LACP Enhancements

Enhancements needed for LACP

Rev. 1

Norman Finn

nfinn@cisco.com

LACP Enhancements

- This contribution is available at: new-nfinn-LACP-improvements-1010-v01.pdf.
- The purpose of this contribution is to describe a work program to enhance LACP with three new features:
  - Symmetrical frame distribution by VLAN.
  - Improved Flush.
  - Graceful name change.

Symmetrical frame distribution by VLAN

Symmetrical distribution by VLAN

- Many operators of Provider Bridged or Provider Backbone Bridged networks would like the ability to distribute frames over the physical links of an Aggregation by VLAN. Many vendors support this ability.
- However, it is also desired by many operators that both aggregating systems make the same decision, placing a given VLAN on the same physical link in both directions.
- An automatic means of accomplishing this is desirable.

Symmetrical distribution by VLAN

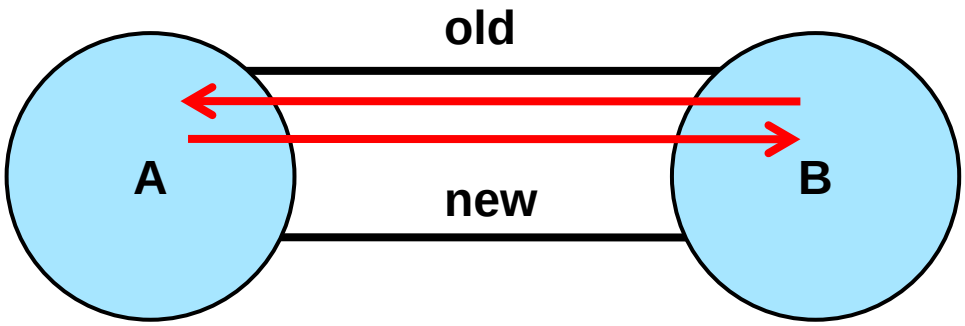
- It is not unreasonable to go through a negotiation phase when an Aggregation is first brought up to decide the VLAN-to-link assignment.
- But, there are 4K services on a S-tagged interface, and 16M services on an I-tagged interface. One cannot, therefore, express all possible preferences for VLAN assignment in one LACPDU.
- It is undesirable, after a link failure or recovery event, to go through a renegotiation for VLAN-to-link assignment.

Symmetrical distribution by VLAN

- My suggestion would be to pre-configure a VLAN-to-link assignment for every possible combination of more physical links present.
- The LACPDUs can carry only a configuration name and checksum, as is done for the MSTP configuration.
- It is also possible to carry two names/checksums, so that the configuration can be changed.

Improved Flush

Improved Flush



- To ensure against out-of-order frame delivery, 802.1AX says that, if bridge B wants to switch a stream from an old physical link to an new physical link:
  1. B stops transmitting the stream on the old link.
  2. B sends a Marker PDU on the old link and starts a timer.
  3. A returns the Marker as a Marker Reply PDU.
  4. B either receives the Marker Reply PDU or times out.
  5. B starts sending the stream on the new PDU.

Improved Flush

- The problem with this is that the time taken by the Marker reply to get from A to B is wasted – A could be transmitting frames on the new Link during that time.
- The time taken for the first frame on the new link to get through B's output queues and across the wire from A to B is also wasted – that frame could have been sent earlier.
- But, that is all you can do if neither system knows anything about the manner in which the other system distributes the frames.

Improved Flush

- But, if A knows that frames are distributed by VLAN, the B could do the following:
  1. Both systems know what the VLAN distribution is. Neither system accepts frames on the “wrong” physical link.
  2. B sends a “Stop receiving frames for VLAN (or VLAN Group) X” message to A on the old link and sends no more frames for those VLANs to the old link.
  3. B immediately sends a “Start receiving frames for VLAN (or VLAN Group) X” message to A on the new link and starts sending all frames for those VLANs on the new link.
  4. Either when A receives the Stop or when it receives the Start, it switches over which link it expects to receive the frames from, discarding any received on the old link.

Improved Flush

- In the ideal case, where the Stop message arrives before the Start message, no frames are dropped.
- In the not infrequent case where the Stop message arrives after the Start message, the minimum number of frames are dropped, and no frames are delivered out-of-order.
- Simply repeating the Start messages once (or twice) takes care of the worst case, where the Start message is lost. (Losing the Stop message is OK.)
- Note again that this can work only if the receiving system knows what the sending system's distribution function is, so that it can discard frames from the old link.

Graceful name change

Graceful name change

- There are occasions when a system would like to change its LACP system name without dropping and re-establishing an active aggregation.

  Bridge brain failures resulting in a hot-standby brain taking over.

  Configuration changes.

  The “split brain” scenario.

- By transmitting an “Old Actor_System_Priority” and “Old Actor_System” TLV, the name of a system can be changed without bringing down an aggregation.