DetNet Working Group Y. Ryoo
Internet-Draft ETRI
Intended status: Standards Track J. Joung
Expires: 2 September 2026 Sangmyung University
1 March 2026

On-time Forwarding with Non-Work Conserving Stateless Core Fair Queuing
draft-ietf-detnet-nscore-01

Abstract

This document specifies the framework and operational procedure for deterministic networking that guarantees maximum and minimum end-to-end latency bounds to flows. The solution has non-periodic, asynchronous, flow-level, non-work conserving, on-time, and rate-based functional characteristics, according to the taxonomy suggested by [I-D.ietf-detnet-dataplane-taxonomy].

The packets are stored in the queue in ascending order of the ideal service start time, called Eligible Time (ET), and the ideal service completion time, called Finish Time (FT). The queued packets were forwarded after ET, in ascending ordering of FT, in a non-work conserving manner. The ET and FT are calculated at the entrance node according to the packet size and rate of the flow. All subsequent core nodes are stateless and asynchronously update ET and FT based on metadata received via packet headers. This mechanism is called non-work conserving stateless fair queuing (N-SCORE), which guarantees both E2E latency upper and lower bounds, thus E2E jitter bound.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

A class of schedulers called Fair Queuing (FQ) limits interference between flows to the degree of the maximum packet size divided by the link capacity. In FQ, the ideal service completion time, called Finish Time (FT), of a packet is obtained from an imaginary system that can provide the ideal flow isolation. Applying this technique, the end-to-end (E2E) latency bound of a flow is similar to that of an ideally isolated system.

Since calculating the FT of the current packet, in FQ, requires the FT of previous packets within the flow, this means that nodes must manage the state of the flow. The complexity of managing the state of a large number of flows can be a burden, so the proposed framework for large-scale deterministic networking is called work conserving stateless core fair queuing (C-SCORE) [I-D.joung-detnet-stateless-fair-queuing], which generates FT for packets at the entrance node and marks FT in the packet to operate with stateless in core nodes.

However, C-SCORE is a scheduler of work conserving approach, so it has an in-time characteristic and does not provide a jitter guarantee. Therefore, this draft proposes a non-work conserving scheduler method by extending C-SCORE to have an on-time characteristic, called N-SCORE. The entrance node additionally generates an ideal service start time, called an eligible time (ET), of the current packet based on the FT of the previous packet or the arrival time of the current packet. Nodes admit all eligible packets, defined as those with an ET preceding the current time, into the output queue in non-decreasing order of their FT for subsequent transmission. This makes it in a non-work conserving shceduler. N-SCORE is a method that guarantees not only the upper bound but also the lower bound of E2E latency by adding ET while using the information managed by the entrance node of the existing C-SCORE.

2. Terminology

2.1. Symbols Used in This Document

FQ	fair queuing
FT	finish time
ET	eligible time
Fh(p)	FT of the packet p at the node h
Eh(p)	ET of the packet p at the node h
Ah(p)	arrival time of the packet p at the node h
dh(p)	delay factor function of the packet p at the node h
Ch(p)	service completion time of packet p at the node h
r(p)	service rate of the packet p
L(p)	length of the packet p
L	maximum packet length of flow unger observation
Rh	link capacity at the node h
Lhmax	maximum packet length of the node h
PDh	propagation delay of the link h

2.2. Abbreviations

3. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

4. N-SCORE Packet Scheduler Framework

Utilizing the concept of virtual clock (VC) scheduler, C-SCORE defines FT for packet p, at the packet entrance node, as

F(p) = max{F(p-1), A(p)} + L(p)/r(p) (1)

Where (p-1) and p are consecutive packets of the flow being observed, F(p-1) is the finish time of p-1, A(p) is the arrival time of p, L(p) is the length of p, and r(p) is the flow service rate. Equation (1) adopts a simplified notation by omitting the flow indicators to the mathematical symbols.

In C-SCORE, the entrance node manages F(p-1) and obtains F(p) by comparing it with A(p). Then, it calculates F(p) of the next node and marks it in the packet header. The service period of packet p in each node is defined as (A(p), F(p)]. Assuming the link propagation delay is zero, an example of the packet service period at the entrance node and core node with the C-SCORE scheduler is illustrated as follows:

```
A1(1)           A1(2)A1(3)A1(4)
|               |    |    |
V               V    V    V
<----1---->     <----2---->F1(2)                            node 1
|       F1(1)   |    <-------3------->F1(3)
|               |    |    <----------4---------->F1(4)
|               |    |    |
A2(1)           A2(2)A2(3)A2(4)
|               |    |    |
V               V    V    V
<----------1--------->F2(1)                                 node 2
                <---------2---------->F2(2)
                     <-------------3----------->F2(3)
                          <----------------4------------->F2(4)
```

Figure 1: C-SCORE packet scheduler service period

The proposed N-SCORE framework introduces an additional parameter, ET (Eligible Time), which is used as the earliest possible packet service start time. Without requiring additional state management for ET, N-SCORE utilizes the information already managed by the entrance node in the existing C-SCORE to obtain ET and FT as follows:

$$E(p) = \max\{F(p-1), A(p)\} \quad (2)$$

$$F(p) = E(p) + L(p)/r(p) \quad (3)$$

A packet can join the output link scheduler immediately after its ET. If no other packet is present in the scheduler, the packet is served right away. Otherwise, the packet joins the queue. Packets in the queue are served in ascending order of their FT. Since the FT of N-SCORE is identical to that of C-SCORE, packets in N-SCORE follow the same service order as in C-SCORE. The only difference between the two systems is the existence of ET. However, in N-SCORE, due to the presence of the ET, the service period of packet p, while maintaining the same service order, is defined as (E(p), F(p)].

Consequently, N-SCORE forwards packets in a non-work conserving manner, maintaining a constant interval between E(p) and F(p) in all nodes. The service periods of packets within the same flow do not overlap at each node. Assuming zero link propagation delay, the packet service period at the entrance and core nodes with the N-SCORE scheduler is illustrated as follows:

```
A1(1)           A1(2)A1(3)A1(4)
|               |    |    |
V               V    V    V
<----1---->     <----2----><----3----><----4---->           node 1
E1(1)  F1(1)    E1(2) F1(2)=E1(3) F1(3)=E1(4) F1(4)
|               |          |          |
A2(1)           A2(2)      A2(3)      A2(4)
|               |          |          |
V               V          V          V
...........<----1---->.....<----2----><----3----><----4---->   node 2
           E2(1)  F2(1)    E2(2) F2(2)=E2(3) F2(3)=E2(4) F2(4)
```

Figure 2: N-SCORE packet scheduler service period

5. E2E latency and jitter bound

The end-to-end (E2E) latency of N-SCORE is upper-bounded by:

(B-L)/r + ☒ (h=0 to H){L/r + Lhmax/Rh} (4)

which is the same as that of C-SCORE, which operates based on FT. Here, B, L, and r represent the maximum burst size, maximum packet length, and service rate of the observed flow, respectively. The link propagation delay is neglected.

Unlike C-SCORE, which has no lower bound for E2E latency, the E2E latency of N-SCORE, which operates based on both ET and FT, is lower-bounded by:

☒ (h=0 to H-1){L/r + Lhmax/Rh} + Lmin/RH (5)

where L, Lmin, and r denote the maximum packet length, minimum packet length, and service rate of the observed flow, respectively. The link propagation delay is neglected.

Therefore, unlike C-SCORE, which exhibits jitter ranging from 0 to the E2E maximum delay, the E2E jitter of N-SCORE is bounded by:

B/r + LHmax/RH - Lmin/RH (6)

6. Operational Procedure

The N-SCORE scheduler in all nodes has a deterministic service period of (E(p), F(p)] for packet p. Packets are first pushed into a temporary queue with ascending order of ET. Upon their eligible time, the packets are transferred to the service queue and pushed with ascending order of FT. Overall, the packets are serviced after their eligible times, in the order of finish times. This makes it a non-work conserving scheduler.

N-SCORE manages per-flow state to calculate ET and FT at the entrance node. However, core nodes do not maintain state to accommodate large-scale networks. As a result, N-SCORE calculates and applies ET and FT differently at the entrance node and subsequent core nodes.

Whenever a packet arrives, the entrance node calculates its ET and FT based on the managed per-flow state, updates the state using the calculated FT, and appends ET and FT as metadata to the packet header. Subsequent core nodes retrieve ET and FT from the metadata without maintaining state separately. At the same time, they calculate new ET and FT for the next node and update the metadata accordingly.

6.1. Operational Procedure in Entrance Node

The entrance node manages the per-flow state, including the FT of the previous packet, F(p−1), and the service rate assigned to the flow, r(p). When a packet arrives at the entrance node, its ET, E(p), is determined as max{F(p−1), A(p)}. The entrance node compares each packet’s arrival time, A(p), with the managed F(p−1) and sets the eligible time as E(p). The FT of the arriving packet, F(p), is calculated as E(p) + L(p)/r(p), and the FT of the previous packet is updated with the newly obtained F(p). Packets are stored in a priority queue in ascending order of F(p), after their E(p), and are forwarded. Ovarall this is a a non-work conserving scheduler.

When the packet arrival interval is greater than the service rate, as seen with the first and second packets in Figure 2, the arrival times of these packets at node 1, A1(1) and A1(2), are later than the FT of the previous packet managed by the entrance node, F1(0) and F1 (1), respectively. Therefore, the ET of the first and second packets at node 1, E1(1) and E1(2), are set as A1(1) and A1(2), respectively. In this case, the service period is (A(p), A(p) + L(p)/r(p)], which matches the service period of C-SCORE.

However, when the packet arrival interval is smaller than the service rate, as seen with the third and fourth packets in Figure 2, the arrival times of these packets at node 1, A1(3) and A1(4), are earlier than the FT of the previous packet managed by the entrance node, F1(2) and F1(3), respectively. Consequently, the ET of the third and fourth packets at node 1, E1(3) and E1(4), are set as F1(2) and F1(3), respectively. In this case, unlike C-SCORE’ s service period of (A(p), F(p−1) + L(p)/r(p)], the N-SCORE’s service period is (F(p−1), F(p−1) + L(p)/r(p)]. N-SCORE effectively regulates packet transmission based on the service rate, ensuring a deterministic and non-overlapping service period for all packets.

The entrance node marks metadata in the packet header, including L(p)/r(p), as well as the ET and FT for the next node. The subsequent core nodes then use this metadata to determine their ET and FT.

6.2. Operational Procedure in Core Node

When the ET and FT of a packet are determined at the entrance node, the ET and FT of all subsequent nodes are determined based on the previous node’s ET and FT as follows:

Eligible Time for the next node:

E(h+1)(p) = Eh(p) + dh(p) (7)

Finish Time for the next node:

F(h+1)(p) = Fh(p) + dh(p) (8)

Here, dh(p) represents the maximum delay within node h, which is calculated as:

dh(p) = L(p)/r(p) + Lhmax/Rh (9)

The term Lhmax/Rh accounts for delay factors at node h, where Lhmax is the max packet length at node h across all flows of the output port, and Rh is the link capacity of node h.

The entrance node delivers the metadata, including L(p)/r(p), ET, and FT, through the packet header. Subsequent core nodes obtain their ET and FT from the metadata without per-flow state management. Based on its delay factors and L(p)/r(p) value in the metadata, each core node computes dh(p), determines the ET and FT for the next node, and updates the metadata accordingly.

Packets are stored in a priority queue in ascending order of F(p), after E(p), as derived from the metadata, and can be forwarded in a non-work conserving manner.

7. Characteristics

7.1. Scaling requirements

The data and controller plane operations described in this document have 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. N-SCORE does not require time synchronization. However, in order to apply the eligible time and finish time calculated by the previous node, the time difference between the previous node and the current node must be known.

2. N-SCORE supports large single-hop propagation delays and does not impose any restrictions on the amount of propagation delay.

3. N-SCORE supports the accommodation of the higher link speed. It is considered possible to implement a PIFO queue supporting speeds of 100 Gbps or more.

4. N-SCORE is scalable to the large number of flows as it does not require to maintain flow states in a node.

5. N-SCORE is robust against node and link failures and topology changes, as the PREOF function can be applied.

6. N-SCORE is a fair queuing-based solution that provides the benefit of near-complete isolation between flows. Therefore, it effectively prevents flow fluctuations even when different flows dynamically join or leave the system.

7. The admission condition of N-SCORE depends solely on the service rates of flows. Therefore, the admission checking process is simple, and there are no scalability issues with respect to the number of hops.

8. N-SCORE uses a dedicated PIFO queue and clearly distinguishes the algorithm applied to it from that used for the existing FIFO queue. It supports multiple mechanisms by appropriately mapping each flow to a queue based on its SLA. Furthermore, it can support multiple algorithms across multiple domains by compartmentalizing the end-to-end delay requirements according to sections divided by differences in domain or link speed, and applying an appropriate service rate for each section.

7.2. Taxonomy

According to the taxonomy defined in [I-D.ietf-detnet-dataplane-taxonomy], latency-bound solutions are classified according to functional characteristics such as

* periodicity (periodic, non-periodic)

* network synchronization (phase and frequency synchronous, asynchronous)

* traffic granularity (flow level, flow aggregate level, class level)

* time bound (bounded, left-bounded, right-bounded, unbounded)

* service order (rate-based, time-based, arrival-based, priority-based)

N-SCORE is a non-periodic, asynchronous, flow level, left-bounded, rate-based solution.

[I-D.ietf-detnet-dataplane-taxonomy] also defines seven suitable categories for deterministic networking as follows:

* Right-bounded category

* Flow level periodic bounded category

* Class level periodic bounded category

* Flow level non-periodic bounded category

* Class level non-periodic bounded category

* Flow level rate based unbounded category

* Flow level rate based left-bounded category

N-SCORE belongs to the "Flow level rate based left-bounded category", which is an on-time solution with rate-based service order characteristic that can handle a large number of dynamic flows with simple admission control. Additionally, it has flow-level traffic granularity characteristics that can minimize the effects of other flows' bursts.

8. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

9. Security Considerations

TBD

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

10.2. Informative References

[I-D.joung-detnet-stateless-fair-queuing]
Joung, J., Ryoo, J., Cheung, T., Li, Y., and P. Liu, "Latency Guarantee with Stateless Fair Queuing", Work in

Progress, Internet-Draft, draft-joung-detnet-stateless-fair-queuing-08, 28 February 2026, <https://datatracker.ietf.org/doc/html/draft-joung-detnet-stateless-fair-queuing-08>.

[I-D.ietf-detnet-scaling-requirements]
Liu, P., Li, Y., Eckert, T. T., Xiong, Q., Ryoo, J., zhushiyin, and X. Geng, "Requirements for Scaling Deterministic Networks", Work in Progress, Internet-Draft, draft-ietf-detnet-scaling-requirements-09, 7 September 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-detnet-scaling-requirements-09>.

[I-D.ietf-detnet-dataplane-taxonomy]
Joung, J., Geng, X., Peng, S., and T. T. Eckert, "Dataplane Enhancement Taxonomy", Work in Progress, Internet-Draft, draft-ietf-detnet-dataplane-taxonomy-05, 8 January 2026, <https://datatracker.ietf.org/doc/html/draft-ietf-detnet-dataplane-taxonomy-05>.

Authors' Addresses

Yeoncheol Ryoo
ETRI
Email: dbduscjf@etri.re.kr

Jinoo Joung
Sangmyung University
Email: jjoung@smu.ac.kr