Deep Label Stacks

MPLS part 2, IETF 84

speaker: curtis; voice: kireeti

Deep Label Stacks

- Curtis's exploration of several forwarding chips led him to think more about the problems that deep label stacks pose to some of the chips
- This is to bring the issue to the MPLS WG's attention, and figure out what (if anything) should be done

Connection with Entropy Labels

- The issue is general, and not related to processing entropy labels per se (or load balancing or …)
- However, ELs contribute to deeper stacks
  - The entropy label means one more label
  - ELI means yet another label
  - The idea of inserting multiple ELs in a label stack further increases the depth

Problem 1: deep label **operations**

- Usually, label operations work with just the top label or two
- However, sometimes many labels have to be pushed/popped/swapped
  - Reasonable to limit number of labels that can be processed => potentially expensive ASIC resources
  - As with many such limits (640K anyone?), time overruns the limit

PHP vs. UHP

- Penultimate hop popping (PHP) is a clever way to distribute deep label operations across several boxes
  - However, PHP means that the data plane of an LSP stops short of the control plane
  - Theoretical issue: “MPLS architecture is impure”
  - Practical issue: “I need stats!”
- UHP to get stats means yet more burden to deep label operations: do stats with every pop

Exceeding Limit #1

- What can a chip do if its imposed limit on label depth is exceeded?
  1. Take a performance hit (lower pps)
  2. Drop “violating” packets
  3. Wedge :-)

Problem 2: Searching for BoS

- Even if forwarding decisions require just a few labels, there are times when chips look for the label with the BoS set
- This may be to validate the label stack
- This may be for load-balancing purposes
  - Pick labels near BoS for LB purposes
  - Go beyond label stack for fields for LB

Load Balancing and BoS

- From the PoV of load balancing, labels near the BoS are better than labels near the Top of Stack
  - think pseudowire labels vs. tunnel labels
- Sometimes even this is not enough, and "good" fields for LB come from the payload
  - IP fields inside an Ethernet frame in a PW
- Fat PW labels and entropy labels obviate the need for this, but these are recent advances

Exceeding Limit #2

- Again, what does a chip do if it doesn't find a BoS label in its "window of search"?
  - Discard packet as "malformed"
  - Take a performance hit to search further

Addressing the Problems

- Change chip design to focus on these
  - Need guidelines on what to watch out for
- Change deployment
  - Need “traps” to realize there’s an issue, and knobs to tweak behavior
  - Limit stats on a given LSP
- Change network architecture
  - E.g., add an extra hop to avoid multiple UHPs

Okay, So What?

- Is this a fairly complete list of problems related to depth of stack and BoS?
- Survey: who has which problem?
- Should the IETF care? MPLS WG? If so, how:
  - BCP? For what:
  - Chip implementation guidelines?
  - Deployment guidelines?
  - Network architecture guidelines?

Next Step

- WG chairs + ADs need to tell us *whether* and *how* to proceed, starting with how much of this in the IETF scope and in the WG scope
- Doing a survey might be helpful to evaluate the depth, urm, extent of the problem
  - Logistics may be tricky (NDA, etc.)