

Congestion Control Based on SRv6 Path

draft-liu-rtgwg-srv6-cc-01

Yisong Liu (China Mobile)

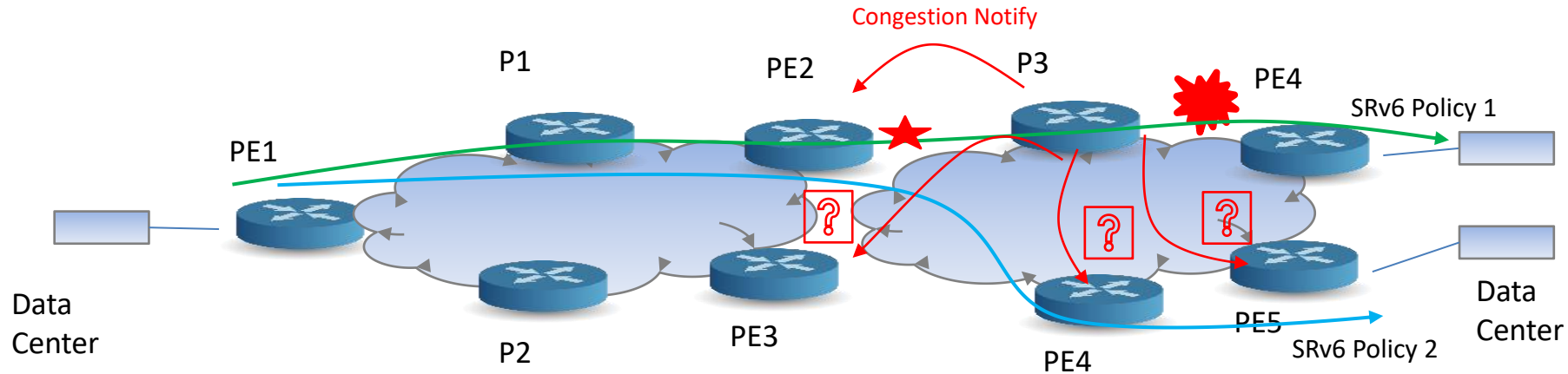
Junda Yao (Huawei)

Changwang Lin (New H3C Technologies)

Xiao Min (ZTE)

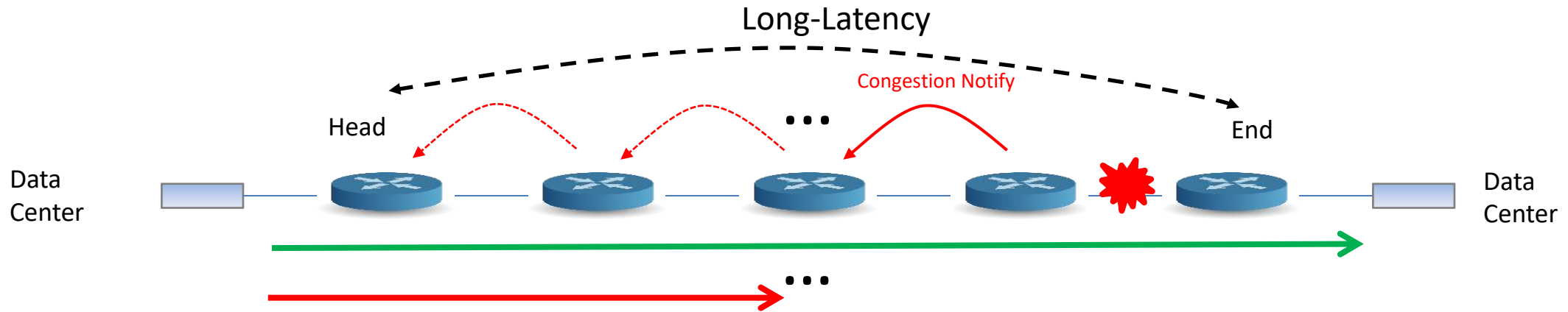
IETF 125

Challenge 1/3: Imprecise Congestion Notification in WANs



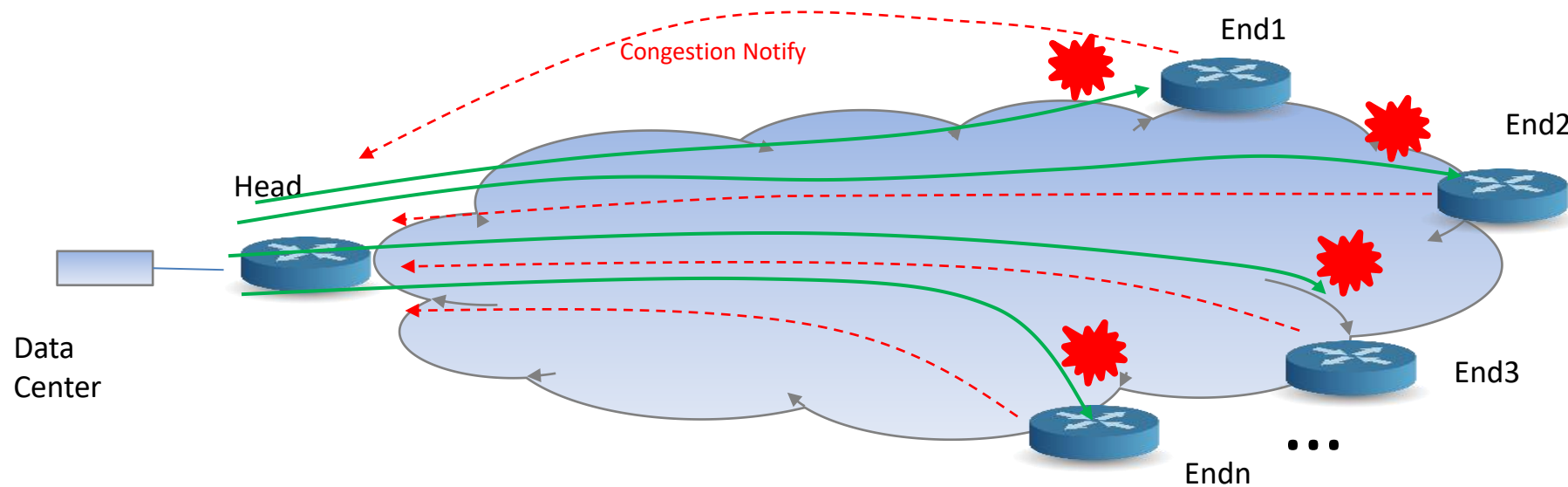
- Traditional PFC relies on ethernet multicast frames to propagate congestion signals
- Problem in WANs
 - In complex, meshed WAN topologies, multicast-based signaling fails to accurately trace the path back to the precise upstream SRv6 source nodes
- Consequences
 - Trigger incorrect suppression of non-offending flows
 - Spreading service impact to unrelated, healthy traffic

Challenge 2/3: Long-Latency SRv6 Paths



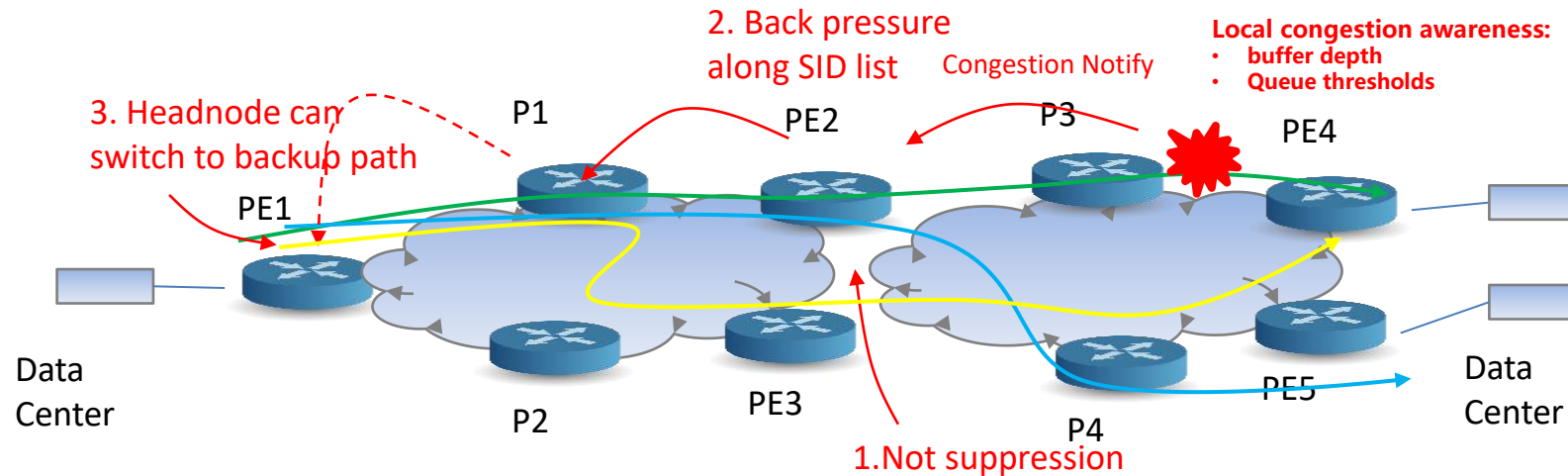
- Inherent WAN Characteristic
 - End-to-end SRv6 paths span large geographic distances, introducing significant, unavoidable latency
- Ineffective End-to-End Backpressure
 - Requiring congestion signals to travel all the way back to the source (e.g., data center) results in an unacceptably slow response
- Critical Requirement
 - Traffic management must be performed directly on the SRv6 data path to enable immediate reaction and prevent prolonged service degradation

Challenge 3/3: Control Overhead at the SRv6 Head Node



- Scale of Responsibility
 - A single head node is responsible for managing and initiating numerous SRv6 paths across the WAN
- Centralized Bottleneck Risk
 - Sending all fine-grained congestion notifications back to this central point creates a severe processing bottleneck
- Distributed Solution
 - Offloading traffic control decisions to intermediate nodes along the path is critical to alleviate head node overload

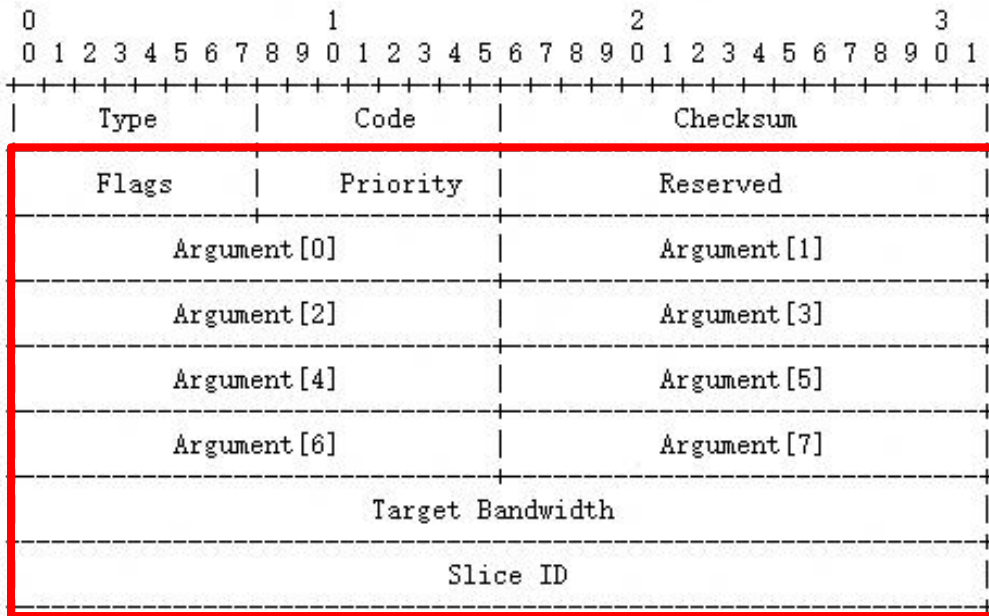
SRv6 Congestion Notification Proposal



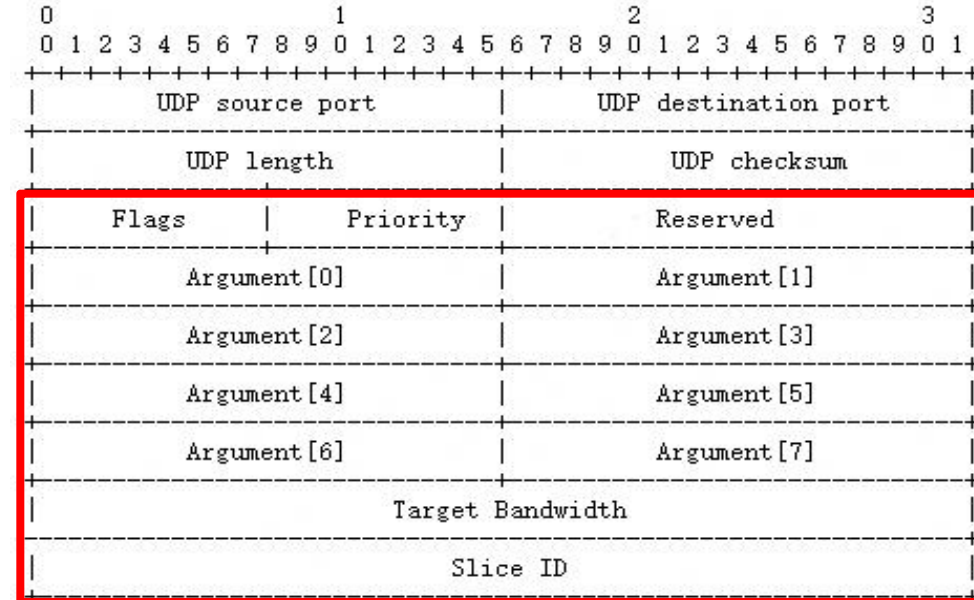
- Head node PE1 encapsulates SID list: P1->PE2->P3->PE4, src=PE1, dst=PE4
- Each node(P1/PE2/P3/PE4) forwards using local SID tables, Verifies slice-related info during forwarding
- Congested node (e.g., P3)
 - Detects priority queue buffer overload
 - Sends a congestion notification message to upstream node (PE2) , Includes: priority queue where congestion occurs, congestion control parameter information (such as pause-time and/or target bandwidth), and slice ID.
- Receiving node (PE2)
 - Receives the notification and adjusts the forwarding rate based on local capacity.
 - Congestion cannot be suppressed, detects local priority queue overload, Sends a notification to upstream node (PE2→P1)
- Head Node Remediation
 - Final fallback: Path rebalancing or Alternate path selection

Congestion Notification Message Format

ICMPv6



UDP



Flags: Contains special flags. not defined.

Priority: Queue priority identifier, each priority queue occupies 1 bit (from high-order to low-order bits representing high priority to low priority respectively).

If set 1, indicates that the priority queue is suppressed due to congestion control; If set 0, indicates that suppression is released from the priority queue.

Argument[] : Congestion control parameter information. By default(when all flag bits are 0), the meaning of argument is pause-time, measured in microseconds.

Target Bandwidth: Indicates the target bandwidth information for expectation suppression. The default value is 0.

Slice ID: The identifier for the slice experiencing congestion.

Next Steps

- Seeking review and feedback from WG
- Welcome involved to help refine and advance this proposal

Thanks!