Delay-aware TDMA Scheduling with Deep Reinforcement Learning in Tactical MANET

Gwangjin Wi, Sunghwa Son, Kyung-Joon Park
Department of Information and Communication Engeenering
Daegu Gyeongbuk Institute of Science and Technology (DGIST)
Daegu, Republic of Korea
Email: {wgj2050, ssh, kjp}@dgist.ac.kr

***Abstract*—In tactical networks, traffic should be delivered in a timely manner satisfying the quality of service (QoS) requirements for survivability and mission success. In this paper, we propose a centralized TDMA slot scheduling based on deep reinforcement learning (DRL) to guarantee the QoS requirements by minimizing end-to-end delay. We consider situations in which mission criticality of tactical traffic is dynamically changing. We introduce a DRL actor-critic algorithm to find a TDMA scheduling policy to minimize the weighted end-to-end delay which is a new metric reflecting the mission criticality of tactical traffic. The simulation results verify that the proposed scheduling policy can guarantee QoS requirements in tactical networks.**

***Index Terms*—Tactical networks, quality of service, deep reinforcement learning, TDMA slot scheduling**

I. Introduction

QoS guarantee of tactical traffic such as low latency and reliability is a critical in tactical networks due to high uncertainty and resource constrained environment [1]–[4]. Tactical nodes typically consist of unmanned ground vehicles (UGV) and unmanned aerial vehicles (UAV) in tactical MANET. The mission of UGV and UAV is to monitor and patrol the battlefield. Each tactical node operates missions requested from the center node and sends the results of the mission to the center node. To receive the tactical traffic in a timely manner at the center node, the QoS requirements of tactical traffic should be guaranteed.

Contention based carrier-sense multiple-access with collision avoidance (CSMA/CA) is a widely used MAC protocol. However, CSMA/CA is not suitable for tactical networks that require highly reliable communication because there is a possibility of collision. Instead, the non-contention-based TDMA MAC protocol enables reliable transmission over tactical networks. Tactical traffic has priority depending on mission criticality such as high-priority traffic and low-priority traffic. Tactical traffic with low-priority has no specific QoS requirements. Hence, we focus on improving the performance of low-priority traffic while guaranteeing the QoS requirements of traffic with high-priority.

In this paper, we propose a centralized TDMA slot scheduling using a deep reinforcement learning (DRL) algorithm that minimizes the weighted end-to-end delay. The considered network is a tree topology that is typical in tactical networks. Recently, DRL has been applied recently in many networking problems [5]–[10]. In DRL models, there are two kinds of policies: value-based and actor-based [11]. We use actor-critic algorithm to obtain the advantages of both value-based method and actor-based method. The weighted end-to-end delay is a new metric reflecting the mission criticality of tactical traffic. Considering the mission criticality of tactical traffic that changes dynamically, we focus on finding a low-latency scheduling policy. We formulate the problem as an integer program and solve it using the DRL actor-critic algorithm. Our simulation results show that the proposed scheme guarantees the required QoS of tactical traffic.

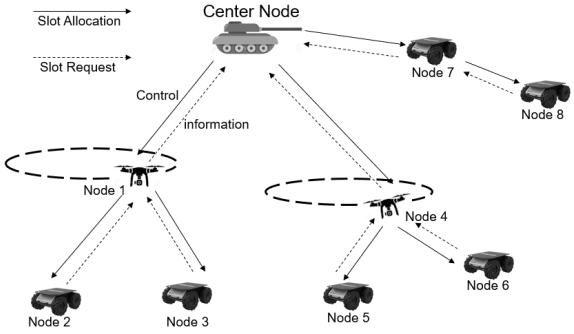

Fig. 1. TDMA based multi-hop tactical MANET.

II. System Model

A. Preliminaries

Fig. 1 describes the network topology of considered tactical multi-hop ad-hoc networks. Let $\mathcal{V} = \{v_1, v_2, \ldots, v_k\}$ denote the set of tactical nodes. Each tactical node is a UGV or UAV for surveillance and reconnaissance. All tactical traffic is transmitted to the center node and we assume two-hop interference model.

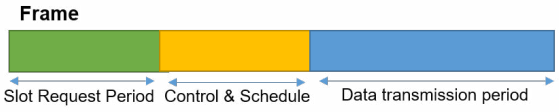

Fig. 2. TDMA frame structure of tactical MANET.

The superframe of the tactical node consists of three parts: slot request period, control and schedule period, data transmission period as shown in Fig. 2. During the slot request

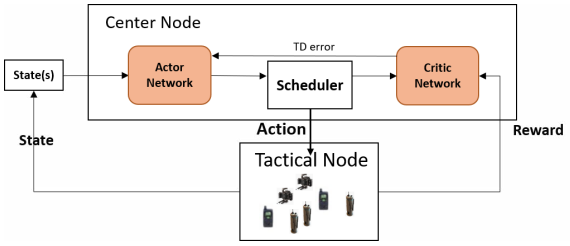

Fig. 3. The structure of the DRL model.

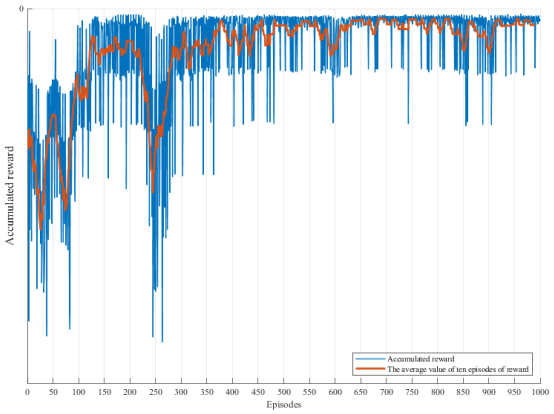

Fig. 4. The accumulated reward of the learning process.

period, each tactical node requests the number of time slots for data transmission of the next period. Let $\mathcal{B} = \{b_1, b_2, \ldots, b_k\}$ denote the number of time slots tactical nodes request. The length of the data transmission period can vary in order to assign all the requests, $\mathcal{B}$, from tactical nodes. In the control and schedule period, the central node transmits the command message and scheduling information to the tactical nodes. Tactical nodes send tactical traffic to the central node based on the scheduling information in the data transmission period.

Tactical traffic has different priority and mission criticality. Let $\mathcal{W} = \{w_1, w_2, \ldots, w_k\}$ denote the set of mission criticality assigned to each tactical node. We define tactical traffic as high-priority traffic, if the weight value, w, exceeds a certain threshold. The weight value changes according to the mission of tactical nodes. For performance evaluation, we introduce the weighted end-to-end delay as a QoS metric.

B. Optimization Formulation

The objective of the proposed TDMA scheduling scheme is to find an efficient scheduling policy to minimize the weighted end-to-end delay. The proposed scheme operates based on a centralized scheduling: the central node gathers information of tactical nodes and determines the scheduling policy for low-latency transmission.

The overall optimization problem of minimizing the weighted end-to-end delay of all tactical nodes can be formulated as follows.

$$\text{minimize} \sum_{j \in Neigh_C} \sum_{t=1}^{T} w_{I_j^t} X_j^t t, \tag{1}$$

where $Neigh_C$ is the neighbor nodes of center node, I_j^t represents the source node of the packet that has the highest mission priority at node v_j in time slot t, $w_{I_j^t}$ is the weight of I_j^t, and X_j^t denotes the scheduling state of v_j in a time slot t and $X_j^t \in \{0, 1\}$. The optimization problem (1) has two constraints: 1) The maximum interference range of the tactical nodes is two-hop. 2) Tactical traffic with high-priority should be transmitted within a specific deadline. Here, we focus on the low-latency scheduling to minimize the weighted end-to-end delay of tactical traffic.

C. Scheduling with Deep Reinforcement Learning

In tactical MANET, tactical nodes operate various missions and the mission criticality of tactical nodes changes dynamically. To solve the problem (1), we use a scheduling policy based on DRL actor-critic algorithm. The architecture of the proposed model is illustrated in Fig. 3. Next, we describe state space, action space, and rewards of tactical network in the proposed DRL actor-critic model.

State Space. A state of tactical network represents a set of status of tactical nodes. A state of tactical nodes in a time slot t includes the number of packets, hop count from the center node, and minimum value of deadline among packets. Let s_t denote the state of the network at time t, $s_t = \{s_{packet,0}, s_{hop,0}, s_{dl,0}, \ldots, s_{packet,k}, s_{hop,k}, s_{dl,k}\}$. Note that $s_{packet,i}$ denotes the number of packets tactical node v_i has and $s_{hop,i}$ denotes hop count from the center node to v_i. $s_{dl,i}$ denotes the shortest time to deadline among packets in tactical node v_i.

Action Space. In tactical network, the center node has k tactical nodes to schedule. Therefore, action space size is set to 2^k. In order to reduce the large action space, we exclude actions which fail to meet the constraints from the assumption of the maximum interference range of the tactical node is two-hop. At each time t, action of the scheduler can be represented as follows, $a_t = \{X_0^t, X_1^t, \ldots, X_k^t\}$.

Rewards. The objective of the proposed scheduling scheme is minimize weighted end-to-end delay of all tactical nodes. In each time slot t, we define reward, $r_t = -\sum_{j \in Neigh_C} \sum_{t=1}^{T} w_{I_j^t} X_j^t t$. In case of tactical traffic with high-priority fails to transmit within the deadline, we give additional negative rewards to the center node.

III. Evaluation

In this section, we provide simulation results to show the performance of the proposed scheduling policy. For performance evaluation, we set the topology as a multi-hop structure with one center node and ten mobile tactical nodes. Each node has a weight value between 0 and 1, and the node that has a weight value larger than 0.9 is defined as high priority traffic. Ten tactical nodes consist of 8 low-priority traffic and 2 high-priority traffic. In order to learn the environment in which mission criticality changes dynamically, we change the weight of tactical nodes dynamically during learning time.

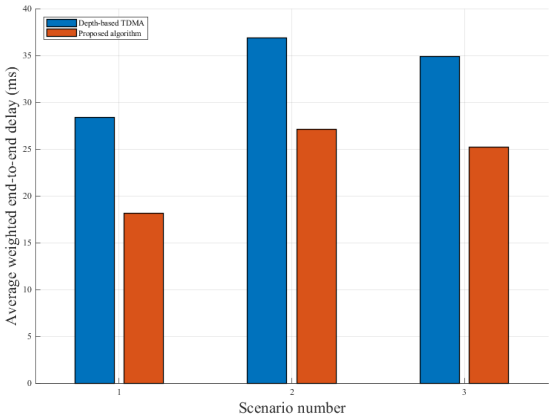

Fig. 5. Average weighted end-to-end delay with three different scenarios.

Fig. 4 shows the learning process of the proposed DRL actor-critic scheduling policy. The x-axis indicates the episode number and the y-axis represents the accumulated reward. The red line represents the average accumulated reward of ten episodes. As the learning process progresses, the proposed scheduling policy tries to minimize the reward, and around episode 400 the reward becomes stable.

Fig. 5 shows average weighted end-to-end delay of tactical traffic for the proposed algorithm and depth-based TDMA according to three scenarios, respectively. We define three different scenarios that a position of tactical nodes which generate high-priority traffic changes in the same network topology. The result of each scenario is an average value for 1000 simulation times. For performance comparison, depth-based TDMA is used. The proposed algorithm uses the DRL actor-critic model which is trained from Fig. 4. The simulation results show that the proposed algorithm can reduce the weighted end-to-end delay of low-priority traffic while the QoS requirement of high-priority traffic is guaranteed.

IV. Conclusion and Future Work

The QoS of tactical traffic such as low-latency and reliability should be guaranteed for survivability and mission success. Since the environment of tactical networks is uncertain and dynamic, we have considered the situation in which the mission criticality changes dynamically. We have introduced a deep reinforcement learning model to schedule TDMA time slots for tactical nodes. The simulation results have shown that the proposed algorithm achieves low-latency transmission under dynamic change in mission criticality. As future work, we will extend our centralized DRL actor-critic algorithm to distributed DRL scheduling.

Acknowledgment

This work has been supported by the Future Combat System Network Technology Research Center program of Defense Acquisition Program Administration and Agency for Defense Development (UD190033ED).

References

[1] H. Chenji, Z. J. Haas, and P. Xue, "Low complexity QoE-aware bandwidth allocation for wireless content delivery," in *IEEE Military Communications Conference (MILCOM)*, 2015, pp. 419–425.

[2] J. P. Hansen, S. Hissam, D. Plakosh, and L. Wrage, "Adaptive quality of service in ad hoc wireless networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2012, pp. 1749–1754.

[3] P. Łubkowski, M. Hauge, L. Landmark, C. Barz, and P. Sevenich, "On improving connectivity and network efficiency in a heterogeneous military environment," in *IEEE International Conference on Military Communications and Information Systems (ICMCIS)*, 2015, pp. 1–9.

[4] N. Boulila, M. Hadded, A. Laouiti, and L. A. Saidane, "QCH-MAC: A QoS-aware centralized hybrid mac protocol for vehicular ad hoc networks," in *IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, 2018, pp. 55–62.

[5] Y. Xu, J. Yu, W. C. Headley, and R. M. Buehrer, "Deep reinforcement learning for dynamic spectrum access in wireless networks," in *IEEE Military Communications Conference (MILCOM)*, 2018, pp. 207–212.

[6] Y. He, F. R. Yu, N. Zhao, H. Yin, and A. Boukerche, "Deep reinforcement learning (DRL)-based resource management in software-defined and virtualized vehicular ad hoc networks," in *Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications*, 2017, pp. 47–54.

[7] A. Ghaffari, "Real-time routing algorithm for mobile ad hoc networks using reinforcement learning and heuristic algorithms," *Wireless Networks*, vol. 23, no. 3, pp. 703–714, 2017.

[8] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1277–1290, 2019.

[9] B. Peng, G. Seco-Granados, E. Steinmetz, M. Fröhle, and H. Wymeersch, "Decentralized scheduling for cooperative localization with deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4295–4305, 2019.

[10] R. Atallah, C. Assi, and M. Khabbaz, "Deep reinforcement learning-based scheduling for roadside communication networks," in *IEEE 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2017, pp. 1–8.

[11] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *4th International Conference on Learning Representations, ICLR*, 2016.