Timeslot Queueing and Forwarding (TQF)

draft-peng-detnet-packet-timeslot-mechanism-06

Shaofu Peng	ZTE
Peng Liu	China Mobile
Kashinath. Basu	Oxford Brookes University
Aihua Liu	ZTE
Dong Yang	Beijing Jiaotong University
Guoyu Peng	Beijing University of Posts and Telecommunications

IETF-119 Mar 2024

Updates (04 -> 06)

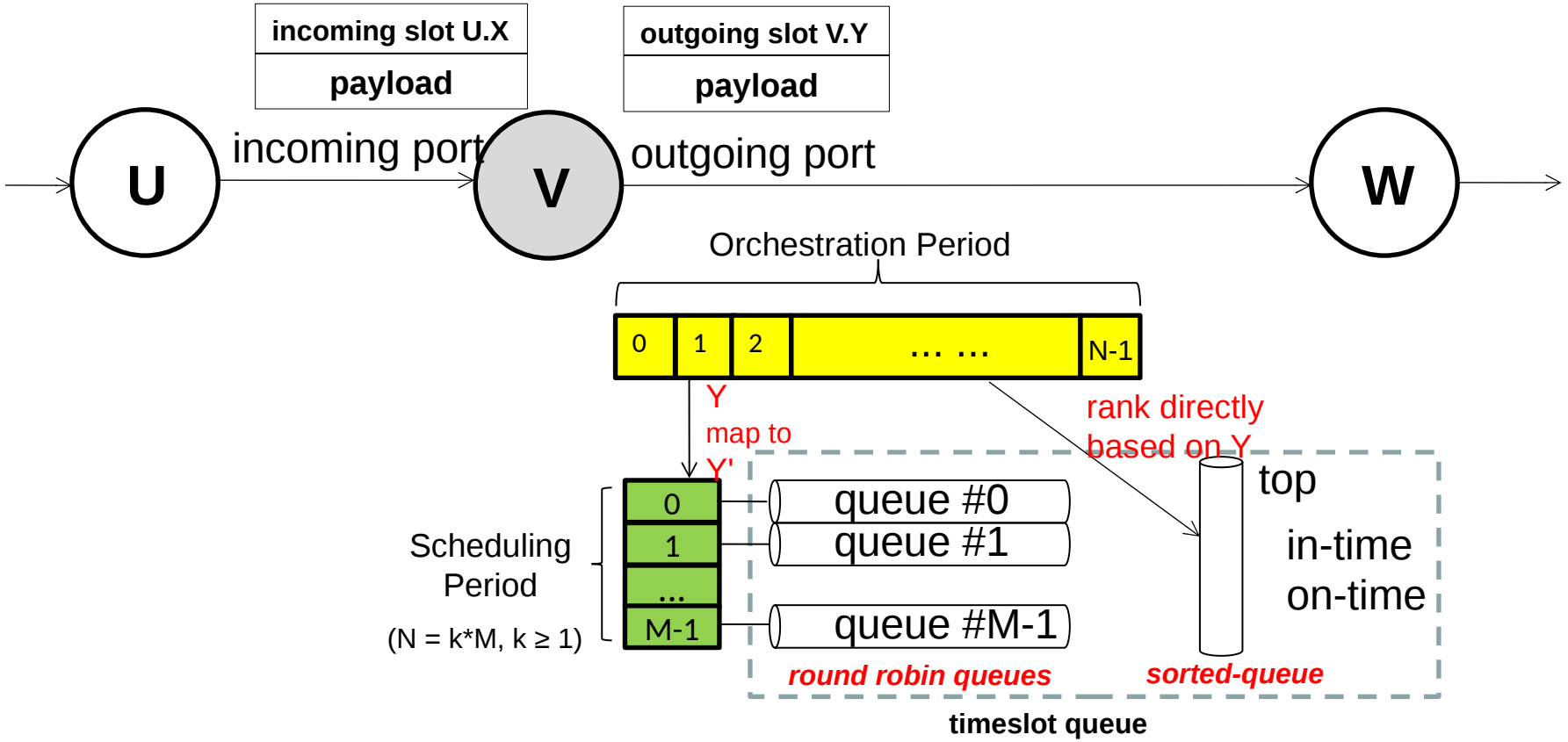
- Clarification of some mapping relationship: BTM, BOM, FTM.
- Runtime policing delay jitter elimination.
- Common topology & service TSpec/RSpec example.
- Taxonomy considerations.

Motivations

- Most scheduling mechanisms have a contradiction between latency performance and service scale. **Flow interleaving** can significantly improve service scale.
  - **TSN TAS** introduces a flow interleaving method based on gate control list (GCL) rotation cycle (i.e., gating cycle) in Ethernet LAN. It calculate when the packets of the Scheduled Traffic arrive at a certain node, then turn on the green light for it. However, it need time synchronization and has **scalability issues** on GCL calculation, updation and installation.

- To meet the large scaling requirements, this document enhance TAS to introduce timeslot resources to layer-3 and related timeslot scheduling on DetNet Data Plane. We call it **TQF** (Timeslot Queueing and Forwarding) mechanism.

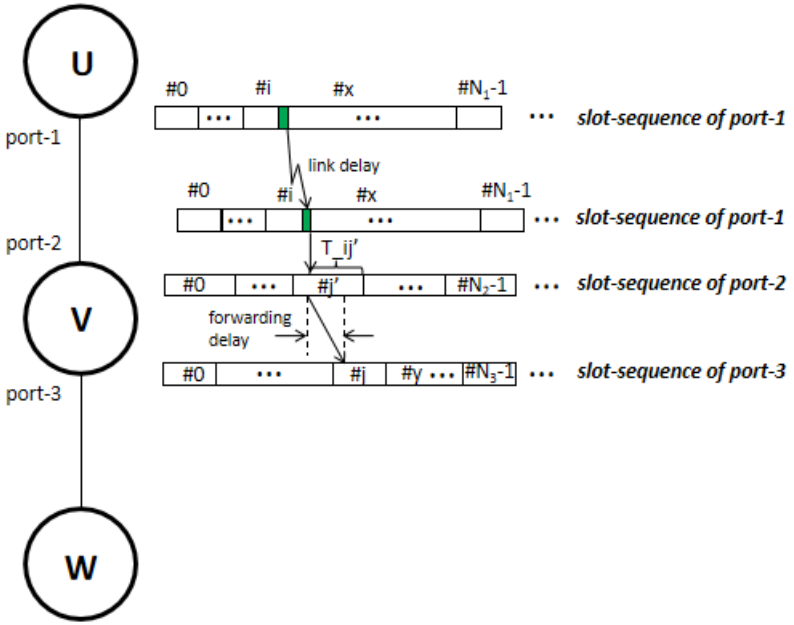
Overview of TQF Mechanism

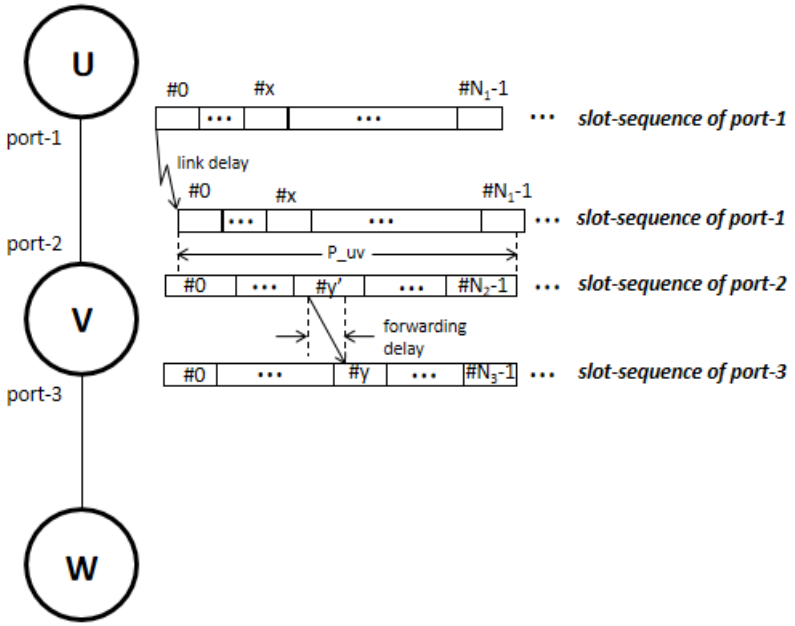


- Decouple the Orchestration Period from the Scheduling Period.
  - **Orchestration Period** includes timeslot resources, euqals to LCM of service burst interval;
  - **Scheduling Period** matches the actual capacity of the device, requiring only a few RR queues or a single sorted-queue.
    - All RR queues have equivalent binary semantic Open/Closed, no other semantics.
- Path calculation based on timeslot resource reservation. Obtain a flexible mapping relationship between the incoming and outgoing timeslots on each node.

BTM, BOM and FTM

- For the pair <local node's outgoing port, downstream node's incoming port>, detect **BTM** (base timeslot mapping), i.e., outgoing timeslot -> ongoing timeslot, or **BOM** (base ochestration-period mapping), i.e, OP phase difference.

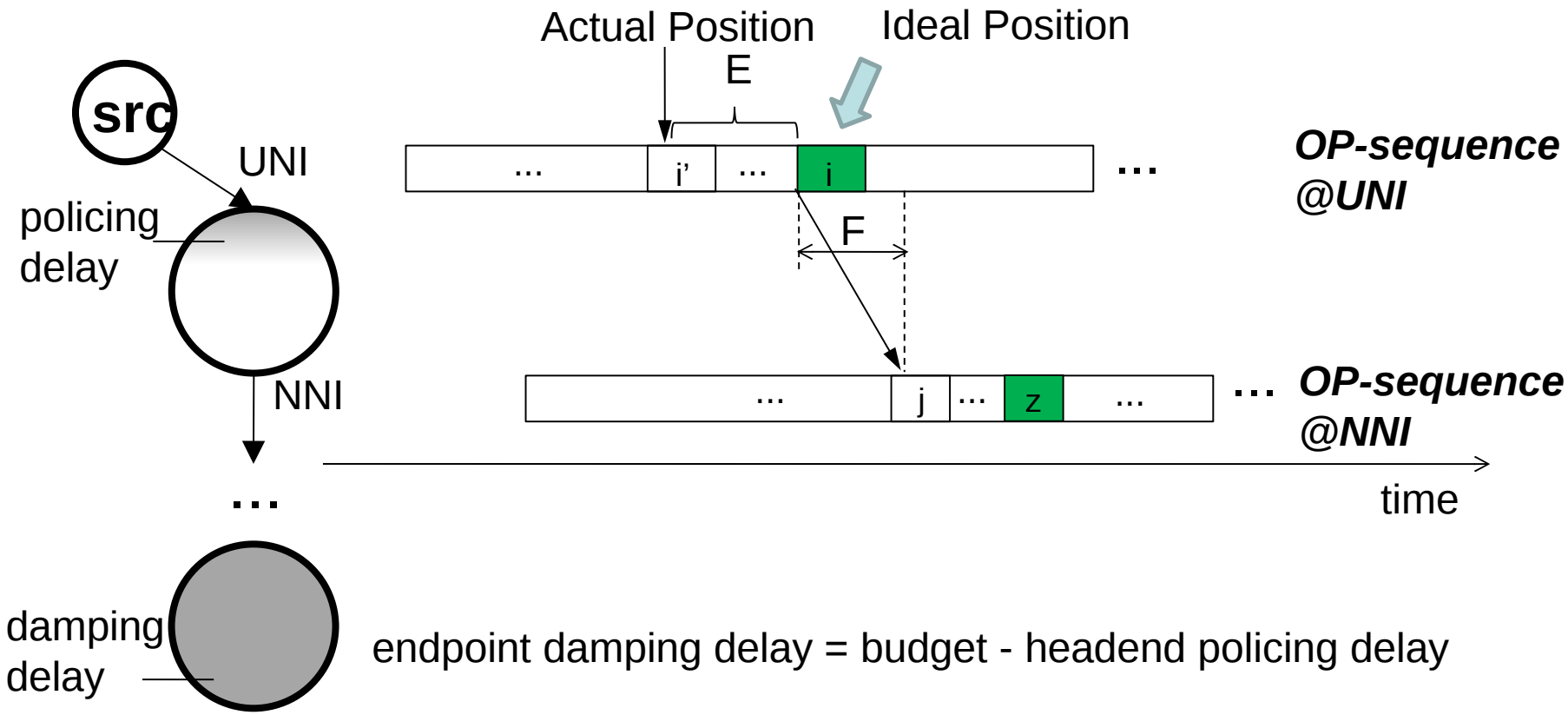




- Based on the topology related **fixed BTM or BOM**, a DetNet flow may allocate timeslot resources on ecah node and obtain **flexible FTM** (forwarding timeslot mapping) that meet its latency requirement.
  - The flexible FTM can achieve flow interleaving.
  - Flow interleaving only on the network entry is hard to avoid congestion on the transit nodes.

Runtime Policing Delay Jitter Elimination

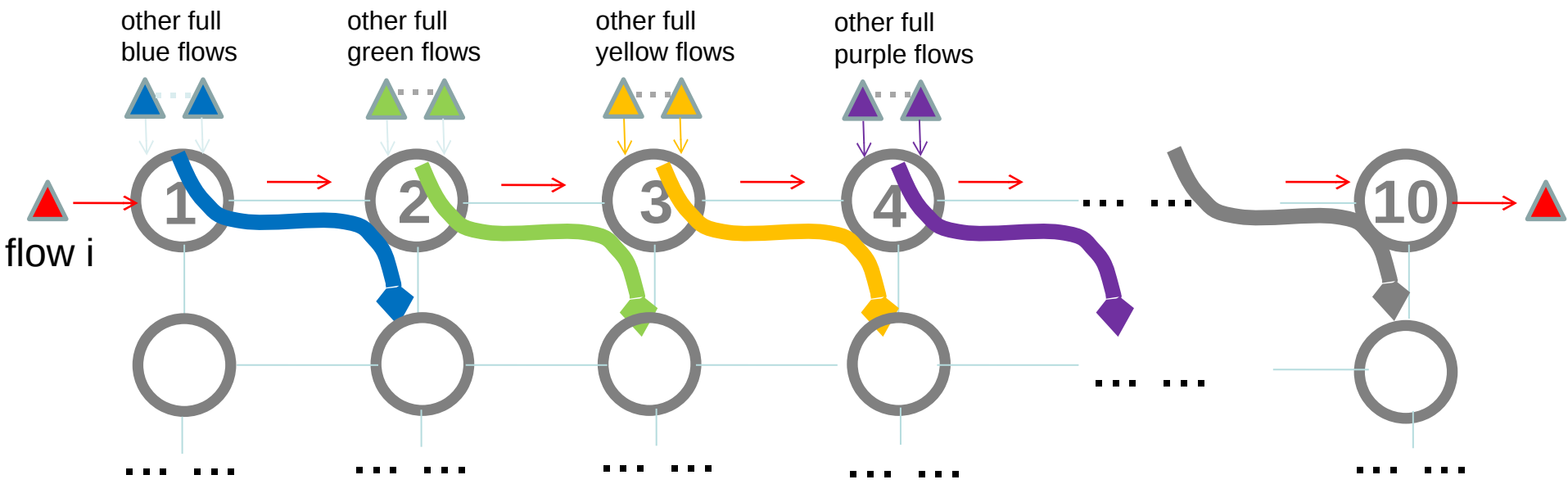
- The **runtime deviation E** (i.e., ideal arrival time minus actual arrival time) on the network entry **is treated as a part of policing delay**.
- Different packets may have different policing delay, so that jitter occurs.
- The policing delay jitter may be eliminated by the calculated endpoint damping delay on the egres node.



endpoint damping delay = budget - headend policing delay

Common Topology Example

- Link speed: **100 Gbps**.
- flows passed through each interface:
  - **TSpec**: each flow has packet size **1000 bits,** average rate **10 Mbps**.
  - **RSpec: (not including link propagation delay)**
    - flow1~flow100 may tolerate E2E latency **100us**, and E2E jitter **10us or 100us**.
    - flow101~flow200 may tolerate E2E latency **200us**, and E2E jitter **20us or 200us**.
    - ... ...
    - flow901~flow1000 may tolerate E2E latency **1ms**, and E2E jitter **100us or 1ms**.
- Topogogy



- OPL-100us resources for each delay level @ link (100 Gbps), supported number of flows (each with 10 Mbps).

Timeslot id	Resources	Offset (e.g)	Admited Flows	E2E Delay & Jitter (10 hops)
#0	b = 1000000 bits	1	1000	delay 100us, jitter 20us(on)/100us(in)
#1	b = 1000000 bits	2	1000	delay 200us, jitter 20us(on)/200us(in)
#2	b = 1000000 bits	3	1000	delay 300us, jitter 20us(on)/300us(in)
#3	b = 1000000 bits	4	1000	delay 400us, jitter 20us(on)/400us(in)
#4	b = 1000000 bits	5	1000	delay 500us, jitter 20us(on)/500us(in)
#5	b = 1000000 bits	6	1000	delay 600us, jitter 20us(on)/600us(in)
#6	b = 1000000 bits	7	1000	delay 700us, jitter 20us(on)/700us(in)
#7	b = 1000000 bits	8	1000	delay 800us, jitter 20us(on)/800us(in)
#8	b = 1000000 bits	9	1000	delay 900us, jitter 20us(on)/900us(in)
#9	b = 1000000 bits	10	1000	delay 1ms, jitter 20us(on)/1ms(in)
			total 10000	

1) **No overprovision**, i.e., addmision check is NOT to let a burst of flow consume all timeslots.
2) **Customized per-hop latency**, i.e., per-hop latency of flow is based on its RSpec and acheived by timeslot reservation, NOT its TSpec (i.e., service flow rate).

Taxonomy Considerations

- ***Per hop latency dominant factor***: Timeslot offset by reservation.
- ***Periodic***: Firstly, there is a time period P containing multiple timeslots, and secondly, a flow is assigned repeatly to a particular set of timeslots in the period..
- ***Frequency synchronous***: Crystal frequency aligned to ensure that all nodes have the same time lapse rate.
- ***Class level***: DetNet Flows is grouped by timeslot id.
- ***Work-conserving/non-work-conserving configurable***: Scheduler configured with in-time mode is work-conserving, while configured with on-time mode is non work-conserving.
- ***In-time/on-time configurable***: Scheduler may be configured with in-time mode or on-time mode.
- ***Time based***: A DetNet flow is scheduled based on its outgoing timeslot, rather on its reserved bandwidth (i.e., rate), to fit well the case of low bandwidth consumption but urgent flows.

Next step

- Any questions/comments ?

Thank you!