Deadline based Forwarding

draft-peng-detnet-deadline-based-forwarding-04

Shaofu Peng	ZTE
Bin Tan	ZTE
Peng Liu	China Mobile

interim 2022.12.12

Motivations

- To find a potential queueing mechanism suitable for large scale networks to provide deterministic QoS.

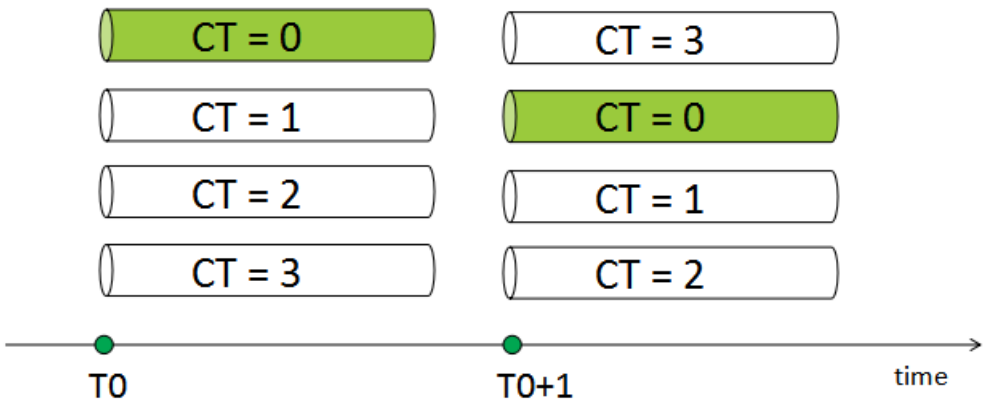
- Some candidate queueing mechanisms:
  - TSN **CBS** and **ATS** come with a high latency variance, as the minimum latency is not affected by them.
  - TSN **CQF** is quite challenging because it requires time synchronization.
  - Traditional strict priority (**SP**) based queueing scheme may give better average latency, but with worst case latency related to hop counts.
  - **EDF** stores packets according to packet's deadline, with costly lookup/insertion operation.

- This document propose **a variants of the existing EDF scheduling**, with **latency compensation**, which 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
- Two sets of isolated queues: **in-time** queues, **on-time** queues.
  - The queue has highest priority if CT = 0, indespite of in-time or on-time mode.
  - The in-time queue has normal priority if CT > 0 (i.e., work-conserving), and the on-time queue has prohibition priority if 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

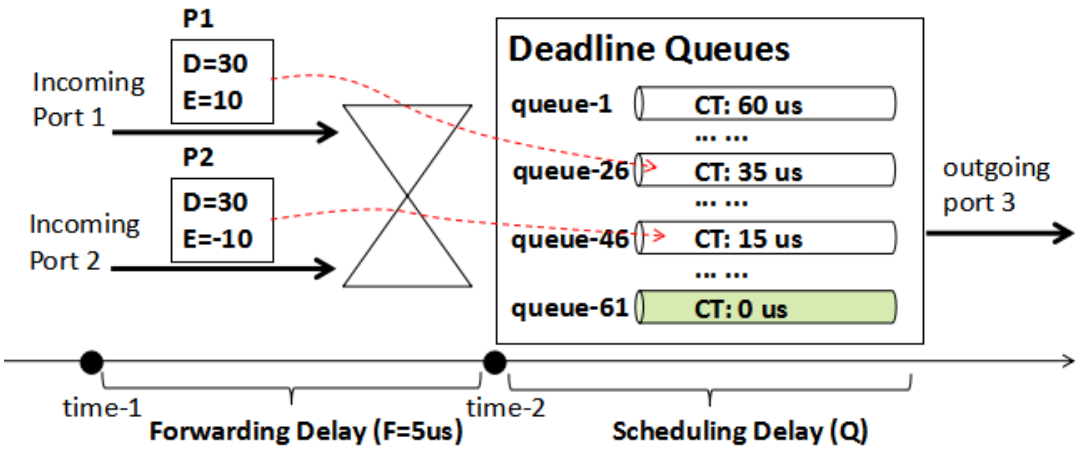
Each CT queue can also be further split into multiple sub-queues with different E/D.

Packets queueing 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.

- Second, put the packet in the specific deadline queue.
  - Allowable queueing Delay (Q) = D + E - F
  - map to a queue with [CT, CT+AT), and meet **CT <= Q-k*AT < CT+AT**

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



Schedulability Condition

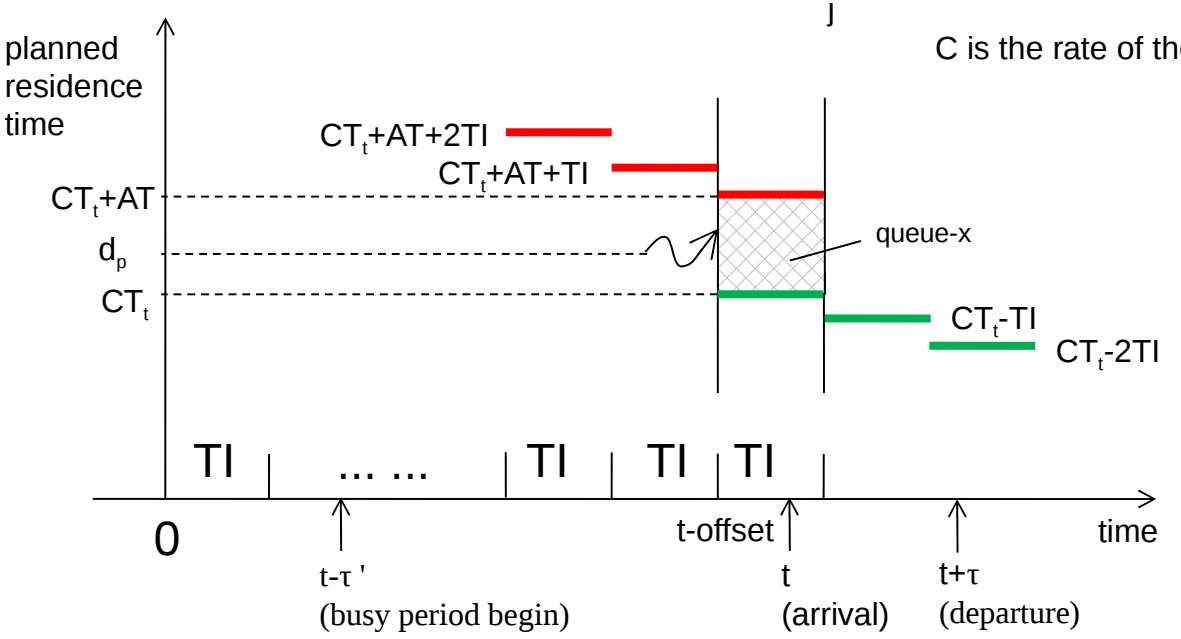
- condition:

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

where, d_j is the planned residence time of class j

Aj is the traffic constraint function of class j

C is the rate of the scheduler.

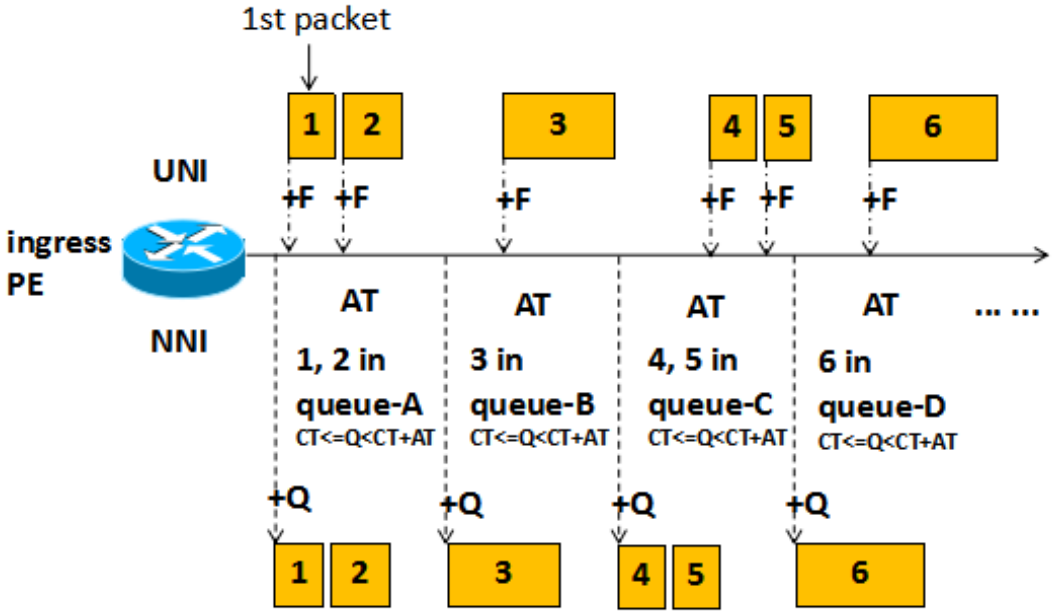


- Take a specific planned residence time as a specific **class**.
- The control plane can provide several level of planned residence time according to the preset traffic constraint function of all classes, or vice versa.
- **No re-shaping** on intermediate nodes, since deadline queue itself based on latency compensation plays the role of shaping queue.

Ingress Regulation and Orchestrating

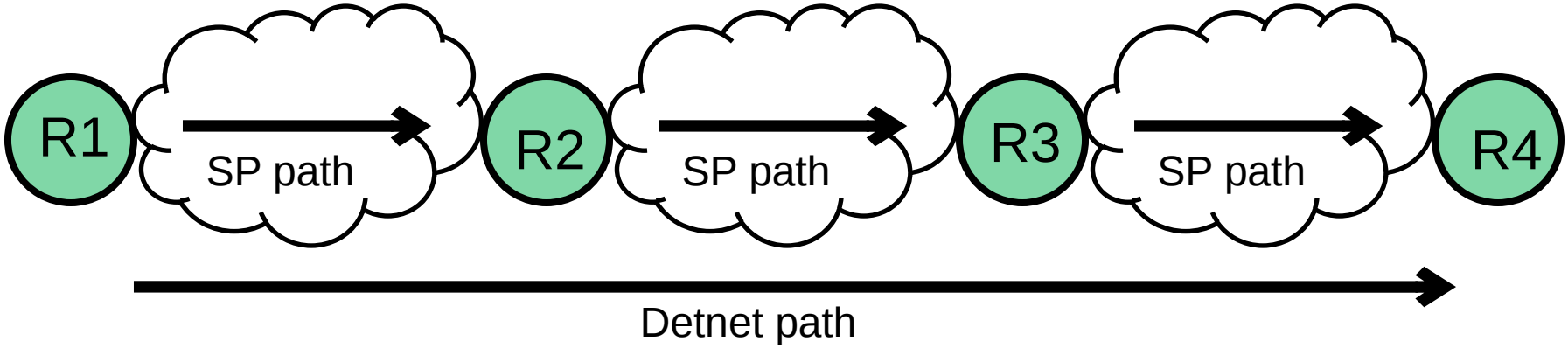
- Regulation and orchestrating per flow on ingress.
  - Regulation on imcoming port:
    - conform to traffic constraint of service such as leaky bucket.
  - Orchestrating on output port:

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;



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, to get a bounded worst-case latency .



- The legacy device in the SP path may or not measure and insert the deadline information into packets.
- In-time mode, compared with SP, can schedule packets by deadline information other than traffic class.
  - Avoid to overestimate the worst-case delay on each hop like SP or EF FIFO.
- The effect of on-time mode is EDF+damper, the packets will be held on the border nodes to achieve the jitter target.

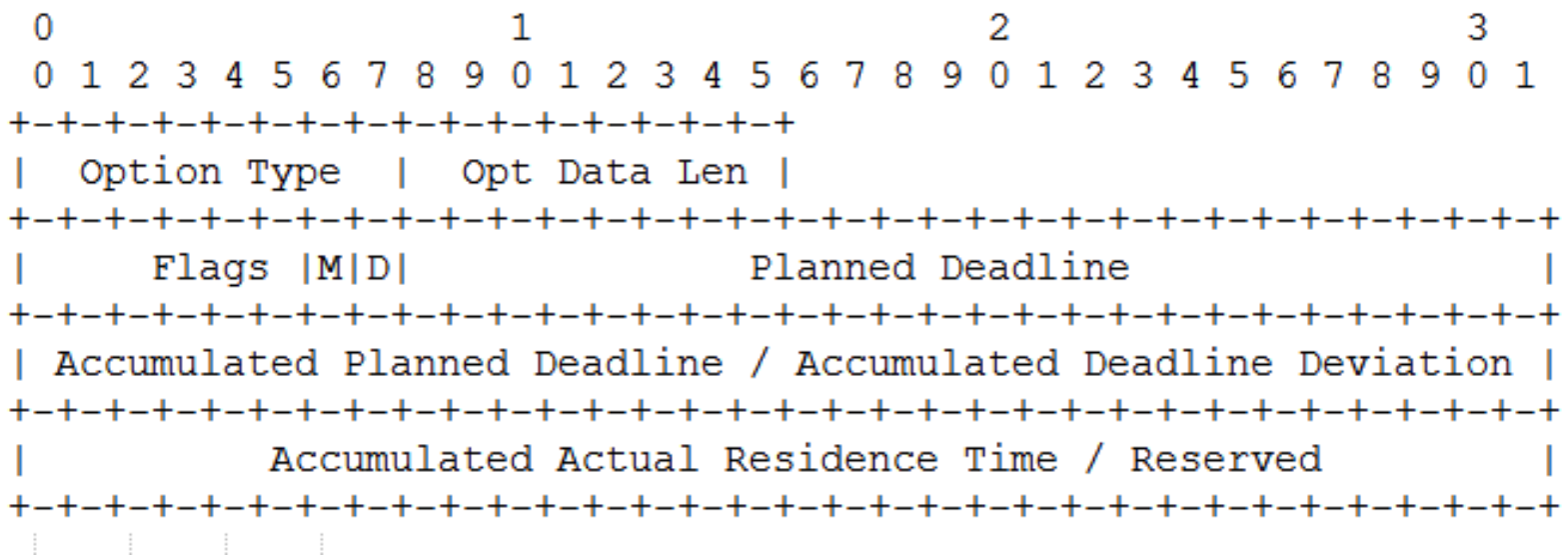
Deployment Considerations

- Get a balance to allocate delay resource (planned residence time) to in-time service and on-time service.
  - If E2E latency requirement is low, provide a small planned residence time.
  - Not suggest a large planned residence time to be used for on-time service, otherwise need more buffers.

- Suggest:
  - **Loose on-time** or in-time scheduling on the transit nodes to meet schedulability condition.
  - **Strict on-time** scheduling on the egress node to achieve jitter target. MAX_CT configured on the egress can be larger than the transit nodes.

Encapsulations

- A new IPv6 HBH option: deadline option, is defined



Flag-M: in-time or on-time scheduling mode

Flag-D:

if 0, the fields below contain Accumulated Planned Deadline and Accumulated Actual Resience Time,
otherwise, contain Accumulated Deadline Deviation and Reserved.

Evaluation of the Mechanism

- Cost:
  - Time synchronization is not required between network nodes. Operate based on local offset time.
  - No re-shaping on the intermediate nodes.
- 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 new service needs.
- Performance:
  - Jitter is a single authorization time.

Next step

- Any questions and comments ?
- Anyone interested in cooperation?

Thank you!