# Layer 3 Configuration and Attribute Discovery for LSVR (and more)

Randy Bush <randy@psg.com>

2019.04.29 IEEE 802.1 TSN TG Weekly Call

## This is about learning Layer 3 and above IP Attributes and Configuration

IIJ is Building a Second Medium Scale Data Center (MSDC) in Shiroi/Chiba Capacity of 6k Racks

## 10,000 rack MSDCs are the new black

# OSPF OK to 500 Nodes IS-IS good to 1,000

# Limited Because They Repeatedly Flood Everything

### IGPs Extremely Noisy

A lot of folk trying to clean that problem up

# BGP Scales Because It Signals Only Changes

# So BGP has become common in MSDCs

## BGP Does Not Repeat (Link) State

But ... How Does BGP in a DC Learn Layer 3 Attributes and Configuration?

# Layer 3 Configuration and Attribute Discovery

What do we really need?

#### Must Haves

- · Discover L3 End-points and Links
- Discover L3 Encapsulations/Addresses:
  - IPv4, IPv6, MPLS4/6
- Maintain Session Liveness
- Exchange config data to start BGP etc.
- Security

#### Discover L3 Nodes and Links

- · Learn Identity of peer node on a link
  - MAC address is the simple case
  - But multiple VLANs run between those two MAC addresses
  - Many real devices use the same MAC address on multiple ports and/or sub-interfaces
- It seems that SystemID and ifIndex solve the general case

#### Discover L3 Encapsulations

Learn peer's supported IP Layer 3
 AFI/SAFIs, V4, V6, MPLS, etc

- Announce my IP Layer 3 Encapsulations
- Learn and announce IPv4, IPv6 addressing, MPLS Labels, ...

### Shhhhhh, Quiet!



McGill University Library, St. Gallen, CH

#### OK, a More Modest Library



A Clos customer who beats us up if there is one extra DHCP request from boot to running

# When you really scale out, you must have a Quiet network

# Therefore TCP-like Sessions

#### Sessions

- A transition from Discovery to State
- So data are not retransmitted, Quiet
- Like BGP, only signal changes
- Ensure that the neighbor/peer has a known same state

#### Maintain Liveness

- Ensure Session Liveness
- Must have tunable timers
- Some ops want fast failure detection
- Some folk want zero noise and may use BFD at Layer 3

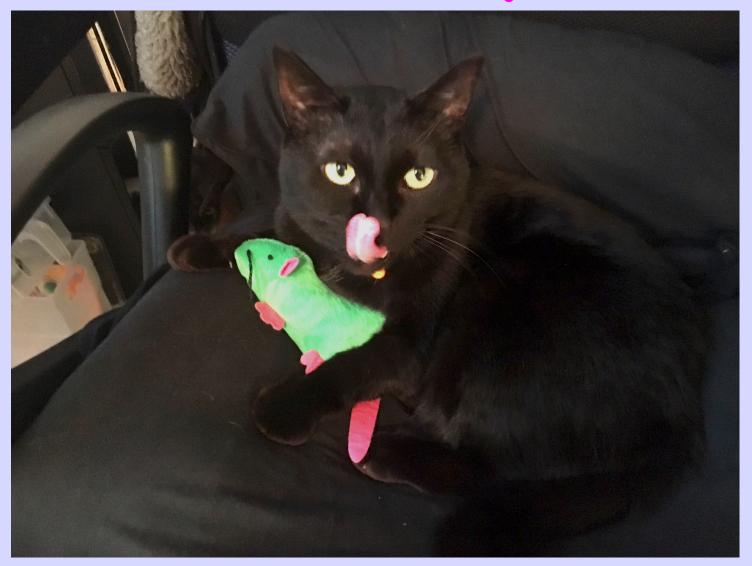
#### [ Aside ]

- So we are looking at L3DP not needing to repeat HELLOs
- The L3DP Session KEEPALIVE rate is tunable, even down to zero
- ACK & retransmits all tunable

## Exchange config data to start BGP etc.

- This is all about Layer 3 and BGP-like protocols
- · Low-touch configuration very desirable
- Discovery protocol also exchanges the BGP configuration attributes to enable BGP OPEN to succeed
- draft-ymbk-lsvr-l3dl-ulpc-00.txt

### Security



### Security

It's hard to motivate and get input from ops. Is it lack of concern or secret sauce?

#### Trust on First Use, TOFU

## Are you the Droid I was Talking to Earlier?

Trivial, Easy to Use, but Weak

Considering ways to bind a Public Key to the Identity Could be X.509 or other syntax

### Transport

- Clean, simple, unreliable datagrams into big
   PDUs'; do not want to hear size restrictions
- We know how to do the reassembly, see RFC791
- We know to ACK, time out, retransmit, back off, etc.
- Though we do not expect congestion when a link has yet to be characterized, we do know how to deal with it should it occur

And, of course, everything is over Ethernet these years; though other L1/2 transport might be used

### Keep it Simple

We want to produce easily understood, implementable, and securable standards, not build resumes.