Dynamic Multichannel Access for URLLC in Industrial Wireless Networks

Sanjay Bhardwaj, Jae-Min Lee and Dong-Seong Kim
Department of IT Convergence Engineering
Kumoh National Institute of Technology
Gumi, Korea
sanjay.b1976@gmail.com, (ljmpaul,dskim)@kumoh.ac.kr

***Abstract*—In industrial wireless networks, real-time transmission increases probability of collision and congestion thus degrading their performance in terms of reliability and latency. In this paper, dynamic multichannel access in industrial wireless networks for ultra reliable low latency communication (URLLC) is considered. For which, partially observable Markov decision process (POMDP) based formulation is carried out for unknown system dynamics of the industrial wireless networks. For the challenges of unknown system dynamics, double deep Q-learning (DDQL) is proposed which deals with two key requirements: reliability and latency Performance evaluation of DDQL is compared with Myopic and Whittle Index based heuristic policy. Simulation results clearly indicate that proposed DDQL finds better and optimal policy that maximizes the rewards in terms of reliability and latency.**

***Index Terms*—Dynamic Multichannel Access, Double Deep Q-Learning, URLLC, Industrial Wireless Networks**

I. Introduction

The dynamic multichannel access is considered crucial to ensuring that the limited spectral resources are correctly allocated to meet users' request, thereby enabling users to select dynamically the available channels. Therefore, authors of [1] explained myopic polices for the collection of the information so that the user senses the channel with highest conditional probability.

Moreover, future of the wireless network is marked by densification, therefore dynamic multichannel access is much more versatile when choosing connectivity, [2]. Multichannel access is foreseen as a potential solution for ultra reliable low latency communication (URLLC) as discussed by [3] particularly in terms of reliability improvement and latency reduction. Authors in [4], [5] proposed deep Q-learning (DQL) for allocation of resources and dragonfly based node identification for URLLC, respectively.

A significant proportion of the allocated channel is sporadically used where a greater section of the channel may be congested under the fixed multi-channel access assignment scheme. The dynamic spectrum access system is considered critical in allowing users to dynamically select the available channels to ensure that scarce spectral resources are correctly distributed to meet user demands.

II. System Model

A dynamic multichannel access is considered, in which a single user makes one choice out of N channels, dynamically, to transmit packets. Each channel can be only in binary state: good (1) or bad (0). Since the user can only senses his selected channel at every slot, the system is not fully observed. Therefore, the problem is formulated as partially observable Markov decision process (POMDP), which finds the exact solution which are having higher complexity due to partial observable states [6]. Thus, dynamic multichannel access problem is modeled as a POMDP. When the channels are separate and distributed identically, the Myopic approach is demonstrated to be efficient under some conditions by the authors of [1]. But, Myopic strategy does not guarantee efficiency when channels are linked or obey specific distributions.

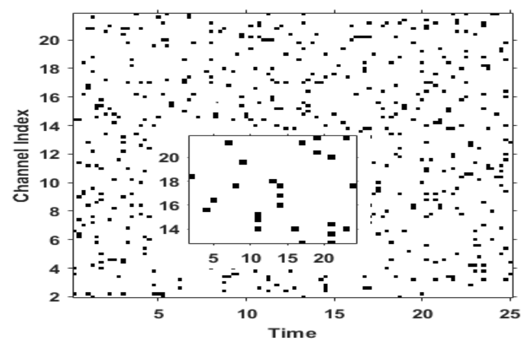

Fig. 1: Multiple Good Channels State (Black Square)

Every time a slot begins, a user selects a channel to sense and send a packet. User receives a reward of +1 or -1, for a successful or failed transmission, respectively. Thus, the rewards describes the channels as good or bad. The objective is to design a policy that maximizes the expected reward, which herein are latency and reliability.

For the learning of the unknown environment double deep Q-learning (DDQL) is proposed which can easily overcome prohibitive computational requirements, as well finds the channel policy (good and optimal) without the knowledge of system dynamics.

III. Proposed Scheme

The goal of DDQL is to define an optimum strategy, i. e., a sequence of actions to maximize reward r_t and to find the Q value of each state and action a_t pair. The policy π for (a_t, s_t) pair, is given by $\pi(a_t, s_t)$, is defined sum of

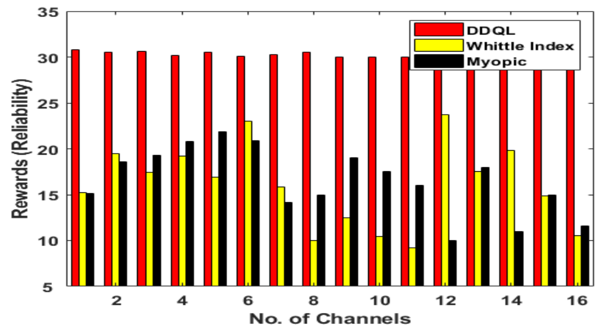

Fig. 2: Reward (Reliability) with Increase in the Number of Channels

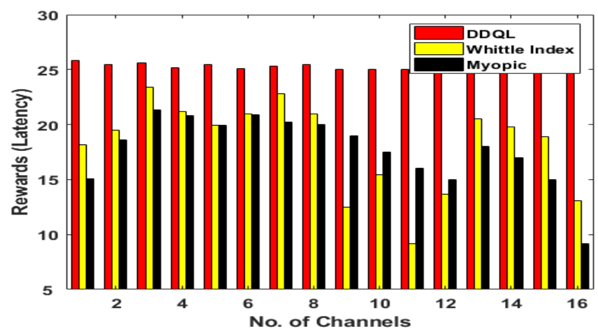

Fig. 3: Reward (Latency) with Increase in the Number of Channels

the reward it receives as it takes action a_t at the state s_t. The Q-value when following the optimal policy π^*, for the pair of state-action is given as $\pi^*(s_t, a_t)$, therefore $\pi^*(s_t) = argmaxa[\pi^*(s_t, a_t)], \forall s_t$. For each time slot, its assumed that the agent takes action at $\in \{1, ..., N\}$ such that for the pair (s_t, a_t), its Q-value is maximized and it gains a reward (r_{t+1}). Then, Q-values with learning rate $0 < \alpha < 1$ is given:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(R_{t+1} + \gamma Q^{'}(s_{t+1}, a) - Q(s_t, a_t) \right), \tag{1}$$

where, a is best action, $Q^{'}$ is expected value of Q and γ is the discount factor.

A. DDQL Architecture

The structure of DDQL is a fully connected neural network with two hidden layers containing 300 neurons each along with rectified linear unit (ReLU) as activation function of each neuron. Along with this, ϵ-greedy policy is applied, with $\epsilon = 0.1$. Minibatch of 40 samples are selected randomly to update weights of DDQL. Along with this, number of channels (N) is 16, and rest of the parameters are shown in Table : I.

IV. Performance Evaluation

For evaluation, it is assumed that only good channel switches according to round robin scheduling. The switching probabilities is shown in Fig. 1, in which for each time slot multiple good channels are available (black squares). The

TABLE I: Simulation Parameters

Parameters	Value
Learning Rate(α)	10^{-5}
Discount Factor (γ)	0.9
Iterations	10^5
No. of re-transmissions	5
Pay load size	1024 Bytes

proposed DDQL is compared with Whittle index and Myopic policy in terms of the rewards for reliability and latency, Figs. 2 and 3, respectively, for $N = 16$. The rewards for DDQL for both reliability and latency is better, highest and consistence in comparison to the other two policies. This clearly shows that DDQL learns the system dynamics including correlations among the channels and thus finds optimal policy.

V. Conclusion

Complex multichannel access is considered in this paper when channels are correlated and URLLC statistics of Industrial wireless networks are unknown. The issue is an unknown POMDP without a traceable solution, which uses the DDQL method explicitly to find an optimal access strategy for round robin channel switching. Through simulations, DDQL is able to achieve the optimal performance in terms of rewards (latency and reliability) even without any prior knowledge. Thus compounds that DDQL based dynamic multichannel access industrial wireless networks provides for high reliability and low latency i.e. supports URLLC.

Acknowledgement

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2018R1A6A1A03024003), the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2020-2020-0-01612) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation).

References

[1] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multichannel opportunistic access: structure, optimality, and performance," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5431–5440, 2008.

[2] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2795–2808, 2010.

[3] N. H. Mahmood, D. Laselva, D. Palacios, M. Emara, M. C. Filippou, D. M. Kim, and I. de-la Bandera, "Multi-channel access solutions for 5g new radio," in *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*. IEEE, 2019, pp. 1–6.

[4] S. Bhardwaj, R. R. Ginanjar, and D.-S. Kim, "Deep q-learning based resource allocation in industrial wireless networks for urllc," *IET Communications*, vol. 14, no. 6, pp. 1022–1027, 2020.

[5] S. Bhardwaj and D.-S. Kim, "Dragonfly-based swarm system model for node identification in ultra-reliable low-latency communication," *Neural Computing and Applications*, pp. 1–44, 2020.

[6] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of markov decision processes," *Mathematics of operations research*, vol. 12, no. 3, pp. 441–450, 1987.