

一种万兆以太网在线数据收发器的设计

牛 贇, 乌力吉, 张向民

(清华大学 微电子研究所, 北京 100084)

摘 要: 一种用于万兆以太网在线安全处理器中的高速数据帧收发器的结构设计. 采用基于中断的共享式缓存加分布式缓存的二级缓冲收发机制, 实现对万兆以太网数据帧的高效在线收发. 通过改进一种调度算法的硬件实现, 完成对可变长度数据包的高效调度. 利用系统级建模语言 systemC 对设计进行建模, 通过仿真优化分布式缓存 FIFO 的数量实现丢包率为 0. 将设计应用到一款万兆在线网络安全处理器中, 实验结果表明, 时钟频率 250 MHz 下的数据吞吐率达到 10.06 Gb/s, 丢包率为 0, 验证了设计性能. 通过配置模块中分布式缓存的数量, 设计可以应用于下一代 40/100 Gb/s 网络安全处理器中满足更高速信息安全的需求.

关键词: 万兆以太网; 网络安全处理器; 调度算法; systemC

中图分类号: TN918

文献标识码: A

文章编号: 1000-7180(2014)03-0027-05

Design of an In-line Packet Transceiver for 10 Gb/s Ethernet

NIU Yun, WU Li-ji, ZHANG Xiang-min

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

Abstract: This paper deals with a high-speed in-line packet transceiver structure design for the 10 Gb/s Ethernet. By adopting a shared and distributed two-stage buffering mechanism, the high-speed packet transfer is achieved real-time. The high-efficient scheduling to the changeable length packets is realized by improving hardware implementation of a scheduling algorithm. The packet transceiver is modeled with the aid of systemC at transaction level. By simulating and optimizing the number of the distributed FIFO buffers, the packet loss rate drop to zero. The design is implemented and simulated at RTL level. To verify the design, it is used in a 10 Gb/s in-line network security processor and simulated. The simulation shows that the system can achieve 10.06 Gb/s throughput at the clock rate 250 MHz without packets loss, this verified the performance of the design. The design can be used in the next generation 40/100 Gb/s Ethernet to protect the information security through configure the number of the distributed FIFO buffers.

Key words: 10 Gb/s Ethernet; network security processor; scheduling algorithm; systemC

1 引言

网络安全处理器是解决高速网络安全的硬件实现方案. 随着万兆以太网在骨干网的广泛部署, 40/100 Gb/s 以太网标准^[1]于 2010 年推出, 研究满足万兆以太网和下一代高速网络的网络安全处理器变得尤为迫切. 网络安全处理器中高速以太网数据帧的在线收发效率直接影响网络安全处理的性能以及

网络服务质量 QoS(Quality of Service). 通常采用 FIFO 队列数据帧收发方式, 在网络数据传输速率不高的情况下, 由于数据帧的接收和内部数据处理速率之间的差异不明显, FIFO 方式尚能满足系统性能的要求. 但是, 当网络速率达到 10 Gb/s 甚至更高时, 容易造成头阻塞(Head of Line, HOL)现象发生, 导致系统性能降低 58.6%^[2], 甚至引起丢包现象. 采用虚拟输出队列(Virtual Out Queue, VOQ)

收稿日期: 2013-05-19; 修回日期: 2013-07-08

基金项目: 核高基重大专项(2011ZX01034-002-002-003)

技术^[3],将目的输出端口不同的数据帧放在不同的队列中缓存,通过使用合理的调度算法,能够 100% 发挥系统性能.但是,这种方式仅适合于需要完成对数据帧的路由,不需要对数据做任何处理的高性能路由器中.在实现网络安全处理时,对数据帧除了进行路由外,还需要对数据帧的有效载荷进行加解密和认证等计算密集型数据处理,造成高速数据帧收发与内部数据处理速率不匹配,因此,VOQ 方式也不适合用于高速特别是万兆网络安全的应用中.另外,目前的调度算法^[4-5]大多是针对固定长度数据帧实现的,对于来自以太网的不确定长度数据帧的收发,现有调度算法的均衡性和效率将受到严重影响.本文提出一种可以并行处理多个数据帧的高速数据收发器结构设计解决现有方法中存在的问题.

2 高速均衡的分配调度机制

以太网数据帧的特点一是数据帧长度可变,最小为 60 字节,最大为 1 500 字节;二是以太网数据流量分布不确定.数据帧的上述特点造成了对数据帧进行处理时间的不确定性.为了达到实时在线数据帧收发的要求,需要选择合理的分配和调度算法,以及数据交换结构.

本文采用一种共享式缓存加分布式缓存的二级缓冲数据收发机制,提出一种基于中断的动态负载主动分配算法,即当分布式缓存模块中有空闲 FIFO 时,通过产生中断信号主动向共享缓冲模块请求读入数据帧,减小了传统方法中发送者主动查询带来的查询周期随着查询分布式缓存数量的增加线性增加,从而影响性能的缺点.具体算法描述如下:

第一步:初始化.共享缓存接收来自高速以太网的数据帧,依次分配数据帧给各个分布式缓存中的 FIFO 缓存.

第二步:中断.当分布式缓存中的 FIFO 空闲时,分布式缓存主动产生中断指示信号,共享缓存根据该信号判断是否存在空闲的 FIFO 缓存器.

第三步:分配.共享缓存通过总线轮询算法完成数据帧分配.如果分布式缓存只有一个中断请求信号,则对该中断请求信号进行处理,把数据帧分配给该请求信号对应的 FIFO 中.如果当前没有中断请求信号,等待直到出现中断请求信号.如果同时有多个中断请求信号,考虑上一个请求信号:如果上一个请求信号是第 i 个 FIFO,则数据帧分配的顺序是 $i+1, \dots, n, 1, 2, \dots, i$;以此类推.

设计中的数据交换采用基于多次迭代 SLIP 调度算法(iSLIP)^[6]的可配置交叉互联结构,实现多通道数据并行传输和多数据帧高效公平调度,算法分为请求,允许,接受三步:

①当多个内部处理单元空闲时,空闲的处理单元分别向存有数据帧的分布式缓存中的 FIFO 发送请求信号;

②接收到请求信号的 FIFO 在所有请求的内部处理单元中,选择当前优先级最高的处理单元发出允许信号,其余处理单元不作响应;

③发送请求信号的处理单元接收到允许信号后,选择当前优先级最高的 FIFO 并发出接受信号,之后开始读取该 FIFO 中的数据帧;

④没有得到接受信号的 FIFO 继续向当前优先级最高的处理单元发出允许信号,直至该 FIFO 被接受.

从上述两种算法可知,当分布式缓存中有空闲的 FIFO 时,下一个数据帧才能接收,因此 FIFO 中同时只存有一个数据帧,解决了传统方法中一个 FIFO 中可能同时存在多个数据帧造成的头阻塞问题.另外,由于 iSLIP 调度算法的交叉互联结构保证了多个数据帧的并行高效传输以及公平合理调度,因此,分布式缓存 FIFO 中的数据帧能得到快速处理,进一步减小了头阻塞的几率.通过配置优化交叉互联结构中端口数量,可以实现丢包率为 0.

3 高速在线收发器结构设计

万兆以太网在线数据收发器结构如图 1 所示,结构包括共享缓存模块、分布式缓存模块和数据帧调度器.其中共享缓存模块包含一个异步 FIFO 实现数据帧的一级缓存,分布式缓存模块中包括多个同步 FIFO 实现数据帧的二级缓存.共享缓存模块和分布式缓存模块间通过基于中断的均衡分配算法实现数据帧的均衡分配.数据帧调度器完成数据帧的高效公平调度.

3.1 建模仿真

通过对分布式缓存模块中的 FIFO 数量进行优化,使得数据帧丢包率为 0,实现 10 Gb/s 高速以太网数据的在线收发.为此,建立高速在线收发器的拓扑模型(如图 2 所示),在该模型中, $F_i(1 \leq i \leq n)$ 表示共享缓存队列中待处理的数据帧; S 表示数据帧分配机制,作用是将待处理的数据帧均衡地分配到 n 个缓冲模块中; $P_j(1 \leq j \leq n)$ 表示系统内部处理单元.建模的目的是优化和确定分布式缓冲模块

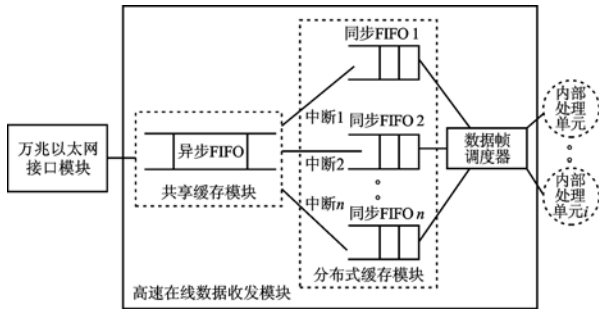


图1 万兆在线数据收发器结构图

中FIFO的个数 n ,确定共享缓冲队列和分布式缓冲模块中FIFO的容量,使得丢包率为0。

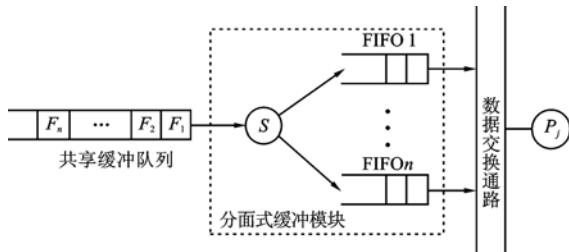


图2 高速数据收发器模型

本文采用基于systemC的事务级建模方法对图2所示的拓扑结构进行建模和仿真。systemC语言采用接口、端口和通道机制,模块之间采用接口调用方法进行通信,是一种适合事务级建模的系统建模语言^[7]。

建模过程中,把一次数据帧的传输当做一次事务级交易,功能由systemC的模块实现,模块之间的通信由通道来实现,并通过接口实现模块间的连接。建模时数据帧分配机制S采用提出的基于中断的动态负载主动分配算法,根据算法可知,分布式缓冲模块中的FIFO的容量为一个最大数据帧的长度字节,共享缓冲队列的容量大小与分布式缓冲模块中缓冲队列的个数 n 有关,为 n 倍的最大数据帧的长度。

仿真时,利用systemC Verification Library (SCV)产生服从泊松分布,长度从64字节到1500字节的测试数据帧,对模型进行事务传输级仿真;通过配置 n 的大小,使得丢包率为0,得到优化的 n 值。图3为 n 与丢包率的关系图。从图中可以得到 n 的最优值为16。

3.2 设计实现

基于对高速收发器模型仿真结果,完成图1中各个模块的RTL级结构设计。模块中异步FIFO和同步FIFO采用自主开发的基于SRAM的同步和异步FIFO设计。这些FIFO的设计都基于数据帧,并通过设置上下流水线控制进出FIFO的流量。

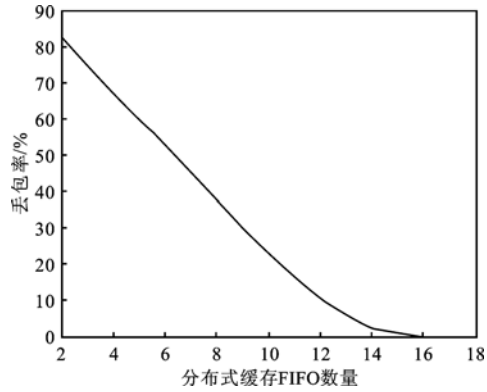


图3 分布式缓存FIFO数量与丢包率关系

图4为设计中各个模块间的接口信号示意图。当pkt_rx_avail信号为高时,表明高速以太网接口模块中有数据帧,共享缓存模块首先根据分布式缓存的Interrupt信号,判断分布式缓存模块中是否有空闲的FIFO缓存,如果有,共享缓存模块则将pkt_rx_ren信号置高,表示可以接收数据帧。共享缓存模块根据Interrupt信号的状态,按照总线轮询的分配算法选择分布式缓存模块中的FIFO缓存,并将对应FIFO的Wren信号置高。分布式缓存模块中对应的FIFO读入数据帧,当Valid_out信号变为低时,Wren信号置为低,表明读完一个数据帧。

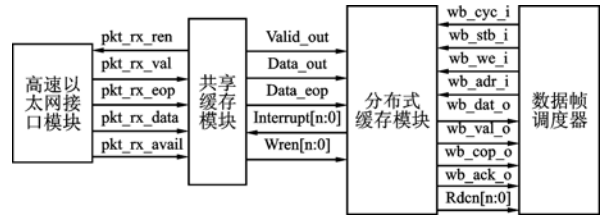


图4 模块间接口信号示意图

分布式缓存模块中如果有空闲FIFO缓存,将该空闲FIFO缓存对应的Interrupt信号置为高表明该FIFO缓存可以接收数据帧,Wren有效(高电平)时开始接收数据帧,并将Interrupt置低;当FIFO缓存器读完一个数据帧时,对应FIFO缓存器的Rden信号置高;内部处理模块根据Rden信号,向分布式缓存模块发出请求并通过数据帧调度器选择FIFO缓存,从选择的FIFO缓存中读取数据帧;当内部处模块读取数据帧结束后(wb_eop_o=1),分布式缓存模块中对应的FIFO缓存将Interrupt置高,表明该FIFO缓存器处在空闲状态。图5为模块间控制信号的典型仿真波形示意图。

在设计中,为解决输入数据帧长度不确定的带来的设计复杂度增加问题,对采用的Wishbone总线协议进行了改进,加入数据帧有效的控制信号wb_val_o和数据帧尾控制信号wb_eop_o,省去了传统

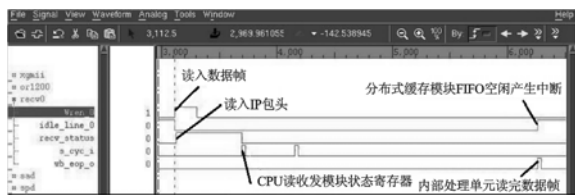


图5 典型仿真波形示意图

FIFO 队列方法中的计数器,减小了设计实现的复杂度。

在对 iSLIP 算法硬件实现时,通过增加数据帧包头和包尾的控制信号,实现对可变长度数据帧的高效调度.采用一种快速轮询仲裁算法实现仲裁器硬件设计^[8]有效减小了仲裁延迟时间,提高了调度器的数据传输速率。

设计中分布式缓存模块中的同步 FIFO 容量大小设计为存放一个最大数据帧的长度 1500 bytes.由建模仿真得到 FIFO 的数量为 16 个,因此,共享缓存模块的异步 FIFO 的容量大小为 24 kbytes.通过 Synopsys 综合工具 Design Compiler 在 SMIC 65 nm 工艺下进行综合比较,该设计的面积比传统 FIFO 队列的方法减小 3%。

另外,通过上述各个功能模块之间控制信号的握手协议,数据帧的收发可以得到有序合理控制,不会造成数据帧得不到及时处理造成头阻塞现象发生,系统的资源利用率可以达到 100%,保证了网络服务质量。

4 仿真验证

设计采用 Verilog 硬件描述语言设计实现,并集成到图 6 所示的一款自主开发的万兆在线网络安全处理器中进行仿真验证.该网络安全处理器系统在一个 SoC 系统上实现 IPsec 网络安全协议处理.系统主要包括一款 32 位 RISC 嵌入式 CPU,高速以太网数据帧收发器(本设计),万兆以太网接口模块,安全联盟数据库 SAD (Security Association Database)SRAM,高速数据库 SPD(Security Policy Database)查找模块,高性能可配置 IPsec 协议处理器、片外算法接口模块,SRAM 控制器。

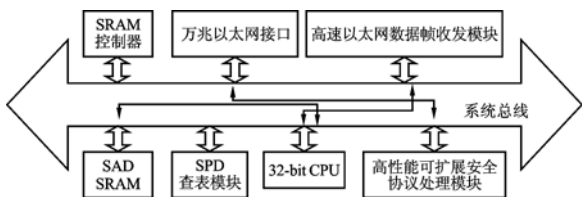


图6 万兆在线网络安全处理器结构

其中,32 位嵌入式 CPU 完成查询输入高速在线数据收发器的状态,解析 SPD 查表模块送出的描述符,对 IPsec 协议处理器进行任务分配和调度;万兆速以太网接口模块完成网络数据帧封装和解封装.高速以太网数据帧收发器完成对输入和输出数据帧接收和发送.高速查表模块对数据帧提取选择符、查找 SPD 数据库,并对安全策略进行解析生成描述符.高性能可配置 IPsec 安全协议处理器实现 IPsec 网络安全协议.该万兆在线网络安全处理器可以应用在高速核心路由器中实现虚拟专用网络 (VPN: Virtual Personal Network)。

利用 65 nm 标准单元库对该系统进行软硬件协同综合仿真,结果表明,在时钟频率为 250 MHz 时,万兆在线网络安全处理器的数据处理吞吐率为 10.06 Gb/s,丢包率为 0,验证了高速以太网数据帧收发器的性能。

基于 Xilinx FPGA 开发板搭建了系统验证平台,验证的功能模块如图 7 所示.测试数据通过开发板上的 UART 接口与 PC 机端相连接,通过 PC 端进行测试数据的发送和接收,进一步验证了高速以太网数据帧收发器的功能正确性。

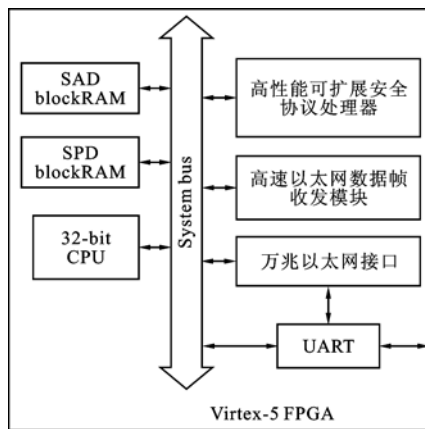


图7 验证结构示意图

5 结束语

文中在分析现有网络数据收发方法存在问题的基础上,提出了一种用于万兆在线网络安全处理器中的高速数据帧收发器结构,该结构中采用了基于中断的二级缓冲的数据收发机制,通过建模仿真优化分布式缓存模块中 FIFO 的数量,得到满足万兆在线网络安全处理器性能的最优值,有效解决万兆以太网数据帧收发时头阻塞问题.对该结构进行硬件设计并应用到一款万兆在线网络安全处理器中进行仿真验证.结果表明,该在线网络安全处理器在

250 MHz 时钟频率下的数据吞吐率可以达到 10.06 Gb/s,并且丢包率为 0,从而验证了该设计可以有效解决高速数据收发和内部数据处理速率不匹配的矛盾,满足万兆以太网网络环境下对高速数据帧的接收和发送需求.通过配置分布式缓存模块中 FIFO 的数量,设计还能够实现下一代 40Gb/100 Gb/s 以太网在线安全处理器中高速数据帧的收发,满足下一代高速网络下的信息安全需求.

参考文献:

- [1] IEEE. IEEE Std. 802.3-2010, LAN/MAN Standards [S]. NY: IEEE Computer Society, 2010.
- [2] Karol M, Hluchyj M, Morgan S. Input versus output queuing on a space division switch[J]. IEEE Trans. on Communications, 1987, 35(12): 1347-1356.
- [3] Elhanany Itamar, Arazi Ortal, Kahane Michael. Virtual input queued packet switches with non-uniform arrivals and bursty service[C]//Proceedings of the 2004 11th IEEE International Conference on Electronics, Circuits and Systems. Israel: IEEE Computer Society, 2004: 93-96.
- [4] Anderson, Owicki. High speed switch scheduling for local area networks[J]. ACM Trans. on Computer Systems, 1993, 11(4): 319-352.
- [5] 张磊磊. 交换机交叉开关调度算法的 ASIC 实现[M]. 成都: 电子科技大学, 2008.
- [6] McKeown Nick. iSLIP scheduling algorithm for input-queued switches[J]. IEEE Trans. on Networking, 1999, 7(2): 188-201.
- [7] David C Black. SystemC: from the ground up[M]. 2nd Ed. New York: Springer, 2010.
- [8] Gupta P, McKeown N. Designing and implementing a fast crossbar scheduler[J]. IEEE Micro, 1999, 19(1): 20-28.

作者简介:

牛 赞 女, (1976—), 博士研究生. 研究方向为高性能网络安全芯片研究与实现.

乌力吉 男, (1965—), 博士, 副教授. 研究方向为信息与网络安全.

张向民 男, (1966—), 硕士, 助理研究员. 研究方向为信息与网络安全.

(上接第 26 页)

作者简介:



袁小霞 女, (1988—), 硕士研究生. 研究方向为电子系统软件测试.

李建红 男, (1986—), 硕士研究生. 研究方向为嵌入式系统.

刘 丹 男, (1981—), 博士, 副教授. 研究方向为网络安全、分布式并行计算.

周 明 男, (1977—), 硕士. 研究方向为信息安全、软件测评.