A proposal for 802.1Qaz Enhanced Transmission Selection

Anoop Ghanwani

(anoop@brocade.com)

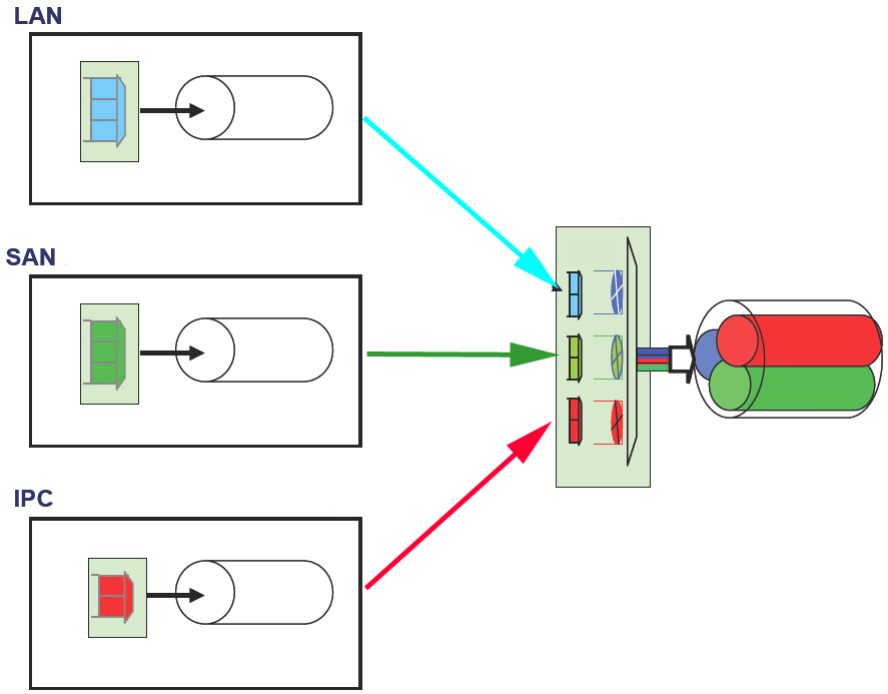
Contributors and Supporters

Amit Shukla – Juniper
Anoop Ghanwani - Brocade
Anjan – Cisco
Anthony Faustini - Cisco
Asif Hazarika – Fujitsu
Avi Godbole – Juniper
Awais Nemat – Marvell
Bruce Klemin – Qlogic
Bruce Kwan - Broadcom
Claudio DeSanti- Cisco
Craig W. Carlson - QLogic
Dan Eisenhauer – IBM
Danny J. Mitzel - Brocade
David Peterson – Brocade
Diego Crupnicoff – Mellanox
Dinesh Dutt - Cisco
Douglas Dreyer – IBM
Ed Bugnion - Nuova
Ed McGlaughlin – Qlogic
Eric Multanen - Intel
Gaurav Chawla - Dell
Glenn Wenig - Brocade
Hemal Purohit - QLogic
Hugh Barrass – Cisco
Ilango Ganga - Intel
Irv Robinson – Intel
J. R. Rivers – Nuova
Jeelani Syed - Juniper
Jeffrey Lynch - IBM
Jim Larsen - Intel
Joe Pelissier - Cisco
John Hufferd – Brocade
John Terry - Brocade
Manoj Wadekar – Qlogic
Menu Menuchehry - Marvell
Mike Ko – IBM
Mike Krause - HP
Parag Bhide - Emulex
Pat Thaler - Broadcom
Ravi Shenoy - Emulex
Renato Recio - IBM
Robert Snively - Brocade
Roger Hathorn - IBM
Sanjaya Anand – Qlogic
Sanjay Sane – Cisco
Shreyas Shah - PLX
Silvano Gai - Nuova
Stuart Berman - Emulex
Suresh Vobbilisetty - Brocade
Taufik Ma - Emulex
Uri Elzur - Broadcom

Overview

- Goals of ETS
- Proposed scheme
- Minimum scheduler requirements
- Configuration recommendations

Goals for ETS



- To support a converged network, need a way to satisfy the requirements of different traffic types
  - IPC – low latency
  - SAN – loss sensitive
  - LAN – latency & loss tolerant
- Ideally a “virtual pipe” for each traffic type so that interference is minimized
- Two pieces
  - Minimum scheduler behavior that allows this
  - Managed objects for consistent configuration in a multi-vendor environment

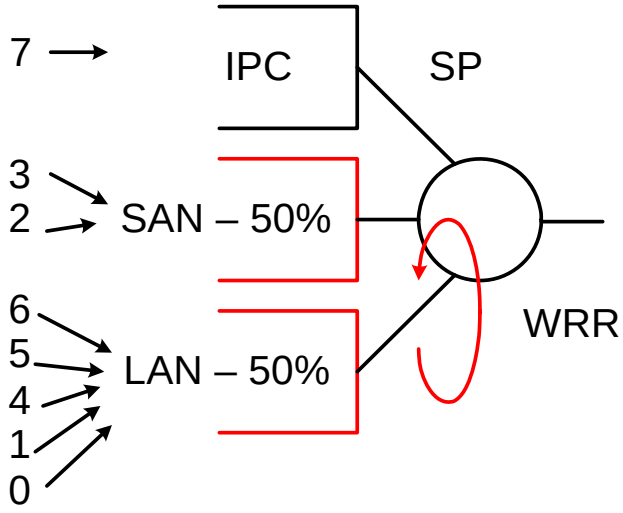
Definitions

- **Priority**
  - 3-bit priority in the 802.1Q tag.
- **Priority Group (PG)**
  - A set of priorities bound together by management for the purpose of bandwidth allocation
  - All priorities in a PG are expected to have similar traffic handling requirements with respect to latency and loss
- **Priority Group ID (PGID)**
  - A 4-bit identifier assigned to a priority group
  - PGID = 15 is a special value that indicates the priority is not managed by ETS; this is used for strict priority and AVB
  - PGID values from 8 to 14 are not used
- **Priority Group BW (PG%)**
  - Percentage of available link bandwidth allocated to a particular PGID

Priority Groups and Bandwidth Assignment – An Example

Priority	PGID	Desc
7	15	IPC
6	1	LAN
5	1	LAN
4	1	LAN
3	0	SAN
2	0	SAN
1	1	LAN
0	1	LAN

PGID	BW%	Desc
0	50	SAN
1	50	LAN
-	-	-



- There are 3 priority groups in use – IPC, LAN & WAN
- First all IPC traffic is serviced (priority 7)
- Next, the available bandwidth is shared equally by LAN (priority 6,5,4,1,0) & SAN (priority 3,2)
- Within a PG, scheduling is not specified

Minimum Scheduler Requirement

- Devices shall support at least 3 Priority Groups
  - One or more priorities with PGID 15
  - At least one PG with bandwidth allocation with all of the priorities within that group having PFC enabled
  - At least one PG with bandwidth allocation with all of the priorities within that group having PFC disabled
- BW configuration with at least 1% granularity.
- Work-conserving transmission selection policy

Configuration Recommendations

- Don’t group priorities having dissimilar traffic handling in the same PG; e.g. PFC traffic should not be grouped with non-PFC traffic
- Don’t map priorities from multiple Priority Groups to the same traffic class; behavior in such instances is undefined

More Information

- This presentation
  - az-ghanwani-ets-proposal-0708-v1.pdf
- Accompanying writeup
  - az-wadekar-ets-proposal-0608-v1.01.pdf