Asynchronous Deterministic Networking (ADN) Framework for Large scale networks

draft-joung-detnet-asynch-detnet-framework-02

Jinoo Joung, Jeong-dong Ryoo, Tae-sik Cheung, Yizhou Li, Peng Liu

IETF 116

Latency guarantee framework revised in version 02

❶ Work Conserving Stateless Core Fair queuing (C-SCORE)

- FT, Finish time F(p) = Service order of packet p. Smaller F gets earlier service.
- At entrance node 0: $F_0(p) = \max\{F_0(p\text{-}1), A_0(p)\} + L(p)/r$;
- At core node h: $F_h(p) = F_{h-1}(p) + d_h(p)$.

- If $d_h(p) = Lmax_{h-1}/R_{h-1} + L/r$,
- Then the E2E latency of p's flow is bounded [Kaur] by

$$\frac{B-L}{r} + \sum_{h=0}^{H} \left(\frac{Lmax_h}{R_h} + L/r\right)$$

  - where H is the last hop of the p's flow.

$\frac{Lmax_h}{R_h}$ is the only term external & can be negligible.

This bound is same as the fair queuing with flow-state (PGPS, etc.)

- E.g., 1Gbps links, 1Kbits max pkt length & Burst, 10Mbps service rate, 10 hops → 10*(1us+100us) = 1.01ms E2E latency bound
- B, L, r are flow specific, which can be controlled according to requirement
  - Need a smaller bound? Then request a larger service rate, and reduce burst & pkt length.
- Implementation:
  - Sorted queue e.g. Queue per Input-port, Push-in First Out manner
  - Meta-data: $F_{h-1}(p)$, $Lmax_{h-1}/R_{h-1}$, L/r. ($Lmax_{h-1}/R_{h-1}$ can be signaled in control plane.)
  - Or $F_h(p)$. (It can be pre-calculated at h-1.)

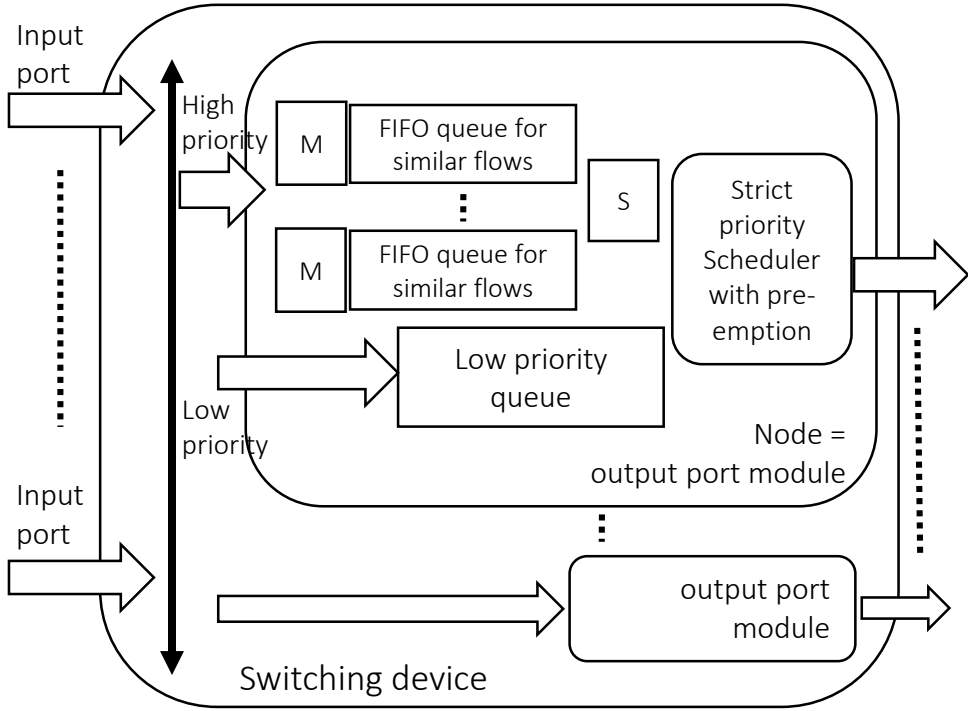
Symbol	Definition
$F_h(p)$	'Finish time' of packet p at node h
$A_h(p)$	Arrival time of p at h
L(p)	Length of p
L	Max Packet Length of p's flow
r	Service rate of p's flow
B	Max Burst size of p's flow
$Lmax_h$	Max Packet Length at node h
R_h	Link capacity of h
H	Last node of p's path
PGPS	Packetized Generalized Processor Sharing

Latency guarantee framework revised in version 02

❷ Work Conserving Stateless Core Fair queuing (C-SCORE) with FIFO queues

- FT, Finish time F(p) = Service order of packet p. Smaller F gets earlier service.
- At entrance node 0: $F_0(p) = \max\{F_0(p-1), A_0(p)\} + L(p)/r$;
- At core node h: $F_h(p) = F_{h-1}(p) + d_h(p)$.

- $d_h(p) = Lmax_{h-1}/R_{h-1} + L/r$.
- Implementation:
  - Assume there are finite number of flow types (e.g. similar L and r)
  - Queue per a set of similar types of flows
  - First in First Out manner
- Simulation shows almost the same E2E bound with the previous one with PIFO queues.



M: Finish time (F) marker.

S: HoQ examine & select the min FT

Simulation setup

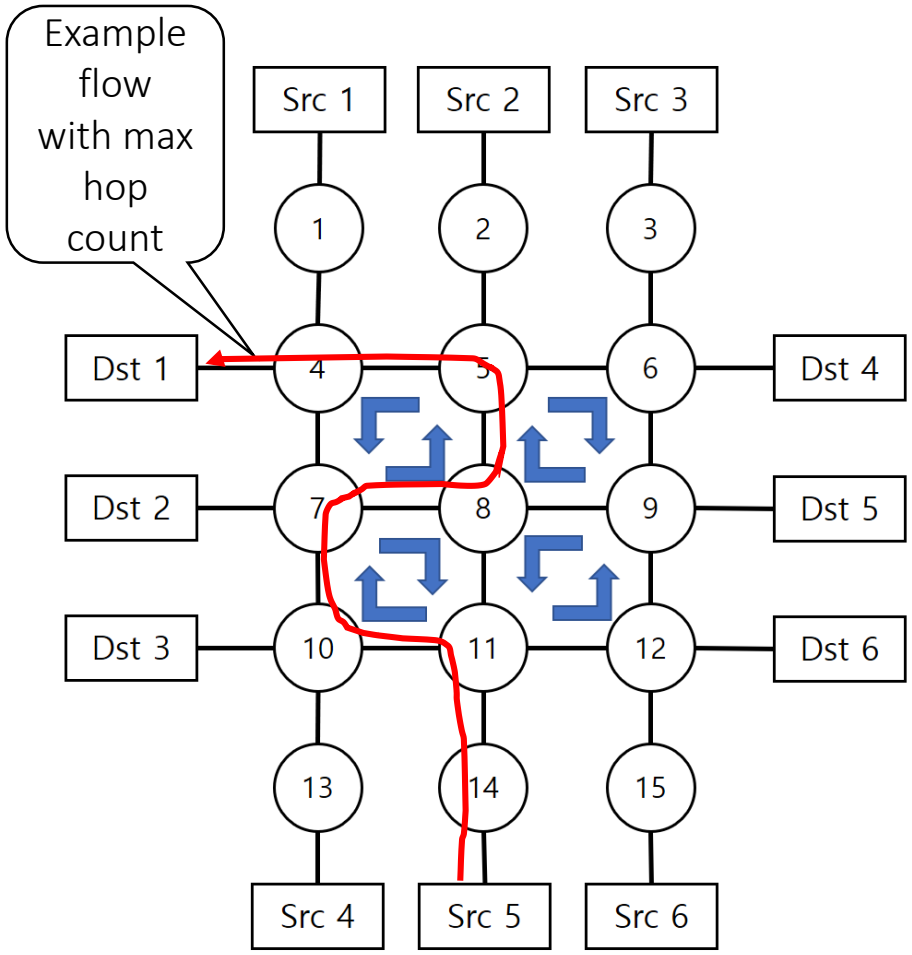
- Three flow types

Flow type	Maximum burst size	Maximum Packet length	Destination
A	200Kbit	10Kbit	0, 5
B	200Kbit	10Kbit	2, 3
C	20Kbit	2Kbit	1, 4

We are interested in Type C because it's smooth. It has to be protected from other bursty flows.

	Service rate [Mbps]		
Utilization	A	B	C
70%	98.571	9.857	98.571
75%	105.714	10.571	105.714
80%	112.619	11.262	112.619
85%	119.762	11.976	119.762
90%	126.667	12.667	126.667
95%	133.81	13.381	133.81

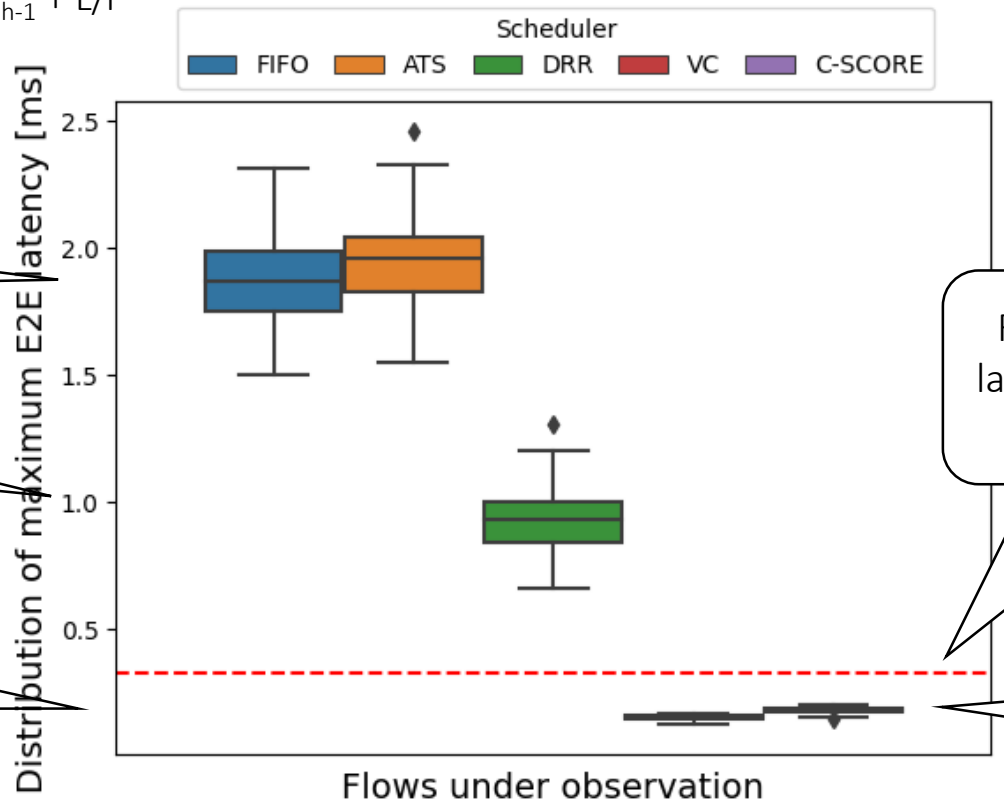
- 1Gbps links
- Flow service rate = arrival rate
- A source generates a flow for each destination, total six flows.
  - A flow's destination decides its type.
- Total 36 flows
- Max 7 hops. A flow's path follows the blue arrows in the right figure.
- Utilization value is of bottleneck links



Simulations for ❶

- 90% Utilization
- In a single simulation run, a source generates 1000 packets.
- Record the observed max E2E latency. One max value per run.
- 100 runs with random seeds → Box plot those max values.
- C-SCORE is with $d_h(p) = Lmax_{h-1}/R_{h-1} + L/r$

Flow type	Maximum burst size	Maximum Packet length	Destination
A	200Kbit	10Kbit	0, 5
B	200Kbit	10Kbit	2, 3
C	20Kbit	2Kbit	1, 4



ATS is of IR+FIFO.

DRR has queue per flow.

VC is of the network having Virtual Clock scheduler at every node.

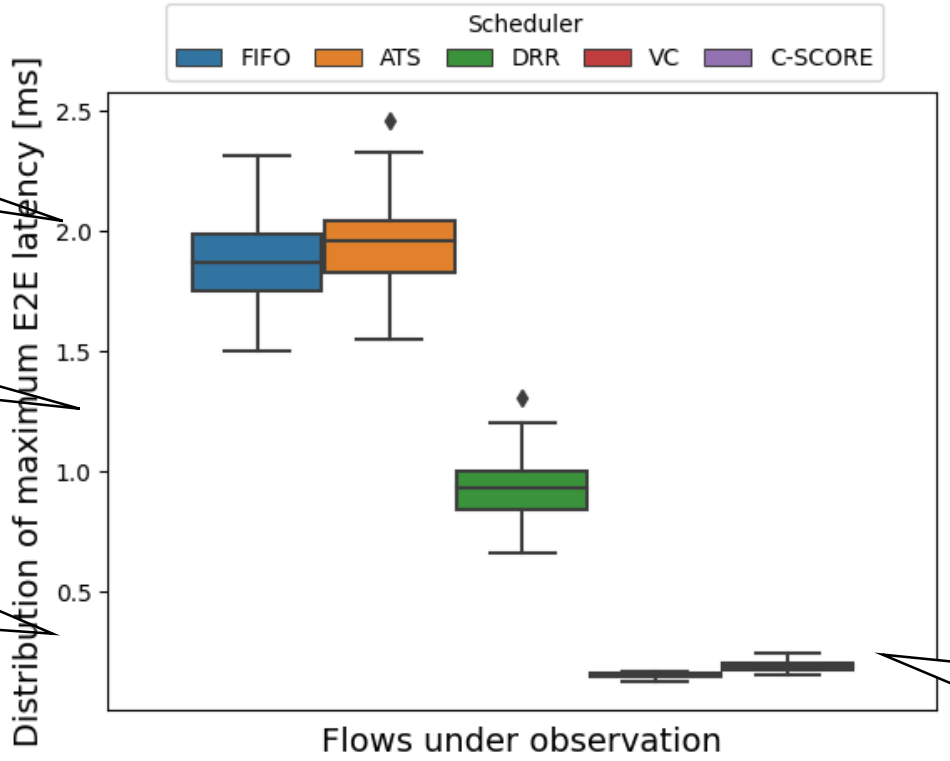
Red line: Theoretical E2E latency bounds of C-SCORE & VC; 0.3226ms.

Maximum observed E2E latency of VC & C-SCORE is 0.27ms

Simulations for ❷

- 90% Utilization
- In a single simulation run, a source generates 1000 packets.
- Record the observed max E2E latency. One max value per run.
- 100 runs with random seeds → box plot those max values.
- C-SCORE is with $d_h(p) = Lmax_{h-1}/R_{h-1} + L/r$ with FIFO per flows of same type

Flow type	Maximum burst size	Maximum Packet length	Destination
A	200Kbit	10Kbit	0, 5
B	200Kbit	10Kbit	2, 3
C	20Kbit	2Kbit	1, 4

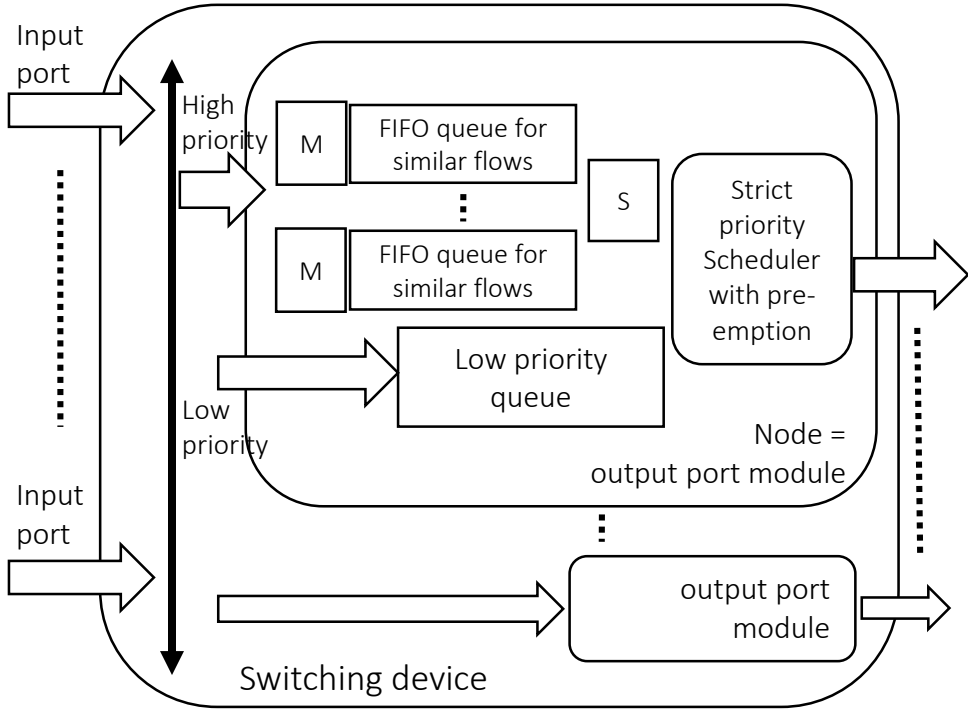


ATS is of IR+FIFO.

DRR has queue per flow.

VC is of the network having Virtual Clock scheduler at every node.

With FIFO, C-SCORE shows indistinguishable E2E max latency statistics.



Discussion

- Consider an ideal slotted scheduler.
  - Flow arrival rate = 10Mbps
  - 100 identical flows
  - Link capacity: R = 1Gbps
  - All the packet's lengths are fixed at L = 10Kbits; No burst (B=L).
  - Then the slot (cycle) length would be 10K/1G = 10us;
  - & the hyper-cycle length = 100*slot length = 1ms, without dead-time or any lost time.
  - Assume every packet arrives exactly at the slot to which it is assigned.
  - With H hops the E2E latency bound is H*L/r. If H = 10 then 10ms.
- C-SCORE's in this case is $\mathrm{H} * (\frac{Lmax}{R} + L/r)$.
  - It is only H*10us larger, thus 10.1ms; with all the robustness, scalability, & superior statistical performance.

Summary

- C-SCORE with $d_h(p) = Lmax_{h-1}/R_{h-1} + L/r$
  - Has a theoretical E2E latency bound,
  - which is the same with the Virtual Clock or PGPS networks
  - Needs a sorted queue
  - Needs a meta-data of $F_{h-1}(p)$, L, r, $Lmax_{h-1}/R_{h-1}$. (Or $F_h(p)$, pre-calculated at h-1.)
  - $Lmax_{h-1}/R_{h-1}$ can be signaled in control plane.
- C-SCORE with $d_h(p) = Lmax_{h-1}/R_{h-1} + L/r$; with FIFO per flow type (of the same L and r)
  - Shows similar performance & needs the same metadata
  - Does not need a sorted queue
- C-SCORE with $d_h(p)$ = any similar value less than $Lmax_{h-1}/R_{h-1} + L/r$; with FIFO
  - Shows similar performance (Not shown in this presentation)
  - Needs meta-data of only $F_{h-1}(p)$, L, r
- They all perform as if the per-flow states are maintained in all the nodes.

Thank you

- Please take a look at

  https://datatracker.ietf.org/doc/draft-joung-detnet-asynch-detnet-framework/

- Comments and Questions are welcome!

- [Kaur] Jasleen Kaur, and Harrick M. Vin. "Core-stateless guaranteed rate scheduling algorithms." In Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No. 01CH37213), vol. 3, pp. 1484-1492. IEEE, 2001.
- [FAIR] Jinoo Joung. "Framework for delay guarantee in multi-domain networks based on interleaved regulators." Electronics 9, no. 3 (2020).
- [ADN] Jinoo Joung, Juhyeok Kwon, Jeong-Dong Ryoo, and Taesik Cheung. "Asynchronous Deterministic Network Based on the DiffServ Architecture." IEEE Access 10 (2022).
- [Zhang] Lixia Zhang. "Virtual clock: A new traffic control algorithm for packet switching networks." In *Proceedings of the ACM symposium on Communications architectures & protocols*, pp. 19-29. 1990.
- [Stoica] Ion Stoica and Hui Zhang. "Providing guaranteed services without per flow management." *ACM SIGCOMM Computer Communication Review* 29, no. 4 (1999): 81-94.
- [Stiliadis] Dimitrios Stiliadis and Varma Anujan. "Rate-proportional servers: A design methodology for fair queueing algorithms." IEEE/ACM Transactions on networking 6, no. 2 (1998): 164-174.
- [BN] Jinoo Joung and Juhyeok Kwon. "Zero jitter for deterministic networks without time-synchronization." IEEE Access 9 (2021).