DCBX TLV Summary

Joe Pelissier

az-pelissier-dcbxtlvs-0309

Parameter Acceptance

- **DCBX provides a common framework that allows a port to configure specific operational parameters of itself to match that of its peer port on the far end of the link**
  - **DCBX limits this capability to specific parameters**
  - **The framework limits this transfer to one direction; i.e. from a "non-Willing" port to a "Willing" port.**
    - For a given parameter, a device may be Willing or non-Willing, but not both
    - Necessary to prevent loops in the exchange
    - In some cases, DCBX prohibits certain devices from being Willing
- **Accomplished through the use of the Willing (W) bit**
  - **Within a TLV, setting W to zero indicates that the device is not-Willing**
  - **If W is set to 1, the device is Willing**
    - Additionally, if the device is not emitting the TLV, it may operate as Willing unless otherwise specified

Parameter Acceptance Rules

- A port operating as Willing may accept the parameters from a port operating as non-Willing

  i.e. A port with W set to 1 in a given DCBX TLV may accept the corresponding parameters from a received TLV with W set to 0

  Additionally, a port that is not emitting a given TLV may accept the corresponding parameters from a received TLV with W set to 0

- If port operating as non-Willing on a given TLV *shall not* configure itself to the parameters received in a corresponding TLV
- If a port is operating as Willing and it receives the corresponding TLV with W set to 1, it *shall not* configure itself to the parameters received in that TL V

  Neither of the above two bullets imply that the port is prohibited from having matching parameters. What is prohibited is the automatic configuration of those parameters based on the received TLVs

Traffic Classes Supported TLV

TLV Type =127	TLV Info String Len=5	802.1 OUI 00-80-C2	802.1 Subtype = 8	Reserved	Traffic Classes Supported
7 bits	9 bits	3 octets	1 octet	5 bits	3 bits

- Generally provided as information
- E.g. may be used to:

  Optimize buffer allocation (especially in a NIC)
- No negotiation – represents physical limitation of device

  No concept of willing

ETS Configuration TLV

TLV Type =127	TLV Info String Len=16	802.1 OUI 00-80-C2	802.1 Subtype = 9	Priority Assignment Table	Priority Group Configured Bandwidth Table
7 bits	9 bits	3 octets	1 octet	4 Octets	8 Octets

- Generally provided as information
- E.g. may be used to:
  - Generate a recommendation for PG Bandwidths
- No negotiation – represents current configuration
  - No concept of willing
- Priority Assignment table has one 4-bit entry per Priority
  - 0-7 indicate the Priority is mapped to the corresponding PG,
  - 13,14 indicate mapped to one of the AVB groups
  - 15 indicates priority operates as specified in 802.1Q (strict priority)
- Priority Group Configured Bandwidth Table has one 8-bit entry per PG (always contains 8 entries).
  - Indicates the current BW configured for each Priority Group
  - Total must equal 100 (implies valid range is 0-100 for each entry)

ETS Recommendation TLV

TLV Type =127	TLV Info String Len=13	802.1 OUI 00-80-C2	802.1 Subtype = 10	Reserved	RV	Recommended Priority Group Bandwidth Table
7 bits	9 bits	3 octets	1 octet	7 bits	1 bit	8 Octets

- **Provides a recommendation of how far end should configure the PG bandwidths**
- **May be done unidirectional or bi-directional**
  - **No need for willing bit to prevent the looping configuration problem**
- **RV (Recommendation Valid), indicates that the TLV contains a recommendation. Set to zero if no recommendation is being provided at this time (although one may be provided in the future, e.g. after receipt of other TLVs)**
- **Priority Group Configured Bandwidth Table has one 8-bit entry per PG (always contains 8 entries) (reserved if RV set to 0).**
  - **Indicates the current BW configured for each Priority Group**
  - **Total must equal 100 (implies valid range is 0-100 for each entry)**
    - If total does not equal 100, TLV is malformed and shall be ignored
  - **If a device is "willing", it may update its configured BWs in accordance with this TLV**
  - **OK for both ends to be "willing"**

Priority-based flow Control TLV

TLV Type =127	TLV Info String Len=6	802.1 OUI 00-80-C2	802.1 Subtype = 11	W	Re-served	PFC Cap	PFC Enable
7 bits	9 bits	3 octets	1 octet	1 bit	4 bits	3 bits	1 Octets

- Provides negotiation and information of PFC enabled / disabled per priority
- PFC Cap indicates the device's limitation of how many priorities may simultaneously support PFC (not negotiated).
- Utilizes Parameter Acceptance Framework
- PFC enable has 8 bits (one per priority)

  A one indicates PFC is enabled on the priority

  A zero indicates that PFC is disabled on the priority

  Local policy in each end of the link decides whether to use the priority if the configuration does not match

Application Priority TLV

TLV Type =127	TLV Info String Len=variable	802.1 OUI 00-80-C2	802.1 Subtype = 12	W	Re-served	Application Priority Table
7 bits	9 bits	3 octets	1 octet	1 bit	7 bits	Multiple of 3 octets

- Provides indication of priority should be used for each protocol
- Utilizes Parameter Acceptance Framework

  A bridge shall not operate as an Acceptor
- Application Priority Table contains zero or more entries:

  3 bits: Priority

  2 bits: Reserved

  3 bits: Sel (0: Ethertype, 1 Port# over TCP, 2 Port# over UDP, 3 Port# over TCP or UDP, 4 Port# over neither UDP nor TCP, 5-7 reserved for future use, devices shall ignore entries with these values)

  16 bits: Protocol ID (based on Sel field)

Congestion Notification TLV

TLV Type =127	TLV Info String Len=7	802.1 OUI 00-80-C2	802.1 Subtype = 13	W	Re-served	CN Mask	RDY Mask
7 bits	9 bits	3 octets	1 octet	1 bit	7 bits	1 octet	1 octet

- Provides synchronization and negotiation of Congestion Notification
- Utilizes the Parameter Acceptance Framework to provide an indication on which Priorities CN should be enabled

  Applies to the CN Mask, *not* the RDY mask
- CN Mask indicates if CN is enabled (1) / disabled (0) for corresponding priority
- RDY Mask indicates if defenses are enabled (0) or disabled (1) for a given priority. Reserved if corresponding CN Mask bit is 0
- See the state machine in CN for more info

Some thoughts on the state machines

- **IMHO:**
  - **These TLVs operating over LLDP provide the functionality necessary to support our PARs**
  - **The proposed Protocol and Feature state machines are therefore unnecessary**
    - And they are complex beyond what is justifiable to support the needs of our PARs
- **There are some that are disappointed that a general framework was not provided; however, again IMHO**
  - **No need for this framework was presented in the context of our PARs**
  - **No proposal that could be justified within the context of our PARs has been brought forward**
- **Extensive work has been done on this proposal (i.e. TLVs that operated over LLDP)**
- **We need to make forward progress**
- **Therefore, I hope we can reach consensus that these TLVs operating over LLDP will be the basis for our work moving forward**

Thank You!