A Fragmentation Strategy for Generic Routing Encapsulation (GRE)

draft-bonica-intarea-gre-mtu-04

R. Bonica, C. Pignataro, J. Touch

Background

- The GRE specification does not describe procedures to address fragmentation
- Vendors have developed implementation-specific fragmentation strategies
  - Because GRE fragmentation procedures are local to the GRE ingress router, devices implementing one fragmentation strategy can interoperate with devices that implement another fragmentation strategy
- Operational experience has demonstrated the relative merits of each strategy

Goals of This Draft

- Specify a GRE tunnel fragmentation strategy
  - Describe current practice and shipping product
  - Specify requirements for implementation supporting current practice
  - Clarify applicability
- Does not UPDATE RFC 2784
  - Ensure that we don't obsolete existing products
  - Should we discuss this?

GRE Fragmentation Alternatives

1. GRE ingress router discards the packet and signals back to the payload source
   - Payload source revises its estimate of the PMTU
2. GRE ingress router fragments the payload and encapsulates it in a non-fragmentable delivery packet
3. GRE ingress router fragments the delivery packet
   - Or allow it to be fragmented downstream

Strategy 1: Discard Payload

- Pros
  - May avoid IP fragmentation altogether
  - When fragmentation is required, packet is reassembled at payload destination. Applicable regardless of GRE egress router's ability to reassemble at required rates
  - Applicable for all payload types
- Cons
  - Requires GRE ingress router to maintain a sufficiently conservative estimate of the PMTU between the GRE ingress and egress
  - Requires payload source to execute PMTUD/PLMTUD procedures
  - Requires the network to deliver ICMP PTB messages from the GRE ingress router to the payload source

Strategy 2: Fragment Payload

- Pros
  - Packet is reassembled at payload destination. Applicable regardless of egress routers ability to reassemble at required rates
- Cons
  - Requires GRE ingress router to maintain a sufficiently conservative estimate of the PMTU between the GRE ingress and egress
  - Applicable only for fragmentable payloads
    - IPv4 with DF = 0 and length > 64
  - Payload is reassembled at payload destination

Strategy 3: Fragment Delivery

- Pros
  - When delivery header is IPv4, does not require GRE ingress router to maintain a sufficiently conservative estimate of the PMTU between the GRE ingress and egress
  - Does not require payload source to execute PMTUD procedures
  - Does not require network to deliver ICMP PTB packets from GRE ingress to GRE egress
- Cons
  - Applicable only when the GRE egress router is capable of reassembling packets at required rate
  - Additional specification required to avoid DoS attack of GRE egress router

Current Behavior

- Default behavior
  - Fragment payload if possible
  - Otherwise, discard payload and signal back to source
  - Do not allow the delivery header to be fragmented
- Support fragmentation of the delivery header
  - Configuration required, not default behavior
  - Not recommended unless the GRE egress router is known to be capable of reassembling at required rates

Applicability

- When GRE is delivered over IPv6, PMTU between GRE ingress and GRE egress MUST be IPv6 MTU + GRE Overhead
- When GRE is delivered over IPv4, PMTU between GRE ingress and GRE egress MUST be IPv4 MTU + GRE overhead

Ask

- Adopt as WG draft