IP Fragmentation Considered Fragile

<draft-bonica-intarea-frag fragile-02>

Ron Bonica, Fred Baker, Geoff Huston, Bob Hinden, Ole Trøan, Fernando Gont

IETF102 Montreal

This Presentation….

- Describes how fragmentation works
  - For IPv4
  - For IPv6
- Describes how IP fragmentation reduces reliability
- Provides recommendations protocol developers and network operators

How Fragmentation Works

- IPv4 Fragmentation [RFC 791]
  - Fragmentation is always allowed at the source
  - DF-bit indicates whether fragmentation is also allowed downstream
- IPv6 Fragmentation [RFC820]
  - Fragmentation is allowed at the source only
- IPv4 and IPv6
  - Upper-layer header appears in first fragment
  - Upper-layer header does not appear in subsequent segments

Fragmentation At The Source Node Only

- Source should refrain from sending packets with length greater than PMTU
  - Packets with length greater than PMTU are dropped
- Approaches
  - Source refrains from sending packets with length greater than the minimum link MTU
  - Source maintains a running estimate of PMTU

PMTU Estimation

- PMTU Discovery (PMTUD)
  - IPv4 – RFC 1191
  - IPv6 – RFC 8201
- Packetization Layer PMTU Discovery (PLPMTUD)
  - RFC 4821 (TCP only)
  - Draft-fairhurst-tsvwg-datagram-plpmtud (other packetization layers
  - Not defined for UDP

PMTUD

- Source produces initial PMTU estimate
  - Estimate may be larger than actual PMTU
- When the source sends a packet that is larger than the actual PMTU
  - Downstream discards the packet and sends ICMP PTB to the source
    - ICMP PTB includes the MTU of the link through which packet could not be forwarded
  - Source updates PMTU estimate accordingly
- Relies on the network to deliver ICMP PTBs

PLPMTUD

- Source produces initial PMTU estimate
- Source sends probe packets of various lengths at the packetization layer
- Source receives acknowledgments at the packetization layer
- Source updates PMTU estimate accordingly
- Does not rely on ICMP Packet Too Big
  - But does rely on timeouts
  - Probe loss can invoke slow start procedures

Fragmentation Reduces Reliability

- Upper-layer header appears in first fragment only
- Impacts
  - Load balancers
  - Firewalls
  - Other middle boxes

Fragmentation Reduces Reliability (continued)

- Security Vulnerabilities
  - Overlapping Fragments
  - Resource exhaustion attacks
  - More…….
- Blackholing due to ICMP loss
  - PMTU fails due to loss of ICMP Packet Too Big messages
- Blackholing due to filtering
  - Widespread dropping of IPv6 packets with extension headers

Transport Layer Solutions

- Select MTU that is unlikely to need fragmentation
- Transport layer solutions
  - PLPMTUD for TCP
  - <draft-fairhurst-tsvwg-datagram-plpmtud> work in progress
  - <draft-ietf-tsvwg-udp-options> work in progress

Recommendations

- Application Developers
  - SHOULD NOT develop applications that rely on IP Fragmentation
- Network Operators
  - MUST NOT filter ICMPv6 Packet Too Big messages
  - SHOULD NOT deploy equipment that discards all packets that contain extension headers
- Meta Recommendation
  - DNSSEC needs a more efficient solution

Next Steps

- Adoption of this document by INTAREA WG?

QUESTIONS / COMMENTS?