DRL-Based Channel and Latency Aware Scheduling and Resource Allocation for Multi-User Millimeter-Wave RAN

Shuyi Shen[1], Ticao Zhang[2], Shiwen Mao[2], and Gee-Kung Chang[1]
1. Georgia Institute of Technology; 2. Auburn University.
ssyzoe@gatech.edu

Abstract: A DRL-based uplink resource allocation algorithm with channel condition and latency awareness is demonstrated for multi-user RAN. The algorithm is verified experimentally with dynamic RoF-mmWave channels, achieving 19% reward improvement compared to conventional scheduling schemes.

1. Introduction

5G new radio access network (RAN) is envisioned to support multiple users with applications such as video streaming, low-latency gaming, and for real-time services including robotics, intelligent factory, telehealth care, etc. 5G is a service-oriented system, therefore the RAN is anticipated to fulfill the quality-of-service (QoS) requirements of various applications [1-2]. This will add more challenges to the scheduling and radio resource management (RRM), with stringent latency requirements and complex QoS objectives. Moreover, millimeter wave (mmWave) links are implemented for 5G RAN, which can result in dynamic channel conditions that add to the RRM complexity [3]. For example, mmWave in Frequency Range 2 (FR2, 24.25 to 52.6 GHz) can be more susceptible to line-of-sight (LoS) blockage with high path loss, as indicated in Fig. 1(a). As a result, resource block (RB) allocation and flow-RB mapping are required to be adaptive to the dynamically changing channels. In this work, photonic-assisted mmWave generation is utilized to achieve wide-bandwidth transmission and experimentally verified channel variations.

In RAN, a scheduler can be deployed at the central unit or distributed unit (CU/DU). In each Transmission Time Interval (TTI), the scheduler must solve a decision-making problem to decide radio resource allocation. In the presence of multiple UEs, complicated QoS requirements, and dynamic wireless environment in 5G RAN, it is challenging to obtain an optimal and adaptive solution using the existing rule-based scheduling schemes with a single objective. Recently, deep reinforcement learning (DRL) has made breakthroughs in fiber-wireless communication systems. In time-varying networks, DRL has proved to be effective in tackling real-time decision-making problems [4].

In this paper, we incorporate DRL to address the problem of delay and channel condition aware packet scheduling and RB allocation in the uplinks of service-oriented mmWave RAN. As shown in Fig. 1(a), the system will consider multi-user multi-service scenarios with different QoS requirements. In contrast to most of the previous DRL-related works with only simulation validated results, the mmWave channel characteristics utilized in the proposed system are experimentally investigated and verified via radio-over-fiber (RoF) mobile fronthaul. In this paper, the channel environment parameters are determined by following 3GPP 5G standards, the implementation of request-grant scheduling cycle, and dynamic mmWave channels, etc. The proposed DRL-based system can achieve improved QoS performance when latency and bit-error-rate (BER) are considered simultaneously.

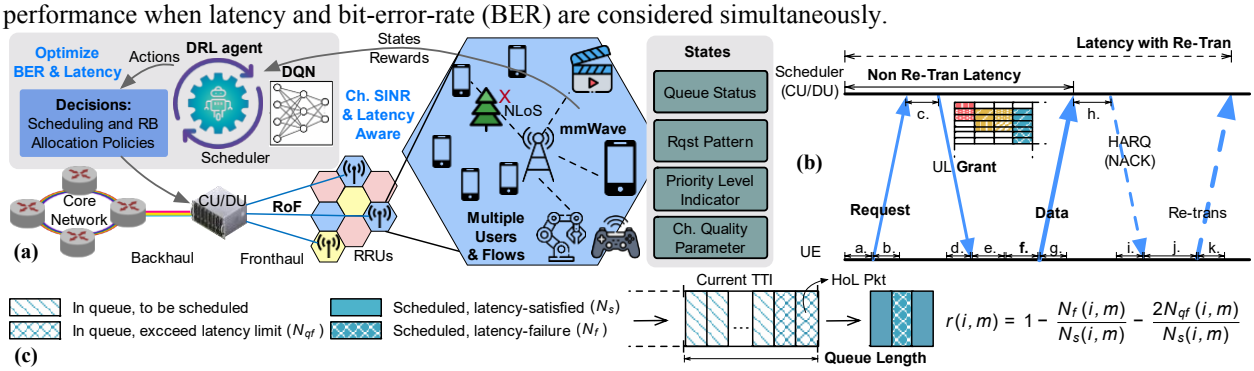

Fig. 1 (a) System architecture. (b) Uplink scheduling process. (c) An example of flow packet and queue status.

2. System Architecture, DRL Design, and Experimentally Verified mmWave Channels

We consider the uplink transmission of a mmWave remote radio unit (RRU) supported by RoF mobile fronthaul as shown in Fig. 1(a). The system is flow-oriented and involves multiple users that are using applications with different QoS requirements. One UE can have multiple active services/flows simultaneously. The packet arrival pattern, QoS priority, latency budget, and other key flow-specific parameters are summarized in Table I. The parameter design is based on [1,3,5]. At each TTI, UEs will firstly request transmission opportunities before data transmission. The scheduler will process the request and then distribute the grants. Not all requests can be satisfied especially when the

traffic load is heavy, which may cost additional queuing delay. The UEs will then prepare and send the data packets in the allocated RBs. Upon receiving the uplink data, the scheduler will check the BER of the received data which determines whether re-transmission is required. For simplicity, the queuing delay is considered only for the original data transmission while not for the re-transmission. Guaranteed re-transmission channels are assumed. The request-grant process is visualized in Fig. 2(b) and the delays of different stages are summarized in Table II. The delay parameters are based on [6], where 2km standard single-mode fiber (SMF) and 50m wireless distance is assumed.

Table I: Flow Characteristics

Service Type	Priority	UE	Pkt Size	Pkt Interval	Speed (Mbps)	Delay Bdgt.
Robotics	30	1	Rand	Cont.	300-350	1ms
Video Streaming	56	1	Log Norm.	Poisson	10	5ms
Gaming/Factory	30	2	Gaussian	Fixed	3	1ms
Health Care	56	2	Poisson	Cont.	300	2ms

Table II: Delay Components

Propagation Delay	9.59µs	b, d, g, i, k
UE Processing	0.32ms	a, e, j
Scheduler Processing	62.98µs (14 sym.)	c
Re-tran Procssing	0.21ms	h
Queuing Delay	Traffic-based	f

Table III: Scheduling Policies (Actions)

Policy	Feature	Objective
Max-SINR	Channel	Best BER
Propotional Fair (PF)	Channel & Speed Aware	Fairness & Throughput
LOG Rule (LOG)	Channel-Speed-Delay Aware	Fairness & Bounded Delay
Exponential Rule (EXP)	Channel-Speed-Delay Aware	Fairness & Bounded Delay

Table IV: OFDM and RG Numerologies

Numerology, μ	4	TTI Duration	0.0354ms
Subc. spacing	240kHz	RB size	12 subcs.
Effective subc. #	840 (/2048)	RG size in freq.	5RB/60 subc.
Effective BW	201.6MHz	RG size in time	2 sym. dur.
Sym. # per TTI	8	Modulation	QPSK/16QAM

At each TTI, the requests of flows, the head-of-line (HoL) latency of queues, the priority level of the flows, and the channel quality parameters of UEs constitute the states/input features of the DRL agent. The action of the DRL-based scheduler is to select the optimal scheduling and RB allocation policy for the current TTI. The candidate action policies are summarized in Table III. Different policies can achieve different scheduling objectives [7]. By choosing different policies TTI-by-TTI adaptively with respect to the channel and traffic conditions, improved QoS performance can be achieved, as opposed to using a single policy over the entire process. Note that the action in the proposed system is not specific RB-flow mapping. In fact, in one of our previous works the DRL action is through direct RB allocation and mapping, which can lead to extreme computational complexity if used in the complicated multi-user high-speed mmWave RAN [4]. The reward of the DRL system is designed as follows: at each TTI for each flow, all packets that have been requested will be categorized to four types as depicted in Fig. 1(c). The reward of flow *m* at TTI *i* is $r(i,m) = 1 - \frac{N_f(i,m)}{N_s(i,m)} - \frac{\mathbf{2}N_{qf}(i,m)}{N_s(i,m)}$, in which N_s is the number of scheduled packets whose overall latency satisfy the delay budget requirements, N_f is the number of scheduled packets whose overall latency exceeds the delay budget, while N_{qf} is the number of packets in queue waiting to be scheduled with their current latency already exceeding the delay budget. In *r(i,m)*, the second term $-\frac{N_f(i,m)}{N_s(i,m)}$ reflects negative feedback if the current scheduling method results in too large latency, whereas the third term with the weight factor *2* indicates more significant negative feedback, to prevent latency-failure packets from queuing up and leading to large queuing delay. The overall reward per TTI will be the weighted sum of all flows: $r(i) = \sum_m w(m)r(i,m)$, where $\sum_m w(m) = 1$.

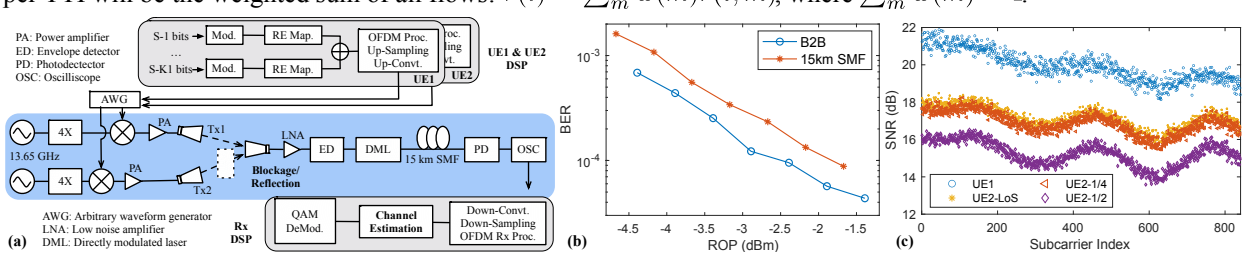

Fig. 2 (a) Experimental setup. (b) Testbed BER performance versus ROP. (c) SNR per subcarrier of both UEs with 16QAM in different scenarios.

The experimental setup to obtain the channel information is depicted in Fig. 2(a), in which two UEs are accessing a remote radio unit (RRU) simultaneously. Due to the devices available in our lab, there are two UEs and four flows in the system without loss of generality. In reality, more UEs can be implemented. UE-flow mapping is indicated in Table I. For each UE, the signal-to-interference-and-noise ratio (SINR) of each RB will be measured and used as channel quality parameter for the scheduling processing. For more efficient processing, RBs are grouped as a resource group (RG) when being allocated. Subcarriers and symbols in one RG have the same QAM modulation. The OFDM numerology and frame design are based on 3GPP 5G specification [8]. The OFDM and RG numerologies are summarized in Table IV. The BER versus received optical power (ROP) performance of the testbed is shown in Fig. 2(b). For the channel characteristics measurement, the testbed is set at the optimal operating condition (ROP = -1.5dBm). To realize the dynamic channel conditions of mmWave links such as reflection, blockage, and reduced transmission power, channel variation is introduced for UE2. The channel of UE2 is measured with three conditions:

1) with LoS link; 2) the link is 1/4 blocked (slightly blocked); 3) the link is 1/2 blocked (severely blocked), while UE1 always has an LoS link. The experimentally measured channel SNR is shown in Fig. 2(c). In the scheduling process, each channel condition will last for 50 TTIs and randomly switch to the next condition. Different channel conditions and policy selection will lead to different flow BER performance. Upon decoding the received signals, the scheduler will check the packet BER per flow. If the BER exceeds the pre-set threshold (6.9×10^{-4} in this paper considering forward error correction [9]), re-transmission will be triggered, and the overall packet latency will become longer.

3. Results and Discussions

We create a deep Q-network (DQN) agent with recurrent neural network (RNN). There are three hidden layers between the input layer and the output layer: two dense layers and one long short-term memory (LSTM) layer, which have 30, 20, 16 neurons, respectively. The training discount factor is 0.99. The experience replay length is 10^6. The DRL agent is trained over 1000 episodes with each episode consisting of 1000 TTIs. The convergence plot of the training process is presented in Fig. 3(a). It can be seen that after around 600-episode training, the reward begins to converge. The fluctuation of the converged reward is caused by the randomness of traffic pattern as indicated in Table I. Generally, the maximum average reward (1000) per episode can be achieved if the traffic is light. However, in that case, the DRL agent can randomly choose any scheduling policy to fulfill the latency requirement. Therefore, the traffic load in the paper is set to a heavier case to exploit the advantages of DRL.

The DRL system is tested over 1000 TTIs with randomly generated flow patterns. The reward results are shown in Fig. 3(b). The case of randomly selecting policies TTI-by-TTI is also presented as a reference. A higher reward value indicates higher percentages of latency-satisfied packets. It can be seen that the proposed DRL system can achieve average $1 - \frac{N_f}{N_s} = 0.87$, indicating the average ratio of latency-satisfied packets $\frac{N_s}{N_f+N_s} = \frac{1}{(N_f+N_s)/N_s} = \frac{1}{N_f/N_s+1} = \frac{1}{(1-0.87)+1} = 88.5\%$. However, among single-policy cases, LOG-rule can achieve the best reward of 0.73. In comparison, the proposed DRL algorithm can achieve 19% reward improvement.

Fig. 3(c) presents the policy selection per TTI with respect to the channel variation of UE2. The blue curve indicates the SNR fluctuation of UE2, from which it is shown that each channel state lasts for 50 TTIs. As the policy selection can be jointly affected by channel variation and flow request pattern, it can be seen that the pattern of policy selection synchronizes well with the channel SNR variation. The results show that the DRL system can react adaptively with channel condition variations.

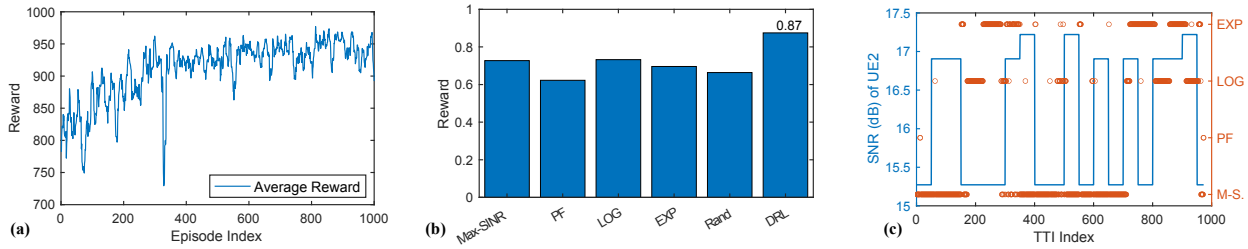

Fig. 3 (a) Reward convergence of the training. (b) Rewards of the test set. (c) UE2 SNR variation and the corresponding policy selection per TTI.

4. Conclusion

A DRL-based scheduler operating with both latency and channel condition awareness is proposed and verified for service-oriented multi-user mmWave RAN. The DRL system is verified with experimental validation of RoF-mmWave channel conditions and variations, as well as various service flows with different QoS requirements. Results show that the proposed DRL system can operate adaptively with channel variations and achieve at least 19% reward improvement compared to conventional single-rule schemes. The proposed DRL system provides a promising AI/ML-based technique that are applicable to the upcoming 6G RAN systems.

Acknowledgement: This work is supported in part by the NSF under Grants CNS-1539976 and CNS-1822055.

References

[1] I. Parvez, et al., "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," in IEEE Commun. Surv. Tutor., vol. 20, no. 4, pp. 3098-3130, Fourthquarter, 2018.
[2] System Architecture for the 5G System (5GS), TS 23.501., V16.2.0., 3GPP, Sept. 2019.
[3] NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone., TS 38.101-2., V16.5.0., 3GPP, Oct. 2020.
[4] T. Zhang, et al., "Delay-aware Cellular Traffic Scheduling with Deep Reinforcement Learning," Proc. IEEE GLOBECOM, Taipei, Dec. 2020.
[5] Greg White, et al. "Low Latency DOCSIS: Technology Overview," DOCSIS Research and Development, CableLabs, Feb. 2019.
[6] Erik Dahlman, et. al., "NR: the next generation wireless access technology," Ericsson.
[7] A. Ahmad, et al., "Resource allocation algorithms in LTE: A comparative analysis," 2015 IEEE India Conference, New Delhi, India, pp. 1-6.
[8] NR; Physical channels and modulation., TS 38.211., V.16.2.0., 3GPP, Jun. 2020.
[9] Juniper, "Forward Error Correction (FEC) and Bit Error Rate (BER)," Dec. 2020.