Issues with LLDP over LAGs

Paul Congdon
Norm Finn
Jeff Lynch

September 19, 2011

IEEE 802.1 Contribution

Maintenance Item – 0009

Disambiguating LLDP over Link Aggregations

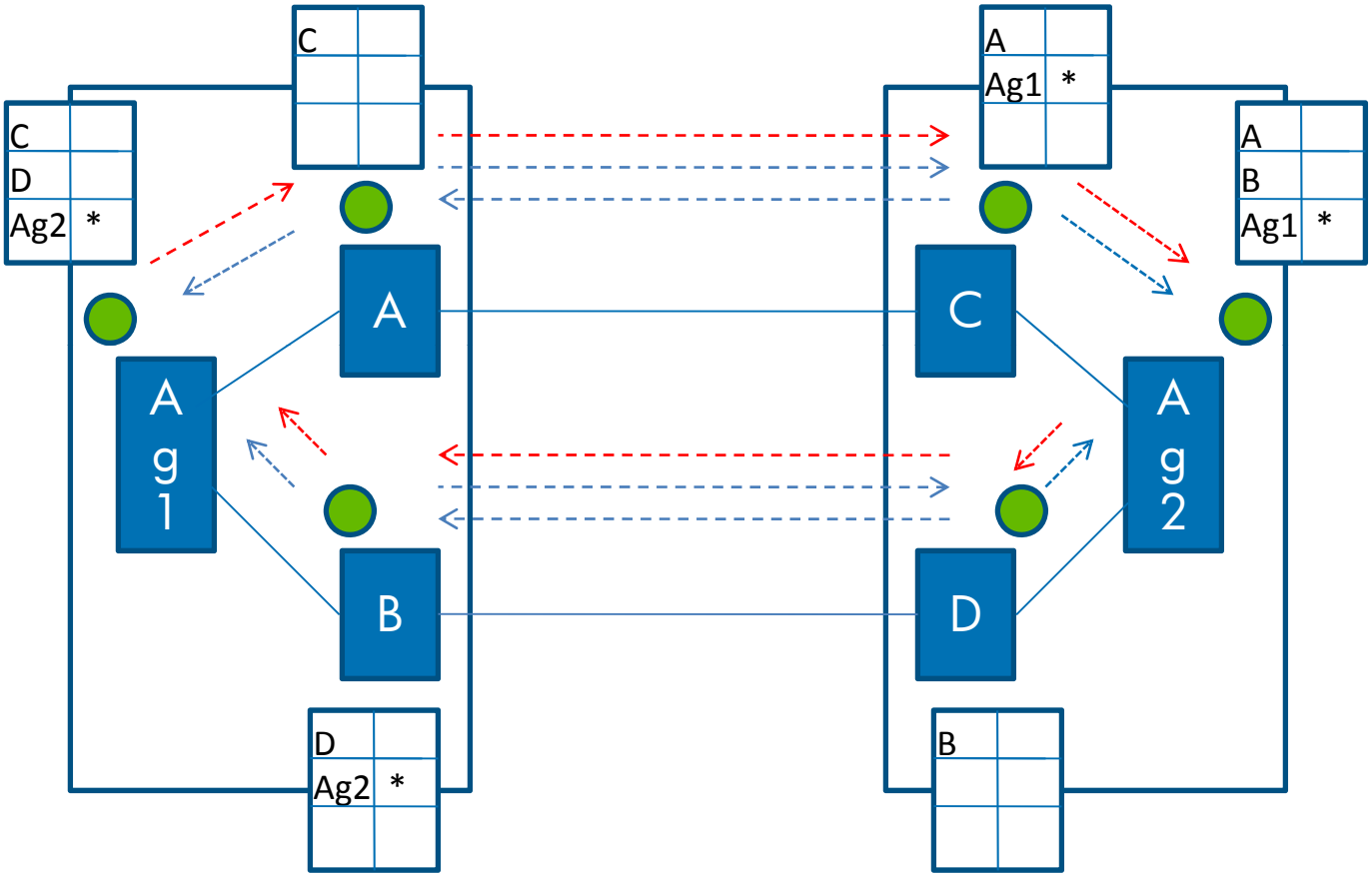
- Submission: Jeffrey Lynch – September 2011
- Issues:
  - It is unclear how LLDP should operate over an aggregation
  - It is currently not possible to determine at the receiver if the LLDP frames were sent from a peer at the physical link or at the aggregate
- Proposed Resolution:
  - Document in IEEE 802.1Q or IEEE 802.1AX the requirements and operational behavior for LLDP in the present of Link Aggregation
    - Either redefine the use of some of the reserve bits or add new fields to the existing LAG TLV (to disambiguate the source & target of LLDP frames)
  - Identify which TLVs should be sent on each individual link or on the aggregate
  - Resolve the architectural addressing issues introduced by running LLDP over a LAG.
- Discussion
  - There is some question as to whether you can actually send/receive LLDP frames at the physical layer because of the way it has been specified originally in 802.3. If that is true, then we have a new feature requirement for 802.1AX and it is needed to send/receive at the physical layer. This aspect could be put into AXbq.
  - How should a Y work in the link aggregation layer? If we want to de-multiplex frames then we would either need new addresses or content specific multiplexing. The way it currently works is that you would see multiple peers at the aggregate layer and one of the physical links would see two peers while others would only see one. Doing the Y based on protocol is a slippery slope (e.g. it would be protocol specific – the LLDP Y as a shim). We would rather not create a new destination address and instead use a TPMR type Y.
  - Discuss this at the Interim via a submission by Jeff, Paul and Norm. Leave in received state for now.

Discussion

Discussion

Discussion

Example LLDP over LinkAgg



Note: Aggregation agent LLDP frames are distributed over the aggregation to one link

- Aggregation agent sees all LLDP agents
- Physical link agent sees physical link peer and possibly the aggregation agent

Alternatives Discussed
(to disambiguate the source & target of LLDP frames)

- Require LinkAgg TLV to be present if running in a Link Aggregation
  - Define the use of some of the reserve bits or add new fields to the existing LAG TLV
- Create a new mandatory TLV that must only be sent by the aggregate agent
- Define a new address to direct frames only to the aggregate (e.g. Nearest-Aggregate Address)
- Create a protocol Y
  - How should a Y work in the link aggregation layer?
  - If we want to de-multiplex frames then we would either need new addresses or content specific multiplexing. The way it currently works is that you would see multiple peers at the aggregate layer and one of the physical links would see two peers while others would only see one. Doing the Y based on protocol is a slippery slope (e.g. it would be protocol specific – the LLDP Y as a shim). We would rather not create a new destination address and instead use a TPMR type Y.

Require LinkAgg TLV

(Define the use of some of the reserve bits to the existing LAG TLV)

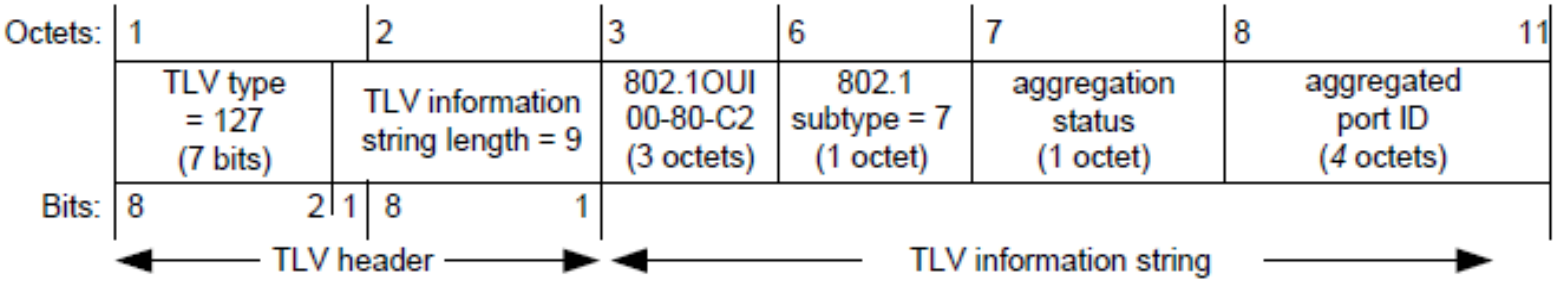


Figure E.7—Link Aggregation TLV format

Table E.3 – Link aggregation capability/status

BIT	FUNCTION	Value/meaning
0	Aggregation capability	0 = not capable of being aggregated 1 = capable of being aggregated
1	Aggregation status	0 = not currently in an aggregation 1 = currently in aggregation
2	Aggregation component	0 = Aggregation port 1 = Aggregation group
3-7	Reserved for future standardization	----

LLDP Agent is on individual port

Disambiguating Agents (cont)

- Agents on physical links
  - Aggregation capability = 1
  - Aggregation status = 1
  - Aggregation component = 0
  - Aggregated PortID = PortID of aggregated link
  - PortID TLV = physical link PortID

- Agents on aggregates
  - Aggregation capability = 1
  - Aggregation status = 1
  - Aggregation component = 1
  - Aggregated PortID = PortID of aggregated link
  - PortID TLV = aggregated link PortID

Issues

- Aggregation portID can not be the same as any physical portID for this to work.
  - Chassis-ID+Port-ID must be unique for each agent to be identified. If the aggregate uses the same Chassis-ID+Port-ID the receiver will wipe-out previously received data.
- This proposal makes it impossible to aggregate an aggregation.

BACKUP

BACKUP

BACKUP

Background

- In the November 2009 IEEE 802.1 meeting, the issue was raised concerning how LLDP is to operate over an aggregated link
  - Is there one LLDP session per physical link?
  - Is there one LLDP session for the aggregated link?
  - Are there both?
- Particular clarification is needed in the context of DCBX

LLDP & Link Aggregation

Consensus of January, 2010 telcon (documented in new-pelissier-laglldp-0110)

- LLDP must be able to run on individual links.
- LLDP may run over the aggregate.
- Necessary to determine topology with link aggregation through TPMRs
  - Norm to provide some material on this (Contribution not yet submitted)
- We need a way to disambiguate the aggregate
  - Paul will have a contribution on this
  - Contribution submitted: new-congdon-linkag-LLDP-0110.pdf "Disambiguating LLDP agents over a Link Aggregation"
- For DCBX:
  - For each TLV, we need to figure out whether to run it on individual links or aggregate.
  - Also need to determine convergence in case of a conflict
- May need to look at other TLVs, e.g. VLAN TLV
  - Does it makes sense to send it on the individual link or the aggregate.
- Manoj brought up the issue of end station problems with link aggregation
  - Not LLDP specific just general problems such as stateful offloads and link aggregation
  - Separate problem from the one under discussion here.

Text in blue added to represent current status

IEEE 802.1 Revision Request

```
+-------------------------------------------------------------------+
| IEEE 802.1 REVISION REQUEST XXXX                                  |
+-----------------=====================================-------------+
DATE:    September 6, 2011
NAME:    Jeffrey Lynch
COMPANY/AFFILIATION:  IBM
E-MAIL:  jjlynch@us.ibm.com

REQUESTED REVISION:
     STANDARD:  IEEE 802.1Q &/or IEEE 802.1AX
                (possibly included into the 802.1AXbq (Distributed LAG) specification/project)
     CLAUSE NUMBER: Appropriate clauses for LinkLAG MIB and TLV
                    definitions such as P802.1Q-REV Annex D "D.2.7 Link Aggregation TLV"
     CLAUSE TITLE:  Appropriate clauses for LinkLAG MIB and TLV
                    definitions such as P802.1Q-REV Annex D "D.2.7 Link Aggregation TLV"

RATIONALE FOR REVISION:
    The current IEEE 802.1 standards are silent on how LLDP is to operate over an aggregated link.  This could lead to
          incompatible implementations in the marketplace.
        •     Is there one LLDP session per physical link or one for the aggregated link?  Or are there both?
        •     Background
              •     new-pelissier-laglldp-0110.pdf “Notes from LLDP over LAG concall”
              •     new-congdon-linkag-LLDP-0110.pdf “Disambiguating LLDP agents over a Link Aggregation”
              •     new-lynch-LLDP-over-LAG-0911.pdf “Eliminating the ambiguity of running LLDP over LAGs”

PROPOSED REVISION TEXT:  (See background presentations above)
1) Document in IEEE 802.1Q or IEEE 802.1AX the requirements and operational behavior for LLDP in the present of Link
       Aggregation
       Either redefine the use of some of the reserve bits or add new fields to the existing LAG TLV to disambiguate the
            source & target of LLDP frames
2) Identify which TLVs should be sent on each individual link or on the aggregate
3) Resolve the architectural addressing issues introduced by running LLDP over a LAG.

IMPACT ON EXISTING NETWORKS:

    None known at this time.  See background presentations (above) for potential backwards compatibility implications.
+-------------------------------------------------------------------------------------------------+
| Please attach supporting material, if any                                                       |
| Submit to:- Tony Jeffree, Chair IEEE 802.1  and copy:  Paul Congdon, Vice-Chair IEEE 802.1      |
| E-Mail: stds-802-1-maint-req@ieee.org                                                           |
|           +------- For official 802.1 use ----------+                                         |
|           | REV REQ NUMBER:                          |                                         |
|           | DATE RECEIVED:                           |                                         |
|           | EDITORIAL/TECHNICAL                      |                                         |
|           | ACCEPTED/DENIED                          |                                         |
|           | BALLOT REQ'D YES/NO                      |                                         |
|           | Status: X                                |                                         |
|           +------------------------------------------+                                        |
+-------------------------------------------------------------------------------------------------+
```