Layer Independent Generic Fragmentation/ESP

draft-zzhang-tsvwg-generic-transport-functions

Jeffrey Zhang, Ron Bonica, Kireeti Kompella

Juniper Networks

IETF 109

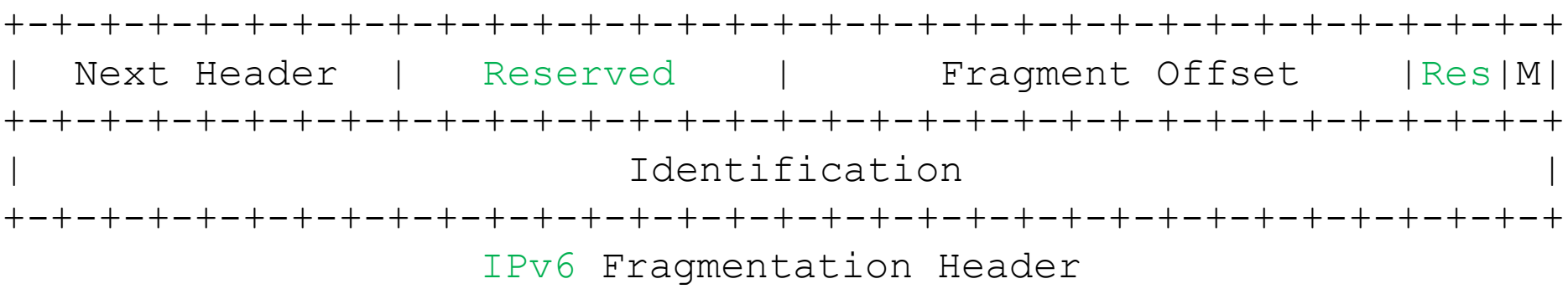
Observations

- With EVPN-MPLS, when a packet received from CE is larger than the underlay path MTU, there is no way to get it through the underlay
  - A workaround is MPLS (service label) over IP – impose service label, encapsulate the packet in IPv4/v6, fragment it, send across, reassemble, decapsulate IP, and then identify the Bridge Domain with the service label
    - MPLS is not used for transport at all
    - Large IPv6 overhead
  - PW/VPLS uses Control Word's sequence number for fragmentation/reassembly – RFC 4623 – though not applicable for EVPN
    - EVPN either does not use CW or uses all-0 CW
    - PW/VPLS fragmentation/reassembly is done in PW context, which EVPN does not have
- IPv6 Fragmentation can be viewed as independent of IPv6
  - As long as context for Identification field in the frag header is available
    - (source, destination) address in case of IP
    - Is "destination" really needed?

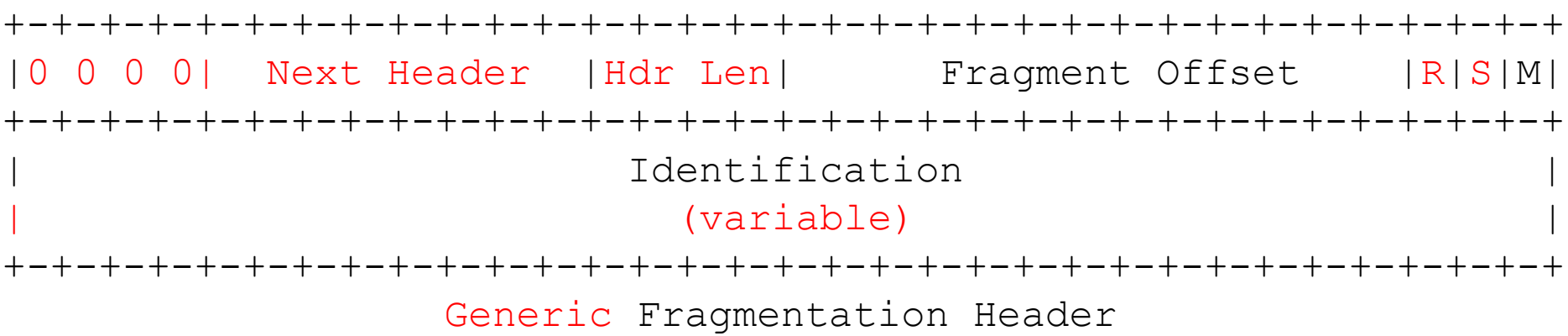
Proposal

- Support independent fragmentation/reassembly function in a shim layer
- In the EVPN-MPLS example:
  - Ingress PE imposes service labels, then fragment the packet w/o IP encapsulation
    - A Generic Fragmentation Header (GFH) is prepended, with Next Header value set to “MPLS” to indicate MPLS is the payload
    - Compared to IPv6 Frag Header, GFH has additional information about the source
  - Ingress PE imposes a GFH label (with semantics “GFH follows”), imposes transport labels and send traffic
    - The GFH label could be individually advertised by the egress PEs or a well-known (but not special) label agreed by all routers for this purpose
  - Egress PE sees the GFH label, reassembles the packet, and then hands to MPLS for further handling based on the service labels
    - Because “next header” is “MPLS”

Fragmentation Header



IPv6 Fragmentation Header



Generic Fragmentation Header

Generic Fragmentation Header

- 0000 nibble: Prevent ECMP hashing from mistaking as IP packet
- Hdr Len: Header length in 8-octet unit
- Identification field in GFH is an arbitrarily sized free-form field
  - If the outer encap header can identify the source, the Identification field can be a simple 32-bit number as in IPv6 Fragmentation header
  - Otherwise the field can additionally encode an IPv4/IPv6 address or any opaque number that can identify the source within the domain where the fragmentation/reassembly happens
  - In case of MPLS transport, the GFH label could also carry additional semantics like identifying the source (e.g., the egress PE could advertise different GFH labels for different ingress PEs)
- S-bit: if set, source identification is embedded in the Identification field
  - Otherwise, source information from outer encap must be used together with Identification field
    - Outer MPLS label, BIER ingress BFR-ID, or Ethernet source mac address

Motivations

- Solve the EVPN-MPLS fragmentation problem w/o incurring IP overhead or requiring IP transportation
- Support fragmentation function (and possibly other functions like ESP) without IP for possible other use cases
  - In theory, if an Ethertype is assigned for GFH, this could be used to fragment Ethernet frames w/o involving IP/MPLS at all.
  - BIER encapsulation has a protocol field that can specify payload type like GFH
- The generic solution works for all layers – MPLS/BIER/Ethernet
  - Can be used for PW/VPLS as well

Next Steps

- Seeking comments
  - Presentations/discussions in BESS/MPLS/BIER/PAL/TSV WGs
- Finding a home – TSVWG/INTAREAWG/?