RTGWG
Internet-Draft
Intended status: Standards Track
Expires: 3 January 2026

Z. Han, Ed.
R. Pang
Z. Ruan
X. Yi
China Unicom

2 July 2025

Usecase and requirement of deploying PFC and fine-grained flow control
draft-han-rtgwg-codeployment-pfc-fgfc-00

Abstract

The demand for lossless network transmission and the application of flow control mechanisms have expanded from DCNs (Data Center Networks) to WANs(Wide Area Networks). To mitigate PFC - related issues in WANs, the fine - grained flow control is proposed. This mechanism aims to achieve precise control at flow / tenant levels, limits flow control to specified paths and slices, and provides intelligent congestion backpressure. As current DCN already adopts PFC mechanisms, the fine-grained flow control in WANs needs to work with PFC in DCNs to achieve end-to-end flow control.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 3 January 2026.

Copyright Notice

Table of Contents

1. Introduction and Background

DCNs are typically characterized by a limited network scale, short path and predictable traffic patterns, so flow control mechanisms like PFC (Priority Flow Control) and ECN (Explicit Congestion Notification) operate effectively. With the growth of AI LLM distributed training and inference, lossless transmission of massive data between geographically separated data centers is required [I-D.hs-rtgwg-wan-lossless-uc], and the flow control mechanisms need to be extended from DCNs to WANs. Unlike DCNs, WANs are large-scale with complex topologies, long paths, and diverse traffic type. PFC based on port-level feedback ensures lossless transmission of RDMA protocol, by pausing/resuming specific priority queues to prevent congestion. When using it in the WANs, the backpressure from PFC will cause head-of-line blocking, deadlocks, and congestion spreading, which degrade network throughput [I-D.hs-rtgwg-wan-lossless-uc]. To mitigate these issues, the fine - grained flow control is required for WANs.

Fine-grained flow control improves upon the coarse-grained port-based PFC mechanism. It enables precise control at the flow, tenant, or other granular levels, limits flow control to specified paths and slices, and provides intelligent congestion backpressure with granular parameters (pausing time, and buffer thresholds etc.). These capabilities collectively contribute to achieving efficient and refined flow control in WANs.

This draft focuses on the scenarios where PFC is employed in DCNs and the fine-grained flow control is utilized in WANs. Usecase and requirements for the interworking deployment of PFC and fine-grained flow control mechanisms are described, achieving end-to-end flow control through coordination and policy mapping between DCNs and WANs.

2. Terminology

PFC: Priority-based Flow Control

DCN: Data Center Network

WAN: Wide Area Network

RDMA: Remote Direct Memory Access

RoCE: RDMA over Converged Ethernet

3. Interworking deployment of PFC and fine-grained Flow Control

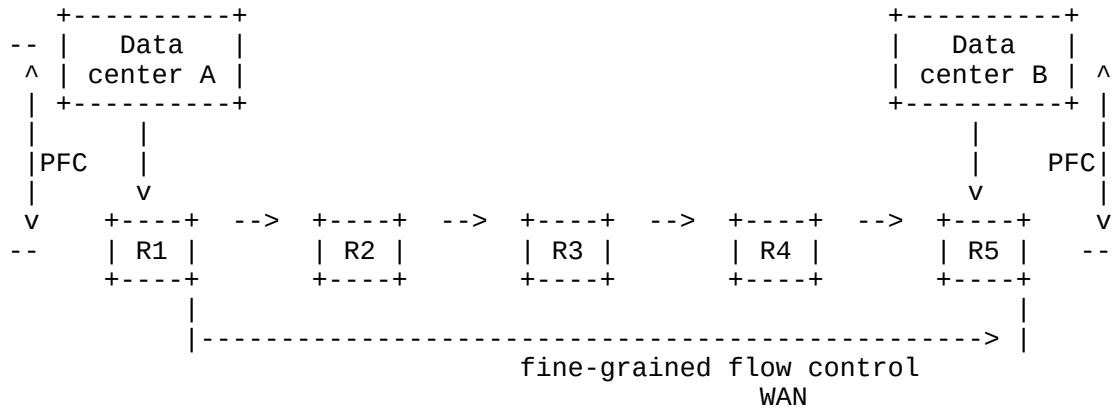

Figure 1: Codeployment of PFC and fine-grained flow control

As shown in Figure 1, there are two data centers, A and B, connected by WAN via nodes R1 -> R2 -> R3 -> R4 -> R5.

The internal nodes of data center A and data center B employ the PFC mechanism. Because most DCN NICs today are optimized for legacy protocols (e.g., Ethernet, DCB) and lack SRv6 processing capabilities. This limitation prevents the direct extension for refined flow control. Hardware/firmware upgrades are needed to enable fine-grained flow control deployment.

WAN nodes R1-R5 deploy fine-grained flow control to avoid PFC backpressure issues, enabling flow/tenant-level congestion handling with granular parameters for precise and intelligent backpressure. WAN nodes support HQOS (Hierarchical Quality of Service) queuing mechanisms and slicing.

Edge nodes R1 and R5 support both PFC and fine-grained flow control, interworking DCN and WAN flow control mechanisms and ensuring seamless end-to-end flow control. The NNI ports of edge nodes R5 and R1 can establish multiple slices, each corresponding to a tenant and supporting 1-8 queues.

4. Procedure of end-to-end flow control

4.1. PFC to fine-grained flow control

TBD

4.2. Fine-grained flow control to PFC

TBD

5. Requirement of joint deployment

TBD

6. Security Considerations

This document does not introduce any new security considerations.

7. IANA Considerations

This document has no IANA actions.

8. Informative References

[I-D.hs-rtgwg-wan-lossless-uc]
Zhengxin, H., He, T., Shi, H., and T. Zhou, "Use Cases and Requirements for Implementing Lossless Techniques in Wide Area Networks", Work in Progress, Internet-Draft, draft-hs-rtgwg-wan-lossless-uc-01, 2 July 2025, <https://datatracker.ietf.org/doc/html/draft-hs-rtgwg-wan-lossless-uc-01>.

Contributors

Authors' Addresses

Zhengxin Han (editor)
China Unicom
Beijing
China
Email: hanzx21@chinaunicom.cn

Ran Pang
China Unicom
Beijing
China
Email: pangran@chinaunicom.cn

Zheng Ruan
China Unicom
Beijing
China
Email: ruanz6@chinaunicom.cn

Xinxin Yi
China Unicom
Beijing
China
Email: yixx3@chinaunicom.cn