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

Z. Han, Ed.
Z. Ruan
R. Pang
Y. Yue
China Unicom
J. Dong
Huawei Technologies
Q. Xiong
ZTE Corporation
2 March 2026

Fine-Grained Flow Control Backpressure Mechanism for Wide Area Networks
draft-han-rtgwg-fine-grained-backpressure-01

Abstract

This document specifies a fine-grained flow control backpressure mechanism for Wide Area Networks (WANs). Leveraging data-plane congestion detection and notification, it enables millisecond-level congestion response. The mechanism enhances Layer 2 PFC by extending network protocols (e.g., ICMPv6) for congestion backpressure messaging in WANs, and leverages network slicing isolation to provide fine-grained flow control at tenant or task granularity. It addresses the limitations of traditional flow control mechanisms in WAN environments through fast and precise backpressure, and supports multi-hop propagation of congestion notifications along the forwarding path.

Table of Contents

1. Introduction and Background

With the rapid development of High-Performance Computing (HPC), remote healthcare, multimedia content production, and AI-Generated Content (AIGC)applications, the volume, velocity, and variety of data are growing exponentially. This poses higher requirements for the efficiency and reliability of massive data transmission across Wide Area Networks (WANs). WANs are characterized by large scale, complex topology, long round-trip times, diverse service types, continuously increasing load, and frequent high-intensity burst traffic, making them prone to congestion.

Traditional congestion control mechanisms [I-D.dong-fantel-problem-statement] like Priority Flow Control(PFC) and Explicit Congestion Notification (ECN) face limitations in WAN environments. PFC provides coarse-grained port-based flow control that can lead to congestion spreading, head-of-line blocking, and deadlocks. ECN requires end-host participation with slow and inaccurate responses, making it unsuitable for long-distance transmission in WANs.

This document proposes a fine-grained flow control backpressure mechanism implemented in the data plane, achieving millisecond-level response times. The controller is involved only in the pre-deployment phase for path planning and static parameters configuration. The mechanism extends network protocols (e.g., ICMPv6) and leverages network slicing isolation to provide precise congestion backpressure at tenant or task granularity. It also supports multi-hop congestion notification along the traffic path.

2. Terminology

* Fine-grained flow control(fgfc): An enhanced PFC mechanism that enables precise flow control at tenant or other granular levels, limits flow control to specified paths and slices, and provides intelligent congestion backpressure to prevent network congestion [I-D.han-rtgwg-wan-lossless-terms]

* Congestion Backpressure Message: Network protocol message carrying congestion information and flow control policies, which can be implemented using ICMP, UDP, or other suitable protocols.

* Dual-Watermark Mechanism: Congestion detection using Xoff (trigger) and Xon (release) thresholds.

* Elastic bandwidth: The bandwidth is dynamically adjusted according to the network conditions to improve network bandwidth utilization and network transmission efficiency[I-D.han-rtgwg-wan-lossless-terms].

* SRv6: IPv6 Segment Routing as defined in [RFC8754].

* SRH: Segment Routing Header as defined in [RFC8754].

3. Fine-grained flow control mechanism in WAN

The fine-grained flow control(fgfc) backpressure mechanism operates entirely in the data plane to achieve millisecond-level congestion response, avoiding buffer overflow caused by control-plane latency.

The controller is involved only in the pre-deployment phase. Based on the network topology and service requirements, controller pre-calculates the forwarding path and distributes static information (such as address information, slice configurations, watermark thresholds, and backpressure message formats) to the routing tables of data-plane devices. It does not participate in real-time congestion handling. The address information includes node addresses, port addresses, and SRv6 behavior addresses[I-D.ruan-spring-priority-flow-control-sid].

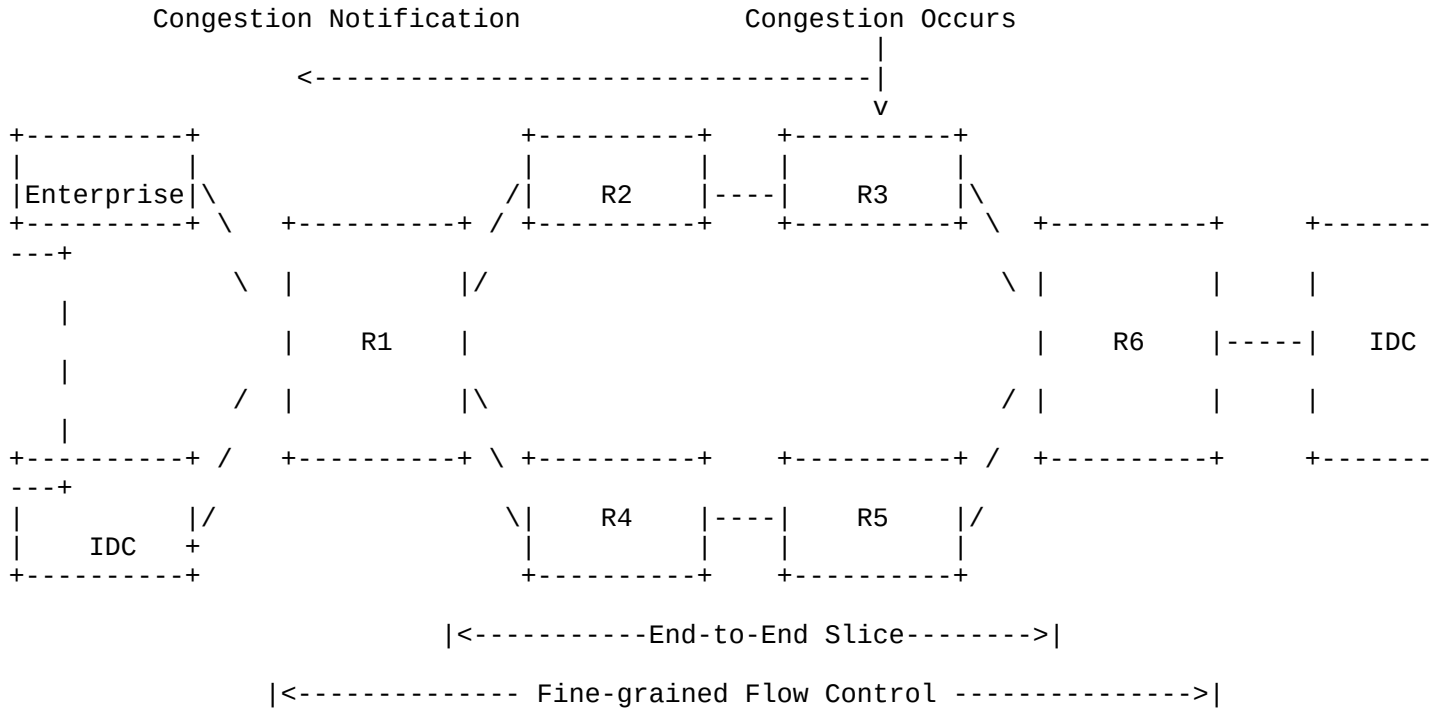

Figure 1: Fine-grained flow control mechanism in WAN

3.1. Network slicing & Elastic bandwidth

End-to-end network slices are established between WAN nodes R1 and R6, with each slice assigned to a single tenant and supporting 1 to 8 PFC queues. Elastic bandwidth is enabled via the configuration of a Committed Information Rate (CIR) and Peak Information Rate (PIR), which optimizes bandwidth resource utilization and improves transmission efficiency.

When the network is idle, tenant traffic may preempt available idle resources to scale up to the PIR. During network congestion, the backpressure mechanism enforces a fallback to the CIR to ensure the minimum guaranteed bandwidth for the tenant. CIR may optionally be set to zero, in which case lossless transmission is ensured entirely by the flow control mechanism.

3.2. Dual-watermark congestion detection

Each node deploys a dual watermark (Xon/Xoff) monitoring mechanism for the queue,which is used to detect congestion and avoid frequent backpressure oscillation.

When the queue length exceeds Xon threshold, the node generates a backpressure message to upstream nodes to reduce the sending rate.When the queue length falls below Xoff threshold, the node may send a release message (or stop sending backpressure) to resume normal forwarding.The threshold can be configured according to the queue buffer depth and business latency requirements.

The dual watermark mechanism avoids the frequent switching of backpressure and cancellation caused by small fluctuations of the queue depth, and ensures the stability of the flow control process.

3.3. Backpressure message generation and propagation

When the R3 node detects congestion (queue > Xon), it triggers a fine-grained flow control backpressure message. The message is sent upstream nodes that support fine-grained flow control(fgfc).

Per-hop backpressure: If all intermediate nodes in WAN support fgfc, the backpressure is applied hop-by-hop, gradually reducing the rate at each node.

Cross-hop backpressure: If only some nodes support fgfc, the backpressure message bypasses non-supporting nodes and is processed only by capable nodes.

The time from congestion detection to backpressure message transmission is controlled within a few milliseconds. Upon receiving the message, the upstream node identifies the affected queue(s) and performs rate limiting within nanoseconds or microseconds, achieving fast congestion response,near-zero packet loss and high throughput.

The backpressure path can be carried in the SRv6 Segment Routing Header (SRH) via encapsulation or insertion. The SRH path is computed by the controller or locally by nodes based on the forward path and pre-configured policies.

4. Congestion backpressure message formats

The Fine-Grained Flow Control mechanism conveys congestion information by extending network protocols. The backpressure message can based on ICMPv6 notification messages, enabling data-plane congestion notification without host involvement. The format is shown in Figure 1.

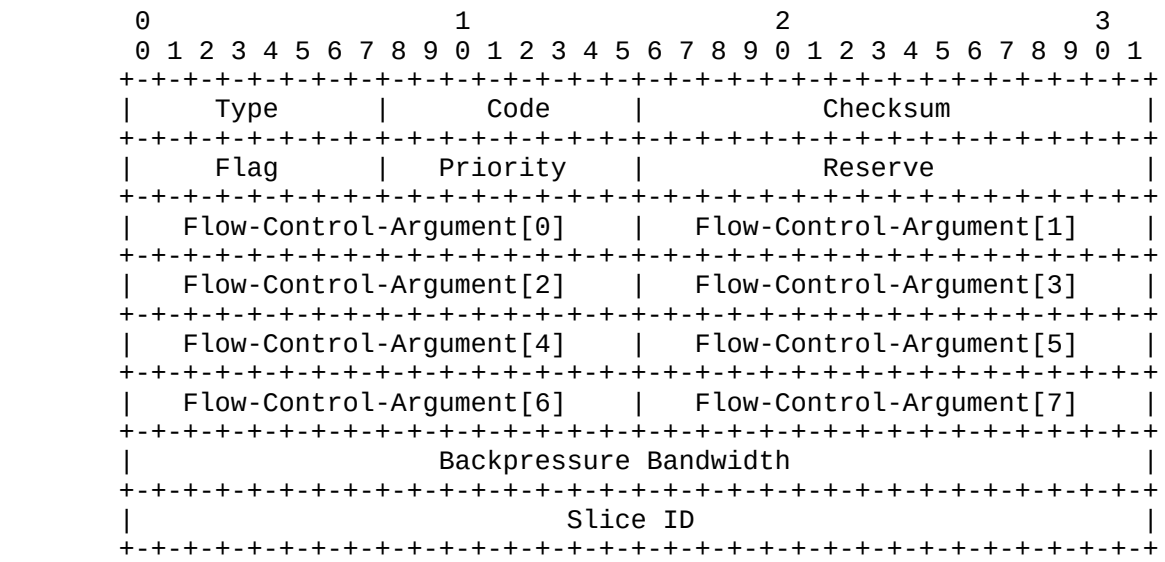

Figure 2: ICMPv6-based Backpressure Message Format

* Type: 1 byte. ICMPv6 message type. A new Type value (170) is suggested to be assigned by IANA to identify this message as a congestion backpressure notification message.

* Code: 1 byte. Set to 0 for fgfc messages.

* Checksum: 2 bytes. Standard ICMPv6 checksum.

* Flag: 1 byte. Flags field. Currently unused and MUST be set to zero on transmission and ignored on receipt.

* Priority: 1 byte. Flow control priority. Each bit corresponds to one of eight queues. A bit set to 1 indicates that the corresponding queue is subject to flow control (backpressure); a bit set to 0 indicates no backpressure.

* Reserve: 2 bytes. Reserved for future use. MUST be set to zero on transmission and ignored on receipt.

* Flow-Control-Argument[]: 16 bytes. Currently used to specify the backpressure duration. It is interpreted as an array of eight 16-bit unsigned integers, each representing the pause time in microseconds for the corresponding queue (matching the priority queues).

* Backpressure Bandwidth: 4 bytes. Specifies the bandwidth throttling value (rate limit) to be applied.

* Slice ID: 4 bytes. Tenant identifier, used to identify the tenant path or slice to which the backpressure applies.

This format allows queue-level pause and tenant-specific backpressure, ensuring congestion does not spread across slices. The message is processed entirely in the data plane without host involvement.

5. Security Considerations

This document does not introduce any new security considerations.

6. IANA Considerations

This document has no IANA actions.

7. References

7.1. Normative References

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

7.2. Informative References

[I-D.han-rtgwg-wan-lossless-terms]
Zhengxin, H., Pang, R., and T. He, "Terminology for Implementing Lossless Techniques in Wide Area Networks", Work in Progress, Internet-Draft, draft-han-rtgwg-wan-lossless-terms-01, 29 June 2025, <https://datatracker.ietf.org/doc/html/draft-han-rtgwg-wan-lossless-terms-01>.

[I-D.dong-fantel-problem-statement]
Dong, J., McBride, M., Clad, F., Zhang, Z. J., Zhu, Y., Xu, X., Zhuang, R., Pang, R., Lu, H., Liu, Y., Contreras, L. M., Mehmet, D., and R. Rahman, "Fast Network

Notifications Problem Statement", Work in Progress, Internet-Draft, draft-dong-fantel-problem-statement-05, 2 February 2026, <https://datatracker.ietf.org/doc/html/draft-dong-fantel-problem-statement-05>.

[I-D.ruan-spring-priority-flow-control-sid]
Ruan, Z., Liu, Y., Han, M., Zhengxin, H., and Ying, "SRv6 behavior extention for Flow Control in WAN", Work in Progress, Internet-Draft, draft-ruan-spring-priority-flow-control-sid-03, 27 February 2026, <https://datatracker.ietf.org/doc/html/draft-ruan-spring-priority-flow-control-sid-03>.

Authors' Addresses

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

Zheng Ruan
China Unicom
Email: ruanz6@chinaunicom.cn

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

Yi Yue
China Unicom
Beijing
China
Email: yuey80@chinaunicom.cn

Jie Dong
Huawei Technologies
Email: jie.dong@huawei.com

Quan Xiong
ZTE Corporation

Email: xiong.quan@zte.com.cn