

4.4 The Slow-receiver symptom

In our data centers, a server NIC is connected to a ToR switch using a point-to-point cable. The NIC is connected to the CPU and memory systems via PCIe. For a 40 GbE NIC, it uses PCIe Gen3x8 which provides 64Gb/s raw bidirectional bandwidth which is more than the 40Gb/s throughput of the NIC. Hence there seems to be no bottleneck between the switch and the server CPU and memory. We thought that the server NIC should not be able to generate PFC pause frames to the switch, because there is no congestion point at the server side. But this was not what we have observed. We found that many servers may generate up to thousands of PFC pause frames per second. Since RDMA packets do not need the server CPU for processing, the bottleneck must be in the NIC. It turned out that this is indeed the case. The NIC has limited memory resources, hence it puts most of the data structures including QPC (Queue Pair Context) and WQE (Work Queue Element) in the main memory of the server. The NIC only caches a small number of entries in its own memory. The NIC has a Memory Translation Table (MTT) which translates the virtual memory to the physical memory. The MTT has only 2K entries. For 4KB page size, 2K MTT entries can only handle 8MB memory. If the virtual address in a WQE is not mapped in the MTT, it results in a cache miss, and the NIC has to replace some old entries for the new virtual address. The NIC has to access the main memory of the server to get the entry for the new virtual address. All those operations take time and the receiving pipeline has to wait. The MTT cache miss will therefore slow down the packet processing pipeline. Once the receiving pipeline is slowed down and the receiving buffer occupation exceeds the PFC threshold, the NIC has to generate PFC pause frames to the switch. We call this phenomenon the slow-receiver symptom. Though its damage is not as severe as the NIC PFC storm, it may still cause the pause frames to propagate into the network and cause collateral damage. The slow-receiver symptom is a 'soft' bug caused by the NIC design. We took two actions to mitigate it. On the NIC side, we used a large page size of 2MB instead of 4KB. With a large page size, the MTT entry miss becomes less frequent. On the switch side, we enabled dynamic buffer sharing among different switch ports. Compared with static buffer reservation, dynamic buffer sharing statistically gives RDMA traffic more buffers. Hence even if the NICs are pausing the switch ports from time to time, the switches can absorb additional queue buildup locally without propagating the pause frames back into the network. Compared with static buffer allocation, our experience showed that dynamic buffer sharing helps reduce PFC pause frame propagation and improve bandwidth utilization.

在我们的数据中心，服务器 NIC 使用点对点电缆连接到 ToR 交换机。NIC 通过 PCIe 连接到 CPU 和内存系统。

对于 40 GbE NIC，它使用 PCIe Gen3x8，提供 64Gb/s 原始双向带宽，超过 NIC 的 40Gb/s 吞吐量。因此交换机与服务器 CPU 和内存之间似乎不存在瓶颈。我们认为服务器网卡不应该能够生成 PFC 暂停帧到交换机，因为服务器端不存在拥塞点。

但这不是我们观察到的。我们发现许多服务器每秒可能会生成多达数千个 PFC 暂停帧。

由于 RDMA 数据包不需要服务器 CPU 来处理，因此瓶颈一定在网卡上。事实证明，确实是这样。

网卡的内存资源有限，因此它将大部分数据结构，包括 QPC（队列对上下文）和 WQE（工作队列元素）放在服务器的主内存中。NIC 仅在其自己的内存中缓存少量条目。

NIC 有一个内存转换表 (MTT)，它将虚拟内存转换为物理内存。MTT 只有 2K 条目。对于 4KB 页面大小，2K MTT 条目只能处理 8MB 内存。

如果 WQE 中的虚拟地址未映射到 MTT 中，则会导致缓存未命中，并且 NIC 必须用新的虚拟地址替换一些旧条目。NIC 必须访问服务器的主内存才能获取新虚拟地址的条目。

所有这些操作都需要时间，并且接收管道需等待才能使用。

因此，MTT 缓存未命中将减慢数据包处理管道的速度。一旦接收管道变慢且接收缓冲区占用超过 PFC 阈值，NIC 必须向交换机生成 PFC 暂停帧。我们将这种现象称为慢接收症状。

虽然它的损害不像 NIC PFC 风暴那么严重，但它仍然可能导致暂停帧传播到网络中并造成附带损害。接收速度慢的症状是由 NIC 设计引起的“软”错误。

我们采取了两项措施来缓解这一问题：

1. 在 NIC 方面，我们使用了 2MB 的大页面大小，而不是 4KB。页面大小较大时，MTT 条目丢失的频率会降低。
2. 在交换机方面，我们启用了不同交换机端口之间的动态缓冲区共享。与静态缓冲区预留相比，动态缓冲区共享在统计上为 RDMA 流量提供了更多缓冲区。因此，即使 NIC 不时暂停交换机端口，交换机也可以在本地吸收额外的队列积累，而无需将暂停帧传播回网络。与静态缓冲区分配相比，我们的经验表明动态缓冲区共享有助于减少 PFC 暂停帧传播并提高带宽利用率。