Deadline based Forwarding

draft-peng-detnet-deadline-based-forwarding-03

Shaofu Peng ZTE
Bin Tan ZTE
Peng Liu China Mobile

IETF-115 November 2022, London

Updates

- Update abbreviations for ease of reading.
  - queue has three attrs: CT (Count-down Time), AT (Authorization Time), TI (rotation Timer Interval).
- Packets queuing rules:
  - queue.CT <= packet.Q < queue.CT + AT
  - or queue.CT <= packet.Q - k*AT < queue.CT + AT, 0<=k<=1
- Queue allocation rules.
- Trrafic regulation and orchestrating on ingress, re-shaping on transit, and schedulability conditions.
- Delay compensation, and deployment considerations

Motivations

- To find a potential queuing mechanism for large scale networks.

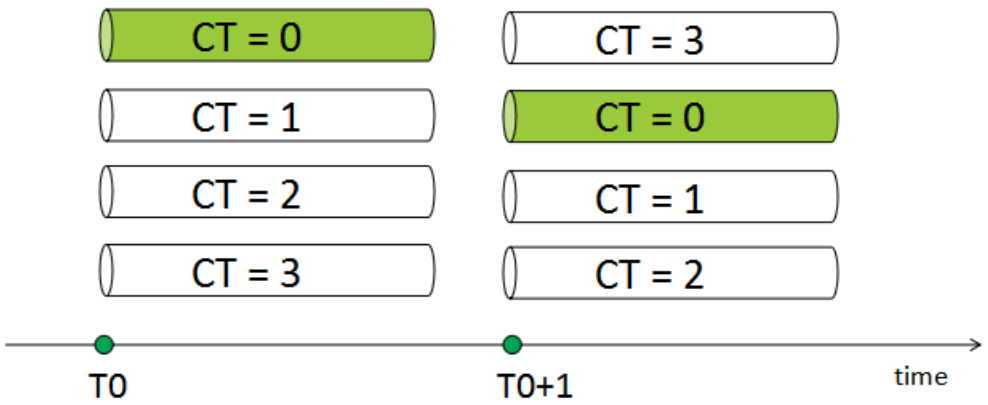
- Issues of existing mechanisms for large scale networks:
  - TSN CBS and ATS come with a high latency variance, as the minimum latency is not affected by them.
  - TSN CQF is quite challenging because it requires time synchronization.
  - The widely used priority based queuing scheme may give better average latency, but with worst case latency.

- This document propose a variants of EDF (Earliest Deadline Forwarding) scheduling, to dynamically rotate the priority of each aggregated FIFO queue and uniformly provide bounded delay/jitter.

Deadline Queue Attributes

- Deadline queues have three attributes.
  - **CT (Count-down Time)**, indicates the countdown waiting for scheduling.
  - **TI (rotation timer interael)**, indicates the decreasing step of CT.
  - **AT (Authorization Time)**, indicates the continuous sending duration when the queue is scheduled.
    - AT = N * TI, where N >= 1
- There are two sets of isolated queues: in-time queues, on-time queues.
  - The queue has highest priority if its CT = 0, indespite of in-time or on-time mode.
  - The in-time queue has normal priority if its CT > 0 (i.e., work-conserving), and the on-time queue has prohibition priority if its CT > 0 (i.e., non-work-conserving).

e.g, 4 deadline queues:



queue length = C * AT - M
where, C = port bandwidth
M = maximum interference size

Each CT queue can also be further split into multiple sub-queues with different planned residence time.

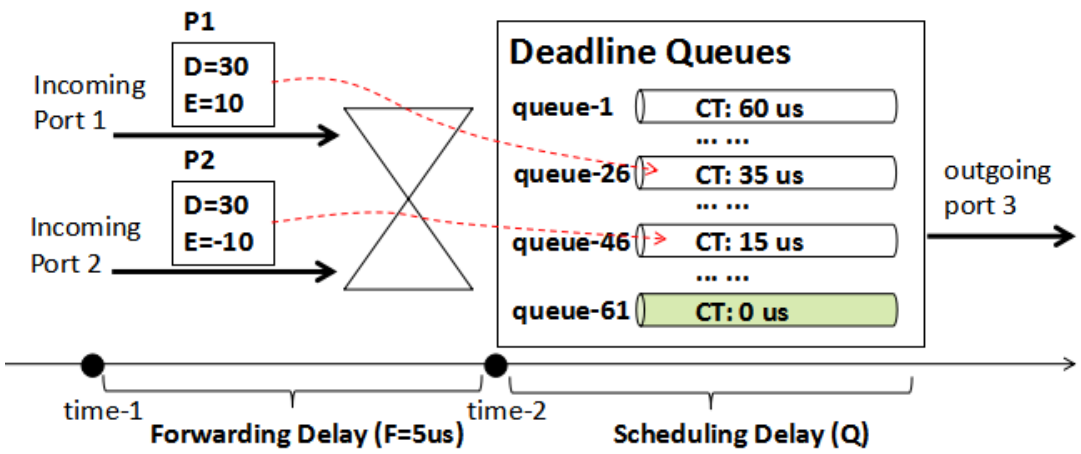
Packets Queuing Rules

- First, get the deadline information of the packet, including:
  - **Planned Residence Time (D)** for the current node
  - **Accumulated Residence Variation (E)** by all upstream nodes.
  - **Forwarding Delay (F)**, from arriving the node to arriving the scheduler.

  can be get from: packet itself, or FIB entry, or policy entry, etc.
    - Please refer to draft-peng-6man-deadline-option.

- Second, put the packet in the specific deadline queue.
  - Allowable Queuing Delay **(Q) = D + E - F**
  - meet CT <= Q < CT+AT

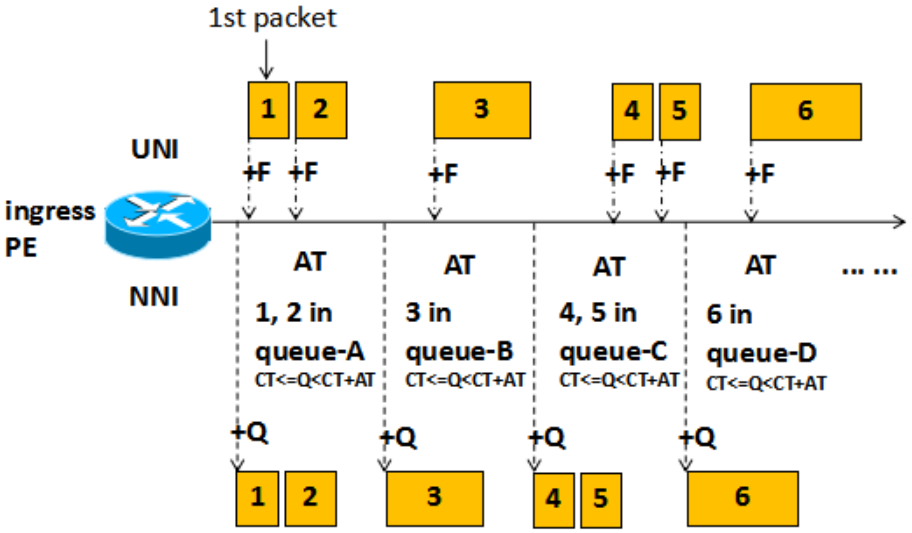
packet 1 to queue-26
packet 2 to queue-46



Traffic Regulation and Orchestrating

- Regulation and orchestrating per flow on ingress.
  - $\sum_{i=1}^{N} \rho_i \leq C$, where ρ_i is reserved bandwidth for service i with leaky bucket constraint function $A_i(t) = \sigma_i + \rho_i(t)$, and C is port bandwidth.
  - Stability condition: $\sum_{i=1}^{N} A_i(t)/t \leq C$
  - Orchestrating:

    Time1:
    packet 1,2 in queue-A with CT<=Q
    Time2:
    packet 3 in queue-B with CT<=Q;
    Time3:
    packet 4,5 in queue-C with CT<=Q
    Time4:
    packet 6 in queue-D with CT<=Q;

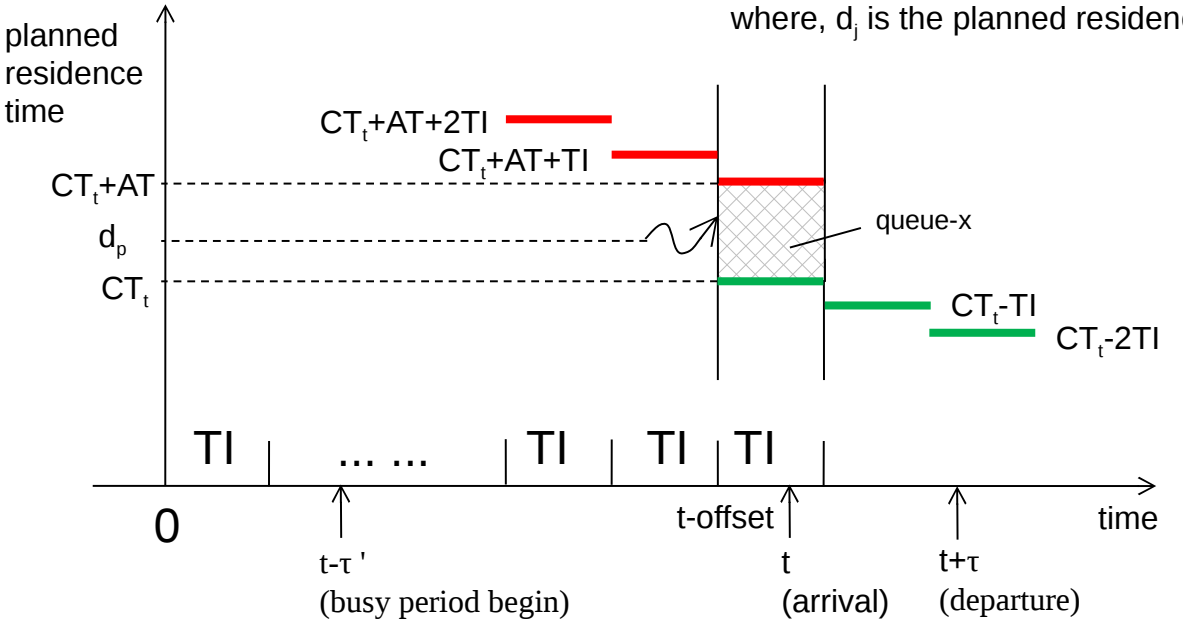


- to get schedulability conditions, re-shaping per flow or per aggregate flows on transit still needed.
  - TBD: re-shaping does not increate worst-case latency bounds for deadline mechanism ?

- then, schedulability condition:

$$\frac{A_1(t - d_1) + \sum_{d_j \geq d_2} A_j(t + AT - d_j)}{C} \leq t$$

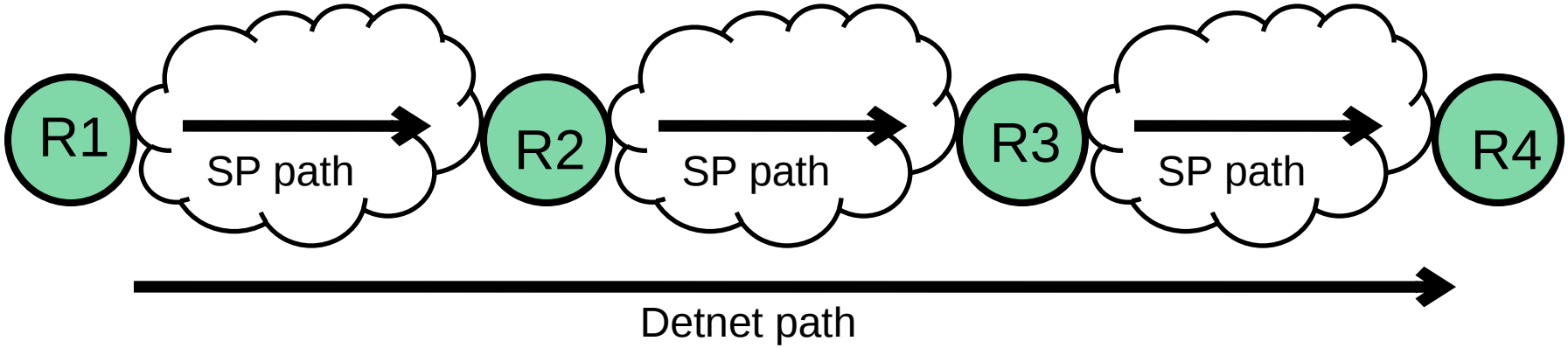
where, d_j is the planned residence time for service j.



- so that the control plane can provide several level of planned residence time according to the preset traffic arrival constraint function of all deterministic services (T-SPEC), or vice versa.

Partially Upgraded Scenarios

A DetNet path consists of several upgraded border nodes (R1, R2, R3, R4) and traditional SP (strict priority) paths in the domain.
SP path should contain fewer hops, with the worst-case latency per hop given by [SP-LATENCY].



- The legacy device in the SP path does not support the deadline scheduling, and may or not measure and insert the deadline information into packets.
  - i.e., deadline-aware or deadline-unaware.
- In-time mode, compared with SP, can schedule packets by deadline information other than traffic class.
  - That means, a packet will not always experience the worst interference delay on every hop, i.e., do NOT overestimate the worst-case delay on each hop like SP.
- The effect of on-time mode is EDF+damper, the packets will be delayed on the border nodes to achieve the jitter target.

Evaluation of the Mechanism

- Cost:
  - Time synchronization is not required between network nodes. Operate based on local offset time.
  - States per flow or flow aggregate may need to be maintained.
- Deployment:
  - Packet multiplexing with several aggregated deadline queues.
  - Each node can independently set the authorization time of the deadline queues, based on self port bandwidth.
  - Support partial upgrade.
- Scalability:
  - A single set of deadline queues supports multiple levels of residence time.
  - Queues with higher Max CT can be created incrementally according to service needs.
- Performance:
  - Good jitter control, just a single authorization time.

Next step

- Any questions and comments ?

Thank you!