

Audio Video Bridging Gen 2 Assumptions

IEEE 802.1 AVB Plenary

May 2012 – York

Green Text = Agreed to at a Plenary (was Blue or Red)

Blue Text = Newly Agreed to (was Red at last Face 2 Face)

Black Text = Not Decided

Changes Marked with Red from last version

Don Pannell

Marvell Semiconductor

dpannell@marvell.com

Revision History

- **avb-pannell-gen2-assumptions-0512-v10: Work done in York & edited afterward**
- avb-pannell-gen2-assumptions-0312-v9: Work done in Hawaii – partial update
- avb-pannell-gen2-assumptions-0112-v8: Work done in Munich
- avb-pannell-gen2-assumptions-1111-v7: Work done in Atlanta
- avb-dolsen-gen2-assumptions-0920-v6: Work done in Nanjing
- avb-pannell-gen2-assumptions-0511-v5: Work done in San Francisco
- avb-pannell-gen2-assumptions-0311-v4: Work done in Singapore
- avb-pannell-gen2-assumptions-0111-v3: Work done in Kauai – not finished
- avb-pannell-gen2-assumptions-1110-v2: Work done in Dallas
- avb-pannell-gen2-assumptions-0910-v1: 1st grouping of all STDs – stolen from below
- at-cgunther-srp-rev2-assumptions: First draft presented July 2010, San Diego, CA

Overview

This document is a collection of concepts and ideas for *possible* inclusion in the next versions of SRP (802.1Qat **Gen 2**) and/or the Gen 2 AVB Shaper (802.1Qbv) **and/or Gen 2 gPTP (802.1ASbt)** or some new standard.

It should not be considered as a Work Item list ~~yet~~ **until the entries are Green**. Each **non-Green** item needs contributions (i.e., presentations) before it can be agreed to and considered an item to be added to a draft standard. These presentations are needed immediately.

PAR Status

- AS Amendment PAR – Approved (IEEE 802.1ASbt)
- AS Corrigendum PAR – Approved (IEEE 802.1AS-2011/Cor 1)
- Time Aware Shaper PAR – Approved (IEEE 802.1Qbv)
- Preemption PARs – Approved in 802.1 (IEEE 802.1Qbu) & In Progress in 802.3 (Yong)
- Multipath/Redundant SRP PAR – (Philippe/Oliver) – Use ISIS
- SRP Amendment PAR (Rodney)
 - MACSec bandwidth & latency issue (other frame overhead)
 - Dynamic changes to bandwidth & latency
 - Report worst cast latency assuming no new reservations
 - Configurable Max Latency – per hop new way to say ‘no’ to a reservation
 - Pre-configure a reservation via MGMT/Flash (lock this down?)
 - Link Aggregation
 - Remove MMRP/MVRP periodic timers
 - Multiple Talkers per Stream
 - More SR Classes ?
 - Configurable SR class priorities and VIDs ? (did we miss the MIB?)
 - Deadlock (Norm)
 - Make SRP an ISIS application
- Use SPB-V or SPB-M?

Requirements

All performance goals are to degrade gracefully over increasing hops

Automotive Needs [7/11]

- Max Latency: 100 uSec w/5 FE hops for Control Frames
- Other Automotive Needs:
 - Max cable hop count: 7
 - Max number of nodes (bridges & end stations): 64 [3/12]
 - Max cable length: 24m
 - Max end to end cable length: 30m
- Control data attributes (assuming Coordinated Transmission) [1/12]
 - Max data size (payload/Layer 2 Data size): 128 bytes (FE), 256 bytes (GE)
 - Max number of simultaneous transmissions: 8 (FE), 32 (GE)
 - Transmission period: 500 uSec
- Payload (Layer 2 Data) size for other traffic:
 - 256 bytes (FE), 1500 bytes (GE)
- Compatibility with Bandwidth reserved Traffic [1/12]
 - Preemption helps extend the use cases [1/12]
- Where these #'s came from [1/12]:
http://www.avnu.org/files/static_page_files/9F0A4E3F-1D09-3519-ADBA4F0C747D7640/Contributed%20Automotive%20Whitepaper_April%202011.pdf

Industrial Needs [7/11]

- Max Latency:
 - Interfering Frames (includes other same PCP frames) + Bridge Latency (not including Store Forward Latency) < 3 uSec / hop (GE only)
- Other Needs:
 - Fixed Transmission Periods – 31.25u Sec to 1 mSec
 - Max 50% of Period for Low Latency Transmissions
 - An HRM (hypothetical reference model) of 64 hops [1/12]
 - At most 512 devices off one controller
 - +/- 1uSec time sync between all nodes w/max 3ppm/sec w/125 MHz gPTP timestamp clock [1/12] (believed to be currently met by AS, but should be verified from Garner simulations)
 - At most 4096 streams
 - 10 to 300 byte control frame size
 - Sending ordering of frames from the Talker needs to be included?
 - Something needed in bridges too? Need a presentation on a proposed solution (Franz)
 - Meet the Redundancy requirements per given presentations [1/12]
 - Need to bring the recovery times requirement from the presentation here from Oliver (Pannell)

Consumer Needs

- Max Latency: Does not need to be better than Gen 1 AVB [1/12]
- Other Needs:
- The maximum time to make or break an SRP reservation in the absence of a topology change or dropped SRP packets is:
 - This goal is defined per hop assuming a max of 7 hops
 - For consumer remote control applications this must not exceed 100 mSec?
- Are there new requirements to enhance interoperability between 802 and CSN networks? (Perhaps an informative annex describing how the control plane is implemented on a CSN) (Philippe's white paper)
 - Need to extend the notion of DMN in CSN to other 802.1 protocols related to AVB [1/12]
- Explicit support of heterogeneous media networks (eg. Multipath & Load Balancing [1/12]) (Philippe)
- A non-access point 802.11 station that is also a bridge to other 802 media (currently not supported in 802.11) [3/12] (Norm)
- Policing? [3/12]

Professional Needs

- Max Latency: Will use whatever gains received from other work [3/12]
- Other Needs:
- The maximum time to make or break an SRP reservation in the absence of a topology change or dropped SRP packets is:
 - This goal is defined per hop assuming a max of 7 hops
 - For professional video applications this must not exceed 20 mSec?
- **Redundancy** – Need to know the time requirement
- **Link Aggregation**
- Policing? [3/12]
- Need presentations in order to proceed [1/12]

gPTP Generation 2 ~~Ideas~~

IEEE 802.1ASbt

gPTP ~~Possible~~ New Work - in PAR

- Support for Link Agg (IEEE 802.1AX)
- Support for other media:
 - IEEE 1901 (if changes are needed)
 - WiFi Direct (if changes are needed)
 - Others?
- Alternate Timescales (e.g., transport time zone information)
- One Step Tolerant on receive
- Look at improving performance for long daisy-chained time-aware systems (or long networks) that may be in a large ring
- Look at Faster Grand Master change over
 - Pre select a failover Grand Master so the selection when needed is faster
- Redundancy
 - Short reconfiguration w/redundant paths when one path fails
 - Look for the holes/issues in a redundant/failover system?
- Automatic measurement of link delay asymmetry

gPTP ~~Possible~~ New Work - in PAR

- Detect buffered repeaters on other than 802.3 copper links
 - Add in a variable latency in the link delay as an enhanced mechanism?
Maybe a MAC address discovery?
 - Need an alternate mechanism for long (fiber) links
- Create an Annex to show Grand Master Re-election time

gPTP Possible New Work

- Reduce BMCA convergence time/work for large (>64 node) networks &/or when a loop exists
 - Use ISIS for this [5/12]
 - See Mick Seaman's work on loop detection – Will not consider unless a contribution is supplied [1/11]
 - Large (64+) node networks force a lot of data examinations
- Provide L2 timing information for 1588v3
 - Update L2 abstract interface information for 1588v3
 - Goal is to meet/liaison with 1588v3 [1/12] Michael is it
- Work with 1588v3 to provide end-to-end quality information
 - Common service interface and information exchange
 - Is there a way to report clock quality and how shall clock quality be defined over the path it took? [1/12]
- Remove the word 'bridge' from 802.1AS [3/12]
 - It can work through routers too and other devices [3/12]

gPTP Possible New Work

- AS Reconfiguration Times [1/12]
 - Define What a reconfiguration is?
 - Death of a Master
 - Loss of a path to a Master
 - Multiple simultaneous Masters?
 - Pre configured paths? (Franz)
 - Grand master change over time is 200 mSec [3/12]
 - Time interval between loss of old Grand Master and locking to the new Grand Master by the slave?
 - Multiple Grand Masters transmit timing trees at exactly the same time?
- Add full (within reasonable limits) support for TLV rate changes from Slave port to its Master port & fix 10-8 accordingly [4/12]
- Each bridge to report its worst case Time Stamp accuracy – i.e. what clock rate is it sampling on and what is the worst case uncertainty of PHYs, etc. [4/12]
- Fig 10-13 problem with current GM when its PRI 1's is downgraded lower than another node in the network? Not sure what it does. [4/12]

gPTP Possible New Work

- <new stuff goes here>

gPTP – Won't Work On

- Security (need the requirements and level of needed security)
- Mapping between NTP & AS (applicable to 1588) – Will not do [1/11]
- No One Step support on transmit
- Hardware Two Step (immediate follow up) – No spec change needed
- How to assess the synchronization performance of a node
 - For certification – Will not do (Jan 2011)

Time Aware Shaper (TAS)

IEEE 802.1Qbv

Qbv Definitions [5/12]

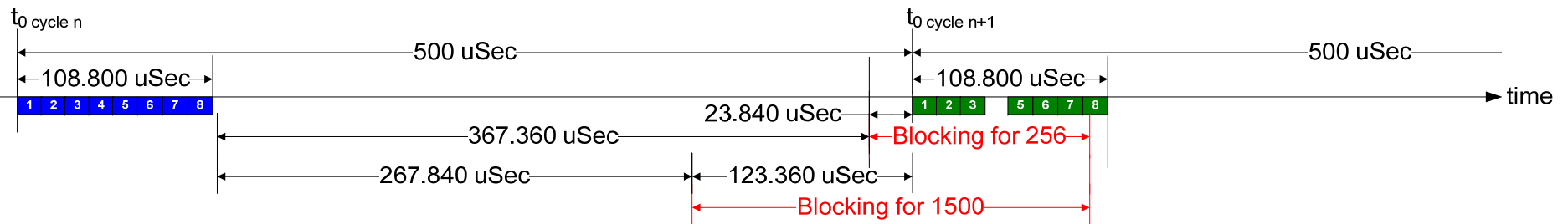
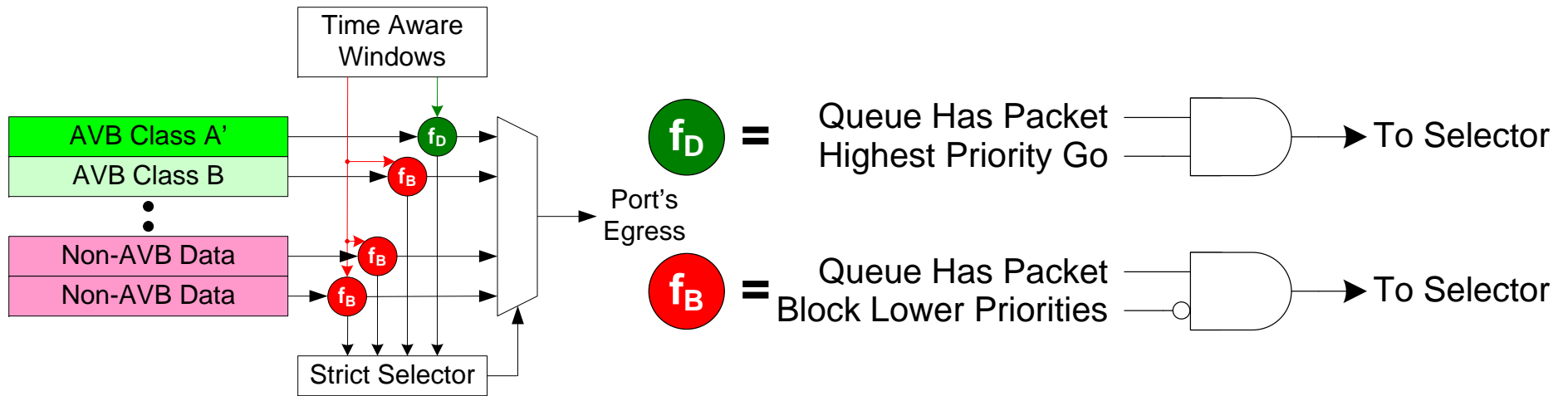
- Gating Cycle – the period of time over which a scheduled sequence of gate-opening and gate-closing events on an egress queue repeats
- Gate-opening Event – an event that connect a queue to transmission selection
- Gate-closing Event – an event that disconnects a queue from transmission selection (this does not interrupt the current frame that is being transmitted)
- Window – the time interval during which a queue is connected or disconnected to/from the transmission selection
- Guard band – the time interval during which ~~all the~~ queues are disconnected from the transmission selection
- Classes of Traffic:
 - Scheduled – Qbv traffic (uses the new IEEE 802.1Qbv shaper)
 - Reserved – SR Class traffic (uses IEEE 802.1Qav shaper)
 - Best Effort – everything else

Qbv Definitions [5/12]

- Shaper Types:
 - Scheduled – IEEE 802.1Qbv
 - Credit based – IEEE 802.1Qav

TAS Ideas [1/12]

- From New-pannell-latency_options-1111-v2:
- Proposed location of the two types of TAS shapers (Blocking and De-Blocking) with example of use of the Blocking Window



TAS Ideas [1 & 5/12]

- Needed Points in time [1/12]:
 - Don't start sending a non-critical (and preempt transmission if needed)
 - Start sending critical (t0)
 - Enable non-critical
- How does a Time Aware Network come up?
- How to handled Scheduled frames that arrive outside the transmission window?
 - Options are to drop or transmit or hold until the next window
 - If its outside the window is it too early or too late?
 - Was this frame for the right window? Will testing for this be in the standard? No
- Do we need one window interval per port or per bridge?
 - Per queue (per port) with 1 to n windows per cycle [5/12]
- Do we need this per stream? This is currently out of scope for a bridge.

TAS Ideas [5/12]

- Extend SRP or ISIS to support simple Qbv configurations?
 - At least one window per cycle per port? (Rodney)
- Need a way to tell SRP how much link bandwidth is left after Scheduled traffic is accounted for
 - SRP to use this new link bandwidth as it does presently, i.e., SR Class streams are to use no more than 75% (default) of the new link's stated remaining bandwidth?
 - Or make sure the SR Class streams always leave at least 25% (default) for Best Effort traffic?
 - The worst case latency needs to be accounted for as well
 - Do all Scheduled flows need to be configured before any plug-n-play SR Classes can be reserved?

-- Bookmark --

- Got this far in March 2012 – the rest of the assumptions were not reviewed nor updated in March 2012.
-

Other Shaper Ideas

- Improve Latency
 - Configurable credit-based Shaper that defaults to the Non-Engineered LAN settings? (i.e., use 802.1Qav)
 - To allow less latency by reducing the spreading out of frames with less than the max (75%) reservations [1/12]
 - Positive Based shaper (MJT)
 - To reduce the permanent delay and/or other pathological cases? [1/12]
 - Need presentation

Preemption

IEEE 802.1Qbu

Qbu Definitions

- Preemptive queue – a queue containing frames that can interrupt the transmission of Preemptable frames with the purpose of reducing latency of the Preemptive frames [1/12]
 - Preemptable queue – a queue containing frames whose transmission can be interrupted by Preemptive frames [1/12]
-

Preemption Ideas [1/12]

- Each queue is configured whether its frames are Preemptive or Preemptable.
 - This allows AVB Gen 1 Class A or Class B can preempt, or Class A only or ...
 - BA ver 2: Does the 75% max AVB bandwidth allocation need to change in any way with preemption?
 - BA ver 2: Is preemption enabled by default?
-

MAC Service I/F Ideas [1/12]

- Number of Levels of Preemption = 1 (i.e. 2 receive queues)
 - Need to tie in TAS to preemption's resumption of the preempted frame (the interface changes to the MAC need to take this into account)
 - The frame including its FCS delivered up the stack for a fragmented frame should not be different from what it would have been had the frame been received unfragmented
-

Redundancy Ideas

IEEE 802.1Q???

Redundancy Definitions

- Multi-Path -
 - Load Balancing –
-

Stream Reconfiguration Times [7/11]

- Gen 1 AVB's Reconfigure time = $T_{rec_routing}$ (RSTP time) + T_{rec_SRP} (SRP time)
- Pre-Reserved – Goal is a reconfiguration time = $T_{rec_routing}$ (RSTP time)
 - For example: Discovery of all possible paths to a Listener such that flows will propagate out all Bridge ports until a Blocked port is reached. Link Cost & Stream Reference Count can then be used to limit the discovered paths to two (MGMT can limit the available paths further).
- Seamless – Goal is a reconfiguration time of = 0
 - For example: Where a Listener receives more than one copy of a stream on more than one port and it can select which one to use in real time.

SRP Generation 2 Ideas

Reviewed to Here [1/12]

SRP – Possible New Work

- Dynamic bandwidth reservations (modify ‘on the fly’)
 - Done by requesting the same Stream ID with a new T-Spec?
 - Will not consider unless a contribution is supplied (Jan 2011)
- Variable bit rate reservations (statistical averaging)
 - Currently a video stream must reserve the max it will use
 - Still want to be able to Guarantee all streams are delivered (by sneaking into the 25%, which is the non-AVB bandwidth?)
 - Will not consider unless a contribution is supplied (Jan 2011)
- Dynamic changes to latency (CG)
 - Due to redundancy
 - Due to MGMT reconfiguration of a bridge
 - Change in Fan-in
 - Class % allocated
 - Due to Multiple Talkers – due to multi-Talker to one Listener
- Add the ability to get current worst case latency assuming no new reservations (CG)
 - Report Max size interfering frame that is smaller than 1522 if that is all a Talker node needs to Tx
 - Lock down the configuration by reporting SRP failed for any new requests?

SRP - Possible New Work

- **Configurable Max Latency parameter that can prevent a reservation**
 - Add support for a new lowest latency Class (i.e, 2 or 3 Classes in one LAN)
 - **On a per port and/or per bridge basis**
- Add a Tear Down Rank bit?
 - So a newer stream can stay when bandwidth is needed elsewhere?
 - Need to consider comments received from previous Qat ballots (Mar 2011)
- **Be able to create or pre-configure a reservation via MGMT/Flash (CG)**
 - For quick boot up or setup via management objects
 - Can this be done and the system remain conformant to the current standard?
 - Is there currently a way to add a reservation through management (MIB)?
- Two-way reservations – decided no need to support
 - Must be handled at a higher layer
- Explicit path reservation – like Talker Advertise pruning to save network & CPU bandwidth by reducing flooding?
 - Needed for Redundancy?
 - Intent is to make things as simple as possible but built on SRP
 - Advertise Pruning on receipt Listener Ready
 - Needed to reduce traffic on Multipath reservations

SRP - Possible New Work

- Link aggregation
 - With and without redundancy
- Redundancy (Philippe)
 - Spanning the range from no single points of failure to up to two completely independent paths with copied data <http://www.ieee802.org/1/files/public/docs2010/at-kleineberg-goetz-AVB-redundancy-1110.pdf>
 - The redundant path may be statistically over subscribed
 - Protocol neutral interface to layer 2 redundancy mechanisms
 - Need to be able to determine stream recovery times & decision metrics
 - Need to be in sync with RSTP, MSRP or allow streams to egress Blocked ports?
 - Need to look at Shortest Path Bridging & ECMP?
- Energy Efficient Ethernet
 - Remove MMRP/MVRP periodic timers on EEE links (or all the links)
 - Rest may be solved in 802.1BA (What does this mean?)
 - Is this a generic MMRP/MVRP issue that needs to be solve in Q, or an AVB specific issue that can be solved in an amendment to BA?
 - Many other protocols interfere with EEE, is fixing just MMRP and MVRP really something that we should be concerned with?
 - The feeling of the group in Nanjing is that we not address this problem.
- Unicast address Stream destination address
 - What is the real problem here? Makes Policing harder.
 - Streaming HTTP on top of TCP use an address passed to it by DNS
 - Can be detected and fixed? (i.e., make it a multicast on the AVB LAN)

SRP - Possible New Work

- Multiple Talkers per Stream (one streaming at a time)
 - Networked video switcher
 - Switch on a bit in a stream or switch at a specific time?
 - Or do the Talkers to all the turning on or off (i.e., the MUX'ing)?
 - Need the concept of a Group Reservation
- Multiple Talkers per Stream (time sliced approach) (Franz)
 - Industrial control
 - <http://www.ieee802.org/1/files/public/docs2010/at-goetz-AVB-lowlatency-part1-0510.pdf>
- More SR Classes - Yes, new 'named' performances need to be defined (CG)
 - Some applications need better than 2mSec over 7 hops of FE
- Support More PCPs? Hopefully no more than 2 SR Classes at a time are needed.
 - Allow 'moving' specific applications to specific performance levels?
 - i.e., have many SR Classes but by default only two PCP's can be in use at one time?
 - Does any environment need more than 2 PCP's?
- Configurable SR class priorities and VIDs
 - Need service primitives (e.g. REGISTER_DOMAIN.request/indication), management (clause 12), or SNMP (clause 17) to do this currently

Other Q Enhancements

- Gateway between conflicting SR Class domains – already solved as Qat requires that both SR Class and PCP match
- Reduce Latency (Yong)
 - Bursting concerns
 - Configure the characteristics of each Class's Qav Shaper
 - Preemption
 - Other options
- Automatic Talker pruning? – this is not easy
 - Should be addressed by 'Explicit path reservation'
- Need to fix SRP to support Ingress VLAN membership checking so a Talker needs to issue an MVRP join request to the VLAN it is using for flows. (CG)

Other Q Enhancements

- Enhance MRP to use difference-based updates rather than complete database updates (reduces bridge CPU overhead and control bus bandwidth usage) (Tony)
 - Another goal is to support a larger attribute set
 - May have periodic updates of a portion of the database
 - Or don't do updates at all – only do Register / Deregister?
- Add in Andre's comments to SRP.

Other Q Enhancements [3/12]

- Need an SRP mechanism to set the Qbv Schedule window & interval globally for easy configuration (don't want to have to touch every bridge manually) (Rodney – 3/12)
- SRP also needs to know the Qbv information so that it can take this into account when calculating a port's remaining bandwidth (Christian – 3/12)
- What do we do with Scheduled frames that miss their window? Drop? Transmit it anyway?

SRP - Other Ideas

- Cloud diagnostics (devices along the path)
 - Perhaps 802.1ag? Need a presentation here!
- **Ingress policing/monitoring**
 - Someone's talking when they shouldn't be
 - Talking without a reservation
 - The stream's DA is not known in the filtering database
 - The frame's PCP is AVB to a unicast
 - Talking too much for the amount reserved
 - Exceeding the reservation
 - Is this perfect policing or best effort?
 - Must it stop a Denial of Service attack?
- 802.1AE (MACsec) environments?
 - Any AVB Streams and PTP & SRP frames can be AE Tagged?
 - Clean up the interface between the link and SRP?
 - One known issues with MACsec is frame expansion that will increase the bandwidth usage
 - Does the MACsec block cipher's variable delays affect gPTP timing accuracy?
 - What is this???
- PONs are currently not specifically supported (Yong)
 - i.e., PON support is dependent on contributions from those that need it

SRP - Other Ideas

- How will MSTP select an SRP path over a CM (Congestion Management) path or a non-SRP/non-CM path using 'out of the box' defaults? (MJT)
 - For AVB with non-AVB devices: Use MSTP with at minimum one spanning tree instance and set AVB to AVB path costs low (match terms in capability vectors) and playing with root costs using MSTP's priority vector?
 - MJT will form an interest group to resolve.
- SRP for Layer 3? IETF issue?
 - Need an Internet Draft how RSVP can use SRP (Subnet Bandwidth Manager – SBM)

SRP - Other New Notes

- As per the discussion with 802.11aa (on March 17 in Singapore) they are targeted to be published in May 2012. We need to replace the duplicate Q Annex C diagrams and discussions relating to 802.11 with appropriate references and text. (CG)
- Preemption to be multi level [7/11]? Two at most? (CG)
 - Large (i.e, Jumbo) frames can be supported with AVB flows with Preemption by inference [7/11].
 - Max need of Two “Latency Controlled” Classes for automotive. Industrial? Others?

SRP - Other New Notes

- Introduce an SRP Class Z? that uses a non-AVB PCP but uses a high (6?) PCP for best effort flows. Needed for flows where the max bandwidth of the flow is not known (i.e., it doesn't have T-Spec). This allows these flows to get gPTP, SRP path selection & Gen 2 redundancy without the delivery guarantee. Target MKT is consumer.
(Philippe)
 - Need presentation

New Ideas Here

- Policing Goals or Issues [1/1 1] (Already covered elsewhere under ingress policing)
 - Someone's talking when they shouldn't be
 - Talking without a reservation
 - The stream's DA is not known in the filtering database (this item is not covered in previous slide)
 - The frame's PCP is AVB to a unicast
 - Talking too much for the amount reserved
 - Exceeding the reservation
- SRP creates its own data path tree?
 - Multipath (Philippe)