2018 IEEE 29th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)

Energy-Efficient Uplink Scheduling for Ultra-Reliable Communications in NB-IoT Networks

*Pei-Yi Liu[1], Kun-Ru Wu[1], Jia-Ming Liang[2], Jen-Jee Chen[3], and Yu-Chee Tseng[1]
[1]Department of Computer Science, National Chiao Tung University, Taiwan
[2]Department of Computer Science and Information Engineering, Chang Gung University, Taiwan
[3]Department of Electrical Engineering, National University of Tainan, Taiwan

***Abstract*—The *3GPP Narrowband Internet of Thing (NB-IoT)* is the promising technology that can provide multiple types of *resource unit (RU)* with a special repetition mechanism to improve the scheduling flexibility and transmission reliability. Since the IoT devices need to operate for a very long time, the energy consumption becomes a critical issue. In this paper, we study how to guarantee the *quality of service (QoS)* while minimizing the energy consumption for IoT devices. We first model the problem and then propose an energy-efficient scheme, which consists of two stages. The first stage tries to incur the lowest energy consumption of devices and satisfy their QoS requirement. The second stage determines the scheduling order to ensure the delay constraint while maintaining energy efficiency. Simulation results show that our scheme can serve more devices while saving their energy.**

***Index Terms*—energy saving, NB-IoT, repetition mechanism, resource unit, uplink scheduling**

I. INTRODUCTION

The *3rd generation partnership project (3GPP)* has developed a new technology, *Narrowband Internet of Thing (NB-IoT)* [1], as the communication standard for IoT, which supports massive connectivity and enhance the benefit of spectrum reuse. Specifically, it supports multiple types of *Resource Unit (RU)* with specific repetitions for data transmission to improve the scheduling flexibility and enhance communication reliability. Since IoT devices need to operate for a very long time [2], energy consumption becomes a key issue. Besides, the reliability of transmission is also a key issue in QoS for uplink transmission especially for mission critical applications, voice applications, and the high timing precision factory automation.

In this paper, we study how to ensure the strict QoS for devices based on the RU scheduling and transmission repetition while minimizing their energy consumption. We first model the problem. Then, we propose an energy-efficient heuristic, which consists of two stages. The first stage tries to conduct the lowest energy consumption and ensure QoS requirements for uplink transmission. The second stage determines the precise scheduling order of uplink requests based on the scheduling emergency and inflexibility. Simulation results show that our scheme can enlarge the number of serving devices while saving energy.

II. RELATED WORK

In the literature, the reference [3] proposes a new procedure for cell search and initial synchronization in NB-IoT which can speed up the access operation for the devices with low SNR. In [4], it proposes a new channel equalization algorithm to optimize the sampling rate of devices when NB-IoT and LTE share the same spectrum. However, they neglect the QoS and reliability of transmissions. The research [5] leverages the *Modulation and Coding Scheme (MCS)* and repetition number to enhance the QoS satisfaction and transmission latency. However, it does not leverage different types of RUs; thus, it will reduce the service coverage of NB-IoT and cannot allocate resource flexibly and effectively. In [6], the authors develop a new detection mechanism for random access procedure to enhance the coverage and access efficiency of NB-IoT. However, it does not consider the transmission reliability and energy efficiency. In [7], it leverages *Non-Orthogonal Multiple Access (NOMA)* to allocate common subcarriers to multiple devices and thus to enhance the spectrum efficiency. However, it does not discuss how to ensure the energy efficiency and transmission reliability, which are the key issues in NB-IoT.

III. PRELIMINARY

A. Resource Unit (RU)

In NB-IoT, the resource is divided into *frames*, where each frame consists of 10 *subframes*. The length of a subframe is 1 ms, which is further divided into two *slots*. For the uplink transmission, data is transmitted through *Narrowband Uplink Shared Channel (NPUSCH)*. The *resource unit (RU)* is the basic transmission resource unit allocated in the bandwidth of 180 KHz. The transmission data can be carried by one or multiple RUs depending on the request size and MCS. Specifically, NB-IoT supports multiple types of RUs based on the subcarrier spacing as shown in Table I. Since the subcarrier spacing of 15 KHz is mandatory in the standard, we focus on it in this paper. For the subcarrier spacing of 15 KHz, there are 4 types of RUs that are classified as *single-tone* (1-tone) or *multi-tone* (3-tone, 6-tone and 12-tone). Each type of RU is with a specific number of subcarriers and time slots, as shown in Fig. 1.

B. Repetition Mechanism

In NB-IoT, one of the key features is the repetition mechanism, which is designed to enhance the reliability of transmission and enlarge the network coverage. According to the NB-IoT standard, the transmission RUs

TABLE I
THE TYPES OF RESOURCE UNITS (RUS) SUPPORTED IN NB-IOT.

Subcarrier spacing	*Number of tones (subcarriers)*	*Classification*	*Number of slots*
15 KHz	1	single-tone	16
	3	multi-tone	8
	6		4
	12		2
3.75 KHz	1	single-tone	16

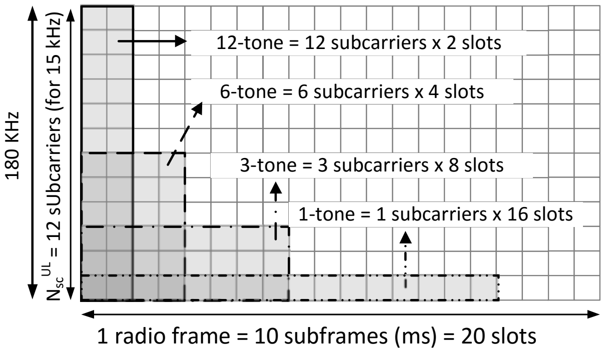

Fig. 1. Multiple types of RUs.

of each device i (also called *user equipment i* (UE_i)) in the following) can associate with a specific number of repetition $N_i^{rep} = 2^l$, where $l \in \{0..7\}$. Thus, each UE can ensure the transmission reliability based on the individual physical status, such as channel quality, path loss, bit-error-rate (BER), and transmission power.

C. Downlink Control Information (DCI)

Downlink Control Information (DCI) is the control message in NPDCCH, which is responsible for describing the scheduling results for both downlink and uplink transmissions. Each DCI is with the length of 1 ms. When the eNB completes the RU scheduling, each scheduling result will be carried by one DCI to inform the corresponding UE about its uplink transmission with the designate RU type, subcarrier set, allocation time, and the number of repetitions. Table II shows the main parameters in DCI (format N0), which is for *uplink grant* and *scheduling* in NPUSCH. Specifically, the *subcarrier indication* (I_i^{sc}) describes the RU type and the corresponding *subcarrier set* to locate RUs. The *resource assignment* (N_i^{RU}) represents the number of allocated continuous RUs for this transmission schedule excluding repetition. The *modulation and coding scheme* (MCS_i) means which MCS is applied on this RU transmission. Note that NB-IoT supports 11 types of modulation and coding schemes for uplink, which depend on the bit-error-rate and received signal-to-noise ratio (this will be clear later on). The *repetition number* (N_i^{rep}) represents the number of repetitions for the scheduled RUs. So, the total amount of RUs assigned to UE_i is $N_i^{RU} \times N_i^{rep}$.

Specifically, subcarrier indication ($I_i^{sc} \in \{0 \sim 63\}$) is used for the description of RU types and their subcarrier set, as shown in Table III. When the subcarrier spacing is 15 KHz, $I_i^{sc} \in \{0 \sim 11\}$ represents that the RU type is single-tone and locates at the subcarrier set of $S_i^{sc} = I_i^{sc}$. Thus, it has 12 possible locations. When $I_i^{sc} \in \{12 \sim 15\}$, the RU type is 3-tone and locates at $S_i^{sc} = 3(I_i^{sc} - 12) + \{0,1,2\}$, which has 4 possible locations, i.e., $S_i^{sc} \in \{\{0,1,2\},\{3,4,5\},\{6,7,8\},\{9,10,11\}\}$. When $I_i^{sc} \in \{16 \sim 17\}$, it indicates the RU type of 6-tone, which has 2 possible locations, i.e., $S_i^{sc} \in \{\{0,1,2,3,4,5\},\{6,7,8,9,10,11\}\}$. Finally, when $I_i^{sc} = 18$, the RU type is 12-tone, which has a unique location, i.e., $S_i^{sc} = \{\{0,1,2,3,4,5,6,7,8,9,10,11\}\}$ and thus $|S_i^{sc}| = 1$.

TABLE II
MAIN PARAMETERS OF DCI (FORMAT N0).

Parameter	*Value*
subcarrier indication (I_i^{sc})	$0 \sim 63$
resource assignment (N_i^{RU})	$0 \sim 7$
modulation and coding scheme (MCS_i)	$0 \sim 10$
repetition number (N_i^{rep})	2^l, $l \in \{0..7\}$

TABLE III
SUBCARRIER INDICATION AND THE CORRESPONDING SUBCARRIER SETS.

Subcarrier indication (I_i^{sc})	*Set of Allocated subcarriers* (S_i^{sc})
0-11	I_i^{sc}
12-15	$3(I_i^{sc} - 12) + \{0,1,2\}$
16-17	$6(I_i^{sc} - 16) + \{0,1,2,3,4,5\}$
18	$\{0,1,2,3,4,5,6,7,8,9,10,11\}$
19-63	reserved

D. Problem Definition

In this paper, we consider an NB-IoT network with a base station (eNB) serving *N* UEs. Each UE_i, $i = 1..N$, has an uplink request with data size $D_i \geq 0$ (bits), required reliability $R_i \in [0,1]$, and strict delay constraint d_i (ms). To guarantee QoS, assume that the arrival time of the UE_i's request is at T_i^{req}-th (ms), then the data must be uploaded to the base station before the delay deadline $T_i^{req} + d_i$. For each UE_i, the transmit power is denoted as P_i (mW) which is constrained by the maximum transmit power P_i^{max}.

When scheduling, each UE_i has to be assigned one type of RUs, $N_i^{sc} \in \{1,3,6,12\}$, according to the designate subcarrier indication $I_i^{sc} \in \{0 \sim 18\}$, i.e.,

$$N_i^{sc} = \begin{cases} 1 & ,\text{if} \quad 0 \leq I_i^{sc} \leq 11 \\ 3 & ,\text{if} \quad 12 \leq I_i^{sc} \leq 15 \\ 6 & ,\text{if} \quad 16 \leq I_i^{sc} \leq 17 \\ 12 & ,\text{if} \quad I_i^{sc} = 18. \end{cases} \tag{1}$$

For each UE_i's RUs, the amount of data that UE_i can carry depends on the modulation and coding scheme $MCS_i \in \{0..10\}$. Specifically, the bit-error-rate of the data received by the base station relies on the *received signal-to-noise ratio* $SNR_{dB}(i)$, i.e., $SNR_{dB}(i) = 10\log_{10}\left(\frac{\frac{\widetilde{P}(P_i)}{N_i^{sc}}}{BN_0+I}\right) \geq SNR_{dB}^{Req}(MCS_i, BER_i)$, where $\widetilde{\boldsymbol{P}}(P_i) = G_i G_{eNB} P_i / L(i, eNB)$ is the received power at base station; G_i, G_{eNB}, and $L(i, eNB)$ are the transmitter gain, receiver gain, and the path loss between UE_i and the eNB, respectively; *B* is the subcarrier bandwidth, i.e., 15 KHz, N_0 is the noise power, I is the interference perceived at the eNB. Note that $SNR^{Req}_{dB}(MCS_i, BER_i)$ is the SNR threshold to apply MCS_i with the measured *bit-error-rate* (BER_i).

According to Table I, the number of required RUs (N^{RU}_i) for each UE_i is

$$N^{RU}_i = \begin{cases} \left\lceil \frac{D_i}{r(MCS_i)\times 16} \right\rceil & \text{, if } N^{sc}_i = 1 \\ \\ \left\lceil \frac{D_i}{r(MCS_i)\times 24} \right\rceil & \text{, otherwise} \end{cases}, \tag{2}$$

where *r*(MCS_i) is the data rate of MCS_i (bits per subcarrier × slot). To guarantee the transmission reliability R_i, we have to leverage the number of repetitions N^{rep}_i and the successful probability of data transmission P^s_i, i.e., $1-(1-P^s_i)^{N^{rep}_i} \geq R_i$, where $P^s_i = (1-BER_i)^{D_i}$ is the successful probability if data D_i is transmitted one time and $1-(1-P^s_i)^{N^{rep}_i}$ is the successful probability after N^{rep}_i repetitions. Thus, to ensure the reliability requirement R_i of D_i, this is the necessary requirement.

Note that the scheduling results will be carried by the DCI message, which is scheduled at T^{DCI}_i (subframe) for each UE_i. Thus, it has to satisfy the delay deadline of UE_i, i.e., $T^{DCI}_i + (N^{RU}_i \times \frac{N^{slot}_i}{2} \times N^{rep}_i) \leq T^{req}_i + d_i$, where N^{slot}_i is the number of slots of single RU (two slots constitutes one ms), which depends on the RU type, i.e.,

$$N^{slot}_i = \begin{cases} 16 & \text{, if} \quad N^{sc}_i = 1 \\ 8 & \text{, if} \quad N^{sc}_i = 3 \\ 4 & \text{, if} \quad N^{sc}_i = 6 \\ 2 & \text{, if} \quad N^{sc}_i = 12. \end{cases} \tag{3}$$

Now, we consider the current scheduling subframe is T^s (ms), the feasible subcarrier set is K (e.g., $|K| = 12$ if subcarrier spacing is 15 KHz), and the available earliest subframe for each subcarrier k to allocate resource to devices is S_k, $k = 1..|K|$. Our problem asks how to optimize the uplink scheduling results for each $UE_i, i = 1..N$, and satisfy the strict QoS while minimizing the total energy consumption.

IV. THE PROPOSED SCHEME

In this section, we introduce an energy-efficient scheme, which consists of two stages. The first stage will determine the scheduling parameters of UEs by quantifying the consumed energy for each UE and then choosing the one with minimal energy cost while satisfying the required reliability. The second stage will determine the scheduling order by considering the emergency level of requests and inflexibility of the scheduling transmission. The details of the scheme are described as follows.

A. Stage 1

The first stage is to determine the default parameters for each UE, including the type of RUs (N^{sc}_i), the number of RUs (N^{RU}_i), the best number of repetitions (N^{rep}_i), and transmit parameters (MCS_i and P_i), to guarantee QoS and the transmission reliability. These operations are described as follows.

Step 1. For each UE$_i, i = 1..N$, we first calculate the required number of RUs N^{RU}_i based on the available RU types and MCS selections. Specifically, the required transmit time to carry the amount of data D_i cannot be greater than the delay requirements. These results are collected as the feasible setting pairs of RU type and MCS setting for each UE$_i$, denoted as set A_i, i.e., $A_i = \{(N^{sc}_{i,j}, MCS_{i,j}) \mid N^{RU}_{i,j} \times \frac{N^{slot}_{i,j}}{2} \leq d_i, N^{sc}_{i,j} \in \{1,3,6,12\}, MCS_{i,j} \in \{0..10\}\}$, where j is the index of feasible setting pair of RU type and MCS for each UE$_i$ and $N^{slot}_{i,j}$ is the number of required slots when the RU type is $N^{sc}_{i,j}$. Note that $N^{slot}_{i,j}$ is divided by 2 because two slots constitutes 1 ms, which is the unit of delay constraint d_i.

Step 2. For each UE$_i, i = 1..N$, consider the feasible RU type and MCS setting pair $(N^{sc}_{i,j}, MCS_{i,j}) \in A_i$, we calculate the allowed repetition numbers $N^{rep}_{i,j}$ in which each $N^{rep}_{i,j,k} \in N^{rep}_{i,j}$ can make UE$_i$ not only satisfy the required reliability R_i but also ensure the corresponding transmission power $P_{i,j,k}$ in the feasible ranges, i.e., $N^{rep}_{i,j} = \{N^{rep}_{i,j,k} \mid 1-(1-P^s_{i,j,k})^{N^{rep}_{i,j,k}} \geq R_i, 0 \leq P(N^{sc}_{i,j}, MCS_{i,j}, BER_{i,j,k}) \leq P^{max}_i, N^{rep}_{i,j,k} \in \{2^l \mid l \in \{0..7\}\}, N^{RU}_{i,j} \times \frac{N^{slot}_{i,j}}{2} \times N^{rep}_{i,j,k} \leq d_i\}$, where $BER_{i,j,k} = 1 - \left(1-(1-R_i)^{\frac{1}{N^{rep}_{i,j,k}}}\right)^{\frac{1}{D_i}}$ and $P(N^{sc}_{i,j}, MCS_{i,j}, BER_{i,j,k})$ is a function which returns the minimum transmit power for the RU type $N^{sc}_{i,j}$, MCS setting $MCS_{i,j}$, and target bit-error-rate $BER_{i,j,k}$ i.e., $P(N^{sc}_{i,j}, MCS_{i,j}, BER_{i,j,k}) = 10^{\frac{SNR^{Req}_{dB}(MCS_{i,j}, BER_{i,j,k})}{10}} \times \frac{(BN_0+I)\cdot L(i,eNB)\cdot N^{sc}_{i,j}}{G_i G_{eNB}}$.

After that, we have all the feasible RU type and MCS setting pairs with each of their allowed repetition numbers $N^{rep}_{i,j}$.

Step 3. Based on the results of steps 1 and 2, we calculate the most energy-saving repetition number $N^{rep^*}_{i,j}$ for each feasible combination pair $(N^{sc}_{i,j}, MCS_{i,j}) \in A_i$, where $N^{rep^*}_{i,j} = \arg\min_{N^{rep}_{i,j,k} \in N^{rep}_{i,j}} E(N^{sc}_{i,j}, MCS_{i,j}, N^{rep}_{i,j,k})$, and $E(N^{sc}_{i,j}, MCS_{i,j}, N^{rep}_{i,j,k}) = P(N^{sc}_{i,j}, MCS_{i,j}, BER_{i,j,k}) \times N^{RU}_{i,j} \times \frac{N^{slot}_{i,j}}{2} \times N^{rep}_{i,j,k}$.

Then, reform A_i as a set of triplets $(N^{sc}_{i,j}, MCS_{i,j}, N^{rep^*}_{i,j})$. Each triplet in A_i is a feasible configuration of RU type, MCS setting, and repetition number.

Step 4. Then, we choose the best triplet of $(\boldsymbol{N^{sc^*}_i}, \boldsymbol{MCS^*_i}, \boldsymbol{N^{rep^*}_i})$ from A_i as the default parameter of UE$_i$, which incurs the minimum energy consumption.

Through the above steps, we can determine the best RU type $\boldsymbol{N^{sc^*}_i}$, MCS setting $\boldsymbol{MCS^*_i}$, and repetition number $\boldsymbol{N^{rep^*}_i}$ that can incur the least energy consumption and meet the reliability requirement $\boldsymbol{R_i}$ of each UE$_i$.

B. Stage 2

The second stage is to optimize the scheduling results of requests from UEs, including the subcarrier set of RUs (S^{sc}_i) and the start time of RUs (T^{sc}_i). In addition, if needed, it can adaptively adjust the transmission parameters of UEs to ensure the delay constraint and enhance spectrum utilization. The detailed steps are depicted as follows.

Step 1. We first define a *score function* to evaluate the emergency and inflexibility for each UE$_i$ with uplink transmission request, i.e., $Score_i = W_1 \times Em_i + W_2 \times \widetilde{Inf}_i$, where $W_1 \in [0,1]$ and $W_2 \in [0,1]$ are the weighting factors of the emergency and inflexibility, respectively, that satisfy $W_1 + W_2 = 1$. Note that Em_i is the urgent level of UE$_i$'s request compared to others, i.e., $Em_i = \frac{\max_j\{T_j^R\} - T_i^R}{\max_j\{T_j^R\} - \min_j\{T_j^R\}}$, where T_j^R is the remaining time from the scheduling subframe T^S to the delay deadline $T_j^{req} + d_j$ of UE$_j$.

$\widetilde{Inf}_i$ is the number of RU types that UE$_i$ can choose, which is defined by $\widetilde{Inf}_i = \frac{Inf_i}{\max_j\{Inf_j\}}$, where

$$Inf_i = \begin{cases} 4 & \text{, if } \Psi_{N^{sc}}(A_i) = 1 \\ 3 & \text{, if } \Psi_{N^{sc}}(A_i) = 2 \\ 2 & \text{, if } \Psi_{N^{sc}}(A_i) = 3 \\ 1 & \text{, if } \Psi_{N^{sc}}(A_i) > 3, \end{cases} \tag{4}$$

and $\Psi_{N^{sc}}(A_i)$ is the number of choices of RU types for the feasible setting pair A_i. That means if the UE$_i$ has fewer choices, its inflexibility is higher and needs to be scheduled earlier.

Now, for each UE$_i, i = 1..N$, we calculate its $Score_i$ and sort them in descending order. For the UEs without any request, define its Score as -∞. Without loss of generality, we use *List* ***L*** to represent the sorted sequence of the UEs.

Step 2. Before determining the subcarrier set of RUs, we first define a function $Waste(i, S_i^{sc})$ to reflect the potential waste of resource if UE$_i$'s RUs are allocated at subcarrier set S_i^{sc}, i.e., $Waste(i, S_i^{sc}) = \sum_{k' \in K - S_i^{sc}}((\max_{k \in S_i^{sc}}\{\widehat{S}_k\} + (N_i^{RU} \times \frac{N_i^{slot}}{2} \times N_i^{rep})) - \widehat{S}_{k'})^+ + \sum_{k \in S_i^{sc}}(\max_{k \in S_i^{sc}}\{\widehat{S}_k\} - \widehat{S}_k)$, where $(\cdot)^+ = \max\{\cdot, 0\}$ outputs the value larger than or equal to 0; $\max_{k \in S_i^{sc}}\{\widehat{S}_k\}$ means the earliest available resource allocation start time of RUs if the subcarrier set is S_i^{sc}, where $\widehat{S}_k = \max\{S_k, T_i^{DCI} + 1\}$ is to ensure allocating RU after DCI.

Then, we choose the best subcarrier set $S_i^{sc^*}$ that makes UE$_i$ have the minimal $Waste(i, S_i^{sc})$ without violating its delay deadline, i.e., $S_i^{sc^*} = \arg\min_{S_i^{sc} \subseteq \Theta(N_i^{sc^*})}\{Waste(i, S_i^{sc}) \mid \max_{k \in S_i^{sc}}\left\{\widehat{S}_k\right\} + (N_i^{RU^*} \times \frac{N_i^{slot^*}}{2} \times N_i^{rep^*}) < (T_i^{req} + d_i)\}$, where $\Theta(N_i^{sc^*})$ is the set of available subcarrier sets when default RU type $N_i^{sc^*}$ is used.

If $S_i^{sc^*} \neq \emptyset$, we set the subframe index of DCI_i by $T_i^{DCI} = T^s$ and start time $T_i^{sc} = \max_{k \in S_i^{sc^*}}\{\widehat{S}_k\}$. Then, update the available scheduling subframe for subcarriers $k \in S_i^{sc^*}$ and $k' \in \Theta(N_i^{sc^*}) - S_i^{sc^*}$ by $S_k = \max\{\max_{k \in S_i^{sc^*}}\{\widehat{S}_k\} + (N_i^{RU^*} \times \frac{N_i^{slot^*}}{2} \times N_i^{rep^*}), T_i^{DCI} + 1\}$ and $S_{k'} = \widehat{S}_{k'}$, respectively. Finally, update $T^s = T_i^{DCI} + 1$ and then remove UE$_i$ from List ***L***. However, if $S_i^{sc^*} = \emptyset$, it means that the current transmission parameter setting is infeasible, then we check whether or not UE$_i$ has other feasible triplet in A_i other than the default parameter. If yes, go to step 3 for further adjusting. If no, we remove such UE$_i$ from List ***L*** and go back to step 2 to schedule the next UE. The above steps are repeated until List ***L*** is empty and then terminate this stage.

Step 3. Here, we try to change the type of RUs and/or MCSs of UE$_i$ by referring A_i and choose the new triplet that can satisfy the delay deadline while incurring the least extra energy consumption and resource as follows.

First, we define a *cost ratio* $\boldsymbol{C_i^{\alpha,\beta}}$ to reflect the results of extra consumed energy over the extra required resource space when the original pair of RU type and MCS, denoted as $\boldsymbol{\alpha} = (\boldsymbol{N_i^{sc^*}}, \boldsymbol{MCS_i^*}, \boldsymbol{N_i^{rep^*}})$, changes to the new pair, denoted as $\boldsymbol{\beta} = (\boldsymbol{N_i^{sc'}}, \boldsymbol{MCS_i'}, \boldsymbol{N_i^{rep'}})$ for $(\boldsymbol{N_i^{sc'}}, \boldsymbol{MCS_i'}, \boldsymbol{N_i^{rep'}}) \in A_i - (\boldsymbol{N_i^{sc^*}}, \boldsymbol{MCS_i^*}, \boldsymbol{N_i^{rep^*}})$, i.e.,

$$C_i^{\alpha,\beta} = \begin{cases} \frac{\Delta E_i^{\alpha,\beta}}{\Delta Area_i^{\alpha,\beta}} & \text{, if } \Delta Area_i^{\alpha,\beta} > 0 \\ \Delta E_i^{\alpha,\beta} & \text{, if } \Delta Area_i^{\alpha,\beta} = 0 \end{cases}, \tag{5}$$

where the extra consumed energy is $\Delta E_i^{\alpha,\beta} = (E(\boldsymbol{\beta}) - E(\boldsymbol{\alpha}))^+$, and the extra resource space is $\Delta Area_i^{\alpha,\beta} = (\boldsymbol{N_i^{sc'}} \times T(\boldsymbol{\beta}) - \boldsymbol{N_i^{sc^*}} \times T(\boldsymbol{\alpha}))^+$.

Then, we choose the new pair $\boldsymbol{\beta^*}$ which incurs the minimal cost ratio and replace the default parameters by $\boldsymbol{N_i^{sc^*}} = \boldsymbol{N_i^{sc'}}$, $\boldsymbol{MCS_i^*} = \boldsymbol{MCS_i'}$ and $\boldsymbol{N_i^{rep^*}} = \boldsymbol{N_i^{rep'}}$ accordingly. Finally, go back to step 2 for further allocation.

Through the above steps, we can determine each UE$_i$'s subcarrier set S_i^{sc}, start time T_i^{sc}, and the corresponding configurations of MCS_i, N_i^{rep} and P_i while ensuring the delay deadline and reducing the waste of spectrum resource and energy.

V. SIMULATION RESULTS

In this section, we develop a simulator in C++ language to verify the efficiency of the proposed scheme. The parameters of the simulation are shown in Table IV. We compare our scheme (***Ours***) with the standard scheme (***Spec***) [1], Narrow-Band Link Adaptation scheme (***NBLA***) [5], random scheduling scheme (***Random***), and Round Robin scheme (***RR***). Note that the weighting factors of our scheme is $W_1 = 0.5$ and $W_2 = 0.5$.

TABLE IV
THE SIMULATION PARAMETERS.

Parameter	***Value***
maximum transmit power ($\boldsymbol{P_i^{max}}$)	23 dBm
antenna gain of transmitter ($\boldsymbol{G_i}$)	-4 dBi
antenna gain of receiver ($\boldsymbol{G_{eNB}}$)	18 dBi
thermal noise density ($\boldsymbol{N_0}$)	-174 dBm/Hz
path loss ($\boldsymbol{L(i, eNB)}$)	120.9+30.76log(d) dB, d in Km
distance from the base station	0 $\sim$ 15 (Km)
number of UEs (***N***)	3000 $\sim$ 30000
request data size ($\boldsymbol{D_i}$)	50 $\sim$ 200 bytes
delay constraint ($\boldsymbol{d_i}$)	50, 100, 150, 300 (ms)
required reliability ($\boldsymbol{R_i}$)	90% $\sim$ 99%

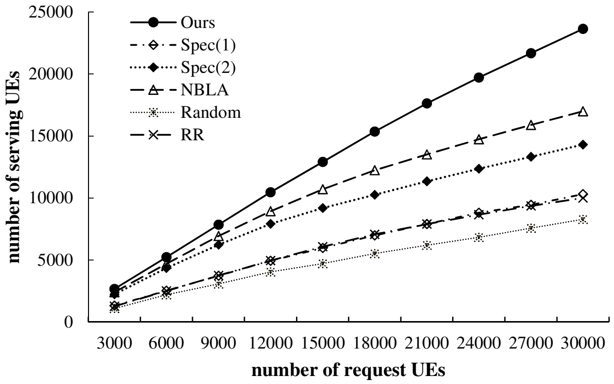

Fig. 2. Comparisons on the number of serving UEs of all schemes.

A. Number of Serving UEs

We first investigate the effects of number of request UEs on number of serving UEs. As shown in Fig. 2, similarly, Random performs the worst because it randomly schedules the UEs with a random repetition number; thus, the QoS and reliability of UEs may not be met. Spec and RR perform slightly better than Random scheme because they prefer to choose single-tone with the fixed repetition number for UEs; thus, it could potentially satisfy more UEs with small data request and lower reliability requirement. Spec(2) is better than Spec(1) because a larger repetition number can achieve higher reliability. NBLA is better than the above schemes because it can adjust the repetition levels and MCSs interactively to satisfy the transmission reliability and delay. Note that our scheme outperforms all others because our scheme can optimize the number of repetitions to satisfy the transmission reliability in stage 1 and apply the best configuration pair of RU type and MCS to ensure QoS while enhancing the spectrum utilization in stage 2.

B. Energy Consumption per UE

Finally, we investigate the effects the number of request UEs on energy consumption per UE. As shown in Fig. 3, we can see that the energy consumption per UE of all schemes increases when the number of request UEs increases. This is because the network is saturated and most satisfied UEs are with higher MCS, which require less resource but consume more energy. Random scheme performs the worst because it randomly choose the number of repetitions that may potentially increase the transmission time, thus consuming more energy. Spec and RR are better than Random scheme because they only serve the UEs with small size request which consumes less energy. NBLA performs slightly better because it can determine the number of repetitions appropriately but neglects to minimize the transmission power. Note that our scheme performs the best because our scheme can choose the best scheduling parameters of RUs with least energy consumption in stage 1, and leverage the cost ratio to reduce energy consumption in stage 2, thus saving energy more efficiently.

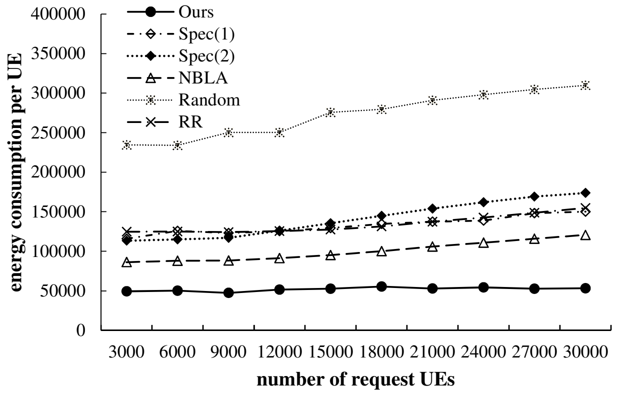

Fig. 3. Comparisons on energy consumption per UE of all schemes.

VI. CONCLUSION

In this paper, we have addressed the problem of energy saving with QoS consideration in NB-IoT networks. We first model this problem and then propose an energy-efficient scheme with two stages. The first stage chooses the default scheduling parameters with least energy consumption and the second stage serves the UEs with least potential resource waste. Simulation results have verified that our scheme can satisfy more UEs while saving their energy.

ACKNOWLEDGMENTS

This research is co-sponsored by MOST 106-2221-E-024-004-, 102-2218-E-182-008-MY3, 105-2221-E-182-051, 106-2221-E-182-015-MY3, 105-2221-E-009-100-MY3, 105-2923-E-009-001-MY2, 104-2221-E-009-113-MY3, MoE Project "Center for Open Intelligent Connectivity", MoE ATU Plan, Delta Electronics, ITRI, Institute for Information Industry, Academia Sinica AS-105-TP-A07, and Chang Gung Memorial Hospital, Taoyuan.

REFERENCES

[1] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA)," *Physical channels and modulation, v14.4.0*, pp. 1–6, Sep. 2017.

[2] J.-M. Liang, J.-J. Chen, H.-H. Cheng, and Y.-C. Tseng, "An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE-advanced networks for Internet of Things," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 1, pp. 13–22, 2013.

[3] N. Mangalvedhe, R. Ratasuk, A. Ghosh, "NB-IoT deployment study for low power wide area cellular IoT," *IEEE International Conference on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2016.

[4] L. Zhang, A. Ijaz, P. Xiao, R. Tafazolli, "Channel equalization and interference analysis for uplink narrowband Internet of Things (NB-IoT)," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2206–2209, 2017.

[5] C. Yu, L. Yu, Y. Wu, Y. He, Q. Lu, "Uplink scheduling and link adaptation for narrowband Internet of Things systems," *IEEE Access*, vol. 5, pp. 1724–1734, 2017.

[6] X. Lin, A. Adhikary, Y-P. Eric Wang, "Random access preamble design and detection for 3GPP narrowband IoT systems," *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 640–643, 2016.

[7] A. E. Mostafa, Y. Zhou, Vincent W. S. Wong, "Connectivity maximization for narrowband IoT systems with NOMA," *IEEE International Conference on Communications (ICC)*, pp. 1–6, 2017.