

引用格式:李翔宇. 基于去协同的抗硬件木马交换机解析器[J]. 微电子学与计算机, 2021, 38(4): 1-5. [LI X Y. Hardware trojans resilient switch parser based on destruction of cooperation[J]. Microelectronics & Computer, 2021, 38(4): 1-5.]

基于去协同的抗硬件木马交换机解析器

李翔宇

(清华大学 微电子学研究所, 北京 100084)

摘 要: 设计了一种既支持软件定义网络又能在有硬件木马的情况下正常工作的交换机包解析器. 使用以查找表为核心的可重构基本处理单元构成包解析器, 同时将输入数据与随机密钥进行异或(去协同变码)后再送入包解析器以破坏包头中潜藏的木马触发指令. 采用模板同步变码的方式进行关键字匹配, 从而避免原文在模块内部出现. 基于该结构实现的 L4 以太网交换机包解析器, 使用 40 nm 工艺库综合的时钟频率可以达到 220 MHz, 基本处理单元的面积约为 176 万门. 测试结果表明, 该包解析器可以破坏不同类型的木马触发指令.

关键词: 硬件木马; 交换机芯片; 包解析器; 软件定义网络

中图分类号: TP309

文献标识码: A

文章编号: 1000-7180(2021)04-0001-05

Hardware trojans resilient switch parser based on destruction of cooperation

LI Xiang-yu

(Institute of Microelectronics, Tsinghua University, Beijing 100084, China)

Abstract: A switch parser that can not only support software-defined network but also work even with hardware Trojans has been designed. It is composed of reconfigurable process elements of which kernels are lookup tables. The input data are XORed with a random key (Destruction of Cooperation transformation) before being fetched into the parser to destruct the possible argots in the head of packets. The key word match is completed in the approach of transforming the pattern templates at the same time, which avoids appearance of the original data. Synthesized with a 40nm process, the L4 switch parser implemented in this structure achieves up to 220 MHz clock frequency, the area of the PE is about 1760,000 gates. The test results show that the parser can destruct different types of Trojan trigger conditions.

Key words: hardware Trojan; switch; parser; software-defined network

1 引言

硬件木马^[1-2]是当今集成电路面临的安全威胁之一. 对网络安全而言, 硬件木马攻击比传统网络攻击更加危险、更加难以发现, 因此, 高安全交换机芯片需要能够避免硬件木马(包括未知的潜在木马)攻击. 包解析器是交换机芯片中的数据入口, 易受木马攻击, 但现有包解析器设计很少考虑

了硬件木马问题. 本文以以太网交换机中的包解析器为研究对象, 运用去协同技术设计了一款对硬件木马具有主动防御能力的硬件可编程的以太网交换机包解析器.

现有的硬件木马防御技术以检测技术为主^[3]; 也有人设计了用来证明 IP 是否存在木马的技术方案^[4]. 这些技术往往都需要对木马或硬件本身进行建模. 可编程芯片使得各部分电路的最

终功能在芯片加工之后定义^[5],虽然可以降低木马植入的成功率,但需要芯片中要有足够的冗余度。去协同技术是对输入被保护硬件的数据进行重新编码(变码),改变输入给内部电路的数值,在数据离开保护域时再将其解码(恢复原来的数值),该过程对外透明,但是保护范围内的数据变成了不确定的其它数值,按照预先设计的触发条件输入的数据(“暗语”)将被破坏,从而使得硬件木马无法被触发^[6]。这一技术不需要参考模型,对于未知木马仍然有效,而且即便已经被植入了木马,硬件仍然能够正常工作。文献[6]给出了一种存储器 IP 的去协同方案,它使用精简后的 PUF-FIN 算法对数据进行变码和解码。但是,这种做法只适用于数值无关的功能模块,包解析器需要根据输入数据各个字段的取值分解包头,产生解析标识,所以不能使用文献[6]的方案。

2 去协同可重构包解析器结构

为了满足软件定义网络(SDN)可以通过软件编程解析不同的协议帧格式的要求,越来越多交换机的数据平面采用了可配置硬件的方式实现^[7-10]。本文在典型的基于可重构基本处理单元的包解析器结构基础上引入去协同技术,设计了一种同时满足硬件木马安全性 SDN 灵活性要求的包解析器结构。

2.1 基于基本处理单元的可重构包解析器工作原理

包解析器的输入数据为来自交换机转发引擎输入端口的数据包头部和首层协议的关键字段偏移量及前级单元发来的端口等信息(中间描述符),其输出数据为对数据包头部中的关键字段进行提取、解析后产生的描述信息(送往后级的中间描述符)和数据包头部的原始数据。本文设计以文献[9]的设计为基础改进而来。原始包解析器结构将每一层的包头解析映射到一级基本处理单元(PE),多个 PE 级联组成解析流水线。文献[9]中的每级 PE 从包头中提取本级所需的关键字段,然后把关键字段和预存的模板进行匹配,根据匹配结果确定当前层的协议类型和下一层包头的协议类型。然后,进一步从动作查找表中检索出所要执行的操作,完成字段提取、描述符赋值等操作。各级 PE 解析的结果存储在中间值寄存器堆(IRF)中。其中的模板匹配单元和动作查找表都是由查找表(LUT)实现的,其内容可以根据所支持的协议进行配置。通过对每级 PE 进行不同的配置就可以实现不同层的协议解析。

2.2 包解析器的去协同方案

本文去协同包解析器的基本设计思想是:(1)使用一个与数据包头等长的随机数(密钥)与包头按位异或,完成变码,再输入解析器,解析器对变码后的数据包头进行解析;(2)解析中与数据值相关的步骤是匹配,解决办法是对模板进行同步变码,然后用变码后的模板和变码后的数据进行匹配;(3)为了保证模板正确变码增加一个和数据流水线对应的密钥流水线,使得变码密钥和被变码的数据包头同步传输,并始终维持数据位和密钥位之间的对应关系不变,关键字段提取时,提取相同位置的密钥比特用来生成变码后模板;(4)对于从包头中提取的结果字段,同时提取对应位置的密钥字段保存到密钥 IRF 中;(5)在包解析输出的最后一级进行解码——把 IRF 中的包头字段和对应的密钥字按位异或,恢复其真实值。

按照上述思路,去协同包解析器的具体结构如下。

2.2.1 去协同包解析器顶层结构

去协同包解析器的结构如图 1 所示,包括:解析器配置单元、数据预处理单元、解析处理流水线以及数据后处理单元。

解析器配置单元的功能是在配置阶段将 CPU 加载的配置文件写入对应的配置寄存器和各级 PE 的查找表中,其与去协同技术无关,本文不展开介绍。数据预处理单元首先生成一个与包头等长的密钥,然后将密钥和包头按位异或完成变码,再将密钥(密钥帧)和变码后的包头(数据帧)输送给解析处理流水线。解析处理流水线中除了传统包解析器的流水线通路(称为数据通路)外,还有一个并行的密钥通路,它们的输入分别是数据帧和密钥帧。

解析处理流水线由结构类似的 PE 级联而成。该 PE 包括数据通路和密钥通路两个部分,我们称这种新的 PE 结构为去协同解析 PE。数据帧、密钥帧和中间描述符沿着流水线传递,前面级的寄存器堆的 IRF 数据会复制到下一级流水线,同时又会产生新的解析结果字段,因此 IRF 和密钥 IRF 的大小随着级数而增加。解析处理流水线所在区域为去协同有效作用域,交换机的后级模块读入的数据应该事先解码。因此,数据后处理单元主要功能是对数据帧解码,并调整中间描述符的格式,发送给后级电路。

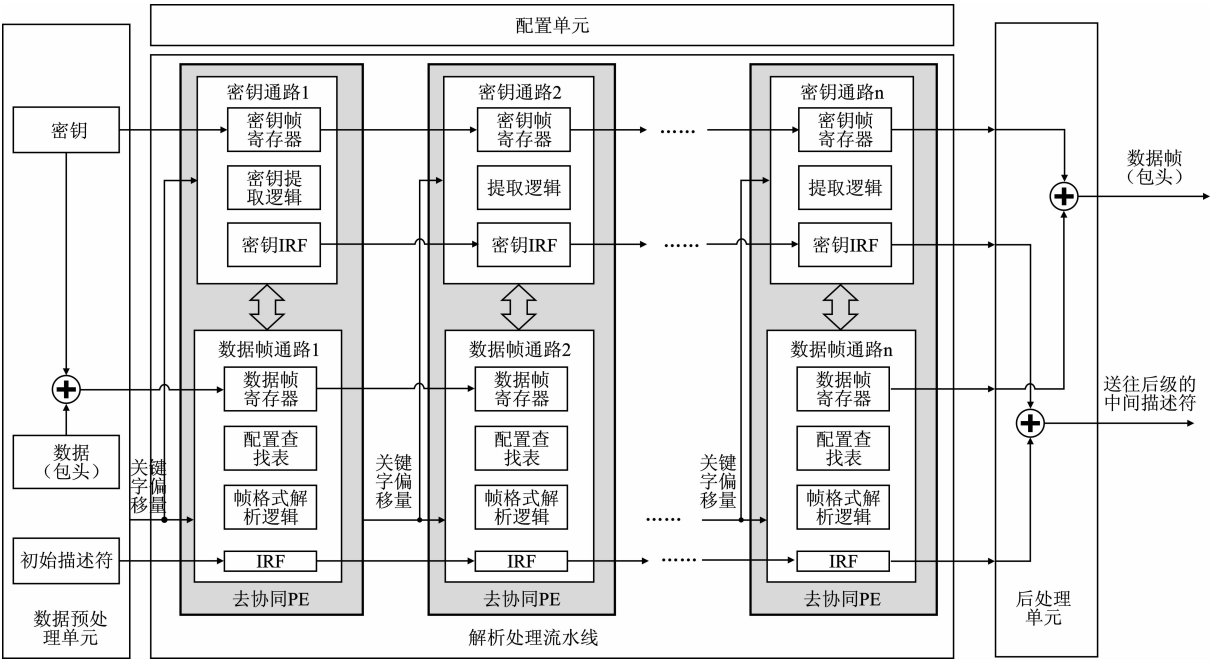


图 1 去协同包解析器结构框图

2.2.2 去协同解析 PE

每级去协同解析 PE 可以解析一层协议. 如图 2 所示,包括数据和密钥两套通路,密钥通路实际上是密钥的同步机制,即在数据通路中对数据进行的操作,在密钥通路中要完成相同操作. 因此二者有很多相似结构和共享信号. 涉及的主要单元如下:

2.2.2.1 帧偏移单元

因为实际数据包中由于每个包的各层网络协议的类型不同,每层封装的头部长度不固定,每层关键字的绝对位置和外面各层的协议类型相关. 为了可以缩小关键字偏移量的变化范围,从而简化逻辑,本包解析器中为每一层协议的封装头都分配了固定的偏移量. 帧偏移单元的功能是将数据包头部对齐到每级 PE 所对应的预设的固定起始位置. 每级 PE 解析完成都要把尚未解析的一段数据帧整体平移到下层封装预设的起始位置. 为了保证密钥和数据对齐,两个数据通路中配有相同的帧偏移单元,使用相同的偏移量值(LYR_offset)——该值由解析逻辑根据本级封装头长度确定.

2.2.2.2 PA 单元

PA 单元是关键字提取单元. 它根据输入的关键字偏移量,从帧寄存器中把相应字段提取出来,送给匹配单元 PB. 在去协同解析 PE 中数据通路和密钥通路各有一个 PA 单元,它们的输入偏移量是同一个,但是连接的帧寄存器分别是数据帧和密钥帧. 其具体结构与文献[9]相同,在此不再赘述.

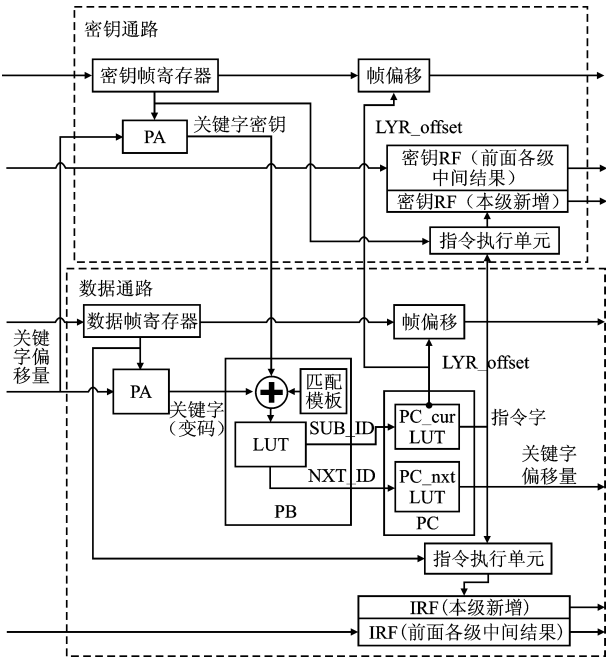


图 2 去协同基本处理单元的结构框图

2.2.2.3 关键字匹配单元 PB

原始的关键字匹配单元的结构中包括一个匹配模板寄存器组和一个匹配查找表. 匹配模板寄存器组中的每条记录对应一个匹配查找表的条目. 将输入的关键字与所有模板进行分别匹配,匹配的结果经过译码逻辑转换成 LUT 的输入地址. PB 的 LUT 中存储的是本层协议操作指令索引(SUB_ID)和下一层协议类型信息(NXT_ID).

在去协同解析 PE 中,输入 PB 的关键字是变码后的数值,无法直接使用原始模板匹配.但如果在进行关键字匹配查询时将数据恢复,又会使得原始输入数据出现在被保护的范围内,木马触发暗语可能渗透进来.所以,在这里采用对模板进行相同的变码(同一密钥异或)的方式解决上述问题,当数据帧原始值和匹配模板一致时,它们使用相同密钥变码后的结果也一定相等.在进行匹配时,先把 PB 单元内部存储的协议匹配模板与来自本级密钥通路的密钥关键字按位异或形成变码后的模板,然后再进行匹配.这样,解决了现有去协同技术不能用于解析器的问题,实现了解析器全范围的去协同防护.(注:当帧格式错误,PB 无法匹配时会输出错误标识或 bypass 标识交由后级处理,由于相关处理机制与本文主题无关,故在文中省去.)

2.2.2.2 节所述的,在密钥通路配置相同的 PA 单元的作用就是提取输入 PB 单元的密钥帧关键字.

2.2.2.4 操作指令及下一层协议类型索引单元 PC

PC 单元中有 PC_{cur} 和 PC_{nxt} 两个查找表,它们表项结构如表 1 所示,其中 PC_{cur} 表可以根据 PB 表输出的 SUB_ID,查找出本层处理单元要执行的操作指令及帧偏移量(LYR_offset).PC_{nxt} 表用来根据 PB 表输出的 NXT_ID 查找出下一层协议的关键字段偏移量(下一层 PE 的 PA 单元的输入).

表 1 PC 表的表项结构

项目	输入	输出
PC _{cur} 表	SUB_ID	操作指令字 LYR_offset
PC _{nxt} 表	NXT_ID	关键字偏移量

PC 单元只存在于数据通路上,它输出的指令字同时送给密钥通路和数据通路中的指令执行单元.

2.2.2.5 指令执行单元

指令执行单元接收来自 PC 的操作指令,填充 IRF.数据通路部分包含 MOV 和 SET 两种指令操作单元;而密钥通路的指令执行单元,只有 MOV 指令操作单元. MOV 指令的功能是将帧寄存器中的某个字段复制到 IRF 的对应寄存器中,用于对包头中的关键字段的保存,例如源 MAC,源 IP 等重要信息. SET 指令是直接设置 IRF 中的特定寄存器的值,例如标记报文的类型是 IPV4 还是 IPV6.在本设计中,每种协议下对应的操作指令会拼接成一条长的指令字,类似于处理器中的超长指令字,发送给

指令执行单元,同时执行多达 8 条的指令.关于指令的格式我们使用了与文献[9]相同的设计,在此不再赘述.我们对 IRF 中的数据做了存储空间预分配,即每个指令槽对应固定的目的 IRF 地址,这样可以缩短指令字长度、简化寻址逻辑.

关于 MOV 类指令,去协同情况下,源操作数是经过变码的值,所以采用的办法是同时提取数据帧的字段和对应的密钥字段,然后分别填入数据通路和密钥通路的 IRF 对应位置中.两部分字段一起沿流水线传递到后处理单元,进行解码.

SET 指令的操作数来自配置信息,而非外部的数据包,因此不需要进行变码,所以其只对数据通路执行,不需要密钥参与.

2.2.3 后处理单元

后处理单元是去协同的解码单元.经过前面各级解析后得到的关键字段和包的各种标识都被记录在了最后一级 PE 输出的 IRF 中.后处理单元的功能包括两项:对数据帧进行解码;生成最终的解码后的描述符.

数据帧解码通过将数据帧和密钥帧按位异或完成.最终解码后的描述符,分如下 2 种情况获得:在解析阶段通过 SET 指令填写到 IRF 中的结果,直接作为最终结果;提取出的关键字段(在解析阶段通过 MOV 指令填充),需要把数据通路的 IRF 寄存器和密钥通路对应的 IRF 寄存器值按位异或后得到最终结果.

3 设计实例

基于上述包解析器结构实现了一个支持常用协议的 L4 以太网交换机包解析器,支持表 2 所示的协议集合.包解析器的解析处理流水线由 5 级 PE 组成,分别对应表中的 5 个协议层.

表 2 数据包解析器支持的协议集合

解析处理层	L2	L2s	L3	L3s	L4
	EthernetII	VLAN	IP-in-IP	IPv6	TCP
协议类型	Vlan	IP-in-IP	IPv4	MPLS	UDP
		MPLS	IPv6		
			MPLS		

使用 SMIC 40nm 工艺标准单元库的 DC 综合得到的最高主时钟频率为 220 MHz,对应的处理一个数据包最短延时为 104 ns.一级 PE 的面积约为 176.28 万等效门.与文献[9]的非安全设计相比,包延时增加 21 ns,一级 PE 增加 28.72 万门.

4 安全测试

4.1 测试方案

对被测包解析器进行木马触发逻辑植入,然后通过外部数据进行木马触发,检查是否能够有效避免木马被激活。以第3节的设计为被测设计,木马逻辑被植入在数据通路的数据帧寄存器处。先后测试了如下的触发条件:

(1)简单固定特征:

特征11:数据包头部包含 Ethernet II 协议,前12字节为固定值 0x5AC300FFE48E5AC300FFE48E。

特征12:数据包头部包含至少两层 MPLS 协议,从第一层 MPLS 协议报头的起始位置开始,连续8字节为固定值 0xEEF770FF77FEE1FF。

(2)高阶特征(特征2):数据包头部包含 IPv6 协议,从 IPv6 协议报头的起始位置开始算起,第9字节到第24字节与第25字节到第40字节按位异或的结果为8个 0xA425 的组合。

4.2 测试结果及分析

特征11和特征12我们分别各测试1024条攻击样本和1024条正常数据,且每条攻击样本和正常数据均作不同密钥和相同密钥的测试。特征2共测试1024条攻击样本,每条攻击样本均作不同密钥和相同密钥的测试。上述数据均未触发木马触发逻辑。

而对无去协同逻辑的包解析器采用相同方法进行测试,所有样本均能触发木马触发逻辑。

仿真实验表明去协同技术确实可以起到提高包解析器安全性的作用。

5 结束语

利用对输入数据帧进行异或变码、在匹配环节采用模板同步变码、后处理单元进行异或解码等关键技术可以实现去协同的包解析器。实验表明该技术对不同触发条件的木马具有主动防御能力。本文工作将去协同技术推广到了非存储器功能模块,今后将进一步围绕安全高效的密钥的产生算法进行研究。

参考文献:

- [1] CHAKRABORTY R S, NARASIMHAN S, BHUNIA S. Hardware Trojan: threats and emerging solutions[C]//Proceedings of 2009 IEEE International High Level Design Validation and Test Workshop. San Francisco, USA: IEEE, 2009: 166-171. DOI: 10.1109/HLDVT.2009.5340158.
- [2] WAKSMAN A, SETHUMADHAVAN S. Silencing hardware backdoors[C]//Proceeding of 2011 IEEE Symposium on Security and Privacy. Berkeley, CA, USA: IEEE, 2011: 49-63. DOI: 10.1109/SP.2011.27.
- [3] 李唐盛晟. 以太网交换机 SoC 硬件木马防御模块及 TRNG 设计与实现[D]. 北京: 清华大学, 2018. LI T S S. Research on hardware Trojans defense module in Ethernet switch SoC and implementation of TRNG design[D]. Beijing: Tsinghua University, 2018.
- [4] HAIDER S K, JIN C L, AHMAD M, et al. Advancing the state-of-the-art in hardware Trojans detection[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 16(1): 18-32. DOI: 10.1109/TDSC.2017.2654352.
- [5] LOVE E, JIN Y R, MAKRI S Y. Proof-carrying hardware intellectual property: a pathway to trusted module acquisition[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(1): 25-40. DOI: 10.1109/TIFS.2011.2160627.
- [6] ZHANG Z M, NJILLA L, KAMHOUA C A, et al. Thwarting security threats from malicious FPGA tools with novel FPGA-oriented moving target defense[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2019, 27(3): 665-678. DOI: 10.1109/TVLSI.2018.2879878.
- [7] ZHAO Y, LI X Y, YIN S J. Reconfigurable parser for software defined network L4 Ethernet switch chip[C]//Proceeding of 2019 IEEE International Conference on Integrated Circuits, Technologies and Applications. Chengdu, China: IEEE, 2019: 81-82. DOI: 10.1109/ICTA48799.2019.9012927.

作者简介:

李翔宇 男, (1977-), 博士, 副研究员. 研究方向为硬件安全、高性能集成电路等. E-mail: xiangyuli@tsinghua.edu.cn