Data plane Enhancement Taxonomy

draft-joung-detnet-taxonomy-dataplane-00

Jinoo Joung, Xuesong Geng, Shaofu Peng, Toerless Eckert

DetNet Interim, Feb. 06. 2024

Overview of the draft

- Purpose
  - To facilitate the understanding of the data plane enhancement solutions, which are suggested currently or can be suggested in the future, for deterministic networking
- Scope
  - To provide criteria for classifying data plane solutions
  - To provide examples of each category, along with reasons where necessary
  - To provide strengths and limitations of the categories
- Out of scope
  - The candidate solutions currently being proposed in DetNet WG are simply listed without any descriptions. The details of the solutions are intentionally omitted.
- Definitions:
  - An enhancement solution can be a combination of multiple data plane functional entities, such as regulators, queues, and schedulers.
  - A solution can also include functional entities across network nodes, e.g. traffic enforcement or regulation functions at the edge.

Taxonomy 1: Per Hop Dominant Factor for Latency Bound

	Category 1	Category 2	Category 3
Criteria	The per hop dominant factor for the latency bound, which is the largest sum term in the expression, when the network and traffic conditions are the worst. The worst condition typically means high network utilization, large packet and burst sizes, and large number of hops.		
Indicator	Max Packet Length / Service Rate	Sum of Max Packet Lengths / Capacity	Sum of Max Burst Sizes / Capacity
Strengths	individual flow isolation	less complex than Category 1	least complex
Limitations	complex		require tighter burst control mechanisms
Example solutions	FQ, C-SCORE	DRR	ATS, CQF and variants

Example data plane solution : ATS

Latency and Backlog Bounds in Time-Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping

Ehsan Mohammadpour, Eleni Stai, Maaz Mohiuddin, Jean-Yves Le Boudec
École Polytechnique Fédérale de Lausanne, Switzerland
{firstname.lastname}@epfl.ch

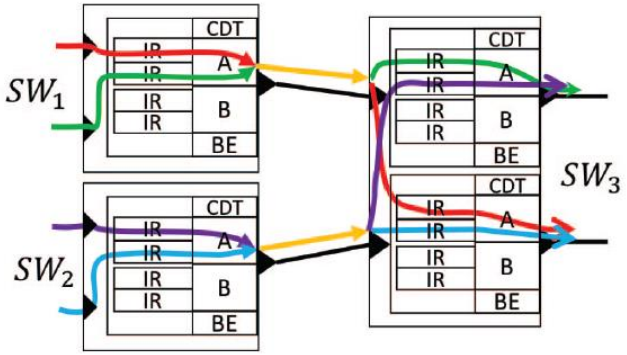


Fig. 2: Illustration of the queuing policy in interleaved regulators (IR) by TSN switches for four flows of class A.

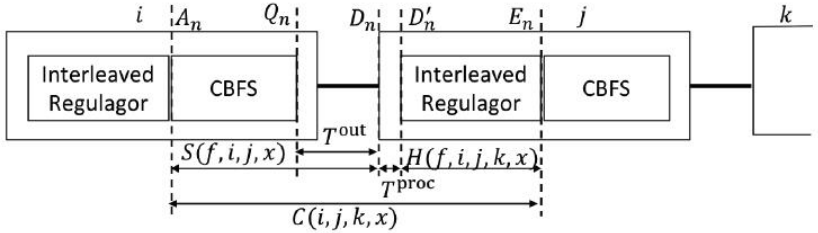


Fig. 3: Timing Model in TSN

$$C(i,j,k,x) = \sup_{f' \in F^x_{ijk}} S(f',i,j,x) + T^{\text{proc,max}}_{ij}. \quad (6)$$

f: flow
x: class
i, j, k: nodes
n: packet

The IR does not increase the worst latency of the class-based FIFO system (CBFS).

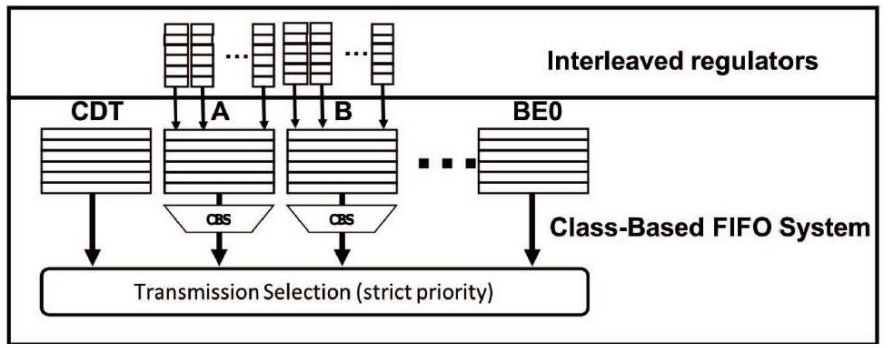


Fig. 1: Architecture of one TSN node output port.

Theorem III.2. *A tight upper bound on the response time in the CBFS of node i (following the interleaved regulator) for flow f of class $x \in \{A, B\}$, going from node i to j, is:*

$$S(f,i,j,x) = T^x_{ij} + \frac{b^{\text{tot},x}_{ij} - \psi_f}{R^x_{ij}} + \frac{\psi_f}{c_{ij}} + T^{\text{var,max}}_{ij}, \quad (5)$$

where the parameter ψ_f depends on the flow f and the type of regulator, namely, for LRQ: $\psi_f = L_f$ and for LB: $\psi_f = M_f$.

$$b^{\text{tot},x}_{ij} = \sum_{f \in F^x_{ij}} b_f$$

for LRQ: $\psi_f = L_f$ and for LB: $\psi_f = M_f$

M: minimum packet length
c: link capacity

The dominant factor is {sum of max bursts of / the allocated rate to} the class.

T^x: the delay due to the upper class traffic. For A class, the dominant factor for T^A is = {total CDT burst / (Link capacity-total CDT rate)}

R^x: the allocated rate to the class

Example data plane solution : ATS

Latency and Backlog Bounds in Time-Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping

Ehsan Mohammadpour, Eleni Stai, Maaz Mohiuddin, Jean-Yves Le Boudec
École Polytechnique Fédérale de Lausanne, Switzerland
{firstname.lastname}@epfl.ch

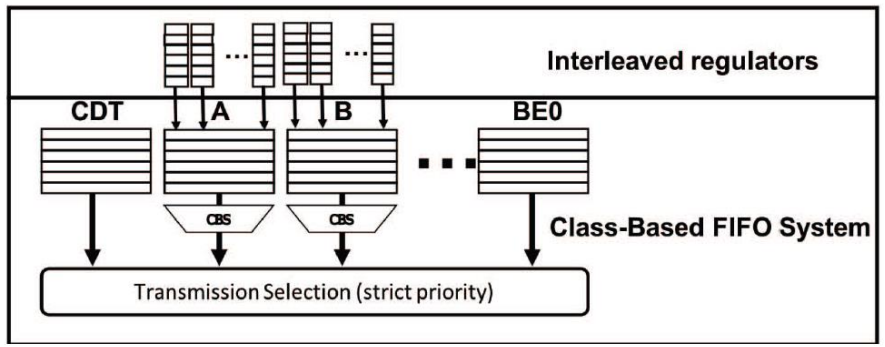


Fig. 1: Architecture of one TSN node output port.

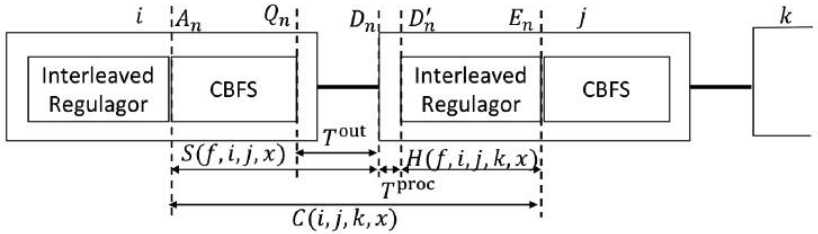


Fig. 3: Timing Model in TSN

The IR does not increase the worst latency of the class-based FIFO system (CBFS).

The dominant factor is {sum of max bursts of / the allocated rate to} the class.

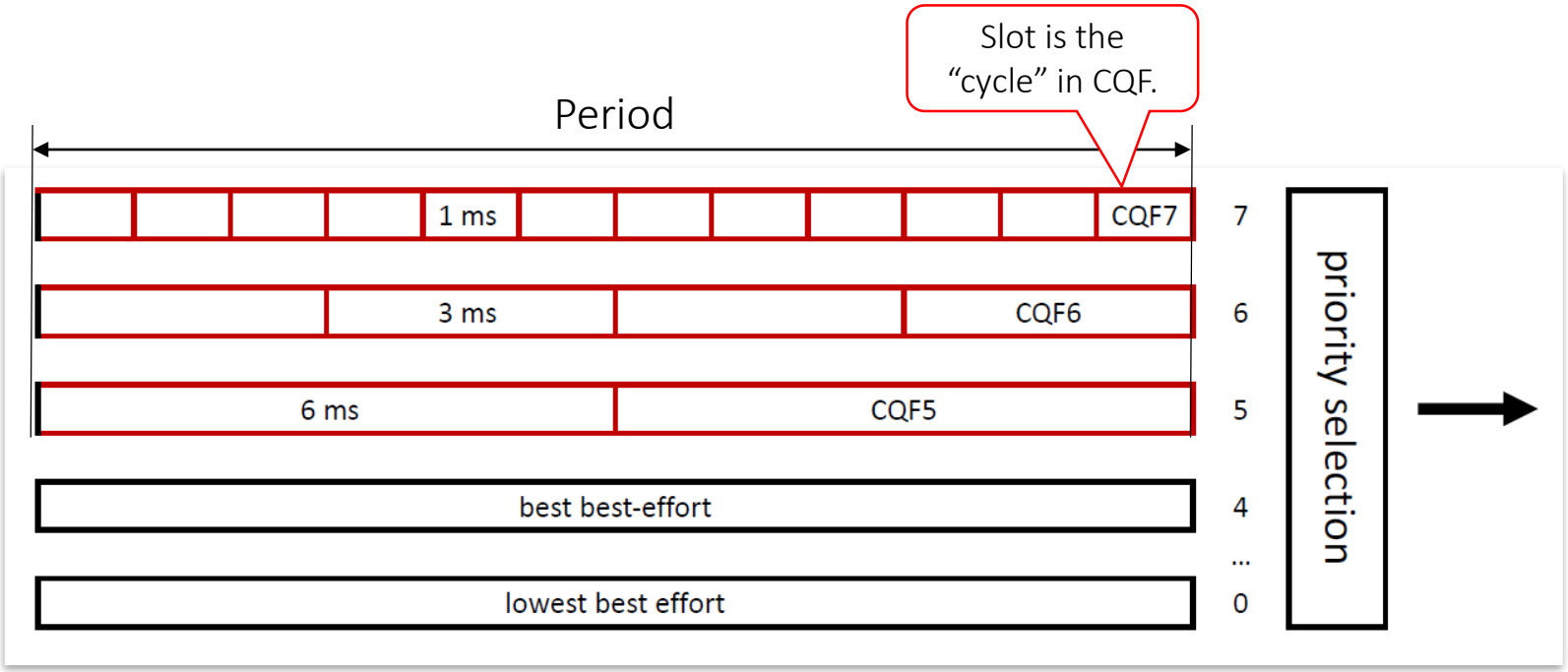
Taxonomy 2: Periodicity

	Periodic	Non-periodic
Criteria	A solution maintains a set of consecutive time slots that are repeated periodically. A packet is assigned to a particular time slot. The slot is decided with a predefined rule based on conventions such as the arrival time, the priority, the flow the packet belongs to, or the time slot it was assigned in the upstream node.	
Indicator	Meets the criteria	Does not meet
Strengths	less jitter	flexible
Limitations		
Example solutions	TAS, CQF, their variants	ATS, C-SCORE, EDF

Example data plane solution : ECQF

IEEE P802.1DC and IEEE P802.1Qdv Time Sensitive Networking Developments

Norman Finn
Huawei Technologies Co. Ltd
nfinn@nfinnconsulting.com
finn-tutorial-802-1DC-1Qdv-2023-12-v02.pdf

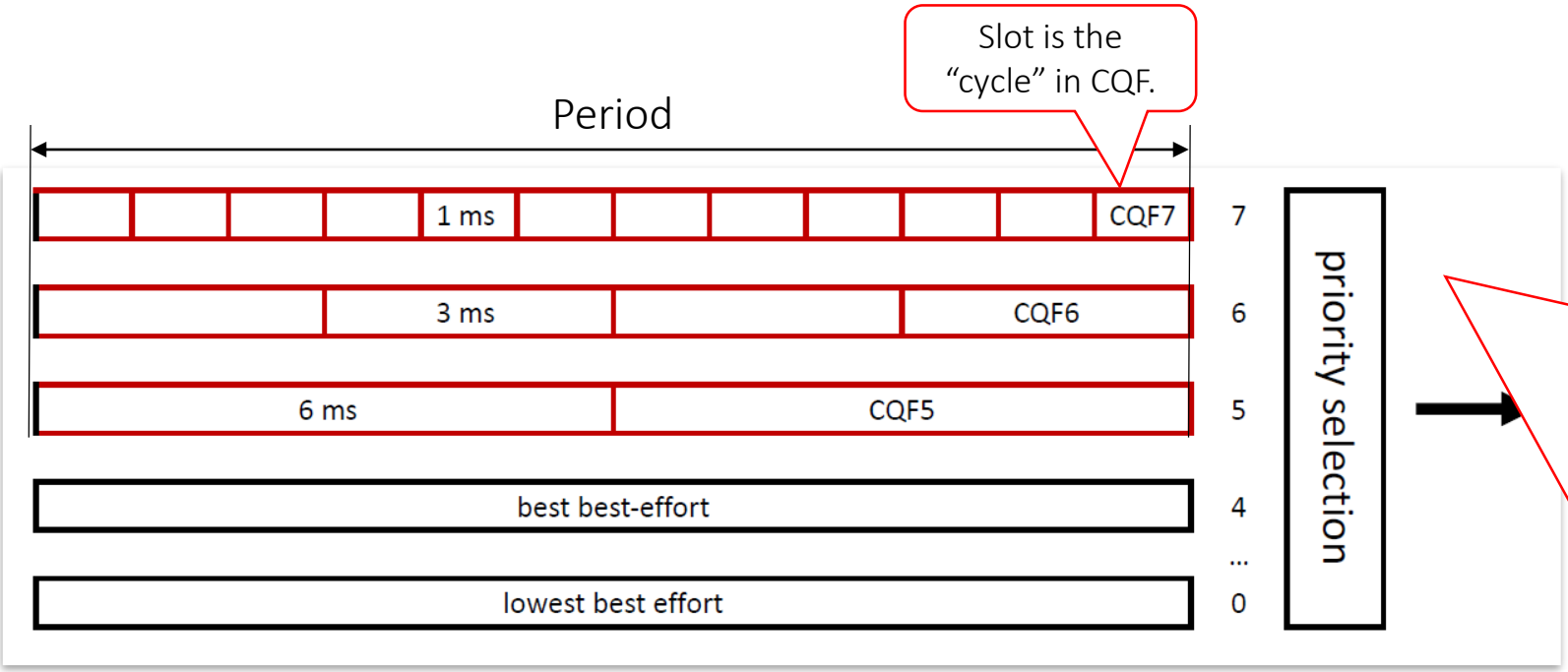


- The time slot pattern in the period is repeated, at least till a reconfiguration.
- A packet is assigned to one of the finite number of slots, with a predefined rule.
- Note that the time slots can be overlapped.

Example data plane solution : ECQF

IEEE P802.1DC and IEEE P802.1Qdv Time Sensitive Networking Developments

Norman Finn
Huawei Technologies Co. Ltd
nfinn@nfinnconsulting.com
finn-tutorial-802-1DC-1Qdv-2023-12-v02.pdf



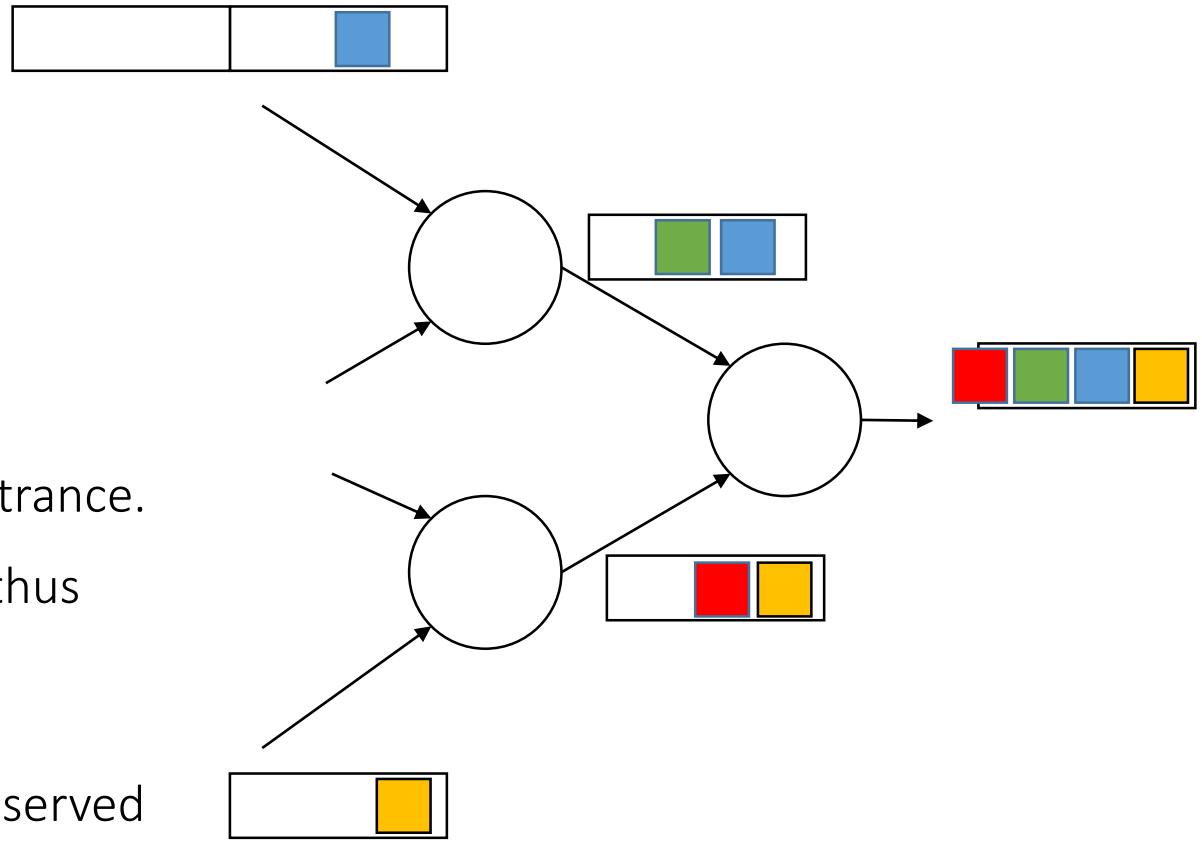
- The time slot pattern in the period is repeated, at least till a reconfiguration.
- A packet is assigned to one of the finite number of slots, with a predefined rule.
- Note that the time slots can be overlapped.

Basic CQF understanding:
E2E latency bound ~ (H+1)*(Slot time), under condition that a packet is served at the NEXT slot in the next node.

The slot time = {slot length / link capacity} is the per-hop dominant factor for the E2E latency bound. But how we determine the slot time is an open problem.

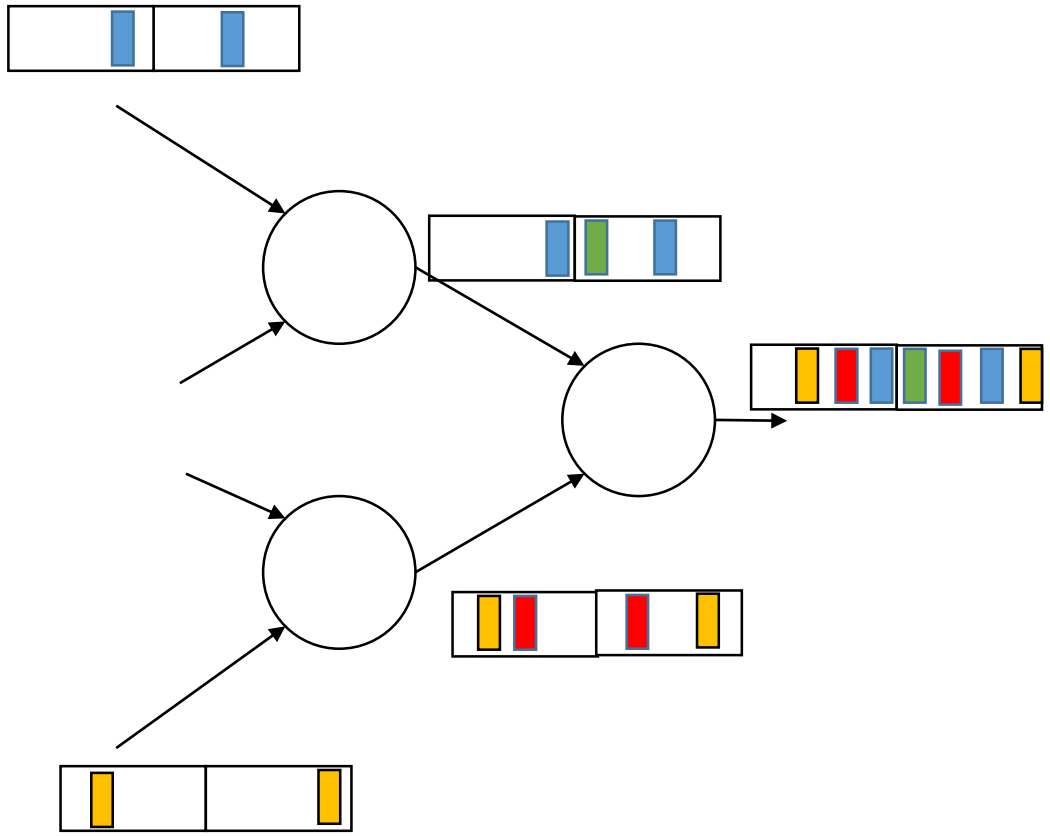
Example data plane solution : ECQF or any CQF variant

- How can we determine the slot time T_S?
- Consider only Class 7 flows.
- Assume
  - A perfectly synched network
  - The packet lengths are all identical to L.
  - All the flow's burst sizes are fixed at b.
  - There is a bottleneck link with utilization ~1.
- Transmit the burst immediately at the network entrance.
- If $\Sigma_{\text{at the bottleneck link}}$ (b) = B <= T_S *C then T_S >= B/C, thus the dominant factor is larger than or equal to B/C. (C: Link capacity)
- Otherwise, T_S < B/C then some packets cannot be served in a single slot. → This contradicts the basic CQF assumption.



Example data plane solution : ECQF or any CQF variant

- How can we determine the slot time T_S?
- Consider only Class 7 flows.
- Assume
  - A perfectly synched network
  - The packet lengths are all identical to L.
  - All the flow's burst sizes are fixed at b.
  - There is a bottleneck link with utilization ~1.
- Or we can regulate the burst at the network edge, and transmit one packet in a slot per flow.
- $\Sigma_{\text{at the bottleneck link}}$ (L) = ΣL can be set as T_S*C, thus the dominant factor is $\Sigma L/C$.
- In this case, the arrival rate from a flow has to be less than the service rate, L/T_S.
- The difference between the arrival rates of flows has to be taken into account, which is NOT trivial, either.



Taxonomy 3: Network Synchronization

	phase synchronous	frequency synchronous	asynchronous
Criteria	Whether network synchronization is required		
Indicator	require network nodes to be both phase and frequency synchronized	require network nodes to be only frequency synchronized	may also require loose phase and frequency synchronizations but with less precision.
	The required level of synch. precision is to be studied further. The level can be determined by an indicator e.g. **MTIE**.		
Strengths	precise jitter control		least complex
Limitations	complex, not scalable		additional jitter control may be necessary
Example solutions	CQF, TAS	variants of CQF & TAS	ATS, C-SCORE, EDF

Example data plane solution : ECQF

1. Bin selection based on bin number from previous hop
   a) Obtained by time-of-arrival of frame (in P802.1Qdv draft 0.4)
   b) Obtained by a field in the frame (*not* in P802.1Qdv draft 0.4)
2. Bin selection based on counting bytes stored in output bin so far (in P802.1Qdv draft 0.4)

By "ECQF", our draft means this method.

Equivalent to TCQF

Not scalable, requires flow states

Bin selection based on previous hop's bin

If the selection of the next-hop bin is derived from the last-hop bin selection, then **no per-flow state machines, and no per-flow configuration**, is required. Furthermore, the end-to-end **delivery time is constant**, modulo one bin rotation cycle.

But, this requires that all hops along the path rotate their bins at **exactly the same frequency** – that is, the difference in the number of bins output between two hops, over an arbitrarily long period of time, must be bounded.

As with other CQF variants, ECQF requires **frequency synchronization.**

How small the bound should be is for further study. MTIE can be an indicator for the synch precision and the criteria.

Taxonomy 4: Traffic Granularity

	Flow level	Flow aggregate level	Class level
Criteria	the granularity of their traffic control target, which refers to the size and specificity of the traffic entity they handle		
Indicator	Each packet is controlled based on its specific flow, which can be identified usually by the 5-tuple.	Flows are grouped by shared characteristics like traffic specification, service requirement, or routing path.	Flows are further grouped by similar service requirements, regardless of specific path or traffic details.
Strengths	more accurate service differentiation among flows		least complex
Limitations	complex		
Example solutions	FQ, C-SCORE	IR, Possible enhancement to TAS with more than 8 queues.	CQF and its variants, EDF
Note	Functional entity with the coarsest granularity is dominant, thus the whole solution belongs to the coarsest granularity category.		

Example data plane solution : ATS

Latency and Backlog Bounds in Time-Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping

Ehsan Mohammadpour, Eleni Stai, Maaz Mohiuddin, Jean-Yves Le Boudec
École Polytechnique Fédérale de Lausanne, Switzerland
{firstname.lastname}@epfl.ch

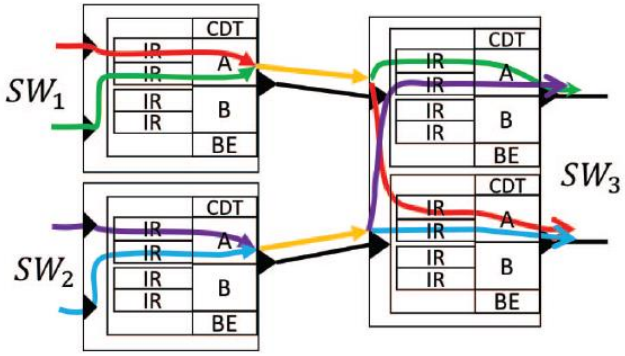


Fig. 2: Illustration of the queuing policy in interleaved regulators (IR) by TSN switches for four flows of class A.

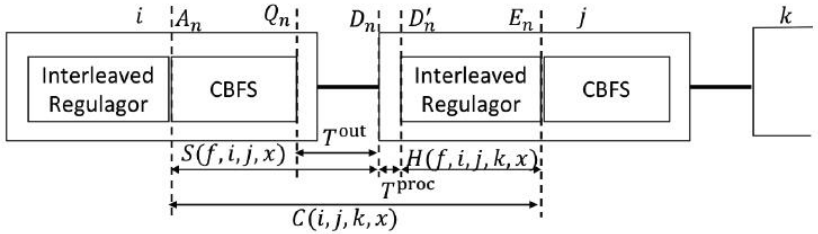


Fig. 3: Timing Model in TSN

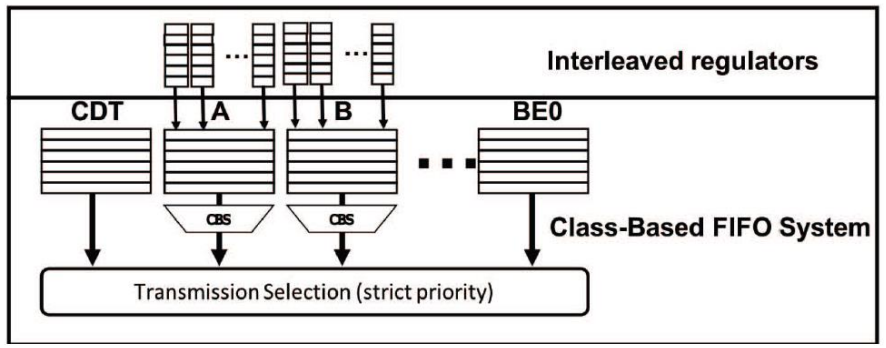


Fig. 1: Architecture of one TSN node output port.

ATS consists of interleaved regulators (IRs) and a strict priority scheduler. An IR has a queue dedicated to a flow aggregate having the same class and the same input port. The regulation function itself is based on a flow. According to the definition, **IR is a flow aggregate level solution**. On the other hand, the strict priority scheduler in ATS is class-based. Therefore, **ATS as a whole is class level**.

Example data plane solution : ECQF

IEEE P802.1DC and IEEE P802.1Qdv Time Sensitive Networking Developments

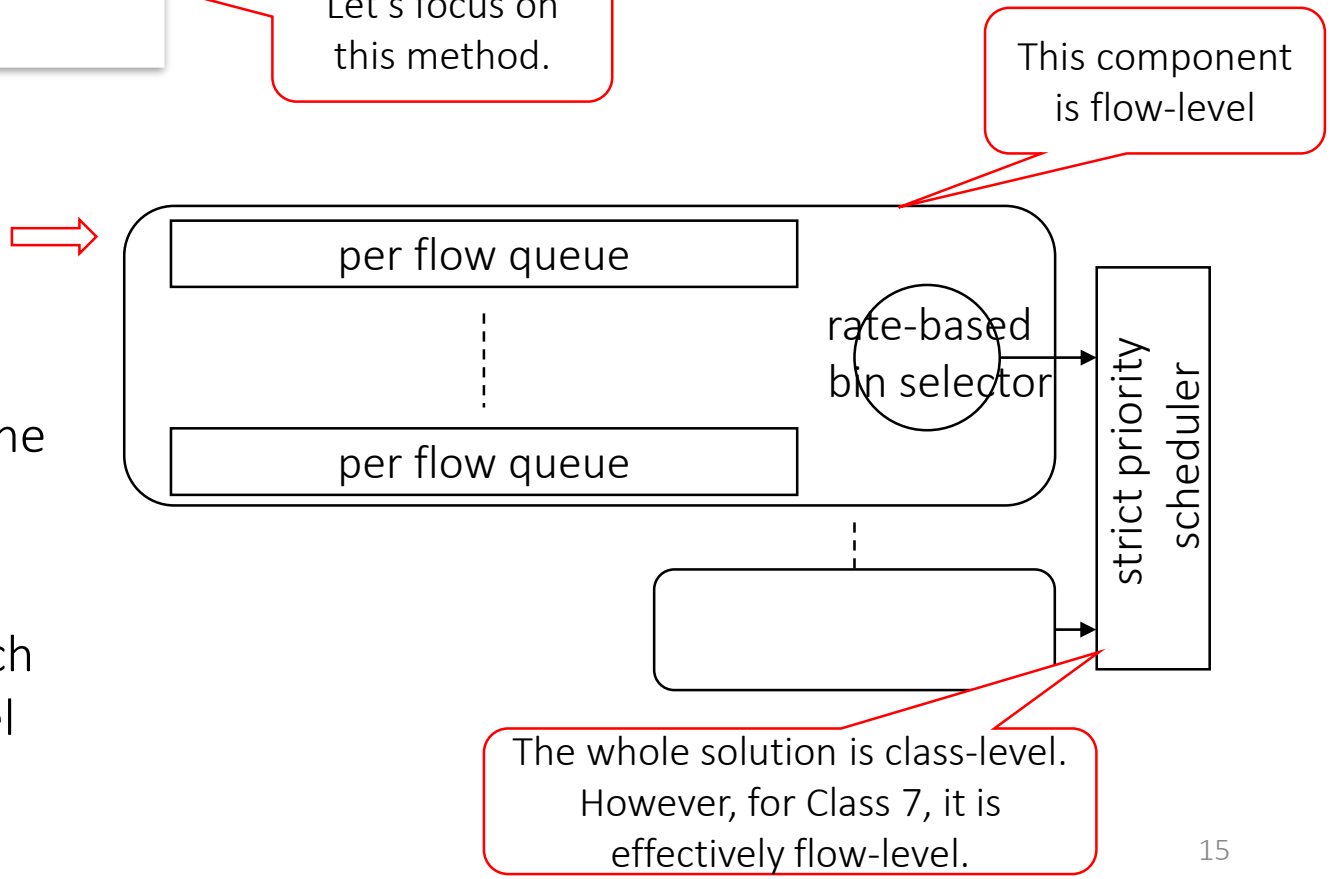
Norman Finn

1. Bin selection based on bin number from previous hop
   a) Obtained by time-of-arrival of frame (in P802.1Qdv draft 0.4)
   b) Obtained by a field in the frame (*not* in P802.1Qdv draft 0.4)
2. Bin selection based on counting bytes stored in output bin so far (in P802.1Qdv draft 0.4)

Mick Seaman's Paternoster algorithm, included in P802.1Qdv Draft 0.4, uses a byte counting state machine, per flow, per output, port to ensure that no flow exceeds its allocation. This provisioning affects scaling negatively, but there is no interaction between flows, so no massive recomputation is ever necessary. These

Let's focus on this method.

- If a bin is selected based on flow's service history, then it is a hierarchical scheduler with rate-based per-flow bin selector.
- It is similar to DRR, which determines a packet's "round" to be served, based on the flow's service history and the rate.
- For Class 7 flows, the achievable per hop latency dominant factor seems ΣL/C, which is the same with DRR (which is a flow-level solution).



Taxonomy 5: Work Conserving

	Work conserving	Non-work conserving
Criteria	If a solution never idle when there is a packet to send.	
Indicator	Meets the criteria	No
Strengths	small average latency, small observed maximum latency than the bound, the statistical multiplexing gain. fit well to bursty traffic, without a need for overprovisioning	avoid burst accumulation, jitter control, simple latency evaluation process
Limitations		
Example solutions	FIFO, round robin schedulers, FQ , C-SCORE	TAS, CQF, ATS, and their variants

Taxonomy 6: Target Transmission Time

	On-time	In-time
Criteria	how closely they adhere to predefined target transmission times for packets	
Indicator	strive to transmit packets as close as possible to their target times without ever exceeding them	transmit packets without a specified target transmission time
Strengths	typically control the jitter as well as latency	less average latency
Limitations	larger average latency	additional jitter control may be necessary
Example solutions	TAS, CQF and their variants, EDF (on-time mode)	ATS, C-SCORE, EDF (in-time mode)
Note	The on-time/in-time taxonomy here is about the scheduling decision, which determines when a packet is transmitted. It is not about the consequence of the scheduling, whether the jitter bound is also guaranteed or not.	

Example data plane solution : ATS

Latency and Backlog Bounds in Time-Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping

Ehsan Mohammadpour, Eleni Stai, Maaz Mohiuddin, Jean-Yves Le Boudec
École Polytechnique Fédérale de Lausanne, Switzerland
{firstname.lastname}@epfl.ch

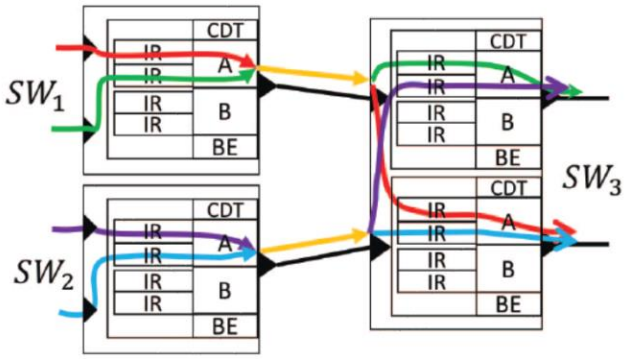


Fig. 2: Illustration of the queuing policy in interleaved regulators (IR) by TSN switches for four flows of class A.

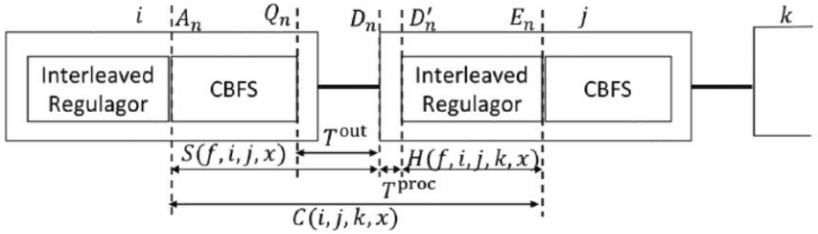


Fig. 3: Timing Model in TSN

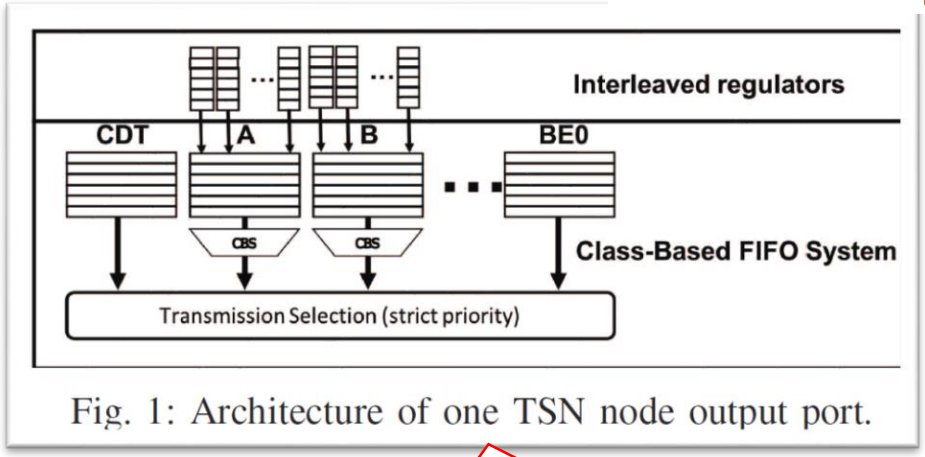


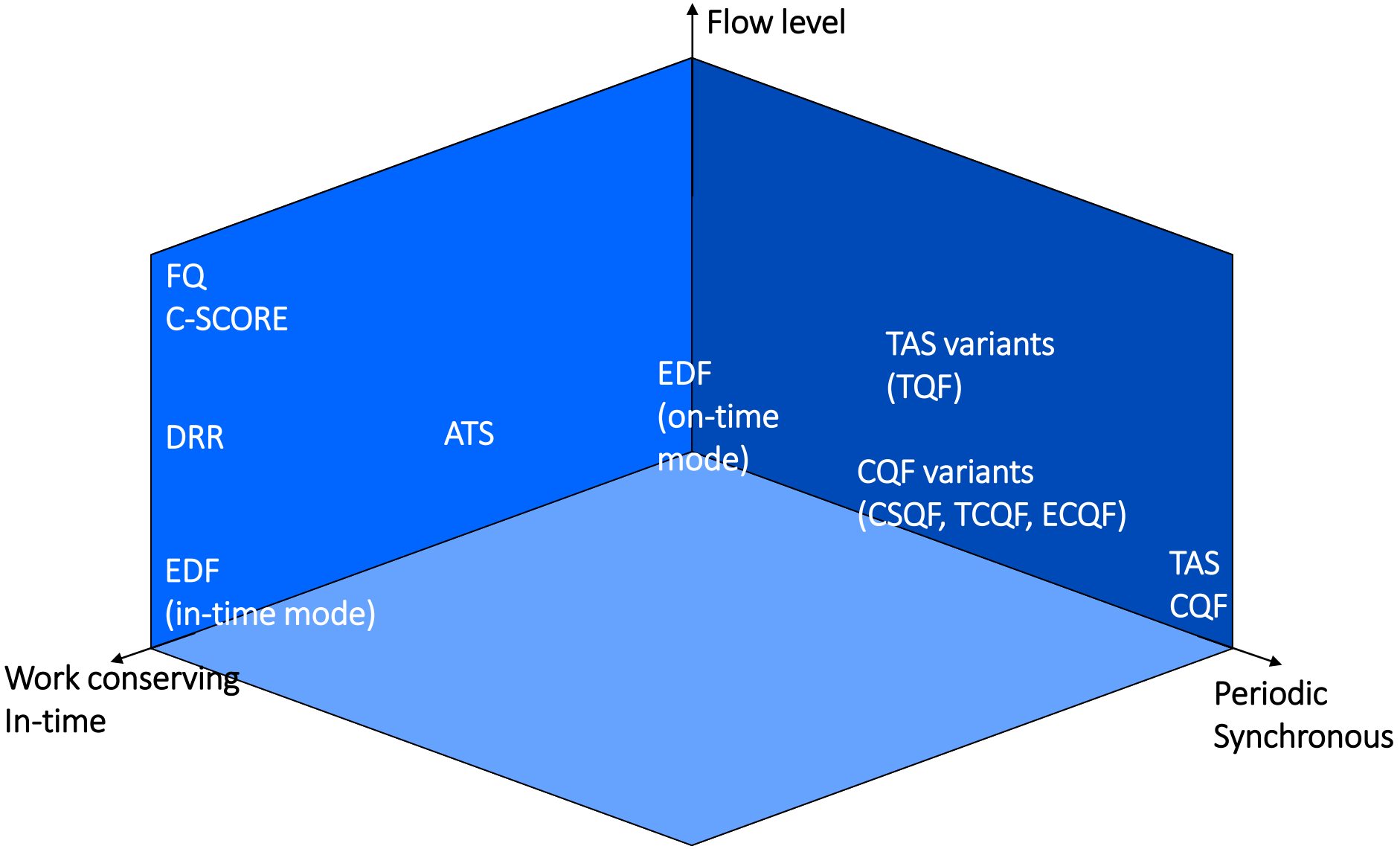
Fig. 1: Architecture of one TSN node output port.

ATS, which includes the interleaved regulator, is an in-time solution. A regulator determines an eligible time for a packet to be transmitted. Packets are always transmitted at or later than their eligible times. **An eligible time is not a target transmission time.** Note that ATS is a non-work conserving but in-time solution.

Taxonomy 7: Service Order

	Rate-based	Time-based	Arrival-based	Priority-based
Criteria	The primary service order decision factor for packets from different flows. (The rule for service order decision can be a combination of multiple factors.)			
Indicator	the allocated service rate of their flows or flow aggregates.	the allowed delay or deadline	the order they arrive	the assigned priorities
Strengths	the “pay burst only once” property, simple admission control process	precise delay control	implementation simplicity	implementation simplicity
Limitations				
Example solutions	DRR, FQ , C-SCORE	EDF	IR	ATS, TAS, CQF, and their variants
Note	If the rule based on the arrival time is combined with the other rules, then the arrival time is considered the secondary factor.			

Visualization of taxonomy



Thank you

- Please take a look at

  https://datatracker.ietf.org/doc/draft-joung-detnet-taxonomy-dataplane/

- Please share your comments and questions.
- References:
  - Mohammadpour, Ehsan, Eleni Stai, Maaz Mohiuddin, and Jean-Yves Le Boudec. “Latency and backlog bounds in time-sensitive networking with credit based shapers and asynchronous traffic shaping.” In 2018 30th International Teletraffic Congress (ITC 30), vol. 2, pp. 1-6. IEEE, 2018.
  - Norman Finn, IEEE P802.1DC and IEEE P802.1Qdv Time Sensitive Networking Developments, finn-tutorial-802-1DC-1Qdv-2023-12-v02.pdf, DetNet WG Interim meeting, Dec. 2023.