Prados-Garzon, Jonathan; Taleb, Tarik; Bagaa, Miloud

LEARNET: Reinforcement Learning Based Flow Scheduling for Asynchronous Deterministic Networks

Published in:
2020 IEEE International Conference on Communications, ICC 2020 - Proceedings

DOI:
10.1109/ICC40277.2020.9149092

Published: 01/06/2020

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Prados-Garzon, J., Taleb, T., & Bagaa, M. (2020). LEARNET: Reinforcement Learning Based Flow Scheduling for Asynchronous Deterministic Networks. In *2020 IEEE International Conference on Communications, ICC 2020 - Proceedings* Article 9149092 (IEEE International Conference on Communications). IEEE. https://doi.org/10.1109/ICC40277.2020.9149092

LEARNET: Reinforcement Learning Based Flow Scheduling for Asynchronous Deterministic Networks

Jonathan Prados-Garzon*, Tarik Taleb*§, and Miloud Bagaa*
jonathan.prados-garzon@aalto.fi, tarik.taleb@aalto.fi, miloud.bagaa@aalto.fi
*Aalto University, Espoo, Finland.
§University of Oulu, 90570 Oulu, Finland.

***Abstract*—Time-Sensitive Networking (TSN) and Deterministic Networking (DetNet) standards come to satisfy the needs of many industries for deterministic network services. That is the ability to establish a multi-hop path over an IP network for a given flow with deterministic Quality of Service (QoS) guarantees in terms of latency, jitter, packet loss, and reliability. In this work, we propose a reinforcement learning-based solution, which is dubbed LEARNET, for the flow scheduling in deterministic asynchronous networks. The solution leverages predictive data analytics and reinforcement learning to maximize the network operator's revenue. We evaluate the performance of LEARNET through simulation in a fifth-generation (5G) asynchronous deterministic backhaul network where incoming flows have characteristics similar to the four critical 5G QoS Identifiers (5QIs) defined in Third Generation Partnership Project (3GPP) TS 23.501 V16.1.0. Also, we compared the performance of LEARNET with a baseline solution that respects the 5QIs priorities for allocating the incoming flows. The obtained results show that, for the scenario considered, LEARNET achieves a gain in the revenue of up to 45% compared to the baseline solution.**

I. Introduction

Time-Sensitive Networking (TSN) and Deterministic Networking (DetNet) are emerging and promising paradigms that come to satisfy the needs of many industries such as professional audio/video, electrical utilities, building automation systems, wireless industrial applications, and cellular transport networks, among others, for deterministic network services [1]–[3]. In other words, TSN and DetNet can establish a multi-hop path for a given flow over the network with deterministic Quality of Service (QoS) guarantees in terms of latency, jitter, congestion loss, and reliability, regardless of the other ongoing flows competing for the same resources [2]. To achieve that, these standards rely on the following functionalities: i) frame/packet sequencing, ii) flow replication/merging, iii) packet encoding/decoding, iv) explicit routes and v) resource reservation. Besides, the network devices considered in these standards for realizing the forwarding plane implement queuing algorithms that are mathematically analyzable. Then, we can derive analytical expressions for predicting the worst-case performance experienced by any flow given the current state of the network (e.g., ongoing flows).

This work addresses the online flow allocation problem (OFAP) in DetNet asynchronous networks. The OFAP consists of finding the optimal configuration for allocating each incoming flow to the network, given an optimization objective. Here, we focus on a DetNet network with an underlying IEEE 802.1 TSN network consisting of IEEE 802.1Qcr Asynchronous Traffic Shaper (ATS) based bridges [4]. The ATS is based on the Urgency-Based Scheduler (UBS) [5], which is an asynchronous queuing algorithm that combines interleaved shaping and strict priority queues to realize per-flow deterministic QoS guarantees (e.g., zero packet loss and bounded latency) in a practical way. In contrast to synchronous queuing algorithms such as Cyclic Queuing and Forwarding (CQF) and Time-Aware Shaper (TAS), asynchronous ones do not depend on network-wide coordinated time (higher scalability) and utilize network bandwidth more efficiently by leveraging statistical multiplexing [1].

Several works address the performance guarantees analysis for an ATS-based network [5]–[7]. However, we believe that the flow allocation problem for ATS-based networks is only tackled in [8] by Specht and Samii. Specifically, they deal with the allocation of a set of flows, considering the path between every source-destination pair is predefined. This process involves determining the assignment of flow-to-queue and queue-to-priority for each involved ATS, i.e., every ATS to be traversed by at least one flow. They employ two approaches to solve the problem: i) a pure Satisfiability Modulo Theories (SMT) solver, and ii) a heuristic method dubbed Topology Rank Solver (TRS) to cope with the combinatorial complexity of the problem. For the latter, they consider the maximization of the delay slack as the optimization goal.

In this work, we propose a Reinforcement Learning (RL)-based solution, which is dubbed LEARNET, for the OFAP in TSN asynchronous networks. The solution leverages data analytics, RL, and ATS performance models to maximize the long-term network operator's revenue. RL is envisioned as a promising approach for handling the complexity of the future networks [9]. We identify the main inputs (observations), potential actions, and reward for the RL agent. We also detail its environment that includes a flow admission control relying on the ATS analytical performance bounds. To the best of the authors' knowledge, this is the first paper proposing a RL-based solution for optimizing the operation of a TSN asynchronous forwarding plane.

The contribution of this article is twofold. First, we define a novel solution for the OFAP in asynchronous TSNs that integrates RL and the performance models of the ATS. Some authors have reported and highlighted a drastic reduction in the amount of data required to train the ML-based models by using approaches that combine data-driven techniques and

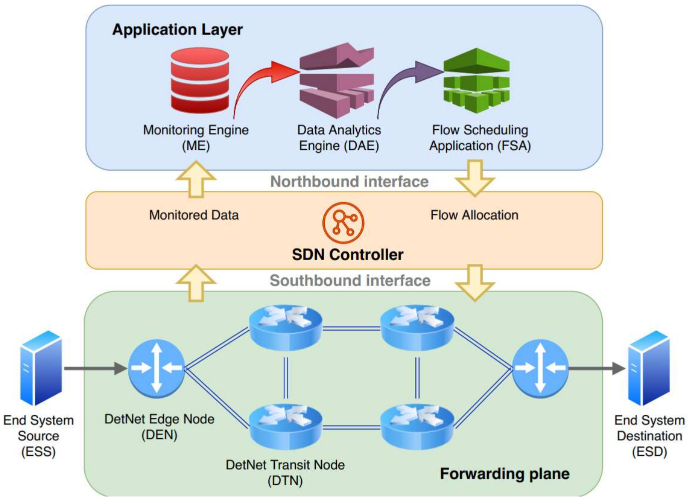

Fig. 1: SDN-like DetNet network architecture considered in this work.

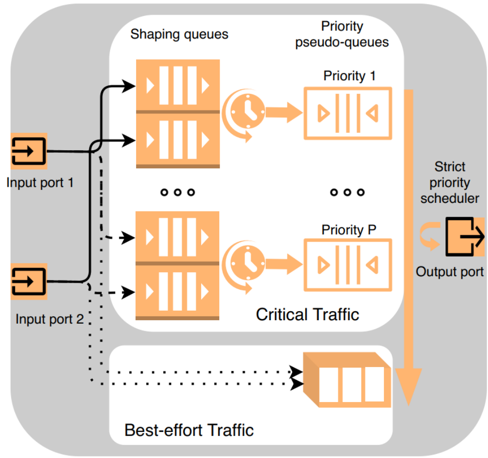

Fig. 2: UBS-based switch model. The illustrated UBS has four shaped queues and two priority levels and receives traffic from two ingress ports.

analytical models [10]. Our solution includes a flow admission control process, which relies on the ATS performance models, to check the feasibility of the actions issued by the agent. The flow admission control rejects the invalid actions, i.e., those that entail the violation of any constraint of the OFAP such as the fulfillment of the QoS requirements of both the ongoing flows and the incoming flow. In this way, the information of the analytical models is transferred to the agent, and, most importantly, the flow allocation process becomes fully reliable. Second, we show the feasibility and the potential benefits of applying RL for fully exploiting the flexibility offered by TSN asynchronous forwarding planes. We evaluate the performance of LEARNET in terms of the operator's revenue through simulation. The considered scenario is a fifth generation (5G) asynchronous deterministic backhaul network where incoming flows have characteristics similar to the four critical 5G QoS Identifiers (5QIs) defined in [11]. TSN and DetNet are appealing technologies for supporting network slicing [12], [13] at the transport network domain. Also, we compared the performance of LEARNET with a baseline solution that respects the 5QIs priorities. The results show that LEARNET achieves a gain in the revenue of up to 45% compared to the baseline solution.

The rest of the article is organized as follows. Section II describes the abstraction considered in this work for a deterministic network whose data plane consists of ATS-based network devices. Section III defines the online flow allocation problem (OFAP) considered in this work. Section IV details the RL-based flow scheduling solution proposed in this work. Section V includes the methods and the discussion of the obtained results. Last, Section VI concludes the paper.

II. System Model

A. DetNet Network Architecture

Let us consider a DetNet network with a Software-Defined Networking (SDN)-like architecture [14] where the Control Plane (CP) and Forwarding Plane (FP) are fully separated, see Fig. 1.

The FP comprises a set of V IEEE 802.1 TSN network devices that include an ATS at each of their interfaces. We will detail the operation of the ATS in the next subsection. There is a total of $E/2$ full-duplex point-to-point links interconnecting the asynchronous TSN nodes. In order to distinguish the different nodes and ATSs in the subsequent notation, we model the FP of the network as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_V\}$ and $\mathcal{E} = \{e_1, e_2, \ldots, e_E\}$ denote the vertices (representing the nodes) and the edges (representing the ATSs or simplex links) of the graph, respectively.

Here, we will distinguish between two types of FP nodes: i) DetNet Edge Node (DEN), and ii) DetNet Transit Node (DTN). The DEN might act as either the starting or the termination point of the deterministic flows. The main functionalities of the DEN include the addition or the removal of packet sequencing, and packet replication and combination. The DTNs will only implement the DetNet forwarding sub-layer, and are responsible for routing the packets from the DEN source to the DEN destination.

The CP consists of a logically centralized SDN controller that is responsible for controlling and monitoring the network devices through the southbound interface. The controller provides the upper application layer with an abstraction of the forwarding plane through the northbound interface.

Three main applications are running on top of the controller: i) the Monitoring Engine (ME), ii) the Data Analytics Engine (DAE), and iii) the Flow Scheduling Application (FSA). The ME is responsible for collecting relevant network statistics such as flow characteristics (e.g., delay constraint, rate demand, and holding time), flow arrival process to each DEN (as source and destination), and times-to-failure of the links, among others. Later, the DAE consumes and analyzes that information for prediction (e.g., estimation of the temporal workload profile). Finally, the FSA leverages that analyzed data for optimizing the flow scheduling process. Particularly, here, we consider the use of Machine Learning (ML) at the FSA to exploit the high degree of flexibility potentially offered by an ATS-based FP.

B. ATS-based Forwarding Plane

As mentioned in the previous subsection, the FP of the DetNet network consists of a set of TSN devices that implement an

IEEE 802.1 Qcr ATS [4] at each of their egress ports. The ATS relies on the UBS proposed by Specht and Samii [5], [8]. In this work, we assume that the ATS follows the operation described in [5], [8]. Then, an ATS consists of two levels queuing hierarchy (see Fig. 2) [5]: i) a set of shaped queues for interleaved shaping, and ii) one queue per priority level in the scheduler. All these queues follow a First Come, First Served (FCFS) discipline. The assignment of flows to shaped queues is subject to the following rules [5]: each shaped queue is associated with only one ingress port (*QAR1 rule*), one priority level in the previous ATS (*QAR2 rule*), and one internal priority level (*QAR3 rule*) at a given time. *QAR2* and *QAR3* rules are required to provide deterministic QoS, whereas *QAR1* isolates the flows from different nodes, avoiding the propagation of non-conformant traffic overloads. These rules also determine the required number of shaped queues to realize P priority levels.

The second stage in the queuing hierarchy includes one pseudo-queue per priority level in the scheduler. Each pseudo-queue merges the output of all shaped queues of the same priority level. The combining form pseudo points out that it is not required to implement them physically when there are few shaped queues [5]. Instead, the packets can be directly and efficiently transmitted from the shaped queues.

The decision on which packet to transmit relies on strict priority levels and the interleaved shaping algorithm considered. Here we will assume the Length Rate Quotient (LRQ) algorithm which enforces an upper bound on the flows of the form $A_f(t) \leq r_f \cdot t + b_f$ [5], [8], [15]. Where $A_f(t)$ is the accumulated amount of transmitted data until the instant t for the flow f, and r_f and b_f are the enforced sustainable rate and burstiness, respectively. LRQ algorithm computes the eligibility time t_f for the next packet of a given flow f as $t_f = l_f / r_f$, where l_f is the length of the head-of-line (HOL) packet of the flow. Then, the HOL packet of the flow f will be eligible for transmission after t_f time units have elapsed.

Under the considerations exposed above, the delay experienced by a packet $D_{f,p}^{(e_l)}$ of a given flow f when pass through an ATS-based link e_l and it has priority p is upper-bounded as [5]:

$$D_{f,p}^{(e_l)} \leq \frac{\sum_{z=1}^{p} b_z + \bigvee_{z=p+1}^{P_{e_l}} l_z}{C_{e_l} - \sum_{z=1}^{p-1} r_z} + \frac{l_f}{C_{e_l}} \tag{1}$$

Where b_z, l_z, and r_z are the aggregated burstiness, maximum packet size, and aggregated data rate at priority level z, respectively. And C_{e_l} denotes the link capacity.

III. Problem Statement

Let us consider the DetNet network with an ATS-based forwarding plane described in the previous section. Each FP node has an ATS at each interface e with $N_{SBs}^{(e)}$ shaping buffers. Consequently, the interface e will have $P_{max}^{(e)} = N_{SBs}^{(e)}$ priority levels at most. Let us also assume that the allocation of each incoming flow f, hereinafter referred as to flow of interest (foi), has an associated income α_f for the network operator. That income might be different for each flow depending on the type of flow or the flow characteristics, e.g., mean sustainable rate r_f, burstiness b_f, maximum packet length l_f, and maximum end-to-end (E2E) delay budget $D_{f,max}$. Besides, as in [8], we suppose that explicit routes between every pair of DENs source and destination in the network is predefined.

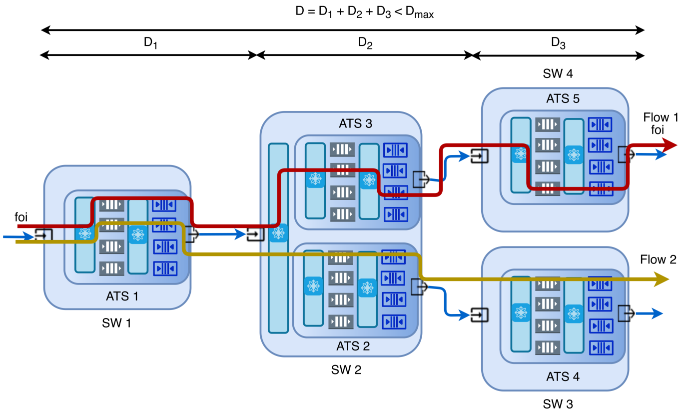

Fig. 3: Two flows are allocated in a network comprising four ATS-based nodes. All ATSs have four shaping queues, and, consequently, four priority levels at most.

The OFAP in an ATS-based FP considered here is defined as the process of choosing the allocation configuration for every incoming flow at every hop all along the predefined path from its source to its destination in order to maximize the network operator's revenue. The flow allocation configuration at every ATS implies to decide the flow to shaping buffer, priority level, and delay budget assignments.

As mentioned previously, the goal is to maximize the long-term operator benefit, which will mainly depend on the flow arrival process, the flow characteristics, pricing, and network setup. Observe that the flow arrival process and the stochastic characterization of the flow will likely vary over time. In this regard, predictive data analytics are ideally suitable for this problem.

A given flow allocation configuration for the foi is valid if only if the following constraints are met:

- C1 The E2E delay D experienced by the foi has to be lower than its E2E delay budget $D_{f,max}$.
- C2 The ongoing flows must keep experiencing an E2E delay lower than their respective E2E delay budgets.
- C3 The aggregated rate allocated to any link must be lower than its capacity.
- C4 The aggregated burstiness allocated to any shaping buffer has to be lower than its size.
- C5 The *QAR1*, *QAR2*, and *QAR3* rules have to be met, i.e., every shaping queue must be either idle or assigned to an only one priority level, only one priority level in the previous hop, and an input port at a given instant.

IV. LEARNET

In this section, we detail the proposed RL-based solution, which is dubbed LEARNET, for the flow scheduling in DetNet asynchronous networks. Figure 4 sketches the operation of LEARNET.

A. Observations

A flow allocation request arrives at the system (step 1 in Fig. 4). Then, the request is parsed in order to identify the foi

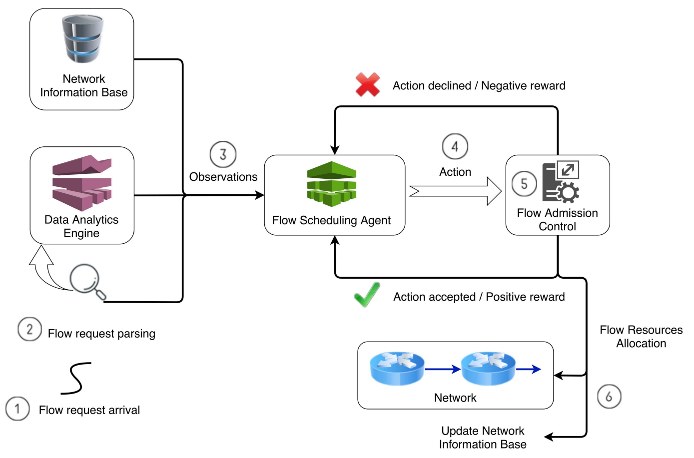

Fig. 4: LEARNET operation.

characteristics, resources demanded, and latency requirements. Specifically, LEARNET needs to know the sustainable rate $r^{(foi)}$, burstiness $b^{(foi)}$, maximum packet size $l^{(foi)}$, and E2E delay budget of the foi. The flow scheduling agent is fed with this information along with predictive analytics, related to the flow arrival and holding time processes, and the current FP network state (step 3 in Fig. 4).

Regarding the required data analytics, LEARNET requires: i) the expected foi lifetime duration $\tau^{(foi)}$, ii) the foreseen arrival rate λ during the next temporal interval of length $\tau^{(foi)}$, and iii) the predicted mean and standard deviation for the different features of the future incoming flows ($\overline{r}$, σ_r, $\overline{b}$, σ_b, $\overline{D}_{max}$, and $\sigma_{D_{max}}$). The derivation of those predictive analytics might also require the flow request information extracted during the parsing process. Please note that the employed predictive analytics reveal three main assumptions taken into account for designing LEARNET: i) the flow lifetime duration obeys an exponential distribution, ii) the flow arrival process is Poissonian, and iii) the flow arrival process and the features of the future incoming flows to the system are stationary. We might consider high-order statistics for the involved stochastic processes as well as their temporal dependence. In this way, we could remove the previous assumptions and improve the generality of the solution.

On the other hand, LEARNET needs the following network information: i) link capacities $C_{e_l}\,\forall\, e_l \in \mathcal{P}$; ii) current state of the shaping buffers at the different ATSs $\mathcal{P}$ in the foi path; and iii) the allocated rate $r_p^{(e_l)}$, allocated burstiness $b_p^{(e_l)}$, minimum delay budget $D_{p,max}^{(e_l)}$, and maximum packet length $l_p^{(e_l)}$ at each priority level of every ATS in the foi path.

B. Actions

The flow scheduling agent takes the information described in the previous subsection as input and produces an action specifying the flow allocation configuration for every ATS in the foi path (step 4 in Fig. 4). Next, we describe the format of the actions.

Let us suppose the predefined path for the foi has N_H hops. Then, the action generated by the agent is an N_H dimensional vector of real numbers $\boldsymbol{u}$, where each component includes two data: i) the chosen priority level $p_a^{(e_l)}$ for the foi at the respective hop (ATS) e_l, ii) the percentage of the total foi E2E delay budget $\eta^{(e_l)} = D_{p,max}^{(e_l)}/D_{max}^{(foi)}$ to be spent at the respective hop. Deciding the maximum delay budget to be spent at erery ATS makes scalable the actions feasibility checking, which is described in the next subsection. The former is included in the integer part of the component, whereas the latter is contained in the decimal part. More precisely, each component of $\boldsymbol{u}$ takes values from a finite set of real values in the interval $\left[1, P_{max}^{(e_l)}+1\right)$. The $P_{max}^{(e_l)}$ is set to the number of shaping queues in the ATS e_l. The decimal part is discretized according to a given granularity. For instance, a granularity of 0.1 means the percentage of the foi E2E delay budget assigned to each hop is 10% or a multiple of it.

Last, it is noteworthy that the set of actions is filtered in order to reduce its size. Specifically, we only consider the actions that meet the following constraints: *c.1*) the sum of the decimal parts of all components of $\boldsymbol{u}$ has to equal one, and *c.2*) the decimal part of every component of $\boldsymbol{u}$ has to be greater than zero. Constraint *c.1* enforces that the foi E2E delay budget is fully consumed along the path. This constraint helps to improve the flow acceptance ratio and, hence, the operator's revenue. By way of illustration, let us assume a path with three hops and $P_{max}^{(e_l)} = 3$ for every ATS. The action vectors $\boldsymbol{u} = (1.5, 3.5, 1.0)$ and $\boldsymbol{u} = (1.2, 3.5, 1.2)$ would be removed from the action set, whereas $\boldsymbol{u} = (2.2, 3.5, 1.3)$ would be valid.

C. Flow Admission Control

The action generated by the agent, along with the foi characteristics, are forwarded to the flow admission control. This block is responsible for accepting or rejecting the flow given the action provided by the agent. To that end, it checks the fulfillment of the constraints listed in Section III (C1-C5). The constraints C1-C3 are checked evaluating the following set of inequalities (please refer to Table I to look up the notation employed):

$$\begin{aligned}&\frac{\sum_{z=1}^{p_{e_l}} b_z^{(e_l)} + b^{(foi)} + \bigvee_{z=p+1}^{P_{max}^{(e_l)}} l_z^{(e_l)}}{C_{e_l} - \sum_{z=1}^{p_{e_l}-1} r_z^{(e_l)}} \leq \left(D_{max}^{(foi)} - \frac{l^{(foi)}}{C_{e_l}}\right)\\ &\wedge \left(D_{p_{e_l},max}^{(e_l)} - \frac{l_{p_{e_l}}^{(e_l)}}{C_{e_l}}\right); \qquad \forall\, e_l \in \mathcal{P} \;\&\; p_{e_l} = p_a;\end{aligned} \tag{2}$$

$$\begin{aligned}&\frac{\sum_{z=1}^{p_{e_l}} b_z^{(e_l)} + b^{(foi)} + \bigvee_{z=p+1}^{P_{max}^{(e_l)}} l_z^{(e_l)}}{C_{e_l} - \sum_{z=1}^{p_{e_l}-1} r_z^{(e_l)} - r^{(foi)}} + \frac{l_{p_{e_l}}}{C_{e_l}} \leq D_{p_{e_l},max}^{(e_l)};\\ &\forall\, e_l \in \mathcal{P} \;\&\; p_{e_l} > p_a;\end{aligned} \tag{3}$$

$$\begin{aligned}&\frac{\sum_{z=1}^{p_{e_l}} b_z^{(e_l)} + \left(\bigvee_{z=p+1}^{P_{max}^{(e_l)}} l_z^{(e_l)}\right) \bigvee l^{(foi)}}{C_{e_l} - \sum_{z=1}^{p_{e_l}-1} r_z^{(e_l)}} + \frac{l_{p_{e_l}}^{(e_l)}}{C_{e_l}}\\ &\leq D_{p_{e_l},max}^{(e_l)}; \qquad \forall\, e_l \in \mathcal{P} \;\&\; p_{e_l} < p_a;\end{aligned} \tag{4}$$

$$\sum_{z=1}^{P_{max}^{(e_l)}} r_z^{(e_l)} + r^{(foi)} \leq C_{e_l}; \qquad \forall\, e_l \in \mathcal{P}; \tag{5}$$

Please note that, in the constraints listed above, we are considering that lower indexes correspond to higher priority levels. It is also noteworthy that thanks to deciding and recording the maximum delay budget spent at every ATS for every

TABLE I: Notation used for describing the operation of LEARNET.

Notation	Description
	Variables
$r^{(foi)}$, $b^{(foi)}$, $l^{(foi)}$, $D_{max}^{(foi)}$	foi characteristics: Sustainable rate, burstiness, maximum packet length, and delay budget.
$r_p^{(e_l)}$, $b_p^{(e_l)}$, $l_p^{(e_l)}$, $D_{p,max}^{(e_l)}$	Aggregated rate, aggregated burstiness, maximum packet length, and the lowest delay budget to be met in the priority level p at ATS e_l.
C_{e_l}	Link capacity at the ATS e_l.
B_s	Size of the shaping buffer s.
$I_{foi}^{(e_l)}$	Ingress port of the foi at the node that includes the ATS e_l.
$p_a^{(e_l)}$	Chosen priority level for the foi at the ATS e_l.
	Sets
$\mathcal{E}$	Set of all ATSs of the DetNet network FP.
$\mathcal{P} \subseteq \mathcal{E}$	Subset of ATSs the foi has to pass through from its DEN source to its DEN destination
$\mathcal{S}_{e_l}$	Pool of shaping buffers at ATS e_l.
	Operators
$\bigvee_{n=1}^{N} a_n$	The greatest value of the set of elements $\{a_n\}_{n=1}^{N}$ ($a_n \in \mathbb{R} \, \forall \, n \in [1, N]$).
$a \bigwedge b$	Minimum of a and b ($a, b \in \mathbb{R}$).

incoming flow, the checking of the constraints C1-C3 is scalable. Otherwise, we will have to check that the network will keep guaranteeing the performance requirements for all the affected ongoing flows after the potential foi allocation.

Algorithm 1 performs the verification of constraints C4-C5 at a given ATS e_l and the foi to shaping buffer assignment at each hop. The algorithm looks for a busy shaping buffer with enough capacity and a valid state. In other words, the shaping buffer is assigned to the same input port, internal priority level $p_a^{(e_l)}$, and priority level $p_a^{(e_{l-1})}$ in the previous hop e_{l-1} as the foi allocation configuration commanded by the action. If not, the algorithm checks whether there is any idle shaping buffer to allocate the flow.

Algorithm 1 Shaping buffers related constraints verification and allocation process for the foi.

Input: $I_{foi}^{(e_l)}$ and $p_a^{(e_l)}$ $\forall p, i \in [1, \cdots, P_h], j \in [1, \cdots, P_{h-1}], \& h \in \mathcal{P}$.
Output: $s_c^{e_l}$ and $d \in$ [ACCEPTED, REJECTED].

```
if there is any busy s ∈ S_{e_l} such that its input port equals I_{foi}^{(e_l)}, its priority level equals p_a^{(e_l)}, its previous priority level equals p_a^{(e_{l-1})}, and B_s is enough to accommodate b^{(foi)} then
    d = ACCEPTED;
    s_c^{e_l} = ChooseBusyBuffer();
else
    if there is any idle s ∈ S_{e_l} then
        d = ACCEPTED;
        s_c^{e_l} = ChooseIdleBuffer();
    else
        d = REJECTED;
    end if
end if
```

If any of the constraints C1-C5 is not met, the admission control block will decline the flow allocation request and will penalize the agent with a negative reward. Otherwise, the flow will have granted access, and a positive reward will be awarded to the agent.

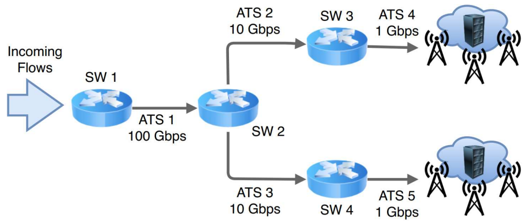

Fig. 5: Simulated Scenario.

D. Reward

If the flow is accepted, the agent will receive an award $1/\tau^{(foi)}$ times the income $\alpha^{(foi)}$ that the operator will obtain for allocating the foi. If the flow is rejected, the agent will be penalized with a reward of $-1/\tau^{(foi)} \cdot \alpha^{(foi)}$. Observe that, if we set $\alpha^{(foi)} = 1$ for all the incoming flows, we will maximize the flow acceptance ratio (flow rejection ratio minimization).

E. Network State Update

Finally, if the flow gets granted access, the allocation setup will take place and the state of the network FP will be updated accordingly (step 6 in Fig. 4). Specifically, $\forall \, e_l \in \mathcal{P} \; \& \; p = p_a^{(e_l)}$, $r_p^{(e_l)} \leftarrow r_p^{(e_l)} + r^{(foi)}$; $b_p^{(e_l)} \leftarrow b_p^{(e_l)} + b^{(foi)}$; $D_{p,max}^{(e_l)} \leftarrow D_{p,max}^{(e_l)} \wedge \eta^{(e_l)} \cdot D_{max}^{(foi)}$; and $l_p^{(e_l)} \leftarrow l_p^{(e_l)} \vee l^{(foi)}$. Besides, the state of the chosen shaping buffers whose state were idle when the foi arrived will be modified from idle to busy, and the buffer will be associated with the respective input port, internal priority level, and priority level in the previous hop.

One last remark, a kind of flow information base is required to keep track of the allocation configuration and characteristics of every ongoing flow in order to properly update the network FP state when the flow leaves the network.

V. Results

A. Experimental Setup

The performance evaluation of LEARNET was carried out by using an event-driven simulator of a 5G DetNet Backhaul Network (BN) with three hops (ATSs) between the source DEN and the destination DENs (see Fig. 5). We considered the flow characteristics of the critical 5QIs defined by Third Generation Partnership Project (3GPP) (see Table II). The actual data rate demanded by each simulated incoming flow follows a Gaussian distribution. The mean of that distribution for each 5QI is included in the third column of Table II, and the standard deviation was set to 15% of the respective mean. The flow lifetime and flows inter-arrival times obey an exponential distribution. The link capacities were set to 100 Gbps, 10 Gbps, and 1 Gbps for the first, second, and third hop, respectively. Like the scenario depicted in Fig. 3, every ATS in the FP includes four shaping buffers and, then, four potential priority levels.

We compared the performance achieved by LEARNET with a baseline solution that respects at every ATS the 5QIs priorities defined by 3GPP in [11]. Then, the fois of 5QIs 82, 85, 83, and 84 will be assigned at every ATS to priority level 1, 2, 3, and

TABLE II: Flow types characteristics. Most of the data included in this table were extracted from [11].

5QI	Prio	Rate (Mbps)	Burstiness (bits)	Dmax (ms)	Income	Avg. Dur. (s)	Lmax (bits)	Ex. service
82	19	0.1	2040	10	2.5	1200	2040	Discrete Automation
83	22	0.2	10832	10	2.5	1200	10832	Discrete Automation
84	24	0.3	10832	30	4	1200	10832	Intelligent transport systems
85	21	0.3	2040	5	3	1200	2040	Electricity distribution HV

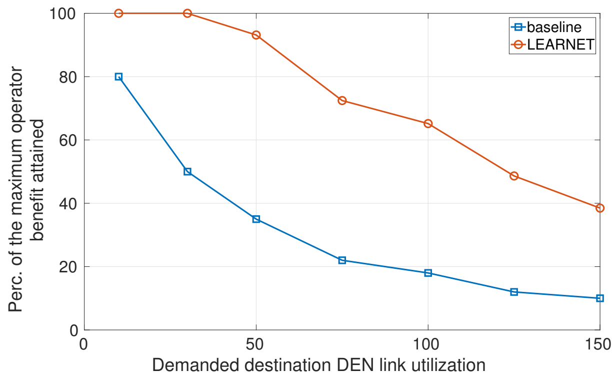

Fig. 6: Percentage of the maximum benefit achieved versus the link utilization of the destination DEN.

4, respectively. Besides, the baseline solution allocates 33% of the E2E delay budget to each hop in the path.

B. LEARNET Performance

Figure 6 shows the percentage of the maximum attainable profit achieved for each solution as a function of the demanded link utilization at the edge. We considered that the aggregated demanded data rate for each 5QI is the same on average as a simple criterion to generate the 5QI of each simulated flow. Moreover, we used a discrete uniform distribution to choose the destination DEN for each simulated flow in the scenario depicted in Fig. 5. In every simulation (every point represented in Fig. 6), we simulated the arrival and departure of one million of flows. As observed, LEARNET outperforms the baseline solution for every destination DEN link utilization considered. Specifically, it achieves a gain in operator benefit up to 45% compared to the baseline solution (see Fig. 6). It shall be noted that we checked out that LEARNET met all the time the delay constraints of all the flows. The flow admission control of LEARNET enforces the fulfillment of the flow performance requirements (see Fig. 4). This block enhances the reliability of the LEARNET, which is crucial for supporting critical flows.

VI. Conclusion

In this article, we have proposed a reinforcement learning-based solution, which is dubbed LEARNET, for the online flow allocation in deterministic asynchronous networks. The solution combines data-driven and analytical model-based approaches to maximize the network operator's revenue. We have evaluated the performance of LEARNET through simulation in terms of operator's achieved income. The considered scenario is a 5G deterministic backhaul where incoming flows have characteristics similar to the four critical 5G QoS Identifiers (5QIs) defined by 3GPP. In addition, we compared the performance of LEARNET with a baseline solution that preserves the 5QIs priorities. The obtained results show that LEARNET achieves a gain in the revenue of up to 45% compared to the baseline solution. These results motivate the investigation of machine-learning approaches to exploit the flexibility offered by TSN and DetNet networks fully.

Acknowledgment

This work is partially supported by Business Finland 5G-FORCE project and the Academy of Finland 6Genesis and CSN projects with grant agreement No. 318927 and No. 311654, respectively.

References

[1] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. ElBakoury, "Ultra-low latency (ull) networks: The ieee tsn and ietf detnet standards and related 5g ull research," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 88–145, Firstquarter 2019.

[2] N. Finn and P. Thubert, "Deterministic Networking Problem Statement," Internet Engineering Task Force (IETF), RFC 8557, May 2019, category: Informational. [Online]. Available: https://tools.ietf.org/html/rfc8557

[3] E. Grossman, "Deterministic Networking Use Cases," Internet Engineering Task Force (IETF), RFC 8578, May 2019, category: Informational. [Online]. Available: https://tools.ietf.org/html/rfc8578

[4] J. Specht, "IEEE Draft Standard for Local and Metropolitan Area Networks–Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks Amendment: Asynchronous Traffic Shaping," *IEEE P802. 1Qcr D*, 2017.

[5] J. Specht and S. Samii, "Urgency-based scheduler for time-sensitive switched ethernet networks," in *2016 28th Euromicro Conference on Real-Time Systems (ECRTS)*, July 2016, pp. 75–85.

[6] J.-Y. Le Boudec, "A theory of traffic regulators for deterministic networks with application to interleaved regulators," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2721–2733, Dec. 2018.

[7] E. Mohammadpour, E. Stai, M. Mohiuddin, and J. Le Boudec, "Latency and backlog bounds in time-sensitive networking with credit based shapers and asynchronous traffic shaping," in *2018 30th International Teletraffic Congress (ITC 30)*, vol. 02, Sep. 2018, pp. 1–6.

[8] J. Specht and S. Samii, "Synthesis of queue and priority assignment for asynchronous traffic shaping in switched ethernet," in *2017 IEEE Real-Time Systems Symposium (RTSS)*, Dec 2017, pp. 178–187.

[9] C. Benzaïd and T. Taleb, "AI-driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions," *IEEE Network*, 2020. [Online]. Available: http://www.mosaic-lab.org/uploads/papers/20edfc7e-02db-4fad-9b63-d584342130d2.pdf

[10] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, ai-based, or both?" *IEEE Trans. on Commun.*, vol. 67, no. 10, pp. 7331–7376, Oct 2019.

[11] 3GPP TS 23.501 V16.1.0. (2019) Technical Specification Group Services and System Aspects; System Architecture for the 5G System.

[12] T. Taleb, B. Mada, M. Corici, A. Nakao, and H. Flinck, "Permit: Network slicing for personalized 5g mobile telecommunications," *IEEE Commun. Magazine*, vol. 55, no. 5, pp. 88–93, May 2017.

[13] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, thirdquarter 2018.

[14] J. Prados-Garzon, O. Adamuz-Hinojosa, P. Ameigeiras, J. J. Ramos-Munoz, P. Andres-Maldonado, and J. M. Lopez-Soler, "Handover implementation in a 5g sdn-based mobile network architecture," in *IEEE 2016 27th Annual Int. Symposium on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–6.

[15] Z. Zhou, Y. Yan, M. Berger, and S. Ruepp, "Analysis and modeling of asynchronous traffic shaping in time sensitive networks," in *14th IEEE Int. Workshop on Factory Commun. Syst. (WFCS)*, June 2018, pp. 1–4.