## -1|1.1|1. CISCO

## Convergence of 802.1Q, PFC, and ETS

## Agenda

- Where we are today
- Some issues with ETS
- Goals
- Simplifications
- A proposed enhancement
- Examples
- Observations


## Where we are today...

- 802.1Q Specifies a Priority to Traffic Class Table

All .1Q switches support 8 priorities
Table maps the priorities into supported traffic classes
Between 1 and 8

- This is highly desirable

Ensures defined interoperability between bridges of different abilities

Bridges do not need to know the number of traffic classes supported by their neighbors

Each bridge exercises its best effort to support the priority scheme

## ETS provides a similar abstraction...

- Priority to Priority Group Table

Each Priority or a set of Priorities mapped to a Priority Group
Neighbor switches are unaware of the mapping

- However:

This is not exactly the abstraction we want
Priority Groups define the bandwidth allocation, therefore this is generally an externally defined behavior

We would like each switch to exercise its best effort to allocate bandwidth as defined by the Priority Groups

While providing opaqueness regarding the number of supported traffic classes

## The Priority Group table is problematic...

- The Priority Group table has dependencies on the programming of Traffic Class table

Scheduling based on Traffic Class
Assume Priorities 1 \& 2 are assigned to TC1
Assume Priority 3 assigned to TC2
Assume Priority 1 assigned to PG1 and 2\&3 assigned to PG2
->Behavior is undefined

- It would be much cleaner if it were not possible to program the tables in such a way that undefined behavior results


## PFC also provides an external abstraction...

- PFC is enabled or disabled per priority

All PFC capable bridges will support 8 priorities and the ability to enable / disable per priority

- Works for switches with less than 8 traffic classes

On Transmit: if one priority is flow controlled, then in general all traffic from the corresponding traffic class halts
On Receive: if a traffic class is becoming full, then all priorities assigned to that traffic class for which PFC is enabled may assert flow control

Or flow control may be asserted based on some local policy

- Again, switch exercises best effort and no drop due to congestion is assured


## Goals:

- Provide a consistent external abstraction for Priority, ETS, PFC, and AVB

Bridges exercise best effort based on number of traffic classes supported, in a defined manner

- Define a deterministic manner to map the external abstraction to the available Traffic Classes
- Eliminate the ability to program a bridge in a conflicting manner


## A new term: Traffic Type

- There are five different Traffic Types that must be considered:

AVB: see 802.1Qav
EP: ETS with PFC enabled
En: ETS without PFC enabled
nP: No ETS with PFC enabled nn: everything else (i.e. non-ETS and non-PFC)

## A few proposed simplifications:

- Disallow nP (i.e. PFC without ETS)

Allowing PFC on a non-bandwidth managed basis seems dangerous

- Require all Priorities within a Priority Group to all be either PFC or not PFC (no mixing of PFC and non-PFC within a Priority Group)
- On an ETS enabled switch, the Priority to Traffic Class mapping is implied via the Priority Group mapping table

The Priority to TC table becomes read only
The mapping of Priority to TC is performed as described in the following slides

- A DCB switch must support at least five Traffic Classes if it supports AVB, otherwise it must support at least three Traffic Classes


## Step 1: Define the Priority to PG table

- Contains eight entries, one for each Priority

All DCB switches must support 8 Priorities and 8 PGs; however, these may be merged into fewer traffic classes as specified later

- Each entry contains one value: the PG to which the Priority is assigned

Each Priority must be assigned to exactly one PG
A PG may have multiple Priorities assigned to it

- Default values: Priority Group = Priority


## Step 2: Define the PG Table

- Contains eight entries, one for each Priority Group

All DCB switches must support 8 Priorities and 8 Priority Groups; however, these may be merged into fewer traffic classes as specified later

- Each entry contains three values:

Traffic Type (AVB, EP, En, nn, or unused)
Bandwidth Allocation (percentage) (reserved for AVB, nn, and unused)
Traffic Class (read only, traffic class to which this PG is assigned based on the algorithm defined in this slide set)

A PG is assigned to exactly one TC
A TC may have multiple PGs assigned to it, all containing the same Traffic Type

## Step 3: Determine the number of TCs for each <br> TT

- It is desirable for each Priority Group to be assigned to a separate Traffic Class, if possible

If there are not enough Traffic Classes, then multiple Priority Groups are merged into Traffic Classes

Similar to the merging of Priorities into Traffic Classes today
All merged Priority Groups within a given Traffic Class must have the same Traffic Type

- Based on the following rules:

A separate Traffic Class is assigned for each AVB Priority Group (up to two)
A Traffic Class is allocated for each Priority Group that contains a Traffic Type that does not appear in any other Priority Group
Considering only the remaining Priority Groups, Traffic Types, and Traffic Classes:
Allocate remaining Traffic Classes in proportion to the number of Priority Groups set to
each Traffic Type
Each set of Priority Groups containing a given Traffic Type is allocated at least one Traffic Class

If exact proportions are not possible (due to integer rounding), then favor Priority Groups containing Traffic Types in this order: EP, En, nn

- These seem confusing, but they become fairly self-evident when one observes the actual possible combinations


## Priority Group to Traffic Class Allocation

| Number of PGs with <br> Traffic Types <br> EP,En,nn <br> respectively | TCs to allocate to Each TT (EP, EN, nn) given Traffic Classes Supported |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 3 | 4 | 5 | - | 7 |
| 000 | 000 | 000 | 000 | 000 | 000 |
| 001 | 001 | 001 | 001 | 001 | 001 |
| 002 | 002 | 002 | 002 | 002 | 002 |
| 003 | 003 | 003 | 003 | 003 | 003 |
| 004 | 003 | 004 | 004 | 004 | 004 |
| 005 | 003 | 004 | 005 | 005 | 005 |
| 006 | 003 | 004 | 005 | 006 | 00 |
| 007 | 003 | 004 | 005 | 006 | 007 |
| 008 | 003 | 004 | 005 | 006 | 007 |
| 010 | 010 | 010 | 010 | 010 | 010 |
| 011 | 011 | 011 | 011 | 011 | 011 |
| 012 | 012 | 012 | 012 | 012 | 012 |
| 013 | 012 | 013 | 013 | 013 | 013 |
| 014 | 012 | 013 | 014 | 014 | 014 |
| 015 | 012 | 013 | 014 | 015 | 015 |
| 016 | 012 | 013 | 014 | 015 | 016 |
| 017 | 012 | 013 | 014 | 015 | 016 |
| 020 | 020 | 020 | 020 | 020 | 020 |
| 021 | 021 | 021 | 021 | 021 | 021 |
| 022 | 021 | 022 | 022 | 022 | 022 |
| 023 | 012 | 022 | 023 | 023 | 023 |
| 024 | 012 | 013 | 023 | 024 | 024 |
| 025 | 012 | 013 | 014 | 024 | 025 |
| 026 | 012 | 013 | 014 | 024 | 025 |
| 030 | 030 | 030 | 030 | 030 | 030 |
| 031 | 021 | 031 | 031 | 031 | 031 |
| 032 | 021 | 022 | 032 | 032 | 032 |
| 033 | 021 | 022 | 032 | 033 | 033 |
| 034 | 012 | 022 | 023 | 033 | 034 |
| 035 | 012 | 022 | 023 | 024 | 034 |
| 040 | 030 | 040 | 040 | 040 | 040 |
| 041 | 021 | 031 | 041 | 041 | 041 |
| 042 | 021 | 031 | 032 | 042 | 042 |
| 043 | 021 | 022 | 032 | 033 | 043 |
| 044 | 021 | 022 | 032 | 033 | 043 |
| 050 | 030 | 040 | 050 | 050 | 050 |
| 051 | 021 | 031 | 041 | 051 | 051 |
| 052 | 021 | 031 | 041 | 042 | 052 |
| 053 | 021 | 022 | 032 | 042 | 043 |
| 060 | 030 | 040 | 050 | 060 | 060 |
| 061 | 021 | 031 | 041 | 051 | 061 |
| 062 | 021 | 031 | 041 | 042 | 052 |
| 070 | 030 | 040 | 050 | 060 | 070 |
| 071 | 021 | 031 | 041 | 051 | 061 |
| 080 | 030 | 040 | 050 | 060 | 070 |
| 100 | 100 | 100 | 100 | 100 | 100 |
| 101 | 101 | 101 | 101 | 101 | 101 |
| 102 | 102 | 102 | 102 | 102 | 102 |
| 103 | 102 | 103 | 103 | 103 | 103 |
| 104 | 102 | 103 | 104 | 104 | 104 |
| 105 | 102 | 103 | 104 | 105 | 105 |
| 106 | 102 | 103 | 104 | 105 | 106 |
| 107 | 102 | 103 | 104 | 105 | 106 |
| 110 | 110 | 110 | 110 | 110 | 110 |
| 111 | 111 | 111 | 111 | 111 | 111 |


| Number of PGs with Traffic Types EP,En,nn | TCs to allocate to Each TT (EP, EN, nn) given Traffic Classes Supported |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| respectively | 3 | 4 | 5 | 6 | 7 |
| 112 | 111 | 112 | 112 | 112 | 112 |
| 113 | 111 | 112 | 113 | 113 | 113 |
| 114 | 111 | 112 | 113 | 114 | 114 |
| 115 | 111 | 112 | 113 | 114 | 115 |
| 116 | 111 | 112 | 113 | 114 | 115 |
| 120 | 120 | 120 | 120 | 120 | 120 |
| 121 | 111 | 121 | 121 | 121 | 121 |
| 122 | 111 | 121 | 122 | 122 | 122 |
| 123 | 111 | 112 | 122 | 123 | 123 |
| 124 | 111 | 112 | 113 | 123 | 124 |
| 125 | 111 | 112 | 113 | 114 | 124 |
| 130 | 120 | 130 | 130 | 130 | 130 |
| 131 | 111 | 121 | 131 | 131 | 131 |
| 132 | 111 | 121 | 122 | 132 | 132 |
| 133 | 111 | 121 | 122 | 132 | 133 |
| 134 | 111 | 112 | 122 | 123 | 133 |
| 140 | 120 | 130 | 140 | 140 | 140 |
| 141 | 111 | 121 | 131 | 141 | 141 |
| 142 | 111 | 121 | 131 | 132 | 142 |
| 143 | 111 | 121 | 122 | 132 | 133 |
| 150 | 120 | 130 | 140 | 150 | 150 |
| 151 | 111 | 121 | 131 | 141 | 151 |
| 152 | 111 | 121 | 131 | 141 | 142 |
| 160 | 120 | 130 | 140 | 150 | 160 |
| 161 | 111 | 121 | 131 | 141 | 151 |
| 170 | 120 | 130 | 140 | 150 | 160 |
| 200 | 200 | 200 | 200 | 200 | 200 |
| 201 | 201 | 201 | 201 | 201 | 201 |
| 202 | 201 | 202 | 202 | 202 | 202 |
| 203 | 102 | 202 | 203 | 203 | 203 |
| 204 | 102 | 103 | 203 | 204 | 204 |
| 205 | 102 | 103 | 104 | 204 | 205 |
| 206 | 102 | 103 | 104 | 204 | 205 |
| 210 | 210 | 210 | 210 | 210 | 210 |
| 211 | 111 | 211 | 211 | 211 | 211 |
| 212 | 111 | 211 | 212 | 212 | 212 |
| 213 | 111 | 112 | 212 | 213 | 213 |
| 214 | 111 | 112 | 113 | 213 | 214 |
| 215 | 111 | 112 | 113 | 114 | 214 |
| 220 | 210 | 220 | 220 | 220 | 220 |
| 221 | 111 | 211 | 221 | 221 | 221 |
| 222 | 111 | 211 | 221 | 222 | 222 |
| 223 | 111 | 112 | 212 | 222 | 223 |
| 224 | 111 | 112 | 212 | 222 | 223 |
| 230 | 120 | 220 | 230 | 230 | 230 |
| 231 | 111 | 121 | 221 | 231 | 231 |
| 232 | 111 | 121 | 221 | 231 | 232 |
| 233 | 111 | 121 | 122 | 222 | 232 |
| 240 | 120 | 130 | 230 | 240 | 240 |
| 241 | 111 | 121 | 131 | 231 | 241 |
| 242 | 120 | 121 | 221 | 231 | 241 |
| 250 | 120 | 130 | 140 | 240 | 250 |
| 251 | 111 | 121 | 131 | 141 | 241 |
| 260 | 120 | 130 | 140 | 240 | 250 |
| 300 | 300 | 300 | 300 | 300 | 300 |


| Number of PGs with <br> Traffic Types EP,En,nn | TCs to allocate to Each TT (EP, EN, nn) given Traffic Classes Supported |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| respectively | 3 | 4 | 5 |  | 7 |
| 301 | 201 | 301 | 301 | 301 | 301 |
| 302 | 201 | 202 | 302 | 302 | 302 |
| 303 | 201 | 202 | 302 | 303 | 303 |
| 304 | 102 | 202 | 203 | 303 | 304 |
| 305 | 102 | 202 | 203 | 204 | 304 |
| 310 | 210 | 310 | 310 | 310 | 310 |
| 311 | 111 | 211 | 311 | 311 | 311 |
| 312 | 111 | 211 | 212 | 312 | 312 |
| 313 | 111 | 211 | 212 | 312 | 313 |
| 314 | 111 | 112 | 212 | 213 | 313 |
| 320 | 210 | 220 | 320 | 320 | 320 |
| 321 | 111 | 211 | 221 | 321 | 321 |
| 322 | 111 | 211 | 311 | 321 | 322 |
| 323 | 111 | 211 | 212 | 222 | 322 |
| 330 | 210 | 220 | 320 | 330 | 330 |
| 331 | 111 | 211 | 221 | 321 | 331 |
| 332 | 111 | 220 | 221 | 222 | 331 |
| 340 | 120 | 220 | 230 | 330 | 340 |
| 341 | 111 | 121 | 221 | 231 | 331 |
| 350 | 120 | 220 | 230 | 240 | 340 |
| 400 | 300 | 400 | 400 | 400 | 400 |
| 401 | 201 | 301 | 401 | 401 | 401 |
| 402 | 201 | 301 | 302 | 402 | 402 |
| 403 | 201 | 202 | 302 | 303 | 403 |
| 404 | 201 | 202 | 302 | 303 | 403 |
| 410 | 210 | 310 | 410 | 410 | 410 |
| 411 | 111 | 211 | 311 | 411 | 411 |
| 412 | 111 | 211 | 311 | 312 | 412 |
| 413 | 111 | 211 | 212 | 312 | 313 |
| 420 | 210 | 310 | 320 | 420 | 420 |
| 421 | 111 | 211 | 311 | 321 | 421 |
| 422 | 210 | 211 | 311 | 321 | 421 |
| 430 | 210 | 220 | 320 | 330 | 430 |
| 431 | 111 | 211 | 221 | 321 | 331 |
| 440 | 210 | 220 | 320 | 330 | 430 |
| 500 | 300 | 400 | 500 | 500 | 500 |
| 501 | 201 | 301 | 401 | 501 | 501 |
| 502 | 201 | 301 | 401 | 402 | 502 |
| 503 | 201 | 202 | 302 | 402 | 403 |
| 510 | 210 | 310 | 410 | 510 | 510 |
| 511 | 111 | 211 | 311 | 411 | 511 |
| 512 | 111 | 211 | 311 | 411 | 412 |
| 520 | 210 | 310 | 410 | 420 | 520 |
| 521 | 111 | 211 | 311 | 411 | 421 |
| 530 | 210 | 220 | 320 | 420 | 430 |
| 600 | 300 | 400 | 500 | 600 | 600 |
| 601 | 201 | 301 | 401 | 501 | 601 |
| 602 | 201 | 301 | 401 | 402 | 502 |
| 610 | 210 | 310 | 410 | 510 | 610 |
| 611 | 111 | 211 | 311 | 411 | 511 |
| 620 | 210 | 310 | 410 | 420 | 520 |
| 700 | 300 | 400 | 500 | 600 | 700 |
| 701 | 201 | 301 | 401 | 501 | 601 |
| 710 | 210 | 310 | 410 | 510 | 610 |
| 800 | 300 | 400 | 500 | 600 | 700 |

## Step 4: Allocate Priority Groups to Traffic Classes

- Priority Groups with Traffic Type EP are allocated the lowest numbered Traffic Classes, followed by En, nn, AVB, and unused
- Merge Priority Groups into Traffic Classes from lowest to highest numbered
- If possible, the same number of Priority Groups are assigned to each Traffic Class for a give Traffic Type

If not, the lowered numbered Traffic Classes are assigned one more Priority Group than the higher numbered Traffic Classes

- If multiple Priority Groups with EP or En Traffic Types are merged into a single Traffic Class, then the bandwidth allocated to that Traffic Class is the sum of the merged Priority Groups


## Step 5

- Populate the read only Priority to Traffic Class table Simply extract from the Priority to PG table and the PG table


## An non-AVB Example

This is what the administrator would set up:

| Priority | Traffic Contents | Priority Group | Priority Group | Traffic Type | BW Allocation | Traffic Class (RO) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 7 | High Priority Management Traffic | 7 |  |  |  |  |
|  |  |  | 7 | nn | na (reserved) |  |
| 6 | Standards Group LAN | 6 | 6 | nn | na (reserved) |  |
| 5 | Engineering LAN | 5 | 5 | nn | na (reserved) |  |
| 4 | IPC | 4 | 4 | EP | 50 |  |
| 3 | Storage | 3 | 3 | EP | 30 |  |
| 2 | iSCSI | 2 | 2 | En | 20 |  |
| 1 | All other LAN | 1 | 1 | nn | na (reserved) |  |
| 0 | Backup | 0 | 0 | nn | na (reserved) |  |

## Determine Number of Traffic Classes

| Priority <br> Group | Traffic <br> Type | BW <br> Allocation | Traffic <br> Class <br> (RO) |
| :--- | :--- | :--- | :--- |
| 7 | nn | na (reserved) |  |
| 6 | nn | na (reserved) |  |
| 5 | nn | na (reserved) |  |
| 4 | EP | 50 |  |
| 3 | EP | 30 |  |
| 2 | En | 20 |  |
| 1 | nn | na (reserved) |  |
| 0 | nn | na (reserved) |  |

- Assume bridge supports 4 Traffic Classes

Only 1 Priority Group is carrying En Traffic so it is allocated one Traffic Class

This leaves two Priority Groups carrying EP traffic and 5 carrying nn traffic, with only 3 Traffic Classes remaining.
Using the table, refer to row $\{E P, E n, n n\}=\{215\}$
From the table, we see that the EP traffic is allocated 1 Traffic Class, En is allocated 1 Traffic Class and the nn traffic gets 2 Traffic Classes (kind of self-evident)

## Assigning the Traffic Classes

- EP is assigned first; one Traffic Class for all EP Priority Groups, which would be Traffic Class 0
- En is assigned next, with one Traffic Class (Traffic Class 1)
- nn is next with two Traffic Classes (Traffic Classes 2 and 3)

Note the lower numbered Traffic Classes \& Priority Groups get greater merging

- Note that since Priority Groups 3 and 4 are merged into one Traffic Class, the BW allocation for that class is the sum of the merged Priority Groups, or 80\%

| Priority <br> Group | Traffic <br> Type | BW <br> Allocation | Traffic <br> Class <br> (RO) |
| :--- | :--- | :--- | :--- |
| 7 | nn | na (reserved) | 3 |
| 6 | nn | na (reserved) | 3 |
| 5 | nn | na (reserved) | 2 |
| 4 | EP | 50 | 0 |
| 3 | EP | 30 | 0 |
| 2 | En | 20 | 1 |
| 1 | nn | na (reserved) | 2 |
| 0 | nn | na (reserved) | 2 |

## An AVB Example

This is what the administrator would set up:

| Priority | Traffic Contents | Priority <br> Group |
| :--- | :--- | :--- |
| 7 | High Priority <br> Management <br> Traffic | 7 |
| 6 | iSCSI | 2 |
| 5 | AVB Type 1 | 5 |
| 4 | AVB Type 2 | 4 |
| 3 | IPC | 3 |
| 2 | Storage | 1 |
| 1 | Backup | 0 |
| 0 | General IP Traffic | 0 |

Priority to Priority Group Table

| Priority <br> Group | Traffic <br> Type | BW Allocation | Traffic <br> Class <br> (RO) |
| :--- | :--- | :--- | :--- |
| 7 | nn | na (reserved) |  |
| 6 | unused | na (reserved) |  |
| 5 | AVB | na (reserved) |  |
| 4 | AVB | na (reserved) |  |
| 3 | EP | 50 |  |
| 2 | En | 20 |  |
| 1 | EP | 30 |  |
| 0 | nn | na (reserved) |  |

## Determine Number of Traffic Classes

| Priority <br> Group | Traffic <br> Type | BW Allocation | Traffic <br> Class <br> (RO) |
| :--- | :--- | :--- | :--- |
| 7 | nn | na (reserved) |  |
| 6 | unused | na (reserved) |  |
| 5 | AVB | na (reserved) |  |
| 4 | AVB | na (reserved) |  |
| 3 | EP | 50 |  |
| 2 | En | 20 |  |
| 1 | EP | 30 |  |
| 0 | nn | na (reserved) |  |

- Assume bridge supports 5 Traffic Classes

AVB always gets separate Traffic Classes, so one Traffic Class each is allocated to Priority Groups 5 and 4

Priority Group 2 is the only one with En traffic, so it is allocated one Traffic Class

There are two Priority Groups with nn traffic and two with EP traffic, but only two Traffic Classes remain, therefore, Priority Group 3\&1 share a Traffic Class and Priority Group 7\&0 share a Traffic Class

## Assigning Traffic Classes

" EP is assigned first; one Traffic Class has been allocated for EP, which would be Traffic Class 0 (start from 0 and work up)

- En is assigned next, with one Traffic Class (Traffic Class 1)
- nn is next with one Traffic Class (Traffic Class 2)
- AVB is last, with a Traffic Class each (Traffic Class 3 and Traffic Class 4)
- Note that since we merged Priority Group 1 and 3 into one Traffic Cass, the bandwidth allocation for that

| Priority <br> Group | Traffic <br> Type | BW <br> Allocation <br> Class <br> (RO) |  |
| :--- | :--- | :--- | :--- |
| 7 | nn | na (reserved) | 2 |
| 6 | unused | na (reserved) |  |
| 5 | AVB | na (reserved) | 4 |
| 4 | AVB | na (reserved) | 3 |
| 3 | EP | 50 | 0 |
| 2 | En | 20 | 1 |
| 1 | EP | 30 | 0 |
| 0 | nn | na (reserved) | 2 | class is the sum of the merged Priority Groups, or 80\%

## Observations

- The Priority to Priority Group table and Priority Group table provide a consistent external abstraction
- The algorithm maps these tables into the available traffic classes in a deterministic manner

Available traffic classes remains opaque externally
Bridges exercise best effort to exhibit behavior defined by the two tables

Reasonable approximation is obtained if there are fewer Traffic Classes than Priority Groups

- The tables cannot be programmed in such a way that results in undefined behavior

Thank You!

