On-time Forwarding with Push-In First-Out (PIFO) Queue

draft-ryoo-detnet-ontime-forwarding-00

Yeoncheol Ryoo ETRI

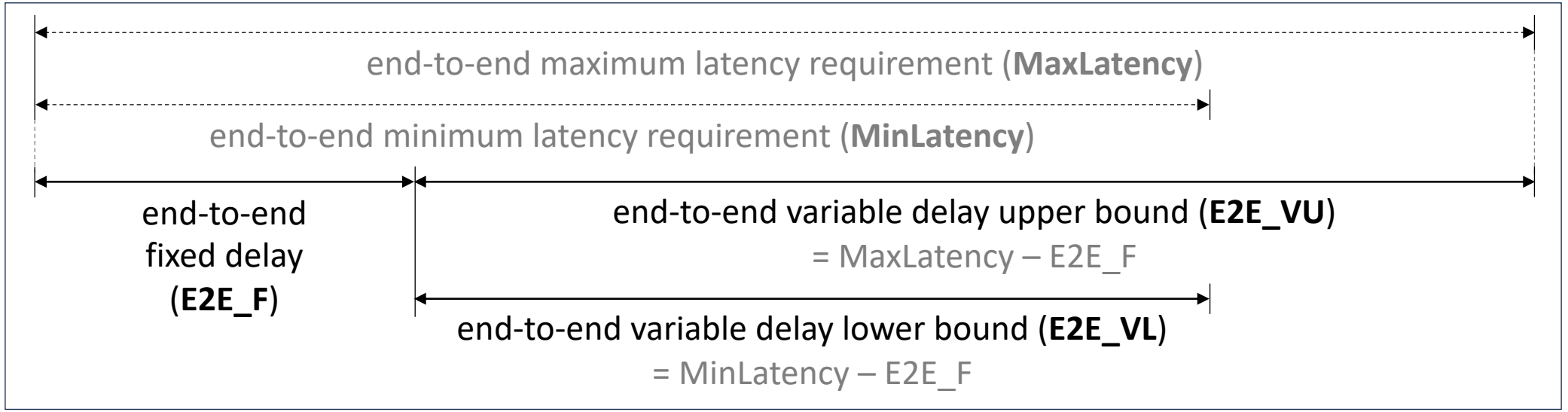
Contents

- Motivations
- Temporal model
- Data plane operation
- Controller plane operation
- Capability analysis

Motivations

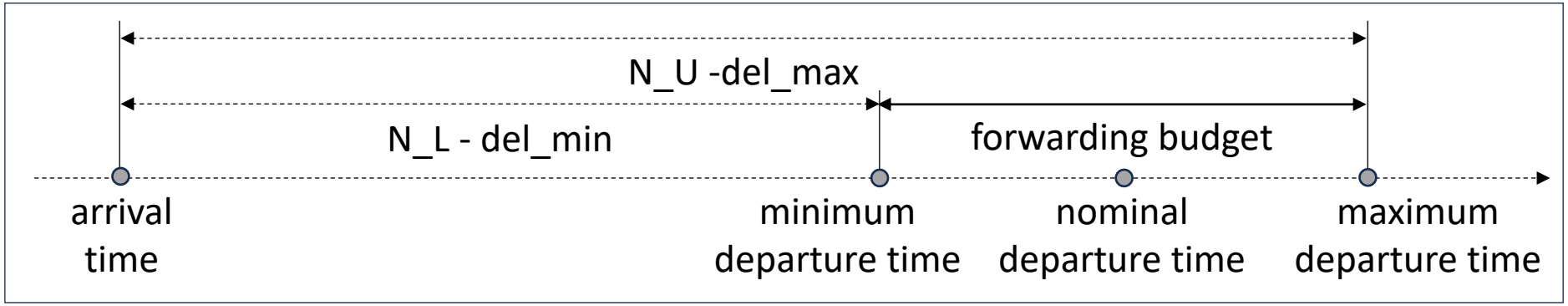
- Deterministic Networking **(DetNet)** [RFC8655]
  - **provides** the capability to carry specified unicast or multicast flows with **extremely low packet loss rates** and **bounded end-to-end latency** (Maximum Latency and Maximum Latency Variation [RFC9016])
- **On-time forwarding is a critical feature** of deterministic networks, especially of networks dealing with **industrial process control signaling**
- Propose a new method that allows for **more accurate on-time forwarding** and **more flexible flow setting** than the many proposed periodic queuing and forwarding methods.

Temporal model



- This document separates end-to-end latency into two components:
  - **end-to-end variable delay**
  - **end-to-end fixed delay**
- The end-to-end variable delay is the sum of variable delays occurring in nodes and links on the path, and has its upper and lower bounds
  - **Queuing delay, processing delay, output delay** (using the terms defined in [RFC9320])
- The end-to-end fixed delay is the sum of fixed delays occurring in links and nodes on the path.
  - **Link propagation delay** (using the terms defined in [RFC9320])

Data plane operation (1)

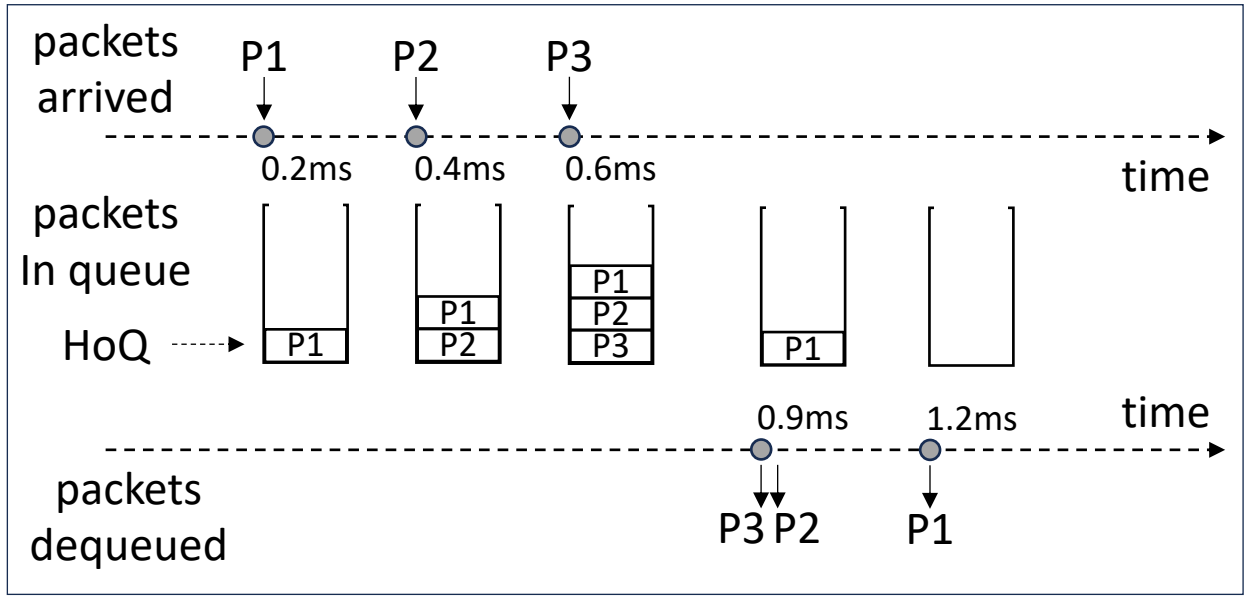


- **E2E_VU and E2E_VL are divided** into all the nodes on the path, and **the controller assigns a node delay upper bound (N_U) and a node delay lower bound (N_L)**
- **N_U and N_L are upper and lower bounds of the time a packet can reside in a node**
- Two new fields, **R_L** and **R_U**, are assumed to be available in the packet header to indicate the **upper and lower bounds of the remaining** end-to-end latency.
- **When a packet arrives at time t, a minimum departure time (t + N_L - del_min) and a maximum departure time (t + N_U - del_max) are calculated**
- del_min and del_max are defined as the minimum and maximum times it takes from the time a packet leaves the queue until it leaves the node
- the packet is placed in the PIFO queue, **where packets are arranged in the ascending order of their nominal departure times** which is the midpoint between the minimum departure time and maximum departure time
- **When the minimum departure time** of the packet in the head of queue (HoQ) has reached or passed current time, the packet is **dequeued**
- **When a packet leaves a node**, **R_L and R_U fields are updated** by subtracting its residence time from received R_L and R_U values

Data plane operation (2)

Assume:

- flow 1 : N_L = 1ms | N_U = 3ms
- flow 2 : N_L = 0.34ms | N_U = 2ms
- flow 3 : N_L = 0.3ms | N_U = 0.5ms
- del_min and del_max = 0



* HoQ : header of queue
* N_L : node delay lower bound
* N_U : node delay upper bound
* del_min : minimum output delay
* del_max : maximum output delay

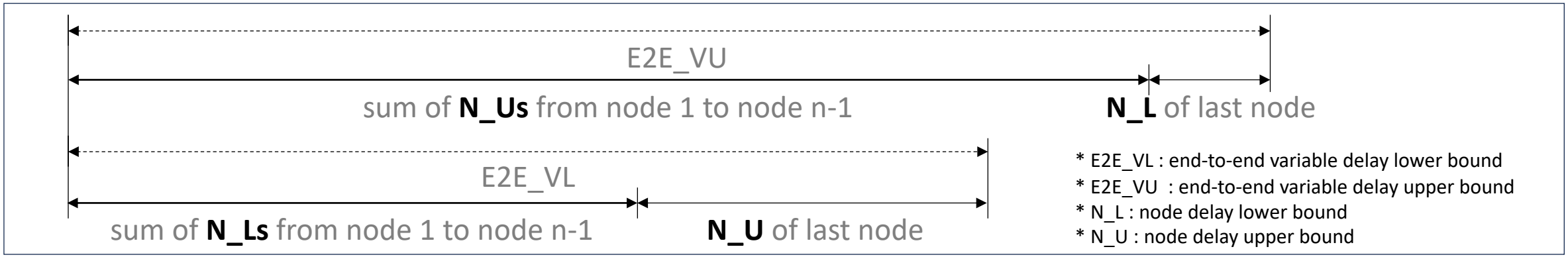
- The first packet, **P1**, arrives at **0.2ms**. The minimum, nominal, and maximum departure times of P1 are calculated as **1.2ms, 2.2ms, and 3.2ms**, respectively. P1 is **placed at the HoQ** and **cannot leave the queue before 1.2ms**.
- The second packet, **P2**, arrives at **0.4ms**. The minimum, nominal, and maximum departure times of P2 are calculated as **0.74ms, 1.57ms, and 2.4ms**, respectively. **Since the nominal departure time of P2 is smaller than that of P1, P2 is placed at the HoQ** and is scheduled to leave the queue at 0.74ms.
- The third packet, **P3**, arrives at **0.6ms**. The minimum, nominal, and maximum departure times of P3 are calculated as **0.9ms, 1.0ms, and 1.1ms**, respectively. **Since the nominal departure time of P3 is smaller than that of P2, P3 is placed at the HoQ** and is followed by P2 and P1.
- At **0.9ms**, **P3 leaves the queue as its minimum departure time** is 0.9ms. Following P3, **P2 immediately leaves the queue as its minimum departure time (0.74ms) has passed**.
- At **1.2ms**, **P1 is dequeued** as its minimum departure time is 1.2ms.

Controller plane operation (1)

- **Collects** resources and delays information
  - Topology, PIFO queue resource, port speed, link delay, maximum and minimum processing and output delay
- **Selects** a path that can satisfy the requirement with traffic specifications, resources, and delay information.
  - The **E2E_F and maximum end-to-end variable non-queuing delay** of the path **MUST NOT exceed the MaxLatency** required for the flow
  - The sum of the available **buffer resources of all nodes** on the path **MUST be large** enough to provide a delay greater than **E2E_VL minus minimum end-to-end variable non-queuing delay**
- **Calculates** the N_U and N_L and **performs** admission control.
  - The available **buffer resource of each node** on the path **MUST be large** enough to provide a delay equal to **N_U minus minimum node variable non-queuing delay**
  - The **forwarding budget** of each node **MUST be larger** than the time required to transmit any preceding packets of all the flows at the speed of the output port

* E2E_F : end-to-end fixed delay
* MaxLatency : maximum end-to-end latency that must be guaranteed
* E2E_VL : end-to-end variable delay lower bound
* N_L : node delay lower bound
* Forwarding budget : difference between (N_U - del_max) and (N_L - del_min)

Controller plane operation (2)



- The **last node is the node that can take final actions to ensure the E2E_VL and E2E_VU** for packets requiring on-time delivery, the value of N_U and N_L of the last node MUST be determined first
- Last node
  - N_U : recommended to set the value as large as possible as long as the buffer resource of the last node allows for the flow
  - N_L : minimum time required to transmit any preceding packets of all the flows at the speed of the output port of the last node
- 1 to n-1 nodes
  - N_U : the remaining value after subtracting N_L of the last node from E2E_VU is divided into all other nodes
  - N_L : the remaining value after subtracting N_U of the last node from E2E_VL is divided into all other nodes

Capability analysis

- **The proposed solution has the following characteristics for the requirements** described in [I-D.ietf-detnet-scaling-requirements]

(The item numbers below correspond to the numbers of the technical requirements in Section 3 of [I-D.ietf-detnet-scaling-requirements])

1. **The solution does not require time synchronization**. However, the solution measures the residence time and passes the remaining end-to-end latency values to the next node. As a specific delay value seen by all nodes must be the same amount, **frequency synchronization is necessary**.
2. **The large single-hop propagation delay is supported**. The solution does not impose any limits on the amount of propagation delay.
3. **Accommodation of the higher link speed is supported**. It is considered possible to implement a PIFO queue supporting speeds of 100 Gbps or more.
4. **The solution is scalable to a large number of flows** as it **does not require to maintain flow states** in a node.
5. **The solution is robust against node and link failures and topology changes**, as the **PREOF function can be applied**.
6. **Since the solution provides on-time forwarding** while complying with the forwarding budget at all nodes, **flow fluctuation inherently does not occur**.
7. **Since each node operates independently** and the operation of the controller does not require any greater burden than existing typical network control, **there are no scalability issues regarding the number of hops**.

Next step

- Update the document and address the comments
- Any questions/comments?

Thank you!