基于双样本 KS 检验的非特定 TVLA 方法

郑震，严迎建，蔡爵嵩，刘燕江
（信息工程大学密码工程学院，河南 郑州 450001）

摘 要：测试向量泄露评估（TVLA）在能耗样本量较少时易出现“假阴性”错误。针对该问题进行了理论推导，发现对非特定 TVLA，能量迹中存在泄露的采样点处得到的检验统计量 *t* 值随能耗样本量变化而变化，非泄露点处的 *t* 值则无显著变化，因此当存在泄露时不同能耗样本量下得到的 *t* 值序列的分布不同。据此提出在不同样本量下实施非特定 TVLA 并对得到的 *t* 值序列实施双样本 KS 检验以评估泄露。分别在无防护对齐仿真能耗数据、加防护对齐能耗数据集 DPA Contest v4_2 和加防护非对齐自测能耗数据上进行了验证，结果表明在对齐的仿真能耗数据和 DPA Contest v4_2 数据集上所提方法检测出泄露所需样本量较其他方法均有所减小，最多分别减小了 46.1%和 39.0%；在非对齐的自测能耗数据进行对齐处理后，所提方法所需能耗样本量较其他方法同样有所减小，最多减小了 29.4%。因此所提方法能够有效减小能耗样本量较小时出现“假阴性”错误的概率。

关键词：侧信道；测试向量泄露评估；假阴性；双样本 KS 检验
中图分类号：TP309
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2023101

Non-specific TVLA method based on two-sample KS test

ZHENG Zhen, YAN Yingjian, CAI Juesong, LIU Yanjiang
Department of Cryptogram Engineering, Information Engineering University, Zhengzhou 450001, China

Abstract: Test vector leakage assessment (TVLA) is prone to “false negative” when the power consumption sample size is small. To address this issue, it was found that for non-specific TVLA, when the power consumption sample size changes, the test statistic t-values obtained at the leakage sampling points in the power trace vary accordingly, while the t-values at the non-leakage sampling points do not significantly vary. Therefore, when there is leakage, the distributions of the t-values obtained under different sample sizes will be different. Based on this, it was proposed to implement non-specific TVLA under different sample sizes and perform two-sample KS test on the obtained t-value sequences to evaluate whether there was leakage. Verifications were carried out based on unprotected-aligned simulation power consumption, protected-aligned power consumption dataset DPA Contest v4_2 and protected-non-aligned self-collected power consumption respectively. The results showed that the sample size required by the proposed method on the aligned simulation power consumption and DPA Contest v4_2 was reduced by at most 46.1% and 39.0% respectively. And after the alignment, the required sample size of the proposed method on the self-collected power consumption is also smaller than that of other schemes, with a maximum reduction of 29.4%. Therefore, the proposed method can effectively reduce the probability of “false negative” when the power consumption sample size is small.

Keywords: side-channel, test vector leakage assessment, false negative, two-sample KS test

收稿日期：2023−02−10；**修回日期**：2023−04−21

0 引言

自 1999 年 Kocher 等[1]开创性地提出差分能量分析攻击以来，大量侧信道能量攻击方法相继被提出[2-4]，对各类密码实现的安全性构成了严重威胁。当前，对侧信道能量信息安全性的评估已经成为密码设备安全性测评中不可或缺的重要环节。最初，评估能量信息安全性是通过实施能量攻击的方法实现的，如果能够攻破说明存在泄露。然而，随着新的能量攻击方法不断被提出，这种评估形式的成本变得越来越大且准确性难以得到保证。因此，不依赖于能量攻击方法的通用型能量信息泄露评估方法逐渐成为更普遍的选择。不同于实施攻击的评估形式，通用型评估方法旨在判断是否存在能量信息泄露，而不关注怎样利用这些泄露破解密钥等敏感信息，其基于成熟的数理统计理论，通过执行确定性的评估步骤来判断明文或密钥等秘密信息是否和能量消耗存在相关性，进而判断泄露情况。

目前，最常用的通用型能量信息泄露评估方法是测试向量泄露评估（TVLA, test vector leakage assessment）[5]技术，该技术将采集的能耗分为两组，并根据两组能耗均值间的差异来判断是否存在泄露，若有显著的统计差异则认为存在泄露。根据分组依据的不同，TVLA 可分为特定 TVLA 和非特定 TVLA，其中特定 TVLA 根据密码算法中间值的一位或多位的取值进行分组，非特定 TVLA 则根据输入明文对能耗数据进行分组，其中一组为固定明文，另一组为服从均匀分布的随机明文。由于可选择的密码算法的中间值数目庞大，特定 TVLA 所需耗费的成本非常大，以 AES-128 算法为例，在考虑圈密钥加、字节置换、行移位和列混合这 4 种密码操作的情况下，仅在算法的一轮运算中就可实施 4×128 种位测试和 4×16×256 种字节测试，因此目前应用较多的是非特定 TVLA。

TVLA 将复杂的泄露检测问题简化为易行的统计步骤，并且不需要评估人员掌握过多的算法实现知识[6-7]，具有简单、高效和可操作性强等优势，近年来被广泛应用于信息泄露的检测中[8-10]。然而，受噪声和自身检验统计量设计等因素影响，TVLA 易出现“假阳性”（实际无泄露却判定存在泄露）和“假阴性”（实际有泄露却判定不存在泄露）的错误，相关领域的学者对此开展了大量研究。总体来说，当前对 TVLA 的研究主要可分为对 TVLA 进行创新和寻找 TVLA 的替代方法两类。文献[11]提出一种基于 HC（higher criticism）检验的 TVLA 方法，对 TVLA 得到的各采样点处的 p 值（由检验统计量 t 值得出的零假设成立，即不存在泄露的概率）实施 HC 检验，通过比较在无泄露情况下预期的 p 值分布和实际检验得到的 p 值的分布之间的差异对泄露情况进行判断。该方法能够检测出信号较弱的泄露，当能量迹中存在多处泄露信号时该方法较 TVLA 能以更小的能量迹样本量检测出泄露。文献[12]针对两组能耗样本的均值差异较小时 TVLA 存在漏检的问题，提出对能耗样本的均值与方差进行综合差异评估，当样本均值间的差异大于方差间的差异时实施多分类 F 检验，当样本均值间的差异小于方差间的差异时实施 Bartlett 检验。该方案能够有效解决两组能耗的均值差异不显著导致的“假阴性”错误。文献[13]提出一种配对 t 检验的方案，对密码算法相邻的两次加密进行配对，由于执行一次加密的时间非常短，可以认为相邻的两次加密是在相同的外界环境下进行的，因此在配对时做差即可减小环境噪声的影响，得到更加稳定的检验统计量，从而以更小的能耗样本量检测出泄露。文献[14]同样对配对 t 检验方案进行了研究。文献[15]通过显著性检验的方法对 TVLA 中的阈值设置问题进行了研究，减小了检验犯“假阳性”错误的概率。文献[16]提出用卡方检验结合 TVLA 对泄露进行检测，该方案利用卡方检验的性质将泄露检测扩展到了多个能耗分组上，且可以捕获多个统计矩中的泄露。当信息泄露在一阶统计矩（均值）上的差异性不大时该方案可以检测出 TVLA 方法漏报的泄露点。文献[17]提出一个基于置信区间的泄露评估框架，该方法对噪声的鲁棒性较强，能够有效避免评估中的“假阳性”错误。2021 年，文献[18]首次将深度学习应用于泄露评估中，使评估人员不必考虑泄露在能量迹中的位置、能量迹是否对齐和泄露的统计矩阶数等问题，且直接覆盖了多变量等泄露情形。文献[19]提出一种针对高阶掩码的基于 Levene 检验的泄露检测方法，首先对能耗数据的分布进行判定，若为正态分布则实施 TVLA，否则实施 Levene 判断多个能耗分组的方差是否相同，进而对泄露情况进行判定。该方法能够以较小的样本量检测出存在的高阶泄露。

本文针对能耗数据样本量较小时 TVLA 易出现“假阴性”错误的问题，推导发现在非特定 TVLA 中，能量迹中泄露点处的 t 值随能量迹样本量的变化

而变化，而非泄露点处的 t 值无显著变化。因此当存在泄露时，非特定 TVLA 在不同样本量下得到的 t 值的分布不同；当不存在泄露时，不同样本量下得到的 t 值的分布无显著变化。据此，本文提出在不同样本量下分别实施非特定 TVLA 并对得到的 t 值序列实施双样本 KS 检验以判断其是否服从相同分布，从而评估泄露情况。本文方法根据 t 值分布的变化情况而非 t 的具体取值对泄露情况进行判断，能够有效减小检测泄露所需的能耗数据样本量。

1　TVLA 概述及技术分析

在各类能量分析攻击中，攻击者试图根据能耗的数学统计特征设计不同的攻击形式来获取与所处理数据相关的信息进而破解密码算法。因此，当所处理的数据发生变化时，若密码设备能耗的统计特征也随之发生显著变化，则说明能耗中有攻击者可利用的信息，即存在能量信息泄露。TVLA 正是基于上述原理，利用 t 检验分析密码设备处理不同数据时产生的能耗均值是否发生变化，从而判断是否存在能量信息泄露，其零假设和备择假设分别为 H_0 表示不存在泄露和 H_1 表示存在泄露。其具体步骤如下。首先将采集到的能耗数据分为两组，分别记为 Ψ_0 和 Ψ_1，将其样本量、样本均值和样本方差分别计为 (n_0,μ_0,S_0^2) 和 (n_1,μ_1,S_1^2)，TVLA 的检验统计量 t 和自由度 v 分别为

$$t=\frac{\mu_0-\mu_1}{\sqrt{\frac{S_0^{\,2}}{n_0}+\frac{S_1^{\,2}}{n_1}}}$$

$$v=\frac{\left(\frac{S_0^{\,2}}{n_0}+\frac{S_1^{\,2}}{n_1}\right)^2}{\frac{\left(\frac{S_0^{\,2}}{n_0}\right)^2}{n_0-1}+\frac{\left(\frac{S_1^{\,2}}{n_1}\right)^2}{n_1-1}} \tag{1}$$

然后，得到对应的 t 分布的概率密度函数 $f(t,v)$ 和零假设成立的概率 p 分别为

$$f(t,v)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

$$p=2\int_{|t|}^{\infty}f(t,v)\mathrm{d}t \tag{2}$$

其中，$\Gamma(\cdot)$ 是伽马函数。由以上过程可知，当 TVLA 得到的 p 值较小或 t 值的绝对值较大时，零假设成立的概率较小，应拒绝零假设，判定存在泄露。TVLA 将复杂的泄露检测问题转化为简捷的计算步骤，具有简单高效等优势，然而该技术同时具有如下缺陷。

1) 对能耗值的分组数过小，无论算法中间值的位数为多少，仅简单地将采集的能耗分为两组。这一缺陷可能导致能耗统计特征的变化被隐藏在其中的一个能耗值分组中。

2) 由式(1)可知，TVLA 检验统计量的构造形式为能耗的一阶矩，这可能与一些场景下能耗泄露的特征并不符合。这一缺陷可能导致能耗统计特征的变化被隐藏在能耗的高阶矩中。

3) 在能量迹样本量较小的情况下，能耗中的噪声等偶然因素对能耗统计特征的影响较大。这可能导致实际上存在泄露的设备在处理不同数据时产生的能耗的统计特征变化并不显著，从而使 TVLA 出现“假阴性”错误。

针对上述第 3 个缺陷，为尽可能地减小能耗样本量较小时噪声的影响，本文对能量迹中得到的 t 值的分布进行研究。经分析推导发现，随着能量迹样本量的变化，泄露点处的 t 值和能量迹样本量存在正相关的比例关系，而非泄露点处的 t 值和能量迹样本量无相关关系。据此，本文提出在 2 个不同能量迹样本量下分别实施非特定 TVLA，然后对得到的 2 个 t 值序列实施双样本 KS 检验判断其是否服从相同分布。当 2 个 t 值序列服从相同分布时说明在不同能量迹样本量下 t 值的分布无明显变化，可以认为能量迹中不存在泄露，否则认为存在泄露。

2　基于双样本 KS 检验的非特定 TVLA 方法

2.1　非特定 TVLA 中能量迹样本量与 t 值关系

密码设备的能量消耗依赖于执行的密码操作和处理的数据，可将能耗中的操作依赖分量记为 P_{op}，数据依赖分量记为 P_{data}；同时，能耗中不可避免地含有与所执行的密码操作和所处理的中间值无关的服从正态分布的噪声分量 P_{noise} 以及由漏电流等产生的常量部分 P_{const}。据此，可用式(3)刻画密码设备的总能量 P_{total} [20]。

$$P_{\mathrm{total}}=P_{\mathrm{op}}+P_{\mathrm{data}}+P_{\mathrm{noise}}+P_{\mathrm{const}} \tag{3}$$

TVLA 是逐采样点实施的过程，本节根据式(3)中的能耗模型对能量迹中的单个采样点进行分析，将该采样点记为 sp，分组 Ψ_0 和 Ψ_1 中的能耗值分别

记为 P_{Ψ_0} 和 P_{Ψ_1}。将能量迹样本量为 n 时 Ψ_0 和 Ψ_1 的样本量、样本均值和样本方差分别计为 (n_0,μ_0,S_0^2) 和 (n_1,μ_1,S_1^2)；能量迹样本量为 kn 时 Ψ_0 和 Ψ_1 的样本量、样本均值和样本方差分别记为 $(n_0',\mu_0',S_0'^2)$ 和 $(n_1',\mu_1',S_1'^2)$。

1) 采样点 sp 是泄露点

a) 分组 Ψ_0

各条能量迹对应相同的明文和密钥，故密码算法的中间值相等，则能耗值中的数据依赖分量 P_{data} 均相等；各能量迹的 sp 采样点对应相同的密码操作，故各能耗值中的操作依赖分量 P_{op} 相等；由定义知各能耗值中常量部分 P_{const} 也相等。因此可将 P_{Ψ_0} 中的 P_{data}、P_{op} 及 P_{const} 之和记为一个常量 C_0，则有

$$P_{\Psi_0}=P_{\text{op}}+P_{\text{data}}+P_{\text{noise}}+P_{\text{const}}=P_{\text{noise}}+C_0 \tag{4}$$

定理 1 随机变量 $X\sim N(\mu,\sigma^2)$，则 X 的线性函数 $Y=aX+b(a\neq 0)\sim N(a\mu+b,(a\sigma)^2)$。

由于噪声分量 P_{noise} 服从正态分布，结合定理 1 可知 P_{Ψ_0} 同样服从正态分布，设 $P_{\Psi_0}\sim N(\mu_{\Psi_0},\sigma_{\Psi_0}^2)$。

b) 分组 Ψ_1

各条能量迹对应服从均匀分布的随机明文和固定密钥，故各能耗值中操作依赖分量 P_{op} 和常量分量 P_{const} 均相等，可将二者之和记为一个定值 C_1，则有

$$P_{\Psi_1}=P_{\text{op}}+P_{\text{data}}+P_{\text{noise}}+P_{\text{const}}=P_{\text{data}}+P_{\text{noise}}+C_1 \tag{5}$$

定理 2 若 $X_i\sim N(\mu_i,\sigma_i^2),i=1,2,\cdots,n$，则它们的和 $Z=X_1+X_2+\cdots X_n$ 仍服从正态分布，且 $Z\sim N(\mu_1+\mu_2+\cdots+\mu_n,\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)$。

当被处理的数据服从均匀分布时，能量消耗中的数据依赖分量 P_{data} 服从正态分布[20]。同时，噪声分量 P_{noise} 服从正态分布且 P_{noise} 和 P_{data} 相互独立。可知 P_{Ψ_1} 同样服从正态分布，设 $P_{\Psi_1}\sim N(\mu_{\Psi_1},\sigma_{\Psi_1}^2)$。

定理 3 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，$\overline{\boldsymbol{X}}$ 和 $\boldsymbol{S}^2$ 分别是样本均值和样本方差，则有

$$\overline{\boldsymbol{X}}\sim N\left(\mu,\frac{\sigma^2}{n}\right),\mathrm{E}(\overline{\boldsymbol{X}})=\mu,D(\overline{\boldsymbol{X}})=\frac{\sigma^2}{n} \tag{6}$$

$$\mathrm{E}(\boldsymbol{S}^2)=\mathrm{E}\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i^2-n\overline{\boldsymbol{X}}^2\right)\right]=\frac{1}{n-1}\left[\sum_{i=1}^{n}\mathrm{E}(\boldsymbol{X}_i^2)-n\mathrm{E}(\overline{\boldsymbol{X}}^2)\right]=\sigma^2 \tag{7}$$

由于 $P_{\Psi_0}\sim N(\mu_{\Psi_0},\sigma_{\Psi_0}^2)$ 且 $P_{\Psi_1}\sim N(\mu_{\Psi_1},\sigma_{\Psi_1}^2)$，则采样点 sp 处 Ψ_0 和 Ψ_1 中的能耗值可分别看作来自 2 个正态分布的样本，由定理 3 可得 Ψ_0 和 Ψ_1 中能耗样本的均值和方差的期望分别为

$$\mathrm{E}(\overline{P_{\Psi_0}})=\mu_{\Psi_0},\mathrm{E}(S^2(P_{\Psi_0}))=\sigma_{\Psi_0}^2 \tag{8}$$

$$\mathrm{E}(\overline{P_{\Psi_1}})=\mu_{\Psi_1},\mathrm{E}(S^2(P_{\Psi_1}))=\sigma_{\Psi_1}^2 \tag{9}$$

定理 4 随机变量 X 具有数学期望 $\mathrm{E}(X)=\mu$，方差 $D(X)=\sigma^2$，则对于任意正数 ε 有

$$P\{|X-\mu|\geqslant\varepsilon\}\leqslant\frac{\sigma^2}{\varepsilon^2} \tag{10}$$

结合式(8)和式(9)可得

$$\mu_0'\approx\mathrm{E}(P_{\Psi_0})=\mu_{\Psi_0},S_0'^2\approx\mathrm{E}(S^2(P_{\Psi_0}))=\sigma_{\Psi_0}^2 \tag{11}$$

$$\mu_1'\approx\mathrm{E}(P_{\Psi_1})=\mu_{\Psi_1},S_1'^2\approx\mathrm{E}(S^2(P_{\Psi_1}))=\sigma_{\Psi_1}^2 \tag{12}$$

将能量迹样本量为 n 时的 t 值记为 t_n，即

$$t_n=\frac{\mu_0-\mu_1}{\sqrt{\frac{S_0^2}{n_0}+\frac{S_1^2}{n_1}}} \tag{13}$$

当样本量变为 kn 时，由于非特定 TVLA 的分组依据不变，因此 Ψ_0 和 Ψ_1 样本量的比例不变，则有 $n_0'=kn_0,n_1'=kn_1$。由式(11)和式(12)可知，样本均值和样本方差的近似值和样本量无关，因此 $\mu_0'\approx\mu_0,\mu_1'\approx\mu_1$ 且 $S_0'^2\approx S_0^2,S_1'^2\approx S_1^2$。将能量迹样本量为 kn 时的 t 值记为 t_{kn}，则有

$$t_{kn}=\frac{\mu_0'-\mu_1'}{\sqrt{\frac{S_0'^2}{n_0'}+\frac{S_1'^2}{n_1'}}}\approx\frac{\mu_0-\mu_1}{\sqrt{\frac{S_0^2}{kn_0}+\frac{S_1^2}{kn_1}}}=\sqrt{k}\frac{\mu_0-\mu_1}{\sqrt{\frac{S_0^2}{n_0}+\frac{S_1^2}{n_1}}}=\sqrt{k}t_n \tag{14}$$

2) 采样点 sp 不是泄露点

根据本文第 1 节对泄露检测原理的分析可知，无论能量迹样本量为多少，Ψ_0 和 Ψ_1 的样本量都相等，样本均值和样本方差之间均不存在显著的统计差异，即

$$n_0'=n_1'=kn_0=kn_1 \tag{15}$$

$$\mu_0\approx\mu_0'\approx\mu_1\approx\mu_1',S_0^2\approx S_0'^2\approx S_1^2\approx S_1'^2 \tag{16}$$

则 $\mu_0-\mu_1$ 和 $\mu_0'-\mu_1'$ 均可看作一个无穷小量 o。

定理 5 无穷小和有界函数之积仍为无穷小。

故能量迹样本量为 n 时的 t 值为

$$t=\frac{\mu_0-\mu_1}{\sqrt{\frac{S_0^{\,2}}{n_0}+\frac{S_1^{\,2}}{n_1}}}=o\frac{1}{\sqrt{\frac{S_0^{\,2}}{n_0}+\frac{S_1^{\,2}}{n_1}}}\to 0 \tag{17}$$

同理可得能量迹样本量为 kn 时的 t 值 $t_{kn}\to 0$，因此 $t_n\approx t_{kn}$。

由以上推导结果可知，当样本量由 n 变为 kn 时，无泄露的采样点处得到的 t 值无显著性变化，而存在泄露的采样点处的 t 值则变为原 t 值的 $\sqrt{k}$ 倍。因此，当存在能量信息泄露时，在不同能量迹样本量下实施非特定 TVLA 得到的 t 值序列所服从的分布不同，而不存在泄露时得到的 t 值序列的分布相同。

2.2　算法设计

本节利用 2.1 节中的结论，引入双样本 KS 检验对非特定 TVLA 在不同能量迹样本量下得到的 t 值序列的分布是否相同进行检验，若分布不同则判定存在泄露，否则判定不存在泄露。双样本 KS 检验用样本的经验分布函数来近似估计总体的累积分布函数，从而对 2 个样本是否服从相同分布进行判断，其零假设和备择假设分别为 H_0 表示 2 个样本服从相同分布和 H_1 表示 2 个样本不服从相同分布。对于某抽样样本的观测值 $s_1,s_2,\cdots,s_n$，其经验分布函数 $F(x)$ 为

$$F(x)=\frac{1}{n}\sum_{i=1}^{n}I_{s_i} \tag{18}$$

其中，当 $s_i\leqslant x$ 时 $I_{s_i}=1$，当 $s_i>x$ 时 $I_{s_i}=0$。双样本 KS 检验的统计量为

$$D=\max\left|F'(x)-F''(x)\right| \tag{19}$$

其中，$F'(x)$ 和 $F''(x)$ 分别为 2 个样本的经验分布函数。检验阈值为

$$T=\sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}\sqrt{\frac{p+q}{pq}} \tag{20}$$

其中，α 为检验的显著性水平，p和q 分别为 2 个样本的样本量。当 $D\leqslant T$ 时接受 H_0，判定 2 个样本服从相同分布；否则拒绝 H_0，判定 2 个样本不服从相同分布。由此本文提出基于双样本 KS 检验的非特定 TVLA 方法：将采集的 n 条能量迹（每条含 p 个采样点）记为能耗矩阵 $\boldsymbol{A}_{n\times p}$，首先依次对矩阵 $\boldsymbol{A}_{n\times p}$ 中能量迹的每个采样点依次实施非特定 TVLA，分别用 num 、mean 和 var 函数求出 2 个分组 Ψ_0^A 和 Ψ_1^A 的样本量、样本均值和样本方差并根据式(1)求出 t 值，将得到的 p 个 t 值记作序列 $\boldsymbol{S}$；再从矩阵 $\boldsymbol{A}_{n\times p}$ 中选择出 $kn(0<k<1)$ 条能量迹，将其记为矩阵 $\boldsymbol{B}_{kn\times p}$，对矩阵 $\boldsymbol{B}$ 实施相同操作得到 t 值序列 $\boldsymbol{R}$（由于能量迹中采样点数量不变，故 $\boldsymbol{S}$ 和 $\boldsymbol{R}$ 中均有 p 个 t 值）；利用 ecdf 函数分别求得序列 $\boldsymbol{S}$ 和 $\boldsymbol{R}$ 的经验分布函数 F_S 和 F_R，利用 abs 和 max 函数求得 F_S 和 F_R 之差的绝对值的最大值 D，并比较 D 和阈值 T 的大小，当 $D\leqslant T$ 时令 $C=0$，判定不存在泄露，反之令 $C=1$，判定存在泄露。上述主要步骤的流程如图 1 和算法 1 所示。

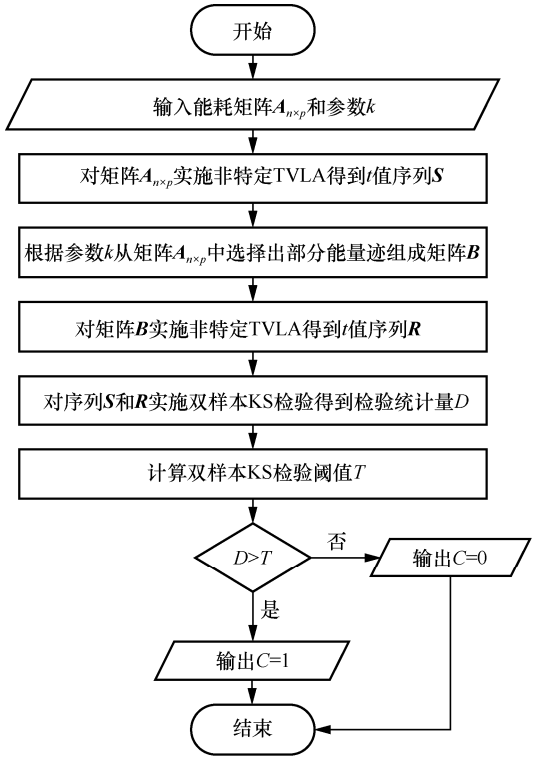


图 1　基于双样本 KS 检验的非特定 TVLA 流程

算法 1　基于双样本 KS 检验的非特定 TVLA

输入　$\boldsymbol{A}_{n\times p}$，$k$；

输出　C；

1) Begin

2) 对 $\boldsymbol{A}_{n\times p}$ 实施 TVLA，将得到的 t 值序列赋给 $\boldsymbol{S}$

3) for $i=1$ to p do

4) $z_A\leftarrow \mathrm{mean}(\Psi_0^A)-\mathrm{mean}(\Psi_1^A)$；

5) $w_0^A \leftarrow \dfrac{\mathrm{var}(\Psi_0^A)^2}{\mathrm{num}(\Psi_0^A)}$;

6) $w_1^A \leftarrow \dfrac{\mathrm{var}(\Psi_1^A)^2}{\mathrm{num}(\Psi_1^A)}$;

7) $w_A \leftarrow (w_0^A + w_1^A)^{\frac{1}{2}}$;

8) $\boldsymbol{S}[i] \leftarrow \dfrac{z_A}{w_A}$;

9) end for

10) 从 $\boldsymbol{A}_{n\times p}$ 中选择出 kn 条能量迹组成矩阵 $\boldsymbol{B}$

11) $\boldsymbol{B} \leftarrow \mathrm{select}(\boldsymbol{A}, k)$;

12) 对 $\boldsymbol{B}$ 实施 TVLA，将得到的 t 值序列赋给 $\boldsymbol{R}$

13) for $i = 1$ to p do

14) $z_B \leftarrow \mathrm{mean}(\Psi_0^B) - \mathrm{mean}(\Psi_1^B)$;

15) $w_0^B \leftarrow \dfrac{\mathrm{var}(\Psi_0^B)^2}{\mathrm{num}(\Psi_0^B)}$;

16) $w_1^B \leftarrow \dfrac{\mathrm{var}(\Psi_1^B)^2}{\mathrm{num}(\Psi_1^B)}$;

17) $w_B \leftarrow (w_0^B + w_1^B)^{\frac{1}{2}}$;

18) $\boldsymbol{R}[i] \leftarrow \dfrac{z_B}{w_B}$;

19) end for

20) 求 $\boldsymbol{S}$ 的经验分布函数

21) $F_S \leftarrow \mathrm{ecdf}(\boldsymbol{S})$;

22) 求 $\boldsymbol{R}$ 的经验分布函数

23) $F_R \leftarrow \mathrm{ecdf}(\boldsymbol{R})$;

24) 求双样本 KS 检验的统计量

25) $D \leftarrow \max(\mathrm{abs}(F_S(x) - F_R(x)))$;

26) 求双样本 KS 检验的阈值

27) $T \leftarrow \left(\dfrac{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)2p}{p^2} \right)^{\frac{1}{2}}$;

28) if $D \leqslant T$ then

29) $C \leftarrow 0$;

30) else

31) $C \leftarrow 1$;

32) end if

33) $C = 0$ 则不存在泄露，$C = 1$ 则存在泄露

34) return C;

35) end

3 实验验证

3.1 实验配置

本节从以下方面对算法 1 展开验证。

1) 用不同平台的能耗数据分别进行验证。分别采用了 MATLAB 仿真能耗数据、DPA Contest v4_2 数据集[21]和基于 Chipwhisperer 开发板的自测能耗数据。此外，本文对这 3 个能耗数据集分别实施了相关能量分析（CPA, correlation power analysis），以明确能耗数据中确实存在能量信息泄露，从而进一步验证本文方法的有效性。

2) 对无防护措施和加防护措施的情况分别进行验证。MATLAB 仿真能耗未加防护措施，DPA Contest v4_2 数据集和自测能耗数据设置了一阶 RSM 掩码防护措施。

3) 用对齐和非对齐的能耗数据分别进行验证。仿真能量迹和 DPA Contest v4_2 数据集中的能耗数据均是对齐的，自测所得能量迹存在抖动。

4) 用不同密码算法分别验证。仿真能耗数据和 DPA Contest v4_2 上的验证针对 AES-128 算法展开，自测能耗数据上的验证针对 SM4 算法展开。

5) 对不同的泄露检测方法进行对比：分别实现本文方法、TVLA[5]、基于 HC 的 TVLA[11]、基于配对的 TVLA[13]、基于深度学习的泄露检测方法[18]和基于 Levene 检验的泄露检测方法[19]并进行了对比。其中，对于 TVLA 和基于配对的 TVLA，为减小噪声等偶然因素影响，在每个样本量下实施两次检验，在各采样处对得到的 2 个 t 值平均值的绝对值进行统计；在基于深度学习的检测方法中，简便起见，使用多层感知器网络模型并将训练过程迭代次数设为 20，将验证集样本量设为 1 000 并使其中 2 个标签分组的能量迹样本量相等，实验中能量迹样本量均指与验证集互相独立的训练集样本量。需要明确的是，基于 Levene 检验的泄露检测方法针对的是加掩码防护的情形，故在无防护的仿真能耗数据中未对比该方法。

本文实验中将显著性水平均设置为 $\alpha = 0.001$，则可由式(20)计算双样本 KS 检验阈值，基于 HC 的 TVLA 检验阈值为 31.65，基于深度学习和基于 Levene 则均以得到的 p 值的对数形式–lg(p)为统计量，阈值均为 5，而在 TVLA 及基于配对的 TVLA

中，将检验阈值设为定值会导致产生误判的概率随能量迹中采样点数量的增加而增大，使对不同长度的能量迹实施评估的精确度不同。因此，为实现对不同长度能量迹的公平评估，应根据能量迹中采样点的数量设置阈值。设能量迹中采样点的数量为l，TVLA 的整体显著性水平为α，单次检验的显著性水平为α_{TH}，则有

$$\alpha = 1-(1-\alpha_{\mathrm{TH}})^{l} \tag{21}$$

结合 t 分布的累积分布函数CDF_t可得阈值 TH 和α_{TH}的关系为

$$\alpha_{\mathrm{TH}} = 2(1-\mathrm{CDF}_t(|\mathrm{TH}|,v)) \tag{22}$$

进一步可得

$$|\mathrm{TH}| = \mathrm{CDF}_t^{-1}\left(\frac{1+(1-\alpha)^{\frac{1}{l}}}{2},v\right)\approx \mathrm{CDF}_{N(0,1)}^{-1}\left(\frac{1+(1-\alpha)^{\frac{1}{l}}}{2}\right) \tag{23}$$

3.2　基于无防护对齐仿真能耗数据的验证

本节利用 MATLAB 工具实现 AES-128 算法，并通过式(24)中的能耗模型仿真能量消耗。

$$P = \mathrm{HW}(M)+N \tag{24}$$

其中，P为仿真能耗值，HW 为汉明重量函数，M为算法中间值，N为服从标准正态分布的噪声。验证过程中，首先采集 6 000 条仿真能量迹，每条能量迹设置 3 000 个采样点，对其实施非特定 TVLA，根据式(23)可计算得到单次 t 检验的阈值为$t=5.103\,5$，双样本 KS 检验阈值为 0.050 3。然后分别挑选出 3 000 条和 1 500 条能量迹重复实施非特定 TVLA。在上述 3 个不同样本量下，能量迹中的部分采样点得到的 t 值如图 2 所示。

具体统计分析如下。

1) 当样本量为 6 000 时，采样点区间(500,750)和(2 100,2 400)内大部分 t 值超过了阈值，经重复实验，该 2 个区间内多数采样点处的 t 值在相同方向（2 次得到的 t 值正负相同）超过了阈值，证明这 2 个采样点区间内存在泄露。

2) 当样本量为 1 500 和 3 000 时，2 个泄露区间内的 t 值均未超过阈值，说明在这 2 个能量迹样本量下存在“假阴性”误报错误。

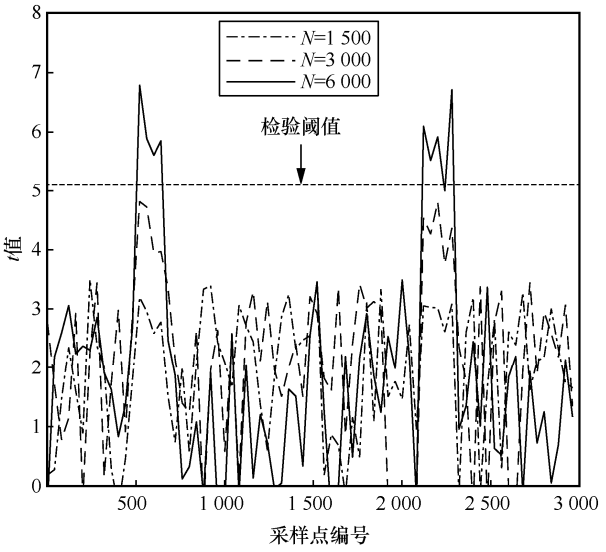


图 2　仿真样本量为 1 500、3 000 和 6 000 时的部分采样点 t 值

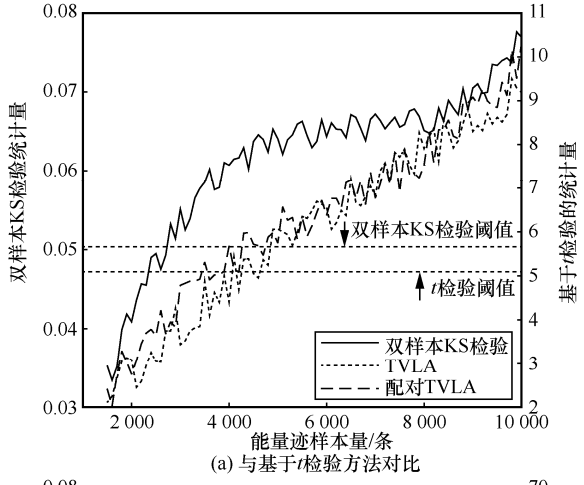
3) 在 3 个泄露区间内，样本量越大，得到的 t 值越大。在泄露区间外，随样本量变化 t 值无明显变化。

从图 2 中可知，在能量迹中泄露点处得到的非特定 TVLA 的 t 值随样本量的增大而增大，而非泄露点处的 t 值无显著性变化，因此当样本量较小时，泄露点处的 t 值会因为样本量不足而无法超过阈值导致“假阴性”误判。

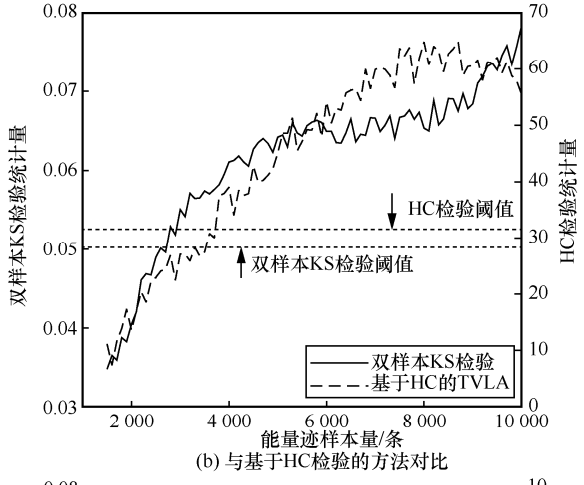
同时，对以上 3 个不同样本量下得到的 t 值序列实施双样本 KS 检验可得$D_{1\,500,3\,000}$、$D_{1\,500,6\,000}$和$D_{3\,000,6\,000}$分别为 0.056 7、0.073 5 和 0.062 3，均大于阈值 0.050 3。因此在以上 3 个能量迹样本量下，算法 1 在样本量为 3 000 时即可检测出泄露，而 TVLA 在样本量为 6 000 时才能发现泄露。

然后在不同能量迹样本量下分别实施 3.1 节所述各泄露检测方法并进行对比，图 3 为对比结果，其中本文方法以左侧 y 轴为参照，其他方法均以右侧 y 轴为参照。

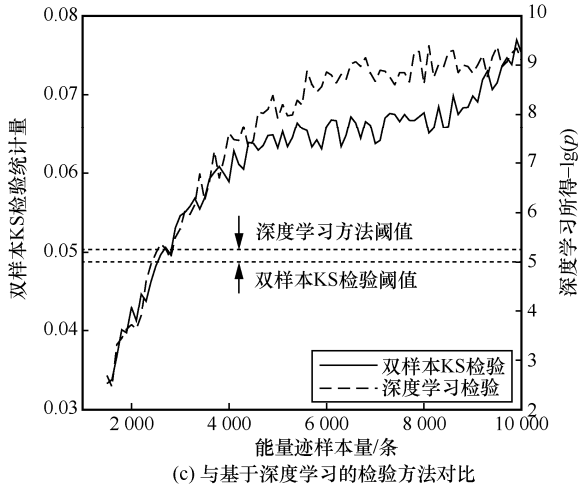
由图 3 可知，在 3 次对比实验中，本文方法最少仅需 2 568 条能量迹即可使统计量超过阈值；TVLA、基于配对的 TVLA、基于 HC 的 TVLA 和深度学习检测方法分别需要 3 980、4 768、3 632 和 2 896 条能量迹。本文方法所需能量迹分别减少了 35.5%、46.1%、29.3 和 11.3%。



(a) 与基于t检验方法对比



(b) 与基于HC检验的方法对比



(c) 与基于深度学习的检验方法对比

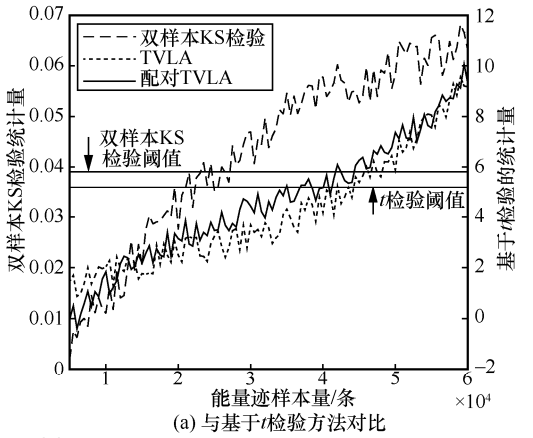
图 3 仿真环境对比结果

3.3 基于加防护对齐能耗数据集的验证

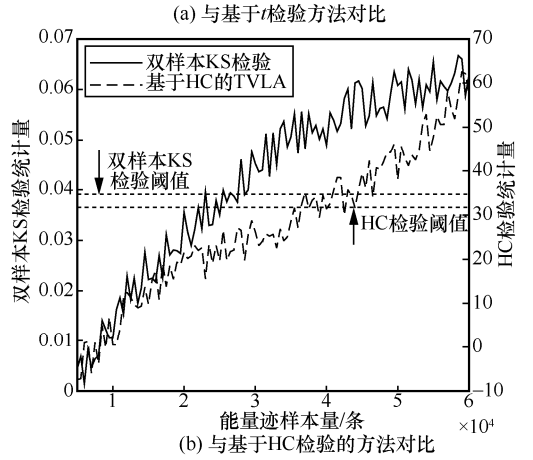
本节利用公开能耗数据集 DPA Contest v4_2 进行验证，该数据集采集的是 Atmel ATMega-163 智能卡上加 RSM 掩码防护 AES-128 算法的能耗，实验中根据式(25)进行预处理[22]。

$$L' = \prod_{i=1}^{d} (L_{t_i} - \mu_{t_i}) \tag{25}$$

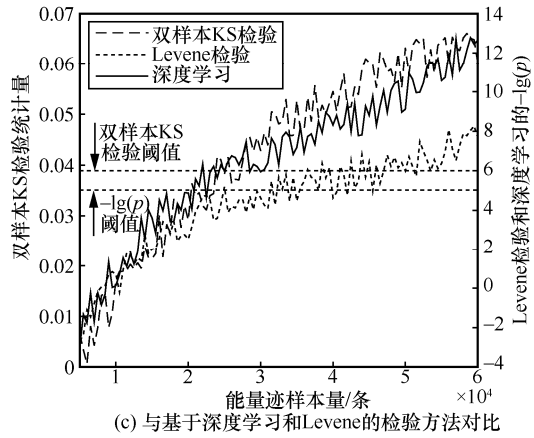
其中，L_{t_i} 为 t_i 时刻的能耗，μ_{t_i} 为 t_i 时刻能耗的均值，d 为共享因子的数量，L' 为处理后的能耗。验证过程中同样地统计各方法在不同能量迹样本量下的检验统计量，结果如图 4 所示。



(a) 与基于t检验方法对比



(b) 与基于HC检验的方法对比



(c) 与基于深度学习和Levene的检验方法对比

图 4 DPA Contest v4_2 数据集对比结果

由图 4 可知，在 3 次实验中本文方法最少需 27 800 条能量迹即可使统计量超过阈值，TVLA、基于配对的 TVLA、基于 HC 的 TVLA、基于 Levene 的检测方法和深度学习检测方法分别需要 45 590、41 750、42 710、41 020 和 31 200 条能量迹，本文方法所需能量迹分别减少了 39.0%、33.4%、34.9%、32.2%和 10.9%。

3.4　基于加防护非对齐实测能耗数据的验证

本节利用 Chipwhisperer 开发板采集 SM4 算法的能耗展开验证，目标板为 CW303 单片机，采集板为 CW1173，通过自带的 OpenADC 模块直接采集能耗。实验前根据式(25)进行预处理，实验中调整采样频率将能量迹中采样点数量设置为 5 000，可求得双样本 KS 检验和 t 检验的阈值分别为 0.039 和 5.199 2。对比结果如图 5 所示。

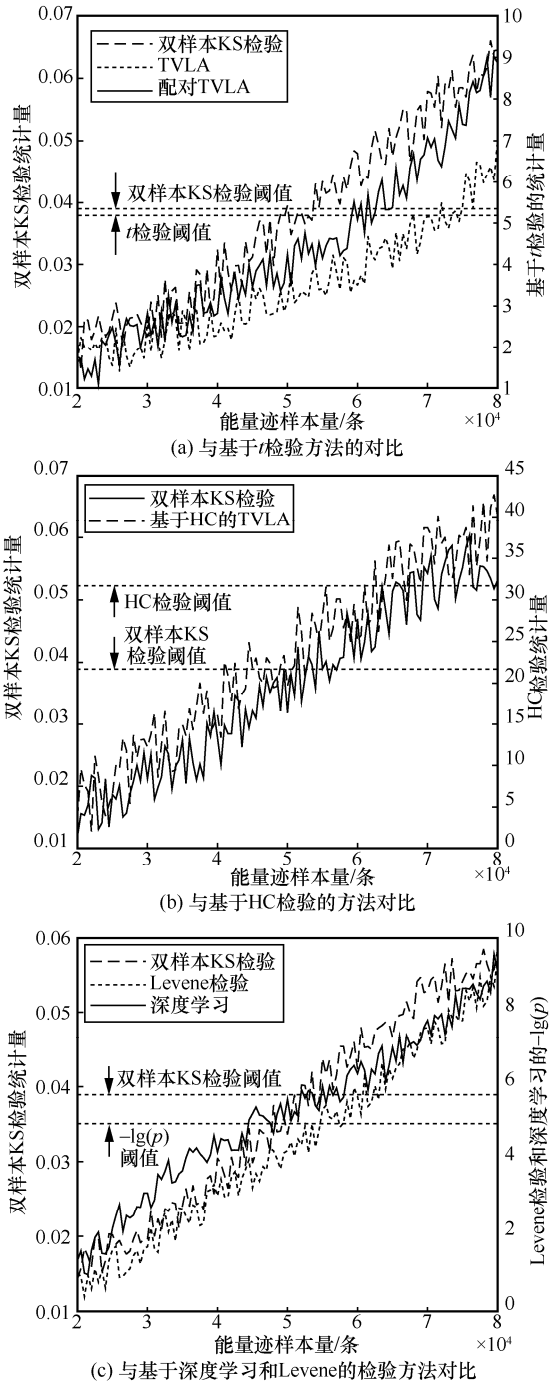


图 5　实测环境实验对比结果

经统计，在 3 次实验中本文方法最少需 53 680 条能量迹可使统计量超过阈值，TVLA、基于配对的 TVLA、基于 HC 的 TVLA、基于 Levene 的检测方法和深度学习检测方法分别需要 76 060、66 840、64 280、59 220 和 49 960 条能量迹，本文方法所需能量迹比深度学习方法增加了 7.4%，比其他方法分别减少了 29.4%、19.7%、16.5%和 9.4%。

为使验证结果更全面，利用相关系数法对能量迹进行了对齐处理，然后在处理后的能耗数据上分别实施本文方法和深度学习检测方法并进行对比，结果如图 6 所示。

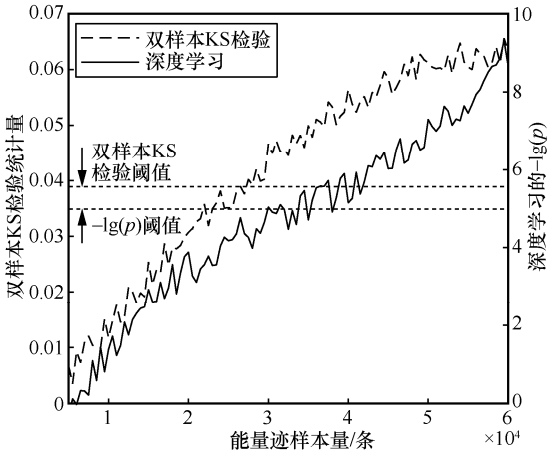


图 6　实测环境下对齐后本文方法与深度学习方法的对比

本文方法需 26 500 条能量迹即可使统计量超过阈值，而深度学习检测方法需 35 280 条能量迹，本文方法所需样本量比深度学习方法少 24.9%。

3.5　攻击验证

本节对上述 3 个能耗数据集分别实施 CPA 攻击以验证本文算法 1 的有效性。在实施攻击前，利用相关系数法对未对齐的自测能耗数据进行了对齐处理，并根据式(25)进行了预处理。

实施攻击过程中，分别根据明文和密钥计算得到 AES-128 算法和 SM4 算法第一轮 S 盒输出，并根据汉明重量模型得到能耗的映射值，然后通过计算该映射值和能量迹中的实际能耗值之间的相关系数实施攻击。

对 3 个能耗数据集实施攻击时，分别以 3 000、20 000 和 30 000 为初始能量迹样本量，以 300、1 000 和 1 500 为样本量增量，不能攻破时即增加样本量，直至攻破密钥。实验结果表明，对 3 个能耗数据集的攻击分别在样本量增加至 3 600、33 000 和 37 500

时破获密钥。

攻击验证的结果表明这 3 个能耗数据集中确实存在泄露，且攻破密钥所需能耗样本量与实施评估检测出泄露所需样本量的大小次序相一致。

综合分析以上各实验结果可得如下结论。

1) 相较深度学习泄露检测方法，在对 AES-128 和 SM4 算法的验证中，在能量迹未对齐的情况下本文方法检测出泄露所需样本量稍大；经对齐处理后，本文方法所需样本量小于深度学习方法。

2) 相较其他泄露检测方法，无论是否施加防护措施，无论能量迹是否对齐，在对 AES-128 和 SM4 算法的验证中本文方法均能以更小的能耗样本量检测出泄露。

综上，本文所提非特定 TVLA 方法能够有效减小因能耗样本量较小而产生的“假阴性”误判错误。

4 结束语

针对能量迹样本量较小时 TVLA 易产生“假阴性”误判的问题，本文提出利用双样本 KS 检验对不同样本量下得到的非特定 TVLA 的 t 值序列进行分析从而判断泄露情况。所提方法利用能量迹中 t 值的分布随能耗样本量的变化对泄露情况进行判断，避免了因能耗样本量偏小造成的泄露点处 t 值不超过阈值的现象，减小了“假阴性”误判概率。

然而所提方法仍存在以下问题：1) 选择能量迹的方法有待优化；2) 需重复实施 TVLA，评估效率有所降低。后续研究将从以下方面开展：1) 选择能量迹的方法；2) 根据推导得到的检验 t 值和能量迹样本量的变化关系对泄露检测所需样本量进行研究。

参考文献：

[1] KOCHER P, JAFFE J, JUN B. Differential power analysis[C]//Advances in Cryptology - CRYPTO’ 99. Berlin: Springer, 1999: 388-397.

[2] BRIER E, CLAVIER C, OLIVIER F. Correlation power analysis with a leakage model[C]//Cryptographic Hardware and Embedded Systems-CHES 2004. Berlin: Springer, 2004: 16-29.

[3] GIERLICHS B, BATINA L, TUYLS P, et al. Mutual information analysis[C]//Cryptographic Hardware and Embedded Systems - CHES 2008. Berlin: Springer, 2008: 426-442.

[4] 吴震, 王燚, 周冠豪. 有学习的高阶 DPA 攻击[J]. 通信学报, 2018, 39(9): 135-146.
WU Z, WANG Y, ZHOU G H. High order DPA with profiling[J]. Journal on Communications, 2018, 39(9): 135-146.

[5] GOODWILL G, JUN B, JAFFE J, et al. A testing methodology for side-channel resistance validation[C]//NIST Non-Invasive Attack Testing Workshop. [S.l.:s.n.], 2011: 115-136.

[6] STANDAERT F X. How (not) to use Welch’s t-test in side-channel security evaluations[C]//Smart Card Research and Advanced Applications. Berlin: Springer, 2019: 65-79.

[7] DAO B A, HOANG T T, LE A T, et al. Correlation power analysis attack resisted cryptographic RISC-V SoC with random dynamic frequency scaling countermeasure[J]. IEEE Access, 2021, 9: 151993-152014.

[8] STEINBAUER T, NAGPAL R, PRIMAS R, et al. TVLA on selected NIST LWC finalists[EB]. 2022.

[9] LU C C, CUI Y J, KHALID A, et al. A novel combined correlation power analysis (CPA) attack on schoolbook polynomial multiplication in lattice-based cryptosystems[C]//Proceedings of 2022 IEEE 35th International System-on-Chip Conference (SOCC). Piscataway: IEEE Press, 2022: 1-6.

[10] JEVTIC R, OTERO M G. Methodology for complete decorrelation of power supply EM side-channel signal and sensitive data[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022, 69(4): 2256-2260.

[11] DING A A, ZHANG L W, DURVAUX F, et al. Towards sound and optimal leakage detection procedure[C]//International Conference on Smart Card Research and Advanced Applications. Berlin: Springer, 2018: 105-122.

[12] 王娅茹, 唐明. 基于Bartlett和多分类F检验侧信道泄露评估[J]. 通信学报, 2021, 42(12): 35-43.
WANG Y R, TANG M. Side channel leakage assessment with the Bartlett and multi-classes F-test[J]. Journal on Communications, 2021, 42(12): 35-43.

[13] DING A A, CHEN C, EISENBARTH T. Simpler, faster, and more robust t-test based leakage detection[C]//International Workshop on Constructive Side-Channel Analysis and Secure Design. Berlin: Springer, 2016: 163-183.

[14] 鹿福祥, 李伟键, 黄娴. 基于配对 t 检验的侧信道泄露评估优化研究[J]. 小型微型计算机系统, 2019, 40(12): 2585-2590.
LU F X, LI W J, HUANG X. Research on optimization of side channel leakage assessment based on paired t test[J]. Journal of Chinese Computer Systems, 2019, 40(12): 2585-2590.

[15] ZHANG L W. Statistics in side channel analysis-modeling, metric, leakage detection testing[D]. Boston: Northeastern University, 2017.

[16] MORADI A, RICHTER B, SCHNEIDER T, et al. Leakage detection with the x2-test[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2018(1): 209-237.

[17] BACHE F, PLUMP C, GÜNEYSU T. Confident leakage assessment-a side-channel evaluation framework based on confidence intervals[C]//Proceedings of 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE). Piscataway: IEEE Press, 2018:

1117-1122.

[18] MOOS T, WEGENER F, MORADI A. DL-LA: deep learning leakage assessment[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021: doi.org/10.46586/tches.v2021.i3.552-598.

[19] WANG Y, TANG M, WANG P, et al. The Levene test based-leakage assessment[J]. Integration, 2022, 87: 182-193.

[20] MANGARD S, OSWALD E, POPP T. Power analysis attacks: revealing the secrets of smart cards[M]. Berlin: Springer, 2007.

[21] BHASIN S, BRUNEAU N, DANGER J L, et al. Analysis and Improvements of the DPA Contest v4 Implementation[C]//International Conference on Security, Privacy, and Applied Cryptography Engineering. Berlin: Springer, 2014: 201-218.

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

[作者简介]

郑震（1996– ），男，陕西宝鸡人，信息工程大学博士生，主要研究方向为侧信道安全评估技术。

严迎建（1973– ），男，河南扶沟人，博士，信息工程大学教授、博士生导师，主要研究方向为芯片安全技术等。

蔡爵嵩（1992– ），男，四川绵阳人，信息工程大学硕士生，主要研究方向为侧信道安全攻防技术。

刘燕江（1990– ），男，河南南阳人，博士，信息工程大学讲师，主要研究方向为芯片安全技术等。