Fine-Grained Management in 5G: DQL Based Intelligent Resource Allocation for Network Function Virtualization in C-RAN

著者	ZHANG Chaofeng, DONG Mianxiong, OTA Kaoru
journal or publication title	IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING
volume	6
number	2
page range	428-435
year	2020
URL	http://hdl.handle.net/10258/00010321

doi: info:doi/10.1109/TCCN.2020.2982886

Fine-grained Management in 5G : DQL based Intelligent Resource Allocation for Network Function Virtualization in C-RAN

Chaofeng Zhang, Mianxiong Dong and Kaoru Ota, *Member, IEEE*

Abstract—Recently, the installation of 5G networks offers a variety of real-time, high-performance and human-oriented customized services. However, the current laying 5G structure is unable to meet all of the growing communication needs by these new emerging services. In this paper, we propose a DQL (Deep Q-learning Network) based intelligent resource management method for 5G architecture, to improve the quality of service (QoS) under limited communication resources. In the environment of network function virtualization (NFV), we aim at improving the efficient usage of spectrum resources. In this two-step solution, our first goal is to guarantee the maximum communication quality with the smallest number of infrastructures. Then, a DQL-based wireless resource allocation algorithm is designed to realize the elaborate operation. Unlike previous studies, our system can provide the allocation policy in a more subdivided way and finally maximize the usage of bandwidth resources. The simulation also shows that our proposed MSIO improves 3.12% in the performance of the maximum coverage importance problem and the ARODQ algorithm improves 4.05% than other standard solutions.

Index Terms—Intelligent control, Intelligent networks, Artificial intelligence, Management decision-making, Mobile communication.

1 Introduction

Nowadays, the conception of 5G has gradually entered people's vision. With the broad deployment of IoT and mobile devices, 5G technology provides more sufficient services than conventional mobile communication technologies. Predictably, billions of devices will be connected to the new mobile communication band in a few years. Users' habits will be changing dramatically because of the support of these mobile wireless communication devices. In 2019 July, the Japanese government has attempted to install 5G-enabled devices on street power poles to assist the realization of a city-level advanced self-driving system [1]. Some social issues, such as elderly drivers and full-fledged trend toward population, would be alleviated by the city-level dedicated band based autonomous systems. As the evolving of artificial intelligence causes more intelligent services to come, more 5G infrastructures for this kind of specific use would be requested in the future [2].

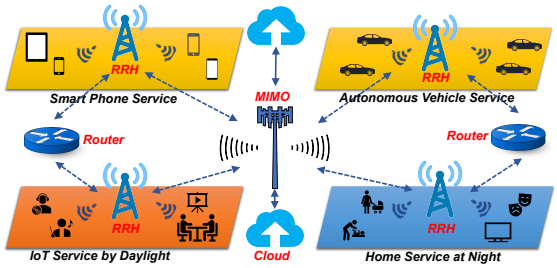


Fig. 1. The architecture of network function visualization based 5G network. The network is divided into specific channels to satisfy different uses.

Different from conventional mobile wireless technologies, 5G leverages denser cellular technology and flexible band configurations to achieve greater transmission throughput [3]. With the increasing of communication demand, 5G may only rely on more infrastructures to provide individual channels or communication devices for the specific use of these services, as shown in Fig. 1, which are limited by the coverage of RRHs (remote radio heads) and the complicated channel conditions [4]. To overcome these problems, the pioneers have accordingly provided a variety of technologies for 5G architecture: C-RAN (cloud radio access network) [5], as a high-profile design for 5G networks, can control and allocate spectrum resources more flexibly [6] from the centralized clouds. WNV (wireless network visualization), which concentrates on a higher level of resource management and virtualization, integrates the network capabilities among the 5G infrastructure architecture [7] [8].

Although the above solutions can use in a more efficient way, they still cannot fulfill the performance requirements of the growing heterogeneous wireless network communications. Specifically, with the advent of the Great Age on the Internet, various devices such as IoT, smart devices and wearable devices are on the rise [9]. To improve the usage efficiency of 5G communication resources, one of the primary problems is how to flexibly allocate existing communication resources [10] [11]. In particular, in the absence of effective expansion of spectrum resources, the existing system has

- *Manuscript submitted in Feb. 21, 2020.*
- *Chaofeng Zhang is in the Department of Information and Electronic Engineering, Advanced Institute of Industrial Technology, Tokyo, Japan. E-mail: zhang-chaofeng@aiit.ac.jp,*
- *Mianxiong Dong and Kaoru Ota are with the Department of Information and Electronic Engineering, Muroran Institute of Technology, Muroran, Japan. E-mail: {mxdong, ota}@mmm.muroran-it.ac.jp*

to not only take care of the QoE of the traditional services [12], but also quickly respond to the needs of new type services. Therefore, how to guarantee the communication infrastructures for new specific channels, and then provide sufficient bandwidth usage under limited spectrum resource is the core problem in our study.

To overcome the resource tradeoff problem in future 5G, this paper focuses on the flexibility and scalability of the 5G resource allocation system, to maximally ensure the quality of service (QoS) in 5G architecture. We propose a deep reinforcement comprehensive solution for 5G architecture, to improve overall usage efficiency under limited communication resources. At first, we formulate the scenario as a maximum coverage problem and the optimization goal is to maximize the communication coverage with the minimum number of infrastructures. Then, a DQL-based network decision-making method is designed to allocate the network resource in real-time. Unlike previous studies, this decision-making method is based on the slicing of the existing channel conditions, which provides resource allocation policies in a more subdivided way to realized maximized bandwidth usage.

The structure of the following part is introduced here: in Section 2, we discuss the state of art technologies in the cross-field of C-RAN and DQL. In Section 3, we formulate the basic service providing problem as a maximum coverage quality problem and develop an algorithm to solve it. Correspondingly, we propose a deep Q-learning method to manage resource allocation policies in Section 4. In the simulation of Section 5, we testify our theory and emphasize the flexibility and efficiency of our proposed solution. Finally, we close our paper with future discussion in Section 6.

2 Related Work

To overcome the spectrum resource inefficiency in 5G architecture, the C-RAN technology is needed to maximize the scheduling of all the message transmission, when the huge computation can afford the AI technologies to achieve a higher performance than pre-defined schedules. The challenge becomes how to realize a stable and joint platform for new 5G architecture [13]. More than that, the heterogeneous communication means in 5G also make the combination of these network resource as new challenges, and only the highly integrated heterogeneous network could be fluently controlled by the C-RAN technology.

C-RAN based Solutions: C-RAN, which is a key technology to enable better capacity and flexibility for the 5G architecture [14], takes advantage of the computation ability in the cloud to realize traffic control in RRHs. In recent years, many studies have developed a large-scale solution to holistically control the transmission platform. Because of the cost and unbalanced spectral allocation issues, Park [15] pays more attention to the front-haul and access links, to maximize the overall throughput of all mobile devices which are active in the clusters. Meanwhile, the limitation of power control is also satisfied in BBUs and RRHs, to achieve a better efficiency performance in a C-RAN architecture. Furthermore, to control the wireless communication system more accurately, Cai [16] overcomes the problem of outdated CSI (channel state information). The authors highlight the topology settings and spectrum allocation problem under the old CSI based framework. Then they propose a more solid policy for the above stochastic optimization, where the computation cost is low in the cloud and the performance is better, due to the online structure of C-RAN and MCC. To explore a better capacity of cloud-based radio access control, Li [17] conducts a QoS oriented computation and communication resource allocations service which aims at minimizes the overall power consumption of the C-RAN system. A Mixed Integer Non-linear Problem is proposed to describe the resource allocation problem, and they develop a geometric map to fix the virtual machine-users.

Challenges in Heterogeneous Networks: As discussed in the previous section, a few new concepts emerge to enable solutions for the specific use of wireless network visualization. The traditional NFV (network function visualization), aiming to implement new network functions with software design in order to satisfy use of [18], [19]. Another new innovated conventional technology, SDN (software-defined network) [20], is used to decouple the control and forwarding of the data stream, which makes all the resources programmable. In C-RAN architecture, network computation intelligence is the advantage of making policies and resource allocation by the centralized cloud. As in a niche of wireless network virtualization, resource blocks (RBs) are considered as a way of spectrum resource allocation solution to guarantee QoE [21], with a higher rate and lower delay. As a kind of network slice, the optimization goal transfers from the data forwarding on the whole spectrum band to achieve better performance on the dedicated vitalized wireless band.

DQL in 5G networking: Computation intelligence is nowadays a widely used approach to improve the decision making the accuracy of the control system, from individual devices to the centralized wireless communication resource control [22] [23]. Sun [24] proposes dynamic reservations systems to make a more efficient resource arrangement. As one of the virtual networks' core problems, the infrastructure has to reserve some resources for future use, and the proposed allocation method can maximize the overall service utility of these resources through deep reinforcement learning. Then, improve the Q-learning model, Qi [25] develops a normalized advantage functions to reduce redundancy of state calculation. The core concept of the solution is using k-means to find an appropriate action in the discrete space to speed up the repetition of a similar policy.

3 Service-oriented Coverage Problem

In Fig. 2, it shows the deep Q-learning-based comprehensive solution for 5G architecture. At first, we transfer the VNF architecture into a channel dividing based service-oriented coverage allocation problem, which aims at the reasonable deployment of RRHs. After solving the maximum coverage problem by proposing a channel division based RRH covering algorithm, we further discuss how to transfer the real communication condition to neural network learn-able problem. A DQL based network resource allocation method

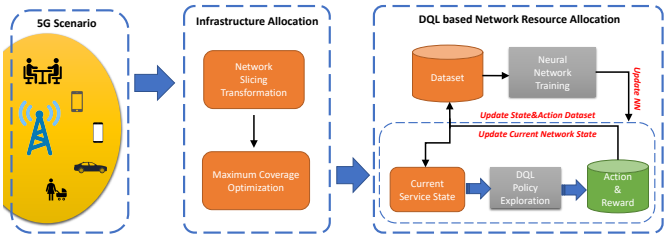


Fig. 2. The system model of network slicing based resource allocation, which use the infrastructure method to solve the maximum coverage optimization problem and a DQL based network resource allocation to complete the allocation policy exploration.

is proposed for policy exploration. Next are the detail steps to realize the maximum coverage optimization.

3.1 Deployment Model

For some special 5G services, such as autonomous driving mentioned above, the RRHs for vehicular signal transmission have to be mounted on the grid along the road in a region. Without the loss of the generality, we consider a plane wherever the RRHs can be distributed to conduct our research. A VNF based 5G architecture can be simplified as a star topology where the sink head is the base station [26], and expanded by several distributed RRHs which timely connected by mobile devices through particular channels.

Next, we consider there exists a set of RRHs $M = \{m_1, m_2, \cdots, m_{max}\}$ on the map. Meanwhile, there will be several channels used to sever specific applications, noted as $C = \{c_1, c_2, \cdots, c_{max}\}$. A channel is considered as pure subband divided from whole 5G spectrum band to provide safe and uninterfered communication services. Thus, for one RRH m, a set of available channels $C^m = \{c_1, c_2, \cdots\}$ is defined for specific use of services. For an arbitrary RRH m holds the non-specific use channel noted as $c_0 = \forall c_i \in C$.

On the real map, the mobile nodes are usually active in the street blocks. We consider a communication requirement map Φ that can be divided into blocks, noted as $\Phi = \{\Phi_1, \Phi_2, \cdots\}$. Then, we preset an importance ratio ϕ to present the service importance in this region (e.g. the government service area with the fire station and police station). We also use the acreage set $S = \{s_1, s_2, \cdots\}$ to present the actual size of each block. Assume there will be $|C|$ services that need to use specific channels for wireless communication. Then, for arbitrary c-th channel, we create a specific channel map of the region Φ^c for each, where ϕ_k^c presents the importance in channel c of block k. If in present the block k is served by at least one RRH using channel c, the binary parameter of X_k^c is considered as 1, otherwise, denoted as 0.

3.2 Maximum Coverage Problem

To guarantee the best service for every channel, the optimization goal is that the RRHs can cover the blocks in all the channels as much as they can. Then, the problem can be formulated as follows: On a communication requirement map Φ divided by blocks, given a set of RRHs M with limited channel service capacity during a period of service time T, in order to serve mobile nodes set N's demands, find a schedule for the maximum coverage problem that in each time slot t the total sum of coverage binary parameter X could be maximum.

To develop a dynamic solution for the multiple-channel scenario, we redefine the channel division based maximum coverage (CDMC) problem as:

On a communication requirement map Φ divided by blocks with importance ratio ϕ and acreage s, given a set of RRH M with limited channel service capacity during a period of time T, in order to serve randomly deployed mobile nodes set N's served by different channels and bandwidth requirement, find a solution for channel division based maximum coverage (CDMC) problem so that during time period T the overall coverage importance for all channels C could be maximum.

In a general condition, if we set ϕ and s as 1, the coverage quality problem can be easily transferred as a problem of service coverage. For an arbitrary map of Φ, we denote the coverage ratio of the map as $CO(\Phi) = \{(\Phi_t, t) | t \in T\}$. Then, we define the Coverage Importance for the time frame T as,

Maximize

$$CO(T) = \frac{\sum_{(\Phi_t,t)\in CO(T)} \sum_{\Phi_j \in \Phi_t} X_j}{|T| \cdot |\Phi|} \quad (1)$$

where the upper part of the fraction means the total number of blocks severed in each time stamp t and the lower part is sum of blocks $|\Phi|$ should be served according to all the time stamp $t \in T$. Notice that it is possible that multiple nodes are located in one block, where $X_j = 1$ is considered in these cases. The RRH m_p served to this block j is noted as $m_p \in \Phi_j$.

Then, if the map Φ is fully covered in any time of T, the coverage importance would be $CIM(CO(T)) = 1$. When the channel resource is limited, the CIM would be less than 1 and it will become a problem to search for the maximum value of CIM; If the communication resource is enough $CIM \geq 1$, there will be surplus resources for flexible use. First, we will discuss the necessary condition of ensuring $CIM = 1$, then develop an algorithm to maximize the coverage importance value. Obviously, the necessary condition of the maximum coverage would be:

(1) Every block belonging to map Φ in arbitrary channel $c \in C$ has been served during T, $\forall X_k^c = 1$.

(2) Each sub-map covering Φ^c is always a connected undirected acyclic graph.

If there is a map Φ served by a set of RRH $n = |M'^c|$, where $M'^c \in M$ for a specific channel c. Let $|\Phi_0|$ be the area of Φ, which divided by the squares with length d. When there exists a positive integer h and $0 < \epsilon < 1$ and we can have

$$\frac{|\Phi|}{d^2} \sum_{i=1}^{|c|} \sum_{j=0}^{h} C_n^j \left(\frac{1}{|\Phi^i|/d^2}\right)^j \left(1 - \frac{1}{|\Phi^i|/d^2}\right)^{n-j} < |c|\epsilon, \quad (2)$$

where the $\frac{|\Phi|}{d^2}$ is the number of square blocks and $|c|$ is the number of channels.

As shown in Fig. 3, in each square, it will be more than h nodes. Based on the Bayesian probability, all the blocks

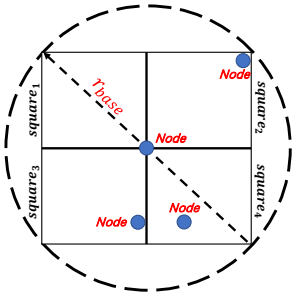


Fig. 3. The worst case of RRH deployment. Square 1 still can be fully covered.

can be served during the whole duration, and finally we get $CO(T) = 1$.

3.3 Maximum Service Importance Optimization Algorithm

Since a map of Φ with above allocation with computation complexity $O((h+1)^M)$ may lead to more than a NP-hard problem, to approximate the optimization, we develop a maximum service importance optimization algorithm through the following two steps: Temporal Maximum Coverage Deployment to find one unique RRH which can optimally improve coverage importance at this time point; Overall Coverage Construction to calculate the overall solution for the whole service period T.

3.3.1 Temporal Maximum Coverage Deployment

This subsection is to discuss how to approximate the maximum coverage importance for all the channels in a specific time t. Without loss of generality, we assume the channel requirement in each block is a dynamic value depending on the time. Thus we develop a more versatile algorithm to search for the solution.

In a time slot t, there will be c independent maps needed to be served. At first, we set up the empty $|c|$ disjoint sets, denoted as $M' = \{M'^1, M'^2, \cdots, M'^c\}$, which is similar with Algorithm 1 (CDRRHC). Then, a temporal maximized coverage importance CIM' is used to present the current value after a node put into the set of M', where we have

$$CIM'(CO(M')) = \frac{\sum_{i=1}^{|c|}\sum_{\Phi_j^i \in \Phi_{M'}} \phi_j s_j X_j^i}{|M'|\sum_{i=1}^{|c|}\sum_{\Phi_j^i \in \Phi} \phi_j s_j}. \quad (3)$$

The upper part of this fraction is the sum of the coverage importance served by RRHs while the lower part is the sum of importance value in all the blocks. If multiple RRHs are allocated in one block j, X_j^i is also considered as 1 in channel i. To simplify the problem, we also consider that one RRH can only serve one channel. Therefore, the arbitrary RRH m_p served to this block j in channel i is noted as $m_p \in \Phi_j^i$.

When a new candidate node $j \in M$ with specific channel c is put into the set of M', there will be a non-negative value if a new block is covered. Let a function $r(j)$ represent there exists a node nearby and connected to the cloud through the network, where we could denote it as $r(j) = \{r_{mobile}(j) \cup r_{base}(j)\} = 1$. Then, for $\forall j \in M$ as a candidate, we calculate the increment value for putting into the M' set as

$$CIM^+(j) = CIM(CO(M' \cup j)) - CIM(CO(M')) \quad (4)$$

where $r(j) = 1$. The value will be positive if a new block is covered, otherwise, it becomes negative. The reason why $CIM^+(j)$ becomes negative is that if node j provides the communication area concluded by other nodes in M, based on Eq. 3, the only increment of denominator would only lead a smaller CIM' in the future, which finally leads to a negative benefit of CIM^+.

To achieve the highest coverage importance, we consider a greedy strategy that focuses on improving the $CIM+$ in every step, which aims at

$$\begin{aligned} argmax(CIM(M,j)) &= (CIM^+(j))' \\ &= \frac{\sum_{i=1}^{|c|}\sum_{\Phi^+} \phi_j s_j X_j}{|M'^2 + M'|\sum_{i=1}^{|c|}\sum_{\Phi_j^i \in \Phi} \phi_j s_j} \end{aligned} \quad (5)$$

where $\Phi^+ = \Phi_{(M' \cup j)} - \Phi_{M'}$. Therefore, this function converges to 0 if no more blocks would be covered. This function is a general indicator for coverage allocation problems, which also can be applied in the former problem if the computation complexity is beyond the polynomial time.

Algorithm 1 Maximum Service Importance Optimization Algorithm

Require: Φ: a service requirement map, concluding block importance ϕ and acreage s; T: a time frame ; M: RRH node sets;
Ensure: M': the disjoint set of nodes for covering blocks; CIM: the overall service importance;
Set empty disjoint node set $M' \rightarrow 0$;
for each $t = 1; t \leq T; t++$; **do**
  Input importance $\phi(t)$ and acreage $s(t)$;
  for each $j = 1; j \leq m; j++$; **do**
    Calculate temporal importance array $ImList(j, |C|) \leftarrow argmax(CIM(M', j))$
  end for
  Select node j in the highest value in $ImList$;
  if highest value in $Max(Imlist) \rightarrow 0$ **then**
    Loop Break;
  end if
  Add Node j into Set $M'_t \leftarrow M'_t \cup j$;
end for
Calculate the total importance $CIM(CO(M'))$;

Following is the complexity description of our proposed Maximum Service Importance Optimization (MSIO) Algorithm: Step.1 and Step.2 are normal set up. Step 3 is a traversing process for all the nodes in M which causes computation complexity as $O(|M||C|)$ and Step 4 makes the complexity of the loopless than $O(|M|^2|C|)$. Through time period T, the overall complexity to output is the final solution would be $O(|M|^2|C|T)$. When considering T and C is a relatively small constant value, the overall computation complexity of MSIO Algorithm becomes $O(M^2)$, which is under polynomial time.

4 Resource Allocation using DQL

As mentioned above, the visualized network functions (VNF) can visualize various resources, including the predefined wireless channels [27] [28]. Although the MSIO algorithm solves the problems of channel slicing and RRH allocation, which contains an independent map and nodes of disjoint set, however, there still allocation problems if one RRH covers more than one block.

4.1 Optimization Goal and Action Reward

In each channel, the balance of bandwidth resources in each block is important due to the service standards such as QoE, and the waste of spectrum would also lead to a lower capacity of the whole network. Assume on a specific channel division map Φ^c, the communication demands in each block is denoted as $D = \{d_1, d_2, \cdots, d_{|\Phi|}\}$, and the allocated bandwidth from all the m RRHs is considered as $B = \{b_1, b_2, \cdots, b_m\}$, where $\forall b_i = [b_{i1}, b_{i2}, \cdots, b_{i|\Phi|}]^T$. Then, the reward for any resource allocation policy is denoted as $R(d, b)$, which is also the optimization goal of our evaluation system, where we have

Maximize

$$R(d,b) = \alpha_1 \cdot \frac{|min\{Bt/p, D\}|}{|D|} + \frac{\alpha_2}{|m|} \sum \cdot Sigmod[\theta_2 \cdot Sigmod(\theta_1 \cdot A)] \tag{6}$$

s.t, $|D| = d_1 + d_2 + \cdots + d_{|\Phi|}$
$|B| = \sum_{i=1}^{|\Phi|} \sum_{j=1}^{m} b_{ij}$
$|B| \leq |D|$
$A^T = [1.0, w_1, w_2]$

In Eq. 6, w_1 is the bandwidth occupancy ratio of real utilization and assigned to this block, denoted as $w_1 = (min\{d_i, b_{\Phi_i}\})/(b_i)$, and w_2 is the overall bandwidth occupancy ratio of the whole assignment, denoted as $w_2 = b_i/B$; p is the packet size of each communication demand.

On the right side of the equation, the first part including α_1 is the QoE value which indicates how many requirements have been satisfied, which also reflects the latency of the waiting queue. The parameter α_1 is adjustable to make the function elastic, which balances the different SLAs (service-level agreement) for channel C. The second part is a $XNOR$ function to emphasize the bandwidth occupation ratio of the whole network, to grantee the overall network capacity and stability.

As shown in Fig. 4, the XNOR function is implemented using a combination of AND, NOR, and OR calculation. Our proposed $XNOR$ neural network has three layers, including the input layer. The numbers 1 in the first and second layers are the bias units of these two layers. The letter θ, on the links, represents the parameter of each layer. The training process is considered as follows:

1) Input: Four training samples are used here, each with two characteristics: $(0,0), (1,0), (0,1), (1,1)$.
2) Target output: Refer to the output between 0 and 1 above, where the two inputs are the same (e.g. $(0,0)$ or $(1,1)$) as 0, and the maximum different inputs (e.g. $(1,0)$) as 1.
3) Parameter Initialization: Random, the range is from (-1, 1).

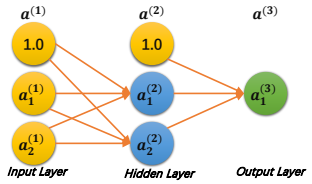


Fig. 4. The neural network architecture for the reward evaluation.

For network occupation ratio $\forall w_1$ and $\forall w_2$, we combine the input as $A^{(1)} = [1.0, a_1^{(1)}, a_2^{(1)}]^T$, and we have parameter inside,

$$\begin{aligned} a_1^{(2)} &= Sigmoid(a_1^{(1)} \cdot \theta_{11}^{(1)} + a_2^{(1)} \cdot \theta_{12}^{(1)} + 1 \cdot \theta_{10}^{(1)}) \\ a_2^{(2)} &= Sigmoid(a_1^{(1)} \cdot \theta_{21}^{(1)} + a_2^{(1)} \cdot \theta_{22}^{(1)} + 1 \cdot \theta_{20}^{(1)}) \\ \textbf{where} \quad & Sigmoid(z) = \frac{1}{1 + exp(-z)}. \end{aligned} \tag{7}$$

Thus the output of the first layer is $[A^{(2)}]^T = [1.0, [Sigmoid(\theta^{(1)} \cdot A^{(1)}]^T] = [1.0, Sigmoid([a_1^{(2)}, a_2^{(2)}]^T]$. Then, the connection between the last two layer of this BP is $A^{(3)} = Sigmoid(\theta^{(2)} \cdot A^{(2)})$. With already trained θ, and the network occupation ratio A, we can have the SE value $A^{(3)}$ as

$$A^{(3)} = Sigmoid[\theta^{(2)} \cdot Sigmoid(1.0; \theta^{(1)} \cdot A^{(1)})]. \tag{8}$$

4.2 DQL based Resource Allocation Decision Making Algorithm

Algorithm 2 Action Reward Optimization Deep Q-learning Algorithm

Require: Q: current Q-learning network with weight z;
Φ: new updated block information, including use demand d and RRHs M;
ϵ: A threshold for convergence
Ensure: π: New policy set $\pi = \{\pi_1, \pi_2, \cdots, \pi_t\}$ based on the target network Q^+;
for each $t = 1; t \leq T; t++$; **do**
Initiate a new network $Q^+ \leftarrow Q$ with weight z;
while $z^{i+1} - z \geq \epsilon$ **do**
Randomly select a user demand state s_i and a corresponding allocation action a_i;
Execute a_i, record the reward based on Eq. 6;
Go forward to the next state s_{i+1} based on the allocation action a_i;
Store basic state transformation record as $s_i, s_{i+1}, a_i, Reward(s,a)$ into the mini-batch set $Mini$;
Update $Q^+(s_i, a_i) = Reward(s_i, a_i) + \sigma \text{max} Q^+(s_{i+1}, a_{i+1})$ using mini-batch set $Mini$;
Update z;
end while
Update $Q \leftarrow Q^+$ using new updated parameter z;
Output the resource allocation policy π_t;
$t \leftarrow t + 1$;
end for

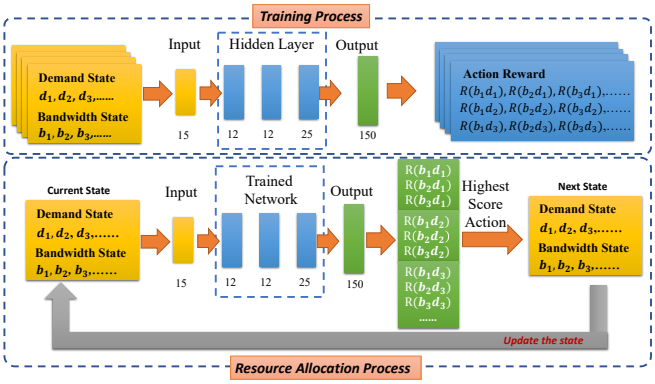


Fig. 5. The network architecture of DNN. In the training process, the input is an array, which including demand state d and bandwidth state b. The output is the action reward scores for each allocation policy, respectively. In the resource allocation process, one current state can be explored for multiple actions. When the highest reward score action is active, it changes to the next state for further policy exploration.

The above pseudocode is our Action Reward Optimization Deep Q-learning (ARODQ) Algorithm. The unique design of the learning model is the input and output of DQL, as shown in Fig. 5. The current state of the waiting queue and bandwidth allocation should be considered as input, which could significantly compress the input and output size in a continuous space. The input size is the sum of blocks and RRHs' current resource allocation percentage. The output size the actions of bandwidth allocation in each block using specific RRH, which is $M \cdot |\Phi|$.

5 Simulations

In this section, we detail how our experiment is conducted, where the whole process was divided into two parts, using the Channel Division based RRH Covering Algorithm to arrange the RRHs in specific channels and using Action Reward Optimization Deep Q-learning Algorithm to decide the allocation policy of bandwidth for users.

5.1 Resource Allocation Efficiency in Coverage Problem

In this part, we conduct the simulations for resource allocation in coverage problems to highlight the performance of our proposed MSIO algorithm. The other two comparison algorithms are considered as follows: convention algorithm represents that the block with high requirement importance should be covered by RRH first, which is a greedy algorithm to pick up the most beneficial block first. The fast networking placement algorithm (FNPA) [29] is an algorithm that covers as many blocks for fast networking in real-time. The core idea of this algorithm is to provide the expansion of radio access in order to serve more users, which leads to better coverage usage, especially in resource-limited environments.

Then, we conduct a platform that there are basically 40 blocks creating communication demands and 150 RRHs serving these blocks. A normal communication range of 50m is used in a square of 1000m. In Fig. 6(a), it shows the coverage importance ratio changes by the communication ranges. The ranges expand from 50 to 150. The overall trend for all three algorithms is the same, which increases with the communication ranges. However, our proposed MSIO is relatively higher due to more careful allocations depending on coverage importance. Similarly, in Fig. 6(b), we add more RRH stations to provide more options for resource allocation. More stations increase the overall coverage importance rate. However, the difference also becomes small because of the overflow of resources and all the blocks cannot be covered due to the limited communication range, which is comparatively practical in a 5G scenario.

Fig. 6(c) shows the coverage importance ratio changes by the extra channels. Since the number of blocks and stations have been predetermined, more channels means the overall coverage importance score increases. If the station resources are limited, the rate decreases due to more uncovered blocks. Fig. 6(d) is the simulation that the influence of pre-defined channels for stations. If more static channels exist, the fewer options of RRHs for resource allocation. It will cause fewer options between the three algorithms, thus the results become closer when the pre-defined ratio increases.

5.2 Resource Allocation Efficiency in Deep Q-Learning

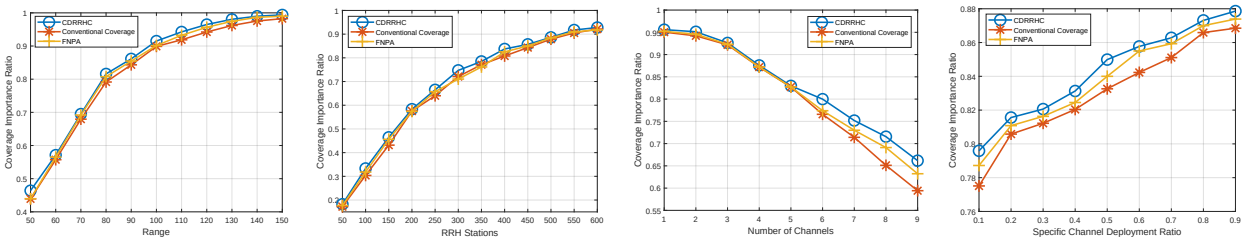
Next, we conduct the DQL based resource allocation simulation using the proposed ARODQ algorithm. The other two comparison algorithms are an average allocation algorithm and a no-gradient algorithm. The basic idea of the average allocation algorithm is that it shares the bandwidth in all blocks based on uniform policy. The no-gradient algorithm is another solution of DQL policy that similar to the water-filling algorithm that balances the bandwidth usage step by step.

In Fig. 7(a), it shows the waiting queue changes by the different number of RRHs. When the number of RRHs is small enough, and cannot satisfy the coming of a new requirement, it leads to the increase of remaining waiting queue, and finally part of the blocks' waiting queue reaches the upper limit. Then, the reward score in Fig. 7(b) shows the same trend in another vision. When the bandwidth resource is not enough, the system performs like the best of its capacity. However, when the bandwidth for the requirement is enough, the policy would arrange all the communication resources to satisfy the requirement in the waiting queue and further handle new demands from blocks.

At last, we test the performance in the same environments using three algorithms. In Fig. 7(c), when the bandwidth is not enough to satisfy all the requirements, every algorithm performs well with the limited bandwidth so that the difference between all the results is not big enough. However, in Fig. 7(d), our proposed result still shows the better result due to the intelligent allocation policy which can well fit the overall service demand in each block.

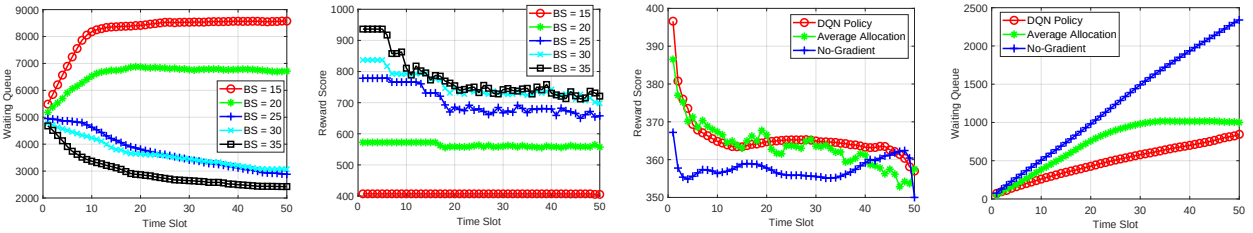
6 Conclusion

In this paper, we propose a deep reinforcement learning-based comprehensive solution for 5G architecture, to solve the problem of quality of service (QoS) using limited wireless communication resources. By ensuring most of the service is responsive, our proposed scheme aims to minimize the redundancy of resource allocation and maximize



(a) Coverge importance ratio changes by communication range (b) Coverge importance ratio changes by RRHs (c) Coverge importance ratio changes by extra channels (d) Coverge importance ratio changes by pre-defined channels

Fig. 6. Coverage performance comparison.



(a) Waiting queue changes by the different requirement ratio (b) Waiting queue changes by the different requirement ratio (c) Reward changes by the time. (d) Waiting Queue changes by the time.

Fig. 7. Allocation performance comparison.

the efficient usage of communication resources in each sliced channel. To overcome the computation complexity, we transfer it into a maximum coverage problem which ensures the quality of each channel for all the blocks. Then, a DQL-based network routing decision-making policy is designed which maximizes the use of bandwidth resources. The simulation results show that our proposed MSIO algorithm improves 3.12% in the performance of the maximum coverage importance problem and the ARODQ algorithm improves 4.05% than other standard solutions. It proves that our proposed scheme can handle such a comprehensive problem and allocate the existing resource sufficiently in a QoE-oriented trade-off scenario.

Acknowledgment

This work is partially supported by JSPS KAKENHI Grant Numbers JP16K00117, JP19K20250, Leading Initiative for Excellent Young Researchers (LEADER), MEXT, Japan and KDDI Foundation. Mianxiong Dong is the corresponding author.

References

[1] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of ran architectures toward 5g mobile communication system," *IEEE Access*, vol. 7, pp. 70371–70421, 2019.

[2] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems," *Journal of Communications and Networks*, vol. 18, pp. 135–149, April 2016.

[3] M. Tao, K. Ota, and M. Dong, "Foud: Integrating fog and cloud for 5g-enabled v2g networks," *IEEE Network*, vol. 31, pp. 8–13, March 2017.

[4] P. Du and A. Nakao, "Deep learning-based application specific ran slicing for mobile networks," in *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*, pp. 1–3, Oct 2018.

[5] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5g heterogeneous cloud radio access networks," *IEEE Network*, vol. 29, pp. 6–14, March 2015.

[6] P. Rost, C. Bernardos, A. De Domenico, M. Di Girolamo, M. Lalam, A. Mader, D. Sabella, and D. Wubben, "Cloud technologies for flexible 5g radio access networks," *IEEE Communications Magazine*, vol. 52, pp. 68–76, May 2014.

[7] M. A. Hossain, R. Md Noor, K. A. Yau, I. Ahmedy, and S. S. Anjum, "A survey on simultaneous wireless information and power transfer with cooperative relay and future challenges," *IEEE Access*, vol. 7, pp. 19166–19198, 2019.

[8] X. Cheng, Y. Wu, G. Min, and A. Y. Zomaya, "Network function virtualization in dynamic networks: A stochastic perspective," *IEEE Journal on Selected Areas in Communications*, vol. 36, pp. 2218–2232, Oct 2018.

[9] K. Han, S. Li, S. Tang, H. Huang, S. Zhao, G. Fu, and Z. Zhu, "Application-driven end-to-end slicing: When wireless network virtualization orchestrates with nfv-based mobile edge computing," *IEEE Access*, vol. 6, pp. 26567–26577, 2018.

[10] T. Wu, P. Yang, H. Dai, W. Xu, and M. Xu, "Charging oriented sensor placement and flexible scheduling in rechargeable wsns," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 73–81, April 2019.

[11] Y. Zuo, Y. Wu, G. Min, and L. Cui, "Learning-based network path planning for traffic engineering," *Future Generation Computer Systems*, vol. 92, pp. 59 – 67, Sept 2019.

[12] Z. Liu, Y. Wu, E. Yuepeng, J. Ge, and T. Li, "Experimental evaluation of consumer mobility on named data networking," in *2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 472–176, July 2014.

[13] I. Sarrigiannis, K. Ramantas, E. Kartsakli, P. Mekikis, A. Antonopoulos, and C. Verikoukis, "Online vnf lifecycle management in a mec-enabled 5g iot architecture," *IEEE Internet of Things Journal*, pp. 1–1, Oct 2019.

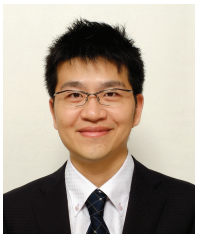
[14] J. Xu, K. Ota, and M. Dong, "Energy efficient hybrid edge caching scheme for tactile internet in 5g," *IEEE Transactions on Green Communications and Networking*, vol. 3, pp. 483–493, June 2019.

[15] S. Park, C. Song, and K. Lee, "Inter-cluster design of wireless fronthaul and access links for the downlink of c-ran," *IEEE Wireless Communications Letters*, vol. 6, pp. 270–273, April 2017.

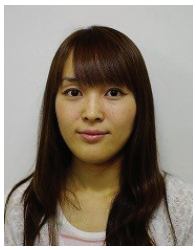
[16] Y. Cai, F. R. Yu, and S. Bu, "Dynamic operations of cloud radio access networks (c-ran) for mobile cloud computing systems,"

IEEE Transactions on Vehicular Technology, vol. 65, pp. 1536–1548, March 2016.
[17] Y. Li, H. Xia, S. Wu, and C. Lu, "Joint optimization of computing and radio resource under outage qos constraint in c-ran," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 107–111, Aug 2017.
[18] I. Farris, T. Taleb, Y. Khettab, and J. Song, "A survey on emerging sdn and nfv security mechanisms for iot systems," *IEEE Communications Surveys Tutorials*, vol. 21, pp. 812–837, Firstquarter 2019.
[19] P. Mekikis, K. Ramantas, L. Sanabria-Russo, J. Serra, A. Antonopoulos, D. Pubill, E. Kartsakli, and C. Verikoukis, "Nfv-enabled experimental platform for 5g tactile internet support in industrial environments," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2019.
[20] M. Labraoui, M. M. Boc, and A. Fladenmuller, "Opportunistic sdn-controlled wireless mesh network for mobile traffic offloading," in *2017 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, pp. 1–7, May 2017.
[21] G. Sun, Z. T. Gebrekidan, G. O. Boateng, D. Ayepah-Mensah, and W. Jiang, "Dynamic reservation and deep reinforcement learning based autonomous resource slicing for virtualized radio access networks," *IEEE Access*, vol. 7, pp. 45758–45772, 2019.
[22] H. Li, K. Ota, and M. Dong, "Eccn: Orchestration of edge-centric computing and content-centric networking in the 5g radio access network," *IEEE Wireless Communications*, vol. 25, pp. 88–93, June 2018.
[23] H. Li, K. Ota, and M. Dong, "Learning iot in edge: Deep learning for the internet of things with edge computing," *IEEE Network*, vol. 32, pp. 96–101, Jan 2018.
[24] G. Sun, G. T. Zemuy, and K. Xiong, "Dynamic reservation and deep reinforcement learning based autonomous resource management for wireless virtual networks," in *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)*, pp. 1–4, Nov 2018.
[25] C. Qi, Y. Hua, R. Li, Z. Zhao, and H. Zhang, "Deep reinforcement learning with discrete normalized advantage functions for resource management in network slicing," *IEEE Communications Letters*, pp. 1–1, 2019.
[26] I. Sarrigiannis, E. Kartsakli, K. Ramantas, A. Antonopoulos, and C. Verikoukis, "Application and network vnf migration in a mec-enabled 5g architecture," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–6, Sep. 2018.
[27] Z. Yan, J. Ge, Y. Wu, H. Zheng, L. Li, and T. Li, "Automatic virtual network embedding based on deep reinforcement learning," pp. 625–631, Aug 2019.
[28] H. Wang, Y. Wu, G. Min, J. Xu, and P. Tang, "Data-driven dynamic resource scheduling for network slicing: A deep reinforcement learning approach," *Information Sciences*, vol. 498, May 2019.
[29] J. Xu, K. Ota, and M. Dong, "Fast networking for disaster recovery," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, Nov 2017.

Chaofeng Zhang received the B.Eng degree in Soochow University, China, in 2011. He also received M.Eng and Ph.D. degree in Muroran Institute of Technology, Japan, in 2016 and 2019, respectively. He is currently an assistant professor in Advanced Institute of Industrial Technology (AIIT), Tokyo, Japan. His research interests include cloud computing, full-duplex communication, wireless positioning technology. He got IEEE VTS Tokyo Chapter 2016 student paper award in 2016, and got the best presentation award in A3 Annual Workshop on Next Generation Internet and Network Security. From March 2017 to April 2017, he was a visiting scholar at Soochow University, China.



Mianxiong Dong Mianxiong Dong received B.S., M.S. and Ph.D. in Computer Science and Engineering from The University of Aizu, Japan. He became the youngest ever Professor of Muroran Institute of Technology, Japan where he currently serves as the Vice President. He was a JSPS Research Fellow with School of Computer Science and Engineering, The University of Aizu, Japan and was a visiting scholar with BBCR group at University of Waterloo, Canada supported by JSPS Excellent Young Researcher Overseas Visit Program from April 2010 to August 2011. Dr. Dong was selected as a Foreigner Research Fellow (a total of 3 recipients all over Japan) by NEC C&C Foundation in 2011. His research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. He has received best paper awards from IEEE HPCC 2008, IEEE ICESS 2008, ICA3PP 2014, GPC 2015, IEEE DASC 2015, IEEE VTC 2016-Fall, FCST 2017, 2017 IET Communications Premium Award and IEEE ComSoc CSIM Best Conference Paper Award 2018. Dr. Dong serves as an Editor for IEEE Transactions on Green Communications and Networking (TGCN), IEEE Communications Surveys and Tutorials, IEEE Network, IEEE Wireless Communications Letters, IEEE Cloud Computing, IEEE Access, as well as a leading guest editor for ACM Transactions on Multimedia Computing, Communications and Applications (TOMM), IEEE Transactions on Emerging Topics in Computing (TETC), IEEE Transactions on Computational Social Systems (TCSS). He has been serving as the Vice Chair of IEEE Communications Society Asia/Pacific Region Information Services Committee and Meetings and Conference Committee, Leading Symposium Chair of IEEE ICC 2019, Student Travel Grants Chair of IEEE GLOBECOM 2019, and Symposium Chair of IEEE GLOBECOM 2016, 2017. He is the recipient of IEEE TCSC Early Career Award 2016, IEEE SCSTC Outstanding Young Researcher Award 2017, The 12th IEEE ComSoc Asia-Pacific Young Researcher Award 2017, Funai Research Award 2018 and NISTEP Researcher 2018 (one of only 11 people in Japan) in recognition of significant contributions in science and technology. He is currently the Member of Board of Governors and Chair of Student Fellowship Committee of IEEE Vehicular Technology Society, and Treasurer of IEEE ComSoc Japan Joint Sections Chapter. He is Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).



Kaoru Ota was born in Aizu-Wakamatsu, Japan. She received M.S. degree in Computer Science from Oklahoma State University, USA in 2008, B.S. and Ph.D. degrees in Computer Science and Engineering from The University of Aizu, Japan in 2006, 2012, respectively. She is currently an Associate Professor with Department of Sciences and Informatics, Muroran Institute of Technology, Japan. From March 2010 to March 2011, she was a visiting scholar at University of Waterloo, Canada. She is an editor of IEEE Transactions on Vehicular Technology (TVT), IEEE Internet of Things Journal, IEEE Communications Letters, Peer-to-Peer Networking and Applications (Springer), Ad Hoc & Sensor Wireless Networks, International Journal of Embedded Systems (Inderscience) and Smart Technologies for Emergency Response & Disaster Management (IGI Global), as well as a guest editor of ACM Transactions on Multimedia Computing, Communications and Applications (leading), IEEE Internet of Things Journal, IEEE Communications Magazine, IEEE Network, IEEE Wireless Communications, IEEE Access, IEICE Transactions on Information and Systems, and Ad Hoc & Sensor Wireless Networks (Old City Publishing). She is the recipient of IEEE TCSC Early Career Award 2017, and The 13th IEEE ComSoc Asia-Pacific Young Researcher Award 2018.She is Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).