Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 23 April 2026

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

20 October 2025

Fine-Grained Flow Control Backpressure Mechanism for Wide Area Networks
draft-han-rtgwg-fine-grained-backpressure-00

Abstract

This document specifies a fine-grained flow control backpressure mechanism for Wide Area Networks (WANs). The mechanism enables precise congestion notification and flow control at tenant or task granularity through extended network protocols and controller-assisted path discovery. It addresses the limitations of traditional flow control mechanisms in WAN environments by providing intelligent backpressure with detailed congestion information, with specific applicability in SRv6-based networks.

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 23 April 2026.

Table of Contents

1. Introduction

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 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 that extends network protocols (including but not limited to ICMP and UDP) to carry detailed congestion information and utilizes controller assistance for optimal path discovery. The mechanism enables precise flow control at tenant or task granularity, provides intelligent backpressure strategies, and supports multi-hop congestion notification along the traffic path. The solution is particularly relevant for SRv6-based networks where precise path control and slicing capabilities are essential.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Fine-Grained Flow Control: Flow control mechanism that operates at tenant or task granularity rather than port or queue level.

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

First Path Forwarding Node: The node experiencing congestion that initiates the backpressure process.

Second Path Forwarding Node: The upstream node capable of handling congestion through fine-grained flow control.

Controller: Central entity that maintains network topology and node capabilities, assisting in optimal backpressure path discovery.

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

SID: Segment Identifier in SRv6 networks.

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

3. System Architecture

The fine-grained flow control backpressure system consists of three main components, with specific considerations for SRv6 networks:

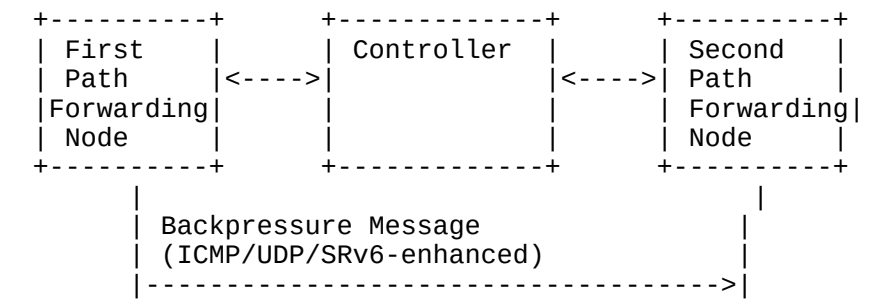


Figure 1: System Architecture

In SRv6 environments, the backpressure mechanism leverages the existing SRv6 infrastructure:

- First Path Forwarding Node: Detects congestion and initiates backpressure process by contacting the controller and generating backpressure messages. In SRv6 networks, this node maintains SID information and SRH processing capabilities.

- Controller: Maintains network topology, node capability information, and SRv6 SID allocations. Determines optimal backpressure paths and policies, considering SRv6 path segments.

- Second Path Forwarding Node: Receives backpressure messages and performs fine-grained flow control on specified traffic. Supports SRv6 processing and slice-aware congestion handling.

4. Fine-Grained Flow Control Backpressure Mechanism

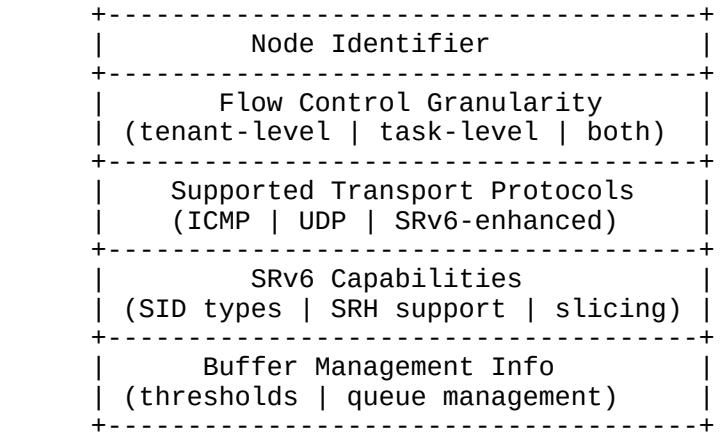
4.1. Flow Control Capability Discovery

Before participating in the fine-grained flow control mechanism, each forwarding node MUST register its capabilities with the controller. The controller maintains a comprehensive view of flow control capabilities throughout the network to facilitate optimal backpressure path selection.

The capability discovery process include the following information:
- Node identifier and addressing information
- Flow control granularity support (tenant-level, task-level, or both)
- Supported flow control mechanisms and algorithms
- Buffer management capabilities and thresholds
- Supported transport protocols for backpressure messages (ICMP, UDP, etc.)
- SRv6-specific capabilities (if applicable), including:
  * Supported SID types and functions
  * SRH processing capabilities
  * Slice awareness and isolation capabilities
  * Path segment manipulation support

The capability discovery follow this procedure:
1. Node initiates capability advertisement to controller
2. Controller validates and stores flow control capabilities
3. Periodic capability updates SHALL be sent to reflect any changes
4. Controller MAY proactively query node capabilities as needed

Example capability advertisement format:



This capability discovery enables the controller to maintain a real-time understanding of nodes capable of fine-grained flow control throughout the network, with specific awareness of SRv6 capabilities and protocol support variations.

4.2. Congestion Detection and Controller Assistance

When a forwarding node detects congestion, it initiate the backpressure process by sending a request to the controller.

The congestion detection be based on monitoring of:
- Queue length
- Link utilization
- Packet loss rate
- Delay
- Resource utilization rate
- In SRv6 networks: per-slice queue utilization and SID-specific congestion metrics

The request to controller include:
- Identity of the congested node
- Request for second path forwarding node that meets congestion handling conditions
- In SRv6 networks: current SID list, slice information, and path segment details

The congestion handling conditions include:
- Ability to perform data traffic control at tenant or task granularity
- Existence of shortest forwarding path between nodes
- In SRv6 networks: compatibility with SRv6 path segments and slicing

Upon receiving the request, the controller:
1. Identify suitable second path forwarding nodes based on network topology and node capabilities
2. Determine the optimal backpressure path, considering SRv6 SID lists when applicable
3. Generate appropriate backpressure policies
4. Return the second path forwarding node address information and backpressure policies to the first path forwarding node

4.3. Backpressure Message Generation and Forwarding

After receiving the controller’s response, the first path forwarding node generate backpressure messages and forward them to the second path forwarding node.

The backpressure message generation follow these steps:

1. Receive backpressure policy information from controller
2. Select appropriate transport protocol (ICMP, UDP, or SRv6-enhanced) based on node capabilities and network configuration
3. Generate backpressure message containing:
   - Backpressure notification path
   - Slice information
   - Backpressure policy information
   - In SRv6 networks: relevant SID information and SRH details
4. Send the message to second path forwarding node using IP-based transmission

The backpressure message includes the following information:

- Message type and protocol identification
- Checksum or integrity verification
- Backpressure notification path
- Number of slices
- Slice identifier
- Control mode (0 for congestion backpressure, 1 for cache release)
- Traffic rate
- Cache capability
- Backpressure policy information
- Protocol-specific extensions (e.g., SRH for SRv6-enhanced messages)

4.4. Multi-hop Backpressure Propagation

If the second path forwarding node cannot fully handle the congestion, the backpressure process propagates further upstream.

The multi-hop propagation follows this process:

1. Second path forwarding node reports its cache occupancy rate to first path forwarding node
2. If cache occupancy rate exceeds threshold (e.g., 80%), first path forwarding node requests updated backpressure policy from controller
3. Controller generates new backpressure policy based on current network traffic information
4. First path forwarding node generates new backpressure message and sends it to third path forwarding node (upstream from second node)
5. Process continues until congestion is resolved or reaches the traffic source

In SRv6 networks, the backpressure path SHOULD follow the reverse direction of the SRv6 SID list to ensure optimal congestion handling along the established path segments.

4.5. SRv6 Integration and Path Handling

For SRv6-enabled networks, the backpressure mechanism leverages SRv6 capabilities for enhanced congestion control:

- SID-based Path Identification: Backpressure messages SHOULD include SID information to precisely identify the congested path segment.

- SRH for Backpressure Routing: When available, SRH MAY be used to ensure backpressure messages follow specific paths, bypassing nodes that do not support fine-grained flow control.

- Slice-aware Congestion Management: SRv6 slicing capabilities SHOULD be utilized to isolate congestion to specific slices and prevent cross-slice interference.

- Path Segment Optimization: The controller SHOULD consider SRv6 path segments when determining backpressure paths to minimize the number of hops and reduce latency.

5. Backpressure Message Formats

5.1. ICMP-based Backpressure Message

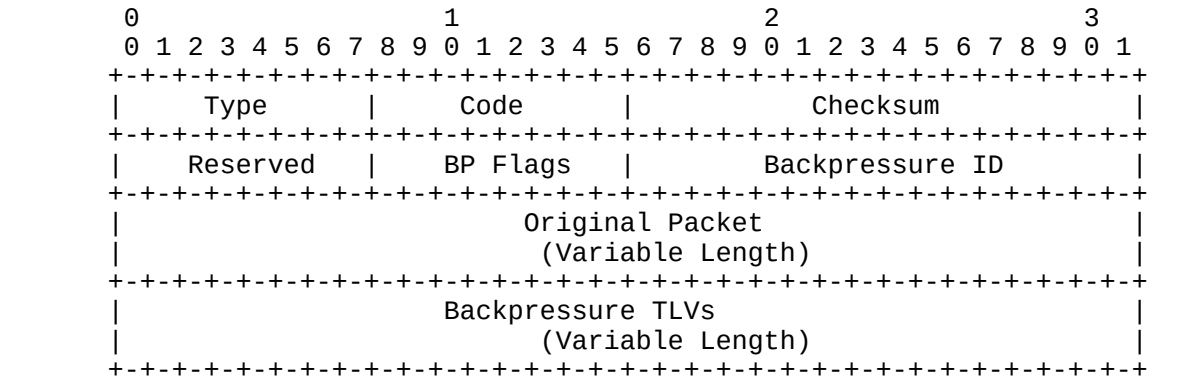


Figure 2: ICMP-based Backpressure Message Format

Fields:
- Type: ICMP message type (to be assigned by IANA)
- Code: Subtype (0 for tenant granularity, 1 for task granularity)
- Checksum: Standard ICMP checksum
- BP Flags: Backpressure flags
- Backpressure ID: Identifier for the backpressure session
- Original Packet: Header of the packet that triggered congestion
- Backpressure TLVs: Type-Length-Value structures carrying detailed backpressure information

5.2. SRv6-enhanced Backpressure Message

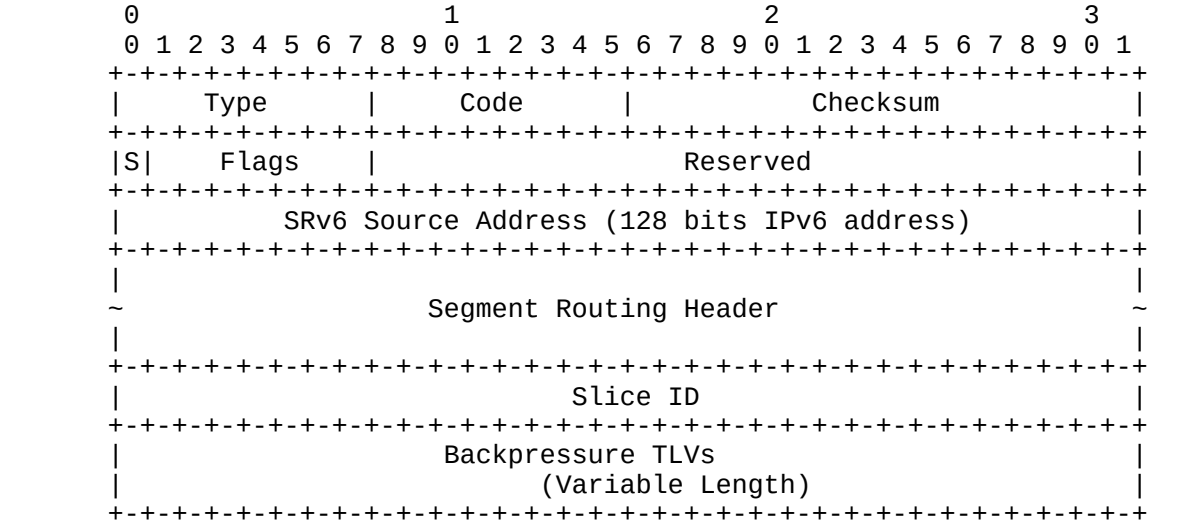


Figure 3: SRv6-enhanced Backpressure Message Format

Fields:
- Type: Message type (ICMP or other as appropriate)
- Code: Message code
- Checksum: Protocol checksum
- Flags: SRv6-specific flags (S bit indicates presence of Slice ID)
- SRv6 Source Address: Source address of the SRv6 message
- Segment Routing Header: SRH as defined in [RFC8754]
- Slice ID: Identifier for the network slice
- Backpressure TLVs: Type-Length-Value structures carrying detailed backpressure information

5.3. Backpressure Path TLV

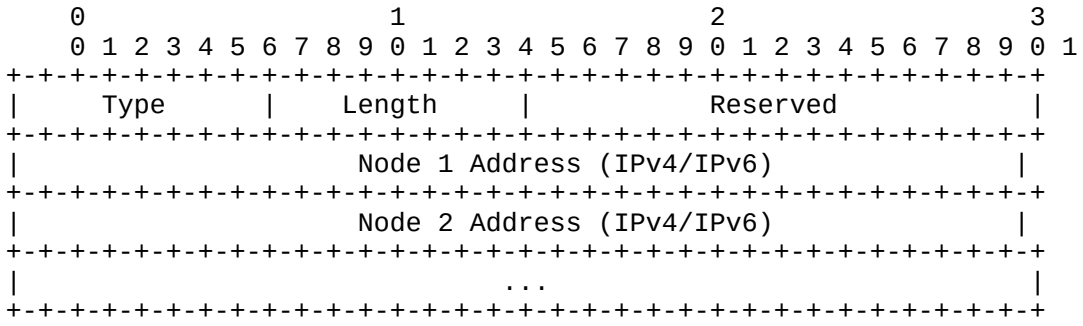


Figure 4: Backpressure Path TLV

Fields:
- Type: TLV type for backpressure path (to be assigned by IANA)
- Length: Total length of the TLV in bytes
- Node Addresses: Sequence of node addresses forming the backpressure path

5.4. Slice Information TLV

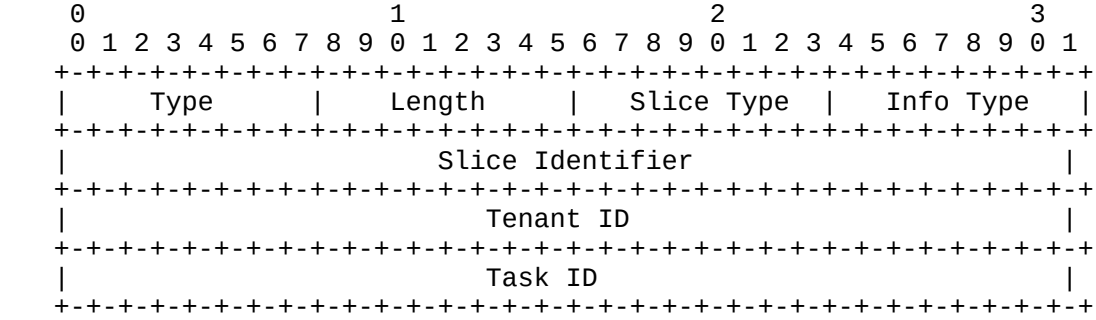


Figure 5: Slice Information TLV

Fields:
- Type: TLV type for slice information (to be assigned by IANA)
- Length: Total length of the TLV in bytes
- Slice Type: Type of network slice
- Info Type: Type of information (tenant or task)
- Slice Identifier: Unique identifier for the slice
- Tenant ID: Identifier for the tenant
- Task ID: Identifier for the task

5.5. Backpressure Policy TLV

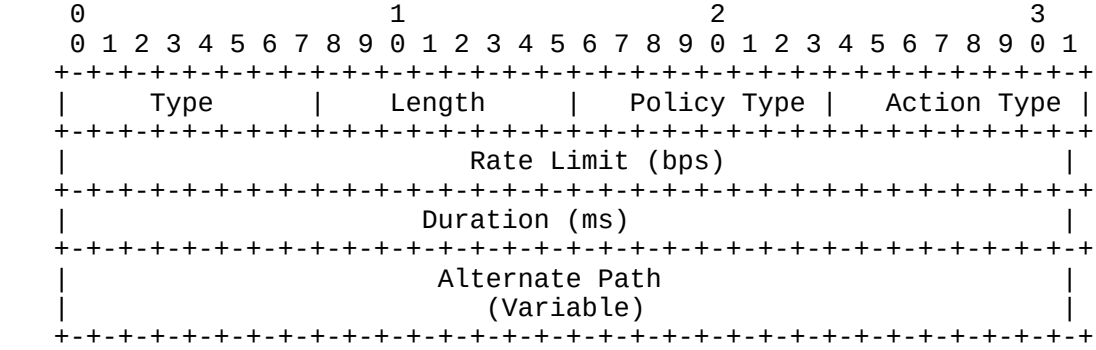


Figure 6: Backpressure Policy TLV

Fields:
- Type: TLV type for backpressure policy (to be assigned by IANA)
- Length: Total length of the TLV in bytes

- Policy Type: Type of backpressure policy
- Action Type: Specific action to be taken
- Rate Limit: Maximum allowed traffic rate in bits per second
- Duration: Time duration for which the policy should be applied
- Alternate Path: Suggested alternative path for traffic (variable length)

6. Security Considerations

TBD;

7. IANA Considerations

TBD;

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017.

[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.

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