DS-Lite Failure Detection and Failover

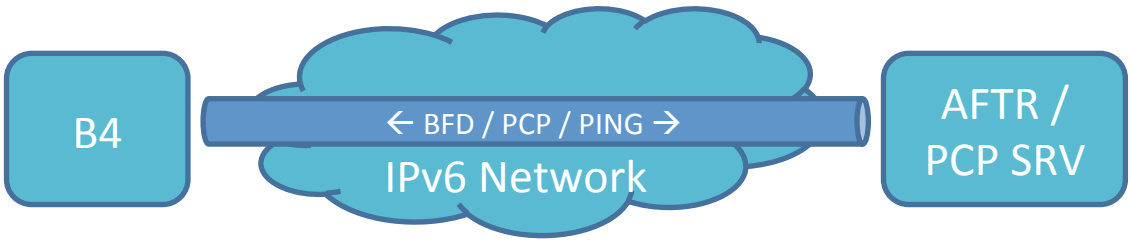
draft-tsou-softwire-bfd-ds-lite-05

Tina Tsou
Brandon Li
Cathy Zhou
Jürgen Schönwälder
Reinaldo Penno
Mohamed Boucadair

Changes from draft -04

- Comments from Ian Farrer
- Add NAT failure detection approach by BFD
- Add state synchronization/session re-establishment
- Comparison of different failure detection solutions

Problem to Solve



- There is no failure detection and failover mechanism for stateless tunnel and stateful CGN function, e.g., tunnel up or down, NAT44 functioning or not, which brings difficulties for operations and maintenance.
- Protocols to resolve this problem: BFD, PCP, ICMP, …

Failover – Anycast

- Upon detecting a failure, the B4 element terminates the current DS-Lite tunnel and creates another tunnel to one of the other addresses.
- AFTR may use anycast address for receiving packets
- There might be issues with failures on the path if for example an ICMP error message fails to get delivered correctly (e.g., it is sent to a different anycast server).

Failover – VRRP

- For active/passive HA in NAT gateways, it's quite common to have a single virtual address offered by VRRP (or a proprietary equivalent) that the upstream routers will use as their next hop.
- In the event that the master CGN fails, the standby takes over the virtual L3 address.
- If a VRRP based virtual address is used as the tunnel endpoint, then the clients wouldn't need to be aware of the failover.

BFD for Tunnel Failure Detection

- BFD auto configuration
  - In DS-Lite, B4 has the AFTR address, which is sufficient to initiate a BFD session
  - Other parameters can be negotiated via signaling or static config, no manual configuration
- BFD packet rate
  - Long time period between BFD packets transmission, e.g., 10s or 30s

BFD for NAT failure detection

- B4 creates PCP mapping.
- BFD at AFTR uses an external public interface (or another external mapping) to send a BFD packet to the public PCP mapping created by B4.
- B4 will reply to the AFTR external interface.

PCP for DS-Lite Failure Detection

- If PCP is available in a DS-Lite deployment, PCP can be used for keep alive testing, and to trigger failover if a failure is detected.
- PCP is used to create a mapping with short lifetime, updates are sent periodically.
- If the PCP client detects a failure, e.g., a NETWORK_FAILURE error code is returned, or there is no response from the PCP server, the client will switch to another PCP server or AFTR

ICMP for DS-Lite Failure Detection

- ICMP pings can be sent periodically, or triggered manually when necessary
- Since ICMP is an integral part of any IP stack, not extra implementation efforts are required on the B4 elements.

Comparison of Three Solutions

	Availablility	Packet format	Additional functionality ontop of keepalives	Configuration /provisioning overheads
BFD	Widely used/ network side, less used/ terminal side	Simple fixed	Bidirectional status synchronization	Similar
PCP	Less than BFD/ICMP	Vari-able	No bidirectional detection	
ICMP	Ubiquitous		Network/CGN initiated detection	

State Synchronization/Session Re-establishment

- BFD link for both active AFTR and backup AFTR should be set up in the initial state.
- In the hot-standby case, the master AFTR and the backup AFTR will synchronize and backup the session. No need to re-establish the TCP session.
- In the cold-standby case, TCP session will need to be re-established by the client if there is a failure.

Next steps

- Adopt by the working group?