Deep Reinforcement Learning for NPDCCH Period Adjustment in NB-IoT Networks

Ya-Ju Yu[1], Ching-Chih Chuang[2], and Yu-Wei Cheng[1]
[1]Department of Computer Science and Information Engineering, National University of Kaohsiung, Taiwan
[2]Department of Information Management, National Pingtung University, Taiwan
E-mail: yjyu@nuk.edu.tw, ccc@mail.nptu.edu.tw, and m1065502@mail.nuk.edu.tw

***Abstract*—In narrowband Internet of Things (NB-IoT) networks, a base station can provide multiple coverage enhancement (CE) levels to accommodate devices with diverse channel qualities for achieving massive connections. The base station can determine a narrowband physical downlink control channel (NPDCCH) period (NP) for each CE level. The base station should allocate uplink and downlink radio resources on an NP-basis for each device in a CE level. The NP length of each CE level significantly affects the efficiency of uplink and downlink radio resource allocation. However, the key challenge in determining an NP length for each CE level is that the NP length should be decided before the uplink and downlink resource allocation and cannot be frequently changed. Therefore, this paper studies the NPDCCH period adjustment problem considering uplink and downlink traffic. The objective is to minimize the radio resource consumption while each device can transmit or receive its data. To overcome the above issue, we propose an algorithm based on deep reinforcement learning to solve the target problem. Compared with the two baselines, the simulation results show the efficacy of the proposed algorithm and useful insights into the design of the NPDCCH period adjustment algorithm for NB-IoT networks.**

***Index Terms*—NB-IoT, NPDCCH period, deep reinforcement learning, cellular networks**

I. Introduction

The 3rd generation partnership project (3GPP) has specified the narrowband Internet of Things (NB-IoT) protocol to support massive IoT devices for cyber-physical systems in fifth-generation (5G) cellular networks [1]–[3]. A cellular base station should serve 480,000 devices [4], and Cisco forecasts that 29.3 billion mobile-connected devices will be in use by 2023 [5]. The amount of data to be collected and transmitted is expected to increase at an exceptional rate [6]. Bursty devices (sensors) may simultaneously send data packets to a base station. Therefore, achieving efficient use of the radio resources of NB-IoT networks for massive connections is crucial.

A base station can provide multiple coverage enhancement (CE) levels to accommodate devices with diverse channel qualities to accomplish high coverage and extensive connection requirements in NB-IoT networks. The base station can determine a narrowband physical downlink control channel (NPDCCH) period (NP), which consists of an NPDCCH and a narrowband physical downlink shared channel (NPDSCH) in downlink frequency for a CE level. The base station should allocate radio resources on an NP-basis for each device in a CE level. No matter whether uplink or downlink transmissions, a device should blindly decode a downlink control indicator (DCI) in an NPDCCH search space. Then the device can receive data in the NPDSCH or transmit data in a narrowband uplink shared channel (NPUSCH). An NP length comprised of NPDCCH length and NPDSCH length significantly affects the efficiency of the uplink and downlink radio resource allocation. Therefore, determining a suitable NP length for each CE level is critical and refers to the *NPDCCH period adjustment problem*. However, the key challenge when it comes to determining an NP length for each CE level is that the NP length should be decided before the uplink and downlink resource allocation and cannot be frequently changed. Tackling this problem would result in a Markov decision process (MDP) with large state and action spaces.

Recently, the uplink resource allocation algorithms for NB-IoT cellular networks have been proposed in [7], [8], and [9]. Feltrin et al. [10] surveyed the downlink and uplink frame structures and procedures for 3GPP NB-IoT networks. Further, Mostafa et al. [7] investigated the joint subcarrier and transmission power allocation problem to maximize the number of low-cost machine-type communication devices, satisfying a desired quality of service. Hsieh et al. [11] proposed a DCI and subcarrier allocation algorithm for uplink transmissions. In [9], Yu et al. considered narrowband physical random access channels (NPRACHs) in the link adaptation and uplink resource allocation problem. They proposed an uplink resource allocation algorithm based on a dynamic programming approach to solve the problem. The downlink resource allocation algorithms for NB-IoT networks have been investigated in [12] and [13]. Huang et al. [12] identified radio resource scheduling issues for NB-IoT systems and proposed a downlink scheduling algorithm for NB-IoT. In [13], Yu proposed an approximation algorithm for the downlink resource allocation problem. The current paper considers that both uplink and downlink resource allocation algorithms are given and can be applied no matter which resource allocation algorithm is adopted for NB-IoT.

The NPDCCH period adjustment problem has been studied in [13]. In [13], Yu proposed a heuristic algorithm to determine an NP for each CE level. The algorithm selects the shortest NPDCCH length that can fulfill the required DCI repetition requirements of all devices and the shortest NPDSCH length that can satisfy the required data repetition requirements of

all devices in a CE level. The algorithm is designed only considering downlink resource allocation, but none simultaneously considers uplink and downlink resource allocation. Because uplink and downlink devices should share the same NPDCCH resource pool, the NPDCCH period adjustment problem should simultaneously consider uplink and downlink traffic.

In this paper, we investigate the NPDCCH period adjustment problem with the consideration of uplink and downlink transmissions in NB-IoT networks. The objective is to minimize the total consumed subframes for satisfying the uplink and downlink data requirements of bursty devices. We summarize the contributions of the present paper as follows.

- This paper completely considers uplink and downlink frame structures and traffic for the NPDCCH period adjustment problem. In the frame structures, we also practically consider the essential signals and NPRACHs.
- We build the reinforcement learning settings including the observation selection, action selection, and reward function for the target problem. Then, given an uplink and downlink resource allocation algorithm, we propose an algorithm based on deep reinforcement learning to solve the target problem.
- We conduct a series of simulations to evaluate the performance of the proposed algorithm. The proposed algorithm is compared using a static method [12] and a baseline designed by [13]. The simulation results show that, compared with the two baselines, our proposed algorithm can reduce the radio resource consumption. The simulation results justify that the NPDCCH period adjustment algorithm design should simultaneously consider uplink and downlink resource allocation, while the results also illustrate an interesting phenomenon, specifically that in determining an NP length there exists a trade-off between the NPDCCH and NPDSCH subframe consumption.

The remainder of this paper is organized as follows. In Section II, we describe the system model and problem formulation. In Section III, we detail the proposed algorithm. Section IV presents simulation results before we conclude this paper in Section V.

II. Modeling and Problem Formulation

A. System Model

This paper considers a single NB-IoT cell serving a set of static NB-IoT devices with uplink and/or downlink transmission requirements. When a device has downlink traffic (e.g., remote reconfigurations of IoT devices, firmware, or software updates) or uplink traffic (e.g., environmental monitoring, or fire alarms), it will receive system signals (e.g., master information block and narrowband primary/secondary synchronization signals). The device will determine its associated CE level by comparing the received power of the broadcast signal to the predefined thresholds of the reference signal received power [14]. The device will then perform the random access procedure according to the predetermined configuration of the associated CE level in the NPRACH. When the device has successfully completed the random access procedure, the base station can select an appropriate NP determined by two parameters, R_{max} and G, for each CE level in the radio resource control connection setup. Each device with different repetition requirements (channel qualities) should transmit or receive data in the defined NP length of its associated CE level.

In determining an NP length, there exists a trade-off between transmission reliability and transmission time-interval. Because three CE levels' devices should share the same radio resources including uplink and downlink resources, the resource allocation efficiency will be significantly affected by three NP lengths of three CE levels. However, the key challenge for determining the three NP lengths of three CE levels is that the three NP lengths should be decided before the uplink and downlink resource allocation and cannot be frequently changed. Tackling this problem would result in an MDP with large state and action spaces. This paper adopts the view that the bursty devices have completed the CE level determination and the random access procedure. We target how to determine an NP for each CE level such that a given downlink and uplink radio resource allocation can consume the minimum radio subframes for satisfying the uplink and downlink data requirements. Readers can refer to [9] and [13] for background knowledge about the uplink and downlink resource allocation of NB-IoT.

B. Problem Formulation

This paper considers that a base station serves a set of V connected NB-IoT devices under L CE levels. The set of devices in the CE level ℓ is denoted as V^ℓ, $V^\ell \bigcap V^{\ell+1} = \emptyset$. The set of V^ℓ devices incudes $|V_U^\ell|$ devices with uplink requests and $|V_D^\ell|$ devices with downlink requests. Device v in CE level ℓ has a data requirement of size ψ_v^ℓ.

The SNR between the base station and device v is η_v. To guarantee the required transmission reliability, device v requires the minimum data repetition number to be $\hat{R}_v$ under the measured SNR. A set of repetition numbers $\Re_U$ and $\Re_D$ can be chosen by the base station for uplink and downlink data transmissions, respectively. The base station should select a repetition number satisfying $R_j \geq \hat{R}_v$, $R_j \in \Re_U$ for uplink and $R_j \in \Re_D$ for downlink.

The length of an NPDCCH period for the CE level ℓ is $R_{max}^\ell \times G^\ell$. Parameter R_{max} can be selected from set $\mathbb{R}_{max}$ and determines the number of NPDCCH subframes in an NPDCCH period. Parameter G is a system parameter and can be chosen from set $\mathbb{G}$. The required DCI repetition number of device v to guarantee transmission reliability is at least $\hat{\mathcal{D}}_v$. There is a set of repetition numbers $\Re_{DCI}$ that can be selected for transmitting a DCI. The repetition number to transmit one DCI should also ensure the reliability (i.e., $R_j \in \Re_{DCI} \geq \hat{\mathcal{D}}_v$). In other words, if a device v requires a higher $\hat{\mathcal{D}}_v$, the base station may require to select a higher R_{max} to increase R_j.

Under the selected R_{max}^ℓ and G^ℓ values for CE level ℓ, a given uplink and downlink resource allocation will consume

TABLE I
SUMMARY OF NOTATIONS

Symbol	Description
L	The number of CE levels
V	The set of NB-IoT devices
V^ℓ	The number of devices in CE level ℓ
ψ_d^ℓ	The data requirement of device v in CE level ℓ
$\hat{R}_v$	The data repetition requirement of device v
$\Re_D$	The set of data repetition numbers can be selected by a base station for downlink transmissions
$\Re_U$	The set of data repetition numbers can be selected by a base station for uplink transmissions
R_{max}^ℓ	The number of NPDCCH subframes for CE level ℓ
$\mathbb{R}_{max}$	The set of R_{max} values
G^ℓ	The system parameter for determining an NPDCCH period for CE level ℓ
$\mathbb{G}$	The set of G values
$\Re_{DCI}$	The set of DCI repetition numbers can be selected by a base station for transmitting a DCI
$\hat{\mathcal{D}}_v$	The required DCI repetition number of device v
C_U^ℓ	The number of consumed subframes under a given resource allocation algorithm for satisfying the devices' uplink data requirements in CE level ℓ.
C_D^ℓ	The number of consumed subframes under a given resource allocation algorithm for satisfying the devices' downlink data requirements in CE level ℓ.

C_U^ℓ subframes to satisfy the uplink data requirements of $|V_U^\ell|$ devices and C_D^ℓ subframes to satisfy the downlink data requirements of $|V_D^\ell|$ devices. The objective of this paper is to minimize the number of consumed subframes to serve V devices for uplink and downlink transmissions by determining R_{max}^ℓ and G^ℓ values for each CE level. The problem can be formulated as the optimization

$$\min_{R_{max}^\ell, G^\ell} \sum_{\ell=0}^{L} C_U^\ell + C_D^\ell. \tag{1}$$

The notations used in this paper are summarized in Table I.

III. NPDCCH Period Adjustment

In this section, the reinforcement learning settings, including observation selection, action selection, and reward function, are introduced in Section III-A. Subsequently, the deep Q-learning-based NPDCCH period adjustment algorithm is proposed in Section III-B.

A. Reinforcement Learning Settings

In Section III-A, we adopt the deep Q-learning (DQN) algorithm, which consists of a deep neural network (DNN) phase and Q-learning phase, as our reinforcement learning algorithm. The Q-learning phase is a value-based reinforcement learning approach [15], where a value function $Q(s, a)$ is used to determine an action a for the state s such that a reward is maximized. In the deep Q-learning algorithm, a Q-agent needs to explore the environment (system state) to select suitable actions for our optimization goal. The Q-agent observes the current state S_t corresponding to a set of previous observations $O_t^\ell = \{O_{t-1}^\ell, O_{t-2}^\ell, ..., O_0^\ell\}$ for CE level ℓ at the start of the t-th time-interval. During a time-interval, the determined NP length of each CE level cannot be changed. In other words, a time-interval can be set according to the required NPDCCH period adjustment time frame. Based on the knowledge of the state S_t, the Q-agent can choose an action A^t in the action set $\mathcal{A}$. Therefore, selecting appropriate observations and actions is important for learning performance.

1) Observation Selection: We design observed information $O_t^\ell = [\mathbb{R}_D^\ell, \mathbb{R}_U^\ell, \phi_D^\ell, \phi_U^\ell]$.

$$\mathbb{R}_D^\ell = \{N_D^{\ell,1}, ..., N_D^{\ell,j}, ..., N_D^{\ell,|\Re_D|}\} \tag{2}$$

$N_D^{\ell,j}$ is the number of devices with downlink data requests in CE level ℓ, whose data repetition requirements can be satisfied by the repetition number $R_j \in \Re_D$.

$$\mathbb{R}_U^\ell = \{N_U^{\ell,1}, ..., N_U^{\ell,j}, ..., N_U^{\ell,|\Re_U|}\} \tag{3}$$

$N_U^{\ell,j}$ is the number of devices with uplink data requests in CE level ℓ. The devices' data repetition requirements can be fulfilled by the repetition number $R_j \in \Re_U$.

$$\phi_D^\ell = \{\varphi_D^{\ell,1}, ..., \varphi_D^{\ell,j}, ..., \varphi_D^{\ell,|\Re_D|}\} \tag{4}$$

$\varphi_D^{\ell,j}$ is the total number of devices' downlink data requirements in CE level ℓ where the devices' data repetition requirements can be fulfilled by the repetition number $R_j \in \Re_D$.

$$\phi_U^\ell = \{\varphi_U^{\ell,1}, ..., \varphi_U^{\ell,j}, ..., \varphi_U^{\ell,|\Re_U|}\} \tag{5}$$

$\varphi_U^{\ell,j}$ is the total number of devices' uplink data requirements in CE level ℓ when the devices' data repetition requirements can be fulfilled by the repetition number $R_j \in \Re_U$.

2) Action Selection: Because we should determine R_{max}^ℓ and G^ℓ values for each CE level ℓ, the action set is

$$\mathcal{A} = \{[R_{max}^0, G^0], ..., [R_{max}^\ell, G^\ell], ..., [R_{max}^L, G^L]\}. \tag{6}$$

The action set is obviously a large action space. In order to reduce the computational complexity, we separate the action set $\mathcal{A}$ into L sets (CE levels), i.e., $\mathcal{A}^\ell = \{[R_{max}^\ell, G^\ell]\}$, $\forall R_{max}^\ell \in \mathbb{R}_{max}$ and $G^\ell \in \mathbb{G}$. Moreover, because some R_{max} values are not suitable for a CE level, we should select appropriate actions for each CE level to avoid the evil actions. For example, $R_{max} = 2$ is not adequate for CE level 2 because the DCI repetition requirements of the devices in CE level 2 generally cannot be satisfied by two NPDCCH subframes. $R_{max} = 2048$ is not proper to CE level 0 because the devices with high channel qualities in CE level 0 only require low repetition numbers and a long NP is not necessary. Therefore, the action set for each CE level should be different. We set $2 \leq R_{max}^0 \leq 32$ for CE level 0, $8 \leq R_{max}^0 \leq 128$ for CE level 1, and $64 \leq R_{max}^0 \leq 1024$ for CE level 2.

3) Reward Function: We focus on minimizing the total uplink and downlink subframe consumption as our objective. Therefore, the reward function for CE level ℓ during t-th time-interval is defined as:

$$\omega_t^\ell = \frac{1}{C_U^\ell + C_D^\ell} \tag{7}$$

Fig. 1 shows the multiple DQN agents and our environment model. We set a DQN agent with a weight of θ^ℓ for CE level ℓ.

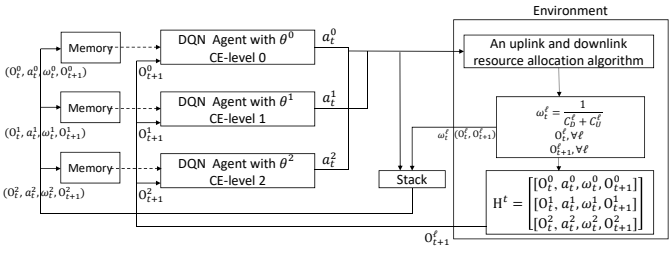

Fig. 1. Three DQN agents for three CE levels and environment interaction.

Algorithm 1: DQN-based NPDCCH period adjustment

Input: L, V, $\Re_D$, $\Re_U$,$\mathbb{R}_{max}$, $\mathbb{G}$, ψ_d^ℓ

1 **for** $t = 0$ *to* T **do**
2 **for** $\ell = 0$ *to* L **do**
3 $O_t^\ell = [\mathbb{R}_D^\ell, \mathbb{R}_U^\ell, \phi_D^\ell, \phi_U^\ell]$
4 **if** $P_\epsilon < \epsilon$ **then**
5 Choose a random action a_t^ℓ from the set $\mathcal{A}^\ell$
6 **else**
7 $a_t^\ell = \arg\max_{a_t^\ell \in \mathcal{A}^\ell} Q(S_t, a_t^\ell, \theta^\ell)$
8 The selected action a_t^ℓ is used for a given uplink and downlink resource allocation algorithm to get the reward ω_t^ℓ
9 Store $(O_t^\ell, a_t^\ell, \omega_t^\ell, O_{t+1}^\ell)$ in the memory
10 Using Eq. (8) and RMSProp to update θ^ℓ
11 **if** *t mod Z = 0* **then**
12 Update target Q-network $\theta'^\ell = \theta^\ell$

The multiple DQN agents are trained in parallel at each time-interval. Each agent of CE level ℓ will decide an action a_t^ℓ from the set $\mathcal{A}^\ell$. The selected action a_t^ℓ is used for a given uplink and downlink resource allocation algorithm. After executing the algorithm, we can obtain the reward of each CE level ℓ at t-th time-interval and a set of observations O_{t+1}^ℓ for each CE level ℓ. Then, $[O_t^\ell, a_t^\ell, \omega_t^\ell, O_{t+1}^\ell]$ are stored in the memory for each CE level ℓ.

B. DQN-based NPDCCH Period Adjustment

The training algorithm for the DQN-based NPDCCH period adjustment algorithm is shown in Algorithm 1. The training algorithm will be trained during T time-intervals (Lines 1-12). At t-th time-interval , we obtain the observed information $O_t^\ell = [\mathbb{R}_D^\ell, \mathbb{R}_U^\ell, \phi_D^\ell, \phi_U^\ell]$ according to Eq. (2)-Eq. (5) for CE level ℓ (Lines 2-3). In the proposed algorithm, an action is selected based on the ϵ-greedy algorithm, which is comprised of exploration and exploitation [16]. In Line 4, ϵ is a predefined threshold, and P_ϵ is a random number which is generated at each time-interval. When $P_\epsilon < \epsilon$, an action is selected by exploration. The Q-agent randomly selects an action a_t^ℓ from the set $\mathcal{A}^\ell$ (Lines 4-5). Otherwise, an action is picked by exploitation. The Q-agent selects the best action $a_t^\ell = \arg\max_{a_t^\ell \in \mathcal{A}^\ell} Q(S_t, a_t^\ell, \theta^\ell)$ according to the previous observations, actions, and rewards at the current state S_t (Lines 6-7). Note that when the training procedure is finished, ϵ is set to 0, and we will not use the exploration phase.

The selected action a_t^ℓ is used for a given uplink and downlink resource allocation algorithm for CE level ℓ. We can obtain the reward ω_t^ℓ including the number of consumed subframes for satisfying the data requirements of uplink and downlink transmissions (Line 8). Values $[O_t^\ell, a_t^\ell, \omega_t^\ell, O_{t+1}^\ell]$ are stored in the memory. In Line 10, the weight θ^ℓ is updated along each training time-interval using the following function [17].

$$E(\theta^\ell) = \left| \left(\omega_t^\ell + \gamma \max_{a_{t+1}^\ell \in \mathcal{A}^\ell} Q(S_{t+1}, a_{t+1}^\ell, \theta'^\ell) \right) - Q(S_t, a_t^\ell, \theta^\ell) \right|^2, \quad (8)$$

where γ is RMSProp learning rate [18]. Based on the gradient of the loss function $E(\theta^\ell)$, we update θ^ℓ. We update target Q-network $\theta'^\ell = \theta^\ell$ every Z time-interval (Lines 11-12).

IV. Performance Evaluation

A. Simulation Setups

In this section, our simulation settings are based on realistic parameters according to 3GPP specifications [3], [19]. The proposed algorithm, namely the *DQN-based NPDCCH Period Adjustment (DNPA)*, is compared with two baselines. The first baseline, named *static* scheme, uses a fixed pair of R_{max} and G values, set by [12]. Specifically, $R_{max} = 8$ and $G = 4$ are used for CE level 0. $R_{max} = 32$ and $G = 4$ are used for CE level 1. $R_{max} = 256$ and $G = 2$ are set for CE level 2. The second baseline, named *NPDCCH period adaptation (NPA)*, chooses the smallest R_{max} and G values that can fulfill the required DCI repetition and required data repetition of all devices in a CE level [13].

We consider a single base station serving a number of devices using an uplink and downlink resource allocation algorithm. The adopted uplink and downlink resource allocation algorithm first allocates radio resources for downlink and then for uplink. The algorithm sequentially allocates each NPDCCH subframe with each scheduling delay value to find an unused NPDSCH/NPUSCH start subframe for a device that has not been satisfied and served in the NP. The algorithm allocates continuous NPDSCH/NPUSCH subframes to the device. For uplink resource allocation, because multi-tones are supported, the algorithm will additionally try each resource unit and each modulation-coding index for each device to find unused radio resources.

We now describe our simulation settings. The RMSProp learning rate γ is set as 0.9. The exploration threshold ϵ is set at 0.1. The number of bursty devices, which complete the random access procedure, varies from 3,000 to 6,000 in a time interval [14]. Each device randomly selects a downlink or an uplink data request. The uplink or downlink data size of each device is randomly selected from 20 bytes to 200 bytes [12]. The transmission power of the base station and each device is, respectively, 32 dBm and 23 dBm [8]. The pass loss model is $120.9 + 30.76 \log(d)$ dB, where d is in kilometers [8]. Each

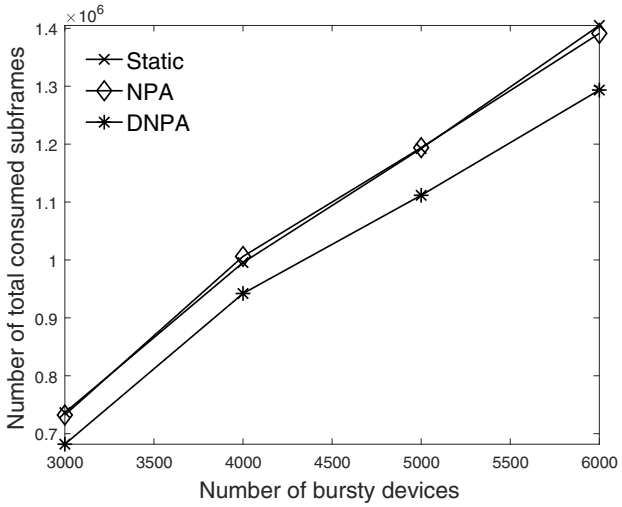

Fig. 2. The impacts of the number of devices on the number of total consumed subframes.

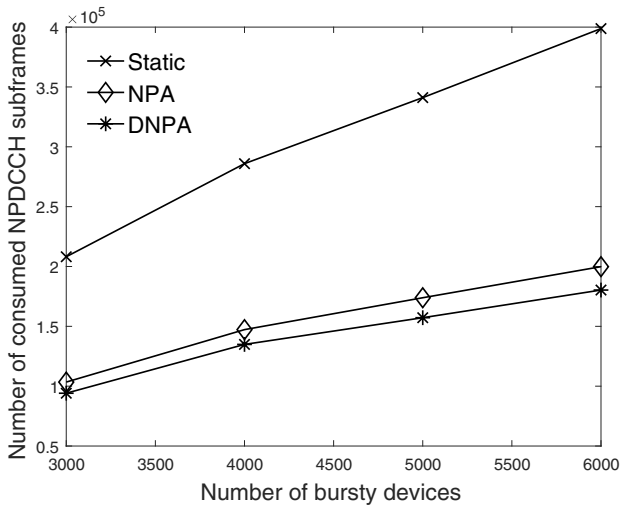

Fig. 3. The impacts of the number of devices on the number of consumed NPDCCH subframes.

device is randomly placed at a distance d between [500, 4000], [4001, 7500], and [75001, 12000] meters from the base station for CE level of 0, 1, and 2, respectively. Each device's signal-to-noise ratio can be derived based on the path loss model, the used resource unit, and its distance d.

The repetition number for uplink data can be selected from: 1, 2, 4, 8, 16, 32, 64, or 128. The repetition number for downlink data can be selected from one of the values 1, 2, 4, 8, 16, 32, 64, 128, 192, 256, 384, 512, 768, 1024, 1536, or 2048. The repetition number for transmitting a DCI can be selected from $R_{max}/8$, $R_{max}/4$, $R_{max}/2$, and R_{max}. The set of R_{max} values is $\{8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$. The set of G values is $\{1.5, 2, 4, 8, 16, 32, 48, 64\}$.

In this simulation, we also consider the signals and NPRACHs. A narrowband physical broadcast channel occupies the subframe with index 1 of every radio frame. Narrowband primary synchronization signals use the subframe with index 6 of every radio frame. Secondary synchronization signals reside at the subframe with index 10 every 20 ms. We set three NPRACHs configured as follows: 48 subcarriers, one repetition, 320 ms periodicity; 24 subcarriers, 16 repetitions, 1280 ms periodicity; and 12 subcarriers, 256 repetitions, 2560 ms periodicity.

B. Simulation Results

Fig. 2 evaluates the effect of the number of devices on the number of total consumed subframes for the three CE levels. The total consumed subframes include uplink subframes and downlink subframes. As the number of bursty devices increases, the number of consumed subframes also increases, because more devices will have more uplink and downlink data requests. The proposed *DNPA* can select a suitable pair of R_{max} and G values for each CE level such that a given uplink and downlink resource allocation algorithm can consume fewer subframes to satisfy the uplink and downlink data requirements. The *Static* method and *NPA* have similar performance in terms of the total subframe consumption. The simulation result justifies that the NPDCCH period adjustment considering both downlink and uplink is important.

Fig. 3 investigates the effect of the number of devices on the consumed NPDCCH subframes. As shown in Fig. 3, when the number of devices increases, the number of used NPDCCH subframes increases. Because each device needs to receive a DCI in NPDCCH subframes, no matter whether an uplink or a downlink request, the base station should allocate more DCIs to schedule the devices. The decided R_{max} value will have a great impact on the NPDCCH subframe consumption. Our proposed *DNPA* consumes the fewest NPDCCH subframes, while the *Static* method consumes the most NPDCCH subframes because the *Static* method does not consider the devices' DCI repetition requirements to determine the R_{max} value for each CE level so that the R_{max} value is higher under the *Static* method. *NPA* consumes fewer NPDCCH subframes than the *Static* method because *NPA* finds the smallest R_{max} value that can satisfy the devices' DCI repetition requirements in each CE level.

Fig. 4 shows the effect of the number of devices on the consumed NPDSCH subframes. Comparing Fig. 3 and Fig. 4, although the *Static* method consumes the most NPDCCH subframes, we can see that the *Static* method consumes the fewest NPDSCH subframes. However, a low NPDSCH subframe consumption does not mean that the total consumed subframes can be minimized. The proposed algorithm can still minimize the total subframe consumption as shown in Fig. 2. In terms of the total downlink subframe consumption including the NPDCCH and NPDSCH subframes, *NPA* can still outperform the *Static* method. This simulation result finds an interesting phenomenon, namely that there exists a trade-off between the NPDCCH and NPDSCH subframe consumption in determining R_{max} and G values. Therefore, we should balance the NPDCCH and NPDSCH subframes according to the repetition and data requirements of devices.

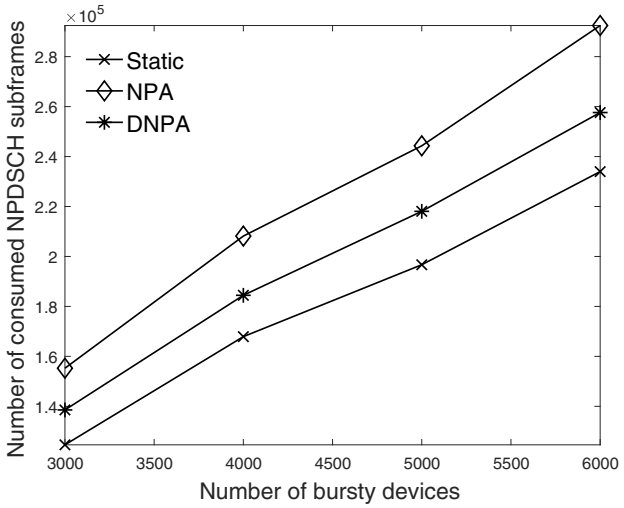

Fig. 4. The impacts of the number of devices on the number of consumed NPDSCH subframes.

V. Conclusion

In this paper, we have studied the NPDCCH period adjustment problem considering uplink and downlink frame structures for NB-IoT cellular networks. The objective is to minimize the number of consumed subframes for satisfying the data requirements of both uplink and downlink. We formulate the target problem as an optimization problem. To solve the problem, we first create the reinforcement learning settings and then propose a DQN-based NPDCCH period adjustment algorithm. The simulation results demonstrate that, compared to a previous NPDCCH period adaptation approach and a static algorithm, the proposed algorithm is very effective in reducing the radio resource consumption, especially for NPDCCH subframes. The simulation results also reveal another phenomenon, specifically that there exists a trade-off between the NPDCCH and NPDSCH subframe consumption in determining R_{max} and G values.

Acknowledgement

The work of Ya-Ju Yu was supported by the Ministry of Science and Technology, Taiwan, under Grants 107-2218-E-390-003-MY3 and 110-2221-E-390-010-MY3. The work of Ching-Chih Chuang was supported by National Pingtung University under Grant C11026 and by MOXA Networking Co., Ltd.

References

[1] Huawei, "NB-IoT - Enabling New Business Opportunities," 2015. [Online]. Available: https://www.huawei.com/minisite/4-5g/img/NB-IOT.pdf

[2] S. Landstrom, J. Bergstrom, and E. Westerberg, and D. Hammarwall, "NB-IoT: A Sustainable Technology for Connecting Billions of Devices," *Ericsson Technology Review*, vol. 93, no. 3, pp. 1–11, 2016.

[3] 3GPP TS 36.213 V16.0.0 Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures, 2020.

[4] A. Ijaz et al., "Enabling Massive IoT in 5G and Beyond Systems: PHY Radio Frame Design Considerations," *IEEE Access*, vol. 4, pp. 3322–3339, June 2016.

[5] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2018–2023," March 2020. [Online]. Available: http://www.cisco.com/

[6] R. Atat, L. Liu, H. Chen, J. Wu, H. Li, and Y. Yi, "Enabling Cyber-Physical Communication in 5G Cellular Networks: Challenges, Spatial Spectrum Sensing, and Cyber-Security," *IET Cyber-Physical Systems: Theory Applications*, vol. 2, no. 1, pp. 49–54, 2017.

[7] A. E. Mostafa, Y. Zhou, and V. W. Wong, "Connectivity Maximization for Narrowband IoT Systems with NOMA," *IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[8] J.-M. Liang, K.-R. Wu, J.-J. Chen, P.-Y. Liu, and Y.-C. Tseng, "Energy-Efficient Uplink Resource Units Scheduling for Ultra-Reliable Communications in NB-IoT Networks," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–17, July 2018.

[9] Y.-J. Yu and J.-K. Wang, "NPRACH-Aware Link Adaptation and Uplink Resource Allocation in NB-IoT Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4894–4906, May 2021.

[10] L. Feltrin, G. Tsoukaneri, M. Condoluci, C. Buratti, T. Mahmoodi, M. Dohler, and R. Verdone, "Narrowband IoT: A Survey on Downlink and Uplink Perspectives," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 78–86, Feb. 2019.

[11] B.-Z. Hsieh, Y.-H. Chao, R.-G. Cheng, and N. Nikaein, "Design of a UE-specific Uplink Scheduler for Narrowband Internet-of-Things (NB-IoT) Systems," *IEEE International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, pp. 1–5, April 2018.

[12] C.-W. Huang, S.-C. Tseng, P. Lin, and Y. Kawamoto, "Radio Resource Scheduling for Narrowband Internet of Things Systems: A Performance Study," *IEEE Network*, vol. 33, no. 3, pp. 108–115, June 2019.

[13] Y.-J. Yu, "NPDCCH Period Adaptation and Downlink Scheduling for NB-IoT Networks," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 962–975, Jan. 2021.

[14] N. Jiang, Y. Deng, A. Nallanathan, and J. A. Chambers, "Deep Reinforcement Learning for Real-Time Optimization in NB-IoT Networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1424–1440, June 2019.

[15] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529–533, Feb. 2015.

[16] M. Tokic and G. Palm, "Value-Difference based Exploration: Adaptive Control Between Epsilon-Greedy and Softmax," *Annual German Conference on Advances in Artificial Intelligence*, p. 335–346, Oct. 2011.

[17] M Tokic and G Palm, "Value-Difference based Exploration: Adaptive Control Between Epsilon-Greedy and Softmax," *Annual Conference on Artificial Intelligence*, pp. 335–346, 2011.

[18] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the Gradient by a Running Average of its Recent Magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 4, no. 2, p. 26–31, Oct. 2012.

[19] 3GPP TS 36.331 V16.0.0 Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification, 2020.