



DCBX Framework

Joe Pelissier

az-pelissier-dcbx-framework-0409v4

DCBX Framework Overview

- **The DCBX Framework provides three capabilities utilizing LLDP:**

Informational Passing: the passing of information that is used in DCB applications or configuration

e.g., passing the number of traffic classes supported by a bridge to an attached end station may enable the end station to optimize its buffer resources

Symmetric Parameter Passing: the passing of a parameter from one port to its peer port with objective of both ports utilizing the same parameter

e.g., Priority Flow Control configuration

The “to” and “from” ports are administratively configured

e.g., passing configuration to and end station from a bridge

In addition, defense mechanisms are supported to ensure that mismatched configurations do not impact the fabric

Optional, required, or non-existent on a feature-by-feature basis

Asymmetric Parameter Passing: the passing of a parameter from one port to its peer port, or between both peer ports. In this case, the resulting configuration of each port may or may not match its peer

e.g., ETS bandwidth configuration

The directions in which the parameters are passed is administratively configured

- **The Framework provides standard state machines for each of the parameter passing methods**

Asymmetric Parameter Passing – Overview

- **Two TLVs:**

Configuration TLV: Provides current operational state and Willing bit

Willing bit is not used by state machine

Provided to higher layers to provide an indication of the expected behavior of the remote port

Recommendation TLV: Provides recommendation for the operational state of remote port

Transmitted only if the local port is configured to make recommendations (in which case it is transmitted in all LLDP PDUs)

Transmitted regardless of the “willingness” of the remote port

Recommendation Valid (RV) bit provides a mechanism for the local port to delay making recommendations until it receives additional information about the remote port (e.g. number of traffic classes supported and priority to priority group assignments).

Asymmetric Parameter Passing State Variables

LocalWilling: Indicates that the local port has been administratively configured to accept recommendations. This value is included in the Willing bit of Configuration TLVs transmitted by this port.

OperParam: The current operational state of the parameter on the local port. This value is included as the parameter in the Configuration TLVs.

LocalAdminParam: The administratively configured state for the parameter. This becomes the operational state of the parameter by default, and may be overridden if the local port accepts a recommendation from the remote port.

RemoteParam: The parameter received in the last Configuration TLV. This variable is set to NULL if the last LLDP PDU received did not contain the Configuration TLV or if no LLDP PDUs have been received.

RV: TRUE indicates that the RV (Recommendation Valid) bit was set in the last Recommendation TLV received. FALSE indicates that the RV bit was not set in the last Recommendation TLV received. NULL indicates either the TLV was not present in the last LLDP PDU or that no LLDP PDUs have been received.

Asymmetric Parameter Passing – Example TLVs

TLV Type =127	TLV Info String Len=17	802.1 OUI 00-80-C2	802.1 Subtype = 9	Reserved	Willing	Priority Assignment Table	Priority Group Configured Bandwidth Table
7 bits	9 bits	3 octets	1 octet	7 bits	1 bit	4 Octets	8 Octets

Configuration 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

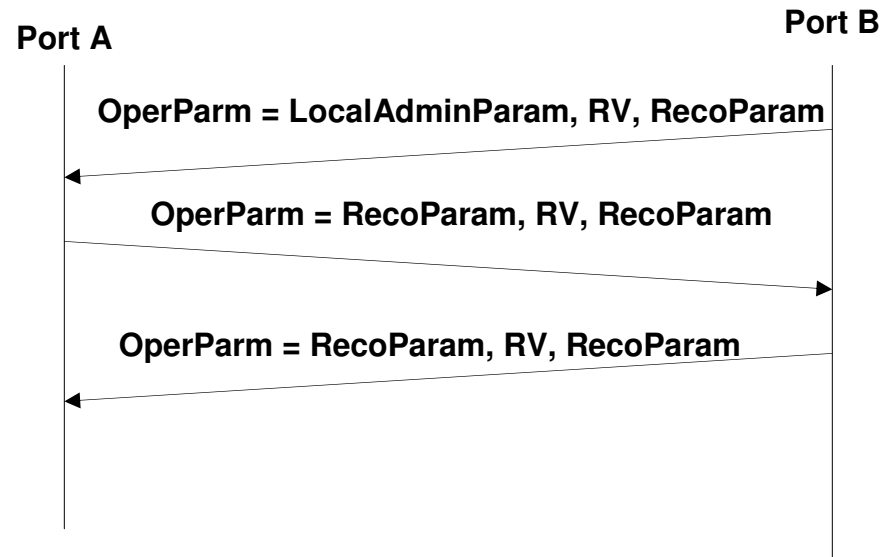
Recommendation TLV

Asymmetric Parameter Passing State Machine



Note: RV can take on three values: TRUE, FALSE, and NULL. NULL indicates that the Recommendation TLV was not part of the last LLDP PDU received, or that no LLDP PDUs have been received.

Asymmetric Parameter Passing Ladder Diagram



Two transfers in one direction and one transfer in the other direction required to transfer the parameters

Symmetric Parameter Passing – Overview

- **Single TLV used**

 - Always carries the current local operational state**

 - Carries a “Willing” (“W”) bit**

 - A port that sets the W is considered “Willing”**

 - A Willing port shall set its operational parameter to that indicated in the received TLV if the received TLV has the W bit set to zero**

Symmetric Parameter Passing State Variables

LocalWilling: Indicates that the local port has been administratively configured to accept the parameter from the remote port. This value is included in the Willing bit of DCBX TLVs transmitted by this port.

RemoteWilling: TRUE indicates that the Willing bit was set in the last TLV received. FALSE indicates that the Willing bit was not set in the last TLV received. NULL indicates either the TLV was not present in the last LLDP PDU or that no LLDP PDUs have been received.

OperParam: The current operational state of the parameter on the local port. This value is included as the parameter in the DCBX TLV.

RemoteParam: Contains the value of the last parameter (i.e. the operational state of the remote port) received in the TLV. NULL indicates either the TLV was not present in the last LLDP PDU or that no LLDP PDUs have been received.

LocalAdminParam: The administratively configured state for the parameter. This becomes the operational state of the parameter by default, and may be overridden if the local port accepts the parameter from the remote port.

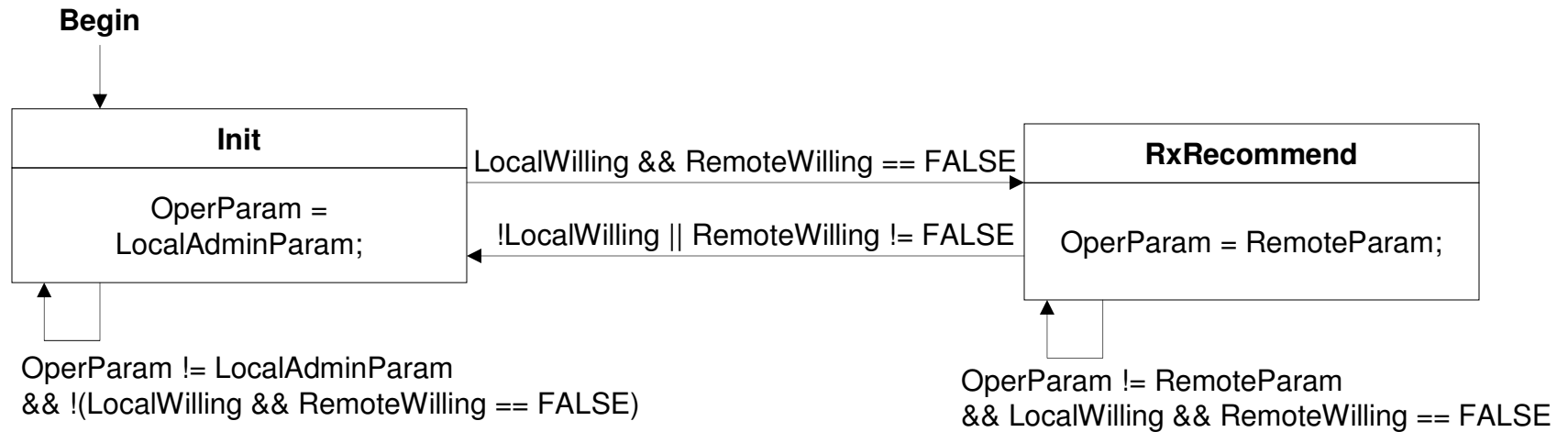
Note – Through observation of the states and state variables it is possible to determine that the state machine is in the process of passing parameters. This knowledge may be useful for clients of DCBX. E.g. a client may wish to delay use of the link while the DCBX state machine is in the process of passing and possibly setting parameters. A “Pending” indication indicating that the state machine is in this process may be created by the following equations:

```
Pending = RemoteParam == NULL  
        || !LocalWilling && RemoteWilling == TRUE && OperParam != RemoteParam;
```

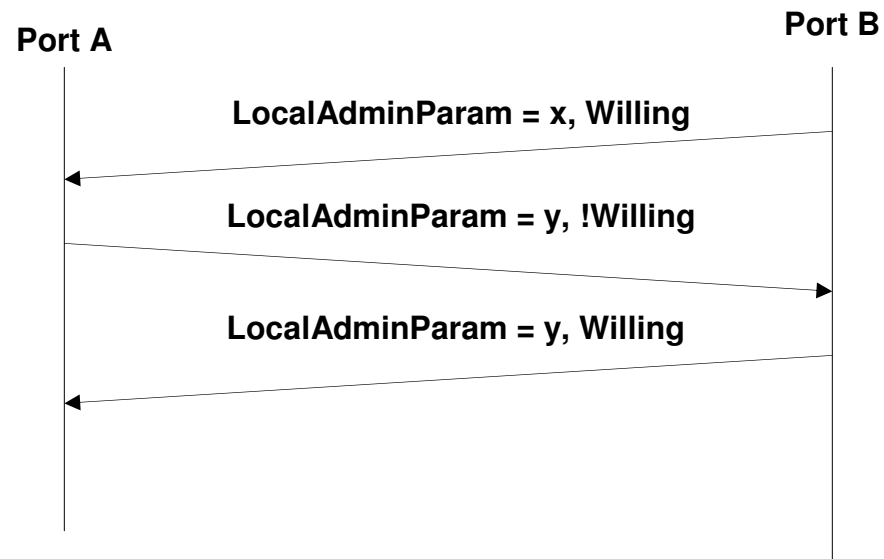
Symmetric Parameter Passing – Example TLV

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

Symmetric Parameter Passing State Machine



Symmetric Parameter Passing Ladder Diagram



Symmetric Parameter Passing – Defense Option - Overview

- **May apply to symmetric parameters that may be turned “on” or “off”**
For example, Congestion Notification, Priority-based Flow Control
Does not apply to all such parameters
Reference the individual feature descriptions to determine if applicable
- **Defense mechanism ensures that received frames do not harm the network if the remote end does not have the feature enabled**
For example, remap frames to different priorities, discard the frame, etc.
The specific mechanism is defined on a feature-by-feature basis
- **If implemented, a DefOn bit is provided in the TLV for each feature that may be individually enabled / disabled**
For example, CN provides one bit for each priority
- **A common state machine is provided for enabling / disabling the defenses**
- **A port may elect to delay transmission of frames until it detects that the remote port has defenses turned off (i.e. DefOn == 0).**

Symmetric Parameter Passing – Defense Option – State Variables and Functions

DefenseEnabled: TRUE indicates that the local port has been administratively configured to enforce defenses. FALSE indicates that the local port has not been administratively configured to enforce defenses. For some features, defense is not optional. In these cases, DefenseEnabled is always TRUE.

OperParam: The current operational state of the parameter on the local port. This value is included as the parameter in the DCBX TLV. TRUE indicates that the feature is enabled. FALSE indicates that it is not.

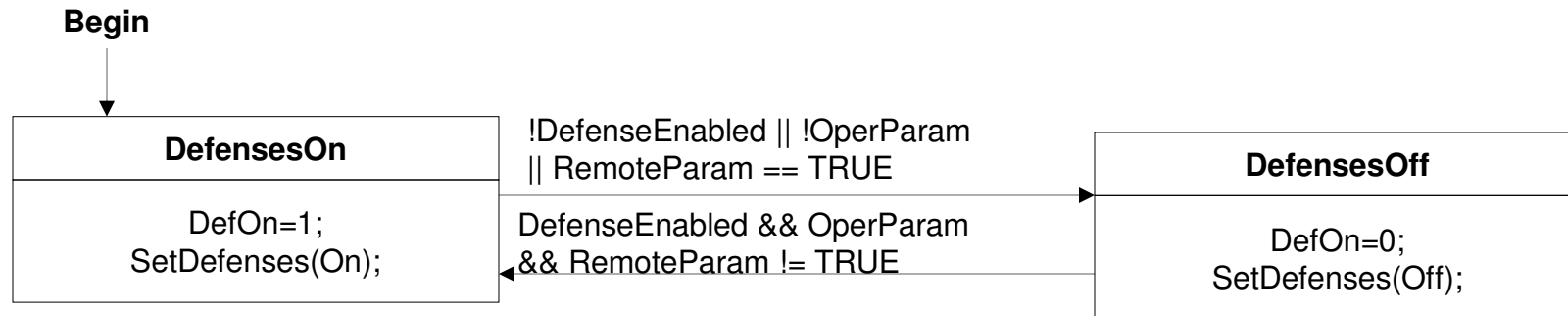
RemoteParam: Contains the value of the last parameter (i.e. the operational state of the remote port) received in the TLV. TRUE indicates that the feature is enabled. FALSE indicates that it is not. NULL indicates either the TLV was not present in the last LLDP PDU or that no LLDP PDUs have been received.

SetDefenses(): SetDefenses(On) enables defenses for this port. SetDefenses(Off) disables defenses for this port.

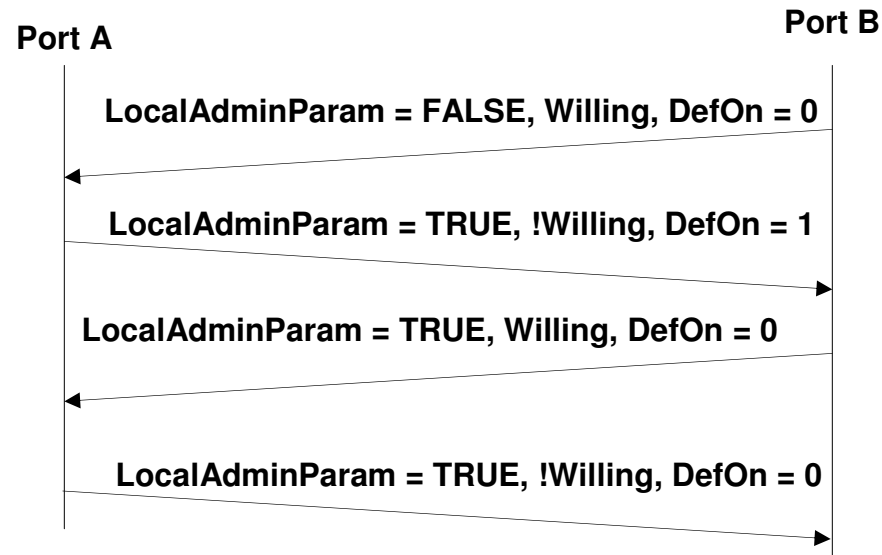
Symmetric Parameter Passing – Defense Option - Example TLV

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

Symmetric Parameter Passing – Defense Option - State Machine



Symmetric Parameter Passing – Defense Option - Ladder Diagram



TLV Summary

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)**
- **Informational Passing**

ETS Configuration TLV

TLV Type =127	TLV Info String Len=17	802.1 OUI 00-80-C2	802.1 Subtype = 9	Reserved	Willing	Priority Assignment Table	Priority Group Configured Bandwidth Table
7 bits	9 bits	3 octets	1 octet	7 bits	1 bit	4 Octets	8 Octets

- **Generally provided as information**
- **E.g. may be used to:**
 - Generate a recommendation for PG Bandwidths
- **Operates as the Configuration TLV with Asymmetric Parameter Passing**
 - The Bandwidth table may be configured using Asymmetric Parameter Passing
- **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
- This is the recommendation TLV used in Asymmetric Parameter Passing
- 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	DefOn
7 bits	9 bits	3 octets	1 octet	1 bit	4 bits	3 bits	1 Octet	1 Octet

- Provides negotiation and information of PFC enabled / disabled per priority
- PFC Cap indicates the device's limitation of how many traffic classes may simultaneously support PFC (not negotiated).
 - Zero indicates no limitation (i.e. PFC may be enabled on all eight priorities)
- Utilizes Symmetric Parameter Passing with *optional* defense mode
- 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
- DefOn has 8 bits (one per priority)
 - A one indicates that defenses are on for that priority on the local port
 - A zero indicates that defenses are off for that priority on the local port
 - Reserved if corresponding PFC enable bit is zero
- When PFC Defenses are enabled for a given priority, the port shall:
 - Discard all received frames on that priority except for Bridge and LLDP PDUs

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
- Uses Symmetric Parameter Passing *without* the defense option
- 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	DefOn Mask
7 bits	9 bits	3 octets	1 octet	1 bit	7 bits	1 octet	1 octet

- Provides synchronization and negotiation of Congestion Notification
- Uses the Symmetric Parameter Passing *with* defense mode *required*
 - i.e. the DefenseEnabled parameter in the Defense Option State Machine is always set TRUE
- CN Mask indicates if CN is enabled (1) / disabled (0) for corresponding priority
- DefOn Mask indicates if defenses are enabled (1) or disabled (0) for a given priority. Reserved if corresponding CN Mask bit is 0

Thank You!