Enhanced DetNet Data Plane: Progress Report

Report on DetNet Open Working Meetings

IETF 117 (San Francisco): July 26, 2023

David Black

Tech Advisor (Transport)

DetNet: Open Working Meetings (7)

- Seven (7) meetings since IETF 116 (Yokohama)
- Two-part agenda (for each meeting):
  1. Process-oriented topics - requirements, draft contents, evaluation, etc.
  2. In-depth presentation of a queueing/scheduling mechanism
- Process-oriented outcomes:
  – Revisions to scaling requirements draft
  – Initial evaluations of TSN mechanisms against requirements (led to more revisions of scaling reqts. draft)
    - TSN Evaluation slides included in this deck (for reference not for presentation @ IETF 117)
- In-depth mechanism presentation at each meeting

Open Working Meeting Plans
(between IETF-117 and IETF-118)

- Evaluations of proposed new mechanisms
  - Draft authors prepare initial evaluation
  - Meeting discussion of each evaluation
- Requirements coverage by mechanisms
  - Meeting discussion
- Proposed logistics:
  - Tuesdays, 8a-10a Eastern, every 3 weeks
    - First meeting: Aug 22, so that Sep 10 second meeting avoids Sep 2-4 US holiday weekend
  - MeetEcho (due to IETF WebEx failure)

TSN Mechanisms: Initial Evaluation Slides

(From Open Working Meetings)

CQF evaluation with DetNet Scaling Requirement

	Requirement	CQF (802.1Qch) evaluation	Remarks
1	Tolerate Time Asynchrony	No	Time sync is mandatory and all nodes rotate transmission buffers according to a wall clock time.
2	Support Large Single-hop Propagation Latency	No	prop latency << cycle interval time Tc. Large prop latency will make Tc extremely large and hard to achieve e2e bounded latency.
3	Accommodate the Higher Link Speed	Partial	Relatively larger time variation (e.g. MTIE, processing latency)=> harder to determine from arrival time which Cycle the packet was sent from
4	(4.1) Be Scalable to The Large Number of Flows (4.2) and Tolerate High Utilization	4.1 Partial 4.2 No	4.1 Transmission control is scalable. Stream gate filtering and policy may not be (can be per stream or some way aggregated). 4.2 Not all cycle time is usable for traffic. Utilization is constrained by ratio of dead time (DT) to Tc. In small Tc cases, hard to achieve high utilization.
5	Prevent Flow Fluctuation from Disrupting Service	Partial	For compliant flows, ok; for non compliant flows, will drop packets since # of buffer is very limited, fundamentally only 2.
6	Tolerate Failures of Links or Nodes and Topology Changes	NA	Not relevant to CQF itself.
7	Be scalable to a Large Number of Hops ~~with Complex Topology~~ (to achieve e2e latency)	No	2-buffer CQF normally supports 100x µs Tc because the buffer needs to be sufficiently large to absorb the converged flows. # of hops ≈ e2e latency/ Tc; for a given e2e latency, # of hops is limited by Tc. Problematic when Tc is small. Jitter bounded by 2* Tc, ok.
8	Support Multi-Mechanisms in single Domain and Multi-Domains	NA	Not relevant to CQF itself.

- CQF = IEEE 802.1Qch-2017
- DetNet Scaling Requirement = draft-ietf-detnet-scaling-requirements

03

TAS Evaluation

- **TAS (Time-Aware Shaper), refer to IEEE 802.1Qbv**

section	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	No	All terminals and network devices need to achieve nanosecond clock synchronization, to ensure that the GCL time of all outgoing ports is synchronized.
3.2	Support Large Single-hop Propagation Latency	Yes	Link delay is naturely cosidered during GCL calculation. GCL is independently installed on each node.
3.3	Accommodate the Higher Link Speed	Partial	More precise time control (smaller TimeInterval of GCL) is required.
3.4(1)	Be Scalable to the Large Number of Flows	Partial	GCL calculation for all flows in the control plane is NP-hard problem. Maintian queues per traffic class. TAS does not restrict the strategy of flow entering traffic class based queues.
3.4(2)	Tolerate High Utilization	Partial	The TimeInterval of GCL, with dedicated bandwidth, reserved for a TSN flow is exclusive. If that TSN flow does not send packets, the bandwidth is waste.
3.5	Prevent Flow Fluctuation from Disrupting Service	No	The re-calculation of the GCLs are more complicated and requires to update GCLs frequently.
3.6	Tolerate Failures of Links or Nodes and Topology Changes		Not related to queuing mechanisms directly.
3.7	Be Scalable to a Large Number of Hops with Complex Topology	Partial	More complexity of GCL calculations due to NP-hard problem. E2E queueing delay is negligible. E2E delay jitter is ultra-low.
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains		Not related to a single queuing mechanism directly.

CBS Evaluation

- **CBS (Credit-Based Shaper), refer to IEEE 802.1Qav**

section	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	Yes	Does not rely on time synchronization or frequency synchronization.
3.2	Support Large Single-hop Propagation Latency	Yes	Link delay does not affect the rate based shaping logic of CBS
3.3	Accommodate the Higher Link Speed	Yes	More buffer space is required to server more service bursts accordingly. It's almost impossible to cause burst storm by a single interfering frame.
3.4(1)	Be Scalable to the Large Number of Flows	Partial	Shaping of CBS is based on serveral traffic class for aggregated flows. May need re-shaping (ATS) to avoid burstiness cascade for each class [1].
3.4(2)	Tolerate High Utilization	Yes	Set the pre-configuration of bandwidth limit for each traffic class. BE flows can use the unused portion of the reserved bandwidth of TSN flows.
3.5	Prevent Flow Fluctuation from Disrupting Service	Yes	Each service flow of class A/B is permitted based on bandwidth reservation. The total amout of bandwidth reservation does not exceed the pre-configuration limit, and not exceed the worst latency.
3.6	Tolerate Failures of Links or Nodes and Topology Changes	N/A	Not related to queuing mechanisms directly.
3.7	Be Scalable to a Large Number of Hops with Complex Topology	Partial	On each node the queueing delay is over-estiamted, basically inversely proportional to the idle slope, due to non work-conserving behavior. Thus the E2E delay is large.
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains	N/A	Not related to a single queuing mechanism directly.

[1] Latency and Backlog Bounds in Time-Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping (https://ieeexplore.ieee.org/document/8493026/)

ATS Evaluation

- **ATS (Asynchronous Traffic Shaping), refer to IEEE 802.1Qcr**

section	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	Yes	Based on the ATS scheduler clocks,which is an implementation specific local system clock function. No need to achieve time synchronization.
3.2	Support Large Single-hop Propagation Latency	Yes	Link delay does not affect asynchronous traffic shaping on per hop
3.3	Accommodate the Higher Link Speed	Yes	More buffer space is required to server more service bursts accordingly.
3.4(1)	Be Scalable to the Large Number of Flows	Partial	ATS only maintain interleaved regulators (IR) per “inport + traffic class”. However, it need to maintain per flow states.
3.4(2)	Tolerate High Utilization	Yes	Can achieve high bandwidth utilization.
3.5	Prevent Flow Fluctuation from Disrupting Service	Partial	The cost of interleaved regulators (IR) maintained per hop is high. The problem of IR head-of-line blocking should be considered.
3.6	Tolerate Failures of Links or Nodes and Topology Changes		Not related to queuing mechanisms directly.
3.7	Be Scalable to a Large Number of Hops with Complex Topology	Partial	Need to consider flow aggregation strategies at intermediate nodes. End-to-end delay upper bound provided by ATS is larger, basically inversely proportional to the reserved bandwidth.
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains		Not related to a single queuing mechanism directly.

CQF Evaluation

- **CQF (Cyclic Queuing and Forwarding), refer to IEEE 802.1Qch**

sectionc	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	No	Rely on nanosecond clock synchronization across the entire network, where all network nodes align their cycle boundaries.
3.2	Support Large Single-hop Propagation Latency	No	Link delay must be much smaller than cycle duration and considered negligible, so 2-buffer mode can work. Otherwise, 3-buffer (or more) mode is needed.
3.3	Accommodate the Higher Link Speed	Partial	More buffer space is required for a specific length of cycle duration.
3.4(1)	Be Scalable to the Large Number of Flows	Partial	Transmission gates are associated with each aggregated queue, but the stream filtering and policing actions per stream should be maintained.
3.4(2)	Tolerate High Utilization	No	The cycle duration includes a time zone called dead time (DT) contributed by Output delay, Link delay, Frame preemption delay, Processing delay, which can not be used to send packets.
3.5	Prevent Flow Fluctuation from Disrupting Service	Partial	Requires corresponding flows setup algorithms to allocate resources appropriately among multiple flows to avoid transmission conflicts.
3.6	Tolerate Failures of Links or Nodes and Topology Changes		Not related to queuing mechanisms directly.
3.7	Be Scalable to a Large Number of Hops with Complex Topology	Partial	It is more difficult to select the cycle time. Need making trade-offs between end-to-end delay and cycle duration.
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains		Not related to a single queuing mechanism directly.