draft-yan-rtgwg-srv6-constrain-analysis-00

IETF-102, Montreal

Shen Yan yanshen@huawei.com

Zhe Chen chenzhe17@huawei.com

July 18, 2018

MOTIVATION

In this document, we analyze the constrains of SRv6's design in some scenarios.

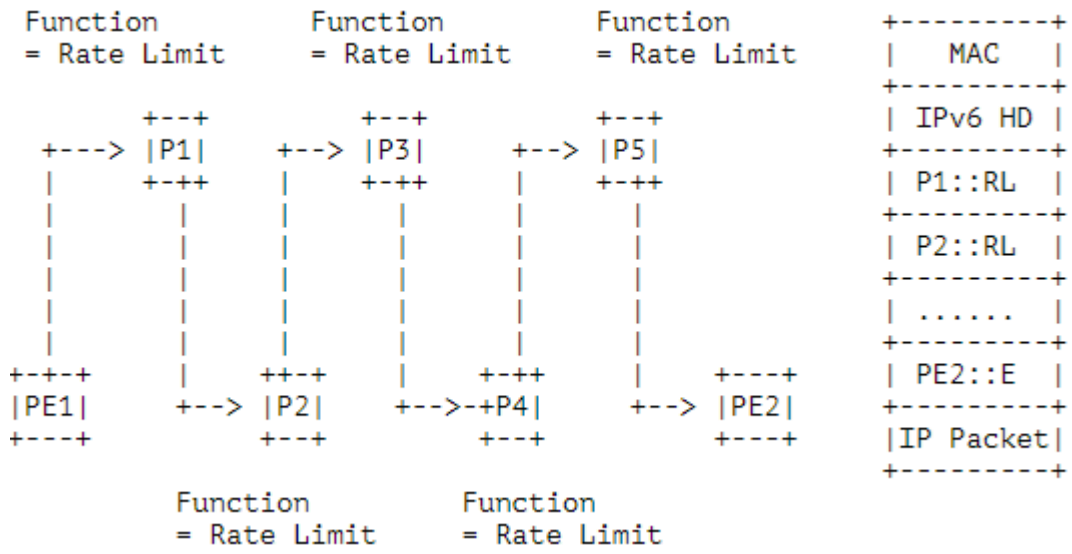
Try to make extensions (improvements) based on SRv6 or design a new protocol encapsulation.

CONSTRAINS - Segment Consumption

The 128 bits SID of SRv6 is a LOC:FUNCT couple design. As shown in figure, PE1 sends a flow to PE2 by shortest path. Each router on the path is required to execute the function of Rate Limit (RL). In this example, the total cost before the original IP Packet will be **158** Bytes (14+40+8+16*6).

The average packet size on the Internet is less than 500 bytes [1] and the design in this case will occupy more than 30%.

- Long header.
- Low efficiency. The current network processor reads normally less than 100 bytes at one time. If the header is too long, it needs more time slot to process.

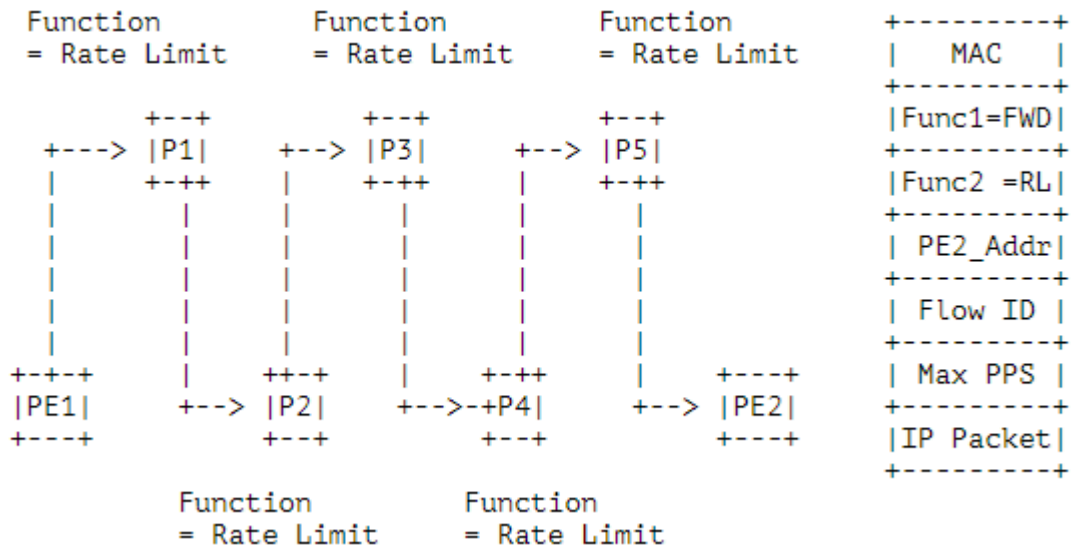


[1] Packet size distribution comparison between Internet links, *https://www.caida.org/research/traffic-analysis/pkt_size_distribution/graphs.xml*

POTENTIAL SOLUTIONS - Segment Consumption

The potential solution is to de-couple instruction and locator carried in the data packet and use globalized instructions instead of the local function code. (Globalized instruction is a universal instruction code that could be recognized by every node in a domain.)

As shown in the figure, to limit packet rate for a specific flow, the packets sent from PE1 to PE2 SHOULD carry two globalized instructions. One has the semantic of “shortest-path forwarding the packet according to PE2‘s address”, and the other has the semantic of “limit the packet rate based on flow identifier and the given maximal packet per second”. The “Func” and “Flow ID”/”Max PPS” may be much shorter than 128 bit. We try to reduce the header length less than **50 Bytes**, which is much shorter than current design.

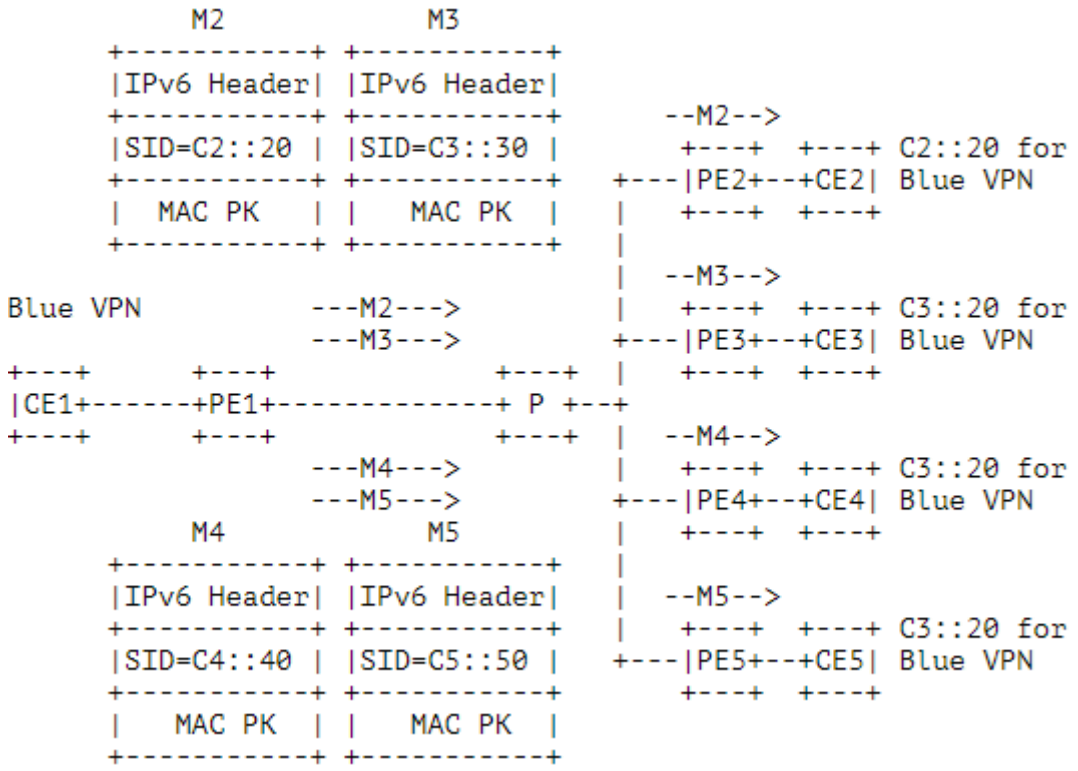


CONSTRAINS - Multicast

In current SRv6 scheme, the multicast packet MUST be replicated at the ingress PE because the locator is contained in SIDs and the egress PE uses different SIDs for the same VPN.

As shown in Figure, CE1, CE2, CE3, CE4 and CE5, construct a Blue VPN. CE1 sends a broadcast frame to all the other CEs.

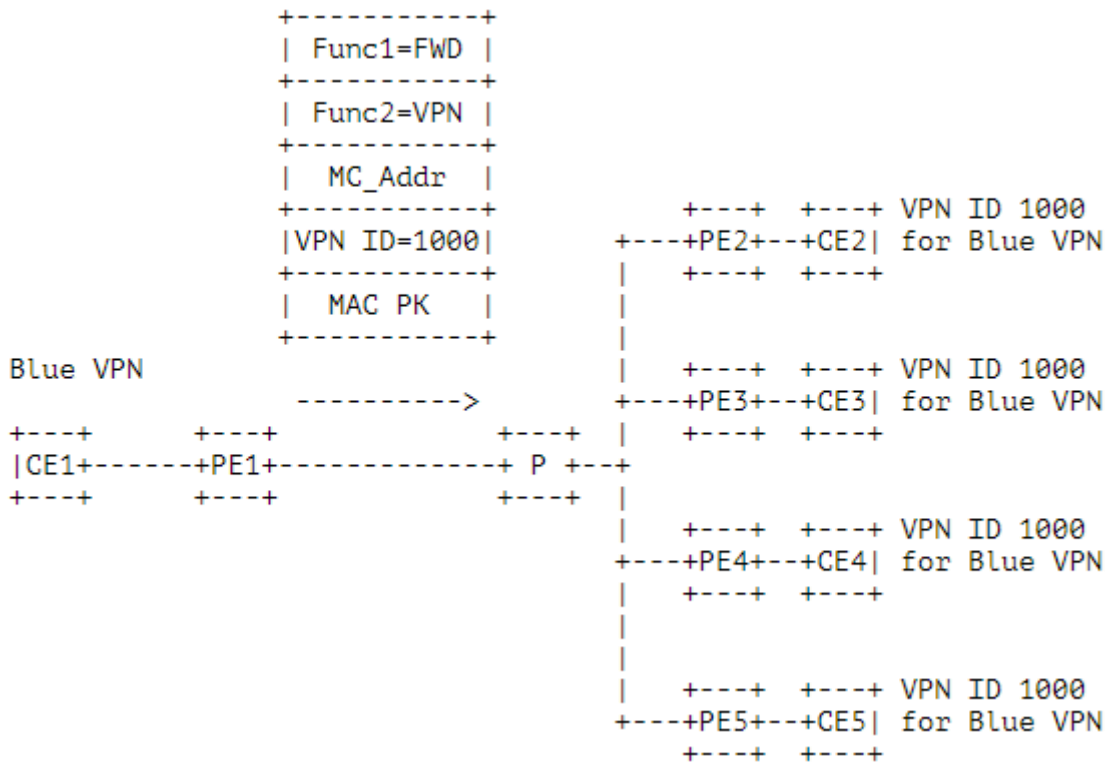
Because of the localized semantics, different routers use different SIDs for the same instruction / metadata. Therefore, the multicast packet MUST be replicated at PE1 instead of any P-routers (P).



POTENTIAL SOLUTIONS - Multicast

This is not an unsolvable problems. If the locator and function can be de-coupled meanwhile all the P-routers perform same operation according to a uniform instruction suite, all the multicast packets will be same in the egress router so that the packet could be replicated at any P-routers.

As shown in the figure, after receiving a broadcast Ethernet frame from CE1, the ingress node (PE1) only need to encapsulate the frame into a single packet, and the packet SHOULD carry two globalized instructions. One has the semantic of "forwarding and replicating (if needed) the packet according to the multicast address of group PE2-PE5", and the other has the semantic of "if there is no next-hop, striping the encapsulated instructions, looking up the VRF of blue VPN and forwarding the packet accordingly". Each transit node in the network forwards and replicates (if needed) the packet based on the multicast address, and the egress nodes perform corresponding VPN actions.



SHORT SUMMERY

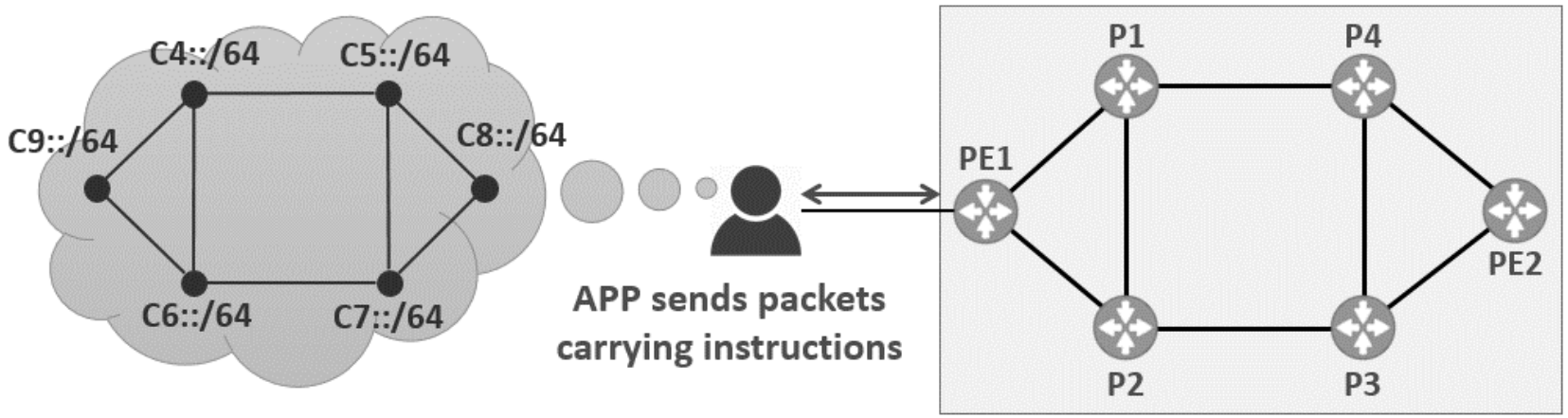
Decouple & Reuse

Globalized Instruction

Header
SID (Locator : Function)
SID (Locator : Function)
SID (Locator : Function)
SID (Locator : Function)
Original Packet

Header
Function
Function
Data (e.g. Locator)
Data (e.g. Flow ID)
Original Packet

UPCOMING – APP and Security



Localized Semantics → **Router's locators are coupled with instructions**

Increase security risks

In case of APP programming, APP/User MUST know the topology/locator information of ISP's network!

Next Step

Continue to evolve the draft

Demo in IETF-103

Welcome discuss and feedback

THANKS