Enhancing Time-Critical Communication for Industrial Applications with Deep Q-Network and Policy Reuse based TDM-PON

Xiang Li
State Key Laboratory of information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
iexlee@bupt.edu.cn

Hui Yang
State Key Laboratory of information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
yanghui@bupt.edu.cn

Qiuyan Yao
State Key Laboratory of information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
yqy89716@bupt.edu.cn

Bowen Bao
State Key Laboratory of information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
baobowen@bupt.edu.cn

Jie Zhang
State Key Laboratory of information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
lgr24@bupt.edu.cn

Mohamed Cheriet
Department of System Engineering
University of Quebec's École de technologie supérieure (ÉTS)
Montreal, Canada
mohamed.cheriet@etsmtl.ca

***Abstract*: This paper proposes a DQN-based scheduling scheme for TDM-PON to optimize industrial communication for time-critical services by introducing Q-function and policy reuse. Simulation results show the proposed scheme effectively reduces the delay and jitter.**

Keywords—industrial park, latency sensitive services, reinforcement learning, policy reuse.

I. INTRODUCTION

Fifth-generation wireless communications (5G) and time-sensitive networking (TSN) technologies, which require efficient network architecture support, are key to the future of industrial manufacturing[1]. Benefit from its passive features, wide bandwidth, and cost-effectiveness, Time Division Multiplexing Passive Optical Network (TDM-PON) has emerged as an attractive and promising communication infrastructure. Existing TDM-PON systems have been optimized for uses in both residential and commercial environments. However, the ultra-reliable low latency communication (URLLC) services such as industrial automation and vehicles with the strict constraints on end-to-end latency and jitter in industrial application systems present a challenge.

Various bandwidth allocation schemes have been proposed to address the uplink bandwidth and delay requirements, including Dynamic Bandwidth Allocation (DBA) scheme and Traffic Load Based Priority Scheduling DBA. The conventional schemes may yield satisfactory results in typical residential and commercial environments. Unfortunately, in an industrial environment, time-critical flows in IIoT devices are typically cyclic, repeating a fixed traffic pattern every period[1]. Mismatch between uplink burst and industrial stream period will cause non-deterministic delay and jitter, which violates strict communication constraints[2].

In this paper, we propose a deep Q-network (DQN) and policy reuse(PR) based upstream scheduling of TDM-PON to optimize frame latency and jitter. In this approach, the Optical Line Terminal(OLT) modeled as an agent interacts with the Optical Network Unites(ONUs). It learns optimal scheduling by iteratively updating Q-function that estimate the expected reward for each state-action pair, with the state defined by the ONU status, the action defined as the allocation of upstream grant to each ONU, and the reward defined as the network performance metrics, such as latency and jitter . In policy reuse, the experience replay buffer of DQN is transferred to cope with the dynamic traffic pattern under the same network environment, so that the scheduling algorithm can quickly adapt to changes and provide the best performance.

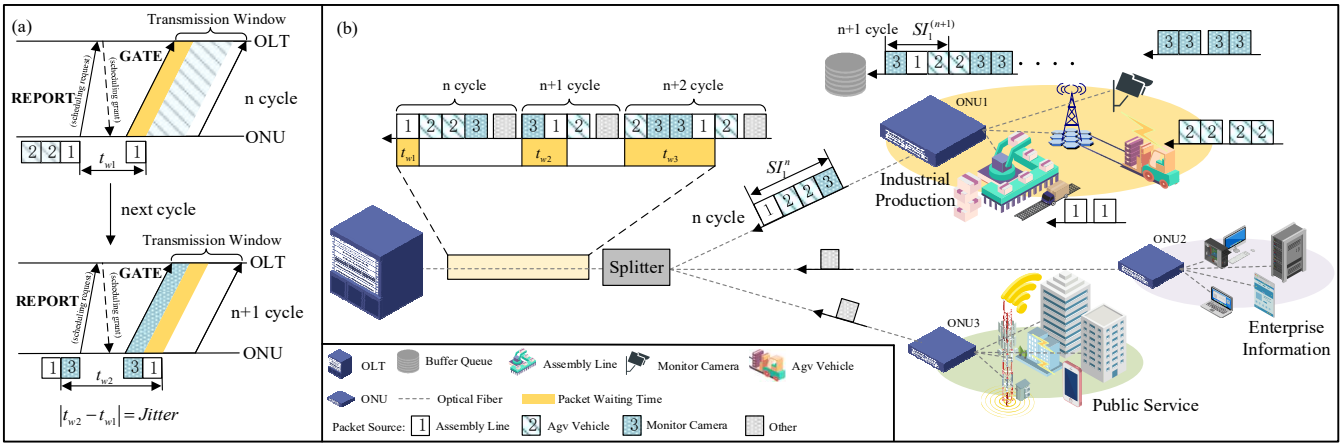

Fig. 1 (a) Packets transmission diagram of upstream under the REPORT/GATE mechanism;(b) an illustration of the industrial park.

II. PROBLEM STATEMENT

Fig. 1(a) presents a typical demonstration of OLT bandwidth scheduling for two transmissions of the same service under the REPORT/GATE mechanism. The occurrence of uncertain delay can be attributed to the discrepant waiting times of packets 1 in the buffer queue, combined with their distinct positions in the transmission window. In this context, we define t_{w1} and t_{w2} as the waiting times for packet 1 during the n^{th} and $(n+1)^{th}$ cycles, respectively. Referring to the formula given in [3], the jitter caused by such uncertainty has the following formula:

$$\left|t_{w2} - t_{w1}\right| = Jitter \tag{1}$$

Fig.1(b) demonstrates a typical industrial park across various dimensions referenced in literature[4], including the networks for industrial production, enterprise information, and public services. Notably, to fulfill the requirements of high bandwidth, ultra-low latency, and high reliability, the TDM-PON is supposed to support and guarantee the performance of multiple services in such specific scenario. Moreover, as depicted in the Fig1.(b), traffic flow aggregated from industrial equipment terminals to ONU can be modeled as cyclic flow[2]. The relative position between the temporal distribution characteristics of flow (i.e., arrival time and period) and the service interval (SI_i^n, denotes service interval for the nth cycle of ONU$_i$) is a factor that affects jitter. Above all, OLT scheduling can be summarized in the following two perspectives:

- ***Horizontal level***: Considering multiple application scenarios in industrial area, PON needs to support multiple service, and enables OLT scheduling to flexibly adapt to the distribution characteristics of flows. Unfortunately, the conventional OLT's single DBA algorithm is primarily designed for fixed equipment and insufficiently flexible to support dynamic demands of modern industrial applications[5].
- ***Vertical level***: Furthermore, in the industrial production network, the cyclic flow converging from the equipment terminal to the ONU fluctuates periodically. To a certain extent, in the same industrial environment, the traffic fluctuations in various time periods have similar rules. Therefore, the scheduling grant of current period can be obtained from the experience replay buffer of the previous period, and reducing the time spent on scheduling.

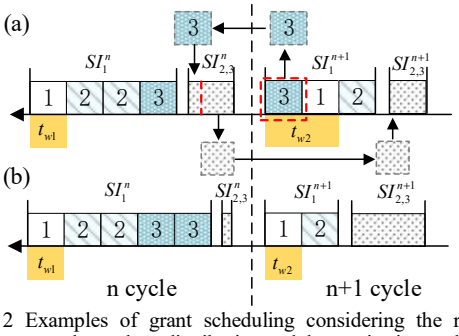

Fig. 2 Examples of grant scheduling considering the relative position between the packets distribution and the service interval, when (a) scheduling and (b) after scheduling, respectively.

III. RL AND PR BASED SCHEME

This paper introduces an algorithm based on policy reuse in reinforcement learning(RL) to optimize industrial streaming delay and jitter. There are two motivations for introducing reinforcement learning into scheduling grant in industrial PON. Firstly, the grant process between OLT and ONU can be modeled as a Markov process, allowing us to apply RL to solve the scheduling problem. Secondly, RL can handle dynamic service indicated in the ***Horizontal level*** across dimensions. The pseudocode of the algorithm is in **Algorithm 1**.

The OLT, acting as an agent, acquires frame information from the buffer queue through the REPORT message. The state space and action space is modeled as follows:

State space: $s \in S\{l_1, l_1, ..., l_m, j_1, j_2, ..., j_3\}$,frame latency and jitters from ONU to OLT.

Action space: $a \in A\{a_1, a_2, ..., a_n\}$,actions taken by the agent in OLT include changing the sending order of the frames in the buffer queue and allocating time slots for the frames. Fig. 2 is an example.

In training process, the action-value function $Q_\pi(s,a)$ is defined as the expected reward of taking action a in state s. Our goal is to find a strategy $\pi(a \mid s)$ to maximize $Q_\pi(s,a)$, optimize relationship between the distribution of packets and the service interval and reduce transmission jitter.

$$\pi = \begin{cases} \pi_{\text{past}}(s), p = \Psi \\ \pi_{new}(s), p = 1 - \Psi \end{cases} \tag{2}$$

Within a certain period of time, OLT grant can be regarded as scheduling problems using the same RL model. As industrial production tasks change, the traffic fluctuation patterns in the network will change accordingly. We assume that their differences in RL model lies in the reward function, which depend on the service-level agreement(SLA)[6]. To, reduce the time for model training

Algorithm: RL and Policy Reuse based algorithm

initialize target values Q', and action-value function Q
initialize Policy Reuse model and replay memory D

For every steps **do:**
  observe current network state s_i and select action a_i with ε
  execute action a_i
  observe new state s_i'
  calculate delay, and throughput in state s_i'
  jitter = abs|delay - delay'|
  if delay > delay_threshold or jitter > jitter_threshold **then**
    reward r_i = -1
  else
    reward r_i = delay_weight * delay + jitter_weight * jitter + throughput_weight * throughput
  store (s_i, a_i, r_i, s_i') in D
  sample random batch of **transitions** (**s, a, r, s'**) from π in formula (2)
  update Q values and DQN model weights by minimizing MSE between Q values and target Q' values
  every fine-tuning steps, fine-tune DQN model on policy reuse model with learning rate
  update target network weights to DQN weights every T steps
  delay' = delay
end For

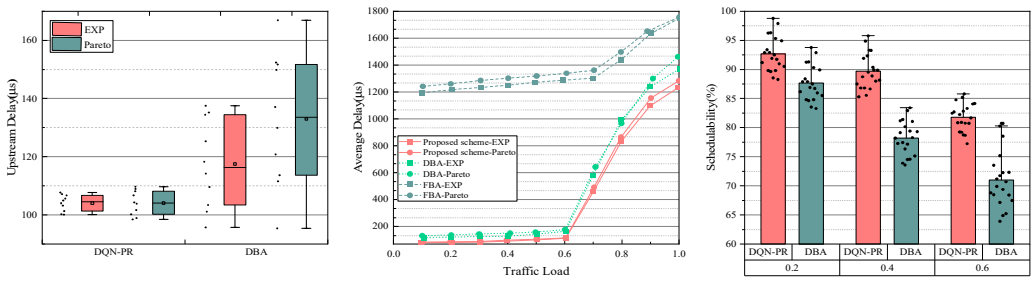

Fig. 3(a) Comparison of upstream delay across different algorithms at 0.1 traffic load; (b) Latency performance of Proposed scheme with different traffic patterns in simulations;(c) Schedulability under different traffic load

we leverage existing value knowledge to accelerate the learning time for similar problems.

The reuse policy in formula (2) aims to gradually reduce the probability of reusing past strategies during the learning process, and correspondingly use $\varepsilon - greedy$ exploration strategy to explore the target task in the increased probability.

IV. Performance Evalution

We conducted a simulation of an XGS-PON system utilizing three ONUs to evaluate the effectiveness of the algorithm, with the uplink rate set to 9953Mbps. To account for the practical conditions of an industrial area environment, the distance between the ONU and OLT was set to 5km. The simulation considers the distribution of two flows: one represents the public service and enterprise traffic, following the exponential distribution of the arrival interval, and the fixed packet size of 100 bytes; the other represents the industrial circulation flow, the packet arrival period is fixed, and the packet size obeys Pareto distribution, with shape parameter k=2, minimum value is 40 Bytes.

Fig. 3(a) presents the distribution of uplink delays among the proposed algorithm, and the DBA algorithm (IPACT [7]) implemented under the REPORTE/GATE mechanism, when the traffic load=0.1. As depicted in the figure, a boxplot with a wider box and longer whiskers may indicate higher jitter. While DBA algorithm assigns scheduling grant to the flows based on the information received in the REPORT message, it ignores the time distribution characteristics of the flows. Especially when the Pareto distribution is used, a higher degree of jitter is observed compared to traffic pattern in Poisson distribution. By contrast, DQN-PR aiming at reducing delay and jitter, considers the distribution characteristics of frames, and adjusts the frame sending sequence. It improves the relationship between the distribution of packets and the service interval, with nice results.

Fig. 3(b) shows the latency of the proposed algorithm, DBA algorithm, and fixed bandwidth allocation(FBA) algorithm increase slowly when the upstream load is low. The worst performance of the latency was observed in the FBA. It can be attributed to the fact that the FBA algorithm allocates a fixed time slot for each ONU, disregarding the dynamics of the flow. As the load on the uplink network increases, the granted for each ONU correspondingly widens, resulting in longer waiting times for data frame transmission and a consequent increase in uplink delay.

Fig. 3(c) shows the schedulability of the network, which is obtained by dividing the number of schedulable traffic by the total traffic. We select three network states as traffic load=0.2, 0.4, 0.6 respectively. The schedulability of the network is obtained by extracting 20 sets of data and calculating the average value under the corresponding algorithm and load. As the load increases, the scheduling of the two algorithms decreases, because the time slot allocated by the OLT to each ONU becomes wider. Comparing the two algorithms, the algorithm proposed in this paper has higher schedulability.

V. Conclution

We propose a DQN-based scheduling scheme for TDM-PON to adapt to time-critical services in industrial area. Q-function and policy reuse are designed to cope with changing traffic patterns and accelerate the learning process, respectively. Simulation results show the proposed scheme effectively reduces the delay and jitter.

Acknowledgment

This work has been supported in part by NSFC project (62122015, 62271075, 62201088) and Fund of SKL of IPOC (BUPT) (IPOC2021ZT04).

References

[1] K. Christodoulopoulos, S. Bidkar, T. Pfeiffer and R. Bonk, "Proof-of-Concept Demonstration of Time Critical Periodic Traffic in Industry-grade Passive Optical Networks," 2022 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2022, pp. 1-3.

[2] C. Su, J. Zhang, H. Yu, T. Taleb and Y. Ji, "Time-Aware Deterministic Bandwidth Allocation Scheme for Industrial TDM-PON," 2022 European Conference on Optical Communication (ECOC), Basel, Switzerland, 2022, pp. 1-4.

[3] M. M. Rahman, M. Hossen and R. Bushra, "Control message scheduling algorithm for improving throughput and jitter performance of the MHSSR DBA scheme of PON," 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 2017, pp. 134-139, doi: 10.1109/ECACE.2017.7912894.

[4] Susan M. Walcott, Industrial Parks, Editor(s): Audrey Kobayashi, International Encyclopedia of Human Geography (Second Edition), Elsevier, 2020, Pages 243-247, ISBN 9780081022962, https://doi.org/10.1016/B978-0-08-102295-5.10084-8.

[5] H. S. Chung, H. H. Lee, K. O. Kim and K. -H. Doo, "Lessons Learn from A Tactile Internet Testbed: An Access Network Perspective," 2021 Optical Fiber Communications Conference and Exhibition (OFC), San Francisco, CA, USA, 2021, pp. 1-3.

[6] Hui Yang, Kaixuan Zhan, Bowen Bao, Qiuyan Yao, Jie Zhang, Mohamed Cheriet,Automatic guarantee scheme for intent-driven network slicing and reconfiguration,Journal of Network and Computer Applications, Volume 190,2021,103163,ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2021.103163.

[7] M. Casoni, "A simulation study of the IPACT Protocol for Ethernet Passive Optical Networks," 2008 10th Anniversary International Conference on Transparent Optical Networks, Athens, Greece, 2008, pp. 309-311, doi: 10.1109/ICTON.2008.4598798.