entropy

MDPI

Article

DRL-Assisted Resource Allocation for NOMA-MEC Offloading with Hybrid SIC

Haodong Li [1,*], Fang Fang [2] and Zhiguo Ding [1]

[1] Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, UK; zhiguo.ding@manchester.ac.uk
[2] Department of Engineering, Durham University, Durham DH1 3LE, UK; fang.fang@durham.ac.uk
* Correspondence: haodong.li@manchester.ac.uk

Abstract: Multi-access edge computing (MEC) and non-orthogonal multiple access (NOMA) are regarded as promising technologies to improve the computation capability and offloading efficiency of mobile devices in the sixth-generation (6G) mobile system. This paper mainly focused on the hybrid NOMA-MEC system, where multiple users were first grouped into pairs, and users in each pair offloaded their tasks simultaneously by NOMA, then a dedicated time duration was scheduled to the more delay-tolerant user for uploading the remaining data by orthogonal multiple access (OMA). For the conventional NOMA uplink transmission, successive interference cancellation (SIC) was applied to decode the superposed signals successively according to the channel state information (CSI) or the quality of service (QoS) requirement. In this work, we integrated the hybrid SIC scheme, which dynamically adapts the SIC decoding order among all NOMA groups. To solve the user grouping problem, a deep reinforcement learning (DRL)-based algorithm was proposed to obtain a close-to-optimal user grouping policy. Moreover, we optimally minimized the offloading energy consumption by obtaining the closed-form solution to the resource allocation problem. Simulation results showed that the proposed algorithm converged fast, and the NOMA-MEC scheme outperformed the existing orthogonal multiple access (OMA) scheme.

Keywords: deep reinforcement learning (DRL); multi-access edge computing (MEC); resource allocation; sixth-generation (6G); user grouping

Citation: Li, H.; Fang, F.; Ding, Z. DRL-Assisted Resource Allocation for NOMA-MEC Offloading with Hybrid SIC. *Entropy* **2021**, *23*, 613. https://doi.org/10.3390/e23050613

Academic Editors: Benjamin M. Zaidel and Ori Shental

Received: 31 March 2021
Accepted: 11 May 2021
Published: 14 May 2021

1. Introduction

With fifth-generation (5G) networks being available now, the sixth-generation (6G) wireless network is currently under research, which is expected to provide superior performance to satisfy the growing demands of mobile equipment, such as latency-sensitive, energy-hungry, and computationally intensive services and applications [1,2]. For example, the Internet of Things (IoT) networks are being developed rapidly, where massive numbers of nodes are supposed to be connected together, and IoT nodes can not only communicate with each other, but also process acquired data [3–5]. However, such IoT and many other terminal devices are constrained by the battery life and computational capability, and thereby, these devices cannot fully support computationally intensive tasks. A conventional approach to improve the computation capability of mobile devices is mobile cloud computing (MCC), where computationally intensive tasks are offloaded to the central cloud servers for data processing [6,7]. However, MCC will cause significant delays due to the long propagation distance between the central server and the user equipment. To address the long transmission delay issue, especially for delay-sensitive applications in the future 6G networks, multi-access edge computing (MEC) has emerged as a decentralized structure to provide the computation capability close to the terminal devices, which is generally implemented at the base stations to provide a cloud-like task processing service [7–10].

From the communication perspective, non-orthogonal multiple access (NOMA) is recognized as a promising technology to improve the spectral efficiency and massive

connectivity, which enable multiple users to utilize the same resource block such as time and frequency for transmission [11,12]. Taking power-domain NOMA as an example, the signals of multiple users are multiplexed in the power domain by superposition coding, and at the receiver side, successive interference cancellation (SIC) is adopted to successively remove the multiple access interference [13]. Hence, integrating NOMA with MEC can potentially improve the service quality of MEC including low transmission latency and massive connections compared to the conventional orthogonal multiple access (OMA).

1.1. Related Works

The integration of NOMA and MEC has been well studied so far, and researchers have proposed various approaches to optimal resource allocation to minimize the offloading delay and energy consumption. In [14], the author minimized the offloading latency for a multi-user scenario, in which the power allocation and task partition ratio were jointly optimized. The partial offloading policy can determine the amount of data to be offloaded to the server, and the remainder is processed locally. The author of [15] proposed an iterative two-user NOMA scheme to minimize the offloading latency, in which two users offload their tasks simultaneously by NOMA. Since one of the users suffers performance degradation introduced by NOMA, instead of forcing two users to complete offloading at the same time, the remaining data are offloaded together with the next user during the following time slot. Moreover, many existing works investigated the energy minimization of NOMA-MEC networks. For example, the joint optimization of central processing unit (CPU) frequency, task partition ratio, and power allocation for a NOMA-MEC heterogeneous network were considered in [16,17]. In [18], the author considered a multi-antenna NOMA-MEC network and presented an approach to minimize the weighted sum energy consumption by jointly optimizing the computation and communication resource.

In addition to the existing works on pure NOMA schemes as previously mentioned, a few works also combined NOMA and OMA together, which is denominated as hybrid NOMA [19]. In this paper, the author proposed a two-user hybrid NOMA scenario, in which one user is less delay tolerant than the other. The two users offload during the first time slot by NOMA, and the user with a longer deadline offloads the remaining data during an additional time duration by OMA. This configuration presents significant benefits and outperforms both OMA and pure NOMA in terms of energy consumption since the energy can be saved for the delay-tolerant user instead of finishing offloading at the same time in pure NOMA networks. In [20,21], the hybrid NOMA scheme was extended to multi-user scenarios, in which a two-to-one matching algorithm was utilized to pair every two users into a group. The whole bandwidth resource was divided into multiple sub-channels, and the users in each group offloaded simultaneously through a dedicated sub-channel.

For the resource allocation in NOMA-MEC networks, user grouping is a non-convex problem, which is solved by exhaustive search or applying matching theory. Moreover, deep reinforcement learning (DRL) is recognized as a novel approach to this problem, which is a powerful tool to solve the real-time decision-making tasks, and only a handful papers have utilized it for user grouping and sub-channel assignment, such as [22,23], which output the user grouping policy for uplink and downlink NOMA networks, respectively.

Moreover, in most of the NOMA works, the SIC decoding order is predetermined either by the channel state information (CSI) or the quality of service (QoS) requirements of the users [24–26]. A recent work [27] proposed a hybrid SIC scheme to switch the SIC decoding order dynamically, which showed significant performance improvement in uplink NOMA networks. The author of [28] integrated the hybrid SIC scheme with an MEC network to serve two uplink users, and the results revealed that the hybrid SIC outperformed the QoS-based decoding order.

1.2. Motivation and Contributions

Motivated by the existing research on MEC-NOMA, in this paper, we investigated the energy minimization for uplink transmission in multi-user hybrid NOMA-MEC networks

with hybrid SIC. More specifically, a DRL-based framework was proposed to generate a user grouping policy, and the power allocation, time allocation, and task partition assignment were jointly optimized for each group. The DRL framework collects experience data including CSI, deadlines, and energy consumption as labeled data to train the neural networks (NNs). In association with the resource allocation, the proposed scheme can dynamically adapt the decoding order to achieve better energy efficiency. The main contributions of this paper are summarized as follows:

- A hybrid NOMA-MEC network was proposed, in which an MEC server is deployed at the base station to serve multiple users. All users are divided into pairs, and each pair is assigned into one sub-channel. The users in each group adopt NOMA transmission with the hybrid SIC scheme in the first time duration, and the user with a longer deadline transmits the remaining data by OMA in the following time duration. We proposed a DRL-assisted user grouping framework with joint power allocation, time scheduling, and task partition assignment to minimize the offloading energy consumption under transmission latency and offloading data amount constraints.
- By assuming that the user grouping policy is given, the energy minimization problem for each group is non-convex due to the multiplications of variables and a 0–1 indicator function, which indicates two cases of decoding orders. The solution to the original problem can be obtained by solving each case separately. A multilevel programming method was proposed, where the energy minimization problem was decomposed into three sub-problems including power allocation, time scheduling, and task partition assignment. By carefully analyzing the convexity and monotonicity of each sub-problem, the solutions to all three sub-problems were obtained optimally in closed-form. The solution to the energy minimization problem for each case can be determined optimally by adaptingthe decisions successively from the lower level to the higher level (i.e., from the optimal task partition assignment to the optimal power allocation). Therefore, the solution to the original problem can be obtained by comparing the numerical results of those two cases and selecting the optimal solution with lower energy consumption.
- A DRL framework for user grouping was designed based on a deep Q-learning algorithm. We provided a training algorithm for the NN to learn the experiences based on the channel condition and delay tolerance of each user during a period of slotted time, and the user grouping policy can be learned gradually at the base station by maximizing the negative of the total offloading energy consumption.
- Simulation results are provided to illustrate the convergence speed and the performance of this user grouping policy by comparing with random user grouping policy. Moreover, compared with the OMA-MEC scheme, our proposed NOME-MEC scheme can achieve superior performance with much lower energy consumption.

1.3. Organizations

The rest of the paper is structured as follows. The system model and the formulated energy minimization problem for our proposed NOMA-MEC scheme are described in Section 2. Section 3 presents the optimal solution to the energy minimization problem. Following that, the DRL-based user-grouping algorithm is introduced in Section 4. Finally, the simulation results of the convergence and average performance for the proposed scheme are shown in Section 5, and Section 6 concludes this paper.

2. System Model and Problem Formulation

2.1. System Model

In this paper, we considered a NOMA-MEC network, where a base station is equipped with an MEC server to serve K resource-constrained users. During one offloading cycle, each user offloads its task to the MEC server and then obtains the results, which are processed at the MEC server. Generally, the data size of the computation outcome is relatively smaller than the offloaded data in practice; thus, the time for downloading the

results can be omitted [18]. Moreover, since the MEC server has a much higher computation capability than the mobile devices, the data processing time at the MEC server can be ignored compared to the offloading time [14]. Therefore, in this work, the total offloading delay was approximated to the time consumption of data uploading to the base station.

We assumed that all K users were divided into Φ groups to transmit signals at different sub-channels, and each group ϕ contained two users such that $K = 2\Phi$. In each group, we denote the user with a shorter deadline by $U_{m,\phi}$ and the user with relevantly longer deadline by $U_{n,\phi}$, which indicates $\tau_{m,\phi} \leq \tau_{n,\phi}$, where $\tau_{i,\phi}$ is the latency requirement of $U_{i,\phi}, \forall i \in \{m, n\}$ in group ϕ. Because $U_{m,\phi}$ has a tighter deadline, it was assumed that the whole duration $\tau_{m,\phi}$ would be used up, which means that the offloading time $t_{m,\phi} = \tau_{m,\phi}$.

In this paper, we adopted the block channel model, which indicates that the channel condition remains static during each time slot. With the small-scale fading, the channel gain of a user in group ϕ can be expressed as:

$$H_{i,\phi} = \tilde{h}_{i,\phi} {d_{i,\phi}}^{-\frac{\alpha}{2}}, \quad \forall i \in \{m, n\}, \forall \phi, \tag{1}$$

where $\tilde{h}_{i,\phi} \sim \mathcal{CN}(0, 1)$ is the Rayleigh fading coefficient, $d_{i,\phi}$ is the distance between $U_{i,\phi}$ and the base station, and α is the pass loss exponent. The channel gain is normalized by the addictive white Gaussian noise (AWGN) power with zero mean and σ^2 variance, which can be written as:

$$h_{i,\phi} = \frac{|H_{i,\phi}|^2}{\sigma^2}, \quad \forall i \in \{m, n\}, \forall \phi. \tag{2}$$

As shown in Figure 1, since those two users have different delay tolerances, it is natural to consider that $U_{n,\phi}$ is unnecessary to finish offloading within $\tau_{m,\phi}$ via NOMA transmission, and potentially to save energy if $U_{n,\phi}$ can utilize the spare time $\tau_{n,\phi} - \tau_{m,\phi}$. Hence, the adopted hybrid NOMA scheme enables $U_{n,\phi}$ to offload part of its data at the same time when $U_{m,\phi}$ offloads its task during $\tau_{m,\phi}$. An additional time duration $t_{r,\phi}$ is scheduled within each time slot to transmit $U_{n,\phi}$'s remaining data. The task transmission for $U_{n,\phi}$ should be completed within $\tau_{n,\phi}$, which means that $t_{r,\phi}$ should satisfy:

$$t_{r,\phi} \leq \tau_{n,\phi} - \tau_{m,\phi}, \forall \phi. \tag{3}$$

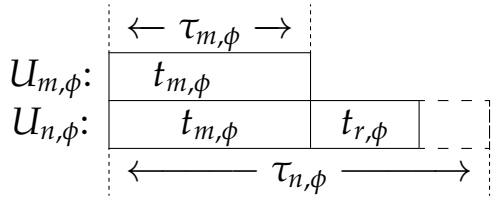

Figure 1. System model.

As previously mentioned, the users in each group will occupy the same sub-channel to upload their data to the base station simultaneously via NOMA. In NOMA uplink transmission, SIC is adopted at the base station to decode the superposed signal. Conventionally, the SIC decoding order is based on either the user's CSI or the QoS requirement [27]. For the QoS-based case, to guarantee $U_{m,\phi}$ can offload its data by $\tau_{m,\phi}$, $U_{n,\phi}$ is set to be decoded first, and the achievable rate should satisfy:

$$R_{n,\phi} \leq B \ln\left(1 + \frac{P_{n,\phi}|h_{n,\phi}|^2}{P_{m,\phi}|h_{m,\phi}|^2 + 1}\right), \tag{4}$$

where B is the bandwidth of each sub-channel. $P_{n,\phi}$ and $P_{m,\phi}$ are the transmission power of $U_{n,\phi}$ and $U_{m,\phi}$ during NOMA transmission, respectively. Based on the NOMA principle, the signal of $U_{m,\phi}$ can then be decoded if (4) is satisfied, and the data rate for $U_{m,\phi}$ can be written as:

$$R_{m,\phi} = B \ln\left(1 + P_{m,\phi}|h_{m,\phi}|^2\right). \tag{5}$$

In contrast, $U_{m,\phi}$ can also be decoded first by treating $U_{n,\phi}$'s signal as interference if the following condition holds:

$$R_{m,\phi} \leq B \ln\left(1 + \frac{P_{m,\phi}|h_{m,\phi}|^2}{P_{n,\phi}|h_{n,\phi}|^2 + 1}\right). \tag{6}$$

Then, the data rate of $U_{n,\phi}$ can be obtained by removing the information of $U_{m,\phi}$, which is:

$$R_{n,\phi} = B \ln\left(1 + P_{n,\phi}|h_{n,\phi}|^2\right). \tag{7}$$

If the same power is allocated to $U_{n,\phi}$ for both decoding sequences, it is evident that the achievable rate for $U_{n,\phi}$ in (7) is higher than that in (4) since the interference is removed by the SIC, and hence the decoding order in (7) is preferred in this case. However, since the constraint (6) cannot always be satisfied, the system has to dynamically change the decoding order accordingly to achieve better performance, which motivated us to utilize the hybrid SIC scheme.

In addition, during $t_{r,\phi}$, $U_{n,\phi}$ adopts OMA transmission, and the data rate can be expressed as:

$$R_{r,\phi} = B \ln\left(1 + P_{r,\phi}|h_{n,\phi}|^2\right), \tag{8}$$

where $P_{r,\phi}$ represents the transmission power of $U_{n,\phi}$ during the second time duration $t_{n,\phi}$.

In this work, the data length of each task is denoted by L, which is assumed to be bitwise independent, and we proposed a partial offloading scheme in which each task can be processed locally and remotely in parallel. An offloading partition assignment coefficient $\beta_\phi \in [0, 1]$ is introduced, which indicates the amount of data offloaded to the MEC server, and the rest can be executed by the local device in parallel. Thus, for each task, the amount of data for offloading to the server is $\beta_\phi L$, and $(1 - \beta_\phi)L$ is the data to be processed locally.

$U_{n,\phi}$ can take the advantage of local computing by executing $(1 - \beta_\phi)L$ data locally during the scheduled NOMA and OMA time duration $t_{m,\phi} + t_{r,\phi}$. Therefore, the energy consumption for $U_{n,\phi}$'s local execution, which is denoted by $E^{loc}_{n,\phi}$, can be expressed as:

$$E^{loc}_{n,\phi} = \frac{\kappa_0 \left[C(1 - \beta_\phi)L\right]^3}{\left(t_{m,\phi} + t_{r,\phi}\right)^2}, \tag{9}$$

where κ_0 denotes the coefficient related to the mobile device's processor and C is the number of CPU cycles required for computing each bit.

The total energy consumed by $U_{n,\phi}$ per task involves three parts, including the energy consumed by local computing and transmission during NOMA and OMA offloading. The power for offloading is scheduled separately during these scheduled two time duration according to the hybrid SIC scheme, and thereby, the offloading energy consumption $E^{off}_{n,\phi}$ can be expressed as:

$$E^{off}_{n,\phi} = t_{m,\phi} P_{n,\phi} + t_{r,\phi} P_{r,\phi}. \tag{10}$$

Hence, the total energy consumption can be expressed as:

$$E^{tot}_{\phi} = E^{loc}_{n,\phi} + E^{off}_{n,\phi}. \tag{11}$$

2.2. Problem Formulation

We assumed that the resource allocation of $U_{m,\phi}$ is given as a constant in each group since $U_{m,\phi}$ is treated as the primary user whose requirements need to be guaranteed in priority, and we only focused on the energy minimization for $U_{n,\phi}$ during both NOMA and OMA duration. Given the user grouping policy, which will be solved in Section 4, the energy minimization problem for each pair can be formulated as:

$$(\mathcal{P}1): \min_{\substack{P_{n,\phi},P_{r,\phi}\\ t_{r,\phi},fi_{\phi}}} \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^2}+\tau_{m,\phi}P_{n,\phi}+t_{r,\phi}P_{r,\phi} \tag{12}$$

$$\text{s.t.} \quad \tau_{m,\phi}R^H_{n,\phi}+t_{r,\phi}B\ln\left(1+P_{r,\phi}|h_{n,\phi}|^2\right)\geq\beta_\phi L \tag{13}$$

$$\tau_{m,\phi}B\ln\left(1+\frac{P_{m,\phi}|h_{m,\phi}|^2}{P_{n,\phi}|h_{n,\phi}|^2+1}\right)\geq \mathbf{1}_{n,\phi}L \tag{14}$$

$$P_{n,\phi}\geq 0, P_{r,\phi}\geq 0 \tag{15}$$

$$0\leq t_{r,\phi}\leq \tau_{n,\phi}-\tau_{m,\phi} \tag{16}$$

$$0\leq \beta_\phi\leq 1, \tag{17}$$

where $R^H_{n,\phi}=\mathbf{1}_{n,\phi}B\ln\left(1+P_{n,\phi}|h_{n,\phi}|^2\right)+(1-\mathbf{1}_{n,\phi})B\ln\left(1+\frac{P_{n,\phi}|h_{n,\phi}|^2}{P_{m,\phi}|h_{m,\phi}|^2+1}\right)$, and $\mathbf{1}_{n,\phi}$ is the indicator function. When $\mathbf{1}_{n,\phi}=1$, $U_{m,\phi}$ is decoded first, and vice verse. The constraint (13) and (14) ensures all the users should complete offloading of the designated amount of data within the given deadline. The constraint (16) limits that the additionally scheduled time slot should not be beyond $U_{n,\phi}$'s delay tolerance. Constraints (15) and (17) set the feasible range of the transmission power and the offloading coefficient.

The problem $(\mathcal{P}1)$ is non-convex due to the multiplication of several variables. Therefore, in the following section, we proposed a multilevel programming algorithm to address the energy minimization problem optimally by obtaining the closed-form solution.

3. Energy Minimization for NOMA-MEC with the Hybrid SIC Scheme

In this section, the problem $(\mathcal{P}1)$ is solved separately for the case $\mathbf{1}_{n,\phi}=1$ and $\mathbf{1}_{n,\phi}=0$. Due to the non-convexity of the original problem for both cases, a multilevel programming method is introduced to decompose the problem $(\mathcal{P}1)$ into three sub-problems, i.e., power allocation, time slot scheduling, and task assignment, which can be solved optimally by obtaining the closed-form solution. By solving those three sub-problems successively, the optimal solutions for both cases can thereby be obtained, which are provided in the subsections below. The solution to the original problem $(\mathcal{P}1)$ can be determined by comparing the numerical result of both cases and choosing the more energy efficient one.

3.1. Power Allocation

Let $t_{r,\phi}$ and β_ϕ be fixed. The problem $(\mathcal{P}1)$ is regarded as a power allocation problem, which can be rewritten as:

$$(\mathcal{P}2): \min_{P_{n,\phi},P_{r,\phi}} \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^2}+\tau_{m,\phi}P_{n,\phi}+t_{r,\phi}P_{r,\phi} \tag{18}$$

$$\text{s.t.} \quad \tau_{m,\phi}R^H_{n,\phi}+t_{r,\phi}B\ln\left(1+P_{r,\phi}|h_{n,\phi}|^2\right)\geq\beta_\phi L \tag{19}$$

$$\tau_{m,\phi}B\ln\left(1+\frac{P_{m,\phi}|h_{m,\phi}|^2}{P_{n,\phi}|h_{n,\phi}|^2+1}\right)\geq \mathbf{1}_{n,\phi}L \tag{20}$$

$$P_{n,\phi}\geq 0, P_{r,\phi}\geq 0 \tag{21}$$

Since there exists an indicator function, $(\mathcal{P}2)$ is solved in two different cases, i.e., when $\mathbf{1}_{n,\phi} = 1$ and when $\mathbf{1}_{n,\phi} = 0$. The following theorem provides the optimal solutions of both cases.

Theorem 1. *The optimal power allocation to* $(\mathcal{P}2)$ *is given by the following two cases according to the indicator function:*

1. *For* $\mathbf{1}_{n,\phi} = 1$, $U_{m,\phi}$ *is decoded first, and the power allocation for this decoding order is presented in the following three offloading scenarios:*

   (a) *When* $P_{n,\phi} \neq 0$ *and* $P_{r,\phi} \neq 0$, $U_{n,\phi}$ *offloads in both time durations, which is termed hybrid NOMA. Given the following two feasible ranges, the optimal power allocation can be expressed as follows:*

   i. *If* $P_{m,\phi} > |h_{m,\phi}|^{-2} e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}} \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)$,

   $$P^*_{n,\phi} = P^*_{r,\phi} = |h_{n,\phi}|^{-2} \left(e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}} - 1 \right). \tag{22}$$

   ii. *If* $|h_{m,\phi}|^{-2} \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right) \leq P_{m,\phi} \leq |h_{m,\phi}|^{-2} e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}} \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)$,

   $$P^*_{n,\phi} = |h_{n,\phi}|^{-2} \left[P_{m,\phi} |h_{m,\phi}|^2 \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)^{-1} - 1 \right], \tag{23}$$

   $$P^*_{r,\phi} = |h_{n,\phi}|^{-2} \left\{ e^{\frac{\beta_\phi L}{Bt_{r,\phi}} - \frac{\tau_{m,\phi}}{t_{r,\phi}} \ln \left[P_{m,\phi} |h_{m,\phi}|^2 \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)^{-1} \right]} - 1 \right\}. \tag{24}$$

   (b) *When* $U_{n,\phi}$ *only offloads during the first time duration* $\tau_{m,\phi}$, *this scheme is termed pure NOMA, and the power allocation is obtained as:*

   if $P_{m,\phi} \geq |h_{m,\phi}|^{-2} e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}} \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)$,

   $$P^*_{n,\phi} = |h_{n,\phi}|^{-2} \left(e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}} - 1 \right). \tag{25}$$

   (c) When $P^*_{n,\phi} = 0$, $U_{n,\phi}$ chooses to offload solely during the section time duration $t_{r,\phi}$, and the optimal power allocation is:

   if $P_{m,\phi} \geq |h_{m,\phi}|^{-2} \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)$,

   $$P^*_{r,\phi} = |h_{n,\phi}|^{-2} \left(e^{\frac{\beta_\phi L}{Bt_{r,\phi}}} - 1 \right). \tag{26}$$

2. *For* $\mathbf{1}_{n,\phi} = 0$, $U_{n,\phi}$ *is decoded first, and similarly, the power allocation for this decoding order is given in three scenarios:*

   (a) *When* $P_{n,\phi} \neq 0$ *and* $P_{r,\phi} \neq 0$, $U_{n,\phi}$, *the hybrid NOMA power allocation is given by:*

   if $P_{m,\phi} \leq |h_{m,\phi}|^{-2} \left(e^{\frac{\beta_\phi L}{Bt_{r,\phi}}} - 1 \right)$,

$$P_{n,\phi}^{*} = |h_{n,\phi}|^{-2}\left(P_{m,\phi}|h_{m,\phi}|^{2}+1\right)\left[e^{\frac{\beta_\phi L - t_{r,\phi} B \ln\left(P_{m,\phi}|h_{m,\phi}|^{2}+1\right)}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}} - 1\right] \tag{27}$$

$$P_{r,\phi}^{*} = |h_{n,\phi}|^{-2}\left[\left(P_{m,\phi}|h_{m,\phi}|^{2}+1\right)e^{\frac{\beta_\phi L - t_{r,\phi} B \ln\left(P_{m,\phi}|h_{m,\phi}|^{2}+1\right)}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}} - 1\right]. \tag{28}$$

(b) *When $P_{r,\phi} = 0$, the pure NOMA case can be obtained as:*

$$P_{n,\phi}^{*} = |h_{n,\phi}|^{-2}\left(P_{m,\phi}|h_{m,\phi}|^{2}+1\right)\left(e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}} - 1\right). \tag{29}$$

(c) *When $P_{n,\phi}^{*} = 0$, the OMA case is the same as (26).*

Proof. Refer to Appendix A. □

Remark 1. *Theorem 1 provides the optimal power allocation for both decoding sequences, i.e., $U_{m,\phi}$ is decoded first when $\mathbf{1}_{n,\phi} = 1$, and $U_{n,\phi}$ is decoded first when $\mathbf{1}_{n,\phi} = 0$. The optimal solution to $(\mathcal{P}1)$ is obtained by a numerical comparison between these two cases in terms of energy consumption. Both cases can be further divided into three offloading scenarios including hybrid NOMA, pure NOMA, and OMA based on different power allocations. For the hybrid NOMA case, $U_{n,\phi}$ transmits during both $\tau_{m,\phi}$ and $t_{r,\phi}$, which indicates $P_{n,\phi} > 0$, $P_{r,\phi} > 0$, and $t_{r,\phi} > 0$. The pure NOMA scheme indicates that $U_{n,\phi}$ only transmits simultaneously with $U_{m,\phi}$ during $\tau_{m,\phi}$, and therefore, $P_{r,\phi} = 0$ and $t_{r,\phi} = 0$. In addition, the OMA case represents that $U_{m,\phi}$ occupies $\tau_{m,\phi}$ solely, and $U_{n,\phi}$ only transmits during $t_{r,\phi}$.*

Remark 2. *Appendix A provides the proof for the case $\mathbf{1}_{n,\phi} = 1$. The proof for the case $\mathbf{1}_{n,\phi} = 0$ similarly, and it can be referred to the previous work in [21]. Thus, the proof for the case $\mathbf{1}_{n,\phi} = 0$ is omitted for this and the following two sub-problems.*

In this subsection, the optimal power allocation for the hybrid NOMA scheme is obtained when $t_{r,\phi}$ is fixed, and then, the optimization of $t_{r,\phi}$ is further studied to minimize $E_{n,\phi}^{tot}$ in the following subsection.

3.2. Time Scheduling

The aim of this subsection is to find the optimal time allocation for the second time duration $t_{r,\phi}$, which is solely utilized by $U_{n,\phi}$ for OMA transmission. As previously mentioned in Theorem 1, the optimal power allocation for the hybrid NOMA scheme is given as a function of $t_{r,\phi}$ and β_ϕ. Hence, by fixing β_ϕ, $(\mathcal{P}1)$ is rewritten as:

$$(\mathcal{P}3): \quad \min_{t_{r,\phi}} \quad \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^{3}}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^{2}} + \tau_{m,\phi}P_{n,\phi}^{*} + t_{r,\phi}P_{r,\phi}^{*} \tag{30}$$

$$\text{s.t.} \quad 0 \le t_{r,\phi} \le \tau_{n,\phi} - \tau_{m,\phi} \tag{31}$$

Proposition 1. *The offloading energy consumption (30) is monotonically decreasing with respected to $t_{r,\phi}$ for both the $\mathbf{1}_{n,\phi} = 1$ and $\mathbf{1}_{n,\phi} = 0$ cases. To minimize the energy consumption, the optimal time allocation is to schedule the entire available time before the deadline $\tau_{n,\phi}$, i.e.,*

$$t_{r,\phi}^{*} = \tau_{n,\phi} - \tau_{m,\phi} \tag{32}$$

Proof. Refer to Appendix B. □

By assuming all the data are offloaded to the MEC server, the following lemma studies the uplink transmission energy efficiency of the two hybrid NOMA-MEC schemes for $\mathbf{1}_{n,\phi} = 0$ and $\mathbf{1}_{n,\phi} = 1$.

Lemma 1. *Assume all data are offloaded to the MEC server, i.e., $\beta_\phi = 1$. The solution in (27) and (28) for the case $\mathbf{1}_{n,\phi} = 0$ has higher energy consumption than the solution in (22) for the case $\mathbf{1}_{n,\phi} = 1$, if* $|h_{m,\phi}|^{-2}\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right) \leq P_{m,\phi} \leq |h_{m,\phi}|^{-2}\left(e^{\frac{L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}} - 1\right)$.

Proof. Without considering local computing, the energy consumption for (22) can be written as:

$$E_1 = \tau_{n,\phi}|h_{n,\phi}|^{-2}\left(e^{\frac{L}{B\tau_{n,\phi}}} - 1\right), \tag{33}$$

and the energy consumption for the power allocation scheme in (27) and (28) is given as:

$$\begin{aligned} E_2 =& \tau_{m,\phi}|h_{n,\phi}|^{-2}\left(P_{m,\phi}|h_{m,\phi}|^2 + 1\right)\left[e^{\frac{L-\left(\tau_{n,\phi}-\tau_{m,\phi}\right)B\ln\left(P_{m,\phi}|h_{m,\phi}|^2+1\right)}{B\tau_{n,\phi}}} - 1\right] \\ &+ \left(\tau_{n,\phi} - \tau_{m,\phi}\right)|h_{n,\phi}|^{-2}\left[\left(P_{m,\phi}|h_{m,\phi}|^2 + 1\right)e^{\frac{L-\left(\tau_{n,\phi}-\tau_{m,\phi}\right)B\ln\left(P_{m,\phi}|h_{m,\phi}|^2+1\right)}{B\tau_{n,\phi}}} - 1\right]. \end{aligned} \tag{34}$$

To prove that $E_2 \geq E_1$, the inequality can be rearranged as:

$$-\tau_{m,\phi}P_{m,\phi}|h_{m,\phi}|^2 + \tau_{n,\phi}e^{\frac{L}{B\tau_{n,\phi}}}\left(P_{m,\phi}|h_{m,\phi}|^2 + 1\right)^{\frac{\tau_{m,\phi}}{\tau_{n,\phi}}} \geq \tau_{n,\phi}e^{\frac{L}{B\tau_{n,\phi}}}. \tag{35}$$

Define $\zeta(x) = -\tau_{m,\phi}x + \tau_{n,\phi}e^{\frac{L}{B\tau_{n,\phi}}}(x+1)^{\frac{\tau_{m,\phi}}{\tau_{n,\phi}}}$. The first-order derivative of $\zeta(x)$ is given as:

$$\zeta'(x) = -\tau_{m,\phi} + \tau_{m,\phi}e^{\frac{L}{B\tau_{n,\phi}}}(x+1)^{\frac{\tau_{m,\phi}}{\tau_{n,\phi}}-1}. \tag{36}$$

Therefore, $\zeta'(x)$ is monotonically decreasing since $\tau_{m,\phi} < \tau_{n,\phi}$, and the following inequality holds:

$$\zeta'(x) \geq \zeta'\left(e^{\frac{L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}} - 1\right) = 0. \tag{37}$$

Hence, for $0 \leq x \leq e^{\frac{L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}} - 1$, $\zeta(x)$ is monotonically increasing, and $\zeta(x) \geq \zeta(0) = \tau_{n,\phi}e^{\frac{L}{B\tau_{n,\phi}}}$, which illustrates that $E_2 \geq E_1$. □

3.3. Offloading Task Partition Assignment

In this subsection, we focused on the optimization of the task partition assignment coefficient for $U_{n.\phi}$ in group ϕ. Given the optimal power allocation and time arrangement, $(\mathcal{P}1)$ is reformulated as:

$$(\mathcal{P}4): \quad \min_{\beta_\phi} \quad \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi} + t^*_{r,\phi}\right)^2} + \tau_{m,\phi}P^*_{n,\phi} + t^*_{r,\phi}P^*_{r,\phi} \tag{38}$$

$$\text{s.t.} \quad 0 \leq \beta_\phi \leq 1, \tag{39}$$

Proposition 2. *The above problem is convex, and the optimal task assignment coefficient can be characterized by those three optimal power allocation schemes for the hybrid NOMA model in* (22), (23), *and* (27), *which is given by:*

$$\beta_{\phi}^{*} = 1 - \frac{2}{z_{2,\phi}} \mathcal{W}\left(\frac{1}{2} z_{1,\phi}^{-\frac{1}{2}} z_{2,\phi} e^{\frac{z_{2,\phi}}{2}} \right), \tag{40}$$

where $\mathcal{W}$ *denotes the single-valued Lambert W function and* $z_{1,\phi}$ *and* $z_{2,\phi}$ *are determined by the different power allocation schemes, which are presented as follows:*

(a) $\mathbf{1}_{n,\phi} = 1$:
If (22) *is adopted:*

$$\begin{cases} z_1 = \frac{3\kappa_0 B C^3 L^2 |h_{n,\phi}|^2}{\tau_{n,\phi}^2}, \\ z_2 = \frac{L}{B\tau_{n,\phi}} \end{cases} \tag{41}$$

If (23) *is adopted:*

$$\begin{cases} z_1 = \frac{3\kappa_0 B |h_{n,\phi}|^2 C^3 L^2 e^{2u_\phi}}{\tau_{n,\phi}^2} \\ z_2 = \frac{L}{B\left(\tau_{n,\phi} - \tau_{m,\phi}\right)} \end{cases} \tag{42}$$

where $u_\phi = \frac{\tau_{m,\phi}}{\left(\tau_{n,\phi} - \tau_{m,\phi}\right)} \ln \left[P_{m,\phi} |h_{m,\phi}|^2 \left(e^{\frac{L}{B\tau_{m,\phi}}} - 1 \right)^{-1} \right]$.

(b) $\mathbf{1}_{n,\phi} = 0$:

$$\begin{cases} z_{1,\phi} = \frac{3\kappa_0 B C^3 L^2 |h_{n,\phi}|^2 e^{\frac{\left(\tau_{n,\phi} - \tau_{m,\phi}\right) \ln\left(P_{m,\phi} |h_{m,\phi}|^2 + 1\right)}{\tau_{n,\phi}}}}{{\tau_{n,\phi}}^2 \left(P_{m,\phi} |h_{m,\phi}|^2 + 1\right)} \\ z_{2,\phi} = \frac{L}{B\tau_{n,\phi}} \end{cases} \tag{43}$$

Proof. Refer to Appendix C. □

Remark 3. *Problem* ($\mathcal{P}4$) *is the lowest level of the proposed multilevel programming method, which provides three task assignment solutions corresponding to the three power allocation schemes* (22), (23), *and* (27), *respectively. The final solution to the energy minimization problem* ($\mathcal{P}1$) *can be obtained by substituting the optimal task assignment into the corresponding power allocation scheme. Then, the most energy-efficient scheme is selected among* (22), (23), *and* (27) *by comparing the numerical energy consumption for each scheme.*

4. Deep Reinforcement Learning Framework for User Grouping

In the previous section, it was assumed that the user grouping is given, and the optimal resource allocation was obtained in closed-form. The user grouping can be obtained optimally by exploring all possible user grouping combinations and finding the one with the lowest energy consumption. Although this method provides the optimal user pairing scheme, the complexity of the exhaustive search method is high, and it is not possible to output real-time decisions. Therefore, we proposed a fast converging user pairing training algorithm based on DQN to obtain the user grouping policy, which is introduced in the following subsection, in which the state space, action space, and reward function are defined. Subsequently, the training algorithm for the user grouping policy is provided.

4.1. The DRL Framework

The optimization of user grouping is modeled as a DRL task, where the base station is treated as the agent to interact with the environment, which is defined as the MEC network. In each time slot t, the agent takes an action a_t from the action space $\mathcal{A}$ to assign users into pairs according to an optimal policy, which is learned by the DNN. The action taken under current state s_t results in an immediate reward r_t, which is obtained at the beginning of the next time slot, and then moved to the next state s_{t+1}. In this problem, the aforementioned terms are defined as follows.

(1) *State space:* The state $s_t \in \mathcal{S}$ is characterized by the current channel gains and offloading deadlines of all users since the user grouping is mainly determined by those two factors. Therefore, the state s_t can be expressed as:

$$s_t = \{h_1[t], h_2[t], ..., h_k[t], ..., h_K[t]; \tau_1[t], \tau_2[t], ..., \tau_k[t], ..., \tau_K[t]\}. \tag{44}$$

(2) *Action space:* At each time slot t, the agent takes an action $a_t \in \mathcal{A}$, which contains all the possible user grouping decisions $j_{k,\phi}$. The action is defined as:

$$a_t = \{j_{1,1}[t], ...j_{k,\phi}[t], ...j_{K,\Phi}[t]\}, \tag{45}$$

where $j_{k,\phi} = 1$ indicates that U_k is assigned to group ϕ. In our proposed scheme, two different users can only be assigned to each group.

(3) *Rewards:* The immediate reward r_t is described by the sum of the energy consumption of each group after choosing the action a_t under state s_t. The numerical result of the energy consumption in each group can be obtained by solving the problem $(\mathcal{P}1)$. Therefore, the reward is defined as:

$$r_t = -\sum_{\phi=1}^{\Phi} E_{\phi}^{tot}[t] \tag{46}$$

The aim of the agent is to find an optimal policy that maximizes the long-term discounted reward, which can be written as:

$$\begin{aligned} R_t &= r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ... \\ &= \sum_{i=0}^{\infty} \gamma^i r_{t+i}, \end{aligned} \tag{47}$$

where $\gamma \in [0, 1]$ is the discount factor, which balances the immediate reward and the long-term reward.

4.2. DQN-Based NOMA User Grouping Algorithm

To accommodate the reward maximization problem, a DQN-based user-grouping algorithm was proposed in this paper, which is illustrated in Figure 2. In conventional Q-learning, the Q-table is obtained to describe the quality of an action for a given state, and the agent chooses actions according to the Q-values to maximize the reward. However, it will be slow for the system to obtain Q-values for all the state–action pairs if the state space and action space are large. Therefore, to speed up the learning process, instead of generating and processing all possible Q-values, DNNs are introduced to estimate the Q-values based on the weight of DNNs. We utilized a DNN to estimate the Q-value denoted by Q-network, for which the Q-estimation is represented as $Q(s_t, a_t; \theta)$, and an additional DNN with the same setting to generate the target network with $Q(s_t, a_t; \theta^-)$ for training, where θ and θ^- are the weights of the DNNs.

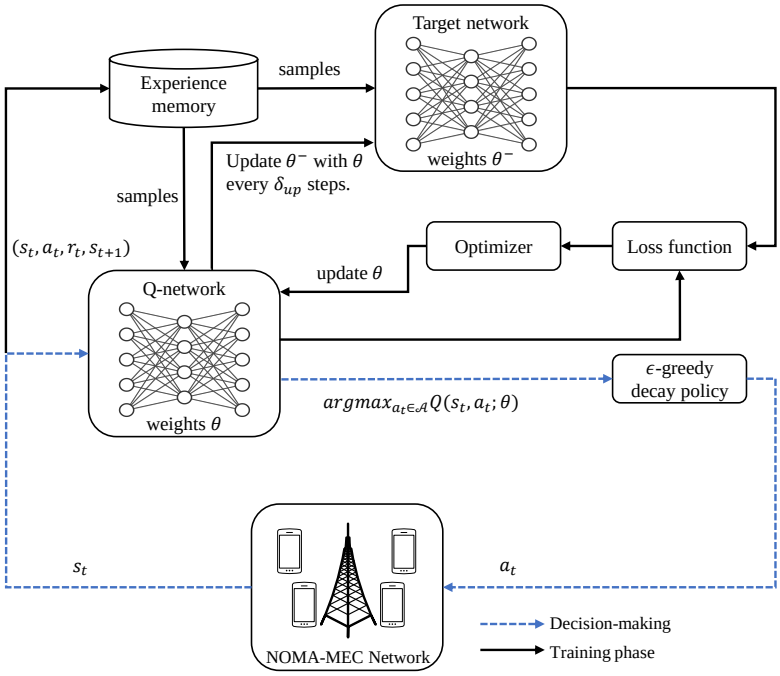

Figure 2. A demonstration of the proposed DQN-based user grouping scheme in the NOMA-MEC network.

We adopted the ϵ-greedy policy with $0 < \epsilon < 1$ to balance the exploration of new actions and the exploitation of known actions by either randomly choosing an action $a_t \in \mathcal{A}$ with probability ϵ to avoid the agent sticking to non-optimal actions or picking the best action with the probability $1 - \epsilon$ such that [29]:

$$a_t = \arg\max_{a_t \in \mathcal{A}} Q(s_t, a_t; \theta). \tag{48}$$

Generally, the threshold ϵ is fixed, which indicates that the probability of choosing a random action remains the same throughout the whole training period. However, this brings a fluctuation when the algorithm converges and may lead to divergence again in extreme cases. In this paper, we adopted an ϵ-greedy decay scheme, for which a large ϵ^+ (greedier) is given at the beginning, and then, it decays with each training step until a certain small probability ϵ^-. The above policy encourages the agent to explore the never-selected actions at the beginning, and then, the agent intends to take more reward-guaranteed actions when the network has already converged.

The target network only updates every certain iteration, which provides a relatively stable label for the estimation network. The agent stores the tuples (s_t, a_t, r_t, s_{t+1}) as experiences to a memory buffer $\mathcal{R}$, and a mini-batch of samples from the memory is fed into the target network to generate the Q-values labels, which is given by:

$$y_i = r_i + \max_{a_{i+1} \in \mathcal{A}} Q(s_{i+1}, a_{i+1}; \theta^-), \quad \forall i \in \mathcal{R} \tag{49}$$

Hence, the loss function for the Q-network can be expressed as:

$$Loss(\theta) = (y_i - Q(s_i, a_i; \theta)), \quad \forall i \in \mathcal{R} \tag{50}$$

The Q-network can be trained by minimizing the loss function to obtain the new θ, and the weights of the target network are updated after δ_{up} steps by replacing θ^- with θ. The whole DQN-based user grouping framework is summarized in Algorithm 1.

Algorithm 1 DQN-based user-grouping algorithm.

Parameter initialization:
Initialize Q-network $Q(s_i, a_i; \theta)$ and target network $Q(s_i, a_i; \theta^-)$.
Initialize reply memory $\mathcal{R}$ with size $|\mathcal{R}|$, and memory counter.
Initialize γ, ϵ^+, ϵ^-, decay step, batch size, target network update interval δ_{up}.
Training Phase:
for *episode* $= 1, 2, ..., N_{ep}$ **do**
  for *time step* $= 1, 2, ..., N_{ts}$ **do**
    Input state s_t into Q-network, and obtain Q-values for all actions.
    Generate a standard uniform distributed random number $\chi \sim \mathcal{U}(0, 1)$
    if $\chi > \gamma$ **then**
      $a_t \leftarrow \arg\max_{a_t \in \mathcal{A}} Q(s_t, a_t; \theta)$
    else
      Randomly select and action a_t.
    end if
    $r_t \leftarrow -\sum_{\phi=1}^{\Phi} E_{\phi}^{tot}[t]$ the observation to next state s_{t+1}.
    Store the experience tuple (s_t, a_t, r_t, s_{t+1}) into the memory $\mathcal{R}$.
    if memory counter $> |\mathcal{R}|$ **then**
      Remove the old experiences from the beginning.
    end if
    Randomly sample a mini-batch of the experience tuples (s_t, a_t, r_t, s_{t+1}) with batch size, and feed into the DNNs.
    Update the Q-network weights θ by calculating the loss function (50).
    Update target network weight $\theta^- \leftarrow \theta$ every δ_{up} steps.
  end for
end for

5. Simulation Results

In this section, several simulation results are presented to evaluate the convergence and effectiveness of the proposed joint resource allocation and user grouping scheme. Specifically, the impact of the learning rate, user number, offloading data length, and delay tolerance is investigated. Moreover, the proposed hybrid SIC scheme is compared to some benchmarks including the QoS-based SIC scheme and other NOMA and OMA schemes.

The system parameters were set up as follows. All users were distributed uniformly and randomly in a disc-shaped cell where the base station was located in the cell center. The total number of users was six, and each of them had a task containing 2 Mbit of data for offloading. As previously mentioned, the delay sensitive primary user $U_{m,\phi}$ was allocated to a predefined power, which was $P_{m,\phi} = 0.5$ W for all groups in the simulation. The maximum delay tolerance for each user was 0.25 s. In addition, the rest of the system parameters are listed in Table 1.

Table 1. System parameters.

Effective capacitance coefficient	10^{-28}
Number of CPU cycles required per bit	10^3
Transmission bandwidth B	2 MHz
Path loss exponent α	3.76
Noise spectral density N_0	-174 dBm/Hz
Maximum cell radius	1000 m
Minimum distance to the base station	50 m

To implement the DQN algorithm, the two DNNs were configured with the same settings, where each of them consisted of four fully connected layers, two of which were

hidden layers with 200 and 100 neurons, respectively. The activation function we adopted for all hidden layers was the rectified linear unit (ReLU), i.e., $f(x) = \max(0, x)$, and the final output layer was activated by tanh, for which the range was $(-1, 1)$ [30]. The adaptive moment estimation optimizer (Adam) method was used to learn the DNN weight θ with the given learning rate [31]. The rest of the hyperparameters are listed in Table 2. All simulation results were obtained with PyTorch 1.70 and CUDA 11.1 on the Python 3.8 platform.

Table 2. Hyperparameters.

ϵ-greedy coefficient	0.5–0.01
ϵ-greedy decay steps	2000
Discount factor γ	0.7
Reply memory size $\mathcal{R}$	20,000
Batch size	64
Target network update interval δ_{up}	10
Number of episode N_{ep}	100
Number of time steps N_{ts}	500

5.1. Convergence of the Framework

In this part, we evaluated the convergence of the proposed DQN-based user-pairing algorithm. Figure 3 compares the convergence rate of the average reward for each episode under different learning rates, which was described by the average energy consumption. The learning rate controls how much the weights of a DNN based on the network loss should be adjusted, and we set the learning rate $= [0.1, 0.01, 0.001]$ to observe its influence on the convergence. The network with a 0.1 learning rate converged slightly faster than the one with a 0.01 learning rate, and both of them converged much faster than the network with a 0.001 learning rate. However, when the learning rate was 0.1, even though the higher learning rate had a better convergence, it overshot the minimum and therefore had higher energy consumption after convergence than the other two plots. Moreover, if the learning rate were too low, the network converged slower because it took more episodes to improve the loss function. Therefore, the most suitable learning rate for our proposed DQN network was 0.01, which was adopted to obtain the rest of the simulation results in this paper.

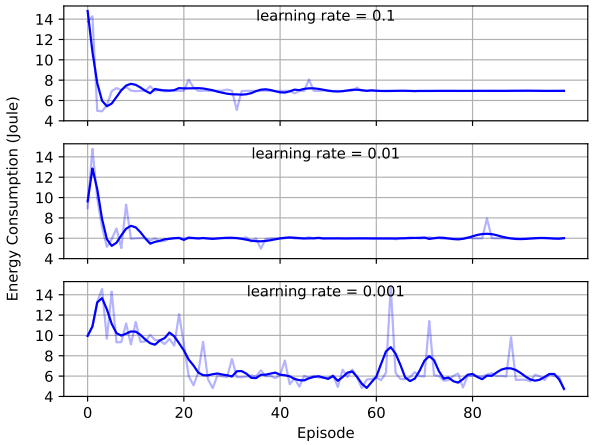

Figure 3. Average energy consumption versus training episodes with different learning rates.

Figure 4 illustrates the effectiveness of the DQN user-grouping algorithm proposed in this paper. By setting the numbers of users to $[6, 8, 10]$, the algorithm showed a similar

performance that the average energy consumption decreased over training. Although the performance may be worse than the random scheme at the beginning of the training, which was due to the random actions and unstable NN weights, it converged within the first 20 episodes for all three cases. Moreover, more users in the network can result in higher energy consumption, and the algorithm showed superior performance over the random policy, which reduced the energy consumption significantly.

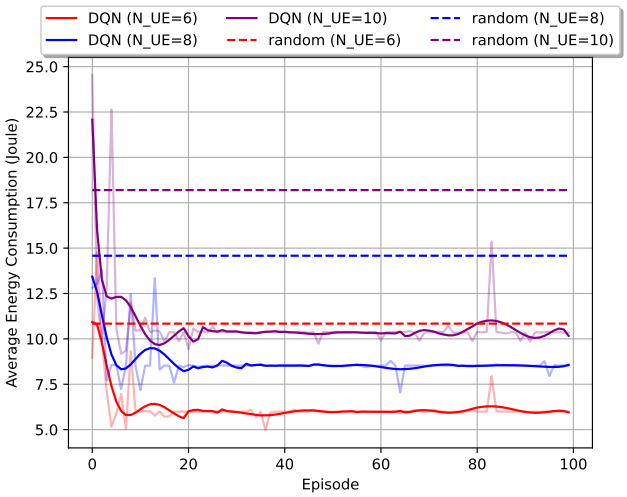

Figure 4. Average energy consumption versus training episodes with different numbers of users.

The application of the ϵ-greedy decay policy to the convergence performance is further investigated in Figure 5. The ϵ-greedy coefficient for the blue curve was set to 0.1, which indicated that the probability of the NN to choose a random action was 0.1, and the probability of choosing the action based on (48) was 0.9. The red curve adopted the ϵ-greedy decay policy with the parameters in Table 2. Since the decay policy started with large ϵ, the network was more likely to choose the random action at the beginning, and hence, the energy consumption was higher at the beginning. With ϵ decaying over the episodes, the network chose the actions that were selected before that guaranteed large rewards, and therefore, it was more stable afterwards. Meanwhile, the network without the decay policy had significant fluctuations during training. It had more of a chance to choose the random actions throughout the training, even when the NN had already converged, which may lead to the NN becoming divergent again. However, if a very small ϵ were adopted, the network would be less likely to explore some actions, which may result in being stuck in non-optimal actions.

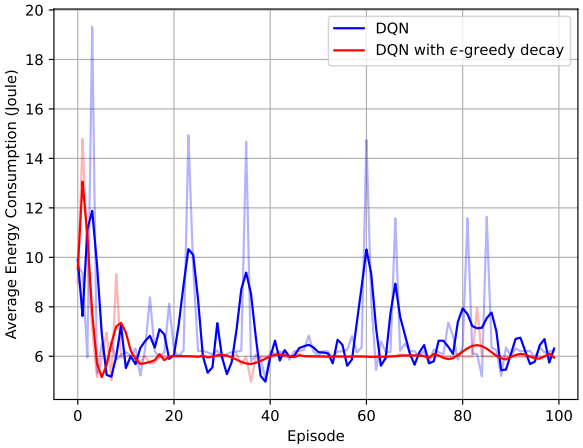

Figure 5. Average energy consumption versus training episodes.

5.2. Average Performance of the Proposed Scheme

In this part, we present the average performance of the proposed NOMA-MEC scheme to show the impact of $P_{m,\phi}$, the offloading data length, and the maximum delay tolerance. Meanwhile, our proposed scheme is compared with the one without task assignment and OMA offloading to show the superior performance. As shown in Figure 6, the energy consumption of both hybrid-SIC schemes rose and then decreased as $P_{m,\phi}$ increased. Since $P_{m,\phi}$ was relatively small at the beginning, $U_{m,\phi}$ was not likely to be decoded first to satisfy the constraint (14) in the case $\mathbf{1}_{n,\phi} = 1$. Therefore, $U_{n,\phi}$ was more likely to be decoded with priority, and increasing $P_{m,\phi}$ caused more interference to $U_{n,\phi}$ according to (4). With $P_{m,\phi}$ continuing to increase, the power allocation schemes in (22) and (23) became feasible, and more groups in the system could adopt different decoding sequences where $U_{m,\phi}$ was decoded first. Then, the energy consumption decreased with the increase of $P_{m,\phi}$, which verified Lemma 1. Moreover, the hybrid-SIC scheme with task assignment outperformed the one without task assignment, shown with the blue line. The one with task assignment had a wider lower bound of the feasible range of power allocation for case $\mathbf{1}_{n,\phi} = 1$ in (22), which means that it could adopt the $\mathbf{1}_{n,\phi} = 1$ case with smaller $P_{m,\phi}$. In addition, both hybrid SIC schemes had lower energy consumption than the OMA scheme.

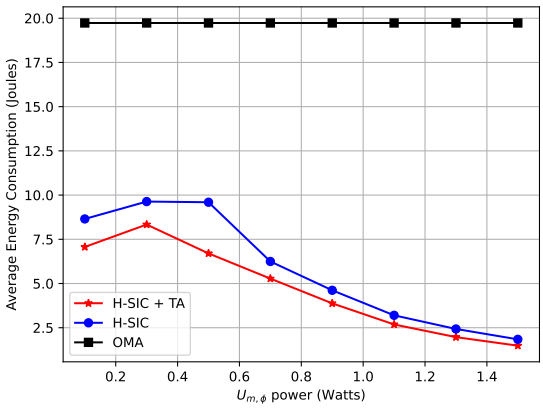

Figure 6. Average energy consumption versus $U_{m,\phi}$'s power.

In Figure 7, the energy consumption is presented as a function of the offloading data length. As the data length increased, the average energy consumption also grew. Our proposed hybrid-SIC scheme reduced the energy consumption significantly especially when the data length was large. Moreover, Figure 8 reveals the energy consumption comparisons versus the maximum delay tolerance for $U_{n,\phi}$. With tighter deadlines, the energy consumption of the hybrid-SIC scheme was much lower than the OMA scheme, and a greater portion of the data was processed locally to save energy compared to the fully offloaded curve.

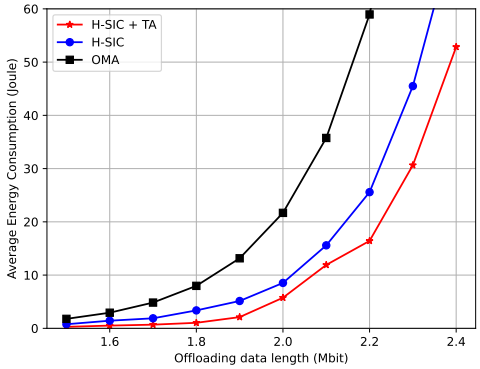

Figure 7. Average energy consumption versus the offloading data length.

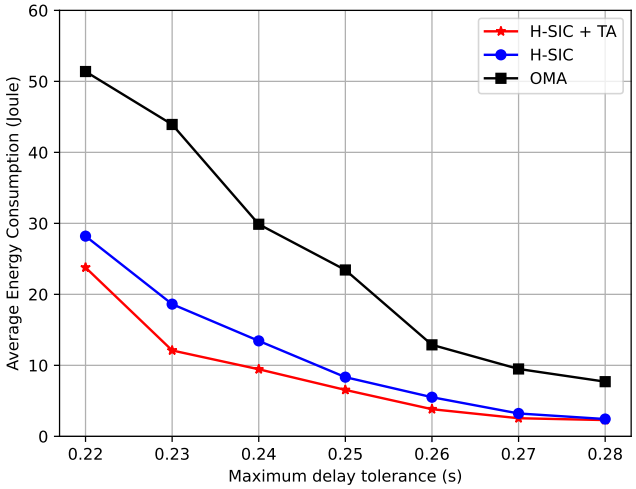

Figure 8. Average energy consumption versus maximum delay tolerance.

6. Conclusions

This paper studied the resource allocation problem for a NOMA-assisted MEC network to minimize the energy consumption of users' offloading activities. The hybrid NOMA scheme had two durations during each time slot, in which NOMA was adopted to serve both users simultaneously during the first time duration, and a dedicated time slot was scheduled to solely offload the remaining part of the more delay-tolerant user by OMA. We assumed the user grouping policy was given at the beginning, the non-convex problem was decomposed into three sub-problems including power allocation, time allocation, and task assignment, which were all solved optimally by studying the convexity and monotonicity. The hybrid SIC scheme selected the SIC decoding order dynamically by the numerical comparison of the energy consumption among different decoding sequences. Finally, after solving those sub-problems, we proposed a DQN-based user-grouping algorithm to obtain the user grouping policy and minimize the long-term average offloading energy consumption. The convergence simulation results showed that the proposed DQN algorithm had similar convergence performance when different numbers of users were chosen, and the ϵ-decay policy was effective at stabilizing the network after convergence. In addition, by comparing with various benchmarks, the partial offloading scheme could reduce the energy consumption compared to full offloading, and the hybrid NOMA transmission outperformed the conventional OMA transmission. Hence, it proved the superiority of the proposed NOMA-MEC scheme in terms of energy consumption.

Author Contributions: Conceptualization, H.L. and Z.D.; formal analysis, H.L. and F.F.; methodology, H.L.; project administration, H.L. and Z.D.; supervision, Z.D.; validation, H.L.; writing—original draft, H.L.; writing—review and editing, F.F. and Z.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the U.K. EPSRC under Grant Number EP/P009719/2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

By fixing $t_{r,\phi}$ and β_ϕ, the above problem in the case $\mathbf{1}_{n,\phi} = 1$ can be rewritten as:

$$(\mathcal{P}5): \min_{\mathrm{P_{n,\mathbb{C}}},\mathrm{P_{r,\mathbb{C}}}} \quad \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^2} + \tau_{m,\phi}P_{n,\phi} + t_{r,\phi}P_{r,\phi} \tag{A1a}$$

$$\text{s.t.} \quad \tau_{m,\phi}B\ln\left(1+P_{n,\phi}|h_{n,\phi}|^2\right) + t_{r,\phi}B\ln\left(1+P_{r,\phi}|h_{n,\phi}|^2\right) \geq \beta_\phi L \tag{A1b}$$

$$\tau_{m,\phi}B\ln\left(1+\frac{P_{m,\phi}|h_{m,\phi}|^2}{P_{n,\phi}|h_{n,\phi}|^2+1}\right) \geq L \tag{A1c}$$

$$P_{n,\phi} \geq 0, P_{r,\phi} \geq 0 \tag{A1d}$$

It is evident that the problem is convex, and by rearranging (A1d) as:

$$P_{n,\phi}|h_{n,\phi}|^2 - P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}}-1\right)^{-1} + 1 \leq 0, \tag{A2}$$

the Lagrangian function can be obtained as follows:

$$\begin{aligned}\mathcal{L}(P_{n,\phi},P_{r,\phi},\boldsymbol{\lambda}) =& \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^2} + \tau_{m,\phi}P_{n,\phi} + t_{r,\phi}P_{r,\phi} - \lambda_1 P_{n,\phi} - \lambda_2 P_{r,\phi} + \lambda_3\beta_\phi L \\ &- \lambda_3\tau_{m,\phi}B\ln\left(1+P_{n,\phi}|h_{n,\phi}|^2\right) - \lambda_3 t_{r,\phi}B\ln\left(1+P_{r,\phi}|h_{n,\phi}|^2\right) \\ &+ \lambda_4\left(P_{n,\phi}|h_{n,\phi}|^2 - P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}}-1\right)^{-1}+1\right),\end{aligned} \tag{A3}$$

where $\boldsymbol{\lambda} \triangleq [\lambda_1,\lambda_2,\lambda_3,\lambda_4]$ are the Lagrangian multipliers. The stationary conditions are given as:

$$\frac{\partial\mathcal{L}}{\partial P_{n,\phi}} = \tau_{m,\phi} - \lambda_1 - \lambda_3\tau_{m,\phi}B\frac{|h_{n,\phi}|^2}{P_{n,\phi}|h_{n,\phi}|^2+1} + \lambda_4|h_{n,\phi}|^2 = 0 \tag{A4}$$

$$\frac{\partial\mathcal{L}}{\partial P_{r,\phi}} = t_{r,\phi} - \lambda_2 - \lambda_3 t_{r,\phi}B\frac{|h_{n,\phi}|^2}{P_{r,\phi}|h_{n,\phi}|^2+1} = 0 \tag{A5}$$

The Karush–Kuhn–Tucker (KKT) conditions [32] can be obtained as:

$$\beta_\phi L - \tau_{m,\phi} B \ln\left(1 + P_{n,\phi}|h_{n,\phi}|^2\right) - t_{r,\phi} B \ln\left(1 + P_{r,\phi}|h_{n,\phi}|^2\right) \leq 0 \tag{A6}$$

$$P_{n,\phi}|h_{n,\phi}|^2 - P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)^{-1} + 1 \leq 0 \tag{A7}$$

$$-P_{n,\phi} \leq 0, -P_{r,\phi} \leq 0 \tag{A8}$$

$$\lambda_i \geq 0, \quad i \in \{1, 2, 3, 4\} \tag{A9}$$

$$\lambda_1 P_{n,\phi} = 0 \tag{A10}$$

$$\lambda_2 P_{r,\phi} = 0 \tag{A11}$$

$$\lambda_3\beta_\phi L - \lambda_3\tau_{m,\phi} B \ln\left(1 + P_{n,\phi}|h_{n,\phi}|^2\right) - \lambda_3 t_{r,\phi} B \ln\left(1 + P_{r,\phi}|h_{n,\phi}|^2\right) = 0 \tag{A12}$$

$$\lambda_4\left(P_{n,\phi}|h_{n,\phi}|^2 - P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)^{-1} + 1\right) = 0 \tag{A13}$$

$$\tau_{m,\phi} - \lambda_1 - \lambda_3\tau_{m,\phi} B\frac{|h_{n,\phi}|^2}{P_{n,\phi}|h_{n,\phi}|^2 + 1} + \lambda_4|h_{n,\phi}|^2 = 0 \tag{A14}$$

$$t_{r,\phi} - \lambda_2 - \lambda_3 t_{r,\phi} B\frac{|h_{n,\phi}|^2}{P_{r,\phi}|h_{n,\phi}|^2 + 1} \tag{A15}$$

The power allocation schemes can be obtained by different Lagrangian multipliers decisions as follows:

- Hybrid NOMA: $\lambda_1 = 0$, $\lambda_2 = 0$, and $\lambda_3 \neq 0$.
  - If $\lambda_4 = 0$:

$$P^*_{n,\phi} = P^*_{r,\phi} = |h_{n,\phi}|^{-2}\left(e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi} + t_{r,\phi}\right)}} - 1\right) \tag{A16}$$

$$P_{m,\phi}|h_{m,\phi}|^2 \geq e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi} + t_{r,\phi}\right)}}\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right) \tag{A17}$$

  - If $\lambda_4 \neq 0$:

$$P^*_{n,\phi} = |h_{n,\phi}|^{-2}\left[P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)^{-1} - 1\right], \tag{A18}$$

$$P^*_{r,\phi} = |h_{n,\phi}|^{-2}\left\{e^{\frac{\beta_\phi L}{B t_{r,\phi}} - \frac{\tau_{m,\phi}}{t_{r,\phi}}\ln\left[P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)^{-1}\right]} - 1\right\}, \tag{A19}$$

    where $e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}}\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right) \geq P_{m,\phi}|h_{m,\phi}|^2 \geq e^{\frac{L}{B\tau_{m,\phi}}} - 1$.

- Pure NOMA: $\lambda_1 = 0$, $\lambda_2 \neq 0$:

$$P_{n,\phi} = |h_{n,\phi}|^{-2}\left(e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}} - 1\right), \tag{A20}$$

$$P_{r,\phi} = 0, \tag{A21}$$

  where $P_{m,\phi}|h_{m,\phi}|^2 \geq e^{\frac{\beta_\phi L}{B\tau_{m,\phi}}}\left(e^{\frac{L}{B\tau_{m,\phi}}} - 1\right)$.

- OMA: $\lambda_1 \neq 0$, $\lambda_2 = 0$:

$$P_{n,\phi} = 0, \tag{A22}$$

$$P_{r,\phi} = |h_{n,\phi}|^{-2}\left(e^{\frac{\beta_\phi L}{t_{r,\phi}B}} - 1\right). \tag{A23}$$

Appendix B. Proof of Proposition 1

The total energy consumption can be expressed as:

$$\begin{aligned} E_{\mathrm{H1}} =& \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^2} + \tau_{m,\phi}|h_{n,\phi}|^{-2}\left[P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}}-1\right)^{-1}-1\right] \\ &+ t_{r,\phi}|h_{n,\phi}|^{-2}\left\{e^{\frac{\beta_\phi L}{Bt_{r,\phi}} - \frac{\tau_{m,\phi}}{t_{r,\phi}}\ln\left[P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}}-1\right)^{-1}\right]}-1\right\}, \end{aligned} \tag{A24}$$

where

$$a_\phi = \frac{\beta_\phi L - B\tau_{m,\phi}\ln\left[P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}}-1\right)^{-1}\right]}{B}.$$

$$\frac{\partial E_{\mathrm{H1}}}{\partial t_{r,\phi}} = -\frac{2\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^3} + |h_{n,\phi}|^{-2}\left(e^{\frac{a_\phi}{t_{r,\phi}}} - \frac{a_\phi}{t_{r,\phi}}e^{\frac{a_\phi}{t_{r,\phi}}} - 1\right). \tag{A25}$$

Define $g(x) = e^{\frac{a_\phi}{x}} - \frac{a_\phi}{x}e^{\frac{a_{\phi,1}}{x}} - 1$,

$$g'(x) = \frac{a_{\phi,1}^2 e^{\frac{a_{\phi,1}}{x}}}{x^3} \geq 0, \quad \forall x > 0. \tag{A26}$$

Hence, $g(x)$ is monotonically increasing for $x > 0$, and $g(t_{r,\phi}) \leq g(\infty) = 0$.

Therefore, $\frac{dE_{\mathrm{H1}}}{dt_{r,\phi}} \leq 0$, which is monotonically decreasing. Hence, the larger $t_{r,\phi}$ scheduled, the less energy is consumed, and the optimal situation is when $t^*_{r,\phi} = \tau_{n,\phi} - \tau_{m,\phi}$.

For the power allocation scheme in (23), the energy consumption is given as:

$$E_{\mathrm{H2}} = \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^2} + \left(\tau_{m,\phi}+t_{r,\phi}\right)|h_{n,\phi}|^{-2}\left(e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}}-1\right). \tag{A27}$$

By obtaining the derivative with respect to $t_{r,\phi}$,

$$\frac{\partial E_{\mathrm{H2}}}{\partial t_{r,\phi}} = -\frac{2\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t_{r,\phi}\right)^3} + |h_n|^{-2}\left(e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}} - t_{r,\phi}\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+t_{r,\phi}\right)}} - 1\right). \tag{A28}$$

Define $g_2(x) \triangleq e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+x\right)}} - x\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+x\right)}e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+x\right)}} - 1$, and the derivative of $g_2(x)$ is:

$$g_2'(x) = \frac{\left(\beta_\phi L\right)^2}{B^2\left(\tau_{m,\phi}+x\right)^3}e^{\frac{\beta_\phi L}{B\left(\tau_{m,\phi}+x\right)}} \geq 0, \quad \forall x > 0. \tag{A29}$$

Thus, $g_2(x)$ is monotonically increasing for $x > 0$, and $g(t_{r,\phi}) \leq g(\infty) = 0$, which indicates $\frac{dE_{\mathrm{H2}}}{dt_{r,\phi}} \leq 0$. Similar to the previous case, the energy function is monotonically decreasing with respect to $t_{r,\phi}$, and the optimal time allocation is $t^*_{r,\phi} = \tau_{n,\phi} - \tau_{m,\phi}$.

Appendix C. Proof to Proposition 2

$$\min_{\beta_\phi} \quad \frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t^*_{r,\phi}\right)^2}+\tau_{m,\phi}P^*_{n,\phi}+t^*_{r,\phi}P^*_{r,\phi} \tag{A30}$$

$$\text{s.t.} \quad 0\le\beta_\phi\le 1. \tag{A31}$$

The Lagrangian is given as:

$$\mathcal{L}(\beta_\phi,\lambda_5,\lambda_6)=\frac{\kappa_0\left[C(1-\beta_\phi)L\right]^3}{\left(\tau_{m,\phi}+t^*_{r,\phi}\right)^2}+\tau_{m,\phi}P^*_{n,\phi}+t^*_{r,\phi}P^*_{r,\phi}-\lambda_5\beta_\phi+\lambda_6\left(\beta_\phi-1\right) \tag{A32}$$

- For the case $P_{m,\phi}=P_{n,\phi}$ in (22), the stationary condition is obtained as:

$$\frac{\partial\mathcal{L}}{\partial\beta_\phi}=\frac{-3\kappa_0(CL)^3\left(1-\beta_\phi\right)^2}{\tau^2_{n,\phi}}+\frac{L}{B}|h_{n,\phi}|^{-2}e^{\frac{\beta_\phi L}{B\tau_{n,\phi}}}-\lambda_5+\lambda_6=0. \tag{A33}$$

Therefore, the KKT conditions can be written as follows:

$$-\beta_\phi\le 0 \tag{A34}$$

$$\beta_\phi-1\le 0 \tag{A35}$$

$$\lambda_5\beta_\phi=0 \tag{A36}$$

$$\lambda_6\left(\beta_\phi-1\right)=0 \tag{A37}$$

$$\frac{-3\kappa_0(CL)^3\left(1-\beta_\phi\right)^2}{\tau^2_{n,\phi}}+\frac{L}{B}|h_{n,\phi}|^{-2}e^{\frac{\beta_\phi L}{B\tau_{n,\phi}}}-\lambda_5+\lambda_6=0 \tag{A38}$$

For $\beta_\phi>0$, $\lambda_5=\lambda_6=0$, and (A38) can be rewritten as:

$$\frac{3\kappa_0(CL)^3\left(1-\beta_\phi\right)^2}{\tau^2_{n,\phi}}=\frac{L}{B}|h_{n,\phi}|^{-2}e^{\frac{\beta_\phi L}{B\tau_{n,\phi}}}. \tag{A39}$$

Define $z_{1,\phi}=\frac{3\kappa_0BC^3L^2|h_{n,\phi}|^2}{\tau^2_{n,\phi}}$, $z_{2,\phi}=\frac{L}{B\tau_{n,\phi}}$, and $b_\phi=\left(1-\beta_\phi\right)$. The optimal task assignment coefficient can be derived as:

$$z_{1,\phi}b_\phi^2=e^{z_{2,\phi}\left(1-b_\phi\right)}, \tag{A40}$$

$$b_\phi=\frac{2}{z_{2,\phi}}\mathcal{W}\left(\frac{1}{2}z_{1,\phi}^{-\frac{1}{2}}z_{2,\phi}e^{\frac{z_{2,\phi}}{2}}\right). \tag{A41}$$

The optimal task assignment ratio can be expressed as:

$$\beta^*_\phi=1-b=1-\frac{2}{z_{2,\phi}}\mathcal{W}\left(\frac{1}{2}z_{1,\phi}^{-\frac{1}{2}}z_{2,\phi}e^{\frac{z_{2,\phi}}{2}}\right). \tag{A42}$$

- For the case $P_{m,\phi}\neq P_{n,\phi}$ in (23):
The stationary condition can be expressed as:

$$\frac{\partial\mathcal{L}}{\partial\beta_\phi}=\frac{-3\kappa_0(CL)^3\left(1-\beta_\phi\right)^2}{\tau^2_{n,\phi}}+|h_{n,\phi}|^{-2}\frac{L}{B}e^{-u}e^{\frac{\beta_\phi L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}-u}-\lambda_5+\lambda_6=0, \tag{A43}$$

where $u_\phi = \frac{\tau_{m,\phi}}{\left(\tau_{n,\phi}-\tau_{m,\phi}\right)} \ln\left[P_{m,\phi}|h_{m,\phi}|^2\left(e^{\frac{L}{B\tau_{m,\phi}}}-1\right)^{-1}\right]$.

$$-\beta_\phi \leq 0 \tag{A44}$$

$$\beta_\phi - 1 \leq 0 \tag{A45}$$

$$\lambda_5\beta_\phi = 0 \tag{A46}$$

$$\lambda_6\left(\beta_\phi - 1\right) = 0 \tag{A47}$$

$$\frac{-3\kappa_0(CL)^3\left(1-\beta_\phi\right)^2}{\tau_{n,\phi}^2} + |h_{n,\phi}|^{-2}\frac{L}{B}e^{-u_\phi}e^{\frac{\beta_\phi L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}-u_\phi} - \lambda_5 + \lambda_6 = 0 \tag{A48}$$

For $\beta_\phi > 0$, $\lambda_5 = \lambda_6 = 0$, constraint (A48) can be rearranged as:

$$\frac{3\kappa_0(CL)^3\left(1-\beta_\phi\right)^2}{\tau_{n,\phi}^2} = |h_{n,\phi}|^{-2}\frac{L}{B}e^{-u_\phi}e^{\frac{\beta_\phi L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}-u_\phi}, \tag{A49}$$

$$\frac{3\kappa_0 B|h_{n,\phi}|^2(CL)^3e^{2u_\phi}\left(1-\beta_\phi\right)^2}{\tau_{n,\phi}^2 L} = e^{\frac{\beta_\phi L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}}. \tag{A50}$$

Define $z_{1,\phi} = \frac{3\kappa_0 B|h_{n,\phi}|^2C^3L^2e^{2u_\phi}}{\tau_{n,\phi}^2}$, $z_{2,\phi} = \frac{L}{B\left(\tau_{n,\phi}-\tau_{m,\phi}\right)}$. The above equation can be rewritten as:

$$z_{1,\phi}b_\phi^2 = e^{z_{2,\phi}\left(1-b_\phi\right)}, \tag{A51}$$

$$b_\phi = \frac{2}{z_{2,\phi}}\mathcal{W}\left(\frac{1}{2}z_{1,\phi}^{-\frac{1}{2}}z_{2,\phi}e^{\frac{z_{2,\phi}}{2}}\right), \tag{A52}$$

Hence, the optimal task partition assignment ratio is:

$$\beta_\phi^* = 1 - b_\phi = 1 - \frac{2}{z_{2,\phi}}\mathcal{W}\left(\frac{1}{2}z_{1,\phi}^{-\frac{1}{2}}z_{2,\phi}e^{\frac{z_{2,\phi}}{2}}\right). \tag{A53}$$

References

1. Nduwayezu, M.; Pham, Q.; Hwang, W. Online Computation Offloading in NOMA-Based Multi-Access Edge Computing: A Deep Reinforcement Learning Approach. *IEEE Access* **2020**, *8*, 99098–99109. [CrossRef]
2. Sun, W.; Zhang, H.; Wang, R.; Zhang, Y. Reducing Offloading Latency for Digital Twin Edge Networks in 6G. *IEEE Trans. Veh. Technol.* **2020**, *69*, 12240–12251. [CrossRef]
3. Zhao, R.; Wang, X.; Xia, J.; Fan, L. Deep reinforcement learning based mobile edge computing for intelligent Internet of Things. *Phys. Commun.* **2020**, *43*, 101184. [CrossRef]
4. Li, L.; Cheng, Q.; Tang, X.; Bai, T.; Chen, W.; Ding, Z.; Han, Z. Resource Allocation for NOMA-MEC Systems in Ultra-Dense Networks: A Learning Aided Mean-Field Game Approach. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 1487–1500. [CrossRef]
5. Bai, T.; Pan, C.; Deng, Y.; Elkashlan, M.; Nallanathan, A.; Hanzo, L. Latency Minimization for Intelligent Reflecting Surface Aided Mobile Edge Computing. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2666–2682. [CrossRef]
6. Dinh, H.T.; Lee, C.; Niyato, D.; Wang, P. A survey of mobile cloud computing: Architecture, applications, and approaches. *Wirel. Commun. Mob. Comput.* **2013**, *13*, 1587–1611. [CrossRef]
7. Abbas, N.; Zhang, Y.; Taherkordi, A.; Skeie, T. Mobile Edge Computing: A Survey. *IEEE Internet Things J.* **2018**, *5*, 450–465. [CrossRef]
8. Huang, Y.; Liu, Y.; Chen, F. NOMA-Aided Mobile Edge Computing via User Cooperation. *IEEE Trans. Commun.* **2020**, *68*, 2221–2235. [CrossRef]
9. Chen, A.; Yang, Z.; Lyu, B.; Xu, B. System Delay Minimization for NOMA-Based Cognitive Mobile Edge Computing. *IEEE Access* **2020**, *8*, 62228–62237. [CrossRef]
10. Mao, Y.; You, C.; Zhang, J.; Huang, K.; Letaief, K.B. A Survey on Mobile Edge Computing: The Communication Perspective. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2322–2358. [CrossRef]
11. Dao, N.N.; Pham, Q.V.; Tu, N.H.; Thanh, T.T.; Bao, V.N.Q.; Lakew, D.S.; Cho, S. Survey on Aerial Radio Access Networks: Toward a Comprehensive 6G Access Infrastructure. *IEEE Commun. Surv. Tutor.* **2021**. [CrossRef]

12. Makki, B.; Chitti, K.; Behravan, A.; Alouini, M.S. A Survey of NOMA: Current Status and Open Research Challenges. *IEEE Open J. Commun. Soc.* **2020**, *1*, 179–189. [CrossRef]
13. Vaezi, M.; Aruma Baduge, G.A.; Liu, Y.; Arafa, A.; Fang, F.; Ding, Z. Interplay Between NOMA and Other Emerging Technologies: A Survey. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 900–919. [CrossRef]
14. Fang, F.; Xu, Y.; Ding, Z.; Shen, C.; Peng, M.; Karagiannidis, G.K. Optimal Resource Allocation for Delay Minimization in NOMA-MEC Networks. *IEEE Trans. Commun.* **2020**, *68*, 7867–7881. [CrossRef]
15. Zeng, M.; Nguyen, N.; Dobre, O.A.; Poor, H.V. Delay Minimization for NOMA-Assisted MEC Under Power and Energy Constraints. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1657–1661. [CrossRef]
16. Song, Z.; Liu, Y.; Sun, X. Joint Radio and Computational Resource Allocation for NOMA-Based Mobile Edge Computing in Heterogeneous Networks. *IEEE Commun. Lett.* **2018**. *22*, 2559–2562. [CrossRef]
17. Xu, C.; Zheng, G.; Zhao, X. Energy-Minimization Task Offloading and Resource Allocation for Mobile Edge Computing in NOMA Heterogeneous Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 16001–16016. [CrossRef]
18. Wang, F.; Xu, J.; Ding, Z. Multi-Antenna NOMA for Computation Offloading in Multiuser Mobile Edge Computing Systems. *IEEE Trans. Commun.* **2019**, *67*, 2450–2463. [CrossRef]
19. Ding, Z.; Xu, J.; Dobre, O.A.; Poor, H.V. Joint Power and Time Allocation for NOMA–MEC Offloading. *IEEE Trans. Veh. Technol.* **2019**, *68*, 6207–6211. [CrossRef]
20. Zhu, J.; Wang, J.; Huang, Y.; Fang, F.; Navaie, K.; Ding, Z. Resource Allocation for Hybrid NOMA MEC Offloading. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 4964–4977. [CrossRef]
21. Li, H.; Fang, F.; Ding, Z. Joint resource allocation for hybrid NOMA-assisted MEC in 6G networks. *Digit. Commun. Netw.* **2020**, *6*, 241–252. [CrossRef]
22. Wang, X.; Zhang, Y.; Shen, R.; Xu, Y.; Zheng, F.C. DRL-Based Energy-Efficient Resource Allocation Frameworks for Uplink NOMA Systems. *IEEE Internet Things J.* **2020**, *7*, 7279–7294. [CrossRef]
23. He, C.; Hu, Y.; Chen, Y.; Zeng, B. Joint Power Allocation and Channel Assignment for NOMA with Deep Reinforcement Learning. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2200–2210. [CrossRef]
24. Ding, Z.; Yang, Z.; Fan, P.; Poor, H.V. On the Performance of Non-Orthogonal Multiple Access in 5G Systems with Randomly Deployed Users. *IEEE Signal Process. Lett.* **2014**, *21*, 1501–1505. [CrossRef]
25. Ding, Z.; Fan, P.; Poor, H.V. Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions. *IEEE Trans. Veh. Technol.* **2016**, *65*, 6010–6023. [CrossRef]
26. Zeng, M.; Hao, W.; Dobre, O.A.; Ding, Z.; Poor, H.V. Power Minimization for Multi-Cell Uplink NOMA With Imperfect SIC. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 2030–2034. [CrossRef]
27. Ding, Z.; Schober, R.; Poor, H.V. Unveiling the Importance of SIC in NOMA Systems—Part 1: State of the Art and Recent Findings. *IEEE Commun. Lett.* **2020**, *24*, 2373–2377. [CrossRef]
28. Ding, Z.; Schober, R.; Poor, H.V. Unveiling the Importance of SIC in NOMA Systems—Part II: New Results and Future Directions. *IEEE Commun. Lett.* **2020**, *24*, 2378–2382. [CrossRef]
29. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.A.; Fidjeland, A.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]
30. Agarap, A.F. Deep Learning using Rectified Linear Units (ReLU). *arXiv* **2018**, arxiv:1803.08375.
31. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
32. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.