# NTSocket 方案性能测试

### 实验环境介绍

对于本文中所有测试的实验环境，其设备信息如下：

测试环境：

|  |  |
| --- | --- |
| CPU | Intel(R) Xeon(R) Platinum 8180 CPU @ 2.50GHz |
| Memory | 48GB DDR4 LRDIMM 2666MHz |
| Linux Kernel version | 5.1.3-050103-generic |
| NTB Device Information | PCI-E Gen 3 Width x4 Max Payload Size = 128B |
| Network Adapter | Intel Corporation Ethernet Connection X722 for 10GBASE-T |
| RDMA Adapter | MCX516A-CDAT ConnectX-5 (100Gbps) |

### 性能开销

**CPU开销：**由于DPDK驱动使用CPU Core是独占式，目前的测试NTSocket拥有一个发送线程和一个接收线程，各占用一个CPU Core，因此CPU开销为200%，即两个CPU Core。

**内存开销：**为NTSocket所属的DPDK进程分配了2个1GB的大页内存，因此内存开销为2GB

### NTB硬件性能测试

#### NTB顺序读写性能

1. 测试内容
   1. 针对单核下Memcpy函数R/W内存长度不同的情况进行测试
   2. 针对多核情况下内存使用Write-Back/Write-Combining模式的R/W性能进行测试
2. 测试方法

NTB Reserved Memory Size默认设置为512MB，读/写均为顺序读/写，

1. CPU开销

由于DPDK驱动使用CPU Core是独占式，目前的测试均独占式的使用一个CPU Core，CPU利用率100%。

1. 测试结果
2. 针对单核下Memcpy函数R/W内存长度不同的情况进行测试

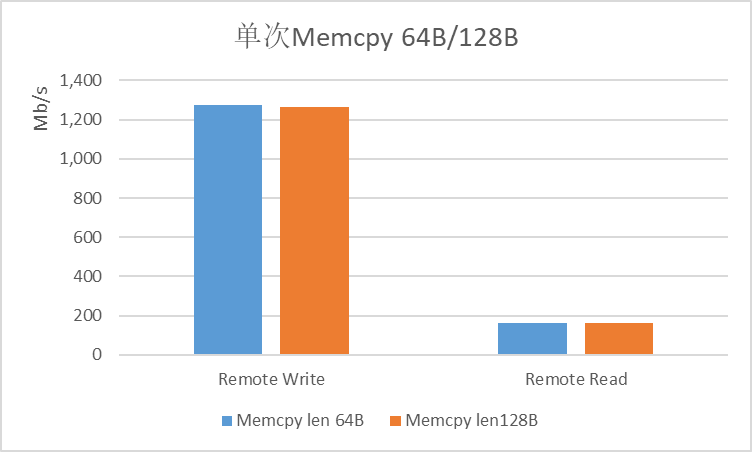


图5.3 Write-Back模式下Memcpy长度对速率影响

**测试结果：**改变Memcpy函数的单次读/写长度对性能没有提升。之后的测试默认使用64B的单次读/写长度。

1. 针对多核情况下分别使用Write-Back/Write-Combining模式的R/WS性能进行测试

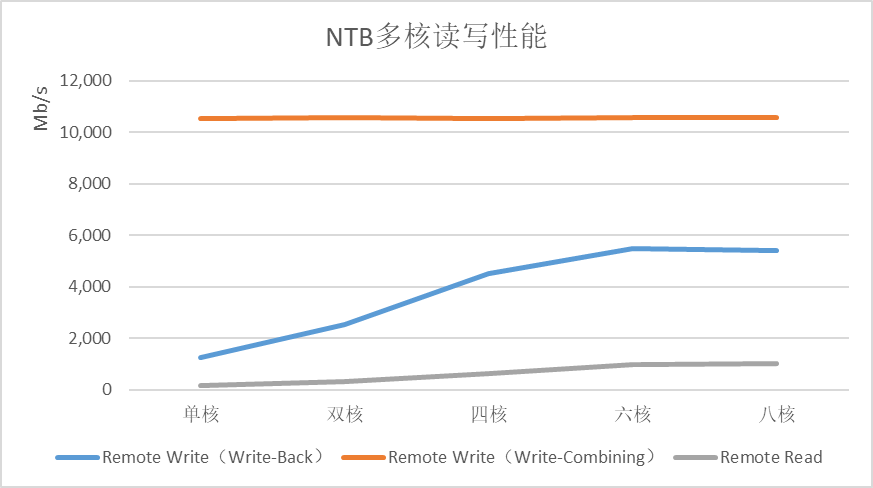


图5.4 NTB多核读写性能测试

**测试结果**：Write-Back模式中，RW、RR吞吐量均在6核左右达到最大，RW速率为5Gb/s，RR速率为1Gb/s。开启Write-Combining模式之后，RW速率在单核时即可达到最大10Gb/s，RR速率在6核时达到最大1Gb/s，与Write-Back下速率相同，图线重合。

**总结：**在开启O3编译优化的情况下，Memcpy长度不影响读写性能，内存区域开启Write-Combining模式后，Remote Write速率在单核时即可达到最大10.2Gb/s，Remote Read速率在6核时达到最大1.1Gb/s，单核170Mb/s。

#### NTB随机写入性能

NTB提供的最大顺序Remote Write性能在10.3Gb/s，顺序Remote Read性能在1.2Gb/s。由于其顺序Remote Read性能远小于顺序Remote Write，并且我们设计NTSocket协议的时候避免了使用Remote Read操作，本节测试随机Remote Write性能，没有再对随机Remote Read进行测试。

下图为单核情况下对NTB进行随机Remote Write测试的情况：

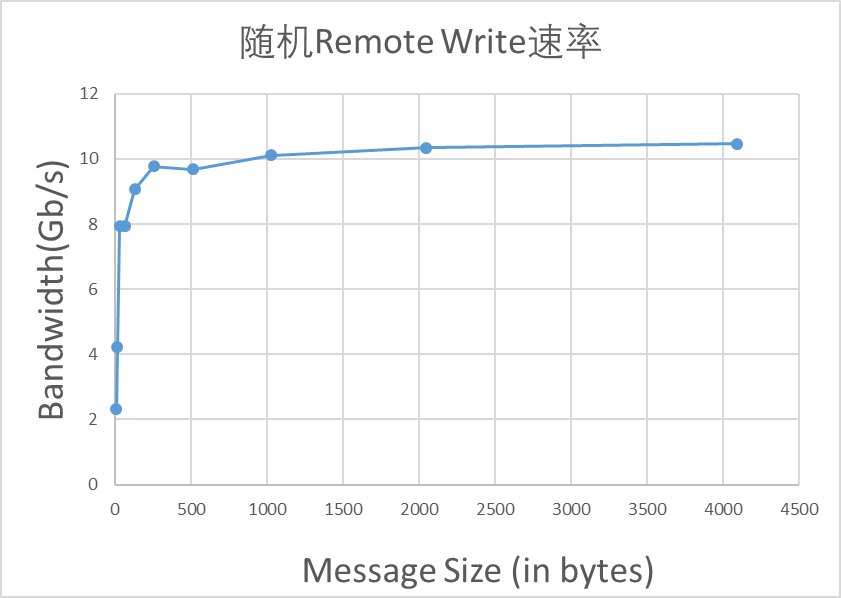
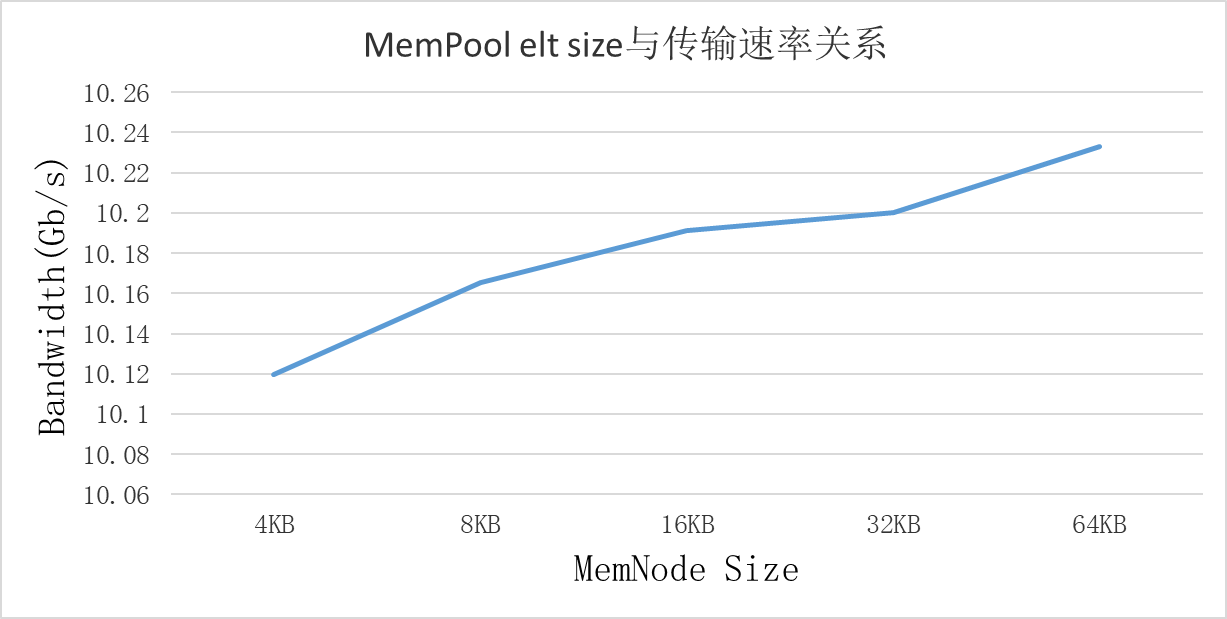


图5.5 NTB随机Remote Write速率

**测试结果**：在Message Size大小为8B时，Bandwidth为2.3Gb/s。在Message Size增加到32B之前，Bandwidth呈线性增长趋势，在Message Size大小为32B时，Bandwidth为7.9Gb/s。之后随着Message Size的增长，Bandwidth呈缓慢增长趋势，在Message Size为256B时，Bandwidth为9.77Gb/s；4096B时，Bandwidth为10.3Gb/s，基本等于顺序读写速率。

### Mempool elt size对吞吐量的影响

本节针对MemPool中Elt Size的不同大小，进行文件传输测试，来评估设计的NTSocket传输性能。

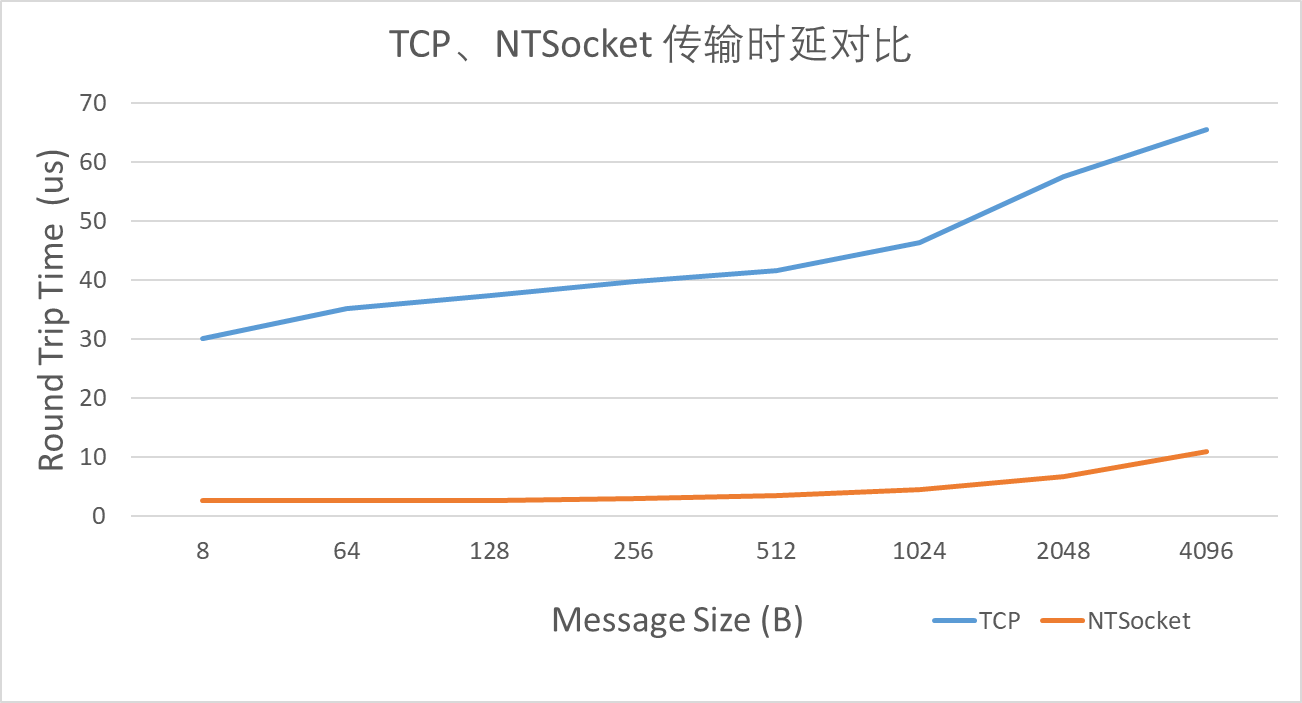


MemNode Size与传输速率关系图

**总结：**当MemPool的内存节点MemNode大小从4KB增加到64KB时，传输速率从10.12Gb/s上升至10.23Gb/s，提升约1%。本协议传输效率为（1-首部长度/最大包长）=96.875%。选择4KB作为MemPool节点默认大小时，Remote Write数据传输效率为9.8Gb/s。MemNode Size的增加对传输性能提升不大，因此NTSocket默认选择4KB作为MemNode的大小。

### 传输时延对比测试

TCP的传输时延使用惠普公司开源的NetPerf测试工具进行测试，网卡设置均采用默认设置，测试模式为Request/Response模式。NTSocket的传输时延是使用本文实现的NDTP提供的Send/Rev接口实现的测试程序进行测试。



TCP、 NTSocket传输时延对比

**总结：**B2B情况下，两种传输方式时延均随Message Size的增大而增大，TCP传输方式的RTT远高于NTSocket，在Message Size为8B时约为30.5us，4KB时达到65.4us。NTSocket在Message Size为8B到128B时RRT约为2.6us；4KB时为11us.。在任意Message Size下，NTSocket的传输时延均大幅低于TCP。