IETF-61
L2TPEXT Workgroup

Washington, 09-Nov-04

Graceful Restart Mechanism for L2TPv3

draft-galtzur-l2tpext-l2tpv3-gr-02.txt

Sharon Galtzur

Axerra networks

What This Is About

- **Terminology:**
  - **L2TPv3 Control State:**
    - Control connection state (including its reliable delivery mechanisms)
    - Association between a control connection and sessions
  - **L2TPv3 Forwarding State (per session):**
    - Association between the session and its local attachment circuit
    - Source and Destination IP addresses (from the control connection)
    - Session State (Idle, Established etc.)
    - Session ID (local and remote)
    - Cookies (local and remote)
    - Sequence numbers (next number to send, next expected number to receive

What This Is About (2)

- **Use Case:**
  - **L2TPv3 Control State is lost, but L2TPv3 Forwarding State is preserved**
- **Normal Situation:**
  - **Forwarding state is reset due to loss of control state**
    - **The traffic stops**
  - **Traffic will only be resumed after the control connection and all the sessions have been re-established**
- **Objective of the Graceful Restart**
  - **Restart the control plane without affecting the forwarding state**
    - Ideally, the customer traffic should not be disturbed at all
  - **Follow the trend set in RFC 3478 for LDP**

How This Can Be Achieved

- **Behavior is split between the two peers:**
  - **LCCE that has lost its control state ("restarting LCCE")**
  - **The peer of the restarting LCCE**
- **The peer of the restarting LCCE detects loss of control connection**
  - **Postpones breaking the associated sessions:**
    - The timeout is agreed upon by the restarting LCCE and its peer during establishment of the (now lost) control connection
    - All the sessions are marked as "stale"
  - **Once the control connection is restarted, allows to re-establish "stale" sessions in the new control connection without affecting the forwarding state**
    - No need to restore the broken control connection itself!

How This Can Be Achieved (2)

- **Restarting LCCE:**
  - **Collects the forwarding state of all the sessions**
    - **If it did not survive the restart, there is nothing to preserve**
  - **Establishes a new control connection to the peer(s) and indicates that it has preserved the forwarding state**
- **Re-establishment of stale sessions can be initiated by any of the two peers**
  - **Mainly, follow the the normal procedure for session establishment with the following modifications:**
    - **Stale sessions transit to Established upon completion without any changes in the Forwarding plane but become associated with the new control connection**
    - **Stale sessions that have not been re-established within the agreed upon timeout will be broken**

How This Can Be Achieved (3)

- **Protocol changes:**
  - **Two new AVPs**
    - Graceful Restart AVP [SCCRQ, SCCRP]
    - Graceful Restart Session AVP [ICRQ, OCRQ]
  - **One new Session state ("stale")**
  - **Only minimal changes to the control connection and session state machines**
  - **No new security issues**
    - **The same security methods as in the "native" L2TPv3**
    - **The forwarding state will not undergo any change**
- **Partial Graceful Restart Support**
  - **It is possible to participate in the graceful restart of a peer even if local forwarding state cannot be preserved across the local control plane restart**

Optimized For

- Separation between Control and Forwarding plane elements in the restarting LCCE
- Zero down time for the customer traffic
- L2TPv3 running directly over IP
  - No need to preserve the Tunnel/Control Connection Ids or UDP ports
  - The typical case for PW applications
- Minimal intervention in the protocol operation
- Hub-and-Spoke topologies:
  - Low-end Spokes support high-end Hub graceful restart
- Minimization of security risks
  - Graceful restart uses the same set of security procedures as the normal operation

Not Optimized For

- The overall Control Plane restart time
  - Not so important as long as the customer traffic is not affected
- The Control Plane CPU consumption
  - Not really important if the Control and Forwarding plane are separated
  - Not important in the Hub-and-spoke topologies
    - High-end Hub benefits from separation
    - Low-end Spokes do not maintain so many sessions each
- The Control Plane traffic bandwidth consumption
  - The amount of traffic is the same for graceful restart as for the LCCE power-up
- L2TPv2 and L2TPv3 over UDP
  - Simply will not work

Questions?