doi:10.3969/j.issn.1000-436x.2013.04.014

基于投影寻踪分析的芯片硬件木马检测

张鹏，王新成，周庆
（信息保障技术重点实验室，北京 100072）

摘 要：提出一种利用芯片旁路泄漏信息的硬件木马无损检测方法，通过基于绝对信息散度指标的投影寻踪技术，将芯片运行过程中产生的高维旁路信号投影变换到低维子空间，在信息损失尽量小的前提下发现原始数据中的分布特征，从而实现芯片旁路信号特征提取与识别。针对示例性高级加密标准（AES-128）木马电路的检测实验表明，该技术可以有效分辨基准芯片与硬件木马测试芯片之间的旁路信号特征差异，实现硬件木马检测。
关键词：集成电路；硬件木马；旁路分析；投影寻踪；木马检测
中图分类号：TN918 **文献标识码**：B **文章编号**：1000-436X(2013)04-0122-05

Hardware Trojans detection based on projection pursuit

ZHANG Peng, WANG Xin-cheng, ZHOU Qing
(Science and Technology on Information Assurance Laboratory, Beijing 100072, China)

Abstract: A novel hardware Trojans detection technique using the side channel signals of chips was proposed. Based on the projection pursuit with absolute information divergence index, this technique could find out the data structure enables reflect high dimension special rules without obvious information loss, so as to attain the goal of feature abstraction and identification on side channel signals of IC chips. The detection experiment against an exemplary AES-128 hardware Trojan circuit showed that the technique could distinguish the difference of side channel signal's feature between the genuine chip and tested chip, and consequently could detect the existence of the hardware Trojan.
Key words: integrated circuit; hardware Trojans; side channel analysis; projection pursuit; Trojans detection

1 引言

当前，伴随着集成电路（IC, integrated circuit）芯片设计与制造全球化的发展趋势，IC 芯片日益受到各种类型的恶意攻击。例如，芯片外包制造过程中，不可信制造商可能通过对芯片的原始设计进行更改，嵌入所谓的“硬件木马”（hardware Trojans）电路，并在特定的激活条件下实现破坏性功能或泄漏芯片内部秘密信息。

对芯片硬件木马进行检测十分困难，传统方法是对芯片进行解剖后利用电子显微扫描来进行探查，这类技术不仅代价昂贵，而且使待测试芯片受到破坏，只能进行抽样性检测。近年来出现一类利用旁路信号分析（side channel analysis）的硬件木马检测技术[1]（以下简称硬件木马旁路检测），通过对原始电路与待测试电路之间旁路测量信号（例如芯片功耗[2,3]、内部时延[4]、电磁辐射等）的差异来检测木马，由于该类技术对芯片无损害，可以实现逐片检查，因此日益成为国际学术界与产业界的研究热点。硬件木马旁路检测的关键之处在于如何对基准芯片与测试芯片的旁路信号特征进行刻画与差异判别。当前典型方法包括：信号变换法与门级电路特征刻画（GLC, gate-level characterization）法。前者是对旁路信号轨迹进行整体处理，通过对旁路信号进行空间变换与压缩以提取信号特征，而基本不涉及硬件电路设计细节，其优点是实现简便。如 Agrawal 等人[5]通过 Karhunen-Loève 变换（K-L 变换）对旁路功耗信号建立“指纹”并

收稿日期：2012-07-03；修回日期：2012-11-26
基金项目：中国博士后科学基金资助项目(2012M512073)
Foundation Item: China Postdoctoral Science Foundation (2012M512073)

进行比对检测；Gwon 等人[6]利用压缩感知（compressive sensing）信号处理方法来感知电路功耗的异常变化。GLC 法[7,8]是利用多次旁路信号测量形成方程组，通过线性规划与奇异值分解以及约束方程实现公式化的门级电路特征刻画与硬件木马检测。该方法对电路的刻画精度高，可以对小规模硬件木马进行检测，但对于较大规模电路来说计算代价过高。

由于芯片级的功耗等旁路信号十分微弱，而一般硬件木马电路的规模很小，其有效旁路信号相对于原始电路信号来说甚至相差几个数量级。同时，硬件木马电路与芯片原始电路之间的旁路信号不是简单的叠加，往往是通过一些耦合方式融合在一起，这些都导致硬件木马电路的旁路信号分布十分复杂，对其特征进行提取与识别十分困难。文献[5]中的 K-L 变换，本质上是主成分分析（PCA, principal component analysis），作为一种传统的证实性多元分析（CDA, confirmatory data analysis）方法，对于微弱、高维、并且分布复杂的芯片旁路信号来说效果并不理想。本文提出一种新的信号变换分析硬件木马检测方法，通过投影寻踪（PP, projection pursuit）技术将芯片高维旁路信号投影到低维子空间，在信息损失尽量小的前提下发现原始数据中的分布特征，从而实现芯片旁路信号特征提取与识别，并有效实现硬件木马检测。

2　投影寻踪技术

2.1　概述

投影寻踪是分析和处理高维观测数据，尤其是非线性、非正态高维数据的一种新兴统计方法。其基本思想是把高维数据投影到低维（一般 1~3 维）子空间，寻找能反映原高维数据结构或特征的投影（或称为“令人感兴趣”的投影），通过对投影数据的分析达到研究分析原始数据的目的[9]。它既是一种新工具，更是探索性数据分析（EDA, exploratory data analysis）的新思维方式。

衡量投影数据令人感兴趣的程度是通过称为“投影指标”的函数来实现。设 $\boldsymbol{X}$ 是 d 维随机向量，分布函数为 $F_{\boldsymbol{X}}$。$\boldsymbol{A}$ 是从 R^d 到 R^k 的一个线性投影（矩阵）（$d\leqslant k$），$\boldsymbol{Y}=\boldsymbol{AX}$ 是一个 k 维随机向量，分布为 $F_{\boldsymbol{Y}}$。投影指标是定义在 $\boldsymbol{Y}$ 上的实值函数。在实际中，$\boldsymbol{X}$ 的分布一般难以直接得到，只能得到它的样本，因此对于投影方向 $\boldsymbol{A}$，投影指标记为 $Q(\boldsymbol{Y})$或 $Q(\boldsymbol{AX})$，此时 $\boldsymbol{X}$ 是指得到的随机变量样本数据。PP 的目标就是找到一个或几个投影矩阵，使指标值达到最大或最小。

最常见的情形是一维投影(k=1)，此时矩阵 $\boldsymbol{A}$ 简化为一个向量 $\boldsymbol{\alpha}$($\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}$=1)，投影指标简化为 $Q(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X})$。如果将指标取为样本方差，即令 $Q(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X})$= var ($\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}$) ，那么使 var($\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}$)取最大值的方向 $\boldsymbol{\alpha}_1$ 就是数据协方差阵的最大特征根对应的特征向量，即第 1 主成分；如果继续作投影，在与 $\boldsymbol{\alpha}_1$ 垂直的空间里求单位向量 $\boldsymbol{\alpha}_2$，即在约束条件 $\boldsymbol{\alpha}_2\perp\boldsymbol{\alpha}_1$ 下，使得 var($\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}$)取最大值的方向 $\boldsymbol{\alpha}_2$ 是第 2 主成分…依次下去，可以证明 PCA 就是以样本方差为指标，寻找一系列正交投影的 PP[9]。即文献[5]中的信号分析技术实际是 PP 方法的一种特例。

投影寻踪的过程一般采用迭代模式，即：选定投影指标→寻找最佳投影方向→将投影后的数据结构从原数据中去除，得到改进新结构。然后重复上述寻优过程，直到数据的投影不再显著含有感兴趣的结构为止。

2.2　信息散度指标

投影指标是 PP 成功与否的关键因素。对于硬件木马检测来说，方差指标可能并非是反映旁路信息的最佳投影指标。信息散度（ID, information divergence）是在 Shannon 互信息的基础上提出的，它很好地度量了 2 个分布之间的距离。一般认为，服从正态分布的数据含有的有用信息最少，因而通常受到关注的是与正态分布差别大的结构。多元正态分布的任何一维线性投影仍然服从正态分布，因此如果一个数据在某个方向上的投影与正态分布差别较大，那它就一定含有非正态的结构。高维数据在不同方向上的一维投影与正态分布的差别是不一样的，它显示了在这一方向上所含有的有用信息的数量，因此可以用投影数据的分布与正态分布的差别作为投影指标[9]。

对于 2 个连续的概率分布 $p(x)$与 $q(x)$、$p(x)$对 $q(x)$的 ID 定义为

$$d(p;q)=\int_R p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \tag{1}$$

由于信息散度是非对称的，因此定义 $p(x)$、$q(x)$间的绝对信息散度（AID, absolute ID）指标为

$$Q_{\mathrm{c}}(p;q)=\left|d(p;q)\right|+\left|d(q;p)\right| \tag{2}$$

$Q_c(p;q)$刻画了 2 个分布间的偏离程度：当 $p(x)=q(x)$时，$Q_c(p;q)=0$；当 2 个分布偏离增加，$Q_c(p;q)$的值也增加。由于根据样本估计 $p(x)$、$q(x)$很麻烦，因此更简便有效的指标是用离散化的概率分布 p、q 分别代替连续密度函数 $p(x)$、$q(x)$。此时定义

$$D(p;q)=\sum p_i \log\frac{p_i}{q_i} \tag{3}$$

其中，p_i、q_i 分别对应于 p、q 中第 i 个元素。这样，离散 AID 指标就为

$$Q(p;q)=\left|D(p;q)\right|+\left|D(q;p)\right| \tag{4}$$

其中，$\left|D(p;q)\right|=\sum\left|p_i \log\frac{p_i}{q_i}\right|$。若定义 q 为正态分布，则投影指标的值越大，那么意味着 p 越偏离正态分布，因而是本文感兴趣的结构。

3 硬件木马检测方法

利用旁路信息分析，对同一生产批次的全部芯片进行硬件木马检测的典型流程为[5]：1) 从该批芯片中随机选择少量样本作为基准；2) 对基准芯片进行足够多的 I/O 测试以触发所有预期的电路工作，同时获取旁路信号；3) 对旁路信号进行特征提取；4) 对基准芯片进行剖片检测以确认其与原始设计一致；5) 对其余芯片（以下称测试芯片）进行相同的 I/O 测试、旁路信号采集与特征提取，并与基准芯片的结果进行比对，从而在不破坏测试芯片的情况下确定其中是否含有硬件木马。

基于上述检测流程，下面给出采用投影寻踪信号分析技术的硬件木马检测详细方案。

Step1 旁路信息采集。选择 n 个测试向量，输入基准芯片使之正常运行，对芯片产生的旁路信号（如功耗）进行采样形成 $n\times m$ 阶轨迹矩阵，记为 $\boldsymbol{B}=\{b(i,j)|i=1,2,\cdots,n;\ j=1,2,\cdots,m\}$，其中，$m$ 表示每条轨迹的采样长度。类似的，以相同测试向量集针对测试芯片形成 $n\times m$ 阶轨迹矩阵 $\boldsymbol{T}=\{t(i,j)|i=1,2,\cdots,n;\ j=1,2,\cdots,m\}$。

Step2 对基准矩阵 $\boldsymbol{B}$ 进行投影，构造服从正态分布的投影值。令投影方向为 $\alpha=\{\alpha(1),\ \alpha(2),\cdots,\alpha(m)\}$，则一维投影值 $ZB(i)$可综合为

$$ZB(i)=\sum_{j=1}^{m} a(j)b(i,j),\ i=1,2,\cdots,n \tag{5}$$

由于基于正态分布的函数偏度等于 0，峰度等于 3，因此最小化指标函数为

$$\min\ F(k_3,k_4)=\left|k_3\right|+\left|k_4\right| \tag{6}$$

约束条件：s.t. $\sum_{j=1}^{m}\alpha^2(j)=1$

其中，$k_3=\frac{\mathrm{E}(ZB-\overline{ZB})^3}{\sigma^3}$；$k_4=\frac{\mathrm{E}(ZB-\overline{ZB})^4}{\sigma^4}-3$

即可得到矩阵 $\boldsymbol{B}$ 最接近于正态分布的一维投影值 $ZB(i)$。

Step3 对 $\boldsymbol{B}$ 进行投影，使之包含最多有用信息（即投影后分布与正态分布差异最大）。令投影方向为 $\beta=\{\beta(1),\beta(2),\cdots,\beta(m)\}$，则一维投影值 $zb(i)$可综合为

$$zb(i)=\sum_{j=1}^{m}\beta(j)b(i,j),\ i=1,2,\cdots,n \tag{7}$$

根据 2.2 节，可以通过求解 AID 指标函数最大化问题来估计最佳投影方向，即

$$\max\ Q(zb;ZB)=\left|D(zb;ZB)\right|+\left|D(ZB;zb)\right| \tag{8}$$

约束条件：s.t. $\sum_{j=1}^{m}\beta^2(j)=1$

其中，$D(zb;ZB)=\sum\left|zb_i\log\frac{zb_i}{ZB_i}\right|$

$$D(ZB;zb)=\sum\left|ZB_i\log\frac{ZB_i}{zb_i}\right|$$

可得最佳一维投影 $zb(i)$及最佳投影方向 β_{zb}。

Step4 与 Step2 类似，得到测试矩阵 $\boldsymbol{T}$ 最接近于正态分布的一维投影值 $ZT(i)$。

Step5 与 Step3 类似，得到测试矩阵 $\boldsymbol{T}$ 含有最多有用信息的一维投影值 $zt(i)$及最佳投影方向 β_{zt}。

Step6 利用公式(4)，分别求 $zb(i)$与 $zt(i)$之间的 AID 值 $Q(zb;zt)$及 β_{zb}与 β_{zt}之间的 AID 值 $Q(\beta_{zb};\beta_{zt})$。前者可以衡量矩阵 $\boldsymbol{B}$、$\boldsymbol{T}$ 分别进行最佳一维投影后所得分布之间的偏离程度；后者可以衡量投影方向分布之间的偏离程度。若偏离程度十分明显，则意味着基准芯片与测试芯片的旁路信号存在明显差异，则可判断测试芯片中存在硬件木马。偏离程度是否明显可以通过阈值范围判定，而阈值的确定可以通过对基准芯片进行多次测量计算先验获取。

Step7 若上一步的阈值判断不明确，则重复 Step3~Step5，分别在与前述最佳投影方向垂直的空间里（即增加约束条件：与前述最佳投影方向正交）估计矩阵 $\boldsymbol{B}$、$\boldsymbol{T}$ 的最佳投影方向及对应的投影值，并类似 Step6 进行硬件木马的存在判定。

Step8 上述 Step7 重复进行，多次投影，直到

明确作出硬件木马存在判断或者不再存在令人感兴趣的投影结构为止（此时硬件木马检测失效）。

在上述检测方案中，AID 具有双重作用：一是作为确定高维旁路信号投影方向的投影指标；二是作为投影后判断硬件木马是否存在的判定指标。

4　硬件木马检测实验

4.1　实验配置

实验基准芯片是一个带串行通讯接口的高级加密标准（AES, advanced encryption standard）加密器。AES 具有多种算法参数与不同的硬件实现方式。不同参数与实现的选择将导致硬件电路速度与规模产生很大差异，并不可避免对硬件木马电路的检测效率产生影响。为更好体现上述硬件木马检测方法的有效性，本着从简入手的原则，本文选择的是密钥与数据分组均为 128bit，10 轮加密的 AES-128。通过流水线方式实现并去掉解密电路，使最终实验基准电路模块规模适中。测试芯片包括 2 片，其中，一片与基准芯片完全相同，而另一片中加入可对密钥进行扩频调制发射的秘密泄漏型硬件木马电路（如图 1 所示），设计细节可参见文献[10]。这是一种常开（always-on）型硬件木马，其特点是始终处于活动状态，因此可以排除激活因素对信号分析结果的影响。硬件木马部分约占整体电路规模的 10%。对于特定激活型硬件木马，当采用适当的测试向量进行部分或全部激活后，本检测方法同样适用。

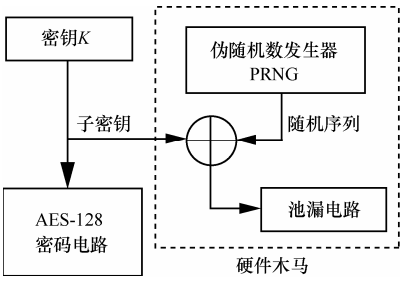


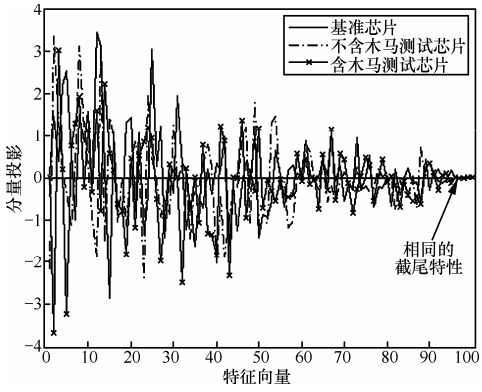
图 1　硬件木马电路示意

实验芯片采用 Xilinx 公司的 FPGA－XC3S500E 芯片，设定运行频率 50 MHz。采用与文献[11]类似的芯片功耗信息采集平台，采样频率为 500MSa/s，采样时长 0.2μs，因此每条轨迹采样长度为 100。采用规模为 50 的随机测试向量集，基准芯片与测试芯片均生成 50×100 阶的功耗信号矩阵。为减少噪声影响，实际每个相同测试向量均重复采样 20 次，平均化后得到对应轨迹。

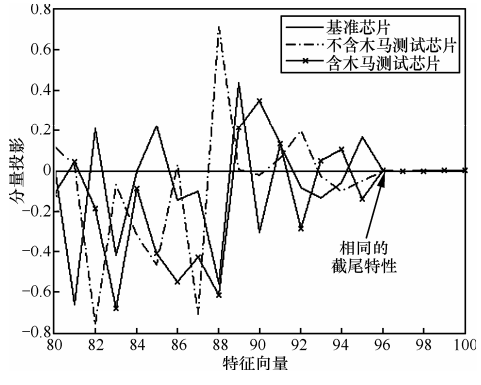
投影过程中寻找最佳投影方向是一个复杂的带约束的优化问题，选取何种寻优算法直接影响寻优效率，甚至影响能否获得最优解。本文采用 MATLAB7 优化工具箱中的有约束非线性优化工具 fmincon 函数执行这一寻优过程。fmincon 函数有 4 种算法实现，优化选项参数比较复杂。实验选用其中的序列二次规划 SQP 法来执行寻优，其他选项参数（如最大迭代次数等）选用默认值或根据实际情况进行适当调整。由于初始值的选取对寻优结果影响很大，为增加寻优精度，本文设计了一种初始值加速处理方式，即首次初始值为随机选取，然后将第一次的寻优结果设置为初始值进行二次寻优，依此类推，将执行 10 次初始值加速寻优过程后的结果作为本次寻优的最终结果。

4.2　实验结果

图 2 为采用 K-L 变换信号分析[5]的实验结果，横坐标表示按主成分排序的特征向量，纵坐标表示对应的投影坐标值。理想情况下，基准芯片与含硬件木马的芯片具有不同的投影分布，表现为“截尾”子空间不同（即二者趋近于 0 的速度不同）。但图 2 中没有表现出这种分布差异，这意味着该方法无法判断是否存在硬件木马电路。



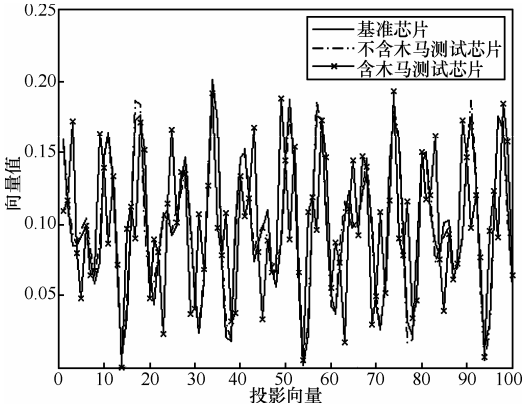
(a) 整体分布



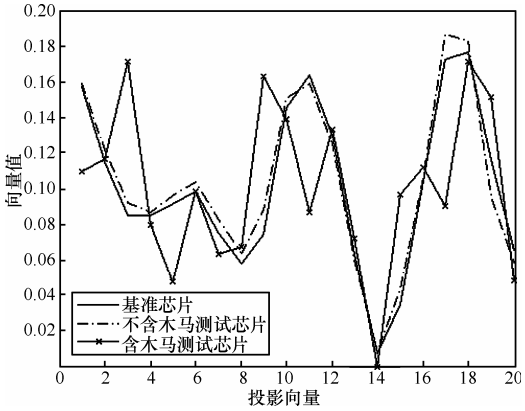
(b) 局部放大

图 2　采用 K-L 变换分析的硬件木马检测结果

图 3 为采用 PP 信号分析的实验结果。图中显示的是基准信号与 2 种测试信号的最佳一维投影方向分布。可以看到，基准信号与不含硬件木马测试信号投影方向分布明显更为接近（AID 值为 0.612 6），远小于基准信号与含硬件木马测试信号投影方向之间的偏离程度（AID 值为 3.696 0）。



(a)整体分布



(b)局部放大

图 3　采用 PP 分析的硬件木马检测结果

实验表明，以 AID 为指标的 PP 对芯片旁路信息特征提取与识别的能力比 PCA（K-L 变换）更高，但计算代价有较大幅度增加。以实验针对的 50×100 维原始数据为例，在普通 PC 机（CPU 为 Intel　Core i5-2400 3.10GHz，内存为 2GB）上 K-L 变换法分析耗时约 100ms，PP 法分析耗时约 2s。当原始数据增加到 50×500 维时，K-L 变换法分析耗时约 1s，而 PP 法耗时增加到约 350s，增加幅度比较明显。由于芯片硬件木马检测对实时性要求不高，相对于芯片的安全性需求来说，PP 检测方法的计算代价仍然可以接受。

5　结束语

基于 AID 指标的 PP 技术，能够将高维、分布复杂的芯片旁路信号数据映射到低维子空间，便于对样本数据分布特征进行分析与识别，为实现芯片硬件木马无损对照检测提供了一条值得探索的新途径。为了使检测效果更为理想，需要针对不同类型的硬件木马研究触发与测试向量生成技术，从而使测试样本中尽可能地包含硬件木马特征信息。同时需要研究有效的全局优化算法，降低 PP 寻优耗时，提高硬件木马检测效率。

参考文献：

[1] TEHRANIPOOR M, KOUSHANFAR F. A survey of hardware Trojan taxonomy and detection[J]. IEEE Design & Test of Computers, 2010, 27(1): 10–25.

[2] RAD R M, WANG X X, TEHRANIPOOR M, *et al*. Power supply signal calibration techniques for improving detection resolution to hardware Trojans[A]. ICCAD 2008[C]. San Jose, USA, 2008. 632-639.

[3] KOUSHANFAR F, MIRHOSEINI A. A unified framework for multi-modal submodular integrated circuits Trojan detection[J]. IEEE Transactions on Information and Security, 2011, 6(1): 162-174.

[4] JIN Y, MAKRIS Y. Hardware Trojan detection using path delay fingerprint[A]. IEEE International Workshop on Hardware-Oriented Security and Trust(HOST)[C]. Anaheim, USA, 2008. 51–57.

[5] AGRAWAL D, BAKTIR S, KARAKOYUNLU D, *et al*. Trojan detection using IC fingerprinting[A]. IEEE Symposium on Security and Privacy[C]. Berkeley, USA, 2007. 296–310.

[6] GWON Y L, KUNG H T, VLAH D. DISTROY: detecting integrated circuit trojans with compressive measurements[A]. The 6th USENIX Workshop on Hot Topics in Security(HotSec'11)[C]. San Francisco, USA, 2011.

[7] POTKONJAK M, NAHAPETIAN A, NELSON M, *et al*. Hardware Trojan horse detection using gate-level characterization[A].Design Automation Conference –DAC[C]. San Francisco, USA, 2009. 688- 693.

[8] WEI S, POTKONJAK M. Scalable hardware Trojan diagnosis[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2012, 20(6):1049-1057.

[9] 付强, 赵小勇. 投影寻踪模型原理及其应用[M]. 北京:科学出版社, 2006.
FU Q, ZHAO X Y.　Projection Pursuit Model, Principle and Application[M]. Beijing: Science Press,2006.

[10] 邹程, 张鹏, 邓高明等. 基于功率旁路泄露的硬件木马设计[J]. 计算机工程, 2011,37(11): 135-137.
ZOU C, ZHANG P, DENG G M, *et al*. Design of hardware Trojan based on power side-channel exposure[J]. Computer Engineering, 2011,37(11): 135-137.

（下转第 137 页）