Text Proposal for: A (sub-)section on Shaper Interoperability in IEEE 802.1DG

Max Turner
Ethernovia
(BMW, Ruetz System Solutions)
June 2021

1 Glossary

1.1 Latency

The time it takes for a full frame to pass from a transmitter to a receiver. It is often referred to as the time “first bit out to last bit in”.

1.2 Delay

The time it takes for a particular element of a frame (usually the start of the fame) to pass from a transmitter to a receiver. It is often referred to as the time “first bit out to first bit in”.

1.3 TAS – Time Aware Shaper

Originally specified in IEEE Std. 802.1Qbv Enhancements to Traffic Scheduling.

1.4 CBS – Credit Based Shaper

Originally specified in IEEE Std. 802.1Qav as Forwarding and Queuing Enhancements for Time-Sensitive Streams (FQTSS).

1.5 ATS – Asynchronous Shaper

Originally specified in IEEE Std. 802.1Qcr Asynchronous Traffic Shaping.

2 IEEE802.1Q Egress Model

Within the transmission selection model of IEEE802.1Q, it is possible to logically couple CBS, TAS, and SP in series. This means in order for a frame to be transmitted, all three elements must allow the transmission at the exact same time and a frame must be available for transmission.

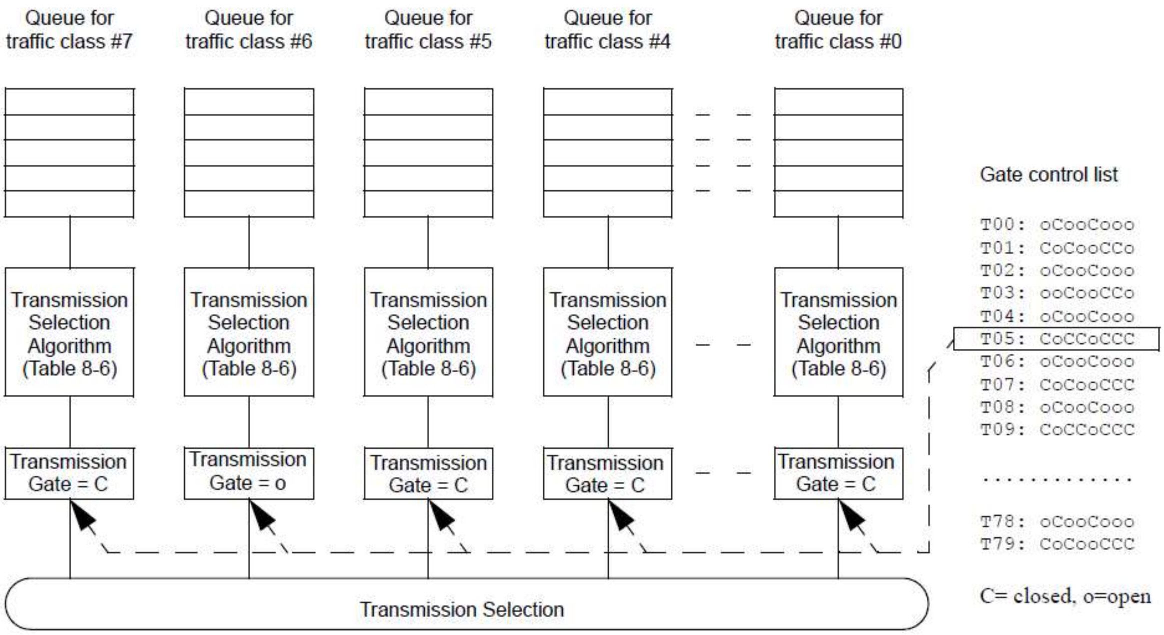

Figure 8-14—Transmission selection with gates

The TAS gate is completely unaware of the state of the queue (is a frame available?). It is controlled by a clock cycle, where two basic patterns of control are discussed below.

The CBS has a (dynamically) configured bandwidth expressed as the operIdleSlope, which accumulates credit only if a frame is in the queue and the TAS gate is open.

g) *sendSlope*. The rate of change of *credit*, in bits per second, when the value of *credit* is decreasing (i.e., while *transmit* is TRUE). The value of *sendSlope* is defined as follows:

$$sendSlope = (idleSlope - portTransmitRate)$$

h) *transmitAllowed*. Takes the value TRUE when the *credit* parameter is zero or positive; FALSE when the *credit* parameter is negative.

The ATS determines a transmission time based on teh arrival time of the frame, potentially using a different time base than the TAS.

3 Time Aware Shaper (TAS)

3.1 TAS Bus Mode

In what this document refers to as the Bus Mode, the gate control of all ports on a LAN are controlled by a single Gate Control List, who's Gate Open Times and Gate Open Duration is identical. This means the whole LAN is reserved for the transmission of a single set of messages. It is called Bus Mode as it resembled the operation of a Bus systems with a TDMA schema. In Bus Mode care must be taken to choose the Gate Open Duration long enough for the frame to propagate the maximum number of hops required. The Gate Open Duration must be at least the Latency of the set of messages to be transferred. This will usually mean having to increase the Gate Open Duration when a hop is added to the network and thereby potentially changing the whole TAS Gate Schedule in each hop of the LAN. Clock synchronisation inaccuracies across all hops need to be taken into account when determining the schedule. As a rough approximation the Gate Open Duration will be at least the number of hops times the length of the set of messages to be transported. For

100Mbit/s link speeds this will likely be in the 100's of µs, for 1Gbit/s link speeds this will likely be in the 10's of µs. These numbers are given to give an indication on the required resolution of the required time synchronization.

The number of such protected gate times is assumed to be limited as the whole LAN is blocked for the duration. If Bus Mode is allowed, the Gate Control List shall have at least 255 entries with a resolution of at least 10µs.

While the fames in question are allowed to pass undisturbed through the LAN, it must be considered, that this time is forced onto other frames as delay.

3.2 TAS Phased Mode

In what this document refers to as the Phased Mode, the gate control of each port a set of messages traverses is timed to allow the transmission without any queuing delay after reception. In a daisy chain this means the port gate of the next hop opens about the transmission latency plus the store and forward latency later than on the previous hop. The naming has been derived form phased traffic lights, which allow vehicles to get to the next intersection when the traffic lights are green. This requires exact a priory knowledge of the path the messages to be protected will travel along. It also means that any minute change in the transmission timing needs to be considered along the full path and will usually lead to a recalculation of the schedule for each hop. To increase the net usage of any egress port it will also require more accurate time resolution than Bus Mode.

The number of such protected gate times is assumed to be high compared to Bus Mode. If Phased Mode is allowed, the Gate Control List shall have at least 4095 entries with a resolution of at least 64 byte times of the line rate of the port in question (64*8bit/100Mbit/s = 5µs, 64*8bit/1Gbit/s = 0.5µs).

3.3 TAS Blockage Protection

As TAS will check for the length of a frame to ensure it can be transmitted in the allocated gate open cycle, it must be ensured that no frame with a length larger than the maximum gate open time within a Gate Control List is allowed in the egress queue. As such a mismatch could constitute an unrecoverable blockage of the port, it is required to required for any node implementing TAS to also implement a policer on frame length.

4 Buffer Overrun Protection

Misconfiguration of any TSN shaping algorithm may lead to a (slow) build up of residual queue occupancy. It is required for any node to implement a watch dog on queue occupancy, which is able to clear the queue (loss of data!) in case it reaches a threshold occupancy. While this will lead to the loss of data it prevent the overall flow of data from being completely stopped.

Further action and error recording of such an event is left to the implementer.

5 Combination of CBS and TAS

5.1 Contiguous Control List Entries

If a queue is controlled by CBS and TAS, the TAS gate control entries shall be configured in such a way that the gate for the CBS controlled queue only opens and closes at most once per TAS cycle. This means the GateOpenTime can not be fragmented across the OperCycleTime, but must be one contiguous time block, even if it extends over multiple control list entries.

5.2 General Limitations

If at least one queue of an egress port is controlled by CBS and TAS, it is important to note three possible overall behaviours:

5.2.1 Burst Operation

If the gate open time allocated in the TAS results in an identical bandwidth to what is reserved by idleSlope, the output will be a single burst wile the TAS gate is open and no effect of the CBS will be visible to the receiver on the link. This requires for no other TAS gate to be open while the queue in question is allowed to transmit. No interfering frame needs to be considered for the latency calculation.

idleSlope = Rate * (GateOpenTime / OperCycleTime)

operIdleSlope * (OperCycleTime / GateOpenTime) = Rate * (GateOpenTime / OperCycleTime)

5.2.2 Singular Operation

If the queue in question is the only one allowed to transmit, but the bandwidth allowed by the TAS gate timing is larger than that reserved for the CBS, the net bandwidth used on the link is reduced to the CBS reserved bandwidth for the time the TAS gate is open. No interfering frame needs to be considered for the latency calculation.

5.2.3 Multi Gate Operation

If at least one more gate is open along with the CBS controlled queue, it is important to configure the gate open times to allow for any interfering traffic along with the bandwidth reserved for CBS.

5.3 No-Interference Stability Condition

For the Burst and Singular operation, there is a simple stability condition that needs to be met, in order to not overrun the buffer of the queue. In these two cases no interfering frame must be considered. The mor complex cases with interference are covered further down in this document.

The CBS stream can deliver a maximum of (operIdleSlope * OperCycleTime) byte during one TAS gate cycle. There are assumed to be quantized in n Frames of equal length L. While the gate is open (Rate * GateOpenTime) byte can be transmitted.

From (operIdleSlope * OperCycleTime) = (n * L) one can find an

$n = \lceil (\text{operIdleSlope} * \text{OperCycleTime}) / L \rceil$

and due to $(n * L) \leq (\text{Rate} * \text{GateOpenTime})$, the configuration is stable if and only if:

$\lceil (\text{operIdleSlope} * \text{OperCycleTime}) / L \rceil * L \leq (\text{Rate} * \text{GateOpenTime})$

5.4 TAS: Maximum CBS bandwidth

IEEE Std. 802.1Q suggests to only reserve a maximum of 75% of a port's bandwidth for CBS (operIdleSlope ≤ 75% * Rate). Considering the Contiguous Control list requirement, the bandwidth available to the TAS gate is TASBand = Rate * GateOpenTime / OperCycleTime.

IEEE Std 802.1Q-2018 page 201

The following external parameters are associated with each queue that supports the operation of the credit-based shaper algorithm:

c) ***portTransmitRate.*** The transmission rate, in bits per second, that the underlying MAC Service that supports transmission through the Port provides. The value of this parameter is determined by the operation of the MAC.

d) ***idleSlope***. The rate of change of *credit*, in bits per second, when the value of *credit* is increasing (i.e., while *transmit* is FALSE and the transmission gate for the queue is open [8.6.8.4]). The value of *idleSlope* can never exceed *portTransmitRate*. If the enhancements for scheduled traffic (8.6.8.4) are not supported, or if GateEnabled is FALSE (8.6.9.4.14), the value of *idleSlope* for a given queue is equal to the value of the *operIdleSlope(N)* parameter for that queue, as defined in 34.3. If the enhancements for scheduled traffic (8.6.8.4) are supported, and GateEnabled is TRUE (8.6.9.4.14), then

$$idleSlope = (operIdleSlope(N) \times \text{OperCycleTime} / \text{GateOpenTime})$$

where OperCycleTime is as defined in 8.6.9.4.20 and GateOpenTime is equal to the total amount of time during the gating cycle that the gate state for the queue is Open.

NOTE 1—When scheduled traffic operation is enabled, *credit* is accumulated only while the gate is open; therefore, the effective data rate of the *idleSlope* is increased to reflect the duty cycle for the transmission gate associated with the queue; however, the value of *operIdleSlope(N)* for the queue remains unchanged.

IEEE Std. 802.1Q increases the idleSlope by the same factor the available port rate is reduced. In Singular or Multi Gate operation it seems appropriate to set idleSlope ≤ 75% * TASBand.

idleSlope ≤ 75% * Rate * (GateOpenTime / OperCycleTime)

idleSlope / Rate ≤ 75% * (GateOpenTime / OperCycleTime)

This can be achieved in two ways: Either (GateOpenTime / OperCycleTime) is set to 75% of the maximum intended bandwidth to be reserved for CBS or the (GateOpenTime / OperCycleTime) must be adjusted for every new (de-)reservation of a CBS stream. This document does not mandate the second but strongly suggests for the implementer to configure (GateOpenTime / OperCycleTime) to the maximum intended CBS bandwidth.

5.5 TAS: Maximum CBS Latency

IEEE Std 802.1BA-2011 page 15

$$\text{Max Latency} = t_{Device} + t_{MaxPacketSize+IPG} + (t_{AllStreams} - t_{StreamPacket+IPG}) \times \text{Rate/MaxAllocBand} + t_{StreamPacket}$$

where

$\mathbf{t_{Device}}$ = the internal delay of the device (in increments of 512 bit times)

NOTE 3—t_{Device} is an integral multiple of 512 bit times so that it scales with the speed of the media.

$\mathbf{t_{MaxPacketSize+IPG}}$ = the transmission time for a maximum size interfering frame (1522 octets to 2000 octets) plus its preamble and start of frame delimiter (SFD) (8 octets), and the following inter-packet gap (IPG) (12 octets)

$\mathbf{t_{StreamPacket}}$ = the transmission time for the maximum frame size of the stream that is being reserved, plus its preamble and SFD (8 octets)

$\mathbf{t_{StreamPacket+IPG}}$ = the transmission time for the maximum frame size of the stream that is being reserved, plus its preamble and SFD (8 octets) and the following IPG (12 octets)

Rate = the transmission rate of the port

MaxAllocBand = maximum allocatable bandwidth, the maximum amount of bandwidth the AVB system is able to allocate for Class A streams on the port

$\mathbf{t_{Interval}}$ = the Class A observation interval or 125 µs

$\mathbf{t_{AllStreams}}$ = (MaxAllocBand × $t_{Interval}$) / tRate = the sum of the transmission times of all Class A stream frames the AVB System is able to allocate in an observation interval (125 µs) on a port

5.5.1 Gate Open Delay

The Gate Open Delay is determined by the time it takes for the TAS gate to open, which is given by GateOpenDelay = OperCycleTime – GateOpenTime, if Contiguous Control List Entries are observed. It must be added to the formula given in IEEE Std. 802.1BA, if a TAS is combined on a queue with CBS.

5.5.2 Minimum CBS Throughput

As stated in the Multi Gate Operations section, the gate must be open long enough to allow the CBS flows to pass along with interfering traffic, this means:

GateOpenTime ≥ MaxLatency

With MaxAllocBand = idleSlope, as this determines the spread of all Frames transmitted before the Stream Packet, and tAllStreams ≤ 75% * observationInterval, as the amount of data remains unchanged:

GateOpenTime ≥ tDevice + tMaxPacketIPG + (75% * observationInterval – tStreamPacketIPG) * Rate / idleSlope + tStreamPacket

Assuming a full reservation from above:

idleSlope / Rate = 75% * (GateOpenTime / OperCycleTime)

GateOpenTime ≥ tDevice + tMaxPacketIPG + (75% * observationInterval – tStreamPacketIPG) * (OperCycleTime / GateOpenTime) / 75% + tStreamPacket

Results in a somewhat circular relation, since GateOpenTime determines idleSlope, which in turn influences MaxLatency, which puts a lower limit on GateOpenTime.

6 Combination of Preemption with CBS and TAS

6.1 Latency considerations

If frames are allowed to pre-empt the CBS shaped stream controlled by a TAS gate in Multi Gate Operation, the MaxLatency becomes strongly dependent on the maximum rate and size of the Frames in the pre-empting flow. In the extreme case the interfering frame may be pre-empted, as may any frame in the CBS stream, including the stream packet under consideration.

MaxLatency = tDevice + tMaxPacketIPG + tPrePacket + (tAllStreams – tStreamPacketIPG) * Rate / idleSlope + N * tPrePacket + tStreamPacket + tPrePacket + (OperCycleTime – GateOpenTime)

Where N is the number of pre-emption events during (tAllStreams - tStreamPacketIPG) and all pre-empting packets are of equal size (Rate * tPrePacket), including pre-emption overhead and IPG.

This document therefore mandates to limit the rate of pre-empting packets to one per GateOpenTime for a CBS controlled queue.

If a policer is put in place to ensure this limitation, this results in a very unfortunate limitation, as pre-empting frames are considered high priority. As they now have to be policed in order to not cause congestion losses in the CBS queues, we are somewhat introducing a priority inversion.

Alternatively the Buffer Overrun Protection will drop frames from the CBS stream, if the pre-empting traffic takes up too much bandwidth.

6.2 GateOpenTime Configuration

As discussed above:

GateOpenTime ≥ tDevice + tMaxPacketIPG + tPrePacket + (tAllStreams – tStreamPacketIPG) * Rate / idleSlope + N * tPrePacket + tStreamPacket + tPrePacket

GateOpenTime ≥ tDevice + tMaxPacketIPG + (75% * observationInterval – tStreamPacketIPG) * (OperCycleTime / GateOpenTime) / 75% + tStreamPacket + (N +2) * tPrePacket

Which basically leads to the circular relation for GateOpenTime above, but also implies the CBS latency as well as the GateOpenTime are now dependent on at least 2 other flows or queues, potentially making system design very difficult.