Tunnels in the Internet Architecture

draft-ietf-intarea-tunnels
IETF 123 - Madrid

Joe Touch, Independent Consultant
Mark Townsley, Cisco

Some history

- Requested by the IESG, March 2007
  - First version draft-touch July 2007
  - Discussions begin…

- INTAREA adoption July 2010
  - Significant revision 2015, with figures and revised terminology
  - Another significant revision 2016, with more figures and expanded terminology

- Largely complete by 2018
  - Then a hiatus due to job shift

Why so long?

- Years to converge on "tunnels as a link"
  - Currently widely-accepted
  - Implications are less completely understood, e.g., on signaling, hopcount processing, etc.

- Years to converge on "who does what"
  - Currently widely-accepted that ingress/egress must do the heavy lifting, not downstream receivers
  - Implications similarly less widely appreciated, e.g., that ingress/egress processing and space limits are not a "free pass" to ignore DF, etc.

- Years to clarify the issues with other approaches
  - Every tunnel WILL be tunneled again
  - I.e., a tunnel that can't be used recursively isn't designed correctly

- Well, that and a change of employers…

Key contribution: terminology

- Use existing terms where possible
  - **MTU** is *link payload (RFC 1122, 1812)*
  - E**MTU**_S is *link payload* sender limit (may or may not avoid src frag, RFC 1122)
  - E**MTU**_R is *link payload receiver (*reassembly) limit (RFC1122)
  - P**MTU** (path MTU) is max(MTU), defines largest atomic packet or fragment over a path
  - **Link packet - l**ink layer message, which may be broken into one or more **frames**
  - **Atomic packet** is not (source) fragmented and not fragmentable (on path) (RFC 6864)

- Add new terms in the style of old ones
  - **MFS** = maximum frame size, the link equivalent of MTU
  - E**MFS**_S = link equivalent of EMTU_S
  - E**MFS**_R = link equivalent of EMTU_R
  - P**MFS** (path MFS) = link equivalent of PMTU
  - Tunnel **MTU** = MTU of tunnel as a link
  - Tunnel **link packet** = link packet of tunnel as a link
  - Tunnel maximum **atomic packet** = largest unfragmented/unfragmentable packet of a tunnel

- And only a few new terms that don't have a corresponding style
  - Tunnel **transit** packet = IP that transits a tunnel
  - **Inner/outer** fragmentation (as commonly used)

LEGEND:
Blue = link payload (MTU)
Red = phy layer payload (MFS)
Green = copy style (add "tunnel")
Purple = new terms

Key issues noted (Sec 5.2)

- IP-in-IPv4
  - Fragment after encapsulation, not before
  - Incorrect hopcount drop, ingress ICMP generation
  - Routing lookup that should not occur

- IP-in-IPv6
  - Incorrect MTU, hopcount decrement, fragmentation, ingress ICMP generation

- IPsec tunnel mode
  - Selects tunnel only by security rather than as a link via routing (see also RFC3884)
  - Incorrect ingress ICMP generation

- Other tunnel protocols
  - GRE, LISP, L2TP, PWE, GUE, Geneve, etc.
  - Incorrect MTU, copying of DF, ICMP generation at ingress

Rationale to continue to RFC

- Important clarifications
  - A tunnel *IS* a link
  - A tunnel MTU is the *egress reassembly MTU*, not the tunnel path MTU
  - Fragmentation happens *after* encapsulation, not before

- Useful terminology
  - RFCs
    - Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines", BCP 145, RFC 8085, March 2017
    - Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, March 2020
    - Fairhurst, G., Jones, T., Tüxen, M., Rüngeler, I., and T. Völker, "Packetization Layer Path MTU Discovery for Datagram Transports", RFC 8899, September 2020
    - Bonica, R., Baker, F., Huston, G., Hinden, R., Troan, O., and F. Gont, "IP Fragmentation Considered Fragile", BCP 230, RFC 8900, September 2020
    - Gross, J., Ed., Ganga, I., Ed., and T. Sridhar, Ed., "Geneve: Generic Network Virtualization Encapsulation", RFC 8926, November 2020
    - Sarikaya, B., von Hugo, D., and M. Boucadair, "Subscriber and Performance Policy Identifier Context Headers in the Network Service Header (NSH)", RFC 8979, February 2021,
    - Robles, M., Richardson, M., and P. Thubert, "Using RPI Option Type, Routing Header for Source Routes, and IPv6-in-IPv6 Encapsulation in the RPL Data Plane", RFC 9008, April 2021
  - IDs
    - draft-ietf-idr-bgp-ct-39
    - draft-ietf-rtgwg-net2cloud-problem-statement-42