关于 RDMA 架构改进的研究综述

王文才[1)]

[1)](华中科技大学计算机科学与技术学院 武汉市 中国 430074)

摘 要 随着高性能计算、人工智能、大数据分析以及物联网技术的高速发展，各类行业应用对于网络传输性能的要求越来越高，与软件传输 TCP 相比，RDMA 以高吞吐量、低延迟和低 CPU 开销特性被广泛运用在各种数据中心当中。
然而 RDMA 需要 PFC 来维持无损网络结构，其带来了管理风险和网络可扩展性挑战，严重影响了整个集群网络的可用性和网络可扩展性。随着 RDMA 部署的不断增加，现代数据中心多采用了多种异构 RNIC，异构网络设备的共存会导致严重的带宽不平衡现象，并增加产生 PFC 暂停帧的风险。为了保持较低的 CPU 使用率和较高的吞吐量，RDMA 将大量网络任务卸载到 RNIC 中，导致 RNIC 需要同时管理数据传输和维护连接元数据，但 RNIC 中有限的片上存储器大小限制了 RNIC 可支持的活动连接的数量，导致 RNIC 只能维持有限数量的连接，限制了其连接可扩展性。针对这些问题，本文对近几年针对 RDMA 架构改进的论文工作进行分析总结。
关键词 远程直接内存访问；网卡；架构设计；可拓展性问题

Research Review on RDMA Architecture Improvementsoman

Wencai Wang[1)]

[1)] (Department of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074)

Abstract With the rapid development of high-performance computing, artificial intelligence, big data analysis, and IoT technologies, various industry applications demand higher network transmission performance. In comparison to software-based TCP transmission, RDMA, known for its high throughput, low latency, and low CPU overhead characteristics, has been widely utilized in various data centers. However, RDMA requires PFC (Priority Flow Control) to maintain a lossless network structure, which brings about management risks and challenges in network scalability. This seriously affects the availability and scalability of the entire cluster network. As RDMA deployments continue to increase, modern data centers often incorporate multiple heterogeneous RNICs (RDMA Network Interface Cards). The coexistence of heterogeneous network devices can lead to significant bandwidth imbalances and increase the risk of generating PFC pause frames. In order to maintain low CPU utilization and high throughput, RDMA offloads a large number of network tasks to the RNIC, causing the RNIC to simultaneously manage data transmission and maintain connection metadata. However, the limited on-chip memory size in the RNIC constrains the number of active connections that the RNIC can support, limiting its connection scalability. To address these issues, this paper analyzes and summarizes recent research efforts in RDMA architecture improvements over the past few years.
Key words Remote Direct Memory Access; Network Interface Card; Architecture Design; Scalability Problem

1 引言

随着高性能计算、人工智能、大数据分析以及物联网技术的高速发展，各类行业应用对于网络传输性能的要求越来越高，传统的 TCP/IP 协议架构存在数据处理和网络传输延迟较大、TCP/IP 协议处理复杂等问题。目前，许多云服务提供商，例如 Microsoft 和 Amazon 等，已转向 RDMA （远程直接内存访问）来优化其数据中心网络。

RDMA 技术是一种用于解决网络传输中服务器数据处理延迟的技术，其支持远程直接读写异地内存，且无需双方 CPU 和操作系统的介入。用户空间应用程序通过将工作队列元素（WQE）发布到队列对（QP）中来向 RNIC 发起数据传输请求。RNIC 发送数据后，生成完成队列元素(CQE)到完成队列(CQ)中，作为用户的发送完成信号。 RDMA 支持三种传输类型：可靠连接（RC）、不可靠连接（UC）和不可靠数据报（UD）。相应地，它提供了两组原语：SEND/RECV、WRITE/READ，其中 SEND/RECV 是所有传输都支持的双向操作，WRITE 是 RC 和 UC 支持的单边操作，而 READ 仅受 RC 支持。

与传统的软件传输 TCP 不同，RDMA 是一种硬件传输，完全在 NIC 硬件中实现拥塞控制和丢失恢复等传输功能，并向用户应用程序提供内核旁路和零拷贝接口。因此，与软件传输 TCP 相比，RDMA 实现了高吞吐量、低延迟和低 CPU 开销。RDMA 这些特性能够用于为许多应用提供服务，例如键值存储、分布式事务、分布式内存、远程过程调用（RPC）、存储系统、图计算和机器学习系统，因此 RDMA 被广泛地运用在数据中心当中。

2 问题与挑战

2.1 PFC风暴带来的运营挑战

RDMA 最初是为无损 Infiniband 设计和简化的，为了使 RDMA 能够在以太网中工作，RoCEv2 (RDMA over Converged Ethernet Version 2)依靠 PFC（基于优先级的流量控制）将以太网转变为无损结构。

然而，PFC 带来了管理风险和网络可扩展性挑战[1]（例如，队头阻塞、拥塞扩散、偶发死锁以及大规模集群中的 PFC 风暴），严重影响了整个集群网络的可用性，这是 RDMA 中最著名的问题。此外，使用 PFC 时，无损网络规模还会受到交换机缓冲区大小的限制，数据中心通常会为此限制 RDMA 网络的规模，导致网络可扩展性问题。

可以通过用更高效的选择性重传(SR) 替换 go-back-N 来消除 PFC 的有损 RDMA。然而 SR 的引入需要添加特定的数据结构，从而增加了内存消耗，无法避免连接扩展性问题[2]。

2.2 异构RNIC的不适配性

随着 RDMA 部署的不断增加，现代数据中心采用了不同世代和供应商的 RDMA 功能 NIC（RNIC），例如 Mellanox ConnectX-(CX)4/5/6、BlueField、Intel E810，以及云提供商定制的 RNIC。一方面，采用多个供应商可避免供应商锁定，即依赖于特定供应商的设备。另一方面，存储和计算系统的分解部署将后端服务与前端服务分离到不同的集群中，也可使每个集群针对其需求使用不同类型的 RNIC。

但数据中心中异构网络设备的共存也带来了新的挑战[3]：首先，设备可能采用不同实现方式的 RDMA 引擎，这种情况不仅存在于不同的供应商之间，也存在于同一供应商的不同代设备之间，在多种不同 RNIC 的混合部署集群中，会观察到严重的带宽不平衡现象。其次，不同供应商的网卡的拥塞控制行为的差异，如 Broadcom RNIC 实现 DCTCP 作为拥塞控制算法，Intel E810 RNIC 实现基于窗口的 DCQCN 算法，会进一步放大带宽的不平衡。此外，异构 RNIC 同样会增加产生 PFC 暂停帧的风险。

2.3 连接拓展性问题

数据中心对 RDMA 连接的可扩展性要求很高。与 HPC 不同，非对称通信经常发生在数据中心网络中。这些通信模式一般是 in-cast 或 out-cast，即一个节点（服务器）连接到多个其他节点（客户端）。在这些网络中，服务器节点必须与客户端保持数千个连接，因此随着 RDMA 在数据中心的大规模部署，对 RDMA 连接扩展性的需求在不断增长。

然而为了保持较低的 CPU 使用率和较高的数据包吞吐量,RDMA 通过将大量网络任务卸载到 RNIC 中，为其他数据中心应用程序释放更多 CPU 资源。因此，RNIC 不仅需要管理数据传输，还需要维护连接元数据，以便及时响应通信请求。然而，RNIC 中有限的片上存储器大小限制了 RNIC 可支持的

活动连接的数量，导致 RNIC 只能维持有限数量的连接，限制了其连接可扩展性[4]。

3 研究现状分析

这一部分主要是对近段时间提出的几个 RDMA 改进架构进行介绍，了解目前针对 RDMA 存在的问题与挑战所做的优化。

3.1 Flor[3]

鉴于异构 RNICs 之间的合作问题，以及 PFC 风暴的根本原因，本研究的动机是提出一个开放和统一的框架，以应对数据中心设备的多样性，并为用户提供 RDMA 编程的灵活性，以减少大规模数据中心网络的运维复杂性。通过在硬件和软件之间重新划分功能，Flor 框架实现了数据路径和控制路径的分离，从而提供了高性能的数据传输和灵活的拥塞控制和可靠性管理。通过在硬件层面保持 RDMA 数据路径规范，Flor 能够实现高吞吐量、低延迟和低 CPU 开销。而将相对较慢的控制路径加载到软件层，以实现灵活性，同时对系统效率影响较小。

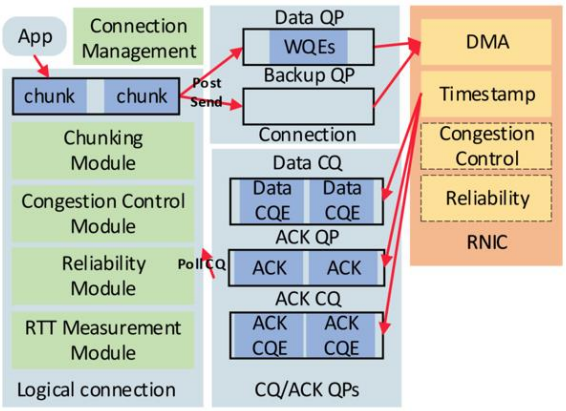


图 1　Flor 架构

Flor 架构主要包括负载感知的动态分块模块、与 RDMA 语义兼容的可靠性和拥塞控制模块以及基于不可靠连接（UC）的选择性重传模块。Flor 用户可以根据不同的场景选择这些软件模块的组合，软件拥塞控制和可靠性模块可以被硬件功能绕过或替代。如下表所示，在禁用 PFC 和 ECN 的 CX-4 集群中，Flor 通过基于硬件 (RC) 或基于软件 (UC) 的可靠传输启用分块和软件拥塞控制，从而提供 RDMA 服务。对于跨 Pod 应用程序，可以使用基于软件 (UC) 的可靠性来容忍数据包丢失。

Scenarios	Chunking	Reliability	CC
Intra-pod, PFC-enabled	No	*RC*	*HW*
Intra-pod, PFC- and ECN-disabled	Yes	*RC* or *UC*	*SW*
Across-pod Applications	Yes	*UC*	*SW*
CX-4/5 Hybrid	Yes	*RC* or *UC*	*SW*

图 X　图片说明 *字体为小 5 号，

3.1.1 动态分块

分块算法决定了 RDMA 请求的粒度。大粒度可能会导致流量突发，导致拥塞，并且在丢包的恢复成本较高，而小粒度 CPU 成本更高。因此，设计的关键点是根据当前网络状况动态调整分块大小。更具体地说，在丢失率较低时采用硬件大粒度分块来减少 CPU 开销，而发生数据包丢失时通过软件切片和重传，其中主要包含 RTT 测量和分块策略两部分。

测量得到的 RTT 除了动态分块算法的反馈信号，还可用于拥塞控制和选择重传的控制信号，因此 RTT 测量的准确性直接影响所有这些组件的性能。与基于每个数据包时间戳测量 RTT 的方法不同，Flor 采用了 RoGUE 方法来测量 RC 传输的 RTT。图 4 展示了这种 RTT 测量方法。

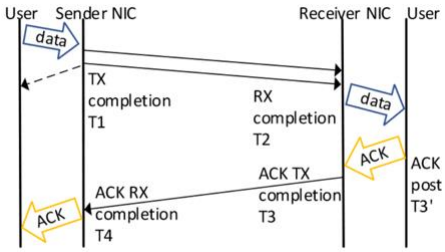


图 2　Flor 中 UC 的 RTT 测量。T 1 - T 4 为事件的时间戳。,

在发送方，可从硬件上获取数据 WQE 发送完成（T1）和相应的 ACK 接收完成（T4）的时间戳。在接收端，数据 WQE 到达的时间戳（T2）可从硬件读取，相应的 ACK WQE 发布时间（T′ 3）可通过软件获取。最终 RTT 由这四个时间戳计算得到。

然而，由于 RNIC 中发送请求的排队，测量并不绝对准确，因为 T' 3 是发送完成后的时间而不是 ACK 的实际发送完成时间(T3)。此外，在重负载下，由于数据 WQE 和 QP 调度策略的队头阻塞，ACK 在发送方软件接收时可能会延迟长达几毫秒。其结果是，一方面，测量的 RTT 会因该开销而增加，不能准确地反映网络拥塞情况；另一方面，这也拉长了拥塞反馈循环，使得发送方无法及时响应网络拥塞。

Flor 使用两种优化方法来提高 RTT 测量精度

并缩短反馈环路延迟：首先，使用高优先级 UD QP 来发送和接收 ACK，以避免 RNIC 上的队头阻塞和调度延迟。其次，对 ACK 使用单独的 ACK CQ，并在数据 CQ 之前对其进行轮询。由下图所示，通过实验测量表明该方法可将尾部 RTT 上的 RTT 测量精度提高 10 倍。

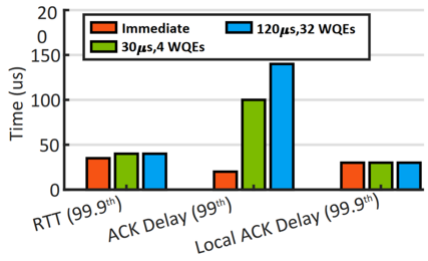


图 3 ACK 设计优化评估结果

分块策略：Flor 将分块算法提取为模块，用户可以指定自己的分块算法进行替换。关键思想是当网络的 RTT 变差时动态减小块大小，当网络状态变好时动态增加块大小。我们通过可用拥塞窗口 (acwnd) 或带宽延迟乘积 (BDP) 的最小值来初始化块大小，后续以 RTT 作为网络反馈信号对块大小进行动态调整。

3.1.2 基于 UC 的选择重传

基于将 RDMA 消息拆分为不同大小的 WQE 的分块机制，Flor 能够在分块 WQE 的粒度上设计可靠性机制（序列号、确认和重传）。

为了避免在接收端将信息块组装成原始信息时造成额外的数据拷贝，Flor 使用特有的序列号空间生成分块到达的信号，使每个 WQE 都拥有两组序列号：全局序列号和可靠性序列号。所有 WQE 都拥有一个唯一的全局序列号，用于识别 QP 中的序列，而可靠性序列号用于识别同类型 WQE 中的序列。此外这两组序列还可用于识别原始 WQE 和重传 WQE。

如果为每个 WQE 生成 ACK 可能会导致小消息流量的高开销，因此 Flor 将多个序列号放入一个软件 ACK 中以减少 CPU 开销。然而，由于 ACK 携带拥塞控制和可靠性机制中使用的 RTT 和 WQE 编号，存在需要及时发送 ACK 的情况。因此，Flor 为接收方立即发送 ACK 设置了触发规则来避免中要信息的延迟。

3.1.3 增强硬件重传

RDMA 操作的硬件重传不受软件拥塞控制的控制，这具有引发网络拥塞的潜在风险，因为在传输中数据的大小可能比软件拥塞窗口大得多。 Flor 通过增加软件重传方案提高了硬件可靠性。 Flor 在 RC QP 中设置较短的重传重试次数，来限制超出软件拥塞控制窗口的数据大小。

如果在重试次数内重传失败，则 RNIC 硬件将 QP 转为错误状态。将 QP 的错误状态转变为工作状态需要很长时间。相反，Flor 将未完成的运行中 WQE 重新提交给另一个预连接的 QP，称为备份 QP，并将备份 QP 翻转为主 QP 以继续数据传输。

非活动备份 QP 不会消耗 RNIC 上的额外缓存资源，因此对性能没有副作用。根据作者的实践经验表明，与 QP 重新连接相比，切换到备用 QP 花费的时间更少。

3.1.4 设计总结

该研究推出了 Flor，一种适用于异构 RNIC 的灵活有损 RDMA 框架，它解决了生产 RoCEv2 集群中出现的一系列问题。这些问题包括 PFC 依赖性、异构 RNIC 的互连性以及硬件绑定的拥塞控制方案。Flor 将 RNIC 的可靠性和拥塞控制功能加载到软件中。Flor 在 RoCEv2 网络中首次提出软件选择性重传，并使用基于软件 RTT 的拥塞控制来处理异构 RNIC 之间的性能差距。通过对测试床和生产集群的评估表明，Flor 在丢包、异构硬件、大规模组播和分布式系统等多种场景下实现了高性能和灵活性。

3.2 SRNIC[5]

RDMA 预计具有高度可扩展性：在数据包丢失不可避免的大型数据中心网络中表现良好，并支持每台服务器的大量高性能连接。但由于商业 RNIC 依赖于无损、规模有限的网络结构，并且仅支持少量的高性能连接，缺乏可扩展性。

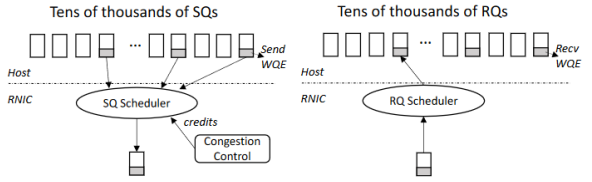
本研究认为，通过仔细的协议和架构协同设计，可以最大限度地减少 RNIC 中的片上数据结构及其内存需求，以提高连接的可扩展性。在这一见解的指导下，论文分析了 RDMA 概念模型中涉及的所有数据结构，并通过 RDMA 协议头修改和架构创新（包括无缓存 QP 调度程序和无内存选择性重复）。

论文提出了 SRNIC，一种可扩展的 RNIC 架构，以解决连接可扩展性问题，同时保留作为商业 RNIC 的传输卸载继承的高性能和低 CPU 开销，并保持源自有损 RDMA 作为 IRN 的高网络可扩展性。SRNIC 通过仔细的协议和架构协同设计，可以消除 RNIC 中的大多数片上数据结构及其内存需求，从而可以显着提高 RNIC 的连接可扩展性。

对于有损 RDMA 概念模型中的典型数据流，分析了所有涉及的数据结构，并将其分为两类：一般 RDMA 所需的通用数据结构，以及选择性重复的特定数据结构有损 RDMA 带来的影响。采取定制的优化策略分别最小化这两类数据结构，以提高连接可扩展性，并提出的无缓存 QP 调度器优化了 RDMA 设计的通用数据结构，引入 RDMA 标头扩展和位图加载优化无内存选择性重复，用于针对有损 RDMA 网络。

3.2.1　无缓存 QP 调度器

SQ 在包含 WQE 时处于活动状态，否则处于非活动状态。SQ 调度器每次都会从主机内存中数万个 SQ 中选择一个活动 SQ 来发送下一个消息。活动 SQ 不能盲目调度，因为它们也受到拥塞控制的影响，一旦 SQ 被调度，如果由于拥塞控制授予的信用不足而不允许其发送消息，则调度不会生效，只会浪费时间并降低性能。RNIC 和主机内存之间的 PCIe 往返延迟很高，并且至少需要两个 PCIe 事务（一个 WQE 获取和一个消息获取）才能执行一个调度决定。如果没有仔细的设计，调度迭代之间的高延迟将显着降低性能。主机内存中有数以万计的 SQ，但 RNIC 内的片上内存非常有限。RNIC 中的不同 SQ 具有单独的 WQE 缓存是禁止的。



(a) SQ scheduler model.　　(b) RQ scheduler model.

图 4　QP 调度器模型

论文设计了一种无缓存的 SQ 调度机制，在 SQ 活跃且具有 credits 时进行调度，并通过适当的批量事务来隐藏 PCIe 延迟，且采用无 WQE 缓存的方式。这种无缓存的 SQ 调度机制可以在数万个 QP 之间进行快速调度，并且片上内存需求最少。它由三个主要部分组成：

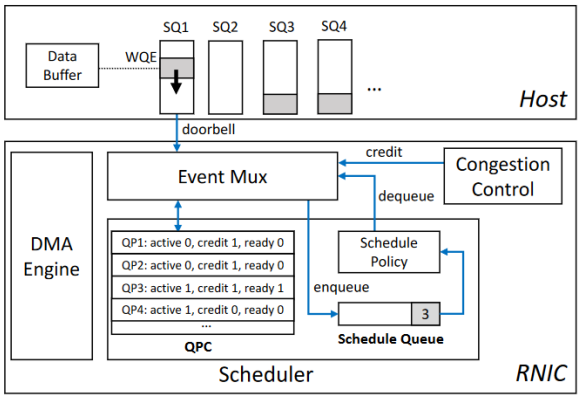


图 5　无缓存的 SQ 调度器

事件复用器（EMUX）：EMUX 模块处理所有与调度相关的事件，当且仅当 SQ 处于活动状态并具有 credits 时才会进行调度。

调度器（Scheduler）：调度器利用调度队列来维护准备调度的 SQ 列表，每次从调度队列的头部弹出一个就绪的 SQ，并从该 SQ 中获取给定数量的 WQE 和消息。

DMA 引擎：当调度 SQ 后，RNIC 中可能剩余未使用的 WQE，未使用的 WQE 将被丢弃而不是缓存在 RNIC 中，并且下次调度其 SQ 时将再次获取它们，以此实现无缓存调度。

3.2.2　无内存选择性重复

选择性重复带来的额外数据结构包括未完成的请求表、重排序缓冲区和位图，其内存需求总共超过了 RNIC 的典型片上 SRAM 大小。为了最大限度地减少选择性重复带来的内存需求，SRNIC 通过 RDMA 协议标头扩展消除了对未完成的请求表和重新排序缓冲区的需求，并通过仔细的软件硬件协同设计将位图加载到主机内存中，而不影响性能。

3.2.3　设计总结

SRNIC 通过无缓存的 QP 调度器消除了 WQE 缓存，通过 SR 友好的标头扩展和位图加载消除了片上存储器中所有与 SR 相关的数据结构，并通过缓存最小化了 MTT 的片上存储器需求，同时将大型 MTT 表保留在主机内存中。该架构解决了连接可扩展性挑战，同时实现了高网络可扩展性、高性能和低 CPU 开销。

4 总结与展望

随着高性能计算、人工智能、大数据分析以及物联网技术的高速发展，各类行业应用对于网络传输性能的要求越来越高。本文对传统 RDMA 网络所暴露出来的问题与挑战进行探讨，如 PFC 风暴带来的运营挑战，异构 RNIC 的不适配性，连接拓展性问题等等。研究了近几年实现 RDMA 改进架构的论文工作。其中 Flor 架构鉴于异构 RNIC 之间的合作问题，以及分析 PFC 风暴产生的根本原因，通过设计不同模块应用于不同场景，从而避免了异构 RNIC 之间的差异性对 RDMA 网络性能的影响，同时消除了 PFC。SRNIC 通过协议和架构的协同设计，在保证了高网络可扩展性、高性能和低 CPU 开销的同时，最大限度地减少 RNIC 中的片上数据结构及其内存需求，以提高 RDMA 的连接扩展性。

参 考 文 献

[1] Gao Y, Li Q, Tang L, et al. When cloud storage meets RDMA[C]//18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21). 2021: 519-533.

[2] Kong X, Zhu Y, Zhou H, et al. Collie: Finding Performance Anomalies in RDMA Subsystems[C]//19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22). 2022: 287-305.

[3] Li Q, Gao Y, Wang X, et al. Flor: An Open High Performance RDMA Framework Over Heterogeneous RNICs[C]//17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 23). 2023: 931-948.

[4] Kang N, Wang Z, Yang F, et al. csRNA: Connection-Scalable RDMA NIC Architecture in Datacenter Environment[C]//2022 IEEE 40th International Conference on Computer Design (ICCD). IEEE, 2022: 398-406.

[5] Wang Z, Luo L, Ning Q, et al. SRNIC: A Scalable Architecture for RDMA NICs[C]//20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23). 2023: 1-14.

附录：论文汇报记录

问题 1：这篇论文最突出的贡献是什么？

该论文的主要工作是设计出了一种适用于异构 RNIC 的灵活有损RDMA框架，它解决了RoCEv2 集群中出现的一系列问题。这些问题包括PFC依赖性、异构RNIC的互连性以及硬件绑定的拥塞控制方案。Flor将RNIC的可靠性和拥塞控制功能加载到软件中，并首次在RoCEv2 网络中提出软件选择重传，并使用基于软件RTT的拥塞控制来处理异构RNIC之间的性能差距。论文通过对测试床和生产集群的测试评估表明，Flor 在丢包、异构硬件、大规模组播和分布式系统等多种场景下实现了高性能和灵活性，此外，将现有RDMA框架升级到Flor的过程对正在运行的应用程序的性能影响很小。

问题 2：怎么增强硬件重传

RDMA操作的硬件重传不受软件拥塞控制的控制，这具有引发网络拥塞的潜在风险，因为在传输中数据的大小可能比软件拥塞窗口大得多。 Flor通过增加软件重传方案提高了硬件可靠性。 Flor 在 RC QP 中设置较短的重传重试次数，来限制超出软件拥塞控制窗口的数据大小。

如果在重试次数内重传失败，则 RNIC 硬件将 QP 转为错误状态。将QP的错误状态转变为工作状态需要很长时间。相反，Flor 将未完成的运行中 WQE 重新提交给另一个预连接的 QP，称为备份 QP，并将备份 QP 翻转为主 QP 以继续数据传输。

非活动备份 QP 不会消耗 RNIC 上的额外缓存资源，因此对性能没有副作用。根据作者的实践经验表明，与 QP 重新连接相比，切换到备用 QP 花费的时间更少。