







# DecAge: Decentralized Flow Scheduling for Industrial 5G and TSN Integrated Networks

Mingyan Li , Songtao Guo , Senior Member, IEEE, Chao Chen , Senior Member, IEEE, Cailian Chen , Senior Member, IEEE, Xiaofeng Liao , Fellow, IEEE, and Xinping Guan , Fellow, IEEE

**Abstract**—The integration of fifth-generation (5G) and time-sensitive networking (TSN) is a key technology to promote the wireless upgrade of manufacturing industry, and some preliminary enhancements on 5G with TSN-integration have been introduced to empower industrial internet ubiquitous connectivity and real-time forwarding. However, at present, the 5G system is only served as a logical bridge with 5G-specific procedures hidden from TSN networks, and thus the deep integration of 5G and TSN is still an open issue. This article focuses on the co-design problem of flow scheduling in 5G and TSN, where 5G grant-free configured-grants (CGs) and TSN cyclic queuing and forwarding mechanisms are adopted. Firstly, we innovatively propose to achieve seamless 5G+TSN transmission by intelligently selecting CGs in 5G, where the starting time offsets of different CGs will influence the flow injection time in the downstream TSN network to mitigate the queue overflow issue. Then, an age-of-information (age) aware decentralized and deterministic scheduling (DecAge) scheme based on actor-critic reinforcement learning is proposed. DecAge guides time-sensitive flow hosts to autonomously make packet sampling-transmission decisions so that timeliness-based age constraints can be respected via collision-free 5G+TSN transmission. DecAge contains multiple local actors learning hierarchically-structured scheduling policies for flow hosts and a central graph attention network-based critic to estimate these policies. Through simulation, DecAge shows the robustness to environmental non-stationarity by elaborately coordinating distributed learning and yields improved performance over non-hierarchical agents on sample-efficient learning.

**Index Terms**—5G+TSN flow scheduling, grant-free CGs, cyclic queuing and forwarding, multi-agent reinforcement learning.

Manuscript received 29 January 2023; revised 5 June 2023; accepted 29 July 2023. Date of publication 9 August 2023; date of current version 8 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grants 62203077, 62272069 and 62172066, in part by the Natural Science Foundation of Chongqing under Grant CSTB2022NSCQ-MSX1029, in part by the Special Research Funding for Chongqing Postdoctoral Researchers under Grant 2021XM2015, in part by the Fundamental Research Funds for the Central Universities under Grants 2022CDJXY-020 and 2023CDJXY-038, and in part by the Open Project of Key Laboratory of System Control and Information Processing, Ministry of Education. Recommended for acceptance by Dr. Tie Qiu. (Corresponding author: Songtao Guo.)

Mingyan Li is with the College of Computer Science, Chongqing University, Chongqing 400044, China, and also with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200241, China (e-mail: limy2021@cqu.edu.cn).

Songtao Guo, Chao Chen, and Xiaofeng Liao are with the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: guosongtao@cqu.edu.cn; cschaochen@cqu.edu.cn; xfliao@cqu.edu.cn).

Cailian Chen and Xinping Guan are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: cailianchen@sjtu.edu.cn; xpguan@sjtu.edu.cn).

Digital Object Identifier 10.1109/TNSE.2023.3301879

## I. INTRODUCTION

INDUSTRIAL Internet of Things (IIoT) is an important support for the fourth industrial revolution, where time-sensitive networking (TSN) technology can promise the deterministic interaction of time-critical tasks based on a series of strict traffic scheduling protocols [1]. To further realize ubiquitous sensing and flexible manufacturing, the standardized enhancement of the fifth-generation (5G) New Radio has been proposed in 3GPP Release 16 (R16) for TSN requirements in the scenario of IIoT. By introducing time-sensitive communication, the 5G system serves as a logical bridge to provide TSN Ethernet ingress and egress ports on the user plane using the “Translator” functions [2], as shown in Fig. 1. However, TSN relies on time-domain access whereas 5G exploits frequency-domain user multiplexing. This scheduling resolution mis-match calls for hold-and-forward mechanisms at the edge of 5G system, which impede the seamless transmission of 5G and TSN.

A key way to integrate 5G and TSN is re-using the ultra-reliable and low-latency communication (URLLC) framework, where the grant-free (GF) transmission has been introduced in R16 for 5G uplinks based on multiple configured-grants (CGs) transmission resources with different starting time offsets [3]. In this way, URLLC users can autonomously choose any of these pre-configured CGs to transmit as soon as data arrives. Compared with the request-followed-by-grant procedure in the conventional centralized dynamic scheduling, this arrive-and-go manner can largely reduce transmission latency [4]. Due to the decentralized nature of GF transmission, users compete CG resources randomly, leading to blind detection of active users in the base station (BS). As a result, transmission conflict may happen. Hence, non-orthogonal multiple access (NOMA) has been proposed to synergize with GF transmission to mitigate network congestion, referred to as GF-NOMA [5], [6], [7], [8], [9], [10].

For the 5G+TSN integrated networks, GF-NOMA enhancements are further required to guarantee the end-to-end timeliness constraints, which calls for the consideration of TSN scheduling. As an ethernet protocol extension developed by IEEE 802.1 TSN task group, TSN introduces many time-aware traffic shapers like time-aware shaper (TAS) to guarantee the deterministic forwarding for periodic time-sensitive flows [11]. TAS assigns time slots to each time-triggered flow and each flow transmits data only during its assigned time slot. Therefore, it requires fine-grained configurations of the gate control list (GCL) attached to

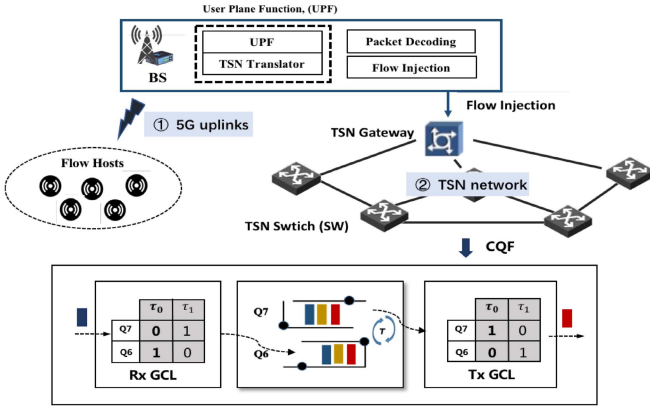


Fig. 1. Illustration of 5G+TSN integrated networks.

each queue in a TSN switch to enforce temporal isolation of flows [12].

To simplify the design of a TSN switch, a Ping-Pong queue-based model named cyclic queuing and forwarding (CQF) was proposed in IEEE 802.1 Qch for aperiodic and periodic flows. CQF installs static configurations on GCL to assign cyclic en-queue and de-queue operations, as shown in Fig. 1 [13]. As a result, the complex time scheduling for each queue can be ignored, and multiple flows are aggregated into the same queue without considering the input and output time sequences of packets as long as the usage of queues on the path of these flows does not exceed their capacity. However, many flows would converge in parts of queues if being forwarded without an elaborated planning of sending time [14]. Therefore, several injection time planning strategies were proposed based on incremental heuristics to avoid queue overflow [15], [16], [17], [18]. By delaying the sending time slots of flows with offsets, a balanced queue utilization can be achieved. Motivated by these CQF-based works, we provide a novel idea for the studies on seamless and collision-free 5G+TSN transmission by jointly designing the time offsets of GF access in 5G and flow injection in TSN. To the best of our knowledge, this is the first work to deal with the flow scheduling resolution mis-match issue for 5G+TSN integration.

#### A. Bi-Objective for Age-Aware 5G+TSN Flow Scheduling

This article addresses the flow scheduling co-design problem of 5G and TSN for industrial monitoring services with status update traffic flows. For timely status update, the age-of-information, referred to as age, is optimized, which captures the timeliness from the perspective of destinations [19]. Age is a crucial metric for industrial services to assist remote controllers taking up-to-date decisions. Different from the metric of delay or latency, age captures the information freshness from destinations' perspective rather than packets' perspective. Specifically, age optimization relies on both packet sampling and transmission decisions of flow hosts (a.k.a. flows), which are generally referred to as scheduling actions. Therefore, in our problem, multiple age-aware flows make scheduling actions

based on the principles of 5G GF-NOMA and TSN CQF under the following bi-objective objective.

*Age-constraint guarantee with minimal power:* The primal objective is to respect flows' age constraints and minimize the sampling-transmission energy consumption. To this end, the secondary transmission objective should be achieved if a packet is sampled.

*Collision-free 5G+TSN transmission:* Flow transmission in both 5G and TSN should be jointly optimized by elaborately selecting the grant-free CGs so that,

- flows can be successfully detected and decoded in 5G;
- based on the no-wait injection principle in the BS, flows can be mapped onto the underlying orthogonal resources of Ping-Pong queues in TSN switches both temporally and spatially to avoid queue overflow.

#### B. Challenge and Contribution

*Challenge:* This age-aware flow scheduling co-design problem is challenging due to the following aspects:

- Stochasticity:* The outcome of age evolution is stochastic under environment uncertainty (unknown channel state and communication interference). Therefore, flows need to consider the expected outcome when deciding which action to execute so as to respect their age constraints.
- Decentrality:* Due to the decentralized nature of GF transmission, individual flows autonomously access the network without any priority of planning, where scheduling policies are optimized with partial observability. As a result, the state transition of a flow is uncertainty due to other concurrently-exploring flows, and thus the environment is perceived as non-stationary.
- Complex reasoning:* According to the above bi-objective, flows infer the effect of scheduling actions from the feedback of age, which is a coupled result of sampling and transmission decisions. Moreover, the transmission performance of 5G and TSN should be jointly optimized by just controlling the operating CGs.

*Contribution:* To deal with these challenges, an age-aware decentralized and deterministic scheduling (DecAge) scheme based on deep reinforcement learning (DRL) is proposed for 5G+TSN integrated networks, and the major contributions of this article are summarized as follows.

- *Age-aware co-design of 5G+TSN scheduling:* To solve Challenge (i), an age-aware flow deterministic scheduling (AgeDS) problem is formulated with flow age constraints and TSN queue constraints. In this problem, we firstly propose to jointly design the collision-free GF access of 5G and flow injection of TSN by optimizing the selection of CGs based on their time slot offsets. Then, primal-dual learning is utilized to deal with environment uncertainty and stochasticity in the dual domain of AgeDS problem.
- *Decentralized multi-agent RL with GAT-based critic:* The challenge of non-stationary non-Markovian environments, i.e., Challenge (ii), may cause instability learning. Based on the centralized training with decentralized execution

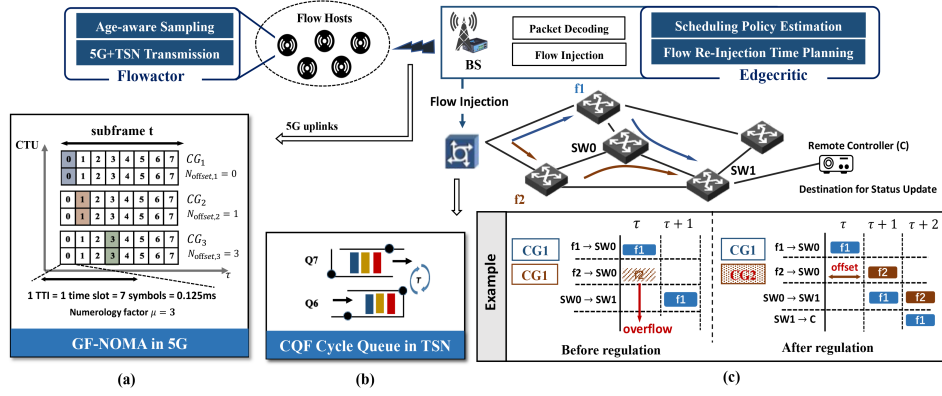


Fig. 2. Illustration of age-aware 5G+TSN scheduling co-design with an example.

TABLE I  
THE LIST OF KEY ABBREVIATIONS

URLLC	ultra-reliable and low-latency communication
GF, NOMA	grant-free, non-orthogonal multiple access
CG, BS, RB	configured-grant, base station, resource block
CTU, MA	contention transmission unit, multiple access signature
CQF	cyclic queuing and forwarding
DRL, DQN	deep reinforcement learning, deep Q network
CTDE	centralized training with decentralized execution
GAT, AAG	graph attention network, augmented adjacency graph

TABLE II  
THE LIST OF KEY NOTATIONS

NOT	DEFN	NOT	DEFN
$i, j$	flow index	$p$	CG index
$g_{i,t}$	age of the flow $f_i$	$r_{i,t}$	packet received indicator
$\phi_{i,t}$	scheduling action	$\kappa_{i,t}, C_{i,t}$	sampling and CTU
$\zeta_{i,t}$	SINR	$q_{e,t}$	queue length
$\lambda$	dual variable	$\mathcal{L}$	dual function or value
$Q(s, a)$	Q network	$\pi_i$	scheduling policy
$\alpha_i, \beta_i$	update stepsizes	$\eta_i$	learning rate of $Q$
$S_i$	adjacent flow set	$\omega_{i,j}$	attention factor

paradigm in multi-agent RL, a central critic with variance-reducing baselines is used in DecAge to estimate the performance of distributed scheduling policies from a global and long-term perspective, which is trained by the graph attention network (GAT) with an augmented adjacency graph. It considers the flows that may impose interference in both 5G and TSN so that the critic can dynamically attend to the relevant information from them.

- *Hierarchical scheduling with adjacent state awareness:* According to Challenge (iii), sampling-transmission actions are jointly optimized. As a result, scheduling policies are learned with sparse rewards of communication. Based on this hierarchy, two-tiered sample-efficient architecture is proposed, containing an age-aware high-level sampling policy and a communication-oriented low-level skill. The former outputs the transmission desire, serving as a sub-goal for the latter. Moreover, local observation is expanded with state prediction of adjacent flows, which further coordinates the behavior of distributed flows.

The key abbreviations and notations are listed in Table I and Table II, respectively. The remainder of this paper is organized as follows. We firstly introduce the 5G+TSN integrated network as well as the scheme overview in Section II. Then, the system model and problem formulation are proposed in Section III with the detail of DecAge introduced in Section IV. The simulation is provided in Section V and Section VI discusses the related work of GF-NOMA. Finally, Section VII concludes this article.

## II. OVERVIEW

### A. 5G+TSN Integrated Network

The 5G+TSN transmission is composed of GF-NOMA uplink transmission in 5G and multi-hop CQF-based forwarding in

TSN. Firstly, the 5G New Radio frame structure and GF-NOMA with multiple CGs are introduced.

*Frame structure and GF-NOMA:* From the temporal dimension, the transmission latency of 5G can be reduced by shortening transmission time interval (TTI) with mini-slots. Depending on the numerology factor  $\mu$ , a subframe, indexed by  $t$ , can have  $2^\mu$  time slots (indexed by  $\tau$ ), as depicted in Fig. 2(a) with  $\mu = 3$ .<sup>1</sup> From the spatial dimension, distributed flows for status sensing are located in the Euclidean plane and their GF uplink transmissions are taken place over a shared spectrum where they may interfere with each other. To mitigate flow collision in GF transmission, the BS pre-configures multiple CGs at each subframe, mainly featured by different resource starting offsets in the unit of time slots. The BS detects flows and decodes their packets in each subframe using NOMA so that the number of served flows can be improved.

*CQF-based forwarding:* According to CQF, there are two Ping-Pong queues in each TSN gateway and switch (e.g., time-sensitive queues Q6 and Q7 in the Fig. 2(b)) performing enqueue and de-queue operations in a cyclic manner. In this way, the packet enqueued in the previous slot will be dequeued in the current slot and received in the next-hop TSN switch at the same time. Based on early-stage analysis, flow injecting time to the TSN network has a critical effect on queue utilization of switches. Hence, for seamless 5G+TSN transmission, the operating CG of flows in the upstream 5G system should be carefully designed to avoid queue overflow in TSN.

<sup>1</sup>With the subcarrier spacing  $\Delta f = 2^\mu \times 15\text{kHz}$ , where  $\mu = 0, 1, 2, 3, 4$ , the slot duration reduces according to  $1/2^\mu$  (ms) [7].



*Example:* To stress it clear, we build an example with queue length set to 1 in Fig. 2(c) where two time-sensitive flows ( $f_1$  and  $f_2$ ) have an overlapping path ( $SW_0 \rightarrow SW_1$ ). If there is no slot offset between  $f_1$  and  $f_2$ , the packet of  $f_2$  will be dropped due to the limited queue length. Nevertheless, if the flow host of  $f_2$  choose to transmit through  $CG_2$  with one slot offset, all packets can reach  $SW_1$  successfully. As the resource utilization of both 5G and TSN improves, the number of flows respecting their age constraints grows up.

### B. Scheme Overview

To this end, we propose our scheduling scheme, DecAge, with its framework delineated in Fig. 2. DecAge consists of multiple local Flowactors and an Edgecritic. On the one hand, Flowactors are in charge of flow scheduling decision making for their corresponding flow hosts, i.e., packet sampling for timely status update (age-aware sampling) and 5G+TSN transmission if a packet sampled. On the other hand, the central Edgecritic is located in the BS with edge computing servers, where the BS decodes packets received via wireless 5G GF uplinks and injects them into the TSN network. Edgecritic is responsible for estimating the scheduling policies of local Flowactors from a global long-term view and re-injecting the packets if overflow events have been foreseen based on the deterministic forwarding property of TSN.

## III. SYSTEM MODEL AND PROBLEM FORMULATION

Before introducing the detail of DecAge, the system model and problem formulation are proposed in the next.

### A. System Model

Consider multiple time-sensitive monitoring systems in an integrated 5G+TSN deployment of industrial environments, where aperiodic age-aware flows are generated by the corresponding hosts, denoted as the flow set  $\mathcal{F}$  with  $|\mathcal{F}| = F$ . Each flow  $f_i \in \mathcal{F}$  generates status packets requiring stringently fresh data. It is captured by the metric of age, which measures the time elapsed since the generation of the packet that was most recently delivered to destination [19]. Formally, the age of a monitoring system (a.k.a. a flow  $f_i$  for simplicity) at subframe  $t$  is defined as  $g_{i,t}$ . Let each flow  $f_i$  be characterized by an age constraint with age limit  $f_i.age$ . Besides, the flow feature also consists of source and destination host IDs and the location of source, as well as the pre-configured 5G multiple access signature (MA) and TSN multi-hop path, denoted as,

$$\begin{aligned} \forall f_i \in \mathcal{F}, i \in \{0, \dots, F-1\}, \\ f_i = \{age, src, dst, loc, ma, path\}. \end{aligned} \quad (1)$$

We consider non-buffer monitoring systems and the generate-at-will strategy is employed such that when  $f_i$  decides to transmit at a certain subframe, it samples the packet at the beginning of that subframe and immediately sends this packet to the remote  $f_i.dst$ . Since the freshest data is sampled, this strategy can largely decrease the age of system and the frequency of packet sampling and transmission for energy saving and low co-channel

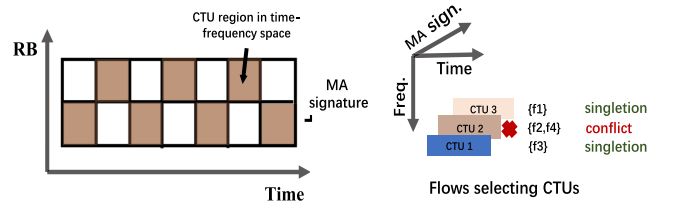


Fig. 3. Illustration of CTU in a time-frequency space.

interference. Then, the age evolution of  $g_{i,t}$  at per-subframe granularity depends on whether a update packet is received within a subframe  $t$ , i.e.,  $r_{i,t} \in \{0, 1\}$ , and thus can be written as,

$$g_{i,t} = \begin{cases} g_{i,t-1} + 1, & \text{if } r_{i,t} = 0, \\ 1, & \text{if } r_{i,t} = 1, \end{cases} \quad (2)$$

where  $g_{i,0} = 0$ . Note that the indicator value  $r_{i,t}$  relies on both the sampling decision and successful transmission probability.

### B. Grant-Free Multiple Configured-Grants NOMA in 5G

For massive low-latency uplink transmissions, GF-NOMA with multiple CGs is considered, where the BS pre-configures  $N_{CG}$  CGs at each subframe  $t$  with the CG set defined as  $\mathcal{N}_{CG}$ . Then, the BS sends radio resource control or downlink control information to activate the static CG configuration [3], [20]. Each CG has the fixed parameter tuple as

$$\begin{aligned} \forall CG_p \in \mathcal{N}_{CG}, p \in \{0, \dots, N_{CG} - 1\}, \\ CG_p = \{N_{RB}, N_{offset}, N_{repe}\}. \end{aligned} \quad (3)$$

- $N_{RB}$  denotes the number of resource blocks (RBs) in this CG.<sup>2</sup>Based on this time-frequency resource, the minimal resource that flows compete for is the contention transmission unit (CTU), which is represented by a RB and an MA signature, as shown in Fig. 3. If multiple flows transmit with the same CTU, the BS cannot differentiate and detect them, and thus cannot decode their packets.
- $N_{offset}$  is the starting time slot of CG within a subframe.
- $N_{repe}$  is the number of repetitions.<sup>3</sup> We assume  $N_{repe} = 1$  for problem simplification and this can be extended to the case with different number of repetitions for our future reliability-based work.

*Definition 1 (Sampling-Transmission Scheduling Action):* For any flow  $f_i \in \mathcal{F}$ , the scheduling action at subframe  $t$  is defined as the tuple  $\phi_{i,t} := \{\kappa_{i,t}, c_{i,t}\} \in \Phi$  composed of the sampling decision  $\kappa_{i,t} \in \{0, 1\}$  and if  $\kappa_{i,t} = 1$ , the operating CTU  $c_{i,t} \in \mathcal{N}_{CTU}$ , where  $\mathcal{N}_{CTU}$  contains the CTUs of all the CGs. Since the MA signature can be pre-defined for a flow, the action space  $\Phi$  of a distributed flow can be reduced with the cardinality of  $|\Phi| = \sum_p CG_p \cdot N_{RB} + 1$ .

At the beginning of each subframe, flows make scheduling actions, where  $f_i$  generates a status update packet according to

<sup>2</sup>We focus on short status packet transmission. Therefore, every time a flow generates one status update packet and it can be delivered through a RB.

<sup>3</sup>The transmission reliability can be enhanced with a pre-defined number of consecutive replicas of the same packet in consecutive time slots.

$\kappa_{i,t}$  and transmits it on  $c_{i,t}$  with power  $P$ . Based on GF-NOMA, the BS, upon each time slot, finds the set of singleton CTUs (as shown in Fig. 3), where each CTU is chosen by only one flow with successful detection. Then, the BS performs the successive interference cancellation (SIC) technique to decode the packets of these flows. Based on NOMA principle, at each iterative SIC stage, the BS decodes the flow with the strongest received power and then subtracts the successfully decoded signal from the received signal [7]. The SIC procedure stops when one iterative stage fails. Thus, the decoding order is in sequence to the received power, and the signal-to-interference-plus-noise ratio (SINR) of a transmitting flow  $f_i$  is derived as,

$$\zeta_{i,t} = \frac{P \times k[c_{i,t}]}{\sigma^2 + \sum_{j \in \{c_{i,t}\}} P \times k[c_{j,t}] + \sum_{m \in \{c'_{i,t}\}} P \times k[c_{m,t}]} \quad (4)$$

$C_{i,t}$  contains the detected flows that transmit on the same RB with  $f_i$  and are decoded in the SIC iterative stages behind  $f_i$ .  $C'_{i,t}$  denotes the flows with detection failure in this RB. We consider channel fading is independent across different RBs. During a time slot, the stochastic process of  $f_i$ 's channel gain over a RB (determined by  $c_{i,t}$ ),  $k[c_{i,t}]$ , is defined as  $\alpha_i^f \hat{k}[c_{i,t}]$ .  $\hat{k}[c_{i,t}]$  is the frequency dependent small-scale fading power component whereas  $\alpha_i^f$  denotes the frequency dependent large-scale fading effect, which consists of path loss and shadowing. For successful decoding, we set the SINR threshold as  $\zeta_{th}$ .

### C. CQF-Based Multi-Hop Forwarding in TSN

To reduce sojourn time, the packet of  $f_i$  will be immediately injected into the TSN network and delivered to its destination once successfully received by the BS for seamless transmission. In the multi-hop TSN network, the latency, is predictable and deterministic, which only depends on the pre-configured time slot size and path length. Here, we set the slot size in TSN to be the same with that in 5G for slot alignment as 5G URLLC has shorter mini-slots for TSN integration.

We formally abstract the TSN topology as a directed graph  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ , where  $\mathcal{V}$  is a set of TSN switches and  $\mathcal{E}$  denotes a set of edges connecting any two switches. For each flow  $f_i$ , its path is described with consequent links as  $f_i.path = (e_1, \dots, e_{hop_i})$ , where  $hop_i$  is the hop length. Since the queue length is limited, CQF-based queues are prone to overflow without elaborated injection time planning. Hence, for a transmitting  $f_i$ , the decisive factors of 5G+TSN end-to-end latency contain the time slot offset  $CG_{c_{i,t}} \cdot N_{offset}$  of the operating CTU  $c_{i,t}$  and the hop length  $hop_i$ . Formally, based on the TSN deterministic forwarding, the receiving time slot for every hop of  $f_i$  can be calculated a priori based on the deduced delaying injection time as,

$$\begin{aligned} \forall f_i \in \mathcal{F}, i \in \{0, \dots, F-1\}, e \in f_i.path \\ \tau_{slot}(f_i, e) = hop[e, f_i] + CG_{c_{i,t}} \cdot N_{offset} + l_{5G}. \end{aligned} \quad (5)$$

$l_{5G}$  is the fixed processing and propagation delay in 5G (at per-time slot granularity).  $hop[e, f_i]$  indicates the hop number of

link  $e$  in  $f_i.path$ . To avoid queue overflow and guarantee deterministic transmission, the following queue resource constraint of CQF-based TSN network should be satisfied.

*Queue resource constraint:* This constraint restricts that the usage of each queue should not exceed the queue length limit  $Q_{size}$ . Formally, we define the mapping indicator  $m_{e,\tau}^i$  to denote whether  $f_i$  occupies the queue resource block  $(e, \tau)$ , which represents the transmission resource in the corresponding port of the upstream switch of link  $e \in \mathcal{E}$  at time slot  $\tau$ . Then, this constraint can be written as,

$$\begin{aligned} \forall f_i \in \mathcal{F}, i \in \{0, \dots, F-1\}, \forall e \in \mathcal{E}, \forall \tau, \\ q_{e,\tau} = \sum_{i=0}^{F-1} m_{e,\tau}^i \leq Q_{size}, \end{aligned} \quad (6)$$

where  $q_{e,\tau}$  indicates the queue length at time slot  $\tau$  of link  $e$ .

### D. Problem Formulation

*1) Primal and Dual AgeDS Problem:* Now we can formulate the cooperative optimization problem, that is the age-aware flow deterministic scheduling (AgeDS) problem in 5G+TSN integrated networks, given as,

$$\begin{aligned} \mathcal{J}: \min_{\phi_t} \sum_i \kappa_{i,t} P_{st} \\ \text{s.t. } \mathbb{E}[g_{i,t}(\kappa_t, c_t)] \leq f_i.age, \quad \forall i, \\ \mathbb{E}[q_{e,\tau}(\kappa_t, c_t)] \leq Q_{size}, \quad \forall e, \forall \tau \in t. \end{aligned} \quad (7)$$

$P_{st}$  denotes the total sampling-transmission energy consumption. Besides the sampling decision vector  $\kappa_t$  of all flows, their transmission decision vector  $c_t$  along with the corresponding delaying injection times, would affect whether packets will be received at the destinations (i.e.,  $r_{i,t}$ ) as well as the age and queue evolution. Thus, we describe these evolution processes as age and queue functions (i.e.,  $g_{i,t}(\kappa_t, c_t)$  and  $q_{e,\tau}(\kappa_t, c_t)$ ) with the expectation  $\mathbb{E}[\cdot]$  denotes that they evolve with respect to the stochastic processes of channel gain and communication interference. This makes AgeDS a stochastic and constrained problem with environment uncertainty.

Therefore, we opt for dual optimization in an ergodic manner by introducing dual variables  $\lambda$ , which are corresponding to the ergodic long-term average age and queue states (the time index  $t$  is dropped based on ergodicity). Formally, the Lagrangian dual problem is deduced as,

$$\begin{aligned} \mathcal{D}: \max_{\lambda} \min_{\phi} \mathbb{E} \left[ \sum_i \kappa_i P_{st} + \sum_i \lambda_i (g_i(\kappa, c) - f_i.age) \right. \\ \left. + \sum_e \sum_{\tau} \lambda_{e,\tau} (q_{e,\tau}(\kappa, c) - Q_{size}) \right] \quad \text{s.t. } \lambda \geq 0. \end{aligned}$$

we denote the process inside the expectation notation as the dual function  $\mathcal{L}(\phi, \lambda)$ , concisely.

*2) Duality Gap Analysis:* Note that the AgeDS problem is non-convex and thus there exists duality gap (i.e.,  $\mathcal{J}^* \geq \mathcal{D}^*$ ). Despite the lack of concavity, the following theorem shows that

the stochastic AgeDS problem has null duality gap under mild conditions permitting their solution in the dual domain.

*Theorem 1:* Let  $\mathcal{J}^*$  denote the optimum value of the primal AgeDS problem and  $\mathcal{D}^*$  that of its dual. Assume there exists a strictly feasible point  $\phi_0$  that satisfies both age and queue resource constraints with strict inequality. If the channel probability distribution  $k$  does not contain the point of positive probability, the duality gap is null, i.e.,  $\mathcal{J}^* = \mathcal{D}^*$ .

*Proof:* According to the Theorem 1 in [21], when satisfying the requirements of constraint qualification and nonatomic measure of channel distribution, the ergodic limit span, containing all the possible values that the expectation of  $g_{i,t}(\kappa_t, c_t)$  and  $q_{e,\tau}(\kappa_t, c_t)$  can take, is convex based on Lyapunov's convexity theorem. The condition on channel distribution not having points of positive probability (i.e., nonatomic measure), is a mild requirement satisfied by practical fading channel models including *Rayleigh*, *Rice*, and *Nakagami*. The detail can be seen in the proof of Theorem 1 in [21], which is not discussed here for the sake of space.

#### E. Primal-Dual Stochastic Gradient Descent

Having null duality gap means that AgeDS can work in dual domain without loss of optimality. Hence, we adopt the stochastic gradient descent method [21] to update both primal and dual variables in an iterative manner as in (8)–(10) by re-introducing the time index  $t$  and stepsizes  $\alpha_t$  and  $\beta_t$ . Since the dual function is convex concerning dual variables, we update  $\lambda$  using the zeroth-order representation  $\tilde{g}_{i,t}, \tilde{q}_{e,t}$ . The latter  $\tilde{q}_{e,t}$  is the number of packet overflows in TSN within a subframe, which is counted to update the reduced dual variables for each queue of link  $e$  rather than  $(e, \tau)$ .

$$\phi_{t+1} = \Pi_{\Phi} \{\phi_t - \alpha_t \nabla_{\phi_t} \mathbb{E}[\mathcal{L}(\phi, \lambda)]\}, \quad (8)$$

$$\lambda_{i,t} = [\lambda_{i,t} + \beta_t \tilde{g}_{i,t}]^+, \text{ where } \tilde{g}_{i,t} = g_{i,t} - f_i.age, \forall i, \quad (9)$$

$$\lambda_{e,t} = [\lambda_{e,t} + \beta_t \tilde{q}_{e,t}]^+, \text{ where } \tilde{q}_{e,t} = \sum_{\tau \in t} (q_{e,\tau} - Q_{size}) \quad \forall e. \quad (10)$$

Based on supermartingale argument, if the stepsizes  $\beta_t$  form a nonsummable but square summable series, i.e.,  $\sum_{t=0}^{\infty} \beta_t = \infty$  and  $\sum_{t=0}^{\infty} \beta_t^2 < \infty$ , it can be shown that  $\lambda_{i,t}, \lambda_{e,t}$  converge to the optimum almost surely [21]. Hence, the left issue is how to solve (8) with respect to  $\phi$  in the decentralized setting under uncertain and non-stationary environment.

#### IV. DECAE: AGE-AWARE SCHEDULING ORCHESTRATOR BASED ON MULTI-AGENT RL (MARL)

The straightforward evaluation of (8) is intractable for individual flows since neither the dual function nor its gradients are explicitly known a priori due to stochastic channel gain and time-varying scheduling policies of concurrently-exploring teammates. To address this issue, the model-free MARL algorithm is a potential approach to be considered.

#### A. Preliminaries: Actor-Critic RL on the CTDE Paradigm

Recall that DecAge deals with the non-stationary learning issue based on the centralized training with decentralized execution (CTDE) paradigm. To elaborate, there are  $F$  local learning agents for each flow  $f_i$ , i.e., Flowactor  $i$ , to learn its scheduling policy  $\pi_i$  and update the corresponding age dual variable  $\lambda_{i,t}$ . Using the near-universal deep neural network (DNN) parametrization<sup>4</sup> with parameter  $\theta_i$ , the scheduling policy, also referred to as  $\pi_i(\theta_i)$  or  $\pi_i(o_i|\theta_i)$ , takes the local observation  $o_{i,t} = (\tilde{g}_{i,t-1}, \lambda_{i,t-1}, f_i.age, f_i.loc, f_i.ma)$  and generates the sampling-transmission action  $\phi_{i,t}$  at each  $t$ . Flowactor  $i$  updates  $\pi_i(\theta_i)$  via trial-and-error learning samples to optimize the ergodic long-term dual function. Formally, based on DNN parametrization and policy gradient theorem [22], the optimization of primal update is performed over  $\theta$  rather than the policy directly as follows,

$$\begin{aligned} \theta_{i,t+1} &= \Pi_{\Theta} \{\theta_{i,t} - \alpha_t \nabla_{\theta_i} \mathbb{E}[\mathcal{L}(\phi, \lambda)]\}, \\ \Pi_{\Theta} \{\theta_{i,t} - \alpha_t \mathbb{E}[Q(s_t, a_t) \nabla_{\theta_i} \log \pi_i(\theta_i)]\}, \end{aligned} \quad (11)$$

where  $Q(s_t, a_t)$  is the estimation value of ergodic long-term dual function at each  $t$  learned by Edgectritic.

In the non-stationary environment, a learning agent's transition probabilities may change as other learning agents update their policies. Therefore, the central Edgectritic based on CTDE paradigm is proposed in the edge-assisted BS to learn the dual function from a global and long-term view. Moreover, the BS can also facilitate to infer the packet overflow possibility based on the deterministic forwarding property of TSN.

Specifically, with all the information collected from Flowactors off-line, Edgectritic generates the following three-tuple:

- 1) *State*: the global state defined as  $s_t = (o_{1,t}, \dots, o_{F,t})$ .
- 2) *Action*: the joint action defined as  $a_t = (\phi_{1,t}, \dots, \phi_{F,t})$ .
- 3) *Cost*: the instantaneous dual value, denoted as  $\mathcal{L}_t = \sum_i (\kappa_i P_{st} + \lambda_i \tilde{g}_{i,t}) + \sum_e \lambda_{e,t} \tilde{q}_{e,t}$ .

Based on this three-tuple at each subframe, Edgectritic estimates the action-state function, a.k.a. Q network, based on the deep Q network (DQN) algorithm [23],

$$Q(s_t, a_t) = \mathbb{E} \left[ \sum_t \gamma^{t-t_0} \mathcal{L}_t | s_0, a_0 \right].$$

This DNN-based  $Q(s_t, a_t)$  denotes the long-term accumulated dual values with discount factor  $\gamma$ . Then, the DNN parameters of  $Q(s_t, a_t)$  are updated using the temporal-difference error with the learning rate  $\eta_t$ , given as,

$$Loss_t = \mathbb{E} [Q(s_t, a_t) - (\mathcal{L}_t + \gamma Q'(s_{t+1}, \pi'(s_{t+1}|\theta_t)))] , \quad (12)$$

where the skills of off-policy learning is utilized to improve training stability. Specifically,  $Q'(\cdot)$  and  $\pi'(\cdot)$  denote the target Q network and scheduling policies, which are updated using the soft target update [23] to slowly track their estimated networks, i.e.,  $Q(\cdot)$  and  $\pi_i(\cdot)$ .

<sup>4</sup>Policy parameterization will bring some duality gap due to the function approximation error. However, the duality gap could be very small when using near-universal parameterizations, e.g., DNN [22].

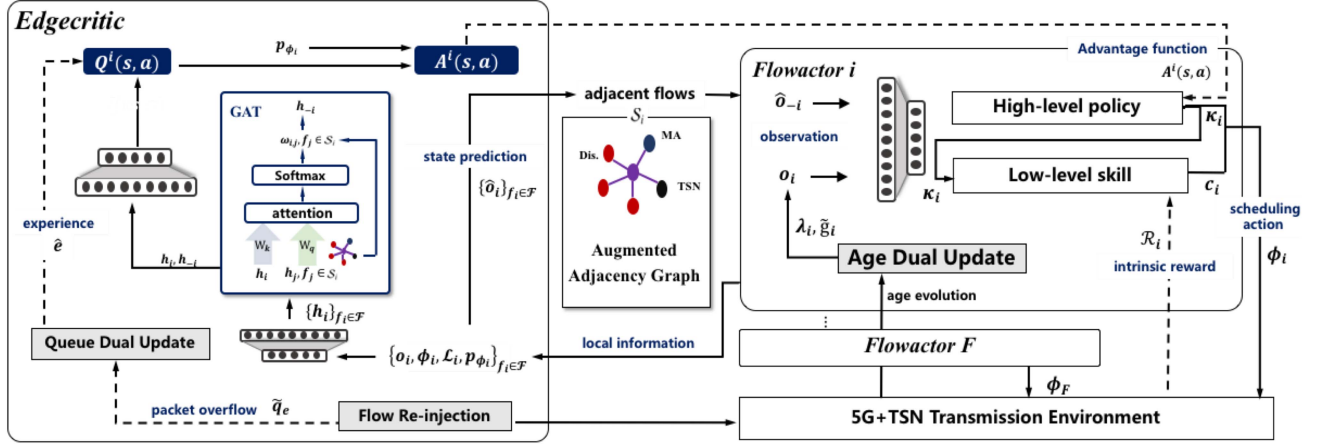


Fig. 4. Framework of DecAge, which contains multiple local Flowactors and a central Edgecritic.

### B. Edgecritic With Augmented Adjacency Graph-Based GAT

The overall architecture of Edgecritic can be seen in Fig. 4. According to the individual-global-min property [24] of AgeDS problem<sup>5</sup>, the Q network is decomposable, which is the sum of individual Q network of each flow as,

$$Q^i(s_t, a_t) = \mathbb{E} \left[ \sum_t \gamma^{t-t_0} (\kappa_{i,t} P_{st} + \lambda_{i,t} \tilde{g}_{i,t} + \sum_{e \in f_i.path} \lambda_{e,t} \mathbb{1}_{e,t}^o) \right],$$

where  $\mathbb{1}_{e,t}^o$  is the indicator denoting whether the packet overflow of  $f_i$  happens in the link  $e$  of its  $f_i.path$ . Therefore, the objective of Edgecritic is to learn  $Q^i(s_t, a_t)$  for each Flowactor. As the operation is done at each subframe, the index  $t$  is dropped in the next for ease of notation.

**Augmented adjacency graph-based GAT:** For the graph-structured flows,  $Q^i(s, a)$  is learned by GAT [25], which can be represented by a function  $\hat{f}$  of  $f_i$ 's feature  $h_i(o_i, \phi_i)$ , and the contributions of others [26], formulated as,

$$Q^i(s, a) = \hat{f}(h_i(o_i, \phi_i), h_{-i}).$$

where  $h_i(\cdot)$  is for feature reconstruction, which contains learnable linear transformation, e.g, multi-layer perceptron (MLP), and embedding layers for discrete encoding.

To capture the impact of others, an augmented adjacency graph (AAG) is proposed. For collision-free transmission, the adjacent flows of  $f_i$  in AAG are defined as the flows with the potential to interfere its packet delivery, i.e., with the same MA or within a certain distance ( $S_{th}$ ), or having overlapping TSN paths. Formally, the adjacent flow set is defined as,

$$\begin{aligned} S_i = \{ & f_j \in \mathcal{F} \setminus f_i \mid (f_i.ma = f_j.ma) \\ & \cup (\|f_i.loc - f_j.loc\| \leq S_{th}) \\ & \cup (f_i.path \cap f_j.path \neq \emptyset) \}. \end{aligned} \quad (13)$$

Based on the features of adjacent flows,  $h_{-i}$  can be calculated by the linear combination of them (after potentially applying a nonlinearity,  $\sigma[\cdot]$ ), formulated as

$$h_{-i} = \sigma \left[ \sum_{f_j \in S_i} \omega_{i,j} h_j(o_j, \phi_j) \right],$$

where  $\omega_{i,j}$  is the attention factor representing the impact degree of  $j$  to  $i$ . It is learned by a shared attention mechanism, which compares the feature  $h_i$  with  $h_j$  using the bilinear query-key mapping system,

$$\omega_{i,j} = \text{softmax}[h_j^T W_k^T W_q h_i],$$

where  $W_q$  transforms  $h_i$  into a query and  $W_k$  transforms  $h_j$  into a key and  $\text{softmax}[\cdot]$  is the normalized exponential function across all choices of  $j$ .

To stabilize learning process, multi-head attention is used, where  $Z$  independent attention mechanisms are learned and the corresponding features are concatenated (i.e.,  $\parallel$ ) to serve as the final output feature for every flow, as

$$h_{-i} = \parallel_{z=1}^Z \sigma \left[ \sum_{f_j \in S_i} \omega_{i,j}^z h_j(o_j, \phi_j) \right].$$

Then,  $h_{-i}, h_i$  are fed into a two-layer MLP to output  $Q^i(s, a)$ .

**Variance-reducing baselines:** To further reduce the variance of gradient estimation for individual Flowactor, an action-dependent advantage function  $A^i(s, a)$  is calculated [27] based on Q network. Specifically, a bias-free baseline is introduced, that only marginalizes out the action  $\phi_i$  of the given Flowactor from  $Q^i(s, a)$  with all other Flowactors fixed, as

$$\begin{aligned} A^i(s, a) &= Q^i(s, a) - b(s, a_{-i}), \text{ where} \\ b(s, a_{-i}) &= \mathbb{E}_{\phi_i \sim \pi_i(o_i | \theta_i)} [Q^i(s, (\phi_i, \phi_{-i}))]. \end{aligned} \quad (14)$$

$A^i(s, a)$  indicates whether the specific action of Flowactor  $i$  will cause an increase in dual value cost or whether any increase in cost is attributed to the actions of other flows. Therefore, this

<sup>5</sup>  $\arg \min_a Q^*(s, a) = (\arg \min_{\phi_1} Q^1(o_1, \phi_1) \dots \arg \min_{\phi_N} Q^N(o_N, \phi_N))$



**Algorithm 1: The Workflow of DecAge Scheme.**


---

**Input:**  $\gamma, M^a, M^c, \alpha_n, \alpha_p, \beta$ ;  
**Initialize:** the scheduling policy  $\pi_i$  and the Q network  $Q$ ;  
*/\* Flowactor with hierarchical hysteretic learning \*/*  
**for**  $t = 1, 2, 3 \dots$  **do**  
  1. Get observation  $o_{i,t}, \hat{o}_{-i,t}$  and make the hierarchical scheduling action according to  

$$\kappa_{i,t} \sim \pi_i^{\kappa}(o_{i,t}, \hat{o}_{-i,t}), \quad c_{i,t} \sim \pi_i^c(\kappa_{i,t}, o_{i,t}, \hat{o}_{-i,t});$$
  
  2. Update  $\lambda_{i,t}$  according to Eqn. (9);  
  3. Get Advantage function  $A_t^i$  from Edgecritic;  
  4. Generate a learning sample  $\{o_{i,t}, \hat{o}_{-i,t}, A_t^i\}$ ;  
  */\* Hierarchical scheduling policy training \*/*  
  **if** Get a batch of  $M^a$  samples for an episode **then**  
    5. Update  $\pi_i^{\kappa}$  based on Eqn. (15);  
    6. Update  $\pi_i^c$  using hysteretic learning according to  

$$\theta_{i,t+1} = \begin{cases} \Pi_{\Theta}\{\theta_{i,t} + \alpha_n \mathcal{R}_{i,t} \nabla_{\theta} \log \pi_i^c(\theta_{i,t})\}, & \text{if } \mathcal{R}_{i,t} < 0, \\ \Pi_{\Theta}\{\theta_{i,t} + \alpha_p \mathcal{R}_{i,t} \nabla_{\theta} \log \pi_i^c(\theta_{i,t})\}, & \text{if } \mathcal{R}_{i,t} \geq 0. \end{cases}$$
  
  **end**  
**end**  
*/\* Edgecritic with AAG-based GAT \*/*  
**for**  $t = 1, 2, 3 \dots$  **do**  
  */\* Flow re-injection with collision-avoiding \*/*  
  1. Keep receiving packets of flows and re-injecting packets with incremental offsets if overflow is foreseen;  
  2. Infer packet overflow numbers  $\tilde{q}_{e,t}$  based on CQF, and update  $\lambda_{e,t}$  according to Eqn. (10);  
  */\* Q network training \*/*  
  3. Collect  $(o_{i,t}, \phi_{i,t}, \mathcal{L}_{i,t})$  from Flowactors along with the action distribution  $p_{\phi_i} \leftarrow \pi_i(o_{i,t}, \hat{o}_{-i,t})$ ;  
  4. Along with the information at  $t + 1$ , generate an experience  

$$\hat{e}_t = \{s_t, a_t, \{\hat{\mathcal{L}}_{i,t}\}_{f_i \in \mathcal{F}}, \hat{s}_{t+1}\};$$
  
  5. Randomly sample a mini-batch  $\{\hat{e}_m\}_{m \in [1..M^c]}$  with size  $M^c$  to train Q network based on Eqn. (12);  
  */\* Output to each local Flowactor \*/*  
  6. Output advantage function  $A_t^i$  based on  $p_{\phi_i}$  according to Eqn. (14) and output state prediction  $\hat{o}_{-i,t+1}$ ;  
**end**

---

baseline can largely reduce the policy estimation variance and assists Flowactor speeding up policy convergence.

### C. Flowactor With Hierarchical Hysteretic Learning

**AAG-assisted hierarchical scheduling policy:** Recall that local Flowactors train scheduling policies based on the customized advantage function  $A^i(s, a)$  from Edgecritic, however, with the local observation  $o_{i,t}$ . To facilitate coordination, we propose to let each Flowactor consider the state prediction  $\hat{o}_{-i}$  of only its adjacent flows in AAG for decision-making,

$$\phi_i \sim \pi_i(o_i, \hat{o}_{-i}), \text{ where } \hat{o}_{-i} = \{o_j\}_{f_j \in \mathcal{S}_i},$$

where  $\hat{o}_{-i}$  can be easily inferred by GAT-based Edgecritic with global information. This neighboring-based message can make Flowactor sensitive to the transmission collision threat from the adjacent flows in  $\mathcal{S}_i$  during both training and inference phases. As a result, optimal scheduling policies would like to choose collision-free CTUs or even be idle when adjacent flows are

likely to transmit. Hence, flows transmit episodically to reduce power consumption and interference to others since there is no need to update if status is fresh enough. This leads to learning with sparse communication rewards due to the hierarchy of scheduling action.

To address this issue, we present the hierarchical hysteretic learning with an auxiliary subgoal in each Flowactor. Specifically, Flowactor has a hierarchically-structured policy, allowing for control concurrently at two-levels. Formally, the joint policy is composed of an age-aware high-level sampling policy  $\pi_i^{\kappa}$  and the communication-oriented low-level skill  $\pi_i^c$ . The high-level action is defined as the sampling decision

$$\kappa_i \sim \pi_i^{\kappa}(o_i, \hat{o}_{-i}),$$

which refers to the inclination to deliver a packet. At each subframe,  $\pi_i^{\kappa}$  passes this instruction to the lower-level skill  $\pi_i^c$ , which serves as the subgoal of whether or not having a collision-free packet delivery. Therefore, the aim of  $\pi_i^c$  is to only transmit successfully, which is conditioned on  $\kappa_i$  that alters states in a consistent way as,

$$c_i \sim \pi_i^c(\kappa_i, o_i, \hat{o}_{-i}).$$

In this way, lower-level skills are supervised with subgoals that are proposed automatically by higher-level policies.

In Flowactor,  $\pi_i^{\kappa}$  and  $\pi_i^c$  partially share the parameters  $\theta_i$  in the previous MLP layers for feature transformation, followed by the hierarchical MLPs for the corresponding  $\pi_i^{\kappa}$  and  $\pi_i^c$ . The detailed architecture is delineated in Fig. 4.

The high-level policy  $\pi_i^{\kappa}$  is trained from the global prospect with the feedback advantage function as,

$$\theta_{i,t+1} = \Pi_{\Theta}\{\theta_{i,t} - \alpha_t A_i(s_t, a_t) \nabla_{\theta_i} \log \pi_i^{\kappa}(\theta_{i,t})\}. \quad (15)$$

Moreover, we design an intrinsic reward, which reflects the realistic transmission result for low-level training, defined as,

$$\mathcal{R}_{i,t} = \psi \kappa_{i,t} (2r_{i,t} - 1), \quad (16)$$

which punishes the low-level skill with  $-\psi$  when collision happens and promotes it for successful packet delivery. As a result, the task of finding the collision-free CTUs is decoupled, which forms a communication game of transmitting flows.

**Hysteretic learning:** With the insight that low returns in this game may be caused by domain stochasticity. In other words, the transmission failure of a flow may be mainly caused by others' exploratory actions [28]. To lag the degradation of the policy underestimated due to the outcomes unrelated to its associated action, hysteretic learning is utilized. Formally, two learning rates are defined, where a nominal learning rate,  $\alpha_p$ , is used when the reward  $\mathcal{R}_{i,t}$  is non-negative and a smaller learning rate,  $\alpha_n$  is used otherwise (where  $0 < \alpha_n < \alpha_p < 1$ ). In this way, more attention will be paid to the positive learning samples (a.k.a. experiences) that occurred due to the successful cooperation among transmitting flows.



#### D. Workflow of DecAge With Flow Re-Injection for TSN

The workflow of both Flowactors and Edgecritic can be concluded in Algorithm 1 and Fig. 4.

1) *Local Flowactor*: Flowactor aims to learn an age-aware hierarchical scheduling policy for each flow. To be specific, at the beginning of each subframe  $t$ , Flowactor firstly gets the observation  $o_{i,t}, \hat{o}_{-i,t}$  to make sampling and transmission decisions on-site through the hierarchical policy  $\pi_i^k, \pi_i^c$ . Then, all flows perform their scheduling actions and have age evolution accordingly. After that, each flow updates its age dual parameter  $\lambda_{i,t}$  locally. After receiving the advantage function  $A_t^i$  from Edgecritic, Flowactor gets a sample  $\{o_{i,t}, \hat{o}_{-i,t}, A_t^i\}$  and its primal parameters of hierarchical scheduling policy will be updated once collecting a batch of samples (with batch size  $M^a$ ). Note that this training is done in an off-line manner and thus does not need timely feedbacks from Edgecritic.

2) *Central Edgecritic*: Edgecritic, located at the edge controller, aims to estimate the performance of Flowactors from a global and long-term view. During each subframe  $t$ , Edgecritic collects local observations from Flowactors along with the action probability distribution  $p_{\phi_i}$  for bias-free baseline calculation. Based on consecutive local observations information, Edgecritic forms an experience  $\hat{e}_t$ . With experience replay skill, Edgecritic accumulates a dataset of experiences and randomly gets a minibatch  $\{\hat{e}_m\}_{m \in \{1 \dots M^c\}}$  experiences with size  $M^c$  to train Q network via DQN. Finally, Edgecritic outputs  $A_t^i, \hat{o}_{-i,t+1}$  to each Flowactor for local training and inference.

*Flow re-injection*: Besides the policy estimation, Edgecritic keeps computing the queue length of CQF-based queues virtually during a subframe based on the deterministic forwarding property for collision-avoiding. With virtual queues, Edgecritic re-injects the packet received in the BS that has been foreseen to overflow according to the offset of its transmitting CG. Specifically, Edgecritic incrementally increases the injection time offset of this packet until it is deduced to be delivered successfully. This step can guarantee the deterministic delivery in TSN at the sacrifice of having additional waiting time in the BS. In this way, the real dual value  $\mathcal{L}_{i,t}$  doesn't contain the punishment of overflow and thus Edgecritic modifies  $\mathcal{L}_{i,t}, s_{i,t+1}$  with the inference of  $\tilde{q}_{e,t}$  (with packet loss) considered, written as  $\hat{\mathcal{L}}_{i,t}, \hat{s}_{t+1}$  in the experience  $\hat{e}$ .

*Complexity and scalability analysis*: Note that the training phase can be done in an off-line manner in Flowactors and Edgecritic, where the most computational tasks (i.e., policy estimation) are offloaded to Edgecritic with powerful computing ability. As for real-time scheduling action inference in local Flowactors, the time complexity of DNN-based scheduling policy linearly scales with the number of DNN parameters (depending on the number of layers as well as the sizes of input and output) and the cardinality of action space (i.e.,  $|\Phi| = \sum_p CG_p \cdot N_{RB} + 1$ ) [29]. Therefore, compared with the centralized scheduling having a joint action space (with the cardinality  $|\Phi|^F$ ), the time complexity of DecAge largely decreases from exponential with  $|\Phi|$  to linear with  $|\Phi|$ .

Moreover, DecAge is scalable and can react to any change of flow number by adding or removing the Flowactor in the corresponding flow host. Then, the scheduling policies in Flowactors

TABLE III  
SIMULATION PARAMETERS

Parameter	Value ( $d[m]$ )
\* MCG-based 5G system*\	
Carrier frequency $f$ , Tx. power	4GHz, 24dBm
Fading, shadowing, noise st. dev.	Rayleigh, 4dB, $-174\text{dBm/Hz}$
LOS path loss	$18.7 \log(d) + 20 \log(f/5) + 46.8$
NLOS path loss	$36.8 \log(d) + 20 \log(f/5) + 43.8 + 5N_{wall}$
$\mu$ , TTI duration, slot number	2, 0.125 ms, 8
The numbers of CG $\times$ RB, MA	$8 \times 2, 6$
The set of offsets of MCGs	$\{0, 1, 2, 3, 4\}$
\* CQF-based TSN system*\	
TSN switch and edge number	16, 24
Queue length, Maximal hop	3 packets, 7

are re-trained using real-time cost feedback. Moreover, the central GAT-based Edgecritic is based on inductive learning, which can apply to completely unseen graphs since only the neighbors' features are utilized and the operation of GAT is parallelizable across node neighbor pairs.

#### V. PERFORMANCE EVALUATION

In this section, we demonstrate the effectiveness of DecAge under different settings of 5G+TSN integrated networks.

##### A. Experimental Setups

The parameters of 5G+TSN networks are listed in Table III, which partially follow the works [20], [30], [31].  $d$  represents the distance between transmitters and receivers.

*Setting for 5G NR and flow feature*: We focus on the 5G system with mini-slots of 7 OFDM symbols using  $\Delta f = 60\text{ kHz}$  ( $\mu = 2$ ) in line with the main guidelines for 3GPP NR performance evaluations in [30]. For age-aware monitoring tasks, 50 flow hosts are randomly positioned inside a  $600 \times 600\text{ m}^2$  industrial service area with their remote destinations randomly selected from the TSN switch set. Their QoS requirements, i.e., age limits, are set from  $[4, 8]\text{ms}$ , which follow the deadline request of industrial applications (e.g., 1ms-20 ms for factory automation) [2]. The BS is located in the center of this area, responsible for SIC decoding and packet injection. All devices operate based on the WINNER II indoor model, a link-level channel model for LTE system design [31]. In the non-line of sight (NLOS) path loss model,  $N_{wall}$  denotes the number of walls between two locations and is set to  $d/50$ . For successful SIC decoding, the SINR threshold  $\zeta_{th}$  is set to be  $0\text{ dB}$ .

*Setting for CQF-based TSN*: A TSN network with mesh topology is established, which contains 16 TSN switches and a TSN gateway in the BS. Every two switches are connected with probability 0.2. We only consider unicast flows with single source and destination as multicast flows can be split into multiple unicast flows. Their TSN paths are pre-configured according to the shortest forwarding paths computed by Dijkstra [32] with maximal hop count 7. Based on the network scale, the length of each CQF-based queue is set to 3. Actually, the queue length should be small enough to avoid long time packet cache [15]. The more realistic network parameters, e.g., link bandwidth, global

time slot size, will be further considered in our future work of building a real 5G+TSN integrated platform for verification.

**Hyperparameters:** In DecAge, both Flowactors and Edgecritic are trained with the Adam optimizer. The local scheduling policy contains 2-layer MLP with hidden size 128. The Q network in Edgecritic is composed of an MLP with hidden size 128 for feature reconstruction, an multi-head attention based GAT network with  $Z = 3$  as well as the output feature size 32, and a 2-MLP output layer with hidden size 64. For Edgecritic, the learning rate is set to 0.001, while for Flowactors, the hierarchical hysteretic learning is applied with  $\alpha_p$  set to 0.008 for high-level training and  $\alpha_p, \alpha_n$  set to 0.001, 0.0001 for low-level training, respectively. As for AAG, the neighbors of  $f_i$ , i.e.,  $\mathcal{S}_i$ , is defined as the flows 1) within the distance of 200 m or the same MA, and 2) with overlapping paths in TSN.

### B. Simulation Results

The following metrics are considered: 1) age violation ratio (VR), the percentage of flows whose age constraints are violated; 2) power efficiency (EFF), which is calculated as the ratio of the total power consumption to the number of flows satisfying age constraints, i.e.,

$$\text{Power EFF} = \frac{\sum_i \kappa_{i,t} P_{st}}{\sum_i \mathbb{1}\{g_{i,t} \leq f_{i,\text{age}}\}};$$

3) communication efficiency, the packet loss ratio of 5G+TSN transmission; and 4) learning capability, that includes the cost prediction error and the MSE (mean squared error) loss of Edgecritic ( $\gamma = 0$ ). Note that the age VR metric relies on both sampling and transmission decisions, and thus it is relevant to the metrics of power and communication efficiency.

1) *Comparison of Different Scheduling Policies:* We first compare the proposed DecAge with the following schemes,

- SRandom [7], where flows access the nearest CGs and randomly choose the CTUs if needed, i.e.,  $g_{i,t} - f_{i,\text{age}} = 0$ ;
- Dec-PG [22], where the hierarchical scheduling policy is learned locally by primal-dual policy gradient with the individual and instantaneous cost  $\kappa_i P_{st} + \lambda_i (g_{i,t} - f_{i,\text{age}})$ .

Fig. 5 delineates the performance of three scheduling policies. It can be clearly seen that DecAge significantly outperforms the traditional strategy SRandom in terms of all the metrics since it can guide flows intelligently to choose collision-free CTUs through learning. Compared with the Dec-PG scheme, the scheduling policies trained by DecAge converge faster with the CTDE learning framework as well as the action-dependent baseline in terms of variance reduction. Moreover, it can be observed in the subfigures that DecAge achieves lower age violation ratio (close to 0) and better communication efficiency after policy convergence, leading to better power efficiency. Thus, flows orchestrated by DecAge can timely deliver status packets with less power consumption in 5G+TSN networks.

2) *Performance of Hierarchical RL of DecAge:* The performance improvement of DecAge with hierarchical learning is shown in Fig. 6, where DecAge-w/oH denotes the policies without hierarchical DNN for action output layers (with the output dimension  $|\Phi|$ ). Obviously, DecAge without HRL experiences

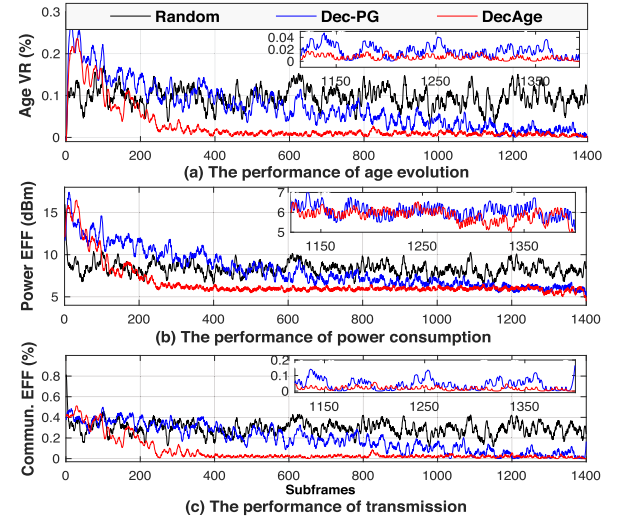


Fig. 5. Performance comparison of scheduling policies.

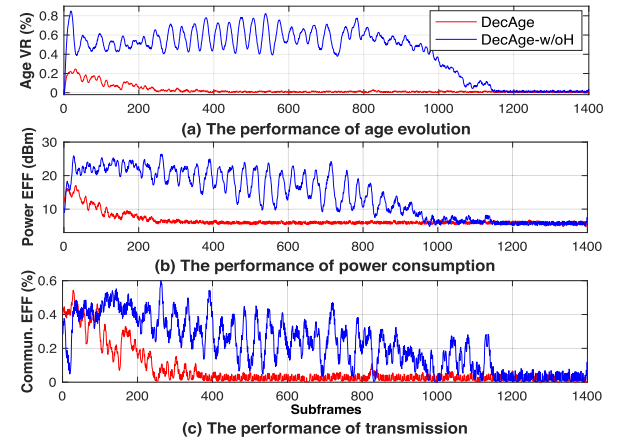


Fig. 6. Performance of hierarchical RL of DecAge.

a long-time performance vibration before policy convergence since most flows will transmit regardless of their age states due to the mixed structure of action output layers. As a result, it takes Flowactors longer time to find the right sampling time and the power consumption of DecAge-w/oH is much higher at first 1000 time slots. Nevertheless, DecAge-w/oH keeps searching for the way of collision-free transmission based on the curve of communication efficiency. Hence, the policies of DecAge-w/oH will finally converge.

3) *Performance of DecAge and Its Action-Dependent Baseline Under Different Environments:* DecAge is evaluated in Fig. 7 under different environments with various numbers of flows and wireless resources (i.e., CG, RB and MA). Moreover, DecAge-w/oB adopting the CDTE learning with traditional action-agnostic baselines is also compared. Note that, more flows need to update their status packets at each subframe with the increasing number of flows and thus transmission collision probability increases, which leads to larger age violation ratio and power consumption. Nevertheless, the elaborately-designed

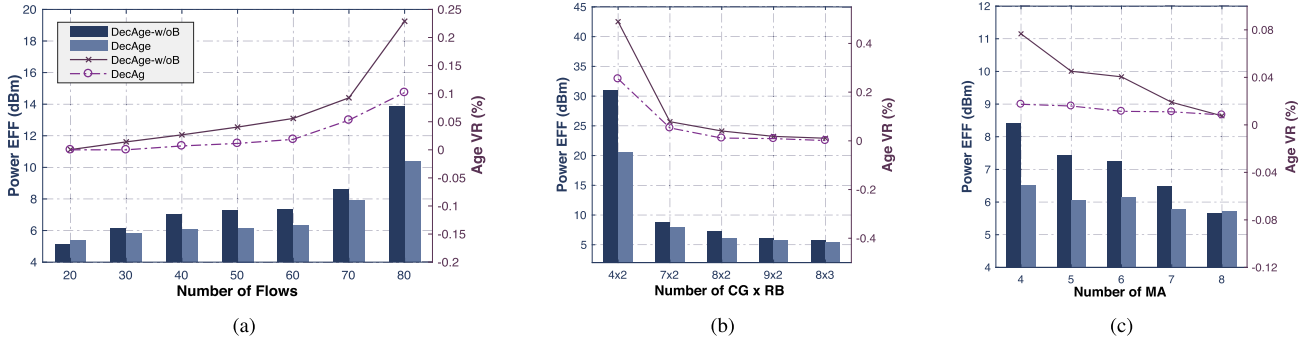


Fig. 7. Performance of DecAge and DecAge without action-aware baseline (DecAge-w/oB) in different environments.

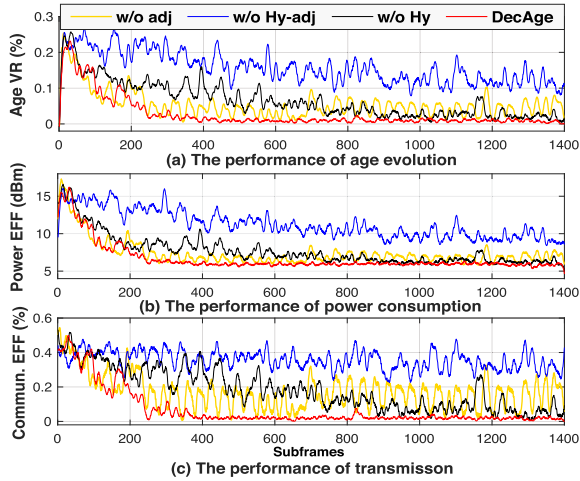


Fig. 8. Ablation study of DecAge.

baselines along with the information of neighbors assist Flowactors perceiving the contributions of others when collision happens and trying to find out collision-free CTUs at the next time with similar situation. Thus, the performance of DecAge will not deteriorate as that of DecAge-w/oB in the environment with large number of flows (e.g., 80) in Fig. 7(a). Another observation can be seen is that with the same resource density (i.e.,  $16/80 = 10/50$ ), the performance of DecAge in  $(80, 8 \times 2)$  in Fig. 7(a) is worse than that in  $(50, 5 \times 2)$  in Fig. 7(b). This is because, with fewer CGs, more packet overflows happen in the CQF-based queues due to the lack of time offset. For the same reason, the less impact of MA number on performance is shown in Fig. 7(c) since flows can form suitable collision-free communication coalitions under the guidance of DecAge even in the shortage of resources.

4) *Ablation Study of DecAge*: The ablation study is conducted as shown in Fig. 8, where the performances of DecAge as well as the DecAge without hysteretic learning (w/o Hy), without adjacent flow state prediction (w/o adj), and without both (w/o Hy-adj) are delineated. It can be clearly seen that DecAge without both skills cannot guide flows to find suitable scheduling policies to meet their age constraints. To analyze separately, the policies learned by w/o Hy will converge despite taking more time since the actions with high costs are over-punished in the

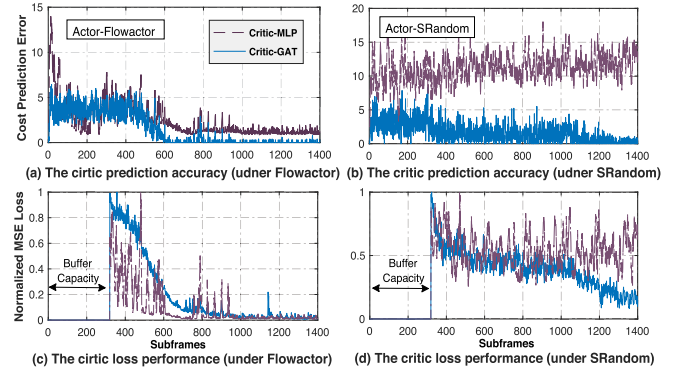


Fig. 9. Approximation performance of the critic.

previous exploratory phase. Moreover, although w/o adj can achieve fast policy convergence, the curves of w/o adj fluctuate severely. This is because, without taking their information into account, flows cannot completely give way to the adjacent flows having collision threat.

5) *Approximation Performance of the Critic*: To test the learning ability of Edgectric, we conduct the prediction task of instantaneous costs ( $\gamma = 0$ ) under the Flowactors following DecAge and SRandom, respectively. Critic-MLP denotes the Edgectric with the Q network composed of MLP while Critic-GAT is the one DecAge adopts. Fig. 9 shows the prediction error (per flow) and MSE loss. As policy converges, the prediction error decreases quickly in both Critic-MLP and Critic-GAT with the policy learned by Flowactor. Nevertheless, Critic-GAT achieves better prediction accuracy, which is much obvious under the situation of SRandom. It's important to have satisfactory performance in SRandom situation since the calculation of action-dependent baseline needs more possibilities taken into considered. Therefore, Critic-GAT can mitigate the over-fitting issue by the mechanisms of graph and attention.

## VI. RELATED WORK

The third generation partnership project (3GPP) has introduced enhanced GF transmission in 5G uplinks with multiple active CGs. In addition, NOMA has been proposed to synergize with GF transmission to mitigate the serious transmission delay and network congestion problems.



In the traditional GF-NOMA schemes, URLLC users simply select the nearest CGs to transmit packets. Therefore, from the BS' perspective, the authors in [7], [8] introduced DRL to optimize the CG configuration parameters, e.g., the number of CTUs and the time offset of CGs, to balance the allocation of channel resources among CGs in order to maximize the number of successful transmissions. From the perspective of users, a DRL-based GF-NOMA algorithm is proposed to learn about the network contention state, and output subchannel and received power level selection to reduce transmission collisions in the frequency domain [9]. Similarly, the authors in [10] proposed distributed Q-learning aided uplink GF-NOMA schemes with bursty traffic considered to maximize the number of accessible devices.

Although these DRL-based GF-NOMA resource allocation schemes have been proposed to optimize network collision and have reduced computational complexity with distributed learning. Nevertheless, the unstable learning issue brought by non-stationary environment and partial observability of local users is not well-addressed. Therefore, we propose DecAge to deal with this issue using centralized training with decentralized execution paradigm in multi-agent RL.

## VII. CONCLUSION

This article addresses the age-aware flow scheduling problem in 5G+TSN integrated networks, where DecAge, a decentralized scheduling scheme based on multi-agent RL, is proposed to guide flows to make sampling-transmission actions. To elaborate, DecAge contains multiple Flowactors for local flows to learn distributed scheduling policies and a central Edgecritic to coordinate the behaviors of flows based on the GAT-based Q network and variance-reducing baselines. Edgecritic is scalable to the number of flows despite centralized training and assists flows to differentiate the impact of others, leading to fast policy convergence. Moreover, faced with insufficient communication samples, Flowactor is extremely sample-efficient with adjacent state aware hierarchical learning, which decouples sampling and transmission actions by the hierarchically-structured DNN. And hysteretic learning is introduced, which lags the degradation of transmission actions that was underestimated due to the outcomes related to others' exploration.

In general, the robustness of Flowactors to non-stationarity and the scalability of Edgecritic make DecAge well-suited to the distributed and partial-observed systems and also provides a potential solution for future studies on the scheduling of graph-structured wireless networks.

## REFERENCES

- [1] A. Nasrallah et al., "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surveys Tut.*, vol. 21, no. 1, pp. 88–145, First Quarter 2019.
- [2] A. Mahmood et al., "Industrial IoT in 5G-and-beyond networks: Vision, architecture, and design trends," *IEEE Trans Ind. Informat.*, vol. 18, no. 6, pp. 4122–4137, Jun. 2022.
- [3] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tut.*, vol. 22, no. 3, pp. 1805–1838, Third quarter 2020.
- [4] Y. Liu, Y. Deng, M. ElKashlan, A. Nallanathan, and G. K. Karagiannis, "Analyzing grant-free access for URLLC service," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 741–755, Mar. 2021.
- [5] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A novel analytical framework for massive grant-free NOMA," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2436–2449, Mar. 2019.
- [6] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, Jun. 2019.
- [7] Y. Liu, Y. Deng, M. ElKashlan, A. Nallanathan, and G. K. Karagiannis, "Optimization of grant-free NOMA with multiple configured-grants for mURLLC," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1222–1236, Apr. 2022.
- [8] Y. Liu, Y. Deng, H. Zhou, M. ElKashlan, and A. Nallanathan, "Deep reinforcement learning-based grant-free NOMA optimization for mURLLC," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1475–1490, Mar. 2023.
- [9] J. Zhang, X. Tao, H. Wu, N. Zhang, and X. Zhang, "Deep reinforcement learning for throughput improvement of the uplink grant-free NOMA system," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6369–6379, Jul. 2020.
- [10] J. Liu, Z. Shi, S. Zhang, and N. Kato, "Distributed Q-learning aided uplink grant-free NOMA for massive machine-type communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2029–2041, Jul. 2021.
- [11] *IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks – Amendment 25: Enhancements for Scheduled Traffic*, IEEE Std. 802.1Qbv-2015, 2016.
- [12] D. Yang, K. Gong, J. Ren, W. Zhang, W. Wu, and H. Zhang, "TC-Flow: Chain flow scheduling for advanced industrial applications in time-sensitive networks," *IEEE Netw.*, vol. 36, no. 2, pp. 16–24, Mar./Apr. 2022.
- [13] *IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks – Amendment 29: Cyclic Queuing and Forwarding*, IEEE Std. 802.1Qch-2017, 2017.
- [14] A. Nasrallah, V. Balasubramanian, A. Thyagaturu, M. Reisslein, and H. ElBakoury, "Cyclic queuing and forwarding for large scale deterministic networks: A survey," 2019, *arXiv:1905.08478*.
- [15] J. Yan, W. Quan, X. Jiang, and Z. Sun, "Injection time planning: Making CQF practical in time-sensitive networking," in *Proc. IEEE Conf. Comput. Commun.*, 2020, pp. 616–625.
- [16] Y. Zhang, Q. Xu, L. Xu, C. Chen, and X. Guan, "Efficient flow scheduling for industrial time-sensitive networking: A divisibility theory based method," *IEEE Trans Ind. Informat.*, vol. 18, no. 12, pp. 9312–9323, Dec. 2022.
- [17] W. Quan, J. Yan, X. Jiang, and Z. Sun, "On-line traffic scheduling optimization in IEEE 802.1Qch based time-sensitive networks," in *Proc. IEEE 22nd Int. Conf. High Perform. Comput. Commun., IEEE 18th Int. Conf. Smart City, IEEE 6th Int. Conf. Data Sci. Syst.*, 2020, pp. 369–376.
- [18] D. Yang, Z. Cheng, W. Zhang, H. Zhang, and X. Shen, "Burst-aware time-triggered flow scheduling with enhanced multi-CQF in time-sensitive networks," *IEEE/ACM Trans. Netw.*, vol. 31, no. 6, pp. 2809–2824, Dec. 2023.
- [19] I. Kadota and E. Modiano, "Age of information in random access networks with stochastic arrivals," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.
- [20] 3rd Generation Partnership Project (3GPP), "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)," 3GPP, Sophia Antipolis, France, Tech. Rep. 38.824 v16.0.0, Mar. 2019.
- [21] A. Ribeiro, "Optimal resource allocation in wireless communication and networking," *Eurasip J. Wireless Commun. Netw.*, vol. 2012, no. 1/19, 2012, Art. no. 272.
- [22] M. Eisen, C. Zhang, L. F. O. Chamon, D. D. Lee, and A. Ribeiro, "Learning optimal resource allocations in wireless systems," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2775–2790, May 2019.
- [23] T. P. Lillicrap et al., "Continuous control with deep reinforcement," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–14.
- [24] Y. Yang et al., "Qatten: A general framework for cooperative multiagent reinforcement learning," 2020, *arXiv:2002.03939*.
- [25] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–13.
- [26] S. Iqbal and F. Sha, "Actor-attention-critic for multi-agent reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2961–2970.
- [27] J. N. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson, "Counterfactual multi-agent policy gradients," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 2974–2982.
- [28] S. Omidshafiei et al., "Deep decentralized multi-task multi-agent reinforcement learning under partial observability," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2681–2690.

- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [30] 3rd Generation Partnership Project (3GPP), "Study on new radio access technology-physical layer aspects," 3GPP, Sophia Antipolis, France, Tech. Rep. 38.802 v14.0.0, Mar. 2017.
- [31] "WINNER II path loss model," Accessed: Jan. 6, 2012. [Online]. Available: <http://www.raymaps.com/index.php/winner-ii-path-loss-model/>
- [32] E. W. Dijkstra et al., "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.



**Mingyan Li** received the B. E. degree in telecommunication engineering from Jinlin University, Changchun, China, in 2015, and the Ph.D. degree from the Department of Electronic Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2021. In 2021, she joined the College of Computer Science, Chongqing University, Chongqing, China, where she is currently a Postdoctoral Research Associate. During 2019–2020, she was a Visiting Professor with the University of Waterloo, Waterloo, ON, Canada. Her research interests include industrial internet, the next generation mobile communication, joint design of sensing, communication and control in industrial cyber-physical systems, software-defined networking and network slicing, ultra-reliable low-latency communication, time-sensitive networking, and deep reinforcement learning. She has received many research grants as a Principal Investigator from the National Science Foundation of China.



**Songtao Guo** (Senior Member, IEEE) received the B.S., M.S. and Ph.D. degrees in computer software and theory from Chongqing University, Chongqing, China, in 1999, 2003 and 2008, respectively. He was a Professor with Chongqing University from 2011 to 2012 and with Southwest University from 2013 to 2018. He is currently a Full Professor with Chongqing University. He was a Senior Research Associate with the City University of Hong Kong from 2010 to 2011, and a Visiting Scholar with Stony Brook University, Stony Brook, NY, USA, from May 2011 to May 2012. He has authored or coauthored more than 200 scientific papers in leading refereed journals and conferences. His research interests include wireless sensor networks, wireless ad hoc networks, and mobile edge computing. He has received many research grants as a Principal Investigator from the National Science Foundation of China and Chongqing and the Postdoctoral Science Foundation of China.



**Chao Chen** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in control science and control engineering from Northwestern Polytechnical University, Xi'an, China, in 2007 and 2010, respectively, and the Ph.D. degree from Pierre and Marie Curie University, Paris, France, and Institut Mines-Télécom/Télécom SudParis, France, in 2014. He is currently a Full Professor with the College of Computer Science, Chongqing University, Chongqing, China. He has authored or coauthored more than 80 papers including 20 ACM/IEEE Transactions. His research interests include pervasive computing, mobile computing, urban logistics, data mining from large-scale GPS trajectory data, and Big Data analytics for smart cities. His work on taxi trajectory data mining was featured by IEEE SPECTRUM in 2011, 2016, and 2020 respectively. He was the recipient of the Best Paper Runner-Up Award at MobiQuitous 2011.



**Cailian Chen** (Senior Member, IEEE) received the B. Eng. and M. Eng. degrees in automatic control from Yanshan University, Qinhuangdao, China, in 2000 and 2002, respectively, and the Ph.D. degree in control and systems from the City University of Hong Kong, Hong Kong SAR, in 2006. Since 2008, she has been with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China. She is currently a Distinguished Professor. She has authored three research monographs and more than 100 referred international journal papers. She is the Inventor of more than 30 patents. Her research interests include industrial wireless networks and computational intelligence, and Internet of Vehicles. Dr. Chen was the recipient of the prestigious IEEE Transactions on Fuzzy Systems Outstanding Paper Award in 2008, IEEE TCCPS Industrial Technical Excellence Award in 2022, and five conference best paper awards. She was awarded N2Women Star in Computer Networking and Communications in 2022, and the Second Prize of National Natural Science Award from the State Council of China in 2018. She was honored National Outstanding Young Researcher by NSF of China in 2020. Dr. Chen has been actively involved in various professional services. She is a Distinguished Lecturer of IEEE VTS. She is the Deputy Editor of the *National Science Open*, and Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.



**Xiaofeng Liao** (Fellow, IEEE) received the B.S. and M.S. degrees in mathematics from Sichuan University, Chengdu, China, in 1986 and 1992, respectively, and the Ph.D. degree in circuits and systems from the University of Electronic Science and Technology of China, Chengdu, in 1997. From 1999 to 2012, he was a Professor with Chongqing University, Chongqing, China. From July 2012 to July 2018, he was a Professor and the Dean with the College of Electronic and Information Engineering, Southwest University, Chongqing. He is currently a Professor and the Dean with the College of Computer Science, Chongqing University. He is also a Yangtze River Scholar of the Ministry of Education of China, Beijing, China. From November 1997 to April 1998, he was a Research Associate with the Chinese University of Hong Kong, Hong Kong. From October 1999 to October 2000, he was a Research Associate with the City University of Hong Kong, Hong Kong. From March 2001 to June 2001 and March 2002 to June 2002, he was a Senior Research Associate with the City University of Hong Kong. From March 2006 to April 2007, he was a Research Fellow with the City University of Hong Kong. He has authored or coauthored four books and more than 300 international journal and conference papers. His research interests include decentralized optimization, machine learning, neural networks, and bifurcation and chaos cryptography. He is an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.



**Xinping Guan** (Fellow, IEEE) received the B.Sc. degree in mathematics from Harbin Normal University, Harbin, China, in 1986, and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, in 1999. He is currently a Chair Professor of Shanghai Jiao Tong University, Shanghai, China, where he is the Dean of School of Electronic, Information and Electrical Engineering, and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. Before that, he was the Executive Director of Office of Research Management, Shanghai Jiao Tong University, a Full Professor and Dean of Electrical Engineering, Yanshan University, Qinhuangdao, China. He has authored and/or coauthored five research monographs, more than 200 papers in peer-reviewed journals, and numerous conference papers. His research interests include industrial network systems, smart manufacturing, and underwater networks. As a Principal Investigator, he has finished/been working on more than 20 National key projects. He is the Leader of the prestigious Innovative Research Team of the National Natural Science Foundation of China. Dr. Guan is an Executive Committee Member of Chinese Automation Association Council and the Chinese Artificial Intelligence Association Council. He was the recipient of the Second Prize of the National Natural Science Award of China in 2008 and 2018, First Prize of the National Science Award from the Ministry of Education of China in 2006 and 2016, and IEEE Transactions on Fuzzy Systems Outstanding Paper Award in 2008 and IEEE TCCPS Industrial Technical Excellence Award in 2022. He is a National Outstanding Youth honored by NSF of China, Changjiang Scholar by the Ministry of Education of China and State-level Scholar of New Century Bai Qianwan Talent Program of China. He was elected a Fellow of IEEE in 2018.