Q-Learning Based Joint Allocation of Fronthaul and Radio Resources in Multiwavelength-Enabled C-RAN

Ahmed Mohammed Mikaeil, Weisheng Hu

Q-Learning Based Joint Allocation of Fronthaul and Radio Resources in Multiwavelength-Enabled C-RAN

Ahmed Mohammed Mikaeil [0000-0002-9621-8316] and Weisheng Hu [0000-0002-6168-2688]

State Key Laboratory of Advanced Optical Communication
Systems and Networks, Shanghai Jiao Tong University
No. 800, Road Dongchuan, Shanghai 200240, China
ahmed_mikaeil@yahoo.co.uk, wshu@sjtu.edu.cn

Abstract. Multi-wavelengths passive optical networks (PONs) such as wavelength division multiplexing (WDM) and time wavelength division multiplexing (TWDM) PONs are outstanding solutions for providing a sufficient bandwidth for mobile front-haul to support C-RAN architecture in 5G mobile network. In this paper a joint allocation framework for multi-wavelength PONs mobile front-haul and C-RAN air interface uplink resources is proposed. From the principle that uplink resource allocation in mobile networks (e.g. 4G and 5G) is an NP-hard optimization problem, this paper contributes with a novel method for uplink scheduling based on a reinforcement learning (RL) algorithm known as Q-Learning. The performance of the algorithm is evaluated with numerical simulations and compared with some other relevant work from the literature such as genetic algorithm (GA) and tabu search (TS). The simulation results show that the new algorithm achieves faster convergence, higher throughput, and minimum scheduling time compared to the two other algorithms. The results also show that RL-based dynamic allocation of front-haul transport block capacity based on actual radio resource block size can greatly reduce front-haul capacity requirement and minimize total end to end uplink scheduling latency.

Keywords: 5G, C-RAN, mobile fronthaul, reinforcement learning, resource allocation, WDM-PON, TWDM-PON.

1 Introduction

Cloud radio access network (C-RAN) is a leading technology for next generation mobile network 5G. In 5G C-RAN the traditional base station functions are split between three entities known as the central unit (CU) which contains a number of virtualized baseband units (vBBUs) pooled in a central location to facilitate signal processing, transmission scheduling and resource sharing, the remote radio units (RRUs) which are remotely deployed at the cell sites, and the distributed units (DUs) which can be independently deployed together with CUs or DUs [1]. The interface connecting

between CU and DU is known as midhaul interface (also known as Fronthaul-II or F1 interface), and the interface connecting between RRU and DU is known as mobile fronthaul interface (also known as Fronthaul-I or Fx interface) [2].

Passive optical networks (PONs) are promising technologies for supporting front-haul and mid-haul interfaces in next generation mobile network (5G). For example, current commercial PONs such as XGS-PON and 10GEPON are capable of supporting mid-haul interface without any modification as the capacity and latency requirement for such an interface is similar to traditional backhaul network [2]. However, for front-haul interface some modifications regarding the latency and bandwidth efficiency are required because such an interface requires a high capacity and low latency transport network solution [2].

There are many proposals in literature that studied the latency and bandwidth efficiency issues of PON based mobile front-haul. The existing popular proposals are: 1- Traffic estimation low-latency PON based mobile front-haul [3], which relies on predictive method to estimate the scheduling grants for the optical network units (ONUs) to minimize mobile front-haul scheduling latency. 2- Mobile–DBA front-haul [4], which utilizes the mobile uplink scheduling information to compute the scheduling grants for the ONUs in order to eliminate the scheduling delay and the waiting time of ONUs. 3- Mobile-PON proposal [5] which relies on PHY-2 split option to increase front-haul efficiency and unifies PON and LTE schedulers by dynamically or statically mapping of LTE radio resource blocks (RBs) into the PON front-haul transport blocks (TBs) to eliminate front-haul latency.

The major limitation of these proposals is that all of them consider single wavelength PONs mobile front-haul; whereas, due to the huge data-rate requirement for front-haul interface in 5G mobile network C-RAN architecture, single wavelength PONs are insufficient for supporting 5G C-RAN. Another limitation is that in Ref [5] the authors assume a fixed front-haul TB size to be allocated to every RB independent of actual RB capacity. However, in practical LTE network the actual capacity of the RB depends on many factors such as user equipment UE request size, channel quality status and modulation and coding (MCS) schemes used during uplink transmission [6]. A fixed TB allocation can decrease front-haul efficiency and increase front-haul uplink latency.

Our major contribution in this paper is that we extend the low-latency PONs based mobile front-haul proposal to the multi-wavelength domain (e.g. WDM and TWDM-PON) and try to overcome the latency and the bandwidth efficiency problems we mentioned earlier. To do that, we propose to jointly allocate C-RAN air interface resources and fronthaul uplink resources to the users at the granularity of LTE media access control (MAC) layer sub-frame cycle which known as transmission time interval (TTI) (i.e. one TTI equals 1ms). We formulate the joint radio and fronthaul resource allocation framework as an optimization problem with the objective of finding an optimum or sub-optimum (RBs/TBs) to UE allocation pattern that minimizes total uplink scheduling latency (as well as fronthaul delay) and improves the total system throughput. Due to the complexity of such an optimization problem, because of the contiguity constraint on single-carrier frequency-division multiple access (SC-FDMA) uplink transmission, we introduce a reinforcement learning algorithm to solve the problem and evaluate its performance against some other heuristic approaches

The rest of this paper is organized as follows. In section II, we present the system model for multi-wavelengths enabled C-RAN and formulate the uplink resource allocation optimization problem. In Section III, we introduce a solution to our resource allocation optimization problem based on Q-leaning algorithm. In section IV we evaluate the performance of our solution, and in section VI we give the conclusion for our paper.

2 Introduction

2.1 Multi-wavelength enabled C-RAN architecture

The system model considers a C-RAN network consists of M RRUs; each RRU is attached to an optical network unit (ONU) (Fig. 1). The ONUs are aggregated over an optical splitter to a TWDM or WDM optical line terminal (OLT) which is connected directly to a DU unit. The DU and CU are co-located together at the central office and connected to each other via a mid-haul network (e.g.,TDM-PON or Ethernet). The CU system is virtualized into M vBBUs. Each vBBU is assigned a fixed wavelength channel to connect to its associated RRU. Each vBBU has a bandwidth equal to N RBs, and total C-RAN system is designed to serve K active mobile users.

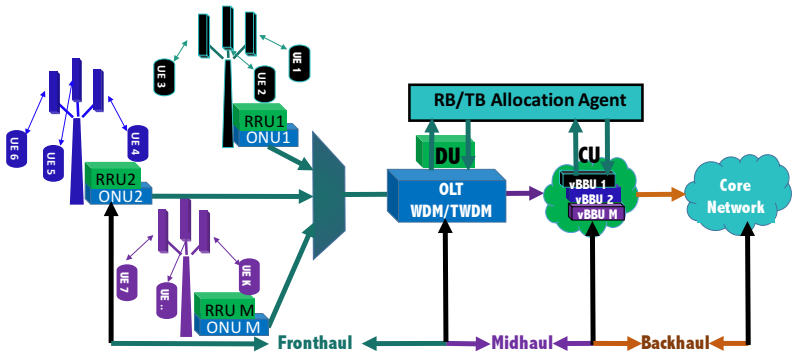


Fig. 1. Multi-Wavelength-Enabled C-RAN Architecture.

We assume that a learning based software agent that coordinates between CU and DU/OLT (assuming a 5G system with dual split as in Fig. 1) is in charge of the scheduling process of uplink air interface and front-haul resources. During the uplink scheduling process, every UEs in the network sends scheduling requests to ONU/ RRU. These requests contain UEs buffer status report (BSR) and channel quality indicator (CQI). The ONU/RRU transmits on single wavelength the UE requests to OLT which passes these requests to the CU unit at the C-RAN center. The scheduling agent at CU utilizes BSR and CQI information to compute the scheduling decision for the radio interface and fronthaul resources (i.e.RB /TBs allocation to UEs) every TTI period.

The final scheduling decision in form of grant allocations is broadcasted over all wavelength channels of the fronthaul aggregation network to ONUs. Each ONU in the network receives these grant allocations; however, its MAC layer protocol permits only the processing of the allocation associated with the RRU that it is connected to. Finally, the RRU sends the scheduling allocation grants to UEs over the air interface.

2.2 Multi-wavelength enabled C-RAN architecture

In C-RAN system described above, we assume that the allocation of air interface resource block (RB) and fronthaul upstream transport block size (TB) to users is done in a slotted scheduling base, with a slot duration equal to one TTI. At each scheduling slot, the RB/TB can be allocated to a one user at most. In order to efficiently utilize RB/TB resources during uplink scheduling while achieving a minimum UE uplink delay in multi-wavelength mobile fronthaul network, we formulate an optimization problem with the objective to minimize the total sum of idle time over the all wavelengths and vBBUs in the network. Fig 2 illustrates the calculation process of sum of wavelength during a TTI duration cycle. In this figure,A_{ij} denotes the j^{th}incoming scheduling requests processing time on the i^{th}wavelength channel of fronthaul network; where, $j \in \{\in \{1, 2,3\ldots, J\}$ is the index of the request with J as the total number of requests. $i \in \{1, 2,3\ldots, M\}$ denotes the index of the wavelength channel with Mas the total number of the wavelengths.B_{ij} denotes the off-scheduling time on the i^{th}wavelength channel ,and λ_i denotes theith wavelength channel $i \in \{1, 2,3\ldots, M\}$.Assuming the above notation and referring to Ref[7] flow-shop scheduling problem, the total sum of ide time as can be written as illustrated in Fig. 2.

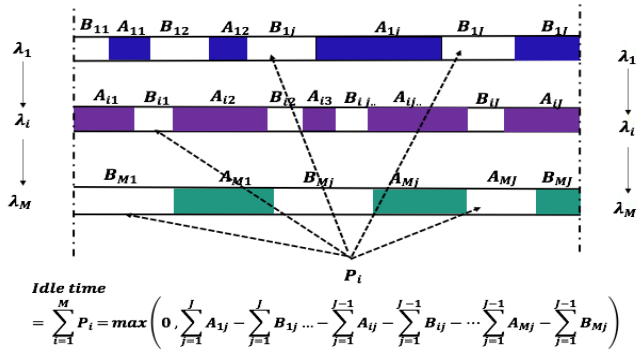


Fig. 2. The of sum of idle time for the wavelengths.

In order to describe our problem define the following notations: K $(k = 1, 3\ldots, K)$ as the number of active user, M $(i= 1, 3 \ldots M)$ as the number of wavelengths/vBBUs in the C-RAN system, N $(n = 1,3, \ldots \ldots N)$ as the number of RBs in C-RAN network , P_i as the sum of idle time on wavelength i when assigning TB/RB n to UE requests k and $y_{i,k,n}$, $y_{i,k,n} \in (0,1)$ as a selection variable that indicates whether the RB/TB n on wavelength i is allocated to UE k or not ($y_{i,k,n} == 1$ if TB_n is allocated to UE k and 0 otherwise). Given the above notations our optimization's objective function can be written as:

$$\underset{p}{\text{minimize}} \sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{n=1}^{N} P_i * y_{i,k,n}$$

subject to the following five constraints:

$$\sum_{k=1}^{K} y_{i,k,n} \leq 1 \quad ,\forall n \in \{1,2 \dots N\},\ \forall i \in \{1,2 \dots M\} \tag{1}$$

$$\sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{n=1}^{N} y_{i,k,n} \leq M * N \tag{2}$$

$$\sum_{i=1}^{M}\sum_{n=n'}^{N} y_{i,k,n} * B_{i,k,n} \leq B_{i,k,max}\ , \forall k \in K^* \tag{3}$$

where $B_{i,k,n}$ is the rate (in bytes) that user k obtains if RB n' is assigned to it, and $B_{i,k,max}$ is the maximum number of bytes requested by user k. K^* is a set contains the UE who has the highest rate over the RB n'.

$$\{ n' \} = \underset{n \in N^*}{argmax} \left(\sum_{K=1}^{K}\sum_{n=1}^{N} W_{k,n}\ y_{k,n} \right) \tag{4}$$

where $W_{k,n} = \frac{\sum_{i=1}^{M} W_{i,k,n}}{M}$ is the matrix that defines the UE over vBBU gain which is calculated based on proportional fair (PF) metric (Note: proportional fair supports high resource utilization and maintains a good fairness among network flows [6]) , $W_{i,k,n} = \frac{R_k\,(n,t)}{T_k(t)}\ \forall i,$ is the proportional fair metric for the UE k over RB n on sub-frame or TTI index (t), $T_k(t)$ is long-term average throughput of user k computed over TTI index (t), $R_k\,(n,t) = log\,(1 + SNR_k(n,\ t))$ is the achievable rate of user k over RB n and at TTI index t , and N^* is a set contains groups of RBs that maximize total PF metric if they allocated to specific UEs (i.e. $k \in K^*$).

$$y_{i,k,n} - y_{i,k\,,(n+1)} + y_{i,k,m} \leq 1\ , m = n+2, n+3 \dots, N \tag{5}$$

The constraint in Eq. (1) is used to limit the allocation of each RB/TB to one user during a single TTI period to avoid the interference (Note: LTE does not allow the allocation of less than one RB to UE). The constraint in Eq. (2) is used to limit the total number of scheduled RBs over all wavelength not to exceed the total capacity of the system (i.e. system stability constraint). The constraint in Eq. (3) is used to avoid over-allocating of RBs/ TBs to the UEs (i.e. ensure that the agent will not assign transport blocks more what the users have requested). The constraint in Eq. (4) is used to ensure that each RB is allocated to the UE that maximizes the total C-RAN PF metric (i.e. ensure each RB is allocated to UE that achieves highest CQI index or SNR value over that specific RB). The constraint in Eq. (5) is SC-FDMA contiguity constraint which is used to ensure all of the allocated RBs to a single UE are adjacent to each other in frequency domain.

The optimization problem we describe above belongs to the class of NP-hard problems due to the constraint given in Eq. (5) (the proof of the NP-hard can be found in [8]). Therefore, classical optimization methods such as branch and bound methods can only be used to solve the small-scale scheduling problems, for large-scale and

complex scheduling problem heuristic approaches or reinforcement learning can be used. Some heuristic approaches such as genetic algorithms [9] and Tabu search [10] have been already evaluated for uplink scheduling problem for disturbed RAN (D-C-RAN case). In this paper a reinforcement learning based solution is presented and its performance is compared with the above-mentioned heuristic methods under C-RAN architecture

3 RL for Resource Scheduling Problem in C-RAN

3.1 Reinforcement learning (Q-Learning algorithm)

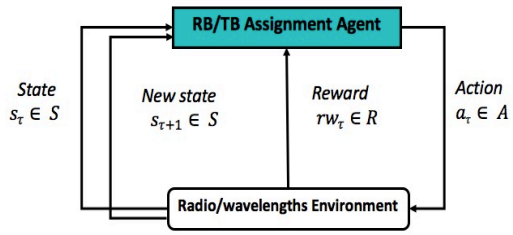


Fig. 3. Reinforcement learning elements

The resource allocation optimization problem in C-RAN is a complex scheduling problem that fits RL context. reinforcement learning, mainly Q-Learning algorithm [11], has shown positive results in solving some resource allocation problems similar to our problem (e.g. [12], and [13]). QL is an iterative model that can be defined by sets of states, actions and a reward function that produce a reward for each state- action interaction. As shown in Fig. 3, at each iteration the learning agent (TB/RB assignment agent) observes the environment state $s_\tau \in S$, then; applies an action $a_\tau \in A$ to the environment according to the strategy π. The environment transits into a new state $s_{\tau+1} \in S$ producing a reward signal $rw_\tau \in R$ to the agent. The agent updates its strategy based on the new state and the received reward. The basic goal of the agent is to choose the best action for each state that maximizes the cumulative reward as

$$Q(s_\tau, a_\tau) := E\left(\sum_{\tau=0}^{\infty}\left(\gamma^\tau * rw_\tau(s_\tau, \pi_s)\right)|s_0 = s\right) \quad (6)$$

where γ is a discount factor that reflects the significance of the upcoming reward relative to the current reward. When the selected action a is the optimal one $\pi^*(s)$, $Q(s_\tau, a_\tau)$is the maximum of the state. The update formula is given as

$$Q_{\tau+1}(s_\tau, a_\tau) := (1-\alpha)\, Q_\tau(s_\tau, a_\tau) + \alpha\left(rw_\tau(s_\tau, a_\tau) + \gamma \max_{a_{\tau+1} \epsilon A} Q_\tau(s_{\tau+1}, a_{\tau+1})\right) \quad (7)$$

where $\alpha \in [0,1]$ is the learning rate that balances new information against previous

knowledge. The Q-learning algorithm does not determine how the actions can be chosen in each state. To determine that, this paper considers ϵ −greedy policy, in this policy ϵ is the exploration rate which is used to choose a random action $a_\tau \in A$ with a probability falling between 0 and 1 (i.e. $\epsilon : 0 < \epsilon < 1$) this known as exploration, in contrast of choosing an action based on previous experience (i.e. selecting action with 1- ϵ probability), which known as exploitation. The exploration rate decays over the course of the learning until it reaches the minimum value.

3.2 The uplink resource allocation scheduling problem in reinforcement learning context

To write the uplink resource scheduling problem we described earlier in reinforcement learning context we can define the states, actions and reward function as follow:

1. ***State:*** S: $\{s_1, s_2, s_3, \dots \dots \dots, s_\tau\}$: as a combination of the total sum of idle time over the all wavelength channels w_τ and the total C-RAN system PF gain G_τ calculated the state transition (i.e. $s_\tau = (G_\tau, w_\tau)$). G_τ and $w_\tau$can be written as :

$$G_\tau = \sum_{K=1}^{K} \sum_{n=1}^{N} W_{k,n,\tau} \; y_{k,n,\tau} \tag{8}$$

, and

$$w_\tau = \sum_{i=1}^{M} P_{i,\tau} \tag{9}$$

2. ***Action:*** $A: \{a_1, a_2, a_3, \dots \dots \dots, a_\tau\}$: as the permutation of RBs allocation strategy to UEs ,and the permutation of sequencing order of the allocated RBs over the wavelength channel TBs as well as the permutation of the wavelength channels order.

3. ***Reward function*****:** $R: \{rw_{s_{1,}a_{1,}}, rw_{s_{2,}a_{2,}}, rw_3, \dots, rw_{s_\tau, a_\tau}\}$ as a function that rewards the unity value if the action has taken by the agent increases the total system PF gain and decreases the total sum of idle time over the past episode, otherwise it rewards the value (-0.1), this function is written as follow

$$rw_{s_\tau, a_\tau} = \begin{cases} 1 & if \quad G_{\tau+1} > G_\tau \; and \; w_{\tau+1} < w_\tau \\ -0.1 & Othewise \end{cases} \tag{10}$$

The optimization objective is to find the optimal/suboptimal RB to UE allocation pattern that maximizes the system PF gain and RB/TB to wavelength scheduling strategy that gives a minimum sum of idle time over the wavelength channels of the fronthaul network. Later on, this allocation pattern and scheduling stagey will be used to update the allocation of RBs to UEs and TBs to ONU/RRUs every TTI scheduling cycle. The complete algorithm for the scheduling is summarized by **Algorithm 1**.

4 Performance Evaluation Results

We evaluate the performance of our uplink scheduling algorithm in NS-3 simulator [14]. Since NS-3 does not support C-RAN and BBU virtualization, we use eNodeBs to

play the role of vBBUs in our simulations. In these simulations, we consider a C-RAN network with 4 RRU connected to over 4 WDM wavelength channels to 4 vBBUs resides in the cloud center. We assume different distances between each RRU/ONU and CU unit at the cloud center as follows: 5,10,15 and 20km. We consider urban propagation environment, where UEs are uniformly distributed in the network, and experience

Algorithm 1

Input: *The initial UE to RB/TB allocation strategy (i.e.* G_0, w_0 *)*
Output: *The optimal allocation strategy.*
Initialize $Q_{(s_0,a_0)} \leftarrow Q_{(s_0=(G_0,w_0),a_0=0)}$, τ =0, γ, α
For *iteration* $\leftarrow$ 1 *to Max-Iterations* **do**
Observe current state: s_τ -> (G_τ, w_τ)
Select an action : $a_\tau \leftarrow$ *Select Action*
While $s_\tau \neg$ *terminal episode* **do**
Perform the action a_τ*:* $(rw_{(s_\tau,a_\tau)}, s_{\tau+1}) \leftarrow$ *Take Action*(a_τ)
Observe new state : $s_{\tau+1}$-> ($G_{\tau+1}$, $w_{\tau+1}$)
Observe reward: rw_{s_τ,a_τ} *(Equation 10)*
Decay exploration rate $\varepsilon \leftarrow$ *max* ($\varepsilon \cdot d$, ε_{min})
Sample r ~ from uniform distribution (0, 1)
if $r \leq \varepsilon$ **then**
Select an action randomly.
else
Select an action a_τ *such that* $a_\tau = \underset{a_{\tau+1}}{argmax}\ Q_\tau(s_\tau, a_{\tau+1})$.
end if
Update the Q matrix using Equation 8.
Update $\tau \leftarrow \tau$ *+1 and the current state* $s_\tau \leftarrow s_{\tau+1}$
End while
End for

different MCS indexes ranging between 2~28. We assume adaptive modulation schemes for the uplink transmission, in which the C-RAN system senses the UEs channel quality condition and accordingly chooses the modulation scheme and the quantization resolution to be used. In this paper, we adopt three modulation schemes namely; QPSK, 16-QAM and 64-QAM, each with different quantization resolution bits as follow, 8 bit with 64QAM, 6 bit with 16-QAM and 4 bit with QPSK. We consider a random walk mobility model with an average UE movement speed equal 3km/h. For the traffic model, we assume a full buffer model with UE traffic load equal to 640kbps. The overall system parameters used during the simulation are summarized in table 1. For the Q-learning scheduling algorithms, we set the following parameters: $\alpha = 0.5$ and $\gamma = 0.5$. We use $\epsilon -$greedy as action selection policy with $\epsilon = 0.90$ at the beginning and decays until became 0.010 when enough number of the episodes have been explored. The complete parameters and settings used for the scheduling algorithms are given in table 2. We choose the total system throughput, total scheduling time, and the

speed of convergence as performance evaluation metrics. To evaluate these metrics, we run multiple. Fig 4 shows the overall performance comparisons.

Fig. 4(a) shows the achieved system throughput by each scheduling algorithm plotted versus the number of the active users during the simulation. From this figure, we can notice that the highest system throughput is achieved by RL algorithm followed by GA whereas TS algorithm achieves the lowest system throughput. We explain RL's superior performance by its ability to produce allocation patterns very close to the optimal as it does not require a long time to simulate the optimization solver as opposed to TS and GA algorithms (see Fig 4(c)).

Fig. 4(b) shows a comparison of the scheduling time consumed by each algorithm. As we can see, the RL algorithm also attains the lowest scheduling time compared to TS and GA algorithms. However, this time TS outperforms GA and achieves lower scheduling time. All of the three algorithms show a total scheduling time of less than 1ms (TTI period) when the number of active users in the system was less than 150 UEs. However, the scheduling time of GA exceeded 1ms when the number of users was 200 active UEs.

Fig 4(c). shows a performance comparison of the three algorithms in term of the speed of convergence considering the objective function given in equation 1. As we can see RL algorithm achieves the fastest speed of convergence on the objective function compared to GA and TS algorithms. In other words, RL algorithm converges to the minimum sum of idle time in the first 50 iterations while GA algorithm converges in about 80 iterations and TS algorithm converges in about 60 iterations; however, the convergence of TS is slightly unstable compared to RL and GA.

Fig 4(d). compares the performance of the total uplink delay for static RB to TB mapping, dynamic RB to TB mapping [5] and our new adaptive-TB allocation method. From this figure we can see that our new adaptive-TB allocation method achieves the lowest total uplink scheduling delay in comparison with static and dynamic RB to TB mapping proposals. The reason behind the improved delay performance achieved by adaptive-TB allocation method is the efficient utilization of fronthaul uplink resources (see Fig 5(a) and (b)). This is due to the fact that adaptive-TB method allocates an adaptive fronthaul TB size equal to the actual RB size calculated by the scheduling algorithm (RL) based on UEs traffic load and channel condition. This method can greatly reduce the capacity required on fronthaul as opposed to static and dynamic RB to TB mapping methods which assume fixed fronthaul TB size for every RB.

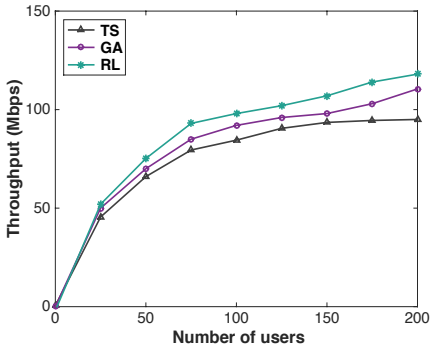
Table 1. Simulation Parameters

Parameter	Value	Parameter	Value
Simulation length	1000TTI	Channel band-width	10 MHZ
Link adaptation	QPSK,16QAM, 64QAM	Number of RB(per TTI)	400 RB
Propagation model	COST 231	Maximum Number of us-ers	200

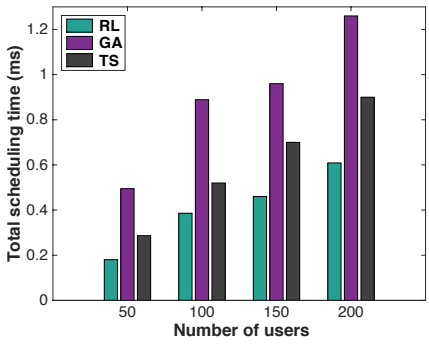
Mobility model	Random-walk	Number wave-length	4(10Gbps)
Transmission power	eNB: 30 dBm; UE: 23 dBm	Front-haul capacity per RB	Equal to RB size
TTI	1 ms	UE speed	3km/h
eNB Antenna Model	Cosine Antenna /3 sectors	eNB Number of MIMO	(4 x4)
Front-haul distance	5km,10km, 15km,20km	UE Traffic model	Full Buffer
Propagation environment	Urban	Load	640kbps

Table 2. Algorithms Settings

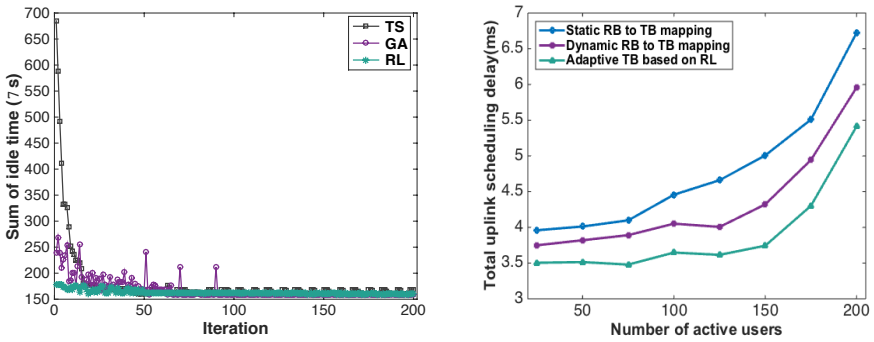
Parameter	Value
Number of Iteration	200
RL Discount Factor	0.5
RL Learning Rate	0.5
RL Exploration Rate	0.90
RL Minimum Exploration Rate	0.010
RL Exploration Rate Decay	0.99
GA Parameters	Same as [9]



(**a**)

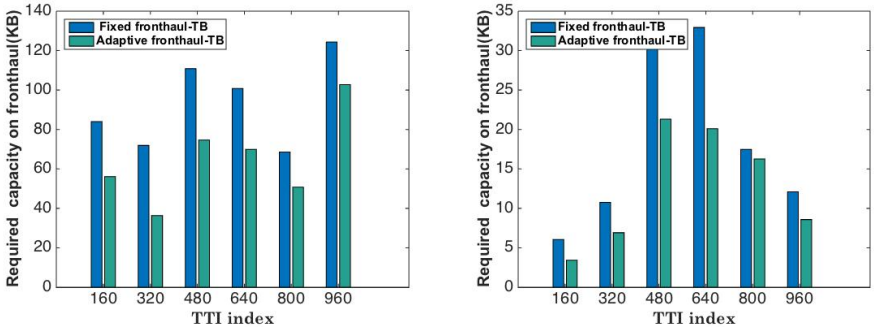


(**b**)



(**c**) (**d**)

Fig .4. Performance comparison: (a) The achieved throughput ;(b) total scheduling time consumed by each algorithm (The convergence speed) ;(d) comparison of UE total uplink scheduling delay with static RB to TB mapping, dynamic RB to TB mapping and adaptive TB allocation methods.



(**a**) (**b**)

Fig .5 The total required capacity on front-haul link with fixed front-haul TB allocation and adaptive front-haul TB allocation: (a) 50 active UEs, (b) 200 active UEs.

5 Conclusion

In this paper a reinforcement learning based scheduling algorithm is proposed to address the resource allocation optimization problem for multi-wavelength Enabled-C-RAN architecture. The performance of the algorithm is validated with simulation and compared with two other heuristic approaches. The simulation results have shown that RL based scheduling is the most promising approach, as it outperforms the two other heuristic methods in all performance evaluation metrics, offerings the highest system throughput, lowest scheduling time and total uplink scheduling latency. The results have also shown that adaptive allocation of fronthaul transport resources with RL based

scheduling which rely on UE traffic load and actual radio condition can greatly enhance the C-RAN system performance in terms of uplink scheduling delay and fronthaul efficiency.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) (61431009, 61371082, and 61521062) and the National Science and Technology Project -China (2015ZX03001021).

References

1. Wey, Jun Shan, and Junwen Zhang. "Passive Optical Networks for 5G Transport: Technology and Standards." *Journal of Lightwave Technology* (2018).
2. Mikaeil, Ahmed, et al. "Traffic-Estimation-Based Low-Latency XGS-PON Mobile Front-Haul for Small-Cell C-RAN Based on an Adaptive Learning Neural Network." *Applied Sciences 8.7 (2018):* 1097.
3. Nomura, Hiroko, et al. "First demonstration of optical-mobile cooperation interface for mobile fronthaul with TDM-PON." *IEICE Communications Express 6.6 (2017)*: 375-380.
4. Siyu, et al. "Low-latency high-efficiency mobile fronthaul with TDM-PON (mobile-PON)." *Journal of Optical Communications and Networking 10.1 (2018)*: A20-A26.
5. Aditya Tiwari, S. S. "LONG TERM EVOLUTION (LTE) PROTOCOL Verification of MAC Scheduling algorithms in NetSim." (2014).
6. Stefan, Peter. Combined use of reinforcement learning and simulated annealing: algorithms and applications. VDM Publishing, 2009.
7. Lee, S-B., et al. "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink." *INFOCOM 2009, IEEE*. IEEE, 2009.
8. da Mata, Saulo Henrique, and Paulo Roberto Guardieiro. "Resource allocation for the LTE uplink based on Genetic Algorithms in mixed traffic environments." *Computer Communications* 107 (2017): 125-137.
9. Khdhir, Radhia, et al. "Tabu Approach for Adaptive Resource Allocation and Selection Carrier Aggregation in LTE-Advanced Network." *Computer and Information Technology (CIT), 2016 IEEE International Conference on*. IEEE, 2016.
10. Watkins, Christopher John Cornish Hellaby. *Learning from delayed rewards*. Diss. King's College, Cambridge, 1989.
11. Gao, Zhibin, et al. "Q-learning-based power control for LTE enterprise femtocell networks." *IEEE Systems Journal* 11.4 (2017): 2699-2707.
12. Ye, Hao, and Geoffrey Ye Li. "Deep reinforcement learning for resource allocation in V2V communications." *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018.
13. GNU GPLv2, "ns-3.25," March, 2016 [Online]. Available: https://www.nsnam.org/ns-3-25..