Service-oriented DU-CU Placement Using Reinforcement Learning in 5G/B5G Converged Wireless-Optical Networks

Yuming Xiao, Jiawei Zhang, Zhengguang Gao, Yuefeng Ji
State Key Lab of Information Photonics and Optical Communications, BUPT, Beijing, China
yumingxiao@bupt.edu.cn, zjw@bupt.edu.cn, gaozg@bupt.edu.cn, jyf@bupt.edu.cn

Abstract: We propose a reinforcement learning based DU-CU placement scheme to accommodate diversified services in 5G/B5G networks. It outperforms ILP model and widely used heuristics in terms of the service-scale and resource-saving respectively.
OCIS codes: (060.4256) Networks, network optimization; (060.4265) Networks, wavelength routing

1. Introduction

5G and beyond are expected to support service scenarios of versatile requirements, such as higher data rates (eMBB), ultra-low latency (URLLC), and massive connections (mMTC). These heterogeneous services will raise new challenges for radio access network (RAN) in terms of capacity, latency, and networking flexibility that accelerates its evolution towards next-generation RAN (NG-RAN). The baseband unit (BBU) has been re-defined as two new entities named distributed unit (DU) and central unit (CU) to make NG-RAN serve as a three-tier architecture, i.e., remote radio unit (RRU)-DU-CU. The large-bandwidth and latency-sensitive BBU functions below PDCP are located at DU, while functions above PDCP are provided in CU. Recently, the DU-CU placement pattern has attracted considerable attention from both academic and industry fields, which will become diversified depending on service requirements. Some telecom operators have submitted proposals on DU-CU placement in their white papers [1]. For eMBB, DUs are placed in processing pools (PPs) near users to alleviate bandwidth burden in fronthaul, while CUs are centralized in remote PPs for processing resource sharing. For URLLC, DUs and CUs are co-located in the proximity of users for latency satisfaction. However, this scheme is designed without a global view of the network state and overall service requests, which may result in resource over-consumption. For instance, if DUs are all distributed close to users, then more edge PPs are activated with extra cost and power consumption. If DU and CU are highly centralized, then more bandwidth and latency are introduced for long-reach and multi-hop transmission. Therefore, how to place DU-CU, both considering versatile service requirements and resource efficiency, becomes a crucial problem.

Previous studies solved DU-CU or BBU placement mainly based on ILP models and heuristic algorithms. Essentially, ILP is a time-consuming solution and not suitable for large-scale service paradigm. The heuristics are extended from ILP and artificially designed for a specific scenario and optimization objective. Once the scenario has changed, heuristics cannot ensure a satisfactory strategy. Thus, we should introduce a more intelligent solution which not only adapts to various service paradigms but also achieves the self-optimization automatically. Reinforcement learning (RL), which handles complicated decision-making problems through efficient exploration, has been introduced to resource optimization in 5G networks [2, 3]. However, these works focused on general requests but ignore diversified services in 5G/B5G. Moreover, the service-oriented DU-CU placement based on RL method hasn't yet been discussed.

In this paper, we propose an RL-enabled DU-CU placement algorithm (RL-PS) adapting to three service scenarios to minimize the PP and bandwidth consumption. We also compare RL-PS with ILP, widely used greedy-based (GBA), and First-Fit algorithms under 3~30 service requests. Results show that RL-PS can generate a satisfactory solution for large-scale service paradigm while economizing the PP and bandwidth than GBA and First-Fit heuristics.

2. Network Architecture and placement scheme

NG-RAN contains three x-haul segments, where fronthaul connects RRU and DU, midhaul connects DU and CU, and backhaul connects CU and metro data center (DC). Optical networks are attractive in providing cost-efficient x-haul solutions. In this paper, we consider an optical transport network (OTN) to interconnect RRUs, PPs, and metro DC in Fig. 1 (a), where fronthaul, midhaul, and backhaul share the common network infrastructures. For each link, several fibers are provided, and each of them consists of multiple wavelengths. For each node, an electronic switch (E-switch) and a reconfigurable add/drop multiplexer (ROADM) are equipped for switching on optical and electronic domains. PP is co-located with E-switch and ROADM, comprising several general-purpose processors (GPPs). DU-CU can be virtually implemented in virtual machines (VM) within GPPs to facilitate the efficient sharing of processing resources. All traffic flows are ultimately aggregated into DC for content processing.

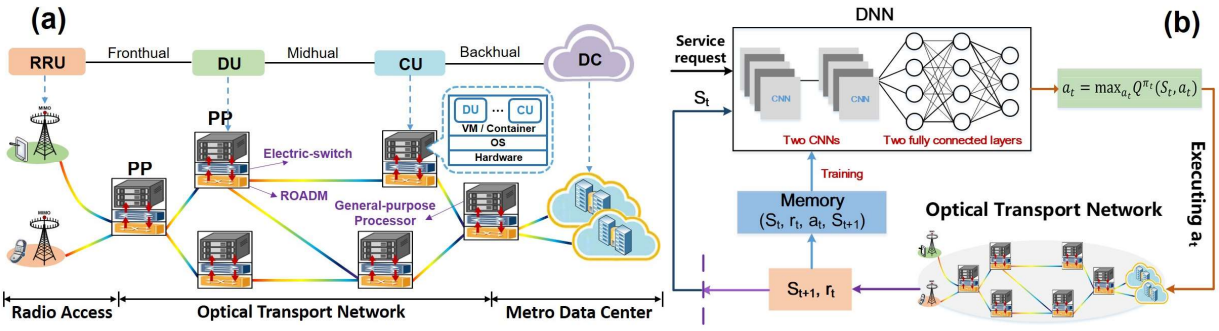

Fig. 1. (a) Optical transport network interconnecting RRUs, PPs, and DC; (b) The structure of RL algorithm model

During the placement, we should select a source RRU and one/two PP(s) to accommodate one service request. DU and CU should be placed once following the processing order, i.e., DU→CU. The DU-CU can be co-located like BBU or be separated into two PPs depending on the residual capacity of PPs. A lightpath should be established to connect selected RRU, selected PPs, and DC: a) we should route this service from selected RRU to DC passing through selected PP(s); b) we should guarantee that wavelength and bandwidth on links of this lightpath are sufficient to carry this service; c) we should guarantee that respective latency for front-/mid-/back-haul segments don't exceed the limitations in Table 1. The latency consists of two parts: 1) transmission latency on fibers for 5μs/km, and 2) OEO conversion, electronic switching, and Fx/F1 interface encapsulation (OSE) for 20μs. The OSE latency is introduced for services when DU-CU processing or E-switching (e.g., grooming) is executed in PP node. We have proposed an ILP model considering the RRU and PP selection, routing, wavelength and bandwidth allocation, and latency control [4]. The detailed formulations and parameters for DU-CU processing and transmission are provided in Ref. [4]. We have also introduced two widely used heuristic algorithms for comparison, i.e., GBA and First-Fit. Both heuristics are based on the K-shortest paths strategy (KSP). First-Fit selects the first appropriate path to accommodate DU-CU. GBA tries to place DU-CU on all candidate paths, and then sort placement results in decreasing order of $O = \alpha \cdot V1 + \beta \cdot V2$ (*V1* and *V2* denote the used PP amount and bandwidth). The path with minimal *O* is selected as the solution.

3. RL-based algorithm

RL algorithm is an episode by episode iterative learning procedure to search for the optimal action under a certain state [2]. As shown in Fig. 1 (b), we have designed an RL-based DU-CU placement algorithm, where the states, actions, and reward function are defined as follows. 1) **State**: The state s_t is the combination of the service type (i.e., eMBB, URLLC, mMTC) and residual capacity of links and PPs at time-step *t*. 2) **Action**: The action a_t describes the location of DU-CU and the selected lightpath. For example, in action [2,3,1,2,3,4], first two numbers denote that DU and CU are respectively placed at *PP 2* and *PP 3*, while latter four numbers represent its lightpath to pass through *RRU 1*, *PP 2,3,* and *DC 4*. 3) **Reward**: The reward r_t=-(α×*P* +β×*B*)/50+5 gives the evaluation on choosing action a_t for state s_t, where α and β are consistent with the weights in objective function of ILP. The variable *P* denotes the number of newly activated PPs for service *sr*, while *B* represents the bandwidth provision for *sr*. The reward function gives small-integer feedback when action a_t obeys the capacity (PP and link) and latency constraints. The deep neural network (DNN) is established to select the action $a_t \in A$ (action space) adaptive to the current OTN resource state s_t, which comprises two convolutional neural networks with five 2×2 convolutional kernels, and two fully connected layers. There are 13 (RRU + PP node) × 13 × 5 (kernel) and 400 hidden neurons in two connected layers. The output of DNN is the estimated value of 198 actions. The execution steps of RL-PS are presented as follows.

First, we should generate the training data for DNN through the interaction between actions and OTN. At time-step *t* in one episode, DNN observes the network state together with the service request and then outputs the state-action value $Q^{\pi_t}(s_t, a_t)$ of all actions under the current strategy π_t. We consider a ε-greedy policy as the action selection method. The ε (exploration rate, 0<ε<1) denotes the possibility of selecting the largest-Q-value action, while 1-ε represents the possibility for a random action. The ε increases with the learning procedure until reaching a maximum value ε_{max}. After executing the selected action, OTN then transits into a new state s_{t+1} (with changed residual capacity for PPs and links) and generates a reward r_t. The data pair (s_t, a_t, r_t, s_{t+1}) is stored into the memory.

Second, we should train DNN with the generated data so that selects the best placement action for service requests. The parameters of DNN are updated each 3000 time-steps with a random mini-batch of data chosen from the memory. The principle to train DNN is the Bellman equation of optimal Q-function as $Q^{\pi_t}(s_t, a_t) = r(s_t, a_t) + \gamma \cdot$

Table 2. Simulation parameters for three defined services in upstream [5]

Category	Wireless RBs	Fronthaul Latency	Mid+Backhual Latency
eMBB	450 (~300Mbps)	100μs (20km)	1000μs
URLLC	50 (~30Mbps)	250μs (50km)	1500μs
mMTC	150 (~100Mbps)	50μs (10km)	450μs

$MaxQ^{\pi_t}(s_{t+1}, a_{t+1})$, where γ is a discount factor that reflects the significance of future rewards compared to the current rewards. The Q-function $Q^{\pi_t}(s_t, a_t, \theta_t)$ with parameters θ_t is introduced to fit Q-value. The Bellman error is defined as $\xi = Q^{\pi_t}(s_t, a_t, \theta_t) - r(s_t, a_t) - \gamma \cdot MaxQ^{\pi_t}(s_{t+1}, a_{t+1}, \theta_t)$, which should be minimized to update DNN parameters through gradient descent method. DNN training continues until the Bellman error converges.

4. Simulation Results and Analysis

We consider a network topology of 8PPs (*Node 1~8*), 4 RRU nodes (*Node 11~13,* each node contains 7 RRUs), 1 DC, and 16 optical links in Fig. 3(a). The distance between RRU and its local PP ranges from 0 to 3km, while optical links range from 10 to 60km. The processing capacity of 6000 Giga operations per second (GOPS) is provided in each PP. Ten wavelengths (10 Gb/s per wavelength) are available on each link. Each RRU contains two antennas and a 100MHz spectrum with 16QAM for upstream. The required resource blocks (RB) and latency constraints for each x-haul segment are detailed in Table. 1. We have compared ILP, RL-PS, and two heuristics under 3 (3 categories ×1) ~30 (3 categories ×10) service requests. Ten candidate paths are pre-calculated for RL-PS, GBA, and First-Fit strategies.

The consumed bandwidth in four strategies is shown in Fig. 3(b), where ILP performs as the benchmark. The lightpath selection and DU location (because fronthaul dominates the bandwidth consumption) significantly influence the placement performance. RL-PS achieves a near-optimal solution followed by two heuristics because of its global optimization, which is benefited from the discount factor γ. GBA is a local-optimal strategy that can only achieve the optimal solution in the small-scale requests. First-Fit selects the first appropriate path and PPs to hold the service that may result in the long-reach and multi-hop fronthaul transmission. As shown in Fig. 3(c), the occupied wavelengths in each strategy are relevant to the bandwidth consumption that RL also performs better than two heuristics. As the service-scale grows, ILP solving becomes prohibitively time-consuming and fails to work at 27 and 30 simulation nodes. We also present the number of activated PPs in Fig. 3(d), where RL-PS achieves optimal performance as ILP, followed by GBA and First-Fit. Simulation results show that RL-PS is the potential to surpass traditional heuristics on resource-saving and ILP model on service-scale.

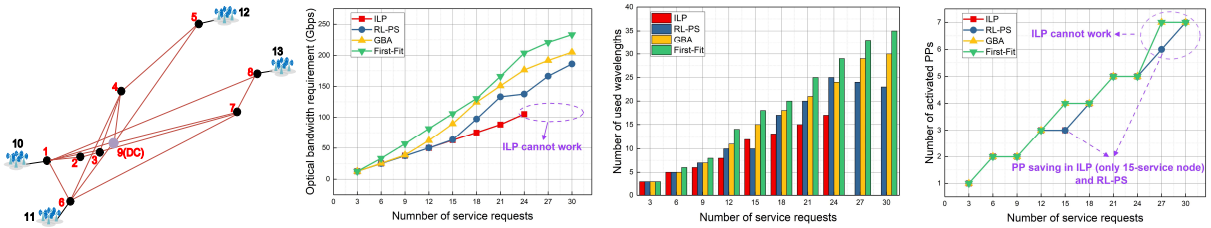

Fig. 3 (a) Simulation topology; (b) Bandwidth vs. (c) Wavelengths vs. (d) Number of activated PPs vs. # of requests

5. Conclusions

We proposed an RL-PS algorithm for DU-CU placement in an optical transport network to achieve both advantages on large-scale service paradigm and resource optimization. Simulation results validated our inspiration that RL-PS accommodated more requests where ILP had collapsed while outperforming GBA and First-Fit heuristics in terms of PP, bandwidth, and wavelength economization.

6. Acknowledgement

This work is supported by the National KeyR&D Program of China (No. 2018YFB1800802), the National Nature Science Foundation of China Projects (61871051, 61971055), the Beijing Natural Science Foundation (No. 4192039), the fund of State Key Laboratory of Information Photonics and Optical Communications, China, IPOC2019ZT05, and the BUPT Excellent Ph.D. Students Foundation (No. CX2019222).

7. References

[1] Netmanias Report. “Survey on 5G RAN Practical Deployment Scenario.” 2018. URL: https://www.netmanias.com/en/?m=board&id=reports.
[2] Mikaeil, A. M., et al. “Joint Allocation of Radio and Fronthaul Resources in Multi-Wavelength-Enabled C-RAN Based on Reinforcement Learning.” Journal of Lightwave Technology, 2019.
[3] Gao, Z., et al. "Deep Reinforcement Learning for BBU Placement and Routing in C-RAN." OFC, 2019.
[4] Xiao, Y., et al. “Resource-Efficient Slicing for 5G/B5G Converged Optical-Wireless Access Networks.” ACP, 2019.
[5] IEEE 1914.1. “Xhaul dimensioning challenges.” 2018. URL:https://sagroups.ieee.org/1914/p1914-1/ieee-p1914-1-tf-teleconference-materials.