IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 978-1-6654-0926-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INFOCOMWKSHPS54753.2022.9797936

Demo: Deep Reinforcement Learning for Resource Management in Cellular Network Slicing

Baidi Xiao, Yan Shao, Rongpeng Li, Zhifeng Zhao, Honggang Zhang

***Abstract*—Network slicing is considered as an efficient method to satisfy the distinct requirement of diversified services by one single infrastructure in 5G network. However, owing to the cost of information gathering and processing, it's hard to swiftly allocate resources according to the changing demands of different slices. In this demo, we consider a radio access network (RAN) scenario and develop several deep reinforcement learning (DRL) algorithms which can keenly catch the varying demands of users from different slices and learn to make an intelligent decision for resource allocation. Besides, in order to implement and evaluate our algorithms efficiently, we have also implemented a platform with a modified 3GPP Release 15 base station and several on-shelf mobile terminals. Numerical analyses of the corresponding results verify the superior performance of our methods.**

I. Introduction

Nowadays, network slicing plays a quite important role in 5G network communication. Empowered with Software Defined Network (SDN) and Network Functions Virtualization (NFV) [1], the physical resource of network can be virtualized in a logical network. Each slice corresponding to a specific service will be allocated different amount resource (e.g., bandwidth) according to the demands. Given the mobility of subscribers and practical characteristics of distinct services, the allocation strategy shall be sufficiently flexible, so as to agily adapt to the incessant variations of users' demands. However, there exists endogenous uncertainty embedded in the demands, which makes the design of a flexible allocation strategy challenging.

Deep reinforcement learning (DRL) is commonly used to solve problems in Markov Decision Process (MDP) by interacting with an unknown environment to learn an optimal policy [2]. In this paper, we primarily leverage DRL for resource management in cellular network slicing. In particular, based on classic DRL algorithms, we have proposed several variants DRL algorithms [3]–[5] to better meet the unique features of a radio access network (RAN). Besides, a platform with a modified 3GPP Release 15 base station (i.e., gNB) and several on-shelf mobile terminals (MTs) is built up to further evaluate these algorithms and verify their effectiveness.

II. The Algorithms for Intelligent Slicing

We consider a RAN scenario where subscribers' demands for services changes frequently due to their mobility and intermittent requests. A gNB dynamically allocates its bandwidth (i.e., resource block (RB)) to different slices, each corresponding to one service. The primary objective function for a resource allocation strategy is to maximize the average service agreement lever (SLA) satisfaction ratio (SSR) as well as spectrum efficiency (SE).

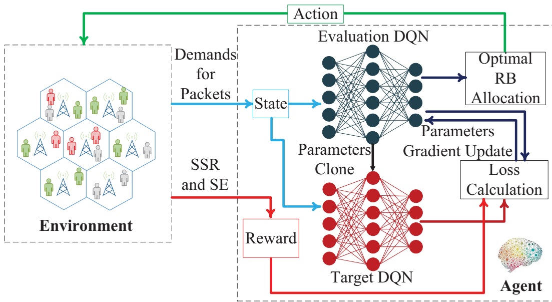


Fig. 1. Interaction between the RAN environment and the DRL agent.

As depicted in Fig. 1, the RAN scenario can be regarded as the environment of MDP with slices demands as *states* and allocated bandwidth for each slice as *actions*. Afterwards, we leverage DRL to design a real-time resource management strategy by continuously perceiving the demands and allocate the resources accordingly, and propose several variants of DRL to better accommodate the characteristics of RAN (e.g., random fading channels, moving subscribers).

- **GAN-DDQN** [3]. For the avoidance of disturbance of random noise when calculating SSR and SE, we use distributional RL to learn the state-action value distribution instead of the expectation. The distributional RL increases robustness to hyperparameter variations and environment noise, as well as avoids the action-value overestimation or underestimation issues. Correspondingly, generative adversarial network (GAN) is resorted to have a better approximation of $\mathcal{Q}$-distribution. The generator network works with output of samples from approximated and target $\mathcal{Q}$-distribution, while the discriminator network tries to distinguish the target action values from the approximated one.
- **LSTM-A2C** [4]. Faced with users' movement between different gNBs, the number of subscribers attached to slices of each gNB becomes relevant and tractable. LSTM-A2C is proposed to utilize this temporal relevancy and unveil the action regularity of users by LSTM, by feeding its output into the A2C network, which is another kind of classic DRL algorithm.
- **GAT-DQN** [5]. Considering a more complicated scenario with multiple gNBs in a dense cellular network, instead of learning the optimal policy by individual gNBs, a graph attention network (GAT)-based multi-agent reinforcement

This work was supported in part by the National NSF of China under Grants 61731002 and 62071425, in part by the Zhejiang Key R&D Plan under Grants 2019C01002 and 2019C03131, in part by Huawei Cooperation Project, and in part by the Zhejiang Provincial NSF of China under Grant LY20F010016.

B. Xiao, Y. Shao, R. Li and H. Zhang are with Zhejiang University, Hangzhou 310027, China. ({xiaobaidi, shaoy, lirongpeng, honggangzhang}@zju.edu.cn); Z. Zhao is with Zhejiang Lab, Hangzhou 310027, China. (zhaozf@zhejianglab.com)

978-1-6654-0926-1/22/$31.00 ©2022 IEEE

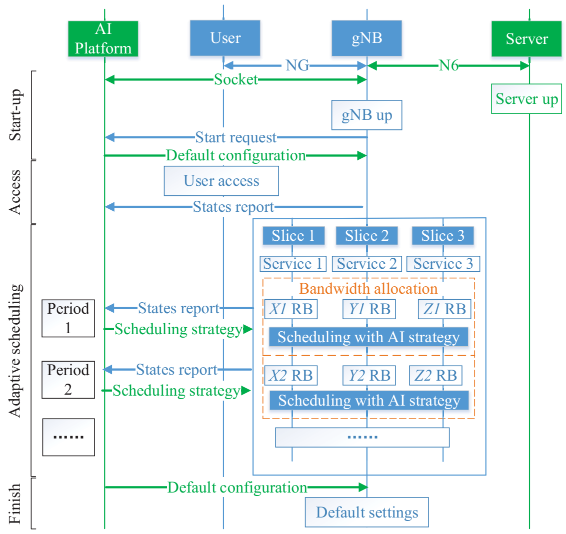


Fig. 2. The sequence diagram of network slicing with AI platform.

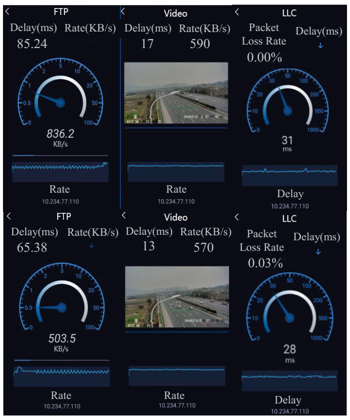
learning is leveraged to process structured data from multiple gNBs as a graph and learn the spatial feature, which will be used to train the DQN for each gNB.

III. The Intelligent Slicing Platform

We have implemented a platform with a modified gNB and six on-shelf MTs for DRL-based resource management in RAN slicing, as depicted in Fig. 2. Meanwhile, the platform implements three types of slices to support services including file transfer protocol (FTP), video streaming (Video) and low-latency communication (LLC), and each slice has at most two subscribed MTs. The whole procedure can be divided into four phases.

- In the "start-up" phase, the server and the gNB will start working with the default configurations in 3GPP.
- Afterwards, subscribers access the RAN, and the gNB starts to report necessary statistical information to the AI platform. In particular, the number of transmission packets in each slice will be the state in MDP. Besides, the modulation and coding scheme (MCS), which is corresponding to SE by Table 5.1.3.1-1 of 3GPP TS 38.214, as well as the downlink rate and delay measured via the radio link control (RLC) layer of gNB, which is essential to calculate SSR, will be received by the platform to calculate the reward.
- During the "adaptive scheduling" phase, the AI platform will use the periodically received statistics to train the algorithms in Section II, and produce a learned policy to guide the bandwidth allocation of each slice accordingly.
- The "finish" phase stops the transmission as normal.

The platform also implements hard slicing and DQN [2] as baselines, where hard slicing implies to divide the total bandwidth equally with a static strategy. Fig. 3(a) provides some snapshots of MTs on the running platform. Moreover, Fig. 3(b) and Fig. 3(c) provide the result for the downlink rate and delay of three slices respectively, indicating all provisioned slices satisfying the SLA with higher downlink rate and lower delay. Meanwhile, Fig. 3(d) gives an illustration of the real-time curve of bandwidth allocation results for each slice. On



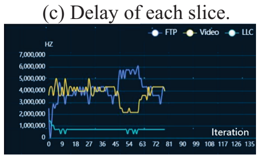
(a) Three kinds of services on six MTs.

	FTP	Video	LLC
Actual Rate(Mbps)	4.78	4.83	0.24
Expected Rate(Mbps)	4.00	4.40	0.20

(b) Rate of each slice.

	FTP	Video	LLC
Actual Delay(ms)	15	18	12
Expected Delay(ms)	25	25	20

(c) Delay of each slice.



(d) Real-time bandwidth allocation.

Fig. 3. Snapshots of the platform.

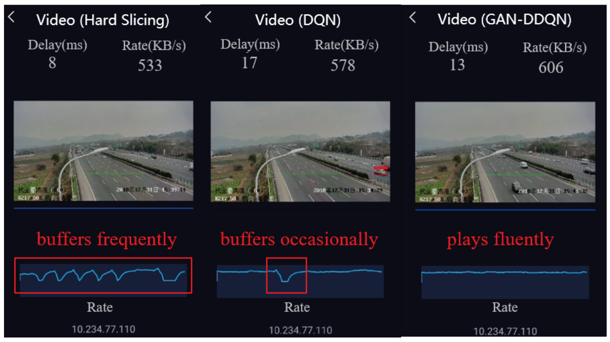


Fig. 4. Performance comparison in Video.

the other hand, we also present the performance comparison between GAN-DDQN and the baselines in Fig. 4 in terms of the performance on Video. It can be observed that the curve of rate is choppy, implying an unsmooth video-play for hard slicing. For DQN, though the curve becomes more stable, there still exists cases with deteriorated performance. Instead, the GAN-DDQN learns a better policy with a much more stable performance curve as the downlink rate of Video increases.

IV. Conclusion

In this paper, we have talked about the algorithms to leverage DRL for intelligent network slicing and built up a platform with a modified gNB and on-shelf commerical MTs to demonstrate the related performance. The platform has demonstrated the effectiveness and superiority of the proposed algorithms (e.g., GAN-DDQN) over classical methods.

References

[1] J. Ordonez-Lucena, *et al.*, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 80-87, May. 2017.

[2] V. Mnih, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, Feb. 2015.

[3] Y. Hua, *et al.*, "GAN-Powered Deep Distributional Reinforcement Learning for Resource Management in Network Slicing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 2, pp. 334-349, Feb. 2020.

[4] R. Li, *et al.*, "The LSTM-Based Advantage Actor-Critic Learning for Resource Management in Network Slicing With User Mobility," *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 2005-2009, Sept. 2020.

[5] Y. Shao, *et al.*, "Graph Attention Network-Based Multi-Agent Reinforcement Learning for Slicing Resource Management in Dense Cellular Network," *IEEE Trans. Veh. Tech.*, vol. 70, no. 10, pp. 10792-10803, Oct. 2021.