Cyclic Queuing and Forwarding for DetNet IP and MPLS Data Plane (TCQF)

...and BIER-TE...

draft-eckert-detnet-tcqf-01

Toerless Eckert, Futurewei USA (tte@cs.fau.de)

Stewart Bryant, University of Surrey ICS (s.bryant@surrey.ac.uk)

Andy Malis (agmalis@gmail.com)

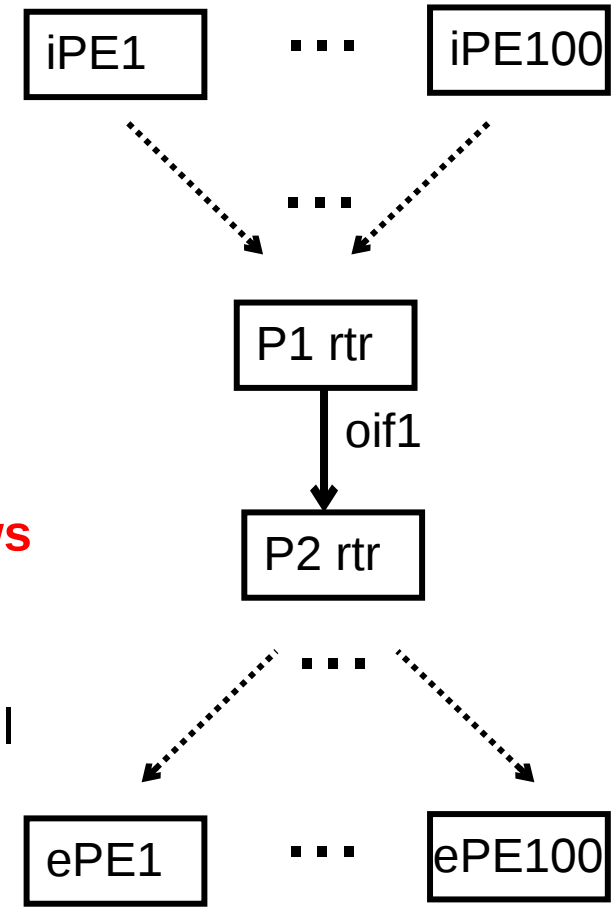
Guangpeng Li <liguangpeng@huawei.com>

IETF DETNET WG, IETF115, 11/07/2022, rev 1.1

Scalability in large DetNets

Realistic reference worst case scenario in large-scale DetNet

- Assume we aggregate all DetNet traffic from one ingres iPEj to one egres ePEk into one aggregate DetNet flow.
  - Most edge-aggregation we can do
  - j=1...100, k=1...100
- Total # flows: j * k = 100 * 100 = 10,000
- These flows may all go through one (core PE) interface
  - oif1 (output interface 1) on P1 in example network



RFC2211 (IntServ): **per-flow shaping for 10,000 flows**

IEEE ATS (Async Traffic Shape): **interleaved regulators for 10,000 flows**

Every time rate and/or burst-size of any of these 10,000 flows changes (because one of its member flow changes):

- Both IntServ and ATS require **signaling of new flow parameters** to all routers affected (P1, P2, ...) – *(limited optimizations possible).*

TCQF: **3 ... 5 cyclic queues** on P1 oif1.

- **No changes in any P router configuration** when member flows change!

Scalable traffic steering support

- DetNet flows need resource reservation along path
  - bandwidth/buffers – *shape of traffic depends on queuing mechanism*
- Result: Standard IGP (re-routing) insufficient
  - *No resources reserved (immediately) on re-routed path*
- Standard DetNet expectations
  - Explicit route for DetNet flow on every hop (IP 5/6-tuple flow, MPLS label path with RSVP-TE)
  - Known scalability, performance issues (RSVP-TE).
  - Per-flow, per-hop static routes in IP less well explored (IMHO: worse !)
- Solution: “source-routing”: Per-hop, per-flow stateless strict-steering options:
  - SR-MPLS, SRv6 (CRH extension header), BIER-TE (multicast)
- Without additional flow-id header, stateless solutions could:
  - (maybe) identify one iPE->ePE DetNet flow from IP/MPLS header, but:
  - Would re-introduce per-iPE/ePE flow, per-hop state that needs to be updated in forarding plane
- With stateless steering: One really wants to use a solution like TCQF not requiring to introduce per-hop, per-flow state for queuing.
  - TCQF is “natural” match for source-routing in forwarding.

Overview

CQF with packet-tagging of cycle to enable large-scale detnets

Benefits from CQF

- Scalable, low-cost hardware implementable – at > 100 Gbps interfaces routers
  - Candidate target: implementable on all programmable forwarding planes supporting CQF
- Bounded Latency with extremely simple latency calculus model
- Tightly bounded jitter – for all existing control-loop applications (industrial etc.)
  - Removes need for clock synchronization on constrained DetNet USER-devices

CQF problems solved with TCQF

- Arbitrary (wide-area network) link propagation latency / jitter
- Reduced clock synchronization accuracy requirement

Changes over CQF

- Use multiple buffers
- Use existing packet header field (MPLS TC / IP/IPv6 DSCP) as cycle-tag
- Replace arrival time based cycle use with packet header tag cycle mapping

~~Can cook coffee~~

Changes since IETF114

draft-eckert-detnet-mpls-tc-tcqf-03 -> draft-eckert-detnet-tcqf-00/01

- IETF114: draft-eckert-detnet-mpls-tc-tcqf-03
  - Described only MPLS support (TC tagging)
    - Authors felt IP support with standardized DSCPs would require different IETF process (TSVWG responsibility ?)
  - @IETF114, David Black reminded us
    - DetNet is controlled domain (single operator or coordinated operations)
    - “private” DSCP space has same applicability as MPLS TC here
- Resulted in adding text for IP/IPv6 with DSCP tagging
  - Simple text enhancements: add DSCP, data model, pseudo-code
  - xxxx11 "EXP/LU" Codepoint space according to [RFC2474], Section 6
  - Aka: with IP up to 16 DSCP
    - not seen significant pre-existing use in my deployment experience (?!)
- Text quality improvements, new co-author

Existing “fields” (DSCP/TC) vs. New header

- Goal of this document: First-to-adopt solution
  - Most easily first standardized/adopted large-scale-detnet bounded latency
    - Existing queuing mechanism (CQF), just with tagging instead of clock-sync
    - Existing header fields/private-semantic for tagging – no new packet header stds. work
- Longer term work !
  - DetNet QoS requiring new header
    - IP/IPv6 (extension header): PREOF (sequence-number, ?flow-id?)
    - Alternative Marking ? Stream-ID ?
    - TCQF (especially MPLS) – more tag values
    - Dampers – various mechanisms (with / without cycle buffers): per-hop-latency value
    - Heuristic bounded latency parameters (per-hop deadline – Jakov Stein) ?
    - Steering together with latency (aka: list of (next-hop, {cycle,deadline,priority,...))
  - Same header across IP and MPLS would be ideal ?!
    - IPv6 extension header: routing? – IMHO better than HbH (could do QoS AND steer)
    - MPLS Design Team work: encap for header shreable with IPv6 ??
  - Good solution could be big step for DetNet, but IMHO not fast standardized.

Summary, Next Steps

- Complete “scalable” / large-detnet bounded QoS with low jitter
- Most easily adoptable option (?!)
  - For both MPLS and IP/IPv6 networks
  - Can leverage experience / hardware from CQF / gated queues
  - PoC validation via 100Gbps WAN network
- Authors / supporters want to ask WG for adoption

- Design-team / more regular meeting / work to revisit / collaborate across all posted bounded latency drafts ?
  - To validate / work out proposed solution strategy ? (short / long term etc... ?)
  - Would be happy to get one started.
- Thank you!

Backup Slide(s)

Reminder: TCQF High Level

With 3 cycles , arbitrary latency links can be supported

Every node has cycle-mapping table from prior hop (calculated by controller)

Clock accuracy: “Maximum Time Interval Error” (MTIE) < 90% cycle time

With 3 or more cycles, additional inaccuracies can be compensated

e.g.: link propagation variation (jitter) and/or higher MTIE