Enhanced DetNet Data Plane: Progress Report

Report on DetNet Open Working Meetings

IETF 118 (Prague): November 6, 2023

David Black

Tech Advisor (Transport)

DetNet: Open Working Meetings (4)

- Four (4) meetings since IETF 117 (San Francisco)
- Two-part agenda (for each meeting):
  1. Process-oriented topics - requirements, draft contents, evaluation structure, etc.
  2. Initial evaluations of new proposed scheduling/queuing mechanisms (7)
- Process-oriented outcomes:
  – Revisions to scaling requirements draft
  – Roughly common initial evaluation templates
- Initial evaluations of new proposed mechanisms against requirements (led to more revisions of scaling reqts. draft)
  - Evaluation slides included in this deck
    (for reference not for presentation @ IETF 118)

Open Working Meeting Plans?
(between IETF-118 and IETF-119)

- Need WG Discussion
  - Meetings have been productive
  - Initial evaluations have been produced
- Now what?
  - Goal: WG selection of mechanisms to standardize
  - How would/should open meetings help achieve that goal?

New Proposed Mechanisms: Initial Evaluation Slides

(From Open Working Meetings)

Description

- Initial evaluations of seven proposed new scheduling/queuing mechanisms, prepared by the authors

- Initial summary slide based on initial evaluations

00

C-SCORE Evaluation

B: Max Burst, L: Max Packet Length, r: Service rate,
H: number of Hops **of the flow**.
Lmax: Max Packet Length, R: Capacity **of a Link**.

C-SCORE E2E Latency bound: (B-L)/r + H*(L/r + Lmax/R)

section	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	Yes	Synch is not necessary. The time difference (TD) between nodes is added to the delay factor, which is robust to inaccuracy of measured TD.
3.2	Support Large Single-hop Propagation Latency	Yes	Propagation delay of a link is a factor of TD, is simply added to the delay factor and to the E2E latency bound. (It is neglected in the bound expression above.)
3.3	Accommodate the Higher Link Speed	Partial	Priority queue can be supported up to 600Gbps Ethernet with 2.5GHz clock ASIC (See next page). The throughput is independent of the queue length.
3.4(1)	Be Scalable to the Large Number of Flows	Yes	Independent of # of flows or Utilization level
3.4(2)	Tolerate High Utilization	Yes	Independent of # of flows or Utilization level
3.5 (now 3.6)	Prevent Flow Fluctuation - Tolerate Dynamic Flows Join/Leave - Burst accumulation	Yes	- Requires admission control & resource reservation (like all the other candidates) - Prevents burst accumulation
3.6 (now 3.5)	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	Yes	Independent of topology, but the E2E latency bound is a linear function of hop counts
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains		It copes well with other asynchronous solutions, such as TSN ATS, deadline-based forwarding, etc.

Deadline Evaluation

Requirement items	Evaluation	Notes
3.1. Tolerate Time Asynchrony	Yes	No full time synchronization needed, only need frequency sync(3.1.3).
3.2. Support Large Single-hop Propagation Latency	Yes	The eligibility arrival of flows is independent with the link propagation delay.
3.3. Accommodate the Higher Link Speed	Partial	The higher service rate, the more burst resource may provided by each delay level, and more buffer space is needed. And, extra instructions ....
3.4. Be Scalable to The Large Number of Flows and Tolerate High Utilization	Partial	Multiple delay levels, each with limited delay resources, can support lots of flows. The unused bandwidth of the high delay level can be used by the low levels or BE flows.
3.5. Tolerate Failures of Links or Nodes and Topology Changes	N/A	No relationship with queueing mechanism...
3.6. Prevent Flow Fluctuation	Yes	Flows are permitted based on the resources reservation of delay levels, and isolated from each other.
3.7. Be scalable to a Large Number of Hops with Complex Topology	Partial	More buffer may be needed when the latency compensation + in-time is used.
3.8 Support Multi-Mechanisms in Single Domain and Multi-Domains	N/A	No relationship with queueing mechanism...

Deadline-07 Evaluation (revised)

Requirement items	Evaluation	Notes
3.1. Tolerate Time Asynchrony	Yes	No full time synchronization needed, only need frequency sync(3.1.3).
3.2. Support Large Single-hop Propagation Latency	Yes	The eligibility arrival of flows is independent with the link propagation delay.
3.3. Accommodate the Higher Link Speed	Partial	The higher service rate, the more burst resource may provided by each delay level, and more buffer space is needed. And, extra instructions ....
3.4. Be Scalable to The Large Number of Flows and Tolerate High Utilization	Yes	Multiple delay levels, each with limited delay resources, can support lots of flows, without overprovision. Utilization may reach 100% link bandwidth. The unused bandwidth of the high delay level can be used by the low levels or BE flows.
3.5. Tolerate Failures of Links or Nodes and Topology Changes	N/A	No relationship with queueing mechanism...
3.6. Prevent Flow Fluctuation	Yes	Flows are permitted based on the resources reservation of delay levels, and isolated from each other.
3.7. Be scalable to a Large Number of Hops with Complex Topology	Yes	E2E latency is liner with hops , from ultra-low to low latency by multiple delay levels. E2E jitter is low by on-time mode.
3.8 Support Multi-Mechanisms in Single Domain and Multi-Domains	N/A	No relationship with queueing mechanism...

TCQF & CSQF evaluation

Sec.	P/PE	Requirements	Evaluation	Notes
3.1.1	PE	Async across TSN subdomains	Yes	TCQF defines async ingres. Same applicable to CSQF
3.1.2	P	Tolerate Clock Jitter/Wander	Yes	E.g.: with 4 cycles, max jitter/wander 1 cycle time. **Main benefit over (E)CQF** *ECQF can not do this: arrival time inaccuracy leads to wrong cycle assumption*
3.1.3	P	No Full Time Sync required	partial	Same as CQF, ECQF, TAS. **Main benefit of gLBF (which has YES) Partial: ca. 90% lower accuracy clock sync req. than with CQF/ECQF**
3.1.4	PE	Support for aperiodic flows	Yes	Via aperiodic burst shape/delay on ingress PE or overprovisioning.
3.2	P	Large Single-hop Propagation Latency	Yes	**Main benefit over TAS, CQF**. Same as ECQF.
3.2.1	P	Support single-hop propagation jitter	Yes	**Benefit over ECQF , CQF, TAS** (processing, RAW/retrans,length deviation)
3.3	P	Support Higher Link Speed	Yes	200km/100Gbps proven, can scale well beyond that. **Not proven for ECQF**
3.4(1)	P	Scalable to Large Number of Flows	Yes	No per-hop, per-flow state, read/write memory access requirement
3.4(2)	P	Tolerate High Utilization	Yes	**Solves CQF issue, equal/better than ECQF**: No dead times
3.5	P	Link/Node failures, Topo Changes	Yes	Can support Segment Routing and hence all re-route, path-fixing options **(TCQF, CSQF). CSQF**: even defined solely for SR!
3.6	P	Prevent Flow Fluctuation from Disrupting Service	Yes	Like CQF/ECQF: No burst accumulation/jitter-increase due to per-hop cycle based reshapeper-hop reshaping.
3.7	P	Be Scalable to a Large Number of Hops with Complex Topology	Good (TCQF) Best (CSQF)	Assuming flow interleaving on edge, like CQF/ECQF **CSQF more flexible** than TSQF, CQF/ECQF
3.8(1)	P	Support tight jitter/sync-control loops	Yes	Network size independent hop-by-hop/end-to-end jittter ~ O(0)

gLBF evaluation **(vs. CQF/ECQF/gLBF highlighted)**

Sec.	P/PE	Requirements	Evaluation	Notes
3.1.1	PE	Async across TSN subdomains	Yes	Naturally async, no special ingress function needed
3.1.2	P	Tolerate Clock Jitter/Wander	Yes	*Max per-hop jitter/wander just needs to be known/over-estimated during config*
3.1.3	P	Tolerate Time Asynchrony	**Yes**	**Main benefit over CQF / ECQF / TCQF / CSQF**
3.1.4	P/PE	Support for aperiodic flows	Yes	Directly via P, or via additional ingress PE timed gates to increase utilization at lower end-to-end latency
3.2	P	Large Single-hop Propagation Latency	Yes	*Max hop propagation latency to be known/over-estimated during net. config.*
3.2.1	P	Support single-hop propagation jitter	Yes	*Requires clock sync only across jittery link – not other links*
3.3	P	Support Higher Link Speed	TBD	Not yet proven in high-speed ASIC implementation. Target timed FIFO implementation model suggested for next-gen high-speed, low-cost.Impl.
3.4(1)	P	Scalable to Large Number of Flows	Yes	No per-hop, per-flow state, read/write memory access requirement
3.4(2)	P	Tolerate High Utilization	Yes	No dead times
3.5	P	Link/Node failures, Topo Changes	Yes	Can support all Segment Routing re-route/path-control options.
3.6	P	Prevent Flow Fluctuation from Disrupting Service	Yes	No burst accumulation/jitter-increase due to per-hop latency based re-shaping.
3.7	P	Be Scalable to a Large Number of Hops with Complex Topology	Yes/TBD	Flow interleaving TBD
3.8(1)	P	Support tight jitter/sync-control loops	Yes	Network size independent hop-by-hop/end-to-end jittter ~ O(0)

TQF Evaluation

Requirement items	Evaluation	Notes
3.1. Tolerate Time Asynchrony	Yes	No full time synchronization needed, but need frequency sync(3.1.3).
3.2. Support Large Single-hop Propagation Latency	Yes	The detection of timeslot mapping covers link propagation delay.
3.3. Accommodate the Higher Link Speed	Partial	The higher service rate, the more buffer needed for the same timeslot length.
3.4. Be Scalable to The Large Number of Flows and Tolerate High Utilization	Partial	Calculating paths for as many flows as possible is an NP-hard problem. The unused bandwidth of the timeslot can be used by best-effot flows.
3.5. Tolerate Failures of Links or Nodes and Topology Changes	N/A	No relationship with queueing mechanism...
3.6. Prevent Flow Fluctuation	Yes	Flows are permitted based on timeslot reservation, isolated from each other through timeslots.
3.7. Be scalable to a Large Number of Hops with Complex Topology	Partial	Calculating TQF paths for all services is NP-hard problem, related to hops count.
3.8 Support Multi-Mechanisms in Single Domain and Multi-Domains	N/A	No relationship with queueing mechanism...

TQF-04 Evaluation (revised)

Requirement items	Evaluation	Notes
3.1. Tolerate Time Asynchrony	Yes	No full time synchronization needed, but need frequency sync(3.1.3).
3.2. Support Large Single-hop Propagation Latency	Yes	The detection of timeslot mapping covers link propagation delay.
3.3. Accommodate the Higher Link Speed	Partial	The higher service rate, the more buffer needed for the same timeslot length.
3.4. Be Scalable to The Large Number of Flows and Tolerate High Utilization	Yes	Multiple OPL instance, each for a set of serivce flows, without overprovision. Utilization may reach 100% link bandwidth. The unused bandwidth of the timeslot can be used by best-effot flows. Calculating paths is NP-hard.
3.5. Tolerate Failures of Links or Nodes and Topology Changes	N/A	No relationship with queueing mechanism...
3.6. Prevent Flow Fluctuation	Yes	Flows are permitted based on timeslot reservation, isolated from each other through timeslots.
3.7. Be scalable to a Large Number of Hops with Complex Topology	Yes	E2E latency is liner with hops , from ultra-low to low latency by multiple OPL. E2E jitter is low by on-time mode. Calculating paths may be NP-hard.
3.8 Support Multi-Mechanisms in Single Domain and Multi-Domains	N/A	No relationship with queueing mechanism...

0n

E2E delay bounds via queue resizing

- **Enforcing end-to-end delay bounds via queue resizing**

section	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	Yes	Delay commitments by on path nodes are evaluated locally, and do not require clock synchronization.
3.2	Support Large Single-hop Propagation Latency	Yes	Propagation delay is considered as a fixed quantity that is taken as part of the overall node delay in the E2E bound calculation.
3.3	Accommodate the Higher Link Speed	Yes	Our mechanism acts on the delay spent by packets in queues, and is independent from link speed.
3.4(1)	Be Scalable to the Large Number of Flows	Partial	The capacity of our mechanism to handle new flows depends on the number of queues in nodes, the granularity of queue reservation and the number of queues.
3.4(2)	Tolerate High Utilization	No	Our mechanism is designed so peak capacity are reserved to enforce end to end delay. This is not compatible with high utilization.
3.5	Prevent Flow Fluctuation from Disrupting Service	Yes	As soon as the flow fluctuates under the capacity reserved in on path nodes' queues, our system tolerates flow fluctuation..
3.6	Tolerate Failures of Links or Nodes and Topology Changes	Needs more work	In the draft's current version, a failure of a node would require a re-negociation of the queue capacity reservations along the alternate path. This can be done but a/ it will not be instantaneous and b/ the failover mechanism needs to be described.
3.7	Be Scalable to a Large Number of Hops with Complex Topology	Partial	While we don't see theoretical limitations, a large scale simulation evaluation is necessary to assess this property.
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains	Partial	In the described mechanism, we mandate that the queueing system is able to accept reservations in terms of capacity allocated to flows, and that the service time of queues is deterministic.

03

Initial Summary v2 (David Black)

- **C-SCORE/Deadline/TCQF/CSQF/gLBF/TQF**

section	Requirements	Evaluation	Notes
3.1	Tolerate Time Asynchrony	Yes – 5, Partial – 2	Partial: TCQF, CSQF (Partial - within domain. Yes - async across domains & tolerates jitter)
3.2	Support Large Single-hop Propagation Latency	Yes – 7	
3.3	Accommodate the Higher Link Speed	Yes – 3, Partial – 3, TBD – 1	Yes: TCQF, CSQF, queue resizing Partial: C-SCORE, Deadline, TQF TBD: gLBF
3.4(1)	Be Scalable to the Large Number of Flows	Yes – 6, Partial – 1	Partial: queue resizing
3.4(2)	Tolerate High Utilization	Yes – 6, No – 1	No: queue resizing (design non-goal)
3.5 (now 3.6)	Prevent Flow Fluctuation from Disrupting Service	Yes – 7	
3.6 (now 3.5)	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	Yes – 5, Partial – 2	Yes: C-SCORE, CSQF, gLBF (flow interleaving TBD), deadline, TQF Partial: TCQF, queue resizing
3.8	Support Multi-Mechanisms in Single Domain and Multi-Domains		Not related to a single queuing mechanism directly.

THANK YOU