湖南大學

HUNAN UNIVERSITY

计算机设计

学 生 姓 名 李博文

学 生 学 号 201708010602

专 业 班 级 智 能 1702

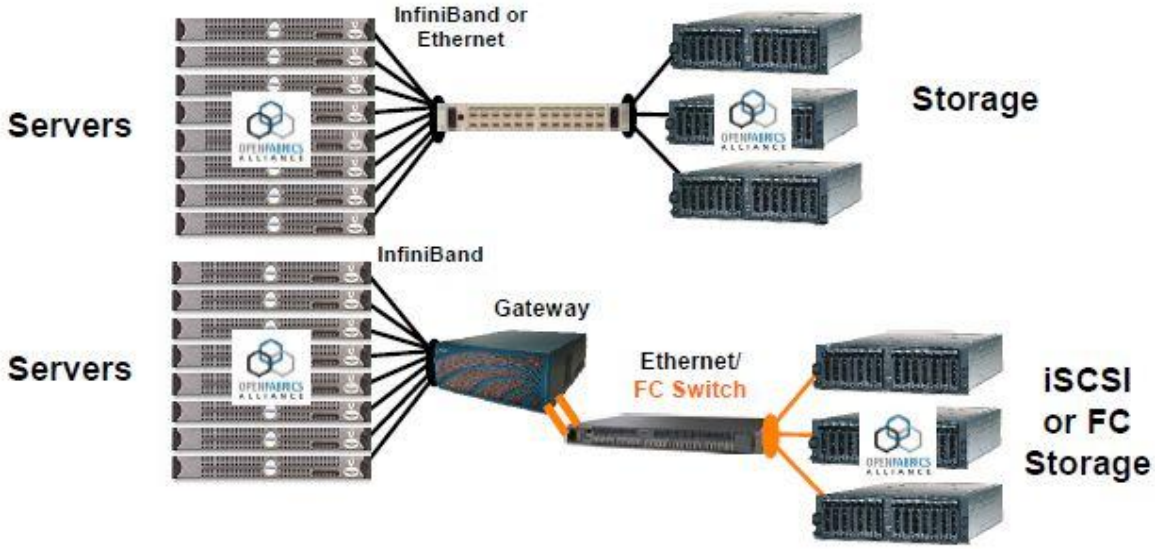
指 导 老 师 吴 强

论 文 题 目 Infiniband网络结构分析

一、infiniband简介

1、InfiniBand网络

InfiniBand是一种网络通信协议，它提供了一种基于交换的架构，由处理器节点之间、处理器节点和输入/输出节点(如磁盘或存储)之间的点对点双向串行链路构成。每个链路都有一个连接到链路两端的设备，这样在每个链路两端控制传输(发送和接收)的特性就被很好地定义和控制了。



InfiniBand通过交换机在节点之间直接创建一个私有的、受保护的通道，进行数据和消息的传输，无需CPU参与远程直接内存访问(RDMA)和发送/接收由InfiniBand适配器管理和执行的负载。

适配器通过PCI Express接口一端连接到CPU，另一端通过InfiniBand网络端口连接到InfiniBand子网。与其他网络通信协议相比，这提供了明显的优势，包括更高的带宽、更低的延迟和增强的可伸缩性。

2、InfiniBand架构

InfiniBand Architecture(IBA)是为硬件实现而设计的，而TCP则是为软件实现而设计的。因此，InfiniBand是比TCP更轻的传输服务，因为它不需要重新排序数据包，因为较低的链路层提供有序的数据包交付。传输层只需要检查包序列并按顺序发送包。

进一步，因为InfiniBand提供以信用为基础的流控制(发送方节点不给接收方发送超出广播 “信用“大小的数据包),传输层不需要像TCP窗口算法那样的包机制确定最优飞行包的数量。这使得高效的产品能够以非常低的延迟和可忽略的CPU利用率向应用程序交付56、100Gb/s的数据速率。

IB是以通道(Channel)为基础的双向、串行式传输，在连接拓朴中是采用交换、切换式结构(Switched Fabric)，所以会有所谓的IBA交换器(Switch)，此外在线路不够长时可用IBA中继器(Repeater)进行延伸。而每一个IBA网络称为子网(Subnet)，每个子网内最高可有65,536个节点(Node)，IBASwitch、IBA Repeater仅适用于Subnet范畴，若要通跨多个IBA Subnet就需要用到IBA路由器(Router)或IBA网关器(Gateway)。至于节点部分，Node想与IBA Subnet接轨必须透过配接器(Adapter), 若是CPU、内存部分要透过HCA (Host Channel Adapter), 若为硬盘、I/O部分则要透过TCA (Target Channel Adapter)，之后各部分的衔接称为联机(Link)。上述种种构成了一个完整的IBA

二、infiniband网络架构

1、IB每一个连接为2.5Gbps，其扩展性非常好，

1X = 2.5Gb/s

4X = 10Gb/s

12X = 30Gb/s

2、互联介质

铜线电缆：10Gbps最长15米

光线电缆：30Gbps最长10米

3、采用8B/10B编码

4、主要硬件部分：

HCA（Host Channel Adapter）主机通道适配器：IB连接的服务器的网络接口，提供虚拟/物理内存的映像，内存直接访问和内存保护。并提供RDMA传送数据。HCA主要功能就是用硬件设备实现了高效的通信功能。

TCA（Target Channel Adapter）目标通道适配器：与HCA类似，但不需要虚拟内存和映像。提供到I/O控制器的链路和传输服务，使I/O设备可脱离主机而直接置于网络中，实现了处理计算、存储I/O、网络I/O等功能的独立。

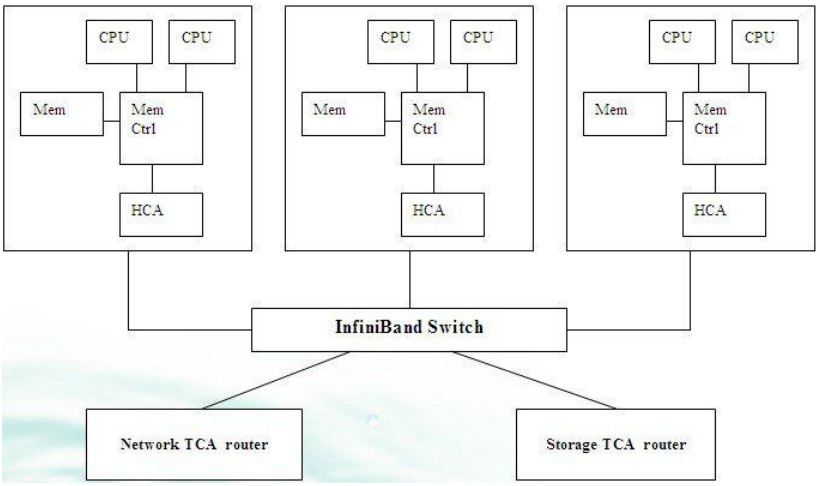
Switch：IBA中提供高集中带宽、负载平衡等的关键部件。一个基本的Switch芯片支持24端口或36端口，可以提供构建成千上万个全交换的高效双向带宽的网络端口，提供无阻塞、低延迟交换功能。

SM（Subnet Manager）子网管理器：IB网络叫做InfiniBand子网，所

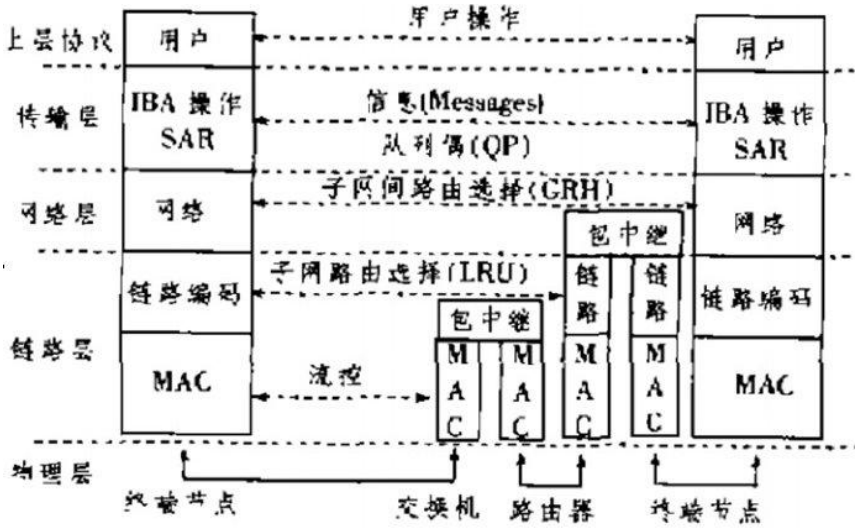
有的设备都在一个SM下控制。负责配置和管理Switch、路由器及通道适配器的应用程序。

三、InfiniBand互联结构

InfiniBand也是一种新的 I/O体系结构 ， 它将 I/0系统与复杂的CPU/存储器分开， 采用基于通道的高速串行链路和可扩展的光纤交换网络替代共享总线结构，提供了高带宽,低延迟,可扩展 的I/O互连，克服了传统的共享 I/O总线结构的种种弊端。



四、InfiniBand分层结构



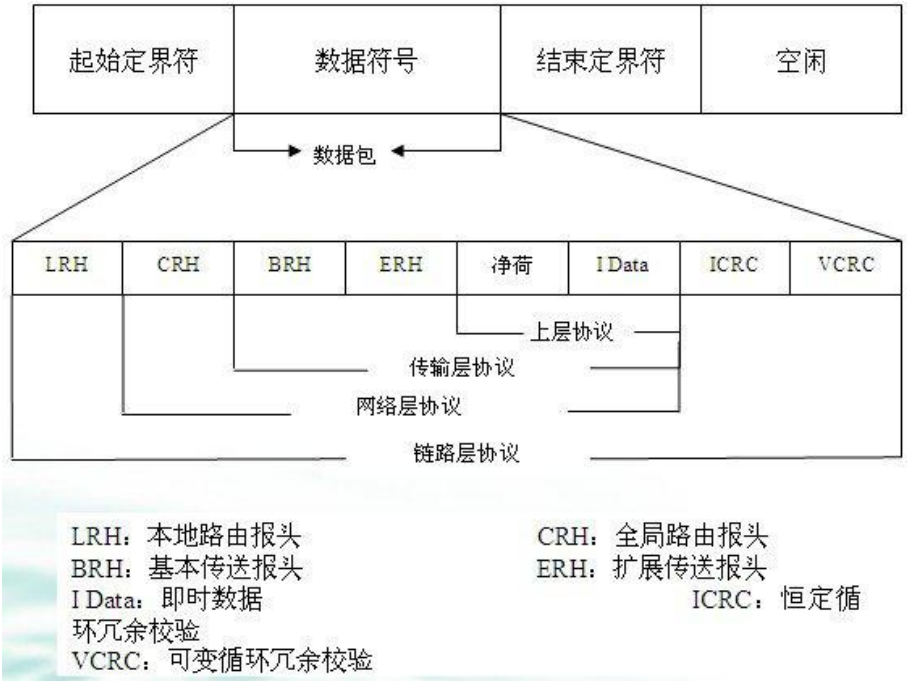
物理层：IBA的物理层定义了在线路上如何将比特组成符号，将符号成帧(如数据包的开始和结束)、数据符号及包间填充(空闲)。详细说明了构建有效包的信令协议，如码元编码、成帧符号排列、起始和结束定界符问的无效或非数据符号、非奇偶性错误、同步方法等。IBA定义了三种物理端口(Port)：背板端口、电缆端口和光缆端口。每个物理端口由4组信号组成：信令组(SignalingGroup)、硬件管理组(Hard—WareManagemenlGroup)、主电源组(BulkPowerGroup)、辅助电源组(AuxiliaryPowetGroup)。

链路层：链路层规定了数据包的格式以及数据包操作的协议，如流控和子网内源端口和目的端口之间的数据包路由选择。在链路/物理接口采用8B/10B编码和解码。IBA定义了两种类型的数据包：链路管理包和数据包。IBA数据包的格式如下图所示。分为链路管理包（用于测试和维护链路，只产生于链路层）和数据包（传送IBA操作）

网络层：为了使数据包在IBA定义的子网(Subnet)之间正确地传送，IBA规定数据包在网络层添加一个全局路由报头(GRH，GlobalRouteHeader)。GRH为4O字节，采用RFC2460定义的IPv6报头格式。

传输层：传输层的功能是将数据包传送到某个指定的队列偶(QP，QHeHePair)中+并指示QP如何处理该数据包。以及当信息的数据净荷部分大于通道的最大传输单元MTU时，对数据进行分段和重组。

五、InfiniBand数据包格式



链路管理包:只产生于链路层，用于测试和维护链路。

数据包：用于传输IBA操作。