Deep Reinforcement Learning Based Joint Edge Resource Management in Maritime Network

Fangmin Xu *, Fan Yang, Chenglin Zhao, Sheng Wu

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
* The corresponding author, email: xufm@bupt.edu.cn

Abstract: Due to the rapid development of the maritime networks, there has been a growing demand for computation-intensive applications which have various energy consumption, transmission bandwidth and computing latency requirements. Mobile edge computing (MEC) can efficiently minimize computational latency by offloading computation tasks by the terrestrial access network. In this work, we introduce a space-air-ground-sea integrated network architecture with edge and cloud computing components to provide flexible hybrid computing service for maritime service. In the integrated network, satellites and unmanned aerial vehicles (UAVs) provide the users with edge computing services and network access. Based on the architecture, the joint communication and computation resource allocation problem is modelled as a complex decision process, and a deep reinforcement learning based solution is designed to solve the complex optimization problem. Finally, numerical results verify that the proposed approach can improve the communication and computing efficiency greatly.

Keywords: maritime network; edge computing; computation offload; computation latency; reinforcement learning; deep learning

I. INTRODUCTION

Recently, maritime communication and maritime network have attracted much attention due to the rapid development of sea financial market. By extending the concept of Internet of Thing (IoT) to the maritime network, and the integration of the fifth generation mobile network(5G), it could provide ubiquitous internet services for emerging maritime applications such as marine environmental monitoring, ocean resource exploration, oceanographic data collection, disaster prevention, and assisted navigation. Therefore, it becomes a new paradigm that attracts lots of attentions in both academic and industrial societies [1].

Future maritime network consists of maritime wireless communication, universal satellite communication and shore-based mobile communication systems, connecting multiple physical things including vessels, underwater sensors, satellites, Unmanned Aerial Vehicles(UAVs), ship-borne Base Stations(BSs) through an integrated network, and enables ubiquitous information communications and novel services [2][3].

Meanwhile, many advanced terrestrial-oriented technologies are introduced in maritime network. However, maritime network has unique characteristics compared with terrestrial network. The marine transmission environment is complicated and unstable under bounded communication distance and vulnerable channel. In addition, explosive demands for large-scale connectivity and com-

Received: Nov. 5, 2019
Revised: Feb. 9, 2020
Editor: Wei Feng

This paper proposed an edge computing based software-defined maritime network architecture.

munications, ultra-low information processing latency, and high reliability in delay-sensitive maritime applications pose challenges for reliable quality of service (QoS) provision in resource-limited maritime network [4][5].

Edge computing that places computing and storage resources at network edges has become an encouraging solution. Therefore, by deploying edge computing servers in maritime network, delay-sensitive tasks could be executed close to vessels to shorten response time and relieve traffic crowding in maritime networks. Due to the various QoS requirements of maritime applications and the dynamic change of computing and communication resource, the allocation of related resource is a challenge task.

In the paper, we studied the issue of joint computation and communication resource management in space-air-ground-sea integrated maritime network, focusing on minimizing the resource utilization and the execution delay of computation tasks. After mathematically modelling and analysis, deep reinforcement learning is utilized to settle the multiple parameters optimization conundrum. This paper has two main contributions which have concluded below.

1) Design a space-air-ground-sea integrated maritime network architecture considering the requirement for edge computing and cloud computing.

2) Apply deep reinforcement learning to develop an intelligent joint resource allocation mechanism for maritime network where the constraints on processing latency can be directly addressed.

The rest of the paper is organised as follows. In Section II, we introduce the background and related working progress. In Section III, the integrated maritime network architecture with the working procedure is discussed to illustrate the intelligent computing scheme. System model and deep reinforcement learning based solution are described in Section IV. In Section V, we conduct numerical simulation and analysis to verify the model correctness. A conclusion is made in Section VI.

II. Background and Related Works

The mainstream internet access technologies used in current maritime network include narrowband maritime radio systems, various satellite systems and on-shore cellular networks. The maritime radio system has limit communication data rate due to bandwidth constraint. Satellite systems could provide worldwide internet access for vessels, but the expensive cost of the satellite terminals and the service is the obstacle for the commercial promotion. Coastal users could benefit from the prosperous development of terrestrial mobile communications, such as Wireless Fidelity (Wi-Fi) and Long-Term Evolution (LTE), even 5G in the future. However, the maximum transmission range of modern terrestrial mobile network is up to 100KM, the vessels users served by on-shore infrastructures are only a small part of all users. Table 1 compared the features of above wireless access technologies in maritime network [6][7].

Because above network technologies have different characteristics, in order to achieve seamless and comprehensive coverage of the

Table I. *Typical wireless technologies for maritime application.*

Technology	Band	Bandwidth	Data rate	Maximum Range	Typical Application
Maritime Radio	VHF(154- 174MHz)/UHF	4.6MHz/Channel	2~10kbps	10- 100KM	Ship-Ship Ship-Shore, Analog
Satellite(Inmarsat-4)	L/C Band	200KHz	4-432kbps	-	Ship-Satellite-CES (Coast Earth Station)
Wi-Fi(802.11n)	2.4/5.8GHz	20MHz	10Mbps	20~50KM(directional antenna)	Shore-Ship
LTE	800,1800,2100,2600MHz	20MHz	10Mbps	50~100KM	Shore-Ship

ocean, the future maritime network should be an integration of these networks. Two fundamental technologies of future networking, Software Defined Network (SDN) and Network Function Virtualization (NFV) facilitate the hybrid network integration.

From the aspect of application, typical applications carried on the maritime network could be classified into the following three types according to the QoS requirements.

- Emergency traffic, which are sensitive to time delay.
- Audio/Video multimedia traffic with strict restriction on time jitter.
- Best effort traffic, such as web browsing and file transfer.

In addition, some new applications require a huge number of computing power, and a lot of data is transmitted from the underlying devices. On one hand, if the application is deployed in the cloud side, it will increase the pressure for the underlying maritime transmission network. On the other hand, more and more maritime applications need fast computational and ultra-reliable response, while long-distance links in maritime applications are difficult to meet this requirement [8].

With the increasing number of connected vessels, users and applications, in an attempt to relieve the pressure on the network, some bandwidth-consuming services can be processed locally or at the edge by the approach of edge computing or local computing, where the raw data do not have to transfer to the cloud centre for further processing.

Generally, vessels have limited computing and storage capacity. By deploying computing and storage resource near the vessels will facilitate the computation-intensive applications, such as video transform, data pre-processing. In future integrated maritime communication networks, vessels are information aggregation nodes which gathering useful data from sensors deployed beneath or floating on the sea surface. On the other hand, the vessel acts as the communications terminals which access the internet service by on-shore BSs, satellite link or UAVs [9][10] depending on the location of the vessels. Figure 1 shows the potential deployment scenario of edge computing in future maritime integrated network.

Recently, with the development of Mobile Edge Computing (MEC) technology, the edge computing schemes based on IoT network and maritime IoT network are proposed. Literature [11] utilized edge computing to enhance the efficiency in the calculation of data contours in marine environmental monitoring IoT system. [12] proposed an open-source fog to cloud platform and the prototype. The related experiments show the feasibility of applying MEC in maritime network.

In [13], a computation offloading scheme based on heuristic algorithm is proposed to realize the minimization of energy consumption of vessel and task execution delay in maritime MEC network. In [14], a cooperative evolution algorithm is introduced to resolve the task scheduling problem in maritime network. [15] discussed a multi-attribute network selection method based on Analytic Hierarchy Process (AHP) and rough set in SDN and fog computing enabled maritime network.

Satellite assistant edge computing is another research direction recently, [16]-[18] discussed the possibility of satellite-based edge computing to promote the service quality of users.

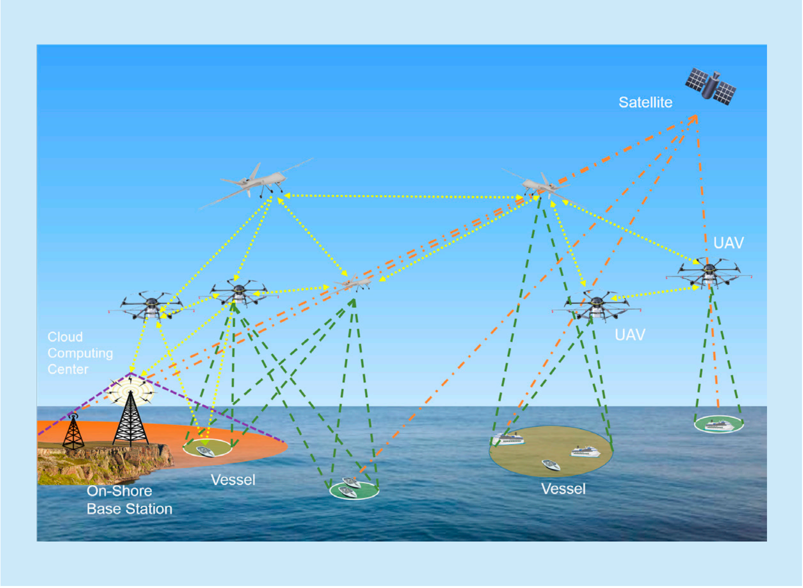


Fig. 1. *Future maritime network with edge computing.*

Existing computation offloading schemes only considers the computing resource status, and the communication resource constraints are not taken into consideration in prior works, but it is the main challenge for maritime network. The status of the distributed and dynamically varying computing and communication resources should be considered jointly when making a computation offloading decision. Thus, the joint resource allocation scheme in maritime network is a complex problem.

Artificial Intelligence(AI) approaches can not only be used in many traditional areas such as image processing, natural language processing, but also in wireless communications fields. Among AI approaches, Reinforcement Learning(RL) is Inspired by operation procedure of human brain, learn how to make optimal decision based on experience which could receive maximum rewards. Therefore, it is suitable for making decision on resource allocation strategy by exploring dynamic unknown environment [19].

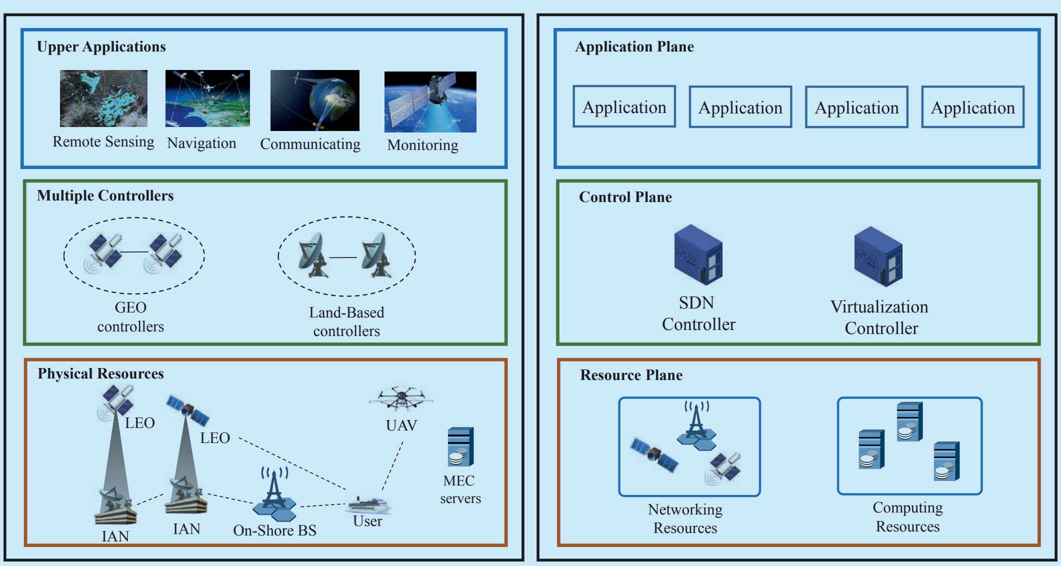
Due to the complexity and high-dimensional characteristic of the joint resource allocation problem. Traditional tabular based RL algorithms is difficult to solve the problem. As one of the RL algorithms, Deep Q-Network(DQN) utilize deep neural networks to estimate the approximate value of tabular items which brings better performance compared with traditional tabular based RL algorithms. And DQN has been deployed in solving the joint resource allocation problem with high environment complexity in various scenarios, such as vehicle network, satellite-to-ground integrated network [20].

Inspired by this idea, this paper discusses the problem of computing resource management in future maritime-oriented integrated networks using DQN-based algorithm to provide better QoS guarantee.

III. Network Architecture and Basic Model

3.1 Network architecture

As described in Section II, the future integrated maritime network has heterogeneous transmission resources and various characteristics. Meanwhile, the computing resources may be located in edge nodes close to the vessels and cloud nodes. Typical edge nodes include UAV,



(a) Physical architecture. (b) Logic architecture.

Fig 2. *The framework of software-defined maritime network.*

LEO (Low Earth Orbit) satellites.

Considering the transmission distance and bandwidth limitations of marine communication, moving vessels are unlikely to acquire the computation service from the cloud centre directly. Nevertheless, the computing capacities of vessels are relatively insufficient to support the complex tasks. Therefore, the vessels have to offload the computation task to nearby edge nodes.

To manage the heterogeneous communication and computing resources effectively and make the integrated network more suitable for future ocean-based application scenarios, SDN-based architecture is utilized in the integrated network to provide centralized flexible and dynamic resource management. Figure 2 shows the architecture of future SDN based maritime network system. The architecture consists of resource plane, control plane and application plane which provides specific applications for system users.

In the resource plane, the network infrastructures (including the data transfer devices and computation processing devices) mainly provide communicating resources and computing resources. Communications resource devices include LEO satellites and UAVs which act as the data transfer pipes between vessels and application servers. Computing resource is provided by near-by MEC servers and far-off cloud servers. Actually the MEC servers are integrated into the on-shore BSs, LEO satellites and UAVs. Since most of the maritime application scenarios are focus on ocean voyages far away from the seashore, only UAV or LEO satellite-assistant edge computing is considered in the paper.

Control plane is generally like the brain of the intelligent network. The designed control policies are embedded in controllers, which may be located in Geostationary Earth Orbit (GEO) satellites and land-based controllers. They are responsible for flexibly allocation of the LEO/UAV communication channels and the computation resource in MEC servers for vessels which request communication and computation resource. The core algorithms in controllers make the communication and computation decisions considering both the scare wireless communication resources and the computation status in MEC servers. The decisions include the choice of UAV/LEO satellites which transfer the data of requesting vessel, and the choice of MEC servers which help the specific vessel to implement the complex computing task based on the collect data. A number of maritime applications are supplied in the upper application plane, such as navigation assistance, ocean environment monitoring.

Based on the proposed architecture, the general work flow is composed of following steps. Firstly, a vessel initiates a specific task request to the domain controller (located in GEO satellite or land). The requested tasks have to occupy or rent certain number of network computation and communication resources. For example, the task of recognize an objects (such as vessels or typhoon) in the media of a certain region may require the transmission of video/image content by the access network and the operation of image recognition algorithm in MEC server. Controller runs the algorithm given in Section IV and transmits the decision back to the user, the decision composes of the choice of access points (APs, including UAVs and LEO satellites) that the user accesses to transmit the data and the choice of MEC server that the user could offload the computation task. After that, user set up the connection with the selected AP based

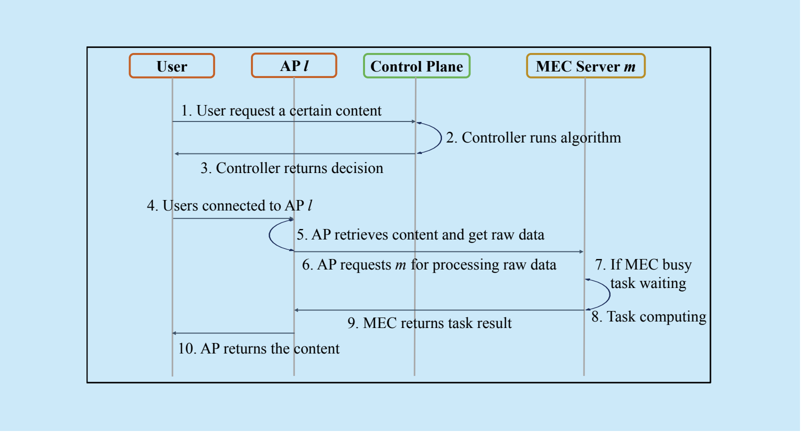


Fig. 3. *Workflow Diagram of computing tasks.*

on the received decision. Access point receives the necessary data and transfer it to the MEC server indicated in the decision. The computation task will be implemented immediately or be placed in a waiting queue depending on the busy/idle state of the MEC server. After the execute of the task, MEC server sends the computation result to the AP, and then the result will be relayed to the user. The related workflow is visualized in Figure 3.

There are two kinds of central controllers (GEO satellites and land-based controllers) managed resources in the architecture where the communication resources are located in LEO satellites and UAVs, and the computation resources are distributed in edge computing servers. Let $\mathcal{U}=\{1,2,...U\}$ denotes the set of vessels (users) in the specific region. And $\mathcal{A}=\{1,2,...A\}$, $\mathcal{M}=\{1,2,...M\}$ is the sets of available APs and MEC severs respectively. Tuple $T_u=\{o_u,n_u\}$ denotes the computation task of user u with data size o_u and required CPU cycle n_u.

3.2 Communication model

The maritime access network consists of several LEO satellites and UAVs in future maritime network. The vessels collect information from deployed sensors and act as data collectors and relays. On the other hand, vessels connected to the LEO satellites and UAVs (APs) with wireless connection. The APs have backhaul that connected to the core network by the relay of GEO/MEO satellites.

Different from general networks, maritime networks access points include LEO satellites, UAVs, and on-shore base stations. And the users are located on vessels, the location of both users, APs and MEC servers are dynamically changing. Moreover, due to the changes of the sea weather and other environmental changes. The wireless channel condition constantly changing sharply. The elements in AP set $\mathcal{A}$ will change during the operation, for example the departure of some UAVs. However, it is assumed that the served AP set will stay unchanged in one resource allocation period, and all APs in the set $\mathcal{A}$ have a possible connection with each user in the set $\mathcal{U}$ with different channel condition. Therefore, the controller will make the decision on choosing appropriate AP to transfer the required data to the MEC server.

The whole spectrum is shared with APs by orthogonal approach (such as Frequency Division Multiplexing). The allocated spectrum bandwidth at AP a is B^a, a proportion of spectrum B_u^a is assigned to connected user u.

The communication channels between users and access points are modelled as finite-state Markov channels (FSMCs). In this model, the channel quality between user u from AP a can be depicted by the received SNR (Signal to Noise Ratio) h_u^a. The discrete SNR value will change according to a certain state transition probability, which represents the change of transmission channel conditions. At time t, the transmission rate of user u is denoted by $v_u^a(t)$. According to Shannon theory,

$$v_u^a(t)=B_u^a(t)\log(1+h_u^a). \tag{1}$$

Considering the choice of AP, the associated communication rate of AP a and user u could be calculated as:

$$\begin{aligned} &ComR_u^a(t)=l_u^a(t)v_u^a(t),\forall u\in\mathcal{U} \\ &s.t.\sum_{u\in\mathcal{U}}ComR_u^a(t)\le Z^a,\forall a\in\mathcal{A}, \end{aligned} \tag{2}$$

where binary $l_u^a(t)$ indicates the decision of whether user u should connects AP a ($l_u^a(t)$=1 represent user u connects certain AP a). And Z^a is the backhaul capacity limitation of AP a.

Assume the SNR h_u^a is not changed in a transmission process. Therefore, the communication latency between user u and AP a equals the data size of the task divided by the communication rate of the chosen AP a, which could be denoted by

$$T_{trans}^u=\frac{o_u}{\sum_{a\in\mathcal{A}}ComR_u^a}. \tag{3}$$

In addition, the Aps and MEC servers are generally deployed in the same physical devices, and the communication delays between MEC servers and APs are neglected.

3.3 Computing model

Assume that the MEC servers are with same processing capability, and each server support multiple thread processing to enable simultaneous computation. The central controller allocates a partition of computation capabilities on MEC server m to AP a.

Since the exactly information on computation capabilities allocated to AP a is difficult to obtain, the computation capabilities of MEC server m assigned for AP a is modelled as finite-state Markov process. The discrete computation capability value Q_a^m will change according to a certain state transition probability, which represents the change of computation workload.

The computation latency T_a^m of computation task from AP a at edge server m can be calculated as

$$T_a^m = \frac{\sum_{u \in \mathcal{U}} l_u^a(t) n_u}{Q_a^m(t)}. \quad (4)$$

The computing rate refers to the speed of computing a task with specific size. So the computing rate could be deduced as:

$$\begin{aligned} CompR_a^m(t) &= d_a^m(t) \frac{\sum_{u \in \mathcal{U}} l_u^a(t) o_u}{T_a^m} \\ &= d_a^m(t) \frac{Q_a^m(t) \sum_{u \in \mathcal{U}} l_u^a(t) o_u}{n_a}, \quad (5) \\ s.t. &\sum_{a \in \mathcal{A}} d_a^m(t) \sum_{u \in \mathcal{U}} l_u^a(t) o_u \le O_m \end{aligned}$$

where binary $d_a^m(t)$ indicates the decision that whether AP a transfers the computation task to MEC server m. ($d_a^m(t) = 1$ denotes the computation task will be handled by MEC server m). O_m is the greatest allowed number of computation tasks that can be parallel executed on MEC server m.

Considering the goal of minimizing the total execute latency (consisting of transmission latency and computation latency) under the communication rate and computation capability constraint (2)(5), the optimization problem of jointly allocating APs and MEC servers strategy could be represented by:

$$Minimize \quad \sum_{u \in \mathcal{U}} T_{trans}^u + l_u^a T_a^m. \quad (6)$$

Above optimization problem is a high-dimensional complicated decision issue. Traditional optimization theory based approaches are difficult to deal with the problem. So artificial intelligent algorithms (such as deep Q-learning) are utilized to learn the inherent rules of the system. In such a way, an intelligent self-learning approach is proposed in next section to resolve the joint computation and communication resource allocation problem in SDN based satellite-terrestrial integrated maritime networks.

IV. PROBLEM FORMULTION

Firstly, the joint resource allocation problem in satellite-terrestrial integrated maritime networks is constructed as a reinforcement learning system. In this section, DQN is introduced briefly, and the operation environment, state space, action space and reward function are defined.

RL algorithms optimize the action choosing behavior by massive offline interaction between agent and environment. In accordance with a predefined policy, the agent selects the optimal action $a_u(t)$ after sensing the environment states. In the next time slot, the environment transits into state $S(t+1)$, meanwhile the agent achieves an immediate reward $R_u(t)$. Traditional RL algorithms are mostly tabular-based, but it is difficult and time-consuming to enumerate the exact value for all possible state-action values $Q(s,a)$ when the dimension of states and actions are relatively large. However, different from traditional tabular based algorithms, DQN utilized deep neural networks to estimate the state-action value. By learning iteration, neural networks update the weight parameters by minimizing the loss $L(w)$. After the iteration update, the neural network is used to approximate real $Q(s,a)$. Moreover, experience replay and fixed target deep networks are two additional features to make DQN more powerful and versatile [21].

4.1 State space

As mentioned above, there are two kinds of resources to manage, so at time slot t, the agent observes the network environment and collects the SNR and computation capability parameters that constitute the system state space. The state space matrix $S(t)$ is therefore defined as:

$$S(t)=\begin{pmatrix} h_u^1(t) & h_u^2(t)\ldots & h_u^A(t) \\ Q_a^1(t) & Q_a^2(t)\ldots & Q_a^M(t) \end{pmatrix}, \tag{7}$$

where the first row of state space is the channel quality state, the second row is the computation capability state.

4.2 Action space

In the learning system, the controller (agent role in reinforcement learning) needs to decide which access point should be selected to provide network access service for the user u, and which MEC server should be assigned to the user u to provide computation assistance service. Let $a_u(t)$ denote the action space at time t which consists of following:

$$a_u(t)=\left\{\mathrm{A}_u^{com}(t),\ \mathrm{A}_u^{comp}(t)\right\} \tag{8}$$

where $ComR_u(t)$ and $CompA_a(t)$ means:

- $ComR_u(t)=[ComR_u^1(t),\ldots ComR_u^A(t)]$, where $ComR_u^a(t), \forall a\in\mathcal{A}$ means whether or not AP a provide network access to user u. $ComR_u^a(t)\in\{0,1\}$, where $ComR_u^a(t)$=1 indicates AP a is associated with user u at the slot otherwise $ComR_u^a(t)$=0. It is restricted that only one AP provides network access for user u in one slot.
- $CompA_a(t)=[CompA_a^1(t),\ldots ComR_a^M(t)]$, where $CompA_a^m(t)\in\{0,1\}, \forall m\in\mathcal{M}$ indicates the decision of whether the computation task of AP a is performed at MEC server m. And $CompA_a^m(t)$=1 means computing at MEC server m. There also exists a constraint that only single MEC server will implement the offloaded computation task for AP a at one time slot.

4.3 Reward and policy

In accordance with prior works in [13], the system ought to recompense the utilization of network access service for AP a which is denoted as δ_a(in the unit of \$/Hz). For the computation, the computation usage charge for MEC server m is defined as η_m (in the unit of \$/Cycle). Besides, the system receives the incomes from user u, which includes the costs of accessing the networks and renting computing resources. Similarly the costs are calculated by τ_u and ϕ_u per unit resource, respectively. Therefore, the reward function is defined as:

$$\begin{aligned} R_u(t) &= \sum_{a\in\mathcal{A}} R_{u,a}^{comm}(t)+\sum_{m\in\mathcal{M}} R_{a,m}^{comp}(t) \\ &= \sum_{a\in\mathcal{A}} ComA_u^a(t)(\tau_u ComR_u^a(t)/\delta_a B_u^a(t)) \\ &\quad + \sum_{m\in\mathcal{M}} CompA_a^m(t)(\phi_u CompR_a^m(t)/\eta_m n_a) \\ &= \sum_{a\in\mathcal{A}} ComA_u^a(t)(\tau_u B_u^a(t) v_u^a(t)/\delta_a B_u^a(t)) \\ &\quad + \sum_{m\in\mathcal{M}} CompA_a^m(t)(\phi_u \frac{Q_a^m(t)o_a}{n_a}/\eta_m n_a) \\ &(utility/resource) \end{aligned}, \tag{9}$$

where the immediate reward function $R_u(t)$ is measured by the utility per resource. It is the proportion between the received incomes of providing service and the costs of renting communication and computation resources. Obviously, the larger reward value means the higher resource utilizing efficiency.

After determining the operation environment, state space, action space and reward function, a DQN based Joint Resource Allocation Scheme is proposed for solving the complex joint resource allocation problem. The algorithm is shown in Algorithm 1.

V. NUMERICAL SIMULATION AND ANALYSIS

In this simulation, the performance of the proposed communication and computation resource allocation algorithm in SDN based maritime network architecture is evaluated.

5.1 Simulation parameters

Simulation environment is constructed in a CPU-based server with Python 3.6.4. With the assistance of tensor-flow, a seven-layer full connected deep neural networks is used in the simulation. ReLU activation function is used in 300 neurons at the hidden layer. AdamOptimizer [22] runs at the learning agent.

To simplify the simulation, the SNR of

wireless channels between user u and AP a during a time instance is discretized into the set (0.2, 2, 5, 10, 20). For example, SNR=0.2 means quite poor channel condition. Similarly, the computation capability $Q_a^m(t)$ of MEC servers m are quantized into following state set (1, 5, 10, 20, 50). In practical, the SNR set and computation capacity set could be more complicated with more discretized levels. The state transition probabilities of both wireless link channel quality and MEC computation capability are assumed as fixed unknown values in the simulation.

There are five APs and five MEC servers in the simulation, and U = 10 users within the coverage. The bandwidth of AP a allocated to user u is 5MHz. The data size and required CPU cycle for computation task is 2Mbits and 5Mcycles, respectively. The unit income of using communication resources and computation resources is 10units/Mbps and 10units/MCycles respectively. The unit cost for using communication resources and computation resources are 2units/MHz and 2units/MCycles respectively.

Algorithm 1. DQN based joint resource allocation scheme.

Initialization:
Initialize evaluate deep networks with the set of weights and biases w.
Initialize target deep networks with the set of weights and biases w'.
for $k = 1:\mathrm{K}$ **do**
  Initialize space-air-ground-sea integrated maritime network environment state S_{ini}, and $S(t) = S_{ini}$.
  while $S(t) != S_{terminal}$ **do**
    Select an AP and a MEC server for executing the task based on $\varepsilon - greedy$ policy, the selection is action $a_u(t)$.
    Obtain immediate reward $R_u(t)$and next observation $S(t+1)$.
    Store the experience $(S(t), a_u(t), R_u(t), S(t+1))$ into the experience replay memory.
    Randomly sample some batches of them from the experience replay memory.
    Calculate target Q-value $Q_{target}(t)$ in target deep networks:
    if s' is $s_{terminal}$
      $Q_{target}(t) = R_u(t)$
    else
      $Q_{target}(t) = R_u(t) + \gamma_q \max_{a'} Q(s', a', w')$
    Train evaluated deep networks to minimize loss function $L(w)$.
    Every some steps, update target deep networks.
    $S(t) \leftarrow S(t+1)$
  end while
end for

5.2 Simulation results

This subsection shows the evaluation result of the proposed scheme, and discuss the convergence situation and performance.

Figure 4 shows the correlation between the total expected rewards (i.e, utilization efficiency of resource) with the training episodes in different schemes. The learning rate of training agent is 0.001. As shown in figure, the proposed DRL-based approach can obtain the total resource utilization rewards fluctuates around 3000 utility/resource after 2500 episodes. By contrary, the rewards achieved by the DRL-based approaches only consider communication resource or only consider edge computation capability fluctuates around 1000 or 250. This comparison proves the benefit of joint considering the optimization of communication and computation resource.

To show the effectiveness of deep networks, the training loss during the training process is given in figure 5. Distinct from supervised learning, DQN utilizes two neural networks (target network and evaluated network). In the initiation stage, these two deep networks have the similar parameter sets, thus the difference of output is quite small which results in litter loss. With the training going further, evaluated networks update the parameters from the interaction with environment, which conduce to the larger loss gaps between evaluated network and target network. Afterward, evaluated network are trained to reduce the training loss. Therefore, after 20000 steps the loss will reduce gradually.

Finally, the task implementation latency performance of different schemes is compared which is the most important metric in the scenario. The required CPU cycles for computation task varies from 5 to 9 million cycles, the expected latencies under different schemes which includes local computing scheme are given in Figure 6. Figure 7 shown the laten-

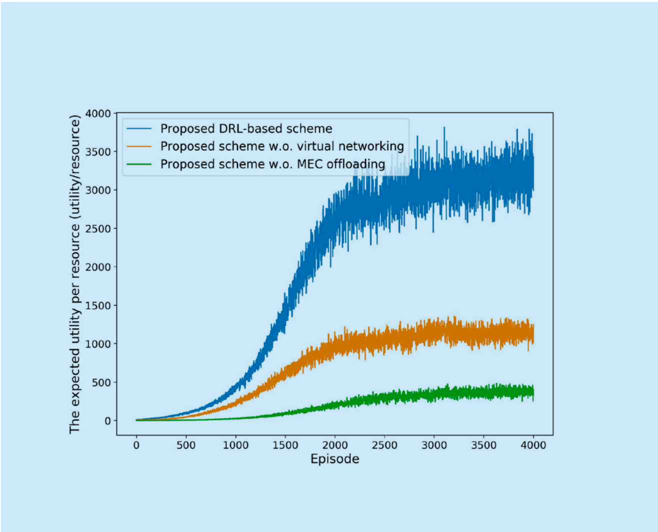


Fig. 4. *The convergence curve of proposed scheme.*

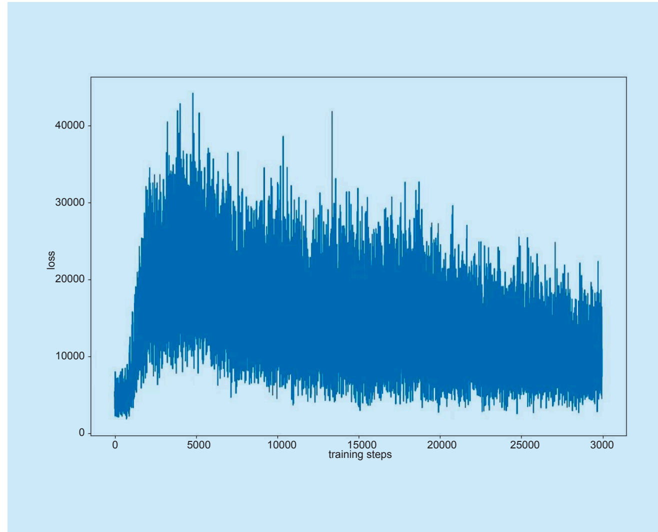


Fig. 5. *Training performance of proposed DQL-based scheme.*

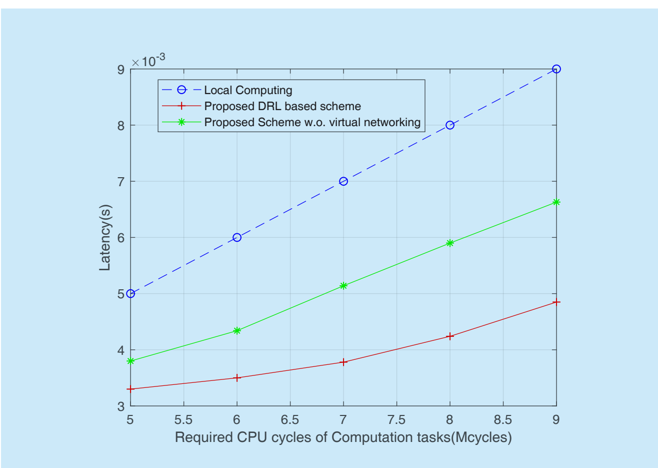


Fig. 6. *The expected latency versus the computation load under different schemes.*

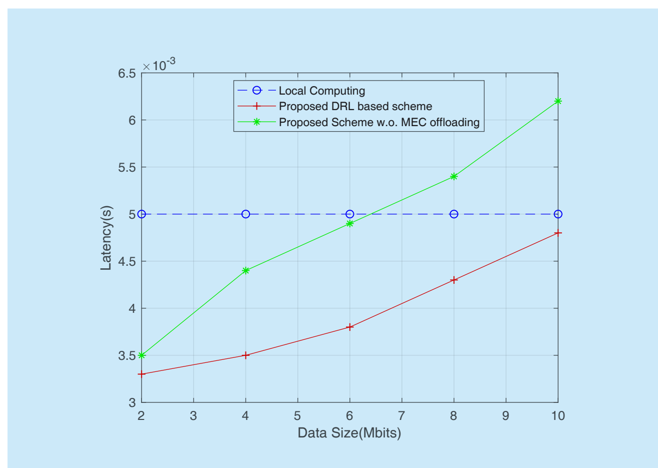


Fig. 7. *The expected latency versus the communication requirement.*

cies with increasing data size. It is apparent that the performance of DQL based scheme is preferable to the other two schemes.

VI. CONCLUSIONS

This paper introduced an edge computing based software-defined maritime network architecture. Firstly, based on the architecture, the joint optimization problem of networking and edge computing resources is devised. Then we adopt a deep reinforcement learning approach to resolve the optimization problem mentioned above. Finally, we gave the performance evaluation of the proposed approach. The results validated the convergence performance and effectiveness of our proposed scheme in a simplified scenario. In future works, we should further consider the mobility scenarios in order to better generalize the proposed approach.

ACKNOWLEDGEMENTS

We gratefully acknowledge anonymous reviewers who read drafts and made many helpful suggestions. This work is supported by the National Natural Science Foundation of China under Grant No. U1805262.

References

[1] Y. Xu, "Quality of service provisions for maritime communications based on cellular networks," *IEEE Access*, vol. 5, 2017, pp. 23881-23890.

[2] T. Wei, W. Feng, J. Wang, N. Ge and J. Lu, "Exploiting the shipping lane information for energy-efficient maritime communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, 2019, pp. 7204-7208.

[3] S. Jiang. "A possible development of marine Internet: A large scale cooperative heterogeneous wireless network." *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, 2015, pp. 481-495.

[4] K. A. Yau, A. R. Syed, W. Hashim, J. Qadir, C. Wu and N. Hassan, "Maritime networking: bringing internet to the sea," *IEEE Access*, vol. 7, 2019, pp. 48236-48255.

[5] T. Xia, M. M. Wang and X. You, "Satellite machine-type communication for maritime internet of things: An interference perspective," *IEEE Access*, vol. 7, 2019, pp. 76404-76415.

[6] S. Jo and W. Shim, "LTE-maritime: High-speed maritime wireless communication based on LTE technology," *IEEE Access*, vol. 7, 2019,pp. 53172-53181.

[7] Y. Kim, Y. Song and S. H. Lim, "Hierarchical maritime radio networks for internet of maritime things," *IEEE Access*, vol. 7, pp. 54218-54227, 2019.

[8] T. Wei, W. Feng, Y. Chen, C.-X. Wang, N. Ge, and J. Lu, "Hybrid satellite-terrestrial communication networks for the maritime Internet of Things: key technologies, opportunities, and challenges," arXiv:1903.11814, Mar. 2019.

[9] W. Feng, J. Wang, Y. Chen, X. Wang, N. Ge and J. Lu, "UAV-aided MIMO communications for 5G internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 2, 2019, pp. 1731-1740.

[10] C. Liu, W. Feng, T. Wei and N. Ge, "Fairness-oriented hybrid precoding for massive MIMO maritime downlink systems with large-scale CSIT," *China Communications*, vol. 15, no. 1, 2018, pp. 52-61.

[11] J. Yang, et al. "Marine depth mapping algorithm based on the edge computing in Internet of things." *Journal of Parallel and Distributed Computing,* vol. 114, 2018, pp. 95-103.

[12] M. Cankar and S. Stanovnik, "Maritime IoT Solutions in fog and cloud," *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, 2018, pp. 284-289.

[13] T. Yang, H. Feng, C. Yang, Y. Wang, J. Dong and M. Xia, "Multivessel computation offloading in maritime mobile edge computing network," *IEEE Internet of Things Journal*, vol. 6, no. 3, 2019, pp. 4063-4073.

[14] T. Yang, Y. Zhang and J. Dong, "Collaborative optimization scheduling of maritime communication for mobile edge architecture," *2018 IEEE International Conference on RFID Technology & Application (RFID-TA)*, 2018, pp. 1-5.

[15] T. Yang, R. Wang, Z. Cui, J. Dong and M. Xia, "Multi-attribute selection of maritime heterogenous networks based on SDN and fog computing architecture," *2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2018, pp. 1-6.

[16] N. Cheng *et al.*, "Space/aerial-assisted computing offloading for iot applications: A learning-based approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 5, 2019, pp. 1117-1129.

[17] Z. Zhang, W. Zhang and F. Tseng, "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques," *IEEE Network*, vol. 33, no. 1, pp. 70-76, January/February 2019.

[18] Y. Wang, J. Zhang, X. Zhang, P. Wang and L. Liu, "A computation offloading strategy in satellite terrestrial networks with double edge computing," *2018 IEEE International Conference on Communication Systems (ICCS)*, 2018, pp. 450-455.

[19] F. Xu, F. Yang, S. Bao and C. Zhao, "DQN inspired joint computing and caching resource allocation approach for software defined information-centric internet of things network," *IEEE Access*, vol. 7, 2019n, pp. 61987-61996

[20] C. Qiu, H. Yao, F. R. Yu, F. Xu and C. Zhao, "Deep Q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, 2019, pp. 5871-5883.

[21] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, 2015, pp. 529.

[22] T. M. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman, "Project adam: Building an efficient and scalable deep learning training system." *OSDI*, vol. 14, 2014, pp. 571–582.

Biographies

Fangmin Xu, received the M.S. and Ph.D. degrees in communication engineering from the Beijing University of Posts and Telecommunication (BUPT), China, in 2003 and 2008, respectively. He is currently an Associate Professor with the School of Information and Communication Engineering, BUPT, China. From 2008 to 2014, he was with Samsung Electronics, where he actively contributed to 3GPP LTE/LTE-A and IEEE 802.16m. He has authored two books, 20 peer-reviewed international research papers, and 50 standard contributions and the Inventor of 15 issued or pending patents among which four have been adopted in the specifications of 4G (3GPP LTE/LTE-A and IEEE 802.16m) standards. His research interests include advance technologies in wireless networks, especially the Internet of Things (IoT) field. Email: xufm@bupt.edu.cn

Fan Yang, received the B.S. degree in e-commerce engineering with law from the Queen Mary University of London, in 2017, and the B.Admin. degree in e-commerce engineering with law from the Beijing University of Post and Telecommunications (BUPT), in 2017. She is currently pursuing the M.S. degree with the Key Laboratory of Universal Wireless Communications, Ministry of Education, BUPT, China. Her research interests include edge computing and blockchain.

Chenglin Zhao, received the bachelor's degree in radio-technology from Tianjin University, in 1986, and the master's degree in circuits and systems and the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications, Beijing, China, in 1993 and 1997, respectively, where he is currently a Professor. His research is focused on emerging technologies of short-range wireless communication, cognitive radios, and industrial internet.

Sheng Wu, received his bachelor's and master's degrees from Beijing University of Posts and Telecommunications in 2004 and 2007, and his Ph.D. degree from Tsinghua university. He is now a lecturer in Beijing University of Posts and Telecommunications. His main research directions include iterative signal processing, millimetre wave communication and space communication.