FC Link Aggregation (FC-LAG)
Control plane aspects – Part 1

Harsha Bharadwaj
Apr 2020

[T11-2020-00073-v001.pdf]
Supported FC-LAG Topologies for “now”

Case: 1

- a: FC Device
- b: FC Device
- c: FC Device
- d: FC N_Port Aggregator (NPV/AG)

- HBA
- FC Switch (or) N_Port Agg

Case: 2

- FC Switch (or) N_Port Agg
- HBA
- FC Device

NOTES:
1) Scope: FC-LAGs spanning one switch and one device only. All other cases in T11-2020-00035-v000.pdf out-of-scope.
2) Case 2 is different since HBA1, HBA2 may need coordination across different software instances to manage single FC-LAG.
3) Case 1d aggregation might be achieved in a proprietary way already. FC-LAG applicable only for interoperability (if feasible from data path perspective).
4) FC-LAG on Switch-to-Switch ISLs is not a focus area now.
FC–LAG Configuration

- FC–LAG members **statically** (explicitly) configured at N_Ports and F_Ports
  - Final LAG formation on “least common denominator” negotiation via protocol
- FC–LAG formation **initiated** from N_Port during FLOGI
  - Applicable in FLOGI and LS_ACC (Class 2,3); Default: 0 (not configured)

<table>
<thead>
<tr>
<th>N_port LAG Bit (FLOGI)</th>
<th>F_port LAG Bit (FLOGI ACC)</th>
<th>Can form LAG?</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Yes, if compatible</td>
</tr>
</tbody>
</table>

**Table 187 – Common Service Parameter applicability (Continued)**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Set to 00h</td>
<td>2</td>
<td>31-24</td>
<td>0</td>
</tr>
</tbody>
</table>
FC-LAG Identifiers

- Locally unique LAG-IDs (numbers) on Switch/Device for multiple LAG config
- Fabric unique FCID for each LAG for frame routing
- Globally unique ID (WWN) to represent each F_Prot FC-LAG
  - Could be assigned from the Switch WWN name space
- Globally unique ID (WWN) to represent each N_Prot FC-LAG for Zoning, Aliasing, etc. in the fabric
  - Not many good options at this point
    - One of the member WWPN ⇒ Member may leave the FC-LAG, HBA may be replaced
    - Device WWNN ⇒ Cannot work when more than one FC-LAG from a device
    - A unique pool of WWNs for assignment to FC-LAGs ⇒ Best suited; Feasibility?
    - Ties into T11-2020-00017-v000.pdf (Names for Enclosures)
Need for a FC-LAG negotiation protocol

- Detect various port misconfig and exclude them from FC-LAG
  - Excluded ports “may” be allowed to operate as individual links

- Synchronized bring up and add/delete of FC-LAG members wrt traffic

- Excluded from LAG

Excluded from LAG

Because (F3-F4) came up first

Both ends ready for traffic in LAG mode

Both ends ready for traffic in LAG mode
Option 1: FC-LAG negotiated as part of EVFP protocol
   - Existing VF negotiation sequence in FC-LS Sec 8.4 runs just after FLOGI
     - The VF bit in first FLOGI ACC does not allocate FCID, but triggers EVFP
     - EVFP negotiates VF via a Sync/Commit procedure with N_Port using FCID FFFFF0h
     - A subsequent FLOGI per VF assigns FCIDs
     - Future VF add/delete also managed via EVFP
   - FC-LAG requires an identical operational semantic
     - Reuse EVFP with new TLVs to negotiate FC-LAG (LAG also “virtualization” scheme)
     - One integrated FLOGI after EVFP when VF + FC-LAG is configured

Option 2: FC-LAG negotiated separate from EVFP protocol
   - New ELS Exchange LAG Parameters (ELAGP) before 1st FLOGI after EVFP
   - ELAGP protocol semantics defined like EVFP (Sync/Commit)
   - Coordination with VF negotiation and FLOGI can become complex

Option 1 preferred due to its simplicity and reusability
FC-LAG first Port Bring up with VF

FC-LAG Config verify

FLOGI Commit (LAG)

LS_ACC (DID= LAG_FCID1, VF_ID1)

LS_ACC (DID= LAG_FCID2, VF_ID2)

FLOGI (LAG-WWN, VF_ID1)

FLOGI (LAG-WWN, VF_ID2)

Note: If VFBit=0 above, EVFP skips VFT phase and only does FC-LAG after the 1st FLOGI
FC-LAG second or subsequent Port Bring up with VF

1st FLOGI (VFBit=1, LAGBit=1, WWPN of port)

LS_ACC (VF Bit=1, LAGBit=1, D_ID = 0)

EVFP_SYNC - Tagging Status, Port VF_ID, Allowed_VFID

LS_ACC (SYNC) - Tagging Status, Port VF_ID, Allowed_VFID

EVFP_COMMIT – VFT_Tagged

LS_ACC (COMMIT) – VFT_Tagged

EVFP_SYNC (LAG) – Local Cfg info

LS_ACC (SYNC: LAG) - Local Cfg Info, Peer Cfg Info

EVFP_COMMIT (LAG)

LS_ACC (COMMIT: LAG)

F-Port (FLOGI WKA FFFFFE)

EVFP_ELS Responder

Tagging Disable

Tagging Enable Rx and Tx

Tagging Enable Rx

Tagging Enable Rx and Tx

FC-LAG Config verify

Member added to FC-LAG

Note: FC-LAG Logical FLOGIs NOT needed for 2nd and subsequent ports bring up
EVFP Descriptors for FC-LAG

- EVFP already has (Sync, Commit) descriptors for negotiating VF
  - Tagging Admin Status, Port VF_ID, Locally enabled VFID_List

- FC-LAG (Sync, Commit) with a new set of descriptors
  - Sync Phase: Determine the “lowest common denominator” of FC-LAG config
  - Commit Phase: Coordinate FC-LAG data path binding before traffic can start
  - Unlike VF, Commit initiated from N_Port to allow for Renegotiate/Abort the LAG if not acceptable to N_Port after Commit ACC.

- **Descriptors**: Local Info (Node WWN, WWPN, LAG-ID), Peer Info (Node WWN, WWPN, LAG-ID) - *Details in future meetings!*

- EVFP initiated based on VF and FC-LAG Config:

<table>
<thead>
<tr>
<th>EVFP</th>
<th>VF only</th>
<th>FC-LAG only</th>
<th>VF and FC-LAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Single stage</td>
<td>Single stage</td>
<td>Two stage</td>
</tr>
<tr>
<td>FLOGI</td>
<td>Single stage</td>
<td>Two stage</td>
<td>Two stage</td>
</tr>
</tbody>
</table>
Two stage FLOGI when LAG Bit=1

• **Physical FLOGI** per link **before** EVFP
  - Done for **every member** that comes up in FC-LAG
  - Fabric does **NOT assign FCID** when LAG Bit = 1
  - Link Config Params like BB_Credit, BB_SCN, Receive_Data_FieldSize applied

• **Logical FLOGI** once for whole FC-LAG **after** EVFP
  - Done for **only 1st member** port that comes up in the LAG
  - FCID **assigned** for the FC-LAG by fabric
  - Based on VF and FC-LAG config

<table>
<thead>
<tr>
<th>Physical FLOGI per:</th>
<th>No VF, No FC-LAG</th>
<th>VF only</th>
<th>FC-LAG only</th>
<th>VF and FC-LAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>N_Port</td>
<td>N_Port</td>
<td>N_Port</td>
<td>N_Port</td>
<td>N_Port</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Logical FLOGI per:</th>
<th>No VF, No FC-LAG</th>
<th>VF only</th>
<th>FC-LAG only</th>
<th>VF and FC-LAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>-NA-</td>
<td>(VFID, N_Port)</td>
<td>FC-LAG</td>
<td>(VFID, FC-LAG)</td>
<td></td>
</tr>
</tbody>
</table>
FC-LAG member Compatibility check

• Compatibility check of parameters among local FC-LAG members
  • **Capability** (Eg: BB_SCN), **Config** (Eg: Port Type) and **Oper** (Eg: Port Oper Speed)
  • Capability and Config should be **identical** among all members
  • Callback in EVFP-LAG protocol for vendor specific Oper compatibility checks Eg:
    1. Switch Vendor-A: All LAG F_ports of same speed for equal split of data path traffic
    2. Storage Vendor-B: All LAG N_ports from same IO Module/SP
  • On a LAG, config changes allowed only on LAG and not on individual members

• Compatibility across peer
  • Same number of members at either end
  • All members compatible at link layer
FC-LAG Status Change Notification

- “Port Changed” RSCN sent for the FC-LAG when FC-LAG comes up or entire FC-LAG goes down
  - No RSCNs sent for the FC-LAG individual member status change
- N_Ports talking to FC-LAG may need to know changes in LAG bandwidth
  - New ELS from the Domain Controller hosting the FC-LAG on member up/down (or) RSCN with new Event Qualifier?
    - ELS_Code = 0xnnnn
    - SID = 0xFFFFCxx; DID = (for every registered device in Zone)
    - Payload = FC-LAG FCID/WWN, New LAG Oper Speed, Reason Code
- Interested N_Ports register with Fabric via new ELS (or) New Reg Function in SCR?
- New SW_ILS for InterSwitch notification distribution (or) New Aff Port in SW_RSCN?
Management view of FC-LAG

- Fabric Config Server hosts all Port Objects in the fabric for topology discovery

- **New FC-LAG Port Object**
  - Port Name: WWPN of the LAG
  - **New Port Types:** LAG-N_port (0x04) LAG-F_port(0x87)
  - Port Speed Capability/Oper: Aggregate Speed
  - Attached Port Name: LAG_Fport-WWPN (or) LAG_Nport-WWPN
  - Port Tx_type, Module_Type, Physical Port Number: **NA** (its logical)

- **New LAG Port Object Queries:**
  - Get Member List Port Objects of LAG. [Response: Member Port Object Names]
  - Get LAG of a WWPN [Response: LAG Port Object Name]

- Each FC-LAG Member will still have a Port Object
  - **New** Port Type=LAG-Member-N_port/F_port (or) Existing N_port/F_port
Future meetings..

- Deep dive into the EVFP-LAG protocol with frame/TLV formats
- Impact analysis on all N_port ↔ F_port protocols (Eg: RDP)
- Data plane guidelines (I/O binding to FC-LAG member, IOD etc)
  - But most of data plane aspects are vendor specific!

- Long Pole: Auto Creation of FC-LAG without explicit config (based on intent)!