电子系统设计
Electronic Design-China
设计实现

利用RapidIO重设计下一代网络

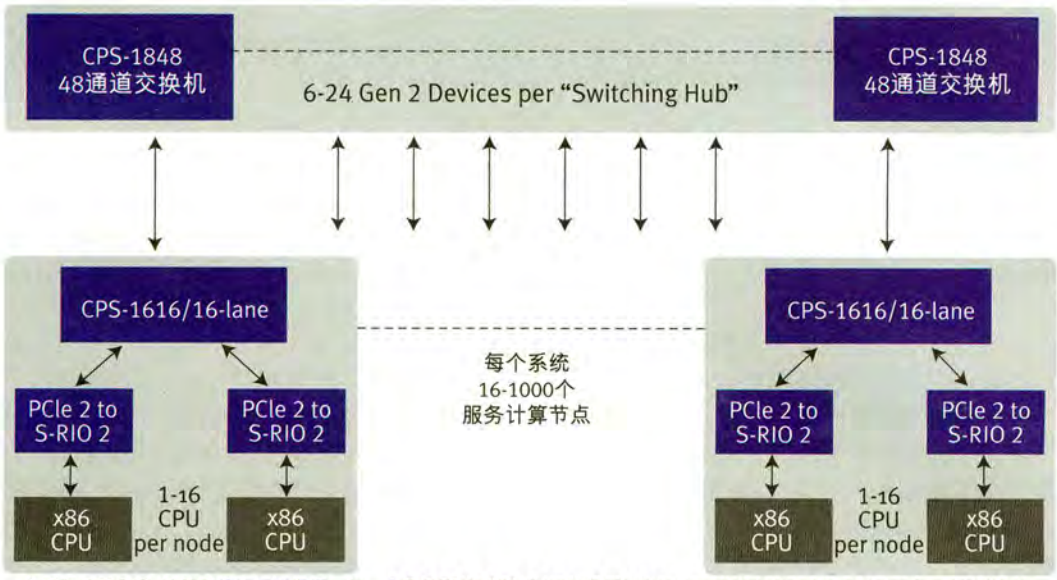
Trevor Hiatt
高级产品经理
IDT公司

企业数据中心、云计算、高性能计算以及嵌入式系统等市场都要求适应下一代网络结构性能要求。这类市场中不断增长的数据需求使得对I/O性能的要求也不断提高。10Gb的端口正在被40Gb甚至更高速的端口所取代。这种演变将导致重新设计硬件，以及重新审视其他竞争网络与目前的以太网技术的优劣。对于普遍使用10Gb以太网的市场，QDR及FDR Infiniband、PCI Express Gen2/3，以及RapidIO Gen2都将是有力的挑战者。

Infiniband已经在交换机、主机总线适配器(HBA)、网络接口卡(NIC)组件、硬件及软件解决方案等众多领域获得巨大成功，成为这类市场中的最高性能系统。RapidIO和Infiniband的成功暴露了以太网技术的潜在弱点，并提供了更低的、更加可预测的网络延迟。RapidIO和Infiniband一样，必须支持以太网和RapidIO在同一系统中共存。能够像以太网应用一样使用本地RapidIO网络，这一性能引发了对在新型系统中使用RapidIO的争论。生态系统对于20Gbps端口的支持及以太网封装的支持，使得RapidIO成为针对下一代吞吐量和性能而重新设计10Gb以太网系统时的一个合适选择。

基于RapidIO的计算/服务器硬件实现示例

IDT公司已经开发出了使用RapidIO Gen2交换器的大端口数非阻塞交换卡。这种交换卡可以应用在数据中心或高性能计算(HPC)环境中。下图所示为一个可用于支持架顶式交换的交换卡的实现。



一个可行的服务器/计算系统实现方框图，每一个服务器/计算节点/刀片通过IDT的Tsi721协议转换器件提供PCIe到RapidIO的桥接，这种架构具有非常好的可扩展性并能比PCIe更有效的机架间连接。

可以使用IDT公司提供的PCIe Gen2至RapidIO Gen2协议转换桥联器件Tsi721作为从互联到服务器节点的适配器。Tsi721将PCIe与RapidIO进行相互转换，并提供20G波特率的全线速桥接。通过使用Tsi721，设计者可以开发既能利用RapidIO的对等网络性能，同时还使用只能在PCIe下实现的多处理器集群的异构系统。目标市场要求大量无需处理器参与的高效数据传输。这可以通过采用全线速分块DMA加上Tsi721的信息传递引擎来实现。支持虚拟化的RapidIO Type9信息传输的优势，以及相比10Gb以太网更高的性能可使线缆数目减少。该示例系统在本文的其余部分将被用作一个讨论的参考。

对于本示例系统，一对节点之间的信息传递是由两种独立的概念所支持的：逻辑I/O和消息传递。逻辑I/O处理由直接桥接转换以及分块DMA引擎来支持。由于DMA和RDMA是针对基于以太网的系统的，因此在系统内也得到支持。信息传递由信息传递引擎来支持。以太网的信息传递是通过包括TCP在内的一系列协议来支持的。RapidIO支持每信道一个单独的IOV，少于现有的以太网解决方案。

协议对比

以太网提供了一种实现对等流量处理器网络(无论是芯片之间、主板之间还是机架之间)的通信方式。但是，以太网是由局域网(LAN)和广域网(WAN)发展而来，因此架构工程师们还在努力寻找一种能在嵌入式系统中更有效应用该网络的途径。由于以太网的局域网及广域网背景，让人自然想到在每个节点放置一个处理器来终结协议栈。这对于局域网还有广域网都是一个合理的设想，但是会在实时嵌入式系统(包括服务器)中造成过大的延迟和功耗。

PCI和PCIe标准提供了一种替代的选择；但这两种标准实际上是为具有根主控概念的单层、单主机处理器系统而设计的。即使使用非透明的桥接，但是在线路卡上(背板之上)利用多主机扩展至多处理器仍然会变得非常困难。这个问题在小规模的终结点或计算结点上还可以加以控制，但是随着系统规模变大，内存映射很快就变得很有难度。

由于RapidIO是针对多处理器对等网络而全新设计的，因此其本身就具有如下特性：可靠的通信、微秒级以下的端到端数据包发送、100ns交换机直通式(cut-through)延迟、无上层处理器终止协议、支持“任意拓扑类型”(即直接互联、网格、星型、双星型等)、面向大量数据传输的高性能信息传递，以及每个处理器都有自己的内存子系统选择的推送架构。

RapidIO已经成为嵌入式互连中的领先者，通过其为背板连接而量身设计的运营级串行通讯，它自身就能够支持一个房间或几个房间内的板卡内、板卡间以及线缆、机架级的连接。

为更好地服务嵌入式领域并扩展到广域网和局域网环境中，针对以太网也制定了补充规范。这些补充是针对共同被定义为数据中心桥接(DCB)的数据中心环境。嵌入式和数据中心领域具有无损传输、优化的流量控制以及低延迟的特性。

QoS和流量控制对比

促使企业数据中心和云端提高带宽的驱动因素之一，是需要把存储网络与服务器间连接网络结合在一起，存储网络通常使用运行速率高达8Gbps的光纤通道，服务器间连接网络通常使用的是1Gbps以太网。这两种网络各自有不同的QoS限制。另外，存储网络一定不能丢包。现在基于RapidIO的系统能够在可预测的QoS条件下实现可靠传输。

有些应用要求更严格、效率更高的QoS，RapidIO为这些应用提供了先进的流量控制和数据平面功能。RapidIO协议在物理层和逻辑层定义了多重流量控制机制。通过在链路层管理物理层的流量控制，利用接受端和发送端控制的流量控制可有效地管理短期阻塞事件。长期阻塞可以通过在逻辑层利用XOFF和XON信息来控制，这使得当在某个特定数据流

探测到拥塞时，接收端可以停止数据包的传输。

虚拟信道支持新的QoS能力。这些功能提供可靠的、尽力服务的传输政策，增强的链路层流量控制和端到端流量管理。虚拟通道还允许在任何两个端点间高达1,600万的独特虚拟流。

以太网还通过采用DCB技术改善了流量控制，从而使链路的一端可以阻止链路的另一端进行传输，以避免缓冲溢出和随之带来的数据包丢失。由VLAN标记而可能实现的数据包简化路由，以及作为VLAN标记一部分的数据包优先级，在缩短延迟时间和提高以太网服务特性方面也发挥了很大的作用。

不过，和RapidIO相比，以太网还存在多重DCB QoS和流量控制限制。例如，以太网的流量控制支持主要由802.3X PAUSE支持。即便使用增强的流量控制机制，阻塞通告成本依然很高，因为通告需要从源头一直传送到网络边缘，而在RapidIO中，通过控制标识传输阻塞通告可以很快发出。以太网机制并未被广泛采用，只有一些供应商为有限的拓扑结构提供专门支持。RapidIO的链路层流量控制使发送端的数据流量能够持续地让接收端保持满负荷状态，从而提升了调度的效率，最终提高了整体交换效率。

性能对比：延迟与吞吐量

以太网交换机设备的延迟时间持续缩短，目前行业领先的以太网交换机能达到的最短时延约为200纳秒。而RapidIO交换机的延迟时间在100纳秒以内，并且同样也在不断缩短。由于企业使用尺寸更小的硅工艺和更高的物理层速度，这种趋势还将继续下去。以太网的端到端数据包终端的延迟可能会大于10微秒，而RapidIO则可能低于1微秒。

RapidIO通过链路层的纠错功能提供有保障的传送。链路层的控制标记使控制环延迟时间降到最低。控制标记也可以嵌入数据包以进一步缩短延迟。无损耗DCE仍然需要卸载引擎和/或者软件堆栈，而这些往往会加大延迟。

RapidIO Gen2交换提供每端口20Gbps的速率。Tsi721用来实现PCIe和RapidIO之间数据的相互转换，并为小至64B的数据包提供16Gbps的全线速桥接。这样的速率比通常提供的10Gb以太网要高，但会比即将出现的40Gb以太网解决方案要低。

从原始带宽的角度来看，以太网要优于RapidIO。但是，一旦RapidIO物理层规范和发展路线实施以后，这个差距应该会缩小。RapidIO性能和协议有效性支持可靠的协议封装。信息传递和/或数据流提供本地以太网封装功能。

安全性

交换结构的安全性由系统主机来保证。除非交换结构的路由表被配置成允许数据包可以在两个节点之间进行路由，否则不可能实现这两个节点之间的通信。每个交换机的端口还有四个过滤器，这些过滤器通过屏蔽和匹配任一数据包开头的20字节，来选择是否丢弃该数据包。这种功能可以用于强制设定地址范围和用于DMA读写操作的destID，而且可以用来阻止任何除了系统主机以外的任何节点查询或者设置交换结构。

设备和系统的功耗和成本

当然，RapidIO的功效也比以太网更高，因为RapidIO的物理层代替了以太网的传输层协议，以确保信息可靠、按顺序地被传送出去。这个上层协议比以太网更为高级，每个被传输的数据可消耗的功率也更多。

以太网供应商对10Gbps以太网端口的定价都很高，对40Gbps端口的定价就更高。10Gbps以太网硅片的单系统端口可能会花费上千美元，同时每个10G的交换设备端口批量售价超过10美元。由于采用了包括较简单的数据包终端、较小的数据包内存和无分类的需求在内的一系列技术，RapidIO的系统端口售价仅为约55美元，交换设备的批量售价为每个10G端口不到4美元。

生态系统

从生态系统角度出发，已使用30多年的以太网比仅有10多年历史的RapidIO的生态系统更有优势。硅片、平台、工具和软件等方面都出现了多个供应商。以太网的硬件生态系统提供了融合的网络卡、交换机和路由器平台、服务器及存储平台。软件生态系统提供网络管理软件、微软视窗(Windows)操作系统、Linux和其他更为多样的选择。另外还包括协议分析、数据包嗅探器、流量发生器、网络测试器和强大的兼容性测试规则。RapidIO具有功能强大的嵌入式操作系统、Linux、OFED、协议分析、系统诊断和多服务器平台。Gen1系统具有很好的兼容性，而Gen2的兼容性还在筹划当中。对于Windows操作系统、VMware、网络适配器解决方案、NPU设备和存储平台的支撑选择仍然十分有限。

本文小结

使以太网从10Gb提升至40Gb而必须进行的重新设计为各种相互竞争的互连结构提供了一个机遇，帮助它们在企业服务器、数据中心、云计算和高性能计算等领域站稳脚跟。在所有竞争者中，具有20Gbps吞吐率的RapidIO Gen2是一个强有力的竞争者。系统设计工程师利用RapidIO更小、欠成熟的生态系统，能够通过交换机、PCIe桥接和目前可用的所有嵌入式CPU解决方案将设计移植到20Gbps上。

这些优势包括一个高度可扩展的、可信赖的、容错的运营商级系统解决方案，同时系统耗损极低。更加出色的流量控制和QoS还能提供更低的功耗和更短的延迟。■

ID号 于www.ed-china.com输入本文ID号可阅读全文及相关文章: 20110952