

基于 FCoE 和 RapidIO 的高速数据交换网设计与实现

贺琨¹, 刘旭波², 刘毅¹, 梁超¹, 郭日成¹

(1. 中国船舶工业系统工程研究院, 北京 100036; 2. 海军研究院, 北京 100073)

摘要: 为适应开放式体系架构信息系统中对海量数据的高速率、低延时和远距离传输需求, 设计了一套基于以太网光纤通道 (FCoE) 和 RapidIO 技术的高速数据交换网络, 网络由光纤交换机和高速接口模块组成。光纤交换机作为数据交换核心实现 FCoE 协议转换, 高速接口模块完成 RapidIO 与 FCoE 接口间数据转换。结合实际工程应用需求, 对高速数据交换网的络功能和性能进行了实验验证, 结果表明: 网络数据交换带宽最大可达 800 MB/s, 并具备低时延、远距离传输和节点规模灵活扩充的特点。

关键词: 以太网光纤通道; RapidIO; 高速数据交换网络; 光纤交换机; 高速接口模块

中图分类号: TN97 **文献标识码:** A **文章编号:** 1002-5561(2019)10-0057-06

DOI: 10.13921/j.cnki.issn1002-5561.2019.10.012

开放科学 (资源服务) 标识码 (OSID):



Design and implementation of high-speed data exchange network based on FCoE and RapidIO

HE Kun¹, LIU Xubo², LIU Yi¹, LIANG Chao¹, GUO Richeng¹

(1. China State Shipbuilding Corporation Limited Systems Engineering Research Institute, Beijing 100036, China;

2. Navy Equipment Research Institute, Beijing 100073, China)

Abstract: In order to meet the open architecture information systems high-speed, low-latency and long-distance data transmission requirement, a high-speed data exchange network based on fibre channel over ethernet (FCoE) and RapidIO technology is designed. it consists of a fiber switch and a high-speed interface module. The optical fiber switch completes the FCoE protocol conversion. the high-speed interface module completes the data conversion between RapidIO and FCoE interfaces. Combined with the actual engineering application requirements, the high-speed data exchange network have carried out functional and performance testing experiments. The results show that the network data exchange bandwidth can reach up to 800 MB/s, and it has the characteristics of low latency, long-distance transmission and flexible expansion of node scale.

Key words: fibre channel over ethernet; RapidIO; high-speed data exchange network; optical fiber switch; high speed interface module

0 引言

随着宽带/超短带、高速跳频扩谱等新体制信号的实际使用, 前端传感器数据呈现爆炸式增长, 数据传输和处理正处在一个蓬勃发展和融合阶段, 而高速数据交换网作为数据通信交换核心单元, 正面临着高效

率、高带宽和低延时的通用要求。

以太网光纤通道 (FCoE)^[1] 技术是在以太网网络协议包基础上封装光纤通道帧, 允许光纤通道 (FC) 使用 10 Gb/s 以太网网络协议, 实现了通过以太网承载光纤通道报文, 使得 FC 和以太网局域网可共享同一个单一的或集成的网络基础设施, 很好地解决了不同类型网络共存所带来的问题, 同时保持 FC 数据速率不变的无损以太网技术。RapidIO^[3-5] 技术是一种基于包交换的交叉开关互连技术, 具备高带宽、低延时、高效率及高可靠性的特性, 主要功能包括系统内高性能的嵌入式微处理器、数字信号处理 (DSP) 模块、现场可编程门阵列 (FPGA) 模块和外设之间高速的数据传输通信, 满

收稿日期: 2019-03-19。

作者简介: 贺琨 (1986-), 男, 四川成都人, 硕士, 工程师。主要研究方向是舰船电子对抗, 负责了国内舰载情报侦察领域重大型号科研任务, 在系统总体设计、通用架构设计、信号和情报处理方面有着深入的理论研究和实践经验。



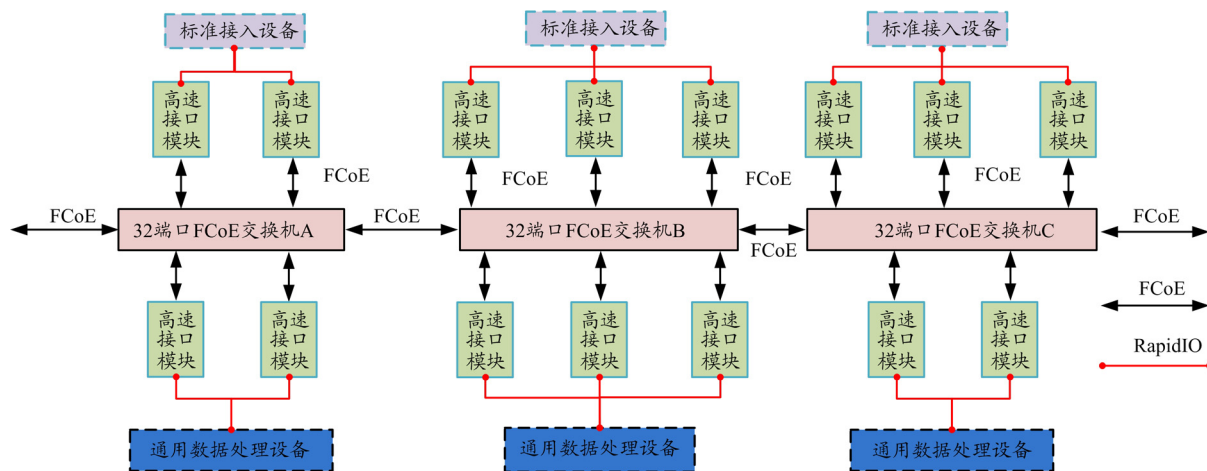


图1 高速数据交换网整体架构

足用户对数据高速实时传输的需求。目前,在实现芯片到芯片之间、板与板之间的高速互连上,RapidIO为内部互连通信提供了良好的解决方案^[2,6]。

结合FC和高速总线技术搭建一套通用的高速数据交换网络,是解决当前实际需求的一条有效的途径。本文设计一种基于FCoE和RapidIO技术的高速数据交换网络。

1 高速数据交换网络设计与实现

1.1 整体架构设计

高速数据交换网络的互联示意图如图1所示,共有3台32端口光纤交换机级联,支持网络中各个标准接入设备和通用数据处理设备数据高速传输与交互,并支持多交换机扩展。标准接入设备内部都是RapidIO总线互联,外部接口符合RapidIO接口规范,通过高速接口模块采用RapidIO总线接入高速数据交换网络,在高速接口模块中实现数据的光/电转换和传输,并接入高速数据交换机。光纤交换机通过FCoE协议完成数据传输,数据传输至高速接口模块,实现FC数据至RapidIO数据交换,并可引接至通用数据处理设备对应RapidIO节点进行数据的特定处理^[7-11]。

1.2 高速数据交换机设计

高速数据交换机由配置管理处理器、交换芯片和光/电转换模块等单元组成。电源模块实现220V交流转换成12V直流输出,为配置管理处理器单元、交换芯片单元和光/电转换模块单元供电;配置管理处理器单元集成了内存、闪存Flash、1个调试串口、1个配置网口和1路PCIe总线接口,实现交换机芯片、端口物理层(PHY)芯片的初始化与配置;交换芯片单元实现

32路FCoE交换功能,PHY芯片、光/电转换模块提供物理通道;配置管理处理器通过1路x2 2.5 Gb/s PCIe接口与交换芯片互连对交换芯片进行初始化配置、管理,通过1路调试串口、1个调试网口实现人机交互接口;交换芯片提供32路万兆以太网交换功能,通过挂载4路FCoE PHY芯片实现FCoE接口,再经光/电转换模块实现光纤接口,引到机箱面板。整机对外接口为1路220V交流电源输入、1路电源开关、1个复位、1个配置串口、1个配置网口、1个电源状态指示灯、1个运行状态指示灯、32路FCoE光纤接口及其指示灯。高速数据交换机逻辑结构如图2所示。

高速数据交换机的软件主要包括嵌入式操作系统、交换机驱动、配置管理和健康管理4个方面,如

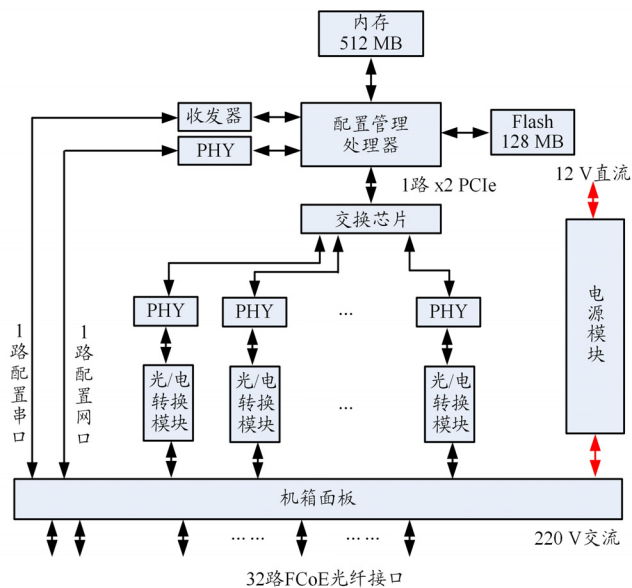


图2 高速数据交换机逻辑结构图

图 3 所示。操作系统选用嵌入式操作系统 Vxworks6.5, 提供 SDK 5.x API, 实现万兆以太网交换芯片和 PHY 芯片的驱动和配置管理。初始化代码、操作系统软件、驱动程序、配置管理和健康管理等软件都固化在板载 Flash 中。加电后, 处理器先从 Flash 加载相应软件, 进行初始化, 引导操作系统启动, 加载驱动程序, 启动配置管理和健康管理相关应用程序, 交换机正常工作。交换机配置管理和健康管理的人机交互通过调试串口或调试网口实现。

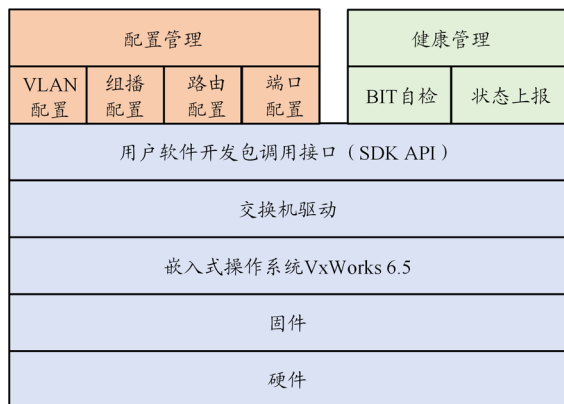


图 3 高速数据交换机软件结构图

1.2.1 交换机驱动

高速数据交换机的驱动设计重点在于实现万兆以太网和 FCoE 协议的数据交换, 即利用以太网的 PHY 及数据链路层替代 FC 协议的物理变换层及传输协议层 (FC-0、FC-1), 并通过网络层 (FC-2) 的一个子层实现底层以太网与上层 FC 层的协议映射。在 FCoE 协议栈的上层, 可实现与 FC 协议的完全兼容, 从而既可以实现原本万兆以太网的数据交换, 又能实现 FCoE 协议之间的数据交换。

1.2.2 配置管理

配置管理软件主要实现交换机的细分功能配置, 如虚拟局域网 (VLAN)、组播、路由和端口配置管理等。通过配置交换控制器, 将配置信息写入交换控制

器相应的寄存器, 从而实现对 VLAN、组播、路由和端口等功能的配置管理。

1.2.3 健康管理

健康管理软件包括机内自检 (BIT) 和状态上报。BIT 自检是指系统对交换机的电源模块、配置管理处理器单元和交换芯片单元等进行上电自检、周期自检和维护自检。状态上报是指对交换机的 BIT 自检结果通过配置端口上报。

1.3 高速接口模块设计

高速接口模块由 PowerPC 处理器和相关外设组成, 采用了新一代高速串行总线 VPX 标准。背板插件接口分为 VPX RP0、RP1、RP3 和 RP4, 管脚定义按照 VITA 总线 SRIO 接口定义。VPX RP0 提供板载 +12V DC 供电输入、I²C 接口和复位接口引接。VPX RP1、RP3 和 RP4 接口为 4x 串行 RapidIO 接口 (SRIO), 速率 (2.5 Gb/s 或 5 Gb/s) 可配置, 遵循 RapidIO Interconnect Specification Revision 1.3。前面板接口由 2 路光纤、1 路网口和 1 路串口组成; 2 路光纤接口是由处理器通过 PCIe 接口上挂接 FCoE 控制器, 通过连接光/电转换模块实现数据转换功能; 1 路网口由处理器通过吉比特介质独立接口 RGMII 连接 PHY 芯片完成; 1 路串口由处理器通过 UART 串行通信口外接 RS232 收发器实现, 通过配置串口、网口实现对接口模块的转换参数设置、信息上报和路径映射等。高速接口模块的逻辑结构如图 4 所示。

高速接口模块运行 Linux 操作系统, 接口模块的软件设计主要包括 FCoE 协议实现、协议转换和动态配置 3 个方面。

1.3.1 FCoE 协议实现

FCoE 协议实现主要包括 FCoE 初始化协议 (FIP) 和 FCoE 协议两部分。设计的协议层次如图 5 所示。物理链路层 (PHY、MAC、FCoE Mapping)、网络层 FC-2、服务层 FC-3 和应用层 (小型计算机系统接口 SCSI 和 SCSI-ML) 是现有 Open-FCoE-SCSI 协议栈的实现层

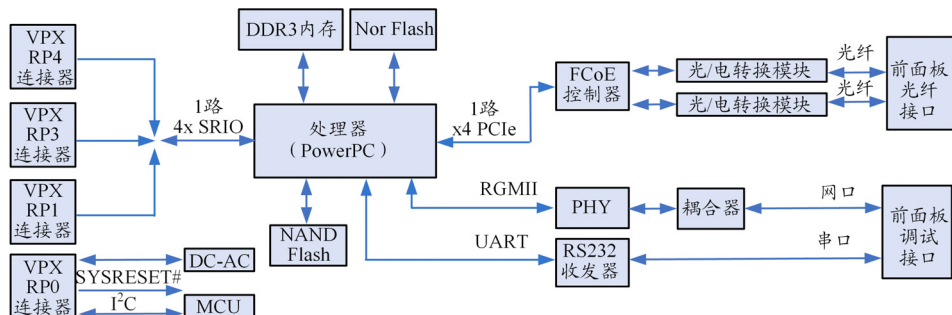


图 4 高速接口模块结构图



图 5 高速接口模块协议层次结构图

次关系。FC 协议原始帧 (FC-RAW)、扩展链路服务 (FC-ELS) 为在此基础上为支持 FCoE 通信协议需要开发实现的部分。整个协议需要设计实现独立的 FC-RAW 的管理接口, 实现通信协议对 FC-ELS 的支持要求, 进而实现对 FIP 的支持, 能够登录高速数据交换机。

FCoE 协议数据交换是以一种全双工、顺序发送方式进行, 在数据可靠性传输方面, 本设计采用了基于流控机制 (PFC) 和传输选择机制 (ETS) 来增强网络避免丢包。其中, PFC 是在同一物理链路中多业务数据传输条件下, 仅对单一拥塞网络的业务数据进行停传机制, 确保其它业务数据交互不受影响; ETS 是一种对链路带宽进行最佳分配的管理机制, 支持根据业务流量情况动态分配业务数据传输链路的带宽。

1.3.2 协议转换

协议转换软件主要实现 RapidIO 协议到 FCoE 协议的转换, 即将 RapidIO 的大块通信与消息门铃机制转换成 FCoE 的消息传输。应用层发送和接收 RapidIO 总线数据包时, 需遵循用户协议进行组包/解包, 单包数据包括了 RapidIO 头部、Payload 字段和 RapidIO 尾部, 其中 Payload 字段是用户应用数据核心字段, 同时约束了交互时采用的链路以及目的地址等信息, 目的地址包括 FCID 和 RapidID 编号信息。

高速数据接口模块共设立了 8 个虚拟链路, 发送通道和接收通道各 4 组, 分别使用 DMA 方式建立内存缓冲区, 缓冲区地址与 RapidIO 总线地址建立了映射关系。FCoE 数据组包过程是: 当主控线程在 RapidIO 总线数据到达时, 获取 Payload 字段数据, 按照链路选择字段编号将数据链路数据与线程池交互; 线程池子线程获取数据后, 解析目的地址中包含的 FCID 和 RapidID 编号, 转换成 FCoE 帧格式中的目的地址 MAC DA, 并将整个 Payload 字段写入到 FCoE 数据帧字段中, 完成数据组包; 数据解包过程相对简单, 对协议字段 FCoE 数据帧进行数据解析即可。

1.3.3 动态配置

在 RapidIO 转换成 FCoE 的接口模块上, 通过配置接口实现 FCoE 的数据流向配置; 在 FCoE 转换成 RapidIO 的接口模块上, 通过配置接口实现 RapidIO 向综合信息处理机内部 CPU 数据的分发, 以及通过配置接口可获取模块参数及配置信息。

1.4 软件及通信协议设计

高速数据交换网络支持的标准接入设备通信接口主要为 SRIO, 不同分机单元 (例如 FPGA、DSP 或

PowerPC 等 SRIO 节点) 只要是遵循 RapidIO 规范, 按照约定的数据格式即可实现与高速接口模块的通信, 并通过高速接口模块实现将 SRIO 数据透明分发至指定通用信息处理设备节点。

1.4.1 RapidIO 初始化及 ID 分配

标准接入设备通过 RapidIO 接口接入 FCoE, RapidIO 节点的初始化可由标准接入设备或者由高速接口模块完成, 各个 RapidIO 节点的 ID 模块需唯一标识、相互区分, 单个高速接口模块支持最大 255 个 RapidIO 节点。

1.4.2 缓冲区设置

高速交换模块独立划分 128 MB 的 DMA 缓冲空间, 用于与标准接入设备和通用处理设备进行数据快速传输交互, 单个缓冲区大小为 1 MB, 缓冲区 RapidIO 地址计算是基地址与偏移地址之和, 在进行多个 SRIO 多种数据节点进行通信时, 使用多线程进行缓冲区的维护^[12-15]。

1.4.3 RapidIO 接入及通信方式

高速接口模块的 RapidIO 软件接口提供 3 种通信方式, 分别为读 (Nread)、写 (Nwrite) 和消息 (Message) 与门铃 (Doorbell) 组合使用, 3 种通信方式只存在主从发起者差别, 数据传输流程一致, 具体 3 个 RapidIO 节点通信流程如图 6 所示。

Nread 和 Nwrite 通信方式适合大块数据通信, 可以将应用层大数据块批量发送到指定的 RapidIO 节点, 同时能够批量接收大块数据, 以使用户处理。采用 RapidIO 的 Nread 和 Nwrite 通信时, 接收端将本地缓存地址通知发送端, 发送端收到缓存地址后, 采用 Nwrite 方式将有效数据部分发送给接收端, 然后发送端通过发送 Doorbell 机制通知接收端。接收端收到数据后, 重复上述工作, 循环处理后完成发送端和接收端的 RapidIO 通信。高速接口模块提供 rioNread、rioNwrite 函数。

Message 通信方式用于小块数据的接收和发送, 其最大的接收和发送长度为 4 kB, 类似于 UDP 通信方式, 其发送和接收需要 4 个参数, 分别为 localport、destid、buff 和 len。高速接口模块提供 rioSendMsg、rioRecvMsg 函数。

2 实验测试与结果

本文依托舰载情报侦察实际应用需求, 结合其通用接入设备特性构建应用场景, 并开展实验测试验证。单个通用接入设备存在 2 种接入高速数据交换网

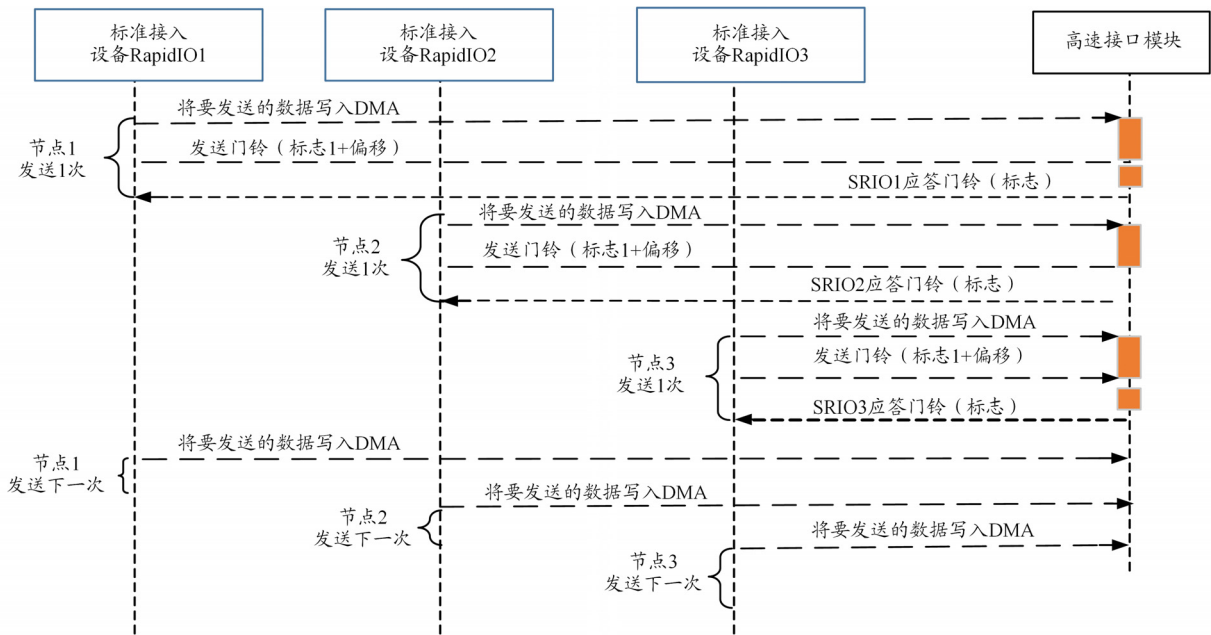


图 6 高速接口模块接入通信方式

络方式,即高速接口模块分别与单个、多个 SRIO 发送节点互联。为充分测试验证高速数据交换网性能,本文采用了单个通用接入设备、不同数量 SRIO 发送节点的接入方式,同时分发数据至后端 2 个通用处理设备节点的信息处理应用构件进行数据处理。具体连接关系结构如图 7 所示。

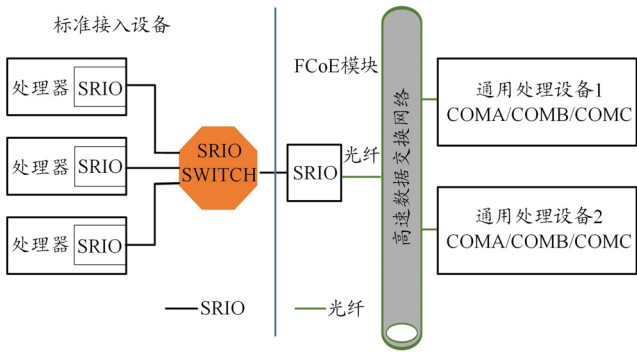


图 7 试验测试连接框图

根据实际应用场景,本文分别测试了单、双和三 RapidIO 节点同时发送不同包大小的数据至通用处理设备节点,测试方案如表 1 所示。

按照高速数据交换网络的使用流程,实验统计高速数据交换网络实际使用场景下的发送时延,定义第 n 包数据发送时刻为 t_n , 连续 2 次数据发送的时间间隔为发送时延 τ , 则发送时延 $\tau = t_{n+1} - t_n$, 传输时延定义为发送时间的平均值。

表 1 实验测试项目

测试项目	测试节点	测试条件包大小			
单 RapidIO 节点性能	单节点	1 MB	512 kB	256 kB	128 kB
双 RapidIO 节点性能	节点 1	1 MB	512 kB	256 kB	128 kB
	节点 2	1 MB	512 kB	256 kB	128 kB
三 RapidIO 节点性能	节点 1	1 MB	1 MB	128 kB	
	节点 2	512 kB	128 kB	128 kB	
	节点 3	128 kB	128 kB	128 kB	

2.1 单 RapidIO 节点传输性能实验

单 RapidIO 节点情况下,分别推送数据包大小 1 MB、512 kB、256 kB 和 128 kB 进行测试,开展了 10 万级以上传输次数交互实验,在网络正常传输不丢包的情况下,传输 1 MB、512 kB、256 kB 和 128 kB 数据包时的传输时延分别为 1206 μ s、974 μ s、892 μ s 和 479 μ s。

2.2 双 RapidIO 节点传输性能实验

双 RapidIO 节点情况下,按照节点 1/2 分别推送至通用处理设备信息处理构建 COMA/COMB 数据传输场景,开展 10 万级以上次数实验。在节点 1/2 分别同时传输 1 MB/1 MB、1 MB/512 kB、1 MB/256 kB 和 1 MB/128 kB 数据包,不丢包情况下,节点 1/2 的传输时延分别是 2745 μ s/2745 μ s、2674 μ s/1426 μ s、2670 μ s/1035 μ s 和 2677 μ s/921 μ s。

2.3 三 RapidIO 节点传输性能实验

三 RapidIO 节点情况下, 按照节点 1/2/3 分别推送至通用处理设备信息处理构建 COMA/COMB/COMC 传输场景, 开展 10 万级以上传输次数实验。节点 1/2/3 分别传输 1 MB/512 kB/128 kB、1 MB/128 kB/128 kB 和 128 kB/128 kB/128 kB 数据包, 不丢包时节点 1/2/3 传输时延分别是 2764 μ s/2687 μ s/1530 μ s、2678 μ s/1477 μ s/1476 μ s 和 1378 μ s/1389 μ s/1389 μ s。

在此基础上, 计算高速数据交换网络传输速率指标, 评估和分析高速数据交换网络的传输能力, 给出单位时间内高速数据交换网络最佳传输次数。传输速

率 $V=P_B \times P_C / S_T$, 即单位时间内标准接入设备数据发送带宽。其中, P_B 表示单包大小, P_C 表示包总数, S_T 表示发送统计时间, 实验结果如表 2 所示。在测试场景条件下, 单 RapidIO 节点最大传输速率可达 829 MB/s, 双 RapidIO 节点最大传输速率可达 728 MB/s, 三 RapidIO 节点最大传输速率可达 628 MB/s。

3 结束语

本文采用了 FCoE 和 RapidIO 总线技术, 设计实现了一种高速数据交换网络, 网络节点间传输带宽可达 800 MB/s, 核心部件包括了光纤交换机和高速接口模块, 具备接入标准化、接口通用化、软件易用化、规模灵活扩展和成本低廉等优势, 具备极大的应用推广价值。

参考文献:

- [1] ZENG Yingni, CHEN Qingchun, WEN Guangliang. Traffic Forwarding Of NPV Switch Based On FCOE [J]. Computer Applications and Software: 2017, 34(12): 189–195.
- [2] ZHU Daoshan. Design and Implementation of Block Data Transmission Based on RapidIO[J]. Modern Radar: 2017, 39(9): 29–32.
- [3] 杨剑. 光纤通信网络传输技术研究 [J]. 通信设计与应用, 2017(3): 77–78.
- [4] 郎为民, 安海燕, 姚晋芳. 大数据中心 SAN 和以太网融合问题研究 [J]. 电信快报, 2018(1): 2–7.
- [5] 杨玻, 王婷, 朱守. 基于星型结构的 RapidIO 交换网络配置方法研究 [J]. 计算机应用, 2018, 48(4): 89–92.
- [6] 李超. RapidIO 高速互联技术研究[J]. 现代导航, 2017(5): 385–390.
- [7] 黎莲花, 农祖技, 曾标. 基于 USB GPIF 模式的高速采集系统的设计与研究[J]. 仪器仪表与分析监测, 2017(1): 26–30.
- [8] 李金猛, 周勇军, 潘庆国, 等. 多通道高速大容量数据采集仪的研制 [J]. 测控技术, 2017, 36(1): 26–36.
- [9] 丁志中, 王定良, 傅银玲, 等. 超高速移动空 空宽带通信网及其物理层关键技术[J]. 数据采集与处理, 2015, 30(4): 725–732.
- [10] 马志刚, 朱思敏, 刘文怡. 基于 LVDS 的高速图像数据存储系统设计 [J]. 数据采集与处理, 2013, 28(3): 45–48.
- [11] 郭慧玉. 基于 CRC 直驱表法的高速数据远距离传输方案的设计与实现[J]. 仪器仪表工程, 2018, 36(3): 57–61.
- [12] 张芳菊. 无线接收机中高速 DMA 数据传输通道的设计与实现[D]. 成都: 电子科技大学, 2018.
- [13] 汪舒. 高速 PCIe 传输 FPGA 设计与 KMDF 驱动实现[D]. 重庆: 重庆大学, 2018.
- [14] 王瀚骏. 无人机高速数据传输技术研究与实现[D]. 成都: 电子科技大学, 2018.
- [15] 张沁馨. 基于 PXIe 高速接口的高速数据传输平台的设计与实现 [D]. 成都: 电子科技大学, 2018.

表 2 试验测试数据结果

测试项目	测试顺序	发送端/接收端	数据包大小	传输时延 (μ s)	传输速率 (MB/s)	传输次数 (次/s)
单 RapidIO 节点性能	1	节点 1/COMA	1 MB	1206	829	829
	2	节点 1/COMA	512 kB	974	513	1026
	3	节点 1/COMA	256 kB	892	280	1121
	4	节点 1/COMA	128 kB	479	261	2087
双 RapidIO 节点性能	1	节点 1/COMA	1 MB	2745	364	364
		节点 2/COMB	1 MB	2745	364	364
	2	节点 1/COMA	1 MB	2674	373	373
		节点 2/COMB	512 kB	1426	350	700
	3	节点 1/COMA	1 MB	2670	374	374
		节点 2/COMB	256 kB	1035	241	966
	4	节点 1/COMA	1 MB	2677	373	373
		节点 2/COMB	128 kB	921	135	1085
三 RapidIO 节点性能	1	节点 1/COMA	1 MB	2764	361	361
		节点 2/COMB	512 kB	2687	186	372
		节点 3/COMC	128 kB	1530	81	653
	2	节点 1/COMA	1 MB	2678	373	373
		节点 2/COMB	128 kB	1477	84	677
		节点 3/COMC	128 kB	1476	84	677
	3	节点 1/COMA	128 kB	1378	90	725
		节点 2/COMB	128 kB	1389	90	720
		节点 3/COMC	128 kB	1389	90	720