26th Optoelectronics and Communications Conference (OECC)

Reinforcement Learning Based Joint Allocation Scheme in a TWDM-PON Based mMIMO Fronthaul Network

Yuansen Cheng and Chun-Kit Chan
Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong, China
cy019@ie.cuhk.edu.hk; ckchan@ie.cuhk.edu.hk

Abstract: We propose a reinforcement-learning based joint allocation algorithm for TWDM-PON based mMIMO fronthaul network. By adopting the combined Pointer Network and Actor Critic algorithm, we realize superior performance in resource block and wavelength utilization efficiencies.

1. Introduction

In next generation mobile networks, centralized or cloud radio access networks (C-RANs) have emerged as the promising approach to cope with the ever-increasing traffic of mass mobile devices. The mobile base station (BS) incorporates three main functional modules, including the central unit (CU), the distributed unit (DU), and the radio unit (RU) [1]. The link between a DU and an RU is called fronthaul, with desirable requirements of high bandwidth efficiency, ultra-low latency and low system costs [2]. Time and wavelength division multiplexed (TWDM) passive optical network (PON) offers an economical solution to realize the fronthaul network infrastructure in C-RAN for its relatively low capital expenditure (CapEx). With the advent of 5G mobile systems, in which massive multiple input multiple output (mMIMO) is deployed to enhance the system capacity, the TWDM-PON-based fronthaul infrastructure has to be further optimized in terms of bandwidth and resource allocation.

In this paper, we propose a general architecture for TWDM-PON based mMIMO fronthaul network and a reinforcement learning (RL) based allocation scheme to jointly maximize the bandwidth efficiency and antenna's RBs utilization (ARU) ratio. We further adopt the combined the Pointer Network and the Actor-Critic (PNAC) algorithm [5] to perform wavelength allocation and the results are compared with Deep Q Network (DQN), First-Fit and Best-Fit algorithms. Compared with the allocation scheme reported in [3], our RL-based allocation scheme attains higher bandwidth efficiency and a higher ARU ratio, simultaneously. The results show that the PNAC algorithm can greatly reduce the number of required wavelengths, which is much closer to the optimal value, while the computation complexity can still be kept low. The training data can also be applied to networks of larger scale in future network upgrade.

2. Framework and Problem Formulation

In this paper, a general TWDM-PON architecture (GEN-ARCH) is adopted as the fronthaul infrastructure for mMIMO. Different from the fixed architecture (FIX-ARCH) adopted in [3], which employed a wavelength de-multiplexer and multiple power splitters, a conventional TWDM-PON with a single splitter is employed in our system to connect the DU to multiple RUs. Each ONU is connected to a RU, forming an ONU-RU group and then connected to multiple antennas from a large antenna array. Each ONU is equipped with tunable transmitters and receivers, which can be tuned to any wavelength for both downstream and upstream directions. Hence, GEN-ARCH enables a more flexible wavelength assignment compared with FIX-ARCH. For the functional split RAN architecture, higher layer functions such as the packet data convergence protocol (PDCP) and the radio link control (RLC) are processed at CU, while the media access control (MAC) function and the forward correction code (FEC) encoding are processed at DU. Modulation, resource mapping, and IFFT with cyclic-prefix-process are incorporated into RU to reduce the required bandwidth of the fronthaul. The bit rate of the fronthaul can be calculated by [4],

$$R_b = N_{mo} \times N_{nsym} \times N_{nsc} \times N_{rb} \times N_{mimo} \quad (1)$$

where N_{mo}, N_{nsym}, N_{nsc}, N_{RB}, N_{mimo} are the modulation order, the number of symbols within transmission time interval (TTI), the number of subcarriers per RB, the number of RBs per user equipment (UE), and the number of MIMO streams. Different allocation schemes may induce different bandwidth demands in the fronthaul, antenna RBs utilization ratio, and various wavelength usage. Our work aims to minimize the required fronthaul bandwidth, wavelength usage as well as the number of active ONU-RU groups. The allocation problem is subjected to some constraints, including ONU-RU group allocation constraints, wavelength allocation constraints, antenna allocation constraints, RBs allocation constraints and antenna & wavelength capacity constraints.

3. RL-based Allocation Scheme

Resource allocation problems are commonly optimized by the integer linear programming (ILP) technique, however, they are not practical due to its prohibitively time-consuming computation [3]. In this paper, an RL based heuristic algorithm is proposed to optimize the required bandwidth for the fronthaul, radio resource utilization, and the required number of wavelengths. For the antenna and RBs allocations, we offer an improved version of the method from [3]. A waiting list was adopted to dynamically change the allocation order for the RBs. The wavelength allocation is modelled as a Markov Decision Process. The combined PNAC algorithm was used for wavelength assignment and the results are compared with other algorithms, including DQN, First-Fit and Best-Fit.

3.1 DQN based wavelength assignment

For the DQN based algorithm, all ONU-RU groups' bandwidth requests are put into a queue first. The DQN model iteratively chooses a wavelength for the first ONU-RU in the waiting queue. Once the first ONU-RU is allocated, it will be popped, thus the previous second ONU-RU becomes the head of the queue. The iterative assignment process continues until all the ONU-RU groups in the waiting queue have been popped. The state space and the action space are modeled as follows, while the reward function is elaborately designed.

State space: The bandwidth demand for all ONU-RU groups and the remaining available candidate wavelength capacity, $s = (r_1, r_2, \ldots, r_m, c_1, c_2, \ldots, c_n.)$

Action space: The candidate wavelengths, $a \in A = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

3.2 PNAC based wavelength assignment

The PNAC model directly attains a sequential index permutation of all ONU-RU groups. Starting from an empty wavelength, we iteratively select the ONU-RU groups according to the permutation indices. Once the current wavelength cannot provide sufficient bandwidth for the current ONU-RU group, a new empty wavelength will be used. The state space and the action space are modeled as follows,

State space: The bandwidth demand for all ONU-RUs, $s = (r_1, r_2, \ldots, r_m)$

Action space: The permutation of the input state indices π.

Given the requests and the sequential permutation indices, $D(\pi|s)$ is defined as the total number of wavelengths after all indices have been exhausted. We aim to find a stochastic policy $p(\pi|s)$ to minimize $D(\pi|s)$. By factorizing the stochastic policy p, the pointer network can use individual SoftMax modules to represent the probability of the *k-th* action taken by the policy. Our pointer network comprises two recurrent neural networks (RNN) modules, namely the encoder and the decoder. It allows the model to point to the specific position of the input sequence. Therefore, the same model can be applied to all cases with a different number of input dimensions. We further optimize the parameters of the pointer network using the Actor-Critic algorithm. The training objective is the expected number of the wavelength usage, which is defined as Eqn. (2). We then resort to policy gradient methods and the stochastic gradient descent to optimize the parameters while the corresponding gradient is shown as Eqn. (3). For sampling a single allocation task, i.e., $\pi \sim p_\theta(\cdot|s)$, the gradient is approximated with Monte Carlo sampling.

$$J(\theta \mid s) = \mathrm{E}_{\pi \sim p_\theta(.|s)} D(\pi \mid s) \tag{2}$$

$$\nabla_\theta J(. \mid s) = \mathrm{E}_{\pi \sim p_\theta(.|s)}[D(\pi \mid s) - b(s)\nabla_\theta \log p_\theta(\pi \mid s)] \tag{3}$$

4. Results

4.1 Simulation setup

For the simulation setup, we adopt a similar setup used in [3], where an RB has a 180 kHz frequency range, and each RB has 12 subcarriers and 7 OFDM symbols within a TTI. The modulation order is 7 and assumes a single-cell scenario with one antenna array. Both large-scale and small-scale networks are considered in this work. For the small-scale network scenario, there are totally 64 antennas equally distributed to 8 sub-antenna arrays. The system bandwidth of the wireless part is 3.2 MHz. In contrast, the total number of antennas in the large-scale network grows up to 512 and are equally distributed to 32 sub-antenna arrays. The corresponding wireless bandwidth is 100 MHz. We divide the space into 36 regions for beamforming, and all user equipment is uniformly distributed among these 36 regions. The simulation results of our proposed GEN-ARCH and RL based allocation algorithm are compared with the FIX-ARCH and the three allocation algorithms (namely, IBBA, SBBA, TSBBA based allocation algorithm) presented in [3].

4.2 Simulation results

Fig. 1 compares the results obtained from various algorithms and architectures for the large-scale network. As shown in Fig. 1(a), both the proposed GEN-ARCH and the RL-based algorithm can improve the bandwidth efficiency. Fig. 1(b) shows that the RL based algorithm can also achieve a high ARU ratio, which is calculated by dividing the employed RBs by the total number of RBs of the active ONU-RU groups. Our proposed algorithm can simultaneously attain a high fronthaul bandwidth efficiency and a high ARU ratio. Fig. 1(c) illustrates the execution times among different algorithms. The proposed RL-based algorithm and the IBBA-based allocation require the least average execution time.

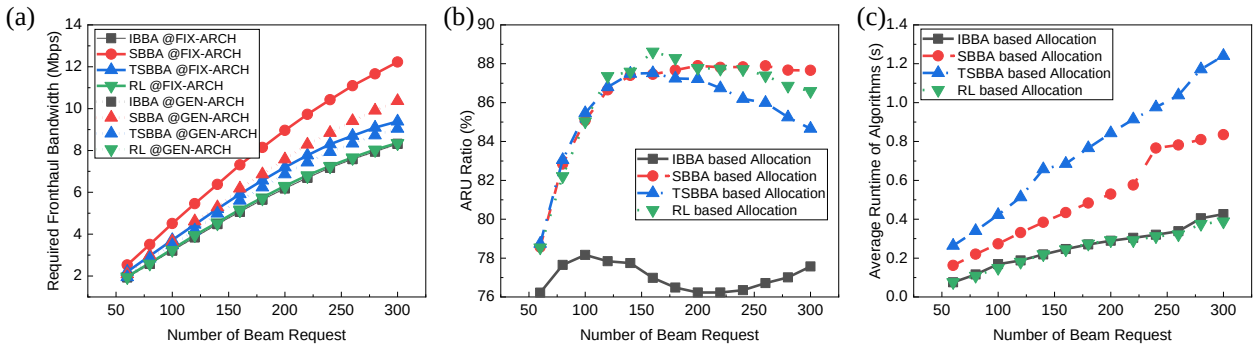

Fig. 1. Simulation results for large-scale networks: (a) required fronthaul bandwidth, (b) ARU ratio, (c) average execution time

Fig. 2 shows the results for wavelength allocation. We compare the results of the PNAC and the DQN with the counterpart from the ILP, First-Fit and Best-Fit algorithm. Fig. 2(a) shows the wavelength demand versus the number of ONUs, where the vertical axis represents the difference between the results of the algorithms from the optimal value (from ILP). The fluctuation in Fig. 2(a) is attributed to the discrete characteristics of the chosen wavelengths. The PNAC based allocation exhibits a superior performance to the DQN, the First-Fit and the Best-Fit algorithms. Due to the point network's characteristics, a single model can deal with a different number of active ONU-RU groups. Moreover, as depicted from Table 1, although the model is trained with 1 to 16 ONU-RU groups, the same set of training data can also be applied to the cases when the active number of ONUs is higher than 16. As shown in Fig. 2(b), the average runtimes of all algorithms increase linearly or quadratically with the number of ONUs. On the contrary, although the ILP can attain the optimal results, the runtime increases exponentially.

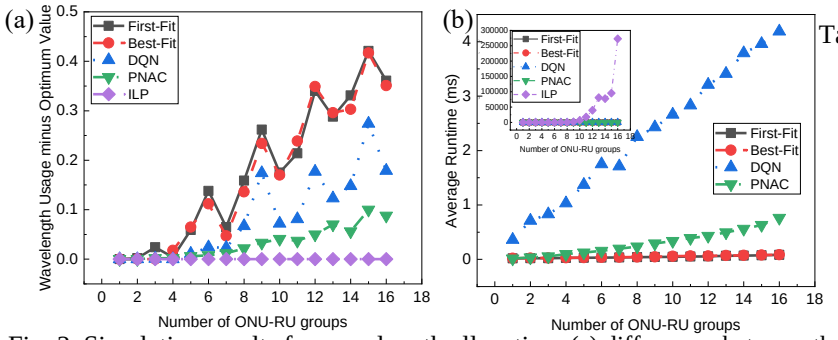

Table 1. Wavelength allocation results

Active ONU-RU Number	PNAC trained with 1-16 groups	Best-Fit
20	7.15	7.47
30	10.46	10.90
50	17.39	17.46

Fig. 2. Simulation results for wavelength allocation: (a) differences between the results of the algorithms from the optimal value (ILP), (b) average runtime

5. Summary

We propose a RL-based joint allocation algorithm to allocate the wavelength, antenna, and RBs in a TWDM-PON based mMIMO fronthaul network. The results demonstrate that our approach can realize a high fronthaul bandwidth efficiency and resource block utilization ratio simultaneously. The wavelength usage can also be further optimized.

6. References

[1] P. Chanclou *et al.*, *Journal of Optical Communications and Networking*, vol. 10, A1-A7, 2018.
[2] N. Gomes *et al.*, *IEEE Vehicular Technology Magazine*, vol. 13, pp. 74-84, 2018.
[3] J. Zhang *et al.*, *J. Lightwave Technol.*, vol. 37, pp. 1396-1407, 2019.
[4] K. Miyamoto *et al.*, *Opt. Express*, vol. 24, pp. 1261–1268, Jan. 2016.
[5] I. Bello *et al.*, in *ICLR Workshop*, 2017.