Heuristic-assisted Deep Reinforcement Learning for Resource-efficient and QoS-guaranteed 5G RAN Slice Migration in Elastic Metro Aggregation Optical Networks

Jiahua Gu [1,2], Min Zhu [1,2*], Yunwu Wang [1,2], Xiaofeng Cai[1,2], Yuancheng Cai [1,2], Jiao Zhang [1,2]

[1] National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210096, China

*[2] Purple Mountain Laboratories, Nanjing, 211111, China, * Corresponding author e-mail address: minzhu@seu.edu.cn*

Abstract: We propose a heuristic-assisted deep reinforcement learning framework for resource-efficient and QoS-guaranteed 5G RAN slice migration in EONs, which can optimize the spectrum resource consumption, traffic migration, and power consumption, simultaneously.

OCIS codes: (060.4251) Networks, assignment and routing algorithms; (060.4256) Networks, network optimization.

1. Introduction

In 5G radio access networks (RAN), thanks to the recent development of software-defined networking (SDN) and network function virtualization (NFV), the baseband processing functions (e.g., DU and CU) in a RAN slice can be realized through virtual machines (VMs) [1] using generic commercial processing pools (PPs) placed in the metro aggregation networks (MAN) [2]. It allows for flexible scheduling of these slice requests. Especially, for time-varying service loads, the low-load PPs may be deactivated during service valley and the corresponding virtualized DUs (vDUs) and vCUs running on them may be migrated to another higher-load PP, to improve the PP utilization and save energy. However, once the vDU/vCU migrates, the light paths connecting them would require re-configuration, leading to temporal service interruption for the slice users, and thus degrading the users' QoS. Therefore, for resource-efficient and QoS-guaranteed RAN slice migration under time-varying loads, a dynamic vDU/vCU scheduling scheme is required that minimizes both the number of active PPs and the amount of migrated traffic. Recently, some works have addressed the similar deployment problem in WDM-based 5G metro aggregation networks [2-3], but the degraded QoS caused by the vDU/vCU migration was not considered. Some papers (e.g., [4]) have optimized the slice migration via traffic prediction and resource reservation, but the relevant energy efficiency issue was not investigated. Based on our previous work [5], we redesigned the heuristic-assisted (HA)-DRL method for the 5G dynamic vDU/vCU scheduling, where the node placement is more crucial than the light path provision. It is because that the different PP selections for the vDU/vCU have great impact on the service latency and spectrum utilization. Specifically, the different PP selections lead to the different routing and length of the light paths connecting the PPs, and thus cause the different service latency and spectrum occupation state. Hence, it is necessary to jointly optimize the spectrum utilization and power consumption for the slice migration, under the latency constraint.

In this paper, we redesigned the HA-DRL to address the above challenges. To ensure the latency requirement of each request, a polling-based update mechanism for the action space is introduced once a slice request is deployed. In the action apace, candidate PP sets are established for all the pending requests, which should satisfy the latency and capacity constraints. To jointly optimize the spectrum utilization and power consumption, a slice decomposition method is proposed, where the computing resource of the candidate PP and the spectrum usage on light paths connecting the PP are provided to the HA-DRL agent as the state input to make the optimal vDU/vCU deployment decision. Simulation results show that the proposed HA-DRL framework can achieve the lowest cost for all the considered cases. The x-haul traffic migration and spectrum resource consumption can be reduced significantly with a slight cost of the power consumption increment.

2. Problem Formulation

We consider an elastic metro aggregation optical network as the substrate network, which offers agile bandwidth management and high spectrum efficiency [6], as shown in Fig.1 (b). It comprises N_P PP nodes, N_D DC (5GC) nodes, and E links. Each PP node hosts several GPPs, which are used for the virtualization of DU/CUs. Generally, an Active Antenna Unit (AAU) is connected to one adjacent PP as a source node of the slice. Each fiber link comprises serval frequency slots (FS') and each FS has a bandwidth of 6.25GHz. Each RAN slice request is denoted as R_i (s_i, d_i, $BW_F^i, BW_M^i, BW_B^i, C_{DU}^i, C_{CU}^i, D_F^i, D_{E2E}^i$), where s_i and d_i are the source and destination node (i.e., DC), respectively. BW_F^i, BW_M^i, BW_B^i are the required bandwidth of fronthaul (i.e., AAU-vDU), midhaul (i.e., vDU-vCU), and backhaul (i.e., vCU-5GC), respectively. C_{DU}^i and C_{CU}^i are the computing resource required for DU and CU processing, respectively. D_F^i, D_{E2E}^i denote the fronthaul and end-to-end latency requirement of the slice request, respectively.

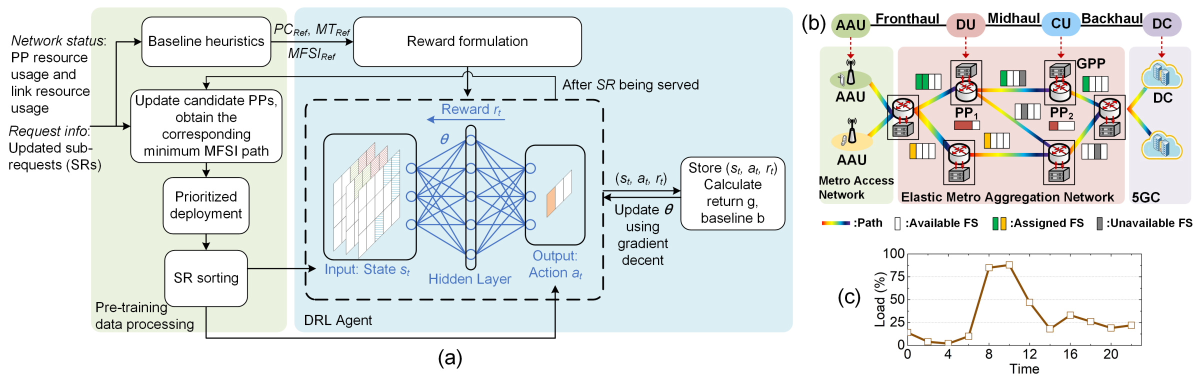

Fig. 1 (a) The training diagram of HA-DRL, (b) System architecture of elastic metro access/aggregation optical networks, (c) traffic profile.

When deploying the slice requests, the routing constraint, PP/GPP and x-haul capacity constraint, service latency constraint, and DU-CU deployment constraint should be all respected.

For each request, its required computing resource and x-haul bandwidth vary with time in Fig.1 (c) [2], where 12 time periods are considered, each lasting for 2 hours. In each time period t, the system power consumption (PC) can be calculated by $PC_t = P_{full}\cdot\sum_n F_n + P_{half}\cdot\sum_n H_n + P_0\cdot\sum_{n,p} G_{n,p}$, where F_n=1 denotes PP n working in full-power mode with P_{full} (2200W) for hosting vDU/vCUs, and H_n=1 means PP n working in half-power mode with P_{half} (1100W) when only data transmission devices are on. $G_{n,p}$=1 when GPP p in PP n is activated with the basic PC P_0 of a GPP (100W). The migration traffic (MT) can be calculated by $MT_t=\sum_i BW_F^{i,t}\times Z_F^{i,t}+BW_M^{i,t}\times Z_M^{i,t}+BW_B^{i,t}\times Z_B^{i,t}$, where $Z_F^{i,t}$, $Z_M^{i,t}$, and $Z_B^{i,t}$ are binary variables denoting if fronthaul, midhaul, and backhaul of R_i are migrated. Maximum frequency slot index (MFSI) is a performance matrix used to represent the spectrum resource consumption in the network [7]. It can be calculated by $MFSI_t=\max_{e\in E} f_e^{\max}$, where $f_e^{\max}$ is the maximum FS index in link e. To jointly optimize the PC, MT and MFSI, the cost function can be formulated as $C=\sum_t\left(PC_t/PC_{ref}^t+MT_t/MT_{ref}^t+MFSI_t/MFSI_{ref}^t\right)$, where PC_{ref}^t, MT_{ref}^t, and $MFSI_{ref}^t$ are the reference results obtained from the baseline heuristic algorithms, which can help the DRL agent quickly learn a better RAN slice migration policy based on the reference algorithms, hence realizing the "HA-DRL".

3. Methodology

For easy configuration, we decompose a slice request R_i into two sub-requests (SRs): 1) R_{DU}^i [s_i (i.e., source PP) , d_i (i.e., CU or destination DC), BW_F^i, BW_M^i, C_{DU}^i, D_F^i, D_{E2E}^i] and 2) R_{CU}^i [s_i (i.e., DU or source PP), d_i (i.e., destination DC), BW_M^i, BW_B^i, C_{CU}^i, D_{E2E}^i]. Fig. 1 (a) shows the training diagram of the proposed HA-DRL method to serve these SRs, where the **state**, **action**, and **reward** are defined as follows: (1) **State**: The state space comprises the resource usages of the candidate PPs and FS usages on the corresponding minimum MFSI paths. Additionally, the request information, including the required computing resources, number of required FS' and history PP indicator are provided. (2) **Action**: The action space consists of the history PP node (if available) and M or M-1 most loaded PPs in K shortest paths connecting (s_i, d_i). Once a slice request is deployed, the action space is updated timely. (3) **Reward**: The reward $r_i^t=\Delta PC/PC_{ref}^t+\Delta MT/MT_{ref}^t+\Delta MFSI/MFSI_{ref}^t$ is calculated based on instant performance metrics, where ΔPC, ΔMT and $\Delta MFSI$ are the PC, MT and MFSI increment caused by SR_i deployment. Algorithm 1 shows the training process in each iteration in detail. The policy gradient (PG) method is adopted for DRL training [8]. A DNN is used as the policy network, where the input is the state s_i and output is the probability distribution of actions $\pi_\theta(a_i\,|\,s_i)$. The θ is the policy network's parameters. We use the REINFORCE with baseline algorithm [8] to iteratively update θ. The update of gradient uses *rmsprop* algorithm.

Algorithm 1 HA-DRL Training Algorithm

Step 1: In each time period t, apply baseline algorithms to acquire the reference data PC_{ref}^t, MT_{ref}^t and $MFSI_{ref}^t$.

Step 2: Apply the polling-based update for each SR_i (R_{DU}^i or R_{CU}^i) to get the available candidate PP set N_i. Check if a special case occurs when the $|N_i|$=1 due to capacity or delay constraints, the corresponding SR can be directly deployed on the designated PP node and the lightpath with the smallest MFSI.

Step 3: Denote the available resource in each PP n as AR(n). Then, calculate the available resources of the N_i for SR_i as $C_{avail}^i=\sum_{n\in N_i} AR(n)$. Sort the SRs by C_{avail}^i in ascending order. For the SR_i with the minimum C_{avail}^i, construct the state space and action space, and acquire the node and corresponding lightpath deployment decision from the DRL agent. Deploy SR_i and obtain rewards. If all the SRs are served, go to **step 4**, otherwise, go to **Step 2** to serve the rest of SRs.

Step 4: Perform routing adjustment for each slice requests: Replace the x-haul's lightpath with the shortest path only if MFSI remains unchanged. Go to **Step 1** for handing SRs in the next time period.

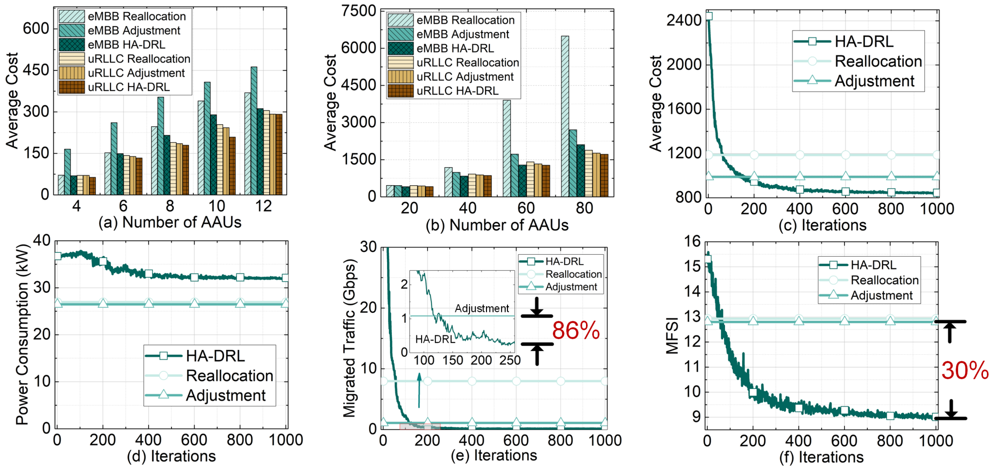

Table I
Latency requirements for different services [2]

service	FH Latency	E2E Latency
eMBB	100μs	11.1ms
uRLLC	50μs	550μs

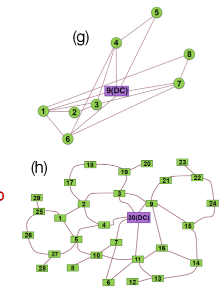

Fig. 2 Simulation results: average cost in (a) 9-node network, (b) 30-node network; Training process for eMBB40: (c) Average cost, (d) PC, (e) MT, (f) MFSI; Network topologies: (g) 9-node network, (h) 30-node network

Two baseline heuristics are designed to assist the DRL training. The **SR Reallocation** presented in Algorithm 2 aims at optimizing the PC and MFSI, because in each time period, all requests need to be re-configured. Note that the **SR Adjustment** can simultaneously optimize the PC, MT, and MFSI. It is due to the fact that in the beginning of each time period, it migrates a part of SRs from the overloaded PP to the un-overload PP with Algorithm 2. Then, it would try to deactivate the least-loaded PP n_{min} to save power by reallocating all the SRs the PP serves. The reallocation process would be terminated once any SR cannot be reallocated. When all SRs are well provisioned, the routing adjustment would be applied too, just as in Step 4, Algorithm 1.

Algorithm 2 Baseline heuristic: SR Reallocation Policy

Step 1: In each time period t, sort SRs in ascending order in turn by s_i-d_i distance and computing resource requirement. For each SR_i, get K shortest paths connecting s_i and d_i. From K paths, get available PP set N_i, where each PP in N_i satisfies all the delay, and capacity constraints. Find the most loaded PP n_{max} in N_i, and deploy the required vDU/vCU in n_{max}. Connect (s_i, n_{max}), (n_{max}, d_i) with the path that leads to the smallest MFSI. If all the SRs are served, go to Step 2, otherwise, repeat **Step1**.
Step 2: Perform similar routing adjustment as **Algorithm 1**. Then, go to **Step 1** for handing SRs in the next time period.

4. Performance Evaluation

The proposed **HA-DRL** is compared against the two baseline heuristics (i.e., **SR reallocation** and **SR adjustment**). We consider a small-scale 9-node network and a large-scale 30-node network for performance evaluation. For both networks, each optical link has a length between 10 and 30 km. Each AAU is set as 4 antennas, 100MHz, two MIMO layers, and MCS 23. In 9-node network, each PP comprises 1 GPP and each GPP is set as 4000 GOPS. Each optical link has a capacity of 20 FS'. In 30-node network, each PP comprises 2 GPPs and each GPP is set as 10000 GOPS. Each optical link has a capacity of 100 FS'. Two types of 5G services – eMBB and uRLLC are considered. The latency requirement of these services is provided in table I.

In Figs. 2(a) and (b), it is observed that, for both eMBB and uRLLC services, the proposed HA-DRL achieves the lowest cost compared with the two baseline heuristics in both networks. Specially, for eMBB40 service, the PC, MT, and MFSI training curves are depicted in Figs. 2(c)-(f). Unlike the heuristic algorithms that only apply fixed policies, through iterative training, HA-DRL is able to learn that the best way to reduce cost is by trading PC for MT and MFSI. Compared with the SR adjustment, up to 86% MT and 30% MFSI reduction are achieved with only slightly higher PC (i.e., 12 % higher PC).

5. Conclusion

An HA-DRL framework is proposed for the resource-efficient and QoS-guaranteed 5G RAN slice migration, which jointly optimizes the PC, MT, and MFSI under the service latency constraint in EONs. Simulation results validate the effectiveness of the proposed HA-DRL with a polling-based update scheme and a slice decomposition method.

References

[1] C.-Y. Chang, et al., IEEE CM, **56**.8: 70-77, 2018.
[2] Y. Xiao, et al., IEEE/OPTICA JLT, **39**.17: 5347-5361, 2021.
[3] L. M. Moreira Zorello, et al.,IEEE TGCN,**6**.3:1683-1696, 2022.
[4] H. Yu, et al., IEEE/ OPTICA JOCN, **12**.12: 403-413, 2020.
[5] M. Zhu, et al., IEEE/ OPTICA JLT, **40**.3, 575-587, 2022.
[6] M. Jinno, et al., ECOC, 1-2, 2008.
[7] W. Fang, et al., IEEE CL, **20**.8: 1539-1542, 2016.
[8] H. Mao, et al., in Proc of HotNets, 50-56, 2016.