基于 Bartlett 和多分类 F 检验侧信道泄露评估

王娅茹[1,2]，唐明[1,2]

（1. 武汉大学国家网络安全学院， 湖北 武汉 430072；2. 武汉大学空天信息安全与可信计算教育部重点实验室，湖北 武汉 430072）

摘 要：为了解决测试向量泄露评估（TVLA）技术进行侧信道泄露检测时，两组功耗样本（固定明文和随机明文）的均值差异较小时，t 检验存在漏检以及可能导致评估出现假阴性的问题。基于此，提出对样本的均值与方差等参数进行差异评估，进而提出基于 Bartlett 和多分类 F 检验侧信道泄露评估（Bartlett-F 检验）方法。在 Bartlett-F 检验中，将 Bartlett 检验用于均值差异小于方差差异的功耗样本以解决漏检问题，将多分类 F 检验用于均值差异大于方差差异的功耗样本以解决评估出现假阴性的问题。在检验中，若 *P* 值小于阈值，则有泄露。实验结果表明，当均值差异小于方差差异时，Bartlett 检验的 *P* 值小于阈值时所需样本量为 1.5×10^4，而 t 检验则需要更大的样本量。当方差差异小于均值差异时，t 检验的 *P* 值小于阈值时所需样本量为 2.0×10^2，而 F 检验所需样本量仅为 t 检验的 1/10 。因此，Bartlett-F 检验可以解决 TVLA 技术在泄露检测中存在的问题。

关键词：泄露检测；Bartlett 检验；F 检验；测试向量泄露评估技术；t 检验；侧信道攻击

中图分类号：TP309

文献标识码：A

DOI: 10.11959/j.issn.1000−436x.2021235

Side channel leakage assessment with the Bartlett and multi-classes F-test

WANG Yaru[1,2], TANG Ming[1,2]

1. School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China

2. Key Laboratory of Aerospace Information Security and Trusted Computing, Wuhan University Ministry of Education, Wuhan 430072, China

Abstract: In order to solve the problem that when test vector leakage assessment (TVLA) technology is used for side channel leakage detection, the mean difference between the two groups of power consumption samples (fixed plaintext and random plaintext) is small, and the t-test may miss detection and lead to false negative evaluation. The Bartlett and multi-classification F-test side channel leakage assessment method (Bartlett-F test) was proposed. In the Bartlett-F test, the Bartlett-test was used to the power samples with greater variance difference than mean difference to solve the problem of missing detection, and the multi-classification F-test was used to the power samples with greater mean difference than variance difference to solve the problem of false negative evaluation. In the test, there is leakage if the *P*-value is less than the threshold. The experimental results show that when the mean difference is less than the variance difference, the sample size required by Bartlett-test is 1.5×10^4 when the *P*-value of Bartlett-test is less than the threshold, while the sample size required by t-test is larger. When the variance difference is less than the mean difference, the sample size required by t-test is 2.0×10^2 when the *P*-value of t-test is less than the threshold, while the sample size required by F-test is only 1/10 of t-test. Therefore, Bartlett-F test can solve the problems of TVLA technology in leak detection.

Keywords: leakage detection, Bartlett-test, F-test, the test vector leakage assessment technology, t-test, side channel attack

收稿日期：2021−09−03；**修回日期**：2021−12−03

基金项目：国家自然科学基金资助项目（No.61972295）；武汉市科技项目应用基础前沿专项基金资助项目（No.2019010701011407）

Foundation Items: The National Natural Science Foundation of China (No.61972295), The Wuhan Science and Technology Project Application Foundation Frontier Special Project (No.2019010701011407)

1 引言

自 Kocher 等[1]指出可以通过侧信道能量分析揭示智能卡中的秘密信息以来，侧信道攻击（SCA, side channel attack）成为加密实现和设备安全的重要威胁之一。多年来，研究者专注于各种侧信道攻击方法和对应防护策略的研究[1-5]，以保护密码实现免受此类攻击。随着防护策略集成到密码实现中，评估密码实现对 SCA 的抵抗能力成为一个必须思考的问题。泄露检测是针对此问题的一种简单、有效的解决方法。泄露检测通过对黑盒模型下侧信道功耗样本是否包含泄露信息进行检测，初步评估密码实现的安全性。泄露检测原理是通过对功耗样本的统计分析，检验不同明文对应的功耗样本分布是否可区分。如果可以区分，则认为功耗样本包含秘密信息，即存在侧信道泄露；否则认为不存在侧信道泄露。

2011—2013 年，Goodwill 等[6-7]给出了测试向量泄露评估（TVLA, test vector leakage assessment）技术进行侧信道泄露检测。在 TVLA 技术中，首次将统计学中的 Welch's t 检验[8]用于泄露检测，以评估不同明文对应的功耗样本分布是否可区分。TVLA 技术提出后，侧信道方向的研究者们利用该方法，进行各种加密设备和实现的泄露检测和安全评估策略的相关研究[9-11]；为进一步提高侧信道安全评估效率，研究者开始对消除泄露检测中环境噪声影响的方法[12]和优化侧信道泄露检测流程[13-16]产生兴趣。随着研究的深入，泄露检测成为侧信道研究中一个重要方面。t 检验仍是目前泄露检测中最常用的检测方法[17-18]，它将明文分成两类：固定明文集和随机明文集。通过比较固定明文（和密钥）集与随机明文（和密钥）集在相同加密实现下，产生的功耗样本分布是否可区分，来判别是否存在侧信道泄露。一般认为如果功耗样本不可区分，即固定明文集和随机明文对应的功耗样本分布相同，则无侧信道泄露。该方法利用 t 检验评估待测设备特定采集的两组功耗曲线均值的差异性，从而判断是否有信息泄露并给出密码实现安全性的初步评估。文献[2,19-20]仔细讨论的 t 检验的主要优势在于：通过比较两类（固定明文集与随机明文集）功耗样本均值，可以将泄露检测问题简化为简单的统计估计。t 检验具有简单、高效的优点且只利用较少的算法实现知识，但是有 2 个缺点[21]：1) 该方法把评估曲线分为两组，而不是按照目标中间值的实际大小来分组，分类有限，可能会导致实际评估时出现假阴性；2) 检测样本的评估结果比较依赖统计距离。t 检验仅考虑固定明文集和随机明文集对应的功耗样本的均值是否区分，而功耗样本的不可区分性不仅要求均值不可区分还要求方差不可区分。本文发现当固定明文集和随机明文集对应的功耗样本分布均值差异越小，P 值越大；当样本均值差异小于 0.01 时，t 检验的 P 值大于阈值，t 检验会有漏检的风险。

基于上述问题，本文提出在泄露检测之前，对样本的均值、方差等参数差异情况进行评估，并根据参数差异度选择特定的统计学假设检验方法而不是按照 TVLA 技术对所有样本均采用 t 检验进行检测，完全不考虑样本分布参数的差异性。进而提出用 Bartlett 检验代替 t 检验对均值差异小于方差差异的功耗样本进行检测，以解决 t 检验可能出现的漏检。此外，Bartlett 检验将功耗样本按明文汉明权重（HW, Hamming weight）分成多类，可以解决分类有限的问题。对于均值差异大于方差差异的功耗样本，虽然依然可以使用 t 检验进行泄露检测，但基于 t 检验分类有限的问题，本文提出用多分类 F 检验代替 t 检验，进而提出基于 Bartlett 和多分类 F 检验侧信道泄露评估（Bartlett-F 检验）方法，旨在解决传统 TVLA 技术中存在的问题。当接受假设 H_0 的概率（P 值）小于阈值 1.0×10^{-5}，则功耗样本有泄露。实验结果表明，对于均值差异（小于 0.01）小于方差差异（大于 0.5）的样本，Bartlett 检验的 P 值小于阈值 1.0×10^{-5} 所需样本量为 1.5×10^{4}。而对 t 检验而言，样本量达到 3.0×10^{4} 时，P 值大于阈值。为使 t 检验的 P 值小于阈值，则需要更大的样本量。因此，在相同样本量下，与 Bartlett 检验相比，t 检验存在漏检。对于方差差异（小于 0.3）小于均值差异（小于 1.0）的样本，P 值小于阈值的样本量为 2.0×10^{2}，而 F 检验扩大了样本分类，P 值小于阈值的样本量仅为 t 检验的 1/10。因此，本文的 Bartlett-F 检验可以解决传统 TVLA 技术在泄露检测中存在的问题。

本文的主要工作如下。

1) 本文发现利用 t 检验进行泄露检测，当样本量相同时，均值差异越小，P 值越大；当两组样本

均值差异小于0.01时，t 检验的 P 值大于阈值，存在漏检风险。

2) 基于 1)发现，本文提出在泄露检测前，添加对样本分布参数（均值和方差）的差异度评估。

3) 根据2)中的评估结果，本文提出采用Bartlett检验代替 t 检验解决 1) 的问题，并通过比较实验验证 Bartlett 检验的有效性。

4) 基于 t 检验分类有限可能导致的假阴性，本文提出用多分类 F 检验代替 t 检验对均值差异大于方差差异的功耗样本进行泄露检测，以降低 t 检验因分类有限导致的假阴性。

2 Welch's t 检验

TVLA 技术利用统计学中的 Welch's t 检验[17-22]来评估加密实现抵抗侧信道攻击的能力。t 检验对随机明文集和固定明文集对应功耗样本的均值差异进行评估，以判定加密实现或设备是否存在信息泄露。t 检验的具体内容如下。

L_0 和 L_1 分别表示固定明文集和随机明文集对应的侧信道功耗样本，L_0 和 L_1 的样本量、样本均值和样本方差分别为 $(n_0,\overline{x_0},s_0^2)$ 和 $(n_1,\overline{x_1},s_1^2)$。假设 H_0 为固定明文集和随机明文集对应的功耗样本均值不存在差异。检验统计量 t 和自由度 v 的计算式分别为

$$t=\frac{(\overline{x_0}-\overline{\mu_1})}{\sqrt{\frac{s_0^2}{n_0}+\frac{s_1^2}{n_1}}},\quad v=\frac{\left(\frac{s_0^2}{n_0}+\frac{s_1^2}{n_1}\right)^2}{\frac{\left(\frac{s_0^2}{n_0}\right)^2}{(n_0-1)}+\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)}} \tag{1}$$

概率密度函数及接受假设 H_0 的概率 P 分别为

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

$$P=2\int_{|t|}^{\infty} f(t,v)\mathrm{d}t \tag{3}$$

其中，$\Gamma(\bullet)$ 是 gamma 函数。

t 检验通常设定 4.5 [23-24]为判定接受或拒绝假设 H_0 的阈值，将 t 检验的统计量 $|t|$ 与阈值 4.5 进行比较。如果 $|t|>4.5$，则拒绝假设 H_0。这是因为当 $v>1\,000$，$P=2\mathrm{tcdf}(4.5,v)<1.0\times10^{-5}$ [20]时，这意味着以小于 0.000 01 的概率接受假设 H_0，即以大于 0.999 99 的概率拒绝假设 H_0。本文将 1.0×10^{-5} 作为阈值，若概率 $P<1.0\times10^{-5}$，则拒绝假设 H_0，表明加密实现或设备有侧信道泄露。

本文根据明文汉明权重对功耗样本进行分类，针对不同的输入类 $I_i\in I$ 进行加密操作，其中 I 是所有可能的输入集合，采集不同输入类 I_i 对应的功耗泄露记为 L_i $(1\leqslant i\leqslant m)$，样本分类数为 m，所有功耗样本集合为 L，总样本量为 N，每类功耗样本的样本均值为 $\overline{x_i}$、样本方差为 s_i^2、均值为 μ_i、方差为 δ_i^2，其中 μ_i、δ_i^2 均未知。

3 关键技术

3.1 问题描述

t 检验一般用于检测两类样本的均值是否存在差异。本文发现在侧信道泄露检测中，检测结果与固定明文集和随机明文集对应功耗样本的均值差异相关。当均值差异较大时，t 检验检测效果较好；当两样本均值相同、方差不同时，统计学判定两样本服从不同分布，而 t 检验结果显示两样本服从相同分布。因此，利用 t 检验对均值近似相同、方差不同的样本进行泄露检测时，存在漏检（加密算法或设备事实上存在泄露的，t 检验未发现泄露）风险。

本文利用常用的实验软件平台 MATLAB 按照 $L_{t_i}=\mathrm{HW}(M_i)+N_{0,\delta}$ 分别仿真侧信道功耗样本，其中，$\mathrm{HW}(M_i)$ 表示明文 M_i 的汉明权重，$N_{0,\delta}$ 表示高斯噪声且信噪比 $\mathrm{SNR}=1.0$ dB。分别仿真方差相同、均值差的绝对值小于 0.01、0.1、0.2 和等于 1 的侧信道功耗样本，如图 1(a)所示。

利用 t 检验分别对该样本进行泄露检测，检测结果如图 1(b)所示。由图 1(b)可知，当样本量相同时，均值绝对值差越小，P 值越大；当样本均值差的绝对值小于 0.01 时，t 检验的 P 值大于阈值，存在漏检风险。为避免出现此种漏检，本文提出在泄露检测前对样本参数差异进行评估。

3.2 样本参数差异评估

本文以每类样本参数差异的绝对值来评估样本参数差异。在 t 检验中，固定明文集和随机明文集对应的功耗样本分别为 L_0 和 L_1，L_0 和 L_1 的样本量、样本均值和方差分别为 $(n_0,\overline{x_0},s_0^2)$ 和 $(n_1,\overline{x_1},s_1^2)$。当将功耗 L 样本分成 m 类样本集 L_i $(1\leqslant i\leqslant m)$ 时，L_i $(1\leqslant i\leqslant m)$ 的样本量、样本均值和样本方差为 $(n_i,\overline{x_i},s_i^2)$。

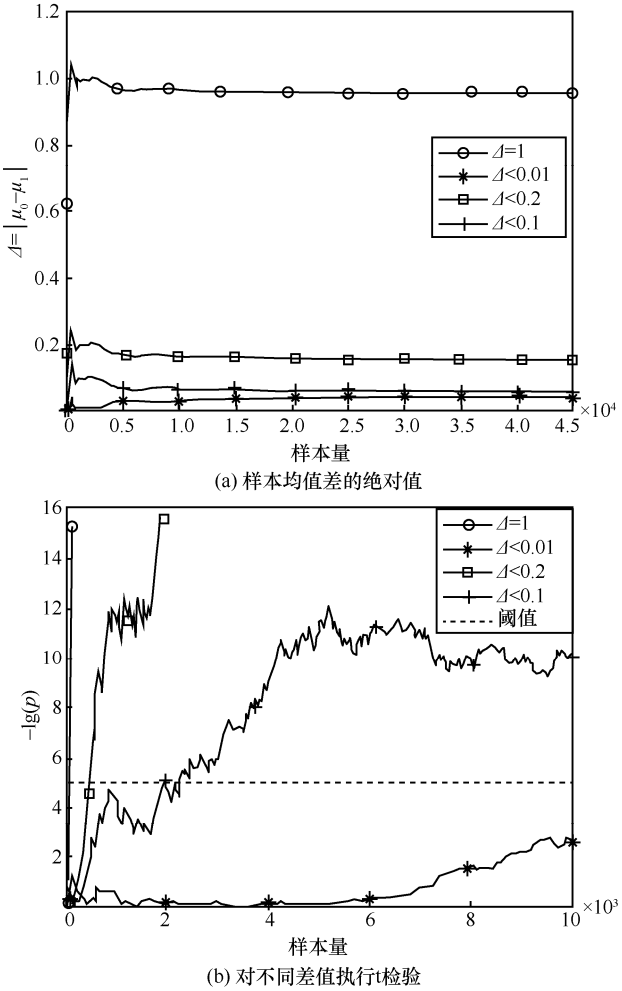


图 1 不同均值差下执行 t 检验

如果 $d_1=\max\left|\overline{x_i}-\overline{x_j}\right|$， $d_2=\max\left|s_i^2-s_j^2\right|$，（其中 $i,j\in[1,m]$ 且 $i\neq j$），若 $|d_1-d_2|$ 趋于 d_1，即 d_2 趋于 0，则两样本方差近似相等，均值差异大于方差差异；若 $|d_1-d_2|$ 趋于 d_2，即 d_1 趋于 0，则两样本均值近似相等，方差差异大于均值差异；若 d_1=d_2，则均值差异与方差差异相同。

3.3 Bartlett-F 检验

本节提出了基于 Bartlett 和多分类 F 检验侧信道泄露检测。Bartlett -F 检验流程如图 2 所示。

Bartlett -F 检验过程如下。

1) 将明文 M_i 按汉明权重分成 m 类，对应功耗曲线记为 L_i。假设每类检测样本集的样本量均为 n（n 可以根据实验需要进行调整）。

2) 从分类中随机选择 j ($1\leqslant j\leqslant m$) 类功耗样本 L_i，计算 L_i 的样本均值和样本方差记为 $(\overline{x_i},s_i^2)$。

3) 对 $|d_1-d_2|$ 进行评估，如果 $|d_1-d_2|$ 趋于 d_2，即 d_1 趋于 0，则样本间方差差异大于均值差异；如果 $|d_1-d_2|$ 趋于 d_1，则均值差异大于方差差异。

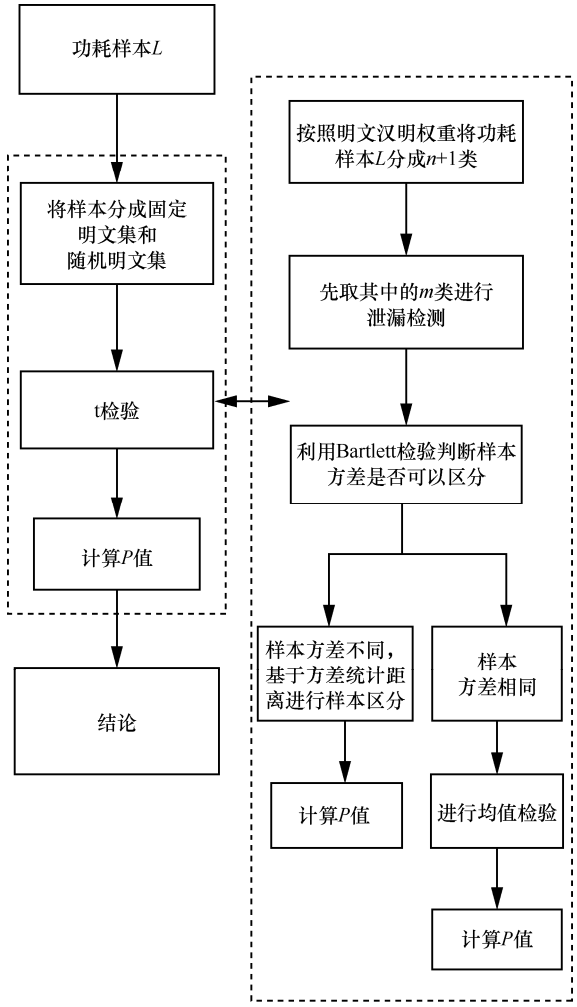


图 2 Bartlett -F 检验流程

4) 如果样本间均值差异小于方差差异，则使用 Bartlett 检验代替 t 检验；如果样本间均值差异大于方差差异，使用多分类 F 检验代替 t 检验。

3.3.1 Bartlett 检验

如果 $|d_1-d_2|$ 趋于 d_2，即 d_1 趋于 0，则两样本均值近似相等，样本间均值差异小于方差差异。利用 Bartlett 检验对样本的方差是否可区分进行检验。假设 H_0 为各样本类的方差相同，则统计量 X^2 和自由度 v 计算式分别为

$$X^2=\frac{\sum_{i=1}^{m}(n_i-1)\ln S^2-\sum_{i=1}^{m}(n_i-1)\ln s_i^2}{C} \tag{4}$$

$$s_i^2=\frac{1}{(n_i-1)}\sum_{j=1}^{n_i}(x_{i,j}-\overline{x_i})^2,i=1,2,\cdots,m \tag{5}$$

$$S^2=\frac{\sum_{i=1}^{m}(n_i-1)S_i^2}{\sum_{i=1}^{m}(n_i-1)} \tag{6}$$

$$C=1+\frac{1}{3(m-1)}\left[\sum_{i=1}^{m}\frac{1}{n_i-1}-\frac{1}{\sum_{i=1}^{m}(n_i-1)}\right] \tag{7}$$

其中，S_i^2 表示样本方差，n_i 表示样本量，$\overline{x_i}$ 表示第 i 类样本均值，$x_{i,j}$ 表示功耗值，i 和 m 分别表示样本类别和类别数。

由于自由度为 $v=m-1$，接受假设 H_0 的概率 $P_0=2\int_{X^2}^{\infty}f(x,v)\mathrm{d}x$，$f(x,v)=\frac{x^{\frac{v}{2}-1}\mathrm{e}^{\frac{-x}{2}}}{\mathrm{e}^{\frac{v}{2}}\Gamma(\frac{v}{2})}$，$v$ 表示自由度，$\Gamma(\bullet)$ 表示 gamma 函数。统计量 X^2 越大，P_0 越小，则以较大概率拒绝 H_0。

3.3.2　多分类 F 检验

如果 $|d_1-d_2|$ 趋于 d_1，即 d_2 趋于 0，则两样本方差近似相等，样本间方差差异小于均值差异。通过多分类 F 检验代替 t 检验对样本的均值是否可区分进行检验。根据明文汉明权重将采集到的功耗分成 m 类，记为 L_i $(1\leqslant i\leqslant m)$，其中每个分类的样本量为 n_i。假设 H_0 为各样本类的均值相同，则统计量 F 为

$$F=\frac{\sum_{i=1}^{m}n_i(\overline{x_i}-\overline{X})^2(N-m)S_i^2}{\sum_{i=1}^{m}\sum_{j}^{n_i}(x_{i,j}-\overline{x_i})^2(m-1)} \tag{8}$$

其中，$\overline{X}$ 为总样本均值，N 为总样本量，n_i 为第 i 类样本的样本量，$\overline{x_i}$ 为第 i 类样本均值，m 为分类数，则接受假设 H_0 的概率 $P_0=\int_{|F|}^{\infty}f(x,v)\mathrm{d}x$。

$$f(v_1,v_2,x)=\frac{\Gamma\left(\frac{v_1+v_2}{2}\right)\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}x^{\frac{v_1}{2}-1}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)\left(1+\frac{xv_1}{v_2}\right)^{\frac{v_1+v_2}{2}}} \tag{9}$$

其中，$v_1=m-1$、$v_2=N-m$ 表示自由度。接受假设 H_0 的概率 P_0 越小，拒绝 H_0 的证据越充分。

此外，设检测有泄露概率均为 P，t 检验仅将功耗样本分为两类，对两类样本均值进行比较，整个检测过程只进行一次比较判定。因此，假阴性的概率为 P。而 Bartlett 检验和多分类 F 检验将按明文汉明权重（也可以按照其他分类标准分类，如汉明距离），将功耗样本按明文汉明权重分成 m 类，比较各类对应的功耗样本的均值或方差是否相同，需要进行 $m-1$ 次判定，假阴性概率为 P^{m-1}。与 t 检验相比，假阴性概率为 t 检验的 1%。因此，扩大分类可以降低 t 检验由于分类有限导致的假阴性。

3.4　推广到多变量

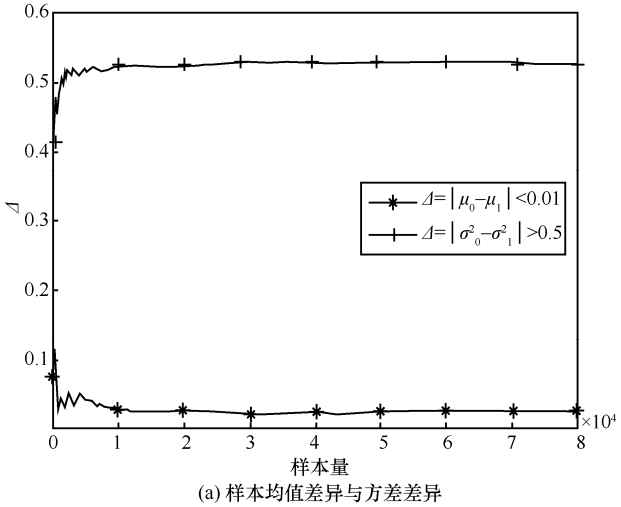
随着高阶掩码应用到软件实现和硬件设计中，每个共享因子泄露产生的时刻不同。所以进行泄露检测时，不能再通过对单个时刻的功耗进行检验来确定是否有泄露，需要利用覆盖所有共享因子泄露时刻的组合功耗来确定是否有泄露。常用的方案是利用组合函数对所有泄露时刻的功耗进行预处理，以获取组合功耗，然后再对其进行泄露检测。文献[25]已证明 $L'=\prod_{i=1}^{d}(L_{t_i}-\mu_{t_i})$ 是汉明权重模型下最优的组合函数，其中 L_{t_i} 表示 t_i 时刻的功耗，μ_{t_i} 表示该时刻的功耗样本均值，d 表示共享因子的数量。本文用此函数对功耗样本进行预处理。

4　实验及结果分析

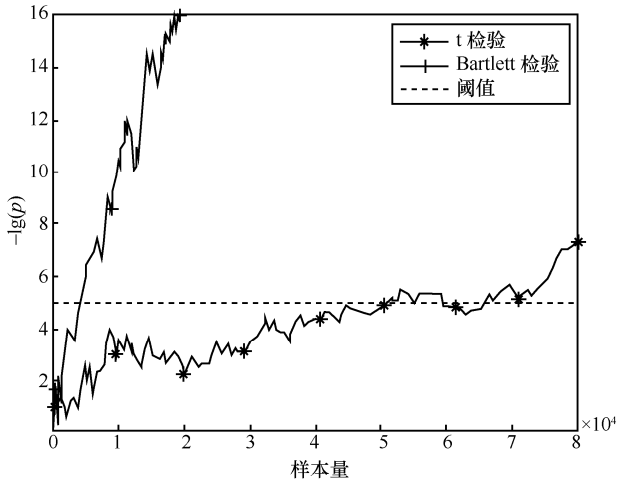
4.1　仿真

本文利用 MATLAB 软件平台按照 $L_{t_i}=\mathrm{HW}(M_i)+N_{0,\delta}$ 仿真侧信道功耗，其中，$\mathrm{HW}(M_i)$ 表示明文 M_i 的汉明权重，$N_{0,\delta}$ 表示高斯噪声且信噪比 SNR = 1.0 dB。由图 3(a)可知，样本均值差异小于 0.01，方差差异大于 0.5。根据差异度的描述，可得 $|d_1-d_2|$ 趋于 d_2，即 d_1 趋于 0，则两样本均值近似相等，样本间均值差异小于方差差异。利用 t 检验和 Bartlett 检验对该样本进行泄露检测，检测结果如图 3(b)所示。

由图 3(b)可知，当样本量为 0.5×10^4 时，Bartlett 检验接受假设 H_0 概率 P 小于阈值，且样本量为 2.0×10^4 时，P 值为 1.0×10^{-16}，远小于阈值。而对 t 检验而言，当样本量大于 5.0×10^4 时，概率 P 小于阈值。由于 $|d_1-d_2|$ 趋于 d_2，即 d_1 趋于 0，则两样本均值近似相等，样本间均值差异小于方差差异。因此，在相同样本量下，与 Bartlett 检验相比，用 t 检验对均值差异小于方差差异的功耗样本进行泄露检测，会存在漏检风险。



(a) 样本均值差异与方差差异



(b) 对样本执行Bartlett检验和t检验

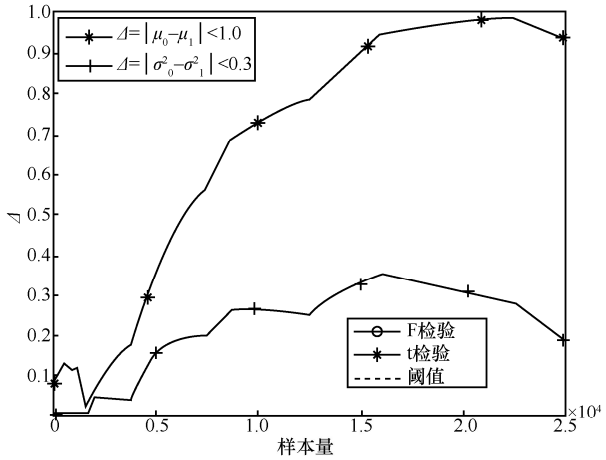
图 3　对均值差异小于方差差异的样本执行 Bartlett 检验和 t 检验

由图 4(a)可知，样本均值差异接近 1，方差差异小于 0.3。根据 3.2 节差异度大小的描述可知，如果$|d_1-d_2|$趋近d_1，则该样本均值差异大于均方差差异。利用多样本 F 检验和 t 检验分别对该样本进行泄露检测，检测结果如图 4(b)所示。

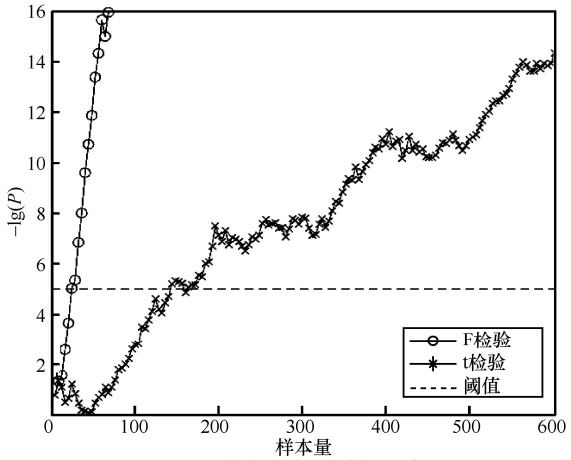
由图 4(b)可知，t 检验的 P 值小于阈值所需的样本量2.0×10^2，而多样本 F 检验所需的样本量是 t 检验的1/10。因此，与 t 检验相比，扩大分类的 F 检能降低因 t 检验分类有限导致的假阴性。

泄露检测时，检测结果与信噪比、样本分类等因素密切相关。在检测时应该考虑这些因素的影响。为了便于控制影响因素的大小，本文利用仿真功耗采样来研究信噪比、样本分类等对检测结果的影响。

为研究扩大样本分类对检测结果的影响，分别将样本分为四类、五类和六类，利用 Bartlett 检验对分类数$m=4$、利$m=5$和$m=6$的情况进行检测，检测结果如图 5 所示。



(a)样本均值差异与方差差异



(b) 对样本执行F检验和t检验

图 4　对均值差异大于方差差异的样本执行 F 检验和 t 检验

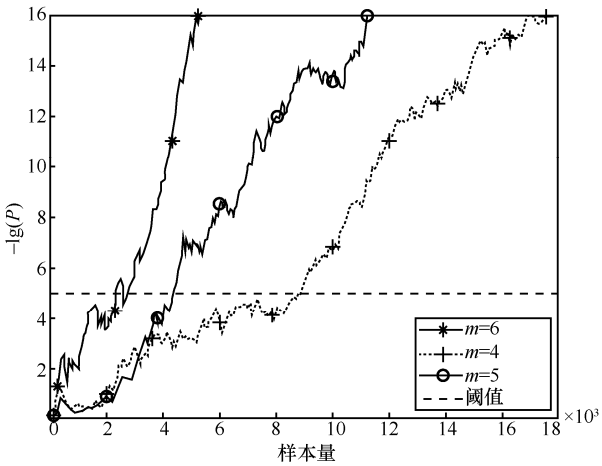


图 5　对分类数不同的样本执行 Bartlett 检验

由图 5 可知，当$m=6$时，Bartlett 检验的 P 值小于阈值的样本量2.5×10^3；当$m=4$和$m=5$时，P 值小于阈值所需样本量分别为8.5×10^3和4.0×10^3。实验结果表明，随着样本分类的增加，Bartlett 检验的 P 值小于阈值所需的样本量降低。

本文利用 Bartlett 检验分别对不同信噪比 SNR = 0.1 dB、SNR = 1.0 dB 和 SNR = 10.0 dB 环境下的功耗样本进行检测，检测结果如图 6 所示。

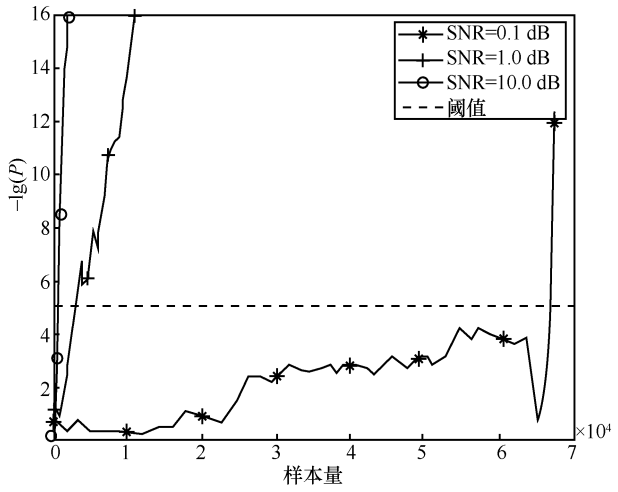


图 6　对不同信噪比环境下的样本执行 Bartlett 检验

由图 6 可知，当信噪比 SNR = 0.1 dB 时，P 值小于阈值所需的检测样本量为 6.8×10^4；SNR = 1.0 dB 时，所需的样本量为 6.0×10^4；SNR = 10.0 dB 时，所需的样本量仅为 2.0×10^3。实验结果表明，在 Bartlett 检验中，SNR=1.0 dB 时，P 值小于阈值所需的样本量大约为 SNR=10.0 dB 时的 34 倍。因此，SNR 的减小降低 Bartlett 检验优势。

4.2　实验

本文还利用公开数据集技术 DPA Contest-V4[26]，按照 Bartlett-F 检验和 TVLA 技术的泄露检测过程进行实验。实验前，首先对数据进行预处理；在实验中，利用峰值提取技术和 CPA 技术，提取出泄露产生时刻。由于 DPA Contest-V4 数据集是 AES 在 RSM 掩码实现下的功耗采集，泄露不在同一时刻发生。因此，在实验中选取 $L^{'}=\prod_{i=1}^{d}(L_{t_i}-\mu_{t_i})$ 作为预处理函数，对功耗样本进行预处理。相关系数与采样时刻如图 7 所示。

随后对预处理后的功耗样本进行均值差异和方差差异评估，如图 8(b)所示。由图 8(a)可知，样本绝对值差趋于 0 且样本均值差异小于方差差异。利用 Bartlett 检验和 t 检验分别进行泄露检测，检测结果如图 8(b)所示。

由图 8(b)可知，Bartlett 检验的 P 值小于阈值所需的样本量 1.5×10^4；当样本量达到 3.0×10^4，P 值小于阈值 1.0×10^{-5}，而对 t 检验而言，样本量达到 3.0×10^4 时，P 值大于阈值。为使 t 检验的 P 值小于阈值，则需要更大的样本量。因此在相同样本量下，与 Bartlett 检验相比，t 检验存在漏检。

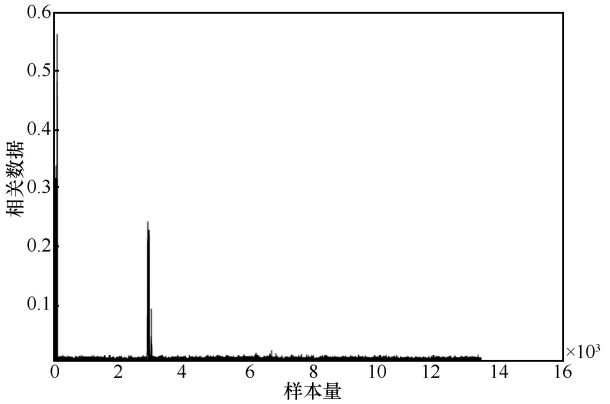
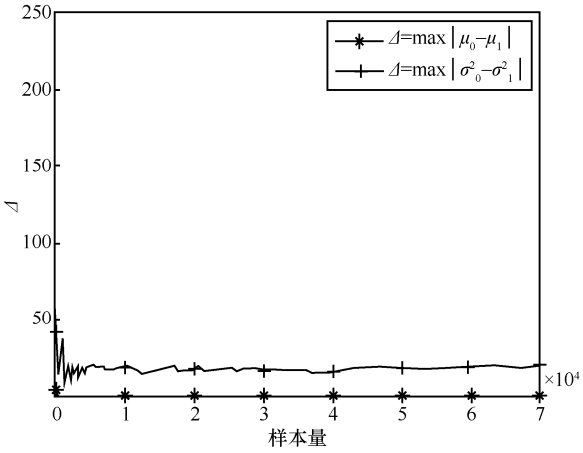
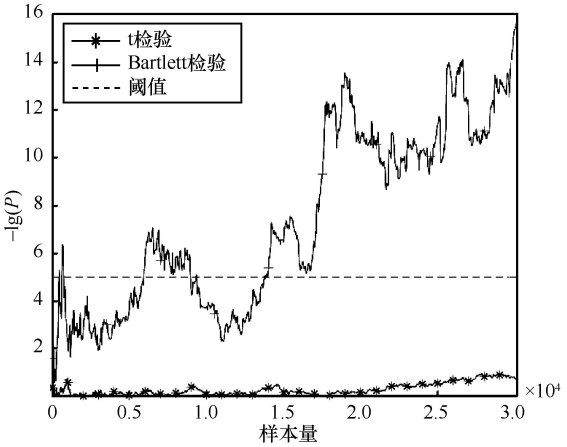


图 7　相关系数与采样时刻



(a) 样本均值差异与方差差异



(b) 对DPA Contest-v4执行Bartlett检验和t检验

图 8　执行 Bartlett 检验和 t 检验

5　结束语

传统 TVLA 技术通过比较固定明文集和随机明文集对应的功耗样本的均值是否存在差异，来

判定样本是否服从相同分布，以说明加密实现是否存在泄露。而事实上，样本服从相同分布，不仅要求样本均值而且要求样本方差均相同。本文发现，当固定明文集和随机明文集对应的功耗样本量相同时，均值差异越小，P 值越大；且均值差异小于 0.01 时，t 检验的 P 值大于阈值，t 检验会有漏检的风险。另外，t 检验将功耗分为两组，而不是按照中间值的大小来分组，样本分类有限，可能会导致评估时出现假阴性。基于上述问题，本文提出在泄露检测前，对样本的均值和方差等参数差异进行评估，并根据参数差异选择检验方法，提出将 Bartlett 检验代替 t 检验用于各样本类的均值差异小于方差差异的功耗样本的泄露检测，以解决 t 检验在检测中存在的漏检风险，同时 Bartlett 检验将功耗样本按明文汉明权重分成多类，可以解决分类有限的问题。此外，对均值差异大于方差差异的功耗样本，本文建议用多分类 F 检验代替 t 检验，扩大样本分类，提出 Bartlett F 检验来解决传统 TVLA 技术中存在的问题。实验结果表明，Bartlett 检验的 P 值小于阈值的样本量，仅为 1.5×10^4。当样本量达到 3.0×10^4 时，P 值小于阈值 1.0×10^{-5}。而对 t 检验而言，当样本量达到 3.0×10^4 时，P 值大于阈值。在相同样本量下，Bartlett 检验可以解决 t 检验出现的漏检。对于方差差异小于均值差异的样本，t 检验的 P 值小于阈值所需的样本量，为 2.0×10^2，而多样本 F 检验所需的样本量是 t 检验的 1/10。当样本量为 2.0×10^2 时，多样本 F 检验的 P 值大于 1.0×10^{-16} 而 t 检验的 P 值仅为 1.0×10^{-5}，多样本 F 检验可以降低 t 检验假阴性概率。因此，在侧信道泄露检测时，本文提出的 Bartlett-F 检验可以有效解决 TVLA 技术存在的问题。

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] MATHER L, OSWALD E, BANDENBURG J, et al. Does my device leak information? an a priori statistical power analysis of leakage detection tests[C]//Advances in Cryptology — ASIACRYPT 2013. Berlin: Springer, 2013: 486-505.

[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] CHARI S, RAO J R, ROHATGI P. Template attacks[C]//Cryptographic Hardware and Embedded Systems — CHES 2002. Berlin: Springer, 2003: 13-28.

[5] SCHRAMM K, WOLLINGER T, PAAR C. A new class of collision attacks and its application to DES[C]//Fast Software Encryption. Berlin: Springer, 2003: 206-222.

[6] 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, 7: 115-136.

[7] BECKER G，COOPER J，DEMULDER E, et al. Test Vector Leakage Assessment (TVLA) methodology in practice[C]//International Cryptographic Module Conference. [S.l.:s.n.], 2013: 20.

[8] WELCH B L. The generalization of 'student's' problem when several different population variances are involved[J]. Biometrika, 1947, 34(1/2): 28.

[9] AZOUAOUI M, BELLIZIA D, BUHAN I, et al. A systematic appraisal of side channel evaluation strategies[C]//Security Standardisation Research. Berlin: Springer, 2020: 46-66.

[10] YU H, HE Z H, WU L J, et al. Power leakage detection for a masked SM3-MAC hardware implementation[C]//Proceedings of 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID). Piscataway: IEEE Press, 2019: 224-228.

[11] GUILLEY S, KARRAY K, PERIANIN T, et al. Side-channel evaluation methodology on software[J]. Cryptography, 2020, 4(4): 27.

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

[13] YANG W, JIA A N. Side-channel leakage detection with one-way analysis of variance[J]. Security and Communication Networks, 2021, 2021: 1-13.

[14] WHITNALL C, OSWALD E. A cautionary note regarding the usage of leakage detection tests in security evaluation[R]. 2019.

[15] MERINO D P S, STANDAERT F X. Getting the most out of leakage detection[C]//Constructive Side-Channel Analysis and Secure Design. Berlin: Springer, 2017: 264-281.

[16] WHITNALL C, OSWALD E. A critical analysis of ISO 17825 ('testing methods for the mitigation of non-invasive attack classes against cryptographic modules')[C]//Lecture Notes in Computer Science. Berlin: Springer, 2019: 256-284.

[17] 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.

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

[19] SCHNEIDER T, MORADI A. Leakage assessment methodology[J]. Journal of Cryptographic Engineering, 2016, 6(2): 85-99.

[20] DURVAUX F, STANDAERT F X. From improved leakage detection to the detection of points of interests in leakage traces[C]//Advances in Cryptology — EUROCRYPT 2016. Berlin: Springer, 2016: 240-262.

[21] CHEN H, XI W, FAN L M, et al. Side Channel Analysis and Evaluation on Cryptographic Products[J]. Journal of Electronics and Information Technology, 2020, 42(8): 1836-1845.

[22] LEI W, WANG L H, SHAN W J, et al. A frequency-based leakage assessment methodology for side-channel evaluations[C]//Proceedings of 2017 13th International Conference on Computational Intelligence and Security (CIS). Piscataway: IEEE Press, 2017: 590-593.

[23] BILGIN B, GIERLICHS B, NIKOVA S, et al. Higher-order threshold implementations[C]//Lecture Notes in Computer Science. Berlin: Springer, 2014: 326-343.

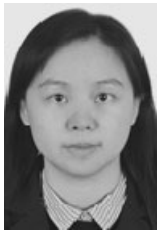
[24] DE CNUDDE T, BILGIN B, REPARAZ O, et al. Higher-order threshold implementation of the AES S-box[C]//Smart Card Research and Advanced Applications. Berlin: Springer, 2016: 259-272.

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

[26] BHASIN S, BRUNEAU N, DANGER J L, et al. Analysis and improvements of the DPA contest v4 implementation[C]//Security, Privacy, and Applied Cryptography Engineering. Cham: Springer International Publishing, 2014: 201-218.

[作者简介]

王娅茹（1989− ），女，河南驻马店人，武汉大学博士生，主要研究方向为侧信道泄露检测、密码学、信息安全等。



唐明（1976− ），女，湖北武汉人，博士，武汉大学教授、博士生导师，主要研究方向为 CPU 安全、信息安全、侧信道攻击与检测等。