要尽可能避免或者减小对能耗的影响。但由于训练设备与被攻击设备的加密实现不可能完全相同，其模板在攻击时往往无效或攻击效率低下。因此现有对加掩防护加密算法的模板攻击技术实施的条件高、难度大、攻击效果不稳定。而且当加密算法采用硬件实现时，由于无法获得训练设备的掩码，因而无法使用现有的技术实施模板攻击。而本文提出的盲掩码模板攻击可以在攻击者完全不了解掩码的情况下，仅使用一台已知密钥的设备作为训练设备，训练或统计出模板，以此实施模板攻击。这一方法显著降低了训练和攻击所需的条件。同时，由于训练设备与被攻击设备的实现完全相同，可以有效保证该方法得到的模板在攻击时的有效性。

本文的实验分析证明了盲掩码模板攻击具有较高的攻击效率。与同样不需要了解掩码的二阶 CPA 攻击相比，盲掩码模板攻击需要的能迹数小 2 个数量级，达到 100%攻击成功率时，二阶 CPA 需要 1 500 条攻击能迹，基于神经网络的盲掩码攻击仅需 8 条能迹，而所需的代价很小，仅需要一台已知密钥的设备。

此外，实验结果表明，基于神经网络的盲掩码攻击比基于统计的一元或多元高斯模型的盲掩码攻击更为有效，前者达到 100%攻击成功率所需的能迹数为 8 条，而后者需要 100 条。但对低信噪比数据的神经网络训练极易出现无法收敛或过拟合，需要采用一系列措施保证神经网络得到有效训练。实验表明，对于信噪比很低的能耗数据，神经网络训练时必须采用全批数据训练才能收敛。本研究认为，这一结论并不局限于对能耗数据的神经网络训练，而是适用于各类极低信噪比数据的神经网络训练。此外，为防止神经网络训练出现过拟合，本文提出了对能耗特征进行分段 PCA 预处理和 L2 规范化的训练方案。该方案能够最大限度地利用低信噪比数据中非常有限的信息特征，同时防止噪声产生的过拟合。

需要说明的是，盲掩码模板攻击，特别是基于神经网络的盲掩码模板攻击，其训练阶段需要较长时间。因为盲掩码模板攻击实施的前提是可以对加掩防护的加密设备成功实施高阶 CPA 攻击。高阶 CPA 攻击在这里起到 2 个作用，一是验证盲掩码攻击所攻击的目标中间组合值是否存在信息泄露，二是提供了盲掩码模板攻击所需要的信息泄露位置。高阶 CPA 非常耗时，此外，神经网络的超参数，如网络隐层大小、学习率，PCA 处理中的保留方差比例、L2 规范化因子 λ 等，也需要通过多次实验才能找到最佳的设置，这些都导致模板学习需要较长的时间。但现有的对加掩防护的模板攻击技术也都需要较长的准备和训练时间，因为训练前需要修改训练设备的代码以便输出加密时使用的随机掩码，而且还要防止修改的代码对加密过程的能耗造成过大的影响，这一准备工作需要耗费大量时间。同时，现有技术中最佳的模板模型也是基于神经网络的，其训练同样需要较长时间。虽然由于其数据不含掩码产生的数字噪声，能耗数据信噪比较高，训练相对容易，但需要创建更多的模板（需要分别对掩码和中间值创建模板），这意味着需要较长的训练时间。由于修改设备代码的时间是不确定的，因此无法对本文的盲掩码模板攻击与现有技术的训练时间进行定量比较，但定性上看，盲掩码模板攻击相对于已知掩码的模板攻击，其效率至少是可比的，如果不是更高的话。在盲掩码模板攻击中，由于密钥已知因而不需要猜测，二阶 CPA 可以在几十分钟到数小时内完成，而已知掩码模版攻击则需要花费数小时甚至数天来改写设备代码。

综上所述，盲掩码模板攻击可以有效地降低对加掩防护加密算法的模板攻击的条件和难度，其中基于神经网络网的盲掩码模板攻击具有很高的攻击成功率。相对于现有技术，本文方案付出的训练时间上的代价并不显著。

参考文献：

[1] 杜之波, 吴震, 王敏, 等. 针对 SM4 轮输出的改进型选择明文功耗分析攻击[J]. 通信学报,2015,36(10):85-91.
DU Z B, WU Z, WANG M, et al. Improved chosen-plaintext power analysis attack against SM4 at the round-output[J]. Journal on Communications, 2015, 36(10):85-91.

[2] 吴震, 王敏, 饶金涛, 等. 针对基于SM3的HMAC的能量分析攻击方法[J]. 通信学报,2016,37(5):38-43.
WU Z, WANG M, RAO J T, et al. Mutual information power analysis attack of HMAC based on SM3[J]. Journal on Communications, 2016, 37(5):38-43.

[3] 杜之波, 吴震, 王敏, 等. 基于 SM3 的动态令牌的能量分析攻击方法[J]. 通信学报,2017,38(3):65-72.
DU Z B, WU Z, WANG M, et al. Power analysis attack of dynamic password token based on SM3[J]. Journal on Communications, 2017, 38(3):65-72.

[4] 王敏, 吴震, 饶金涛, 等. 针对密码芯片频域互信息能量分析攻击[J]. 通信学报,2015,36(s1):131-135.
WANG M, WU Z, RAO J T, et al. Mutual information power analysis attack in the frequency domain of the crypto chip[J]. Journal on Communications, 2015, 36(s1):131-135.

[5] KOCHER P C.Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems[C]//Annual International Cryptology Conference. 1996: 104-113.

[6] KOCHER P. Differential power analysis and related attacks[C]// Annual International Cryptology Conference. 1999: 388-397.

[7] MANGARD S, OSWALD E, POPP T. Power analysis attacks: revealing the secrets of smart cards[M]. Springer Science & Business Media.2008.

[8] BATINA L , GIERLICHS B , LEMKE-RUST K . Differential cluster analysis[C]// International Workshop on Cryptographic Hardware & Embedded Systems. 2009.

[9] BRIER E , CLAVIER C , OLIVIER F . Correlation power analysis with a leakage model[C]// Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge. 2004.

[10] GIERLICHS B , BATINA L , TUYLS P , et al. Mutual Information Analysis[C]// Proceeding Sof the International Workshop on Cryptographic Hardware & Embedded Systems. 2008.

[11] CHARI S , RAO J R , ROHATGI P . Template attacks[M]// Cryptographic Hardware and Embedded Systems - CHES 2002. Springer Berlin Heidelberg, 2002.

[12] SCHINDLER W, LEMKE K, PAAR C. A stochastic model for differential side channel cryptanalysis[M]// Cryptographic Hardware and Embedded Systems–CHES 2005, 2005：30-46.

[13] 刘飚, 孙莹. 基于公共协方差矩阵的实用模板攻击[J]. 计算机应用研究,2016(1):236-239.
LIU B, SUN Y. Practical template attacks based on pooled covariance matrix[J]. Application Research of Computers, 2016(1):236-239.

[14] 崔琦, 王思翔, 段晓毅, 等. 一种AES算法的快速模板攻击方法[J]. 计算机应用研究, 2017, 34(6):1801-1804.
CUI Q, WANG S X, DUAN X Y, et al. Fast tempolate DPA attack against AES algorithm[J]. Application Research of Computers, 2017, 34(6):1801-1804.

[15] CHOUDARY O, KUHN M G. Efficient Template Attacks[M]//Smart Card Research and Advanced Application Conference-CARDIS. Springer, 2013: 253-270.

[16] 杜之波, 孙元华, 王燚. 针对AES密码算法的多点联合能量分析攻击[J]. 通信学报,2016(s1):78-84.
DU Z B, SUN Y H, WANG Y. Multi-point joint power analysis attack against AES[J]. Journal on Communications, 2016(s1):78-84

[17] 王小娟, 郭世泽, 赵新杰, 等. 基于功耗预处理优化的 LED 密码模板攻击研究[J]. 通信学报,2014(3):157-167.
WANG X J, GUO S Z, ZHAO X J, et al. Research of power preprocessing optimization-based template attack on LED[J]. Journal on Communications, 2014(3): 157-167.

[18] ARCHAMBEAU C, PEETERS E, STANDAERT F X, et al. Template attacks in principal subspaces[M]// Cryptographic Hardware and Embedded Systems-CHES 2006. Springer, 2006：1-14.

[19] 王红胜, 徐子言, 张阳, 等. 基于模板的光辐射分析攻击[J]. 计算机应用研究,2017,34(7):2151-2154.
WANG H S, XU Z Y, ZHANG Y, et al. Template based phtonic emission attacks[J]. Application Research of Computers, 2017, 34(7): 2151-2154.

[20] PICEK S, HEUSER A, GUILLEY S. Template attack versus Bayes classifier[J]. Journal of Cryptographic Engineering,2017,7(2):1-9.

[21] BARTKEWITZ T，LEMKE-RUST K. Efficient template attacks based on probabilistic multi-class support vector machines[M].Springer, 2013.

[22] HEUSER A, ZOHNER M. Intelligent machine homicide[M]// Constructive Side-Channel Analysis and Secure Design. Springer, 2012: 249-264.

[23] MARTINASEK Z, ZEMAN V. Innovative method of the power analysis[J]. Radioengineering,2013,22(2):586-594.

[24] MARTINASEK Z, HAJNY J, MALINA L. Optimization of power analysis using neural network[C]// International Conference on Smart Card Research and Advanced Applications, 2013: 94-107.

[25] SCHRAMM K,PAAR C. Higher order masking of the AES[M]// Topics in cryptology–CT-RSA 2006. Springer, 2006:208-225.

[26] MESSERGES T. Using second-order power analysis to attack DPA resistant software[C]// Cryptographic Hardware and Embedded Systems—CHES 2000, 2000:27-78.

[27] JOYE M, PAILLIER P, SCHOENMAKERS B. On second-order differential power analysis[M]// Cryptographic Hardware and Embedded Systems–CHES 2005. Springer, 2005:293-308.

[28] BELGARRIC P, BHASIN S, BRUNEAU N, et al. Time-frequency analysis for second-order attacks[M]//Smart Card Research and Advanced Applications. Springer, 2014:108-122.

[29] OSWALD E, MANGARD S. Template attacks on masking—resistance is futile[M]. Topics in Cryptology–CT-RSA 2007. Springer, 2006: 243-256.

[30] LEMKE-RUST K, PAAR C. Gaussian mixture models for higher-order side channel analysis[C]//Cryptographic Hardware and Em-

bedded Systems–CHES 2007.2007:14-27.

[31] LERMAN L, BONTEMPI G, MARKOWITCH O. A machine learning approach against a masked AES[J]. Journal of Cryptographic Engineering,2015,5(2):123-139.

[32] GILMORE R. Neural network based attack on a masked implementation of AES[J]. Hardware Oriented Security and Trust. 2015(6):5.

[33] CORON J S, PROUFF E, RIVAIN M. Side channel cryptanalysis of a higher order masking scheme[M] Springer, 2007.

[34] HOSPODAR G, MULDER E, GIERLICHS B, et al. Least squares support vector machines for side-channel analysis[J]. Center for Advanced Security Research Darmstadt,2011:99-104.

[35] LERMAN L, BONTEMPI G, MARKOWITCH O. Side channel attack: an approach based on machine learning[J]. Center for Advanced Security Research Darmstadt, 2011: 29-41.

[36] LERMAN L, POUSSIER R, BONTEMPI G, et al. Template attacks vs. machine learning revisited[C]//Constructive Side Channel Analysis and Secure Design COSADE 2015.2015: 20-33.

[37] NASSAR M, SOUISSI Y, GUILLEY S, et al. RSM: A small and fast countermeasure for AES, secure against 1st and 2nd-order zero-offset SCAs[C]//Design, Automation & Test in Europe Conference & Exhibition.2012:1173-1178.

[38] PROUFF E, RIVAIN M, BEVAN R. Statistical analysis of second order differential power analysis[J]. IEEE Transactions on computers,2009,58(6):799-811.

[39] DREXLER H B R M, PULKUS J. Improved template attacks[C]//The Constructive Side-Channel Analysis and Secure Design-First International Workshop.2010:4-5.

[40] BHASIN S, DANGER J L, GUILLEY S, et al. NICV：normalized inter-class variance for detection of side-channel leakage[C]// Electromagnetic Compatibility.2014:310-313.

[41] STANDAERT F X, ARCHAMBEAU C. Using subspace-based template attacks to compare and combine power and electromagnetic information leakages[M]//Cryptographic Hardware and Embedded Systems–CHES 2008. Springer, 2008:411-425.

[42] GIERLICHS B. Signal theoretial methods in differential side channel cryptanalysis[D]. Nordrhein-Westfalen: Ruhr-University Bochum, 2005-2006.

[43] STANDAERT F X, MALKIN T G, YUNG M. A unified framework for the analysis of side-channel key recovery attacks[M]//Advances in Cryptology-EUROCRYPT 2009. Springer, 2009: 443-461.

[作者简介]

王燚（1968− ），男，四川成都人，博士，成都信息工程大学教授，主要研究方向为机器学习、侧信道攻击与防御、自然语言处理。

吴震（1975− ），男，江苏苏州人，成都信息工程大学副教授，主要研究方向为信息安全、密码学、侧信道攻击与防御、信息安全设备设计与检测。

蔺冰（1973− ），男，四川成都人，成都信息工程大学讲师，主要研究方向为信息安全、侧信道攻击与防御、计算机网络。