硬件木马检测方法综述

王立敏
中国科学院信息工程研究所 第五实验室 北京 中国 100093

摘要：随着集成电路的发展和经济全球化的分工协作，使得在芯片设计过程中插入硬件木马的风险变大。硬件木马可能会完全改变芯片的功能或者泄露重要信息，危害十分巨大，应该引起重视。本文主要梳理了目前主流的硬件木马检测技术，并列举了一些具体法例子。在文末做出了总结以及展望。
关键词 硬件木马；安全；检测；

A Survey on the Hardware Trojan Protection

Wang Limin
Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract As the development of IC design and economic globalization , It is more risky to insert hardware Trojan into IC design in each steps. Hardware Trojan will damage the chips or leak the important information in the chips. This paper sorts out the current mainstream hardware Trojan detection technologies, and introduce some cases. Finally, it puts forward the future of Hardware Trojan after summarizing the current researches.
Key words Hardware Trojan; security; detection;

1 引言

近些年来，由于集成电路的设计复杂度不断提升，以及经济全球化浪潮推动的分工细化，使得集成电路设计中的各个环节都可能受到安全威胁。在功能设计环节，我们可能会引用第三方的 IP 核来方便我们的工作，在物理设计环节，恶意 EDA 也可能修改 RTL 或者门级网表，在测试和制造时，由于需要经多个生产商之手，而某些生产商并不一定可信，也会存在被修改或者添加一些额外的部件的可能。

硬件木马主要由触发器和有效载荷两部分组成。触发器用于激活硬件木马，例如当接收到一个外部的有效输入或者检测到内部的某些有效条件，则触发硬件木马。有效载荷则是硬件木马实施破坏功能的部分。

硬件木马按照功能的不同可分为许多不同的类型，且会造成许多不可控制的

后果。后门型的硬件木马可以泄露芯片中的敏感信息，破坏型木马可以改变电路原有的功能，甚至使原有电路不再工作。

由此可见硬件木马的危害极其大，然而硬件木马自身具有隐蔽性，难以触发，而且电路的集成度越来越高使得硬件木马的体积也变得越来越小（难以区分正常电路和硬件木马电路），如何检测出硬件木马成了一个难题。本文将梳理常见的硬件木马检测方式，并比较他们的优缺点。

2 硬件木马检测方案

硬件木马检测大致分为三类，破坏性检测，非破坏性检测和侧信道信号分析方法。[1]

2.1 破坏性验证

逆向工程是最常见的方法，将芯片打开，并且在显微镜下将芯片的每一层结构观察并记录下来，最终再根据自己看到的图还原出原始的设计。逆向工程通常可以获得产品和厂商的各种信息，同时也能获取系统级的信息，例如分析操作，功能，内部连接，分析得到它是如何被生产出来的，以及将芯片解析到晶体管层[2]。通过逆向工程 100%能找到恶意修改，但是所需要的成本过高，分析一块常规规模的芯片很可能需要几个月甚至几年的时间。传统的逆向工程方法耗费的时间太多，并且大多数逆向工程方法都需要得到黄金芯片来与之对比才能知道是否存在硬件木马（有时候黄金芯片并没有那么容易获得），因此有文章提出基于逆向工程的利用机器学习的方法（支持向量机）来识别硬件木马[3]，如图 1 所示，它通过选择特征，并利用 SVM 作为分类器将是否存在硬件木马区分开来。该文虽然宣称不需要黄金芯片，但是在选择特征的时候，还是提到了要与黄金芯片进行对比才行，并且目前硬件木马的有监督学习数据库只有 TrustHub 一个，数据量并不大，而且其中有一些错误，用机器学习的方法虽然新颖，实验结果看起来也十分漂亮，但是距离大规模使用还有一段距离。

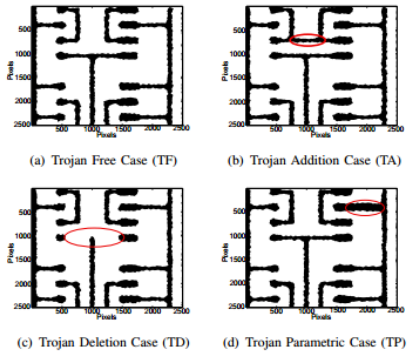

图 1 三种木马的例子，这是扫描电子显微镜下的金属层

2.2 非破坏性验证

逆向工程最大的问题就是需要破坏芯片，这在很多场合是不被允许的，因此还有许多非破坏性检测方法被提出。其中一大类是功能测试，或者称逻辑测试，也就是通过测试向量来激活木马，然后比较输出值和正常值是否匹配来判断电路是否存在问题。ATPG（Auto Test Pattern generation）技术和 DFT（Design For Trust）是相当常用的方法。ATPG[4]基于侧信道分析的测试技术，通过在芯片入口处施加激励，然后检查输出，如果输出与预期不同，则说明是个电路缺陷或者是个木马。但是这个方法的目的主要是为了发现暴露的错误，而不一定是硬件木马。并且它的逻辑测试尽力确保生成的测试向量可以覆盖硬件木马的触发条件，然而硬件木马设计出来就是难以触发的，这使得这一技术实施效果并没有预期的好。还有一个问题就是难以将硬件木马侧信道信号从实验噪声和 CMOS 处理噪声中分离出来。对于木马难以触发这一难题，另一篇文章[5]提出一种新的方法来试图改进它，它通过频繁地翻转 NAND，AND，NOR，OR 门对结果进行排列组合，并且将其作为硬件木马的触发方案，来最大可能地激活硬件木马，与此同时持续的向量仿真可以降低整体电路活动，从而放大了硬件木马带来的额外的翻转。但是我们不可能列举组合逻辑电路的所有的状态，因此这一方法也存在局限性。在线监测技术的机制[6]与 ATPG 类似，但是有所改进，它是硬件木马检测的最后一道防线，由于硬件木马的插入，会影响某个节点（Important Node）以及它周围的其余节点的正确关系，因此它在原始电路中的一些可疑节点（Important Node）插入图 2 所示的检查逻辑（checker）。

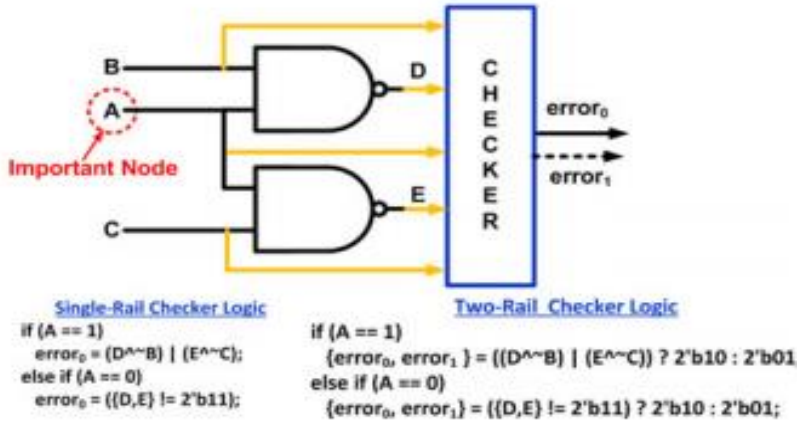

图 2 一个检查逻辑的例子，对于单路的检查器（只有 error0 输出）而言，若可疑点 A 为 1 则 D=~B E=~C 若 A=0，则 D=E=1。而对于双路（error0 和 error1 都输出）而言，若存在 HTH 则输出为 10　否则输出为 01

在线监测技术可以在很大程度上避免电路噪声的干扰，但是它依然无法准确识别是功能设计错误还是硬件木马，而且各个检查逻辑（checker）中的信息汇总输出需要一定的延时，因此可能在木马运行一段时间之后才能得知。

功能测试只适用于一些改变功能的硬件木马，例如[7]中所关注的雷达后门事件，但是对于一些类似于泄露无线加密芯片中的 key[8]，或者芯片正在处理的敏感信息的信息泄露型的木马，功能测试并测不出他们，而且功能测试的测试逻辑较为复杂，比较适合小的硬件木马。

2.3 侧信道分析

还有一类比较常用的方法是侧信道分析的方法，侧信道分析通常是分析待测芯片的功耗，发热，等来判断是否存在硬件木马。一般的侧信道分析主要从一个角度来进行分析，比如[9]利用热学图片来检查硬件木马，由于插入硬件木马将会增大温度，因此它将热图像拍下来然后比较测试芯片和黄金芯片，最终得到温差矩阵，如果没有硬件木马存在则温差几乎为 0，如果存在则基本大于 0。 也有利用多参数来判别硬件木马的，例如[10]通过测试电流和频率，选择相同频率下工作的待测芯片和黄金芯片进行对比，如果不匹配度超出了阈值，则认为存在硬件木马。侧信道分析十分依赖黄金芯片，这使得它的使用范围受限，而且小的硬件木马对温度，电流的影响十分的小，侧信道分析不一定可能分析和判断是否存在硬件木马，因此侧信道分析更适用于检测大的硬件木马。

3 总结

目前硬件木马检测方案都有所限制，或需要黄金芯片，或只能检测大硬件木马或者小硬件木马，还没有切实可用的通用检测方案。而且随着集成电路设计的进步，硬件木马将越来越难以检测，尽管机器学习的发展给了本行业一些新的启发，但是由于硬件样本数据库的缺乏等原因，这一方向也并不成熟。这个领域十分重要，但还有很长的路要走，需要引起我们的重视。

4 国内外安全研究和产业实践

计算机体系结构是整个计算机的基石，中国与国外在这方面的差距十分巨大，前段时间中兴遭到美国制裁就说明，至少中国在芯片行业中，与美国还是有一定的差距的。纵观现在的计算机行业，虽然中国也有很多十分厉害的计算机公司，比如联想，华为等，但是其中的很多重要部件都来自于国外，联想计算机的 CPU 大部分来自 Intel，而操作系统大部分来自于微软。华为的手机稍微乐观一些，

因为华为开始自己设计手机上的 CPU，然而性能等依旧很难一下子追上国外的高通等厂商，而且操作系统依旧是来自国外的安卓。中国在这个领域内依然是受制于人。这便意味着，这些产品若是被生产厂商植入后门，后果将十分严重，必然将造成严重的信息泄露。自主可控的强烈需求变得越来越突出了，大力发展国产芯片和国产操作系统的呼声近些年在国内越来越大。所幸的是像龙芯，兆芯等国产芯片一直在这个行业默默耕耘，目前已经有十分可观的成果了，龙芯目前在我国的很多重要领域都有所应用，比如导弹系统，卫星上，而很多国产的操作系统也被应用到了许多敏感的地方。然而我们不得不承认，我国在计算机领域上与国外还是存在挺大的差距的。

棱镜门事件之后，全世界开始更加重视计算机安全问题。Intel 在 2013 年发布了 SGX（Intel® Software Guard Extensions）技术，并在第 6 代 CPU 中集成了这项技术，它通过给每个软件分配一个 enclave 的容器，将敏感数据独立开来，单独处理，连操作系统都无权篡改，以此来保证软件的运行安全。与此对应的，ARM 上运用了 TrustZone 技术来保证安全，它并没有为单独的软件分配安全空间，而是在直接开辟了一块安全空间，所有需要保护的数据都放在里面。中国则致力于构造一个可信计算平台，通过一小块绝对安全的芯片作为根，并在此基础上延伸扩展最终构建成一个安全体系，这个芯片一般被称作 TPM（Trusted Platform Module）。

国外的发展模式更趋向于从具体实现到体系的自下而上的发展模式，而国内的发展模式则趋向于从体系到具体实现的自上而下的发展模式。这两者并不一定有优劣之分，前者可能在前期看起来会比较混乱，但是当发展到一定程度，形成体系是必然的。而后者虽然在理论上比较完美，在实际应用中也有许多厂商，比如金立，华为，360 等厂商在自己的手机中实现了自己安全模块，但是部分理论在实际应用中还是存在一定的困难的。国内所研究的体系还是应该更贴合实际，以实现作为目标，更能事半功倍。

参考文献

[1] Xiao, K., Forte, D., Jin, Y., Karri, R., Bhunia, S., & Tehranipoor, M. (2016). Hardware Trojans: Lessons Learned after One Decade of Research. *ACM Transactions on Design Automation of Electronic Systems*, *22*(1), 1–23. https://doi.org/10.1145/2906147.

[2] Torrance, R., & James, D. (2009). The State-of-the-Art in IC Reverse Engineering. In C. Clavier & K. Gaj (Eds.), *Cryptographic Hardware and Embedded Systems – CHES 2009* (Vol. 5747, pp. 363–381). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04138-9_26.

[3] Bao, C., Forte, D., & Srivastava, A. (2014). On application of one-class SVM to reverse engineering-based hardware Trojan detection (pp. 47–54). IEEE. https://doi.org/10.1109/ISQED.2014.6783305.

[4] Cruz, J., Farahmandi, F., Ahmed, A., & Mishra, P. (n.d.). Hardware Trojan Detection using ATPG and Model Checking, 6.

[5] Banga, M., & Hsiao, M. S. (2009). VITAMIN: Voltage inversion technique to ascertain malicious insertions in ICs (pp. 104–107). IEEE. https://doi.org/10.1109/HST.2009.5224960.

[6] Chakraborty, R. S., Pagliarini, S., Mathew, J., Rajendran, S. R., & Devi, M. N. (2017). A Flexible Online Checking Technique to Enhance Hardware Trojan Horse Detectability by Reliability Analysis. *IEEE Transactions on Emerging Topics in Computing*, *5*(2), 260–270. https://doi.org/10.1109/TETC.2017.2654268.

[7] Adee, S. (2008). The Hunt For The Kill Switch. *IEEE Spectrum*, *45*(5), 34–39. https://doi.org/10.1109/MSPEC.2008.4505310.

[8] Liu, Y., Jin, Y., & Makris, Y. (2013). Hardware Trojans in wireless cryptographic ICs: Silicon demonstration & detection method evaluation (pp. 399–404). IEEE. https://doi.org/10.1109/ICCAD.2013.6691149.

[9] Zhong, J., & Wang, J. (2018). Thermal images based Hardware Trojan detection through differential temperature matrix. *Optik*, *158*, 855–860. https://doi.org/10.1016/j.ijleo.2017.12.145.

[10] Narasimhan, S., Du, D., Chakraborty, R. S., Paul, S., Wolff, F., Papachristou, C., … Bhunia, S. (2010). Multiple-parameter side-channel analysis: A non-invasive hardware Trojan detection approach (pp. 13–18). IEEE. https://doi.org/10.1109/HST.2010.5513122.