侧信道能量信息测试向量泄漏评估技术

郑 震* 严迎建 刘燕江

(战略支援部队信息工程大学 郑州 450001)

摘 要：侧信道能量分析攻击技术以其计算复杂度低和通用性强等优势，给各类密码产品带来了严峻的安全挑战。抗能量分析攻击能力的评估已经成为密码产品安全性测评的重要环节。测试向量泄漏评估(TVLA)是一种基于假设检验的能量信息泄漏评估方法，具有简单高效和可操作性强等特点，目前被广泛应用于密码产品的安全性评估实验中。为全面把握TVLA技术机理及研究现状，该文首先对TVLA技术进行了概述，阐述了其实现原理并介绍了其实施过程，紧接着对特定和非特定两种TVLA的优势与不足进行了对比，随后参考已有研究，对TVLA的局限性进行了深入分析和归纳，在此基础上重点介绍并分析了已有的TVLA的改进方法，最后对TVLA未来可能的发展方向进行了展望。

关键词：安全评估；侧信道；能量分析攻击；测试向量评估

中图分类号：TP309 **文献标识码**：A **文章编号**：1009-5896(2023)09-3109-09

DOI: 10.11999/JEIT230295

Test Vector Leakage Assessment Technique of Side-channel Power Information

ZHENG Zhen YAN Yingjian LIU Yanjiang

(*Information Engineering University of Strategic Support Force*, *Zhengzhou* 450001, *China*)

Abstract: The side-channel power analysis attack technique, with its advantages of low computational complexity and high generality, poses a critical security challenge to all kinds of cryptographic implementations. The assessment of resistance to power analysis attacks has become an essential aspect of cryptographic product security evaluation. Test Vector Leakage Assessment (TVLA) is a power information leakage evaluation method based on hypothesis testing techniques, which is highly efficient and operable, and is now widely used in security evaluation experiments of cryptographic products. In order to have a comprehensive understanding of the mechanism of TVLA technology and the current status of research, this paper begins with an overview of TVLA technology, including an explanation of its implementation principles and a description of its implementation process, followed by a comparison of the advantages and disadvantages of both specific and non-specific TVLA. The limitations of TVLA are then analyzed and summarized in depth with reference to existing studies, based on which existing approaches for improving TVLA are highlighted and analyzed, and finally the possible future directions of TVLA are prospected.

Key words: Security assessment; Side-channel; Power analysis attack; Test Vector Leakage Assessment (TVLA)

1 引言

1999年Kocher等人[1]通过收集密码设备运行过程中的能量消耗获取了密码算法密钥等信息，该研究催生了新的密码攻击形式——侧信道能量分析攻击。自此之后，经过20余年发展，能量分析攻击已经发展成为包含差分能量分析(Differential Power Analysis, DPA)、简单能量分析(Simple Power Analysis, SPA)和相关能量分析(Correlation Power Analysis, CPA)等多种形式的密码分析技术，对各类密码产品的机密性和其中的数据完整性构成了严重的现实威胁[2,3]。为应对这些威胁，将针对能量分析攻击的防护措施整合到密码产品上已成为必然，因此对能量信息安全性和防护措施有效性的评估已经逐渐成为密码学领域的关注热点之一。目前为止，任何评估方法都无法保证检测出所有类型的能量信息泄漏，抗能量分析攻击评估实验的目的是判断密码产品或模块中是否存在能量信息泄漏，或是否能够提供与所需安全级别相称的防护能力。一个合格的能量信息泄漏评估方案应满足有效性(评

收稿日期：2023-04-18；改回日期：2023-07-11；网络出版：2023-07-13
*通信作者： 郑震 zhengzhen_0917@163.com

估结果应是可重复实现的)和简易性(不应消耗过多的评估时间或要求评估人员掌握过于专业的技能等)两方面要求[4,5]。

在ANSSI和BSI等机构定义的标准[6]中，通过实施能量分析攻击的方式对待测设备抗能量分析攻击的能力进行评估，若攻击成功，则根据攻击所消耗的时间等资源对待测设备的安全级别进行评定。但是这种基于攻击的评估策略所需时间成本较大且通用性较差：在密码产品的使用过程中，任何一种能量分析攻击方法都可能被用于对其发起攻击，因此为确保评估的全面性，应尽可能地遍历更多的能量分析攻击方法，同时兼顾密码算法的各个中间值以及各种能量模型。然而，即使只尝试个别已知的攻击方法，密码算法中大量的中间值和可供选择的不同能量模型仍会使得评估过程相当耗时，且由于保密的能量分析攻击方法的存在，仍无法保证评估结果的全面性。当密码算法被攻破时，也不能保证该攻击方法即是最有效的攻击手段，因此难以建立一个客观的安全级别的标准。同时，基于攻击的评估策略对评估人员的要求较高，一方面要掌握尽可能多的能量分析攻击方法，另一方面，实施攻击时能量模型的选择很大程度上取决于评估人员的专业知识和经验，如果因能量模型选择不当导致攻击没有成功，存在安全隐患的设备就会被误判为安全[7]。随着新型能量分析攻击方法不断被提出，这种评估策略的实施成本变得越来越大，局限性变得越来越突出。因此，当前评估人员更加倾向于标准化的评估程序，通过实施较为固定的步骤对泄漏情况进行评估。

与其寻找针对特定待测设备的最佳攻击方法并以攻破设备所需的资源消耗量来衡量设备的安全级别，可以利用假设检验的统计手段，先对待测设备的泄漏情况作出假设，然后通过分析采集的能耗数据的统计特征来决定是否接受该决定。这种策略不必确定如何利用或量化这些泄漏，可用于决策待测设备是否能够通过某项安全性测评，但不可用于破获明文或密钥等秘密信息。基于这一思路，Goodwill等人[8]在2011年美国国家标准技术研究院主办的研讨会上针对AES算法提出了一种评估泄漏的新方法，使用一组预先指定的输入明文作为测试向量执行加密操作，然后对收集的能耗数据执行假设检验测试并计算得到置信度，对该置信度的数值设置阈值以建立一个明确的失败/通过标准从而判定是否存在泄漏。Cooper等人[9]对该方法进行了整理并将其命名为测试向量泄漏评估(Test Vector Leakage Assessment, TVLA)，TVLA将泄漏评估与不断发展的能量分析攻击技术分离开来，将复杂的泄漏检测问题转化为简捷的数理统计步骤，使用固定的统计步骤来捕获能量信息泄漏。TVLA不需要评估人员掌握过多的密码算法或能量分析攻击知识，且可以通过修改测试向量等方法来捕捉能量信息泄漏，并根据所需的安全要求调整阈值，具有简单高效等优势。

目前，TVLA已被广泛应用于对后量子密码等各类密码实现的泄漏评估中。Wang等人[10]利用TVLA对ARM Cortex-M4单片机上实现的SABER KEM的解密过程进行了评估，结果表明该实现中的增量存储步骤可能导致密钥信息的泄漏。Saarinen等人[11]讨论了TVLA在KEM的硬件实现上的实施并指出，由于后量子密码的各密钥对间的紧密联系，用固定密钥和随机密钥实施的非特定TVLA的效果不理想，实施TVLA前需对测试向量进行谨慎的选择。Krausz等人[12]针对后量子密码提出了一个基于多项式反演的掩码方案，并通过实施TVLA证明了其有效性。此外，TVLA还被用于其他平台上不同密码体制的安全性测评实验中[13–16]。

本文以TVLA为研究对象，对其研究现状进行归纳总结。首先在第2节对TVLA技术进行了概述，对TVLA的实现机制进行了介绍，并对特定和非特定两种TVLA进行了对比分析。然后在第3节根据已有文献中的研究对TVLA的局限性进行了分析和归纳。接着在第4节对TVLA的研究现状进行了归纳整理，对现有的各类TVLA的优化和替代方法的优势与不足进行了阐述和深入分析，最后在第5节对本文进行了总结并对TVLA的发展前景进行了展望。

2 TVLA概述

2.1 TVLA实现机理

在各种能量分析攻击中，攻击者尝试设计各种数学公式以捕捉由明文或密钥的变化等因素造成的能耗数据的统计学差异，进而破解密码算法。因此，设备能耗中任何的由所处理数据变化引起的显著性的统计学差异都可能会被攻击者加以利用，故当密码算法的明文或敏感中间值等操作数发生变化时，若设备的能耗也随之发生显著性变化，即说明能耗中有攻击者可利用的信息，即存在泄漏。TVLA正是基于这样的原理，对能量信息泄漏进行评估。

TVLA使用Welch's t检验[17]来确定操作数和密码设备能耗间是否存在依赖关系，Welch's t检验是针对不同样本量和不同方差的student t检验的扩展形式。具体地，在实施TVLA时，根据操作数将能耗数据分为两组，并通过检验两组能耗数据之间的

均值差异来判断是否存在泄漏。检验的零假设是两组能耗数据均值相同(即相应的操作数对设备能耗没有影响，不存在泄漏)，备择假设是两组能耗数据均值不同(即相应的操作数对设备能耗有影响，存在泄漏)。TVLA的检验统计量t值的计算如式(1)

$$t=\frac{\mu_0-\mu_1}{\sqrt{{S_0}^2/n_0+{S_1}^2/n_1}} \tag{1}$$

其中，(n_0,μ_0,S_0^2)和(n_1,μ_1,S_1^2)分别为两组能耗数据的样本量、样本均值和样本方差。TVLA需要对能量迹中的采样点逐个实施检验，Welch's t检验则用于判断所采集的能耗数据是否提供了足够的证据来拒绝零假设，检验统计量t出现高正值或负值表明零假设不正确的置信度较高，任一处采样点处的t值超出阈值即可判定待测设备或算法存在能量信息泄漏。

对t值选择不同的阈值C可以使得$t>C$或$t<-C$的概率对应于零假设被拒绝的不同置信度。选择一个较大的阈值可以使犯假阳性(实际无泄漏却判定存在泄漏)误判错误的可能性较小，但大阈值会增加犯假阴性(实际有泄漏却判定不存在泄漏)误判错误的概率，因此为平衡泄漏检测的需求，同时控制假阳性和假阴性错误出现的概率，阈值不能过大也不能过小，TVLA可以实施两次相互独立的实验，只有在两次实验中检验统计量t在同一方向上均超过阈值(均大于C或均小于$-C$)，才能判定存在泄漏。如果能量迹的某个采样点处存在能量信息的泄漏，那么在两次独立测试实验中t值均应大于C或均小于$-C$；而如果检验统计量t在某采样点处因噪声等偶然因素超过了阈值，那么这种偶然情况在另一次独立实验中几乎不可能再次出现。

2.2 TVLA分类及对比

根据对能耗数据分组依据的不同，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进行分组，非特定TVLA根据输入明文或密钥对能耗数据进行分组。

特定TVLA选择密码算法中间值的某一位或几位作为分组判定位，根据判定位的值是否等于检测前设置的数值将能耗数据分为两组。特定TVLA判定位的设置和DPA等常见的能量分析攻击的形式较为相似，因此其针对这些常见攻击形式的测试效果较好。然而，由于可选择的密码算法中间值数目庞大，特定TVLA所需耗费的成本非常大。以AES-128算法为例，在仅考虑圈密钥加、字节置换、行移位和列混合4种密码操作的情况下，在第1轮加密中就可以进行4×128种位测试，4×16×256种字节测试，或更多种判定位数为其他值的测试。因此，特定TVLA同基于攻击的测试方法一样，难以保证评估的全面性。

根据输入明文的不同，非特定TVLA又可分为“固定-固定”和“固定-随机”两种，其中前者的两组输入明文(密钥)均为固定值，后者两组能耗对应的输入明文(密钥)分别为固定值和随机值。在“固定-固定”非特定TVLA中，评估人员事先选定两个固定的测试向量，并以随机交错的方式输入密码设备中进行加密，得到两组能耗数据。在“固定-随机”非特定TVLA中，评估人员事先选定一个固定测试向量，并以随机交错的方式依次将该固定和随机测试向量输入密码设备进行加密，得到两组能耗数据。在非特定TVLA中，选择不同的输入向量可能导致非特定TVLA结果不同，为保证评估结果的可靠性，应选择不同的测试向量重复实施TVLA。表1是特定和非特定TVLA的对比情况。

3 TVLA的局限性分析

TVLA的局限性主要体现在以下方面：

(1) TVLA只考虑了1阶原点矩(均值)和单变量(针对能量迹中的单个采样点)泄漏，其可能对存在高阶或多变量泄漏的待测设备作出“假阴性”误判[18,19]。

一方面，以2阶中心矩(方差)上存在泄漏的情况为例，设某采样点能耗X的均值为$\mathrm{E}(X)$，方差为$\mathrm{D}(X)$，分布律为$P\{X=x_k\}=p_k,k=1,2,\cdots$，有

$$\begin{aligned}\mathrm{D}(X)=&\mathrm{E}(X^2)-[\mathrm{E}(X)]^2\\=&\sum_{k=1}^{\infty}x_k^2p_k-[\mathrm{E}(X)]^2\end{aligned} \tag{2}$$

由式(2)可知，即使两个采样点处能耗数据的1阶矩相同，其2阶矩仍可能存在差异。结合式(1)分析可知，TVLA检验统计量t值的构造形式为能耗数据的1阶矩，因此当两组能耗数据间的统计差异只体现在高阶矩上时，TVLA无法探测到存在的泄漏。

表 1　特定和非特定TVLA的对比

	优势	不足
特定TVLA	针对DPA等常用攻击的测试效果较好	对能耗数据分组时需计算算法中间值；可供选择的算法中间值过多，难以保证测试的全面性
非特定TVLA	对能耗数据的分组较为简便，测试结果较为全面	所选择的测试向量对结果影响较大，需使用不同的测试向量重复实施评估

另一方面，以信息泄漏以积的形式分散在两个采样点上的情况为例，设对这两个采样点实施TVLA得到的能耗分组分别为(G_{X1},G_{X2})和(G_{Y1},G_{Y2})，积运算后对该“组合”采样点实施TVLA得到的两个能耗分组为(G_1,G_2)，则结合TVLA的分组原理可知

$$\begin{aligned}\mathrm{E}(G_1)=&\mathrm{E}(G_{X1}\cdot G_{Y1})\\=&\mathrm{E}(G_{X1})\cdot\mathrm{E}(G_{Y1})+\mathrm{Cov}(G_{X1},G_{Y1})\end{aligned}\tag{3}$$

$$\begin{aligned}\mathrm{E}(G_2)=&\mathrm{E}(G_{X2}\cdot G_{Y2})\\=&\mathrm{E}(G_{X2})\cdot\mathrm{E}(G_{Y2})+\mathrm{Cov}(G_{X2},G_{Y2})\end{aligned}\tag{4}$$

当$\mathrm{E}(G_{X1})=\mathrm{E}(G_{X2})$且$\mathrm{E}(G_{Y1})=\mathrm{E}(G_{Y2})$时，TVLA会判定不存在泄漏，而实际上泄漏是存在的：由式(3)和式(4)可知，由于$\mathrm{Cov}(G_{X1},G_{Y1})$与$\mathrm{Cov}(G_{X2},G_{Y2})$不一定相等，仍可能出现$\mathrm{E}(G_1)\neq\mathrm{E}(G_2)$。因此当对加掩码防护措施或对串行运行的软件密码产品实施泄漏检测时，各个共享因子产生泄漏的时刻可能不同，需要利用覆盖所有共享因子泄漏时刻的组合能耗来确定是否有泄漏，而TVLA只针对单个采样点的能耗，会导致漏检。

(2) 根据TVLA的t值仅能判断待测设备是否存在泄漏，而不能进一步解释泄漏的具体情况或量化能量信息泄漏的多少，对后续攻防的参考意义有限[20,21]。

完整的TVLA中，求得t值后，还需根据式(5)—式(7)分别求得自由度v，t分布的概率密度函数$f(t,v)$和t检验中零假设成立的概率p

$$v=\frac{\left({S_0}^2/n_0+{S_1}^2/n_1\right)^2}{\left({S_0}^2/n_0\right)^2/(n_0-1)+\left({S_1}^2/n_1\right)^2/(n_1-1)}\tag{5}$$

$$f(t,v)=\frac{\Gamma\left((v+1)/2\right)}{\sqrt{\pi v}\Gamma\left(v/2\right)}\left(1+t^2/v\right)^{-\frac{v+1}{2}}\tag{6}$$

$$p=2\int_{|t|}^{\infty}f(t,v)\,\mathrm{d}t\tag{7}$$

综合分析式(5)—式(7)以及TVLA的完整过程可知，t值和p值是相互对应的，t值越大或p值越小仅能表示零假设成立的概率越小。因此TVLA的结果并不能用于量化侧信道能量信息泄漏，检验统计量t值的大小不能代表能量信息泄漏的多少，其与能量分析攻击的成功率之间也无确定性关系。即使TVLA判定待测设备存在泄漏，能量分析攻击仍不一定能够攻破设备，导致将实际安全的密码设备判定为存在泄漏而造成资源浪费。

(3) TVLA对能耗数据的信噪比要求较高，噪声较大时TVLA评估效果会受到显著影响[18,22]。

密码设备的能耗依赖于其中执行的密码操作和处理的操作数，将能耗中的操作依赖分量记为P_{op}，数据依赖分量记为P_{da}；同时，能耗中不可避免地含有与执行的密码操作和处理的中间值无关的随机噪声P_{no}，以及由漏电流等产生的常量部分P_{co}。据此可用式(8)刻画密码设备的总能耗P_{to}

$$P_{\mathrm{to}}=P_{\mathrm{op}}+P_{\mathrm{da}}+P_{\mathrm{no}}+P_{\mathrm{co}}\tag{8}$$

TVLA中，评估人员对多条能量迹中相同采样点处的能耗值进行分析，相同采样点对应的密码操作相同，故各能耗值中的P_{op}相等。根据定义，各能耗值中的P_{co}同样相等。可知不同能耗值间的差异是由P_{da}和P_{no}造成的。P_{no}较大时会掩盖P_{da}中存在的统计差异导致漏检，因此实施TVLA前需进行对齐和降噪等预处理步骤。

(4) TVLA只关注了在能量迹的单个采样点处能耗值和设备中操作数的依赖关系，这导致TVLA整体犯误判错误的概率随能量迹中采样点数量的增加而变大[23,24]。

设能量迹中采样点数量为l，单次t检验犯假阳性误判错误的概率为α，则一次完整的TVLA犯假阳性误判错误的概率为$1-(1-\alpha)^l$，犯假阴性误判错误的情况同理。因此采样点数量l越大，TVLA犯误判错误的概率越大。当能量迹中的采样点数量非常庞大时，TVLA很难避免犯误判错误。

(5) TVLA仅将能耗数据分为两组，只能发现这两个能耗分组之间表现出的统计差异性，而无法检测到能耗数据更一般的分布差异，导致泄漏可能被隐藏在其中的一个分组中[19,24]。

在TVLA中选择不同的输入测试向量时，能耗数据的分组情况不同，当所选测试向量不正确时，存在差异的能耗数据可能被分至同一分组中，导致存在的统计差异被隐藏。

4 TVLA的优化研究

本节根据第2节所归纳的TVLA的各项局限性，对已有具代表性的各类TVLA的改进方法进行介绍和分析。

(1) 针对TVLA对高阶多变量的泄漏检测效果不佳的问题

文献[25]提出一种基于Hotelling's T^2检验的泄漏检测方法，Hotelling's T^2检验本质上是t检验由单变量向多变量的扩展形式，利用其代替Welch's t检验，能够显著提高泄漏信号分散在多个能量迹采样点时的泄漏检测率。然而由于需要对能耗数据的协方差矩阵进行求逆运算，Hotelling's T^2检验的计算复杂度随能量迹中采样点数量的增加呈指数增长，当能量迹中的采样点数量非常大时，Hotelling's

T^2检验在计算上可能具有不可实现性。因此该文中又对Hotelling's T^2检验进一步改进，提出对角检验以提升检测效率，对角检验的计算复杂度随采样点数量的增加呈线性增长，但其对多变量泄漏的检测性能低于Hotelling's T^2检验。

文献[26]通过增量算法将泄漏检测的统计矩由1阶扩展到了任意阶，由单变量扩展到了任意变量，使TVLA可以检测任意变量和任意阶的泄漏。增量算法使得泄漏检测过程能够同时进行能耗数据的采集和检验统计量的计算，当检验统计量超过阈值时即可终止整个评估过程，因此可以提升泄漏检测的效率，并减小评估所需成本。

文献[27]针对两组能耗数据在1阶矩上统计差异较小时TVLA存在漏检的问题，提出对两组能耗数据的1阶原点矩均值与2阶中心矩方差进行综合差异评估，当能耗数据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。由于一般的能量信息泄漏主要分布在能耗数据的低阶矩上，因此该方案能够有效控制TVLA犯假阴性误判错误的概率。

文献[28]提出一种基于KS检验的泄漏检测方法。该文证明了KS检验可作为一种基于信息论的泄漏检测工具对任意变量的泄漏进行检测，并提出了基于直方图的快速实现方法提升了KS检验的效率。与TVLA的对比实验结果显示，当实验参数等设置不同时，KS检验的鲁棒性明显强于TVLA。该文指出，KS检验可作为TVLA的补充方案对泄漏进行检测。

此外，文献[29]提出一种基于统计直方图的泄漏检测方法，该方法仅需在初始化时访问能耗数据集，对每个能量迹采样点建立各能耗分组的直方图，并在检测过程中不断对直方图进行更新。由于直方图携带了大量能耗分布的信息，因此该方法可直接对任意阶的统计矩和检验统计量等参数进行统计。由于无需重复访问整个能耗数据集，该方法能够有效提升泄漏检测过程中数据统计的效率。

(2) 针对检验统计量t值的参考意义有限的问题

文献[30]推导并证明了TVLA和能耗数据信噪比之间的等价关系，并参考文献[31]中的成果，用TVLA的结果估计了能量分析攻击成功率的理论界限，从而建立了TVLA结果、信噪比和能量分析攻击成功率之间的联系。由此，由TVLA结果即可得到相应的能量分析攻击的成功率。实验结果表明，这种端到端方法的预测结果与实际攻击结果能较好地吻合。

文献[32]将泄漏检测过程转换为建立一个依赖于密钥的回归模型的过程，当该模型能够对大部分的能耗数据作出解释时，说明能耗数据和密钥间存在依赖关系，此时应判定存在泄漏。同时，根据回归模型的具体情况即可量化泄漏的可利用性，并构造出攻击向量。该方法将泄漏检测的结果与密钥联系了起来，可以回答TVLA所检测出的泄漏是否可以利用的问题。

(3) 针对TVLA对能耗数据的信噪比依赖性较强，需进行降噪和对齐等预处理的问题

通过时域频域结合分析的方法能够减小能量迹在时域上未对齐对TVLA的影响，基于该原理，文献[22]提出先通过快速傅里叶变换将采集到的能耗数据从时域变换到频域进行频谱分析；然后对频谱的每个频率分量实施t检验评估是否存在泄漏。文献[33]指出，能耗中存在信息泄漏的部分是由设备所处理的数据和运行的密码操作引起的，而噪声部分是由环境和元器件间的相互影响等不确定因素引起的，因此将能耗信息由时域转换至频域时，泄漏分量和噪声分量分布在不同的频率分量上。据此，该文提出了多源时频融合信息泄漏检测方案，综合利用多个信道的时域信息及频域信息对泄漏进行检测。该方案无需对能耗数据进行对齐和降噪等预处理，可以提高泄漏检测效率，同时有利于发现单信道检测中的漏检。

文献[34]提出一种基于配对的t检验方案，对密码算法相邻的两次加密进行配对，由于实际中执行1次加密的时间非常短，可以近似地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对能耗的影响，进一步得到更加稳定的检验统计量，从而提高泄漏检测的准确性。文献[35]对配对t检验方案进行了优化研究，该文指出，文献[34]中的方法没有考虑到配对的能量迹组间采样点的相关性对配对t检验的影响，当两者间呈负相关关系时采取配对的方法不仅会导致检测效率下降，还可能导致犯误判错误。因此，应该先对该相关系数进行判断，当其大于0时方可实施配对t检验，否则只能实施Welch's t检验。

文献[36]提出将深度学习技术应用于泄漏评估中，通过有监督学习的方法用训练集构造出一个卷积神经网络，并将其作为待测数据集的区分器。当该区分器对待测能耗数据进行分组的正确率和随机地对待测能耗数据分组的正确率存在显著性区别时，说明训练过程获取了能量信息，故应判定存在泄漏。深度学习中卷积神经网络的特点使得评估人员不必考虑能量迹是否对齐和泄漏的统计矩阶数等

问题，且涵盖了多变量的泄漏情形，大大简化了泄漏检测的预处理步骤。但深度学习检测方法由于需要对神经网络进行训练，所需时间成本较大，训练过程中参数的设置对最终的泄漏检测效果影响较大，并且存在概率适应性和过拟合等问题。

(4) 针对TVLA犯误判错误的概率随能量迹中采样点数量增加而增大的问题

文献[37]中实施了一个“随机-随机”的TVLA实验，实验过程中将同一明文总体中的全部明文随机地分成两组并据此对采集的能耗数据进行分组，因此t检验的零假设是正确的。在能量迹中设置5×10^6万个采样点，多次重复迭代实施TVLA，得到的最大的t值的绝对值为5.608 8。考虑到该实验中采样点数量过大，该文提出将t值的阈值设置为5。但这种通过设置检验统计量阈值来控制误判错误的方法具有一定的片面性，大阈值会导致假阴性错误的概率增加，反之，小阈值会导致假阳性错误的概率增加。

文献[38]提出一种基于HC(Higher Criticism)检验的TVLA方案，对TVLA得到的各采样点处的p值实施HC检验，通过比较p值在无泄漏情况下预期的分布和实际检验得到的p值的分布之间的差异对泄漏情况进行判断，当该差异较为显著时即可判定存在泄漏。该方法综合利用能量迹中全部采样点处的能耗数据的统计特征，而非仅根据单个采样点的t值判定泄漏情况，能够有效控制TVLA犯误判错误的概率。

(5) 针对能量信息泄漏可能被隐藏在能耗数据的某个分组中的问题

文献[39]提出用卡方检验结合TVLA进行泄漏检测，卡方检验将泄漏检测自然而然地扩展到了多个能耗分组上，有效控制了因分组数过小造成的假阴性误判错误。同时卡方检验可以捕获多个统计矩中的泄漏，而非只关注某一个统计矩。然而，实验结果显示，当信噪比较低时，t检验的效果优于卡方检验的效果，因此该文提出卡方检验可以和TVLA技术结合使用以提高评估的准确性。

上述各TVLA改进方法的简要情况如表2。

5 总结与展望

侧信道能量分析攻击以其通用性强、计算成本低和成功率高等优点，目前已被广泛应用于对密码算法的破解中，各类密码产品面临着严峻的安全性挑战。因此，对密码产品的抗侧信道能量分析攻击能力进行评估已经成为密码设计过程中不可或缺的环节，该评估用于判断是否存在能量信息的泄漏，也可以对施加的防护措施的安全等级进行评定。评估可以通过实施攻击的方式来实现，也可通过统计测试的方式实现，由于具体攻击方法的种类繁多，攻击型评估难以保证全面性，因此统计测试方法已成为主流的评估形式。TVLA是目前为止最为常见的一种统计测试型评估方法，本文首先对TVLA的原理进行了分析，对其实现过程进行了介绍，然后

表 2 TVLA改进方法汇总表

所针对问题	对应文献	主要方法	优缺点(研究意义)
TVLA对高阶和多变量信息泄漏容易产生漏检	[25]	Hotelling's T^2检验	能够提高多变量泄漏的检出率，但计算复杂度高
	[26]	增量算法	适用于多变量和高阶泄漏，效率较高
	[27]	多分类F检验和Bartlett检验	2阶以内的泄漏检测准确率较高
	[28]	KS检验	鲁棒性较强
TVLA检验统计量t值的参考意义有限	[29]	统计直方图	效率较高，但初始化较繁琐
	[30]	理论推导和实验验证结合	建立了TVLA结果、信噪比和能量分析攻击成功率之间的联系
	[32]	回归模型	回答TVLA所检测出的泄漏是否可以利用的问题
TVLA对能耗数据的信噪比要求较高	[22]	快速傅里叶变换	减小了能量迹未对齐对TVLA结果的影响
	[33]	多源时频信息融合	避免了对齐和降噪的预处理步骤，检测效率和准确率较高
	[34]	配对t检验	统计结果较稳定准确
	[35]	相关关系	进一步优化了文献[34]中的方法
	[36]	深度学习	不必考虑能量迹是否对齐和泄漏的统计矩阶数等问题，且涵盖了多变量的泄漏情形；但所需时间成本较大，存在过拟合等问题
TVLA犯误判错误的概率随能量迹中采样点数量增加而增大	[37]	将t值的阈值设置为5	导致犯假阴性误判错误的概率增加
	[38]	HC检验	能够有效控制TVLA因仅依赖于单个采样点的t值而犯误判错误的概率
泄漏可能被隐藏在TVLA的某个分组中	[39]	卡方检验	可以和t检验结合使用以提高评估的准确性

对特定和非特定两种TVLA的优点和缺点分别进行了归纳和对比，接着根据目前已有研究中的观点对TVLA的局限性进行了剖析和分类，针对这些不同方面的局限，对各TVLA改进方法的优缺点分别进行了分析。

后期，TVLA可能向以下方向和领域发展延伸：

(1)更加广泛地用于对后量子密码算法的安全性进行评估。鉴于量子攻击的巨大潜在威胁，发展后量子密码体制机制的需求已十分迫切，TVLA可用于对后量子密码算法及其实现过程中的安全性进行评估。

(2)基于深度学习技术的TVLA不断发展，实现准确性更高，耗时更短的评估。泄漏检测的本质在于分类，深度学习技术一经提出就成为分类器的一个很好的选择。深度学习效果的好坏主要取决于神经网络的训练效果，而目前神经网络的训练过程迭代次数较多，所引入的参数较复杂，仍有对其进行优化的必要。

(3)与人工智能技术结合实现智能化泄漏评估。近年来人工智能技术的迅猛发展催化了包括密码学在内的各学术领域的深刻变革，将人工智能技术应用于密码产品的安全性评估中已是可以预见的发展趋势。

参 考 文 献

[1] KOCHER P, JAFFE J, and JUN B. Differential power analysis[C]. Proceedings of the 19th Annual International Cryptology Conference, Santa Barbara, USA, 1999: 388–397. doi: 10.1007/3-540-48405-1_25.

[2] RADHAKRISHNAN R A. Side-channel resistant implementation using arbiter PUF[J]. Cryptology ePrint Archive, 2023.

[3] 赵毅强, 王庆雅, 马浩诚, 等. 基于数据预处理的侧信道分析优化方法[J]. 电子与信息学报, 2023, 45(1): 49–58. doi: 10.11999/JEIT211462.
ZHAO Yiqiang, WANG Qingya, MA Haocheng, *et al*. Side channel analysis optimization method based on data preprocessing[J]. *Journal of Electronics & Information Technology*, 2023, 45(1): 49–58. doi: 10.11999/JEIT211462.

[4] BREUER R, STANDAERT F X, and LEVI I. Fully-digital randomization based side-channel security—toward ultra-low cost-per-security[J]. *IEEE Access*, 2022, 10: 68440–68449. doi: 10.1109/ACCESS.2022.3185995.

[5] PERIN G, WU Lichao, and PICEK S. Exploring feature selection scenarios for deep learning-based side-channel analysis[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2022, 2022(4): 828–861. doi: 10.46586/tches.v2022.i4.828-861.

[6] Common Criteria v3.1. Release 4[EB/OL]. https://www.commoncriteriaportal.org/cc/index.cfm?, 2013.

[7] 陈华, 习伟, 范丽敏, 等. 密码产品的侧信道分析与评估[J]. 电子与信息学报, 2020, 42(8): 1836–1845. doi: 10.11999/JEIT190853.
CHEN Hua, XI Wei, FAN Limin, *et al*. Side channel analysis and evaluation on cryptographic products[J]. *Journal of Electronics & Information Technology*, 2020, 42(8): 1836–1845. doi: 10.11999/JEIT190853.

[8] GOODWILL G, JUN B, JAFFE J, *et al*. A testing methodology for side-channel resistance validation[C]. NIST Non-Invasive Attack Testing Workshop, 2011: 115–136.

[9] COOPER J, DEMULDER E, GOODWILL G, *et al*. Test Vector Leakage Assessment (TVLA) methodology in practice[C]. International Cryptographic Module Conference, Shanghai, China, 2013.

[10] WANG L C, GOLDER A, FANG Yan, *et al*. Power side-channel leakage assessment of reference implementation of SABER key encapsulation mechanism[C]. 2022 Opportunity Research Scholars Symposium (ORSS), Atlanta, USA, 2022: 8–11. doi: 10.1109/ORSS55359.2022.9806031.

[11] SAARINEN M J O. WiP: Applicability of ISO standard side-channel leakage tests to NIST post-quantum cryptography[C]. 2022 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), McLean, USA, 2022: 69–72. doi: 10.1109/HOST54066.2022.9839849.

[12] KRAUSZ M, LAND G, RICHTER-BROCKMANN J, *et al*. Efficiently masking polynomial inversion at arbitrary order[C/OL]. The 13th International Conference on Post-Quantum Cryptography, 2022: 309–326. doi: 10.1007/978-3-031-17234-2_15.

[13] SADHUKHAN R, CHAKRABORTY A, DATTA N, *et al*. Light but tight: Lightweight composition of serialized s-boxes with diffusion layers for strong ciphers[C]. The 12th International Conference on Security, Privacy, and Applied Cryptography Engineering, Jaipur, India, 2022: 28–49. doi: 10.1007/978-3-031-22829-2_2.

[14] KHAIRALLAH M and BHASIN S. Hardware implementation of masked SKINNY SBox with application to AEAD[C]. The 12th International Conference on Security, Privacy, and Applied Cryptography Engineering, Jaipur, India, 2022: 50–69. doi: 10.1007/978-3-031-22829-2_3.

[15] DUAN Xiaoyi, HUANG Ye, SU Yonghua, *et al*. Research on the grouping method of side-channel leakage detection[C/OL]. The 18th International Conference on Security and Privacy in Communication Systems, 2023: 807–818. doi: 10.1007/978-3-031-25538-0_42.

[16] LU Chuanchao, CUI Yijun, KHALID A, *et al*. A novel

combined Correlation Power Analysis (CPA) attack on schoolbook polynomial multiplication in lattice-based cryptosystems[C]. 2022 IEEE 35th International System-on-Chip Conference (SOCC), Belfast, UK, 2022: 1–6. doi: 10.1109/SOCC56010.2022.9908076.

[17] WELCH B L. The generalization of 'STUDENT'S' problem when several different population varlances are involved[J]. *Biometrika*, 1947, 34(1/2): 28–35. doi: 10.1093/biomet/34.1-2.28.

[18] STANDAERT F X. How (not) to use welch's t-test in side-channel security evaluations[C]. The 17th International Conference on Smart Card Research and Advanced Applications, Montpellier, France, 2019: 65–79. doi: 10.1007/978-3-030-15462-2_5.

[19] WHITNALL C and OSWALD E. A cautionary note regarding the usage of leakage detection tests in security evaluation[J]. Cryptology ePrint Archive, 2019.

[20] ROY D B, BHASIN S, GUILLEY S, *et al*. CC meets FIPS: A hybrid test methodology for first order side channel analysis[J]. *IEEE Transactions on Computers*, 2018, 68(3): 347–361. doi: 10.1109/TC.2018.2875746.

[21] DURVAUX F and STANDAERT F X. From improved leakage detection to the detection of points of interests in leakage traces[C]. The 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, 2016: 240–262. doi: 10.1007/978-3-662-49890-3_10.

[22] LEI Wan, WANG Lihui, SHAN Weijun, *et al*. A frequency-based leakage assessment methodology for side-channel evaluations[C]. The 13th International Conference on Computational Intelligence and Security (CIS), Hong Kong, China, 2017: 590–593. doi: 10.1109/CIS.2017.00137.

[23] ZHANG Liwei. Statistics in side channel analysis-modeling, metric, leakage detection testing[D]. [Ph. D. dissertation], Northeastern University, 2017. doi: 10.17760/D20251582.

[24] WHITNALL C and OSWALD E. A critical analysis of ISO 17825 ('testing methods for the mitigation of non-invasive attack classes against cryptographic modules')[C]. The 25th International Conference on the Theory and Application of Cryptology and Information Security, Kobe, Japan, 2019: 256–284. doi: 10.1007/978-3-030-34618-8_9.

[25] BRONCHAIN O, SCHNEIDER T, and STANDAERT F X. Multi-tuple leakage detection and the dependent signal issue[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2019, 2019(2): 318–345. doi: 10.13154/tches.v2019.i2.318-345.

[26] SCHNEIDER T and MORADI A. Leakage assessment methodology: A clear roadmap for side-channel evaluations[C]. The 17th International Workshop on Cryptographic Hardware and Embedded Systems, Saint-Malo, France, 2015: 495–513. doi: 10.1007/978-3-662-48324-4_25.

[27] 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. doi: 10.11959/j.issn.1000-436x.2021235.
WANG Yaru and TANG Ming. Side channel leakage assessment with the Bartlett and multi-classes F-test[J]. *Journal on Communications*, 2021, 42(12): 35–43. doi: 10.11959/j.issn.1000-436x.2021235.

[28] ZHOU Xinping, QIAO Kexin, and OU Changhai. Leakage detection with Kolmogorov-Smirnov test[J]. Cryptology ePrint Archive, 2019.

[29] REPARAZ O, GIERLICHS B, and VERBAUWHEDE I. Fast leakage assessment[C]. The 19th International Conference on Cryptographic Hardware and Embedded Systems, Taipei, China, 2017: 387–399. doi: 10.1007/978-3-319-66787-4_19.

[30] ROY D B, BHASIN S, GUILLEY S, *et al*. Leak me if you can: Does TVLA reveal success rate?[J]. Cryptology ePrint Archive, 2016.

[31] FEI Yunsi, DING A A, LAO Jian, *et al*. A statistics-based success rate model for DPA and CPA[J]. *Journal of Cryptographic Engineering*, 2015, 5(4): 227–243. doi: 10.1007/s13389-015-0107-0.

[32] GAO Si and OSWALD E. A novel framework for explainable leakage assessment[J]. Cryptology ePrint Archive, 2022.

[33] 曹雨晨, 周永彬. 多源融合信息泄漏检测方法[J]. 信息安全学报, 2020, 5(6): 40–52. doi: 10.19363/J.cnki.cn10-1380/tn.2020.11.04.
CAO Yuchen and ZHOU Yongbin. Multi-channel fusion leakage detection[J]. *Journal of Cyber Security*, 2020, 5(6): 40–52. doi: 10.19363/J.cnki.cn10-1380/tn.2020.11.04.

[34] DING A A, CHEN Cong, and EISENBARTH T. Simpler, faster, and more robust t-test based leakage detection[C]. The 7th International Workshop on Constructive Side-Channel Analysis and Secure Design, Graz, Austria, 2016: 163–183. doi: 10.1007/978-3-319-43283-0_10.

[35] 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. doi: 10.3969/j.issn.1000-1220.2019.12.021.
LU Fuxiang, LI Weijian, and HUANG Xian. Research on optimization of side channel leakage assessment based on paired t test[J]. *Journal of Chinese Computer Systems*, 2019, 40(12): 2585–2590. doi: 10.3969/j.issn.1000-1220.2019.12.021.

[36] MOOS T, WEGENER F, and MORADI A. DL-LA: Deep learning leakage assessment: A modern roadmap for SCA

evaluations[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2021, 2021(3): 552–598. doi: 10.46586/tches.v2021.i3.552-598.

[37] BALASCH J, GIERLICHS B, GROSSO V, *et al*. On the cost of lazy engineering for masked software implementations[C]. The 13th International Conference on Smart Card Research and Advanced Applications, Paris, France, 2015: 64–81. doi: 10.1007/978-3-319-16763-3_5.

[38] DING A A, ZHANG Liwei, DURVAUX F, *et al*. Towards sound and optimal leakage detection procedure[C]. The 16th International Conference on Smart Card Research and Advanced Applications, Lugano, Switzerland, 2018: 105–122. doi: 10.1007/978-3-319-75208-2_7.

[39] MORADI A, RICHTER B, SCHNEIDER T, *et al*. Leakage detection with the x2-test[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2018, 2018(1): 209–237. doi: 10.13154/tches.v2018.i1.209-237.

郑　震：男，博士生，研究方向为侧信道安全防护.

严迎建：男，教授，博士生导师，研究方向为芯片安全防护、嵌入式密码系统.

刘燕江：男，博士，讲师，研究方向为芯片安全防护.

责任编辑：马秀强