WILEY | Hindawi

Research Article

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

Jia-Ming Liang (iD),[1,2] **Kun-Ru Wu,**[3] **Jen-Jee Chen** (iD),[4] **Pei-Yi Liu,**[3] **and Yu-Chee Tseng**[3,5]

[1]*Department of Computer Science and Information Engineering, Chang Gung University, Kweishan, Taoyuan 33378, Taiwan*
[2]*Department of General Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan*
[3]*Department of Computer Science, National Chiao Tung University, HsinChu 30010, Taiwan*
[4]*Department of Electrical Engineering, National University of Tainan, Tainan, Taiwan*
[5]*Research Center for Information Technology Innovation, Academia Sinica, Taipei 11574, Taiwan*

Correspondence should be addressed to Jen-Jee Chen; jjchen@mail.nutn.edu.tw

For 5G wireless communications, the *3GPP Narrowband Internet of Things (NB-IoT)* is one of the most promising technologies, which provides multiple types of *resource unit (RU)* with a special repetition mechanism to improve the scheduling flexibility and enhance the coverage and transmission reliability. Besides, NB-IoT supports different operation modes to reuse the spectrum of LTE and GSM, which can make use of bandwidth more efficiently. The IoT application grows rapidly; however, those massive IoT devices need to operate for a very long time. Thus, the energy consumption becomes a critical issue. Therefore, NB-IoT provides discontinuous reception operation to save devices' energy. But, how to further reduce the transmission energy while ensuring the required ultra-reliability is still an open issue. In this paper, we study how to guarantee the reliable communication and satisfy the *quality of service (QoS)* while minimizing the energy consumption for IoT devices. We first model the problem as an optimization problem and prove it to be NP-complete. Then, we propose an energy-efficient, ultra-reliable, and low-complexity scheme, which consists of two phases. The first phase tries to optimize the default transmit configurations of devices which incur the lowest energy consumption and satisfy the QoS requirement. The second phase leverages a weighting strategy to balance the emergency and inflexibility for determining the scheduling order to ensure the delay constraint while maintaining energy efficiency. Extensive simulation results show that our scheme can serve more devices with guaranteed QoS while saving their energy effectively.

1. Introduction

The *Internet of Things (IoT)* is one of the key applications and technologies in the *fifth-generation (5G)* communications. Since IoT is widely used for remote monitoring and reporting, such as smart building, smart transportation, smart grid, e-health, and/or factory automation, it makes our life more convenient and makes industry more efficient. Therefore, the *3rd generation partnership project (3GPP)* develops a new technology, *Narrowband Internet of Things (NB-IoT)* [1, 2], as the communication standard for IoT, which is featured by low cost, low energy consumption, low complexity, and low throughput. Besides, the NB-IoT supports massive connectivity and enhances the benefit of spectrum reuse. Specifically, it supports multiple types of *resource units (RU)*

with specific repetitions for data transmission to improve the scheduling flexibility and enhance communication reliability. In addition, NB-IoT also provides three-operation modes to flexibly reuse the spectrum of LTE and GSM, which can achieve higher spectrum utilization and reduce the extra deployment cost for the operators.

On the other hand, due to the inherent behaviors of IoT applications, such as remote monitoring and reporting, IoT devices need to operate for a very long time [3]. Thus, energy consumption becomes a key issue. Currently, NB-IoT provides discontinuous reception to save devices' energy based on wake-up and sleep operation. However, how to further decrease their transmission energy during wake-up period is an open problem. In addition, the reliability of transmission is also a key issue in QoS for uplink transmission
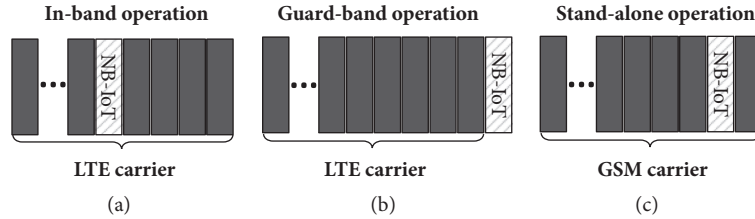
FIGURE 1: Three-operation modes of NB-IoT.

especially for mission critical applications (e.g., e-health, intelligent transport), voice applications, and the high timing precision factory automation [4–6]. Thus, NB-IoT provides the maximum repetition time up to 128 for a scheduling resource allocation to the ultra-reliability issue. But, how to optimize the repetition operation to achieve high reliability while reducing the waste of resource and energy is still an open issue.

In this paper, we study how to ensure the high transmission reliability to guarantee the strict QoS for devices based on the RU scheduling and repetition determination while minimizing their energy consumption in NB-IoT networks. We first model the problem as an optimization problem and prove it to be NP-complete. Then, we propose an energy-efficient and ultra-reliable heuristic, which consists of two phases. The first phase tries to select the primary parameters which conduct the lowest energy consumption and ensure QoS requirements for uplink transmission. The second phase applies a weighting strategy to determine the precise scheduling order of uplink requests based on the scheduling emergency and inflexibility. In addition, it also adjusts the corresponding results appropriately to satisfy the strict delay constraint if needed while considering energy efficiency. Extensive simulation results show that our scheme can enlarge the number of serving devices with guaranteed QoS and decrease the packet drop ratio while saving energy.

The rest of this paper is organized as follows. Related work is discussed in Section 2. Preliminaries are given in Section 3. Section 4 presents our scheme. Simulation results are shown in Section 5. Section 6 concludes the paper.

2. Related Work

In the literature, the studies [7, 9–11] give an overview of NB-IoT and conclude that NB-IoT can enhance bandwidth efficiency and increase network coverage. Reference [12] 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 [13], 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 [14] 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 [15], 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. The study [16] proposes a transmission scheme without connection setup to reduce connectivity latency. But, it may cause extra energy consumption. Reference [17] develops a detection scheme based on maximum likelihood to detect timing acquisition with low delay while reducing energy consumption. However, the QoS satisfaction and the transmission reliability are ignored in this paper. In [18], the authors develop a prediction method to allocate resource in advance based on the occurrence and delay time of uplink transmission. Although it can accelerate the transmission procedure, it may decrease the scheduling flexibility and resource efficiency. In [19], it leverages *Nonorthogonal 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.

Based on the above observation, it motivates us to address the issue of considering both transmission reliability and energy efficiency by scheduling multitypes of RUs with optimal repetition in NB-IoT networks.

3. Preliminary

In this section, we first give an overview of the operation modes of NB-IoT. Then, we introduce the resource unit and the repetition mechanism used in NB-IoT. Finally, we formally define our resource allocation problem and show it to be NP-complete.

3.1. NB-IoT Operation Modes. In NB-IoT, all devices connect with the centralized base station (also called the *Evolved Node B, eNB*). In order to enhance the spectrum utilization and reduce the cost of operators, NB-IoT provides three-operation modes for devices to access the eNB by reusing the existing spectrum of LTE and GSM [20–22]:

3.1.1. Inband. Using the bandwidth of one *resource block (RB)* inside the LTE carrier as the access spectrum is shown in Figure 1(a).

3.1.2. Guard-Band. Using the bandwidth of one RB in the guard-band of LTE carrier as the access spectrum is shown in Figure 1(b).
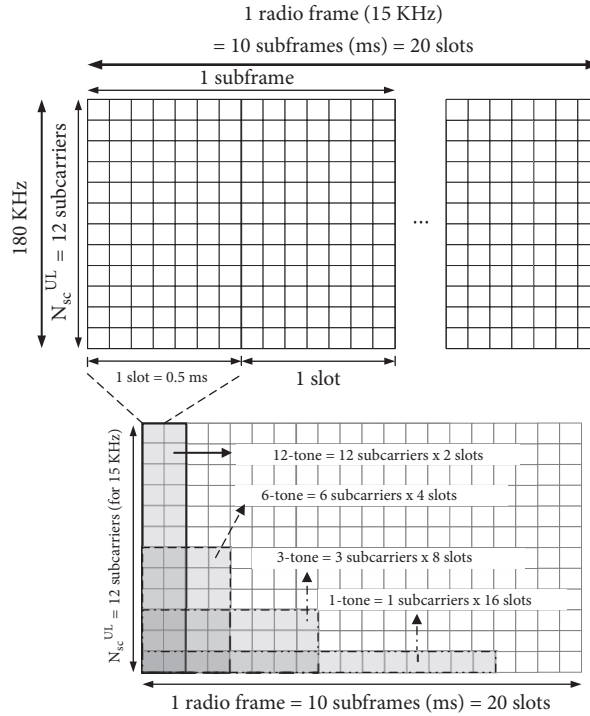
FIGURE 2: Multiple types of RUs.

TABLE 1: 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
15 KHz	3	multi-tone	8
	6		4
	12		2
3.75 KHz	1	single-tone	16

3.1.3. Stand-Alone.
Using the bandwidth of a reframed GSM carrier as the access spectrum is shown in Figure 1(c).

3.2. 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 1. Since the subcarrier spacing of 15 KHz is mandatory in the standard, we focus on it in this paper [1]. For the subcarrier spacing of 15 KHz, there are 4 types of RUs that are classified as *single-tone* (1-tone) or *multitone* (3-tone, 6-tone, and 12-tone). Each type of RU is with a specific number of subcarriers and time slots, as shown in Figure 2.

3.3. 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 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 $l \in \{0 \cdots 7\}$. 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. Figure 3 shows an example of 5 UEs with different repetition numbers. Note that the transmission finish time of each UE should not violate its delay constraint.

3.4. Downlink Control Information (DCI).
Downlink Control Information (DCI) is the control message in *Narrowband Physical Downlink Control Channel (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 2 shows the main parameters
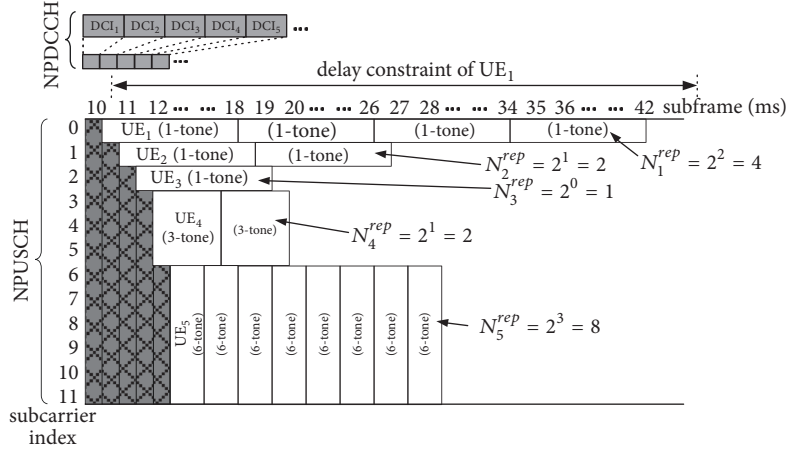
FIGURE 3: The example of repetitive transmissions.

TABLE 2: Main parameters of DCI (format N0).

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

TABLE 3: 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

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 3. When the subcarrier spacing is 15 KHz, $I_i^{sc} \in \{0 \sim 11\}$ represents the fact 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 $|S_i^{sc}| = 1$.

3.5. Problem Definition. In this paper, we consider an NB-IoT network with a base station (eNB) serving N UEs. Each UE$_i$, $i = 1 \cdots N$, 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) and 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}; i.e.,

$$0 \leq P_i \leq P_i^{max}. \tag{1}$$

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.,

$$N_i^{sc} = \begin{cases} 1, & \text{if } 0 \leq I_i^{sc} \leq 11 \\ 3, & \text{if } 12 \leq I_i^{sc} \leq 15 \\ 6, & \text{if } 16 \leq I_i^{sc} \leq 17 \\ 12, & \text{if } I_i^{sc} = 18. \end{cases} \tag{2}$$

For each UE$_i$'s RUs, the amount of data that UE$_i$ can carry depends on the modulation and coding scheme MCS$_i \in \{0 \cdots 10\}$. 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.,

$$\text{SNR}_{dB}(i) = 10 \log_{10} \left(\frac{\tilde{P}(P_i)/N_i^{sc}}{BN_0 + I} \right) \tag{3}$$

$$\geq \text{SNR}_{dB}^{Req}(\text{MCS}_i, \text{BER}_i),$$

where $\widetilde{\mathbf{P}}(P_i) = G_i G_{\text{eNB}} P_i / L(i, \text{eNB})$ is the received power at base station and G_i, G_{eNB}, and $L(i, \text{eNB})$ 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 and I is the interference perceived at the eNB. Note that

$\text{SNR}_{\text{dB}}^{\text{Req}}(\text{MCS}_i, \text{BER}_i)$ is the SNR threshold to apply MCS_i with the measured *bit-error-rate* (BER_i).

According to Table 1, the number of required RUs (N_i^{RU}) for each UE_i is

$$N_i^{\text{RU}} = \begin{cases} \left\lceil \dfrac{D_i}{r(\text{MCS}_i) \times 16} \right\rceil, & \text{if } N_i^{\text{sc}} = 1 \\[2ex] \left\lceil \dfrac{D_i}{r(\text{MCS}_i) \times 24} \right\rceil, & \text{otherwise,} \end{cases} \tag{4}$$

where $r(\text{MCS}_i)$ is the data rate of MCS_i (bits per subcarrier $\times$ slot). To guarantee the transmission reliability R_i, we have to leverage the number of repetitions N_i^{rep} and the successful probability of data transmission P_i^s; i.e.,

$$1 - \left(1 - P_i^s\right)^{N_i^{\text{rep}}} \geq R_i, \tag{5}$$

where $P_i^s = (1 - \text{BER}_i)^{D_i}$ is the successful probability [23, 24] if data D_i is transmitted one time and $1 - (1 - P_i^s)^{N_i^{\text{rep}}}$ is the successful probability after N_i^{rep} repetitions. Thus, to ensure the reliability requirement R_i of D_i, Equation (5) is the necessary requirement.

Note that the scheduling results will be carried by the DCI message, which is scheduled at T_i^{DCI} (subframe) for each UE_i. Thus, it has to satisfy the delay deadline of UE_i; i.e.,

$$T_i^{\text{DCI}} + \left(N_i^{\text{RU}} \times \frac{N_i^{\text{slot}}}{2} \times N_i^{\text{rep}}\right) \leq T_i^{\text{req}} + d_i, \tag{6}$$

where N_i^{slot} is the number of slots of single RU (two slots constitutes one ms), which depends on the RU type; i.e.,

$$N_i^{\text{slot}} = \begin{cases} 16, & \text{if } N_i^{\text{sc}} = 1 \\ 8, & \text{if } N_i^{\text{sc}} = 3 \\ 4, & \text{if } N_i^{\text{sc}} = 6 \\ 2, & \text{if } N_i^{\text{sc}} = 12. \end{cases} \tag{7}$$

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, $k = 1 \cdots |K|$. Our problem asks how to optimize the uplink scheduling results for each UE_i, $i = 1 \cdots N$, by determining (1) the type of RUs (N_i^{sc}), (2) the number of RUs (N_i^{RU}), (3) the subcarrier set of RUs (S_i^{sc}), and (4) the allocation start time of RUs (T_i^{sc}) with (5) the number of repetitions (N_i^{rep}), (6) transmit power of UE_i (P_i), (7) the modulation and coding scheme (MCS_i), and (8) the subframe index of DCI (T_i^{DCI}) to ensure that no two RUs overlap with

each other and the QoS parameters including the request data size (D_i), delay constraint (d_i), and reliability of transmission (R_i) are satisfied while the total energy consumption of UEs, denoted as $\sum_{i=1 \cdots N} E_i$, is minimized, where

$$E_i = P_i \times \left(N_i^{\text{RU}} \times N_i^{\text{slot}} \times N_i^{\text{rep}}\right). \tag{8}$$

Therefore, our problem can be summarized as an optimization-like problem:

$$\min_{P_i, \text{MCS}_i, N_i^{\text{RU}}, I_i^{\text{sc}}, N_i^{\text{rep}}} \sum_{i=1 \cdots N} E_i, \tag{9}$$

subject to (1), (2), (3), (4) (5), (6), and (7).

Table 4 also summarizes the notations used in this paper.

Theorem 1. *The addressed problem is NP-complete.*

Proof. To simplify the proof, we consider the case of subcarrier spacing with 3.75 KHz where the UEs use the single-tone only and the modulation and coding scheme (MCS) is monotonic. So, the number of repetitions with minimal transmission power to meet required reliability of each UE is unique. Thus, the *energy cost* of an UE on each parameter list is also uniquely determined. Then, we formulate the resource allocation problem as a decision problem: *energy-efficient ultra-reliable scheduling decision (EUSD)* problem. Given a NB-IoT network and the UEs with required reliability for uplink transmission, we ask whether or not there exists a set of numbers of repetitions S^{Rep} such that all UEs can conserve the total energy of $\mathbf{Q}$ to satisfy their uplink transmission with reliability. Then, we show that EUSD problem is NP-complete.

We first show that the EUSD problem belongs to NP. Given a problem instance and a solution containing the set of repetition numbers it can be verified whether or not the solution is valid in polynomial time. Thus, this part is proved.

We then reduce the *multiple-choice knapsack (MCK)* problem [25], which is known to be NP-complete, to the EUSD problem. Consider that there are $\boldsymbol{n}$ disjointed classes of objects, where each class $\boldsymbol{i}$ contains $\mathbf{N_i}$ objects. In each class $\boldsymbol{i}$, every object $\mathbf{x_{i,j}}$ has a profit $\mathbf{q_{i,j}}$ and a weight $\mathbf{u_{i,j}}$. Besides, there is a knapsack with capacity of $\boldsymbol{U}$. The MCK problem asks whether or not we can select exact one object from each class such that the total object weight is no larger than $\boldsymbol{U}$ and the total object profit is $\boldsymbol{Q}$.

We then construct an instance of the EUSD problem as follows. Let $\boldsymbol{n}$ be the number of UEs. Each UE_i has $\mathbf{N_i}$ repetition selections to transmit data to the eNB. When UE_i selects the repetition number $\mathbf{x_{i,j}}$, it will conserve energy of $\mathbf{q_{i,j}}$. Note that the conserved energy of an UE with a particular number of repetition is compared to the same UE's number with the most energy cost. Thus, the system should allocate a RU size of $\mathbf{u_{i,j}}$ to transmit UE_i's data to the eNB. The total frame space is $\boldsymbol{U}$. Our goal is to let all UEs conserve energy of $\boldsymbol{Q}$ and satisfy their transmission requirement. We show that the MCK problem has a solution if and only if the EUSD problem has a solution.

Suppose that we have a solution to the EUSD problem, which is a set of repetition parameters S^{Rep} with the conserved

TABLE 4: Summary of notations.

notation	definition
N_i^{rep}	number of repetitions of UE$_i$
I_i^{sc}	subcarrier indication of UE$_i$
N_i^{RU}	number of RUs of UE$_i$
MCS$_i$	modulation and coding scheme of UE$_i$
S_i^{sc}	set of allocated subcarriers of UE$_i$
D_i	uplink request of UE$_i$ (bits)
R_i	required transmission reliability of UE$_i$
d_i	delay constraint of UE$_i$ (ms)
T_i^{req}	data arrival time of UE$_i$ (ms)
P_i	transmit power of UE$_i$
P_i^{max}	maximum transmit power of UE$_i$
N_i^{sc}	RU type of UE$_i$
$\text{SNR}_{\text{dB}}(i)$	received signal-to-noise ratio of UE$_i$ (dB)
$\tilde{P}$	received power
G_i	transmitter gain of UE$_i$
G_{eNB}	receiver gain of UE$_i$
$L(i, \text{eNB})$	path loss between UE$_i$ and the eNB
B	subcarrier bandwidth of the NB-IoT (Hz)
N_o	noise power
I	interference perceived at the eNB
BER$_i$	bit-error-rate of UE$_i$
$\text{SNR}_{\text{dB}}^{\text{Req}}(\cdot)$	SNR threshold (dB)
$r(\text{MCS}_i)$	data rate of MCS$_i$ (bits per subcarrier × slot)
P_i^s	successful probability of data transmission of UE$_i$
T_i^{DCI}	DCI subframe index of UE$_i$ (ms)
T^s	index of current scheduling subframe (ms)
K	feasible subcarrier set
S_k	index of earliest available subframe of subcarrier k
T_i^{sc}	allocation start time of RU of UE$_i$ (ms)
N_i^{slot}	number of slots for a single RU of UE$_i$
notations of the proposed scheme	definition
A_i	feasible setting pairs of RU type and MCS of UE$_i$
Score$_i$	score value of UE$_i$
W_1, W_2	weighting factors
Em_i	urgent level of UE$_i$'s request
T_j^R	remaining time of UE$_j$'s request from the scheduling subframe T^s (ms) to the delay deadline
Waste(i, S_i^{sc})	potential waste of UE$_i$ with its allocated subcarrier set
$\widehat{S}_k$	earliest available allocation subframe (ms) of RUs for subcarrier k
$S_i^{\text{sc}*}$	best subcarrier of UE$_i$
$C_i^{\alpha,\beta}$	cost ratio of UE$_i$
$\Psi_{N^{\text{sc}}}()$	number of choices of RU types

energy of RUs. Each UE can choose exact one repetition number and we need to assign each number to each UE to satisfy their transmission reliability. The total size of RUs cannot exceed the frame space U and the conserved energy of all UEs is Q. By viewing the possible number of repetitions of an UE as a class of objects and the frame as the knapsack, the repetition numbers in S^{Rep} all constitute a solution to the MCK problem. This proves the *if* part.

Conversely, let $\mathbf{x}_{1,\alpha 1}, \mathbf{x}_{2,\alpha 2}, \ldots$ be a solution to the MCK problem. Then, for each UE$_i$, we select a repetition number

such that UE$_i$ conserves energy of $\mathbf{q}_{i,\alpha i}$ and the size of allocated RUs to transmit UE$_i$'s data to the eNB is $\mathbf{u}_{i,\alpha i}$. In this way, the conserved energy of all UEs will be $\mathbf{Q}$ and the overall RU size is no larger than U. This constitutes a solution to the EUSD problem, thus proving the *only if* part. □

4. The Proposed Scheme

Since the EUSD problem is NP-complete, finding the optimal solution is impractical due to the time complexity. Thus, we

propose a low-complexity, energy-efficient, and high-reliable scheme to tackle this problem. This scheme consists of two phases. The first phase exploits the strategy of "*minimal energy cost*" to determine the scheduling parameters of UEs. This scheme first quantifies the consumed energy for each UE and then chooses the one with minimal energy cost and reserves the corresponding parameters as the default transmit setting while satisfying the required reliability. The second phase determines the scheduling order based on the "*score function*" which considers the emergency level of requests and inflexibility of the scheduling transmission. Then, it determines the best resource location of RUs of each UE based on the "*potential resource waste*" function to enhance the resource utilization. Finally, if the predetermined results violate an UE's delay requirements on scheduling, the scheme will calculate a "*cost ratio*" to adaptively adjust the RU assignments. The details of the scheme are described as follows.

4.1. Phase 1. Minimal Energy Cost. The goal of the first phase is to determine the default parameters for each UE, including the type of RUs (N_i^{sc}), the number of RUs (N_i^{RU}), the best number of repetitions (N_i^{rep}), 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, $i = 1 \cdots N$, we first calculate the required number of RUs N_i^{RU} according to (4) 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.,

$$
A_i = \left\{ \left(N_{i,j}^{sc}, MCS_{i,j} \right) \mid N_{i,j}^{RU} \times \frac{N_{i,j}^{slot}}{2} \le d_i, \ N_{i,j}^{sc} \right.
$$
$$
\left. \in \{1, 3, 6, 12\}, MCS_{i,j} \in \{0 \cdots 10\} \right\}, \tag{10}
$$

where j is the index of feasible setting pair of RU type and MCS for each UE_i and $N_{i,j}^{slot}$ is the number of required slots when the RU type is $N_{i,j}^{sc}$, which can be obtained by (7). Note that $N_{i,j}^{slot}$ is divided by 2 because two slots constitute 1 ms, which is the unit of delay constraint d_i.

Step 2. For each UE_i, $i = 1 \cdots N$, consider the feasible RU type and MCS setting pair $(N_{i,j}^{sc}, MCS_{i,j}) \in A_i$; we calculate the allowed repetition numbers $N_{i,j}^{rep}$ in which each $N_{i,j,k}^{rep} \in N_{i,j}^{rep}$ 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.,

$$
N_{i,j}^{rep} = \left\{ N_{i,j,k}^{rep} \mid 1 - \left(1 - P_{i,j,k}^s \right)^{N_{i,j,k}^{rep}} \ge R_i, N_{i,j,k}^{rep} \right.
$$
$$
\in \left\{ 2^l \mid l \in \{0 \cdots 7\} \right\}, N_{i,j}^{RU} \times \frac{N_{i,j}^{slot}}{2} \times N_{i,j,k}^{rep} \le d_i, 0 \tag{11}
$$
$$
\left. \le P \left(N_{i,j}^{sc}, MCS_{i,j}, BER_{i,j,k} \right) \le P_i^{max} \right\},
$$

where

$$
BER_{i,j,k} = 1 - \left(1 - \left(1 - R_i \right)^{1/N_{i,j,k}^{rep}} \right)^{1/D_i} \tag{12}
$$

is derived from (5) and $P(N_{i,j}^{sc}, MCS_{i,j}, BER_{i,j,k})$ is a function which returns the minimum transmit power for the RU type $N_{i,j}^{sc}$, MCS setting $MCS_{i,j}$, and target bit-error-rate $BER_{i,j,k}$; i.e.,

$$
P \left(N_{i,j}^{sc}, MCS_{i,j}, BER_{i,j,k} \right)
$$
$$
= 10^{SNR_{dB}^{Req}(MCS_{i,j}, BER_{i,j,k})/10} \tag{13}
$$
$$
\times \frac{(BN_0 + I) \cdot L(i, eNB) \cdot N_{i,j}^{sc}}{G_i G_{eNB}},
$$

which can be derived from (3).

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

Step 3. Based on the results of Steps 1 and 2, we calculate the most energy-saving repetition number $N_{i,j}^{rep^*}$ for each feasible combination pair $(N_{i,j}^{sc}, MCS_{i,j}) \in A_i$, where

$$
N_{i,j}^{rep^*} = \arg \min_{N_{i,j,k}^{rep} \in N_{i,j}^{rep}} E \left(N_{i,j}^{sc}, MCS_{i,j}, N_{i,j,k}^{rep} \right),
$$
$$
E \left(N_{i,j}^{sc}, MCS_{i,j}, N_{i,j,k}^{rep} \right)
$$
$$
= P \left(N_{i,j}^{sc}, MCS_{i,j}, BER_{i,j,k} \right) \times N_{i,j}^{RU} \times \frac{N_{i,j}^{slot}}{2} \tag{14}
$$
$$
\times N_{i,j,k}^{rep}.
$$

Then, reform A_i as a set of triplets $(N_{i,j}^{sc}, MCS_{i,j}, N_{i,j}^{rep^*})$. 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 $(\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}})$ from A_i as the default parameter of UE_i, which incurs the minimum energy consumption by

$$
\left(\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}} \right)
$$
$$
= \arg \min_{(\mathbf{N_{i,j}^{sc}}, \mathbf{MCS_{i,j}}, \mathbf{N_{i,j}^{rep^*}}) \in A_i} \left\{ \mathbf{E} \left(\mathbf{N_{i,j}^{sc}}, \mathbf{MCS_{i,j}}, \mathbf{N_{i,j}^{rep^*}} \right) \right\}. \tag{15}
$$

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

4.2. Phase 2. Weighting Based Flexible Scheduling. The goal of the second phase is to optimize the scheduling results of requests from UEs, including the subcarrier set of RUs (S_i^{sc}) and the start time of RUs (T_i^{sc}). 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.,

$$\text{Score}_i = W_1 \times Em_i + W_2 \times \widetilde{\text{Inf}_i}, \tag{16}$$

where $W_1 \in [0, 1]$ and $W_2 \in [0, 1]$ are the weighting factors of the degrees 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.,

$$Em_i = \frac{\max_j \left\{ T_j^R \right\} - T_i^R}{\max_j \left\{ T_j^R \right\} - \min_j \left\{ T_j^R \right\}}, \tag{17}$$

where T_j^R is the remaining time from the scheduling subframe T^S (or current subframe) to the delay deadline $T_j^{req} + d_j$ of UE_j; i.e.,

$$T_i^R = \max\left(\left(T_i^{req} + d_i \right) - T^s, 0 \right). \tag{18}$$

$\widetilde{\text{Inf}_i}$ is the number of RU types that UE_i can choose, which is defined by

$$\widetilde{\text{Inf}_i} = \frac{\text{Inf}_i}{\max_j \left\{ \text{Inf}_j \right\}}, \tag{19}$$

where

$$\text{Inf}_i = \begin{cases} 4, & \text{if } \Psi_{N^{sc}}(A_i) = 1 \\ 3, & \text{if } \Psi_{N^{sc}}(A_i) = 2 \\ 2, & \text{if } \Psi_{N^{sc}}(A_i) = 3 \\ 1, & \text{if } \Psi_{N^{sc}}(A_i) > 3 \end{cases} \tag{20}$$

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

Now, for each UE_i, $i = 1 \cdots N$, we calculate its Score_i and sort them in descending order. For the UEs without any request, define its score as $-\infty$. Without loss of generality, we use *ListL* to represent the sorted sequence of the UEs.

Step 2. Before determining the subcarrier set of RUs, we first define a function $\text{Waste}(i, S_i^{sc})$ to reflect the potential waste of resource if UE_i's RUs are allocated at subcarrier set S_i^{sc}; i.e.,

$$\begin{aligned} &\text{Waste}\left(i, S_i^{sc}\right) \\ &= \sum_{k' \in K - S_i^{sc}} \left(\left(\max_{k \in S_i^{sc}} \left\{ \widehat{S}_k \right\} + \left(N_i^{RU} \times \frac{N_i^{slot}}{2} \times N_i^{rep} \right) \right) \right. \\ &\quad \left. - \widehat{S}_{k'} \right)^+ + \sum_{k \in S_i^{sc}} \left(\max_{k \in S_i^{sc}} \left\{ \widehat{S}_k \right\} - \widehat{S}_k \right), \end{aligned} \tag{21}$$

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. Note that (21) sums up the unused resource space before the resource allocation finish time of UE_i if UE_i's RUs are allocated at S_i^{sc}.

Then, based on (21), we choose the best subcarrier set $S_i^{sc^*}$ that makes UE_i have the minimal $\text{Waste}(i, S_i^{sc})$ without violating its delay deadline; i.e.,
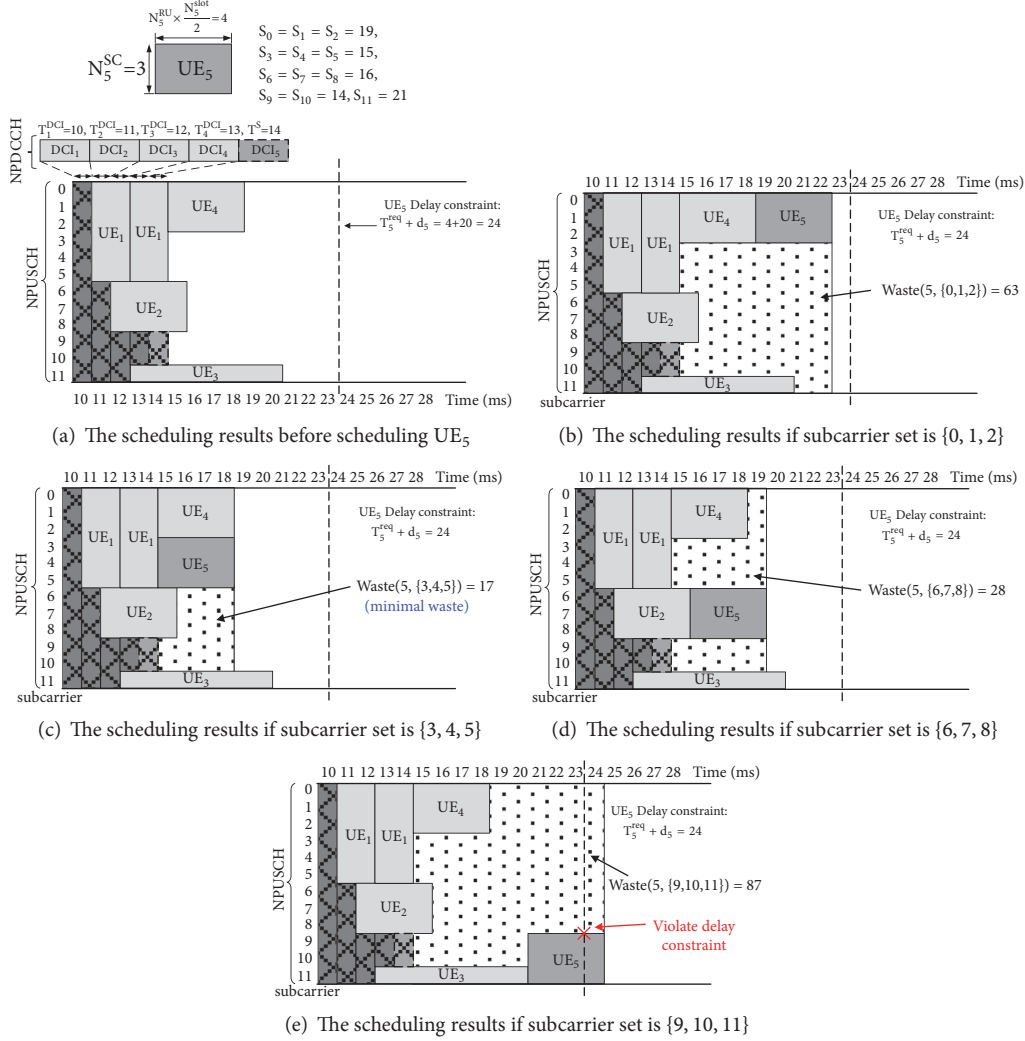
$$\begin{aligned} S_i^{sc^*} = \arg \min_{S_i^{sc} \subseteq \Theta(N_i^{sc^*})} &\left\{ \text{Waste}\left(i, S_i^{sc}\right) \mid \max_{k \in S_i^{sc}} \left\{ \widehat{S}_k \right\} \right. \\ &\left. + \left(N_i^{RU^*} \times \frac{N_i^{slot^*}}{2} \times N_i^{rep^*} \right) < \left(T_i^{req} + d_i \right) \right\}, \end{aligned} \tag{22}$$

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 $S_k = \max\{\max_{k \in S_i^{sc^*}}\{\widehat{S}_k\} + (N_i^{RU^*} \times (N_i^{slot^*}/2) \times N_i^{rep^*}), T_i^{DCI} + 1\}$ 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 current transmission parameter setting is infeasible and 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 phase.

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* $C_i^{\alpha, \beta}$ 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 $\alpha = (N_i^{sc^*}, MCS_i^*, N_i^{rep^*})$, changes to the new pair, denoted

(a) The scheduling results before scheduling UE$_5$

(b) The scheduling results if subcarrier set is {0, 1, 2}

(c) The scheduling results if subcarrier set is {3, 4, 5}

(d) The scheduling results if subcarrier set is {6, 7, 8}

(e) The scheduling results if subcarrier set is {9, 10, 11}

FIGURE 4: Examples to schedule UE$_5$ based on the potential resource waste.

as $\beta = (\mathbf{N_i^{sc'}}, \mathbf{MCS_i'}, \mathbf{N_i^{rep'}})$ for $(\mathbf{N_i^{sc'}}, \mathbf{MCS_i'}, \mathbf{N_i^{rep'}}) \in A_i - (\mathbf{N_i^{sc^*}}, \mathbf{MCS_i^*}, \mathbf{N_i^{rep^*}})$; i.e.,

$$C_i^{\alpha,\beta} = \begin{cases} \dfrac{\Delta E_i^{\alpha,\beta}}{\Delta Area_i^{\alpha,\beta}}, & \text{if } \Delta Area_i^{\alpha,\beta} > 0 \\ \Delta E_i^{\alpha,\beta}, & \text{if } \Delta Area_i^{\alpha,\beta} = 0, \end{cases} \quad (23)$$

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

Then, we choose the new pair β^* which incurs the minimal cost ratio; i.e.,

$$\beta^* = \left(\mathbf{N_i^{sc'}}, \mathbf{MCS_i'}, \mathbf{N_i^{rep'}} \right)$$
$$= \arg \min \left\{ C_i^{\alpha,\beta} \mid \beta \in A_i - \alpha \right\} \quad (24)$$

and replace the default parameters by $\mathbf{N_i^{sc^*}} = \mathbf{N_i^{sc'}}$, $\mathbf{MCS_i^*} = \mathbf{MCS_i'}$, and $\mathbf{N_i^{rep^*}} = \mathbf{N_i^{rep'}}$ 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.

Below, we give an example in Figure 4, where there are four scheduled UEs (UE$_1$~UE$_4$) and one UE (UE$_5$) to be scheduled. The subframe indexes of DCIs for the four scheduled UEs are $T_1^{DCI} = 10$, $T_2^{DCI} = 11$, $T_3^{DCI} = 12$, and $T_4^{DCI} = 13$ (ms), separately. The current earliest subframe index for DCI is subframe 14 and scheduling subframe is $T^s = 14$ (ms). Now, we consider schedule UE$_5$, whose RU type is 3-tone, number of RUs is $N_5^{RU} = 1$, total length is $N_5^{RU} \times (N_5^{slot}/2) = 1 \times (8/2) = 4$ (ms), arrival time is $T_5^{req} = 4$ (ms), and delay constraint is $d_5 = 20$ (ms). From

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

the current scheduling results, as shown in Figure 4(a), the available start time for each subcarrier is $S_0 = 19$, $S_1 = 19$, $S_2 = 19$, $S_3 = 15$, $S_4 = 15$, $S_5 = 15$, $S_6 = 16$, $S_7 = 16$, $S_8 = 16$, $S_9 = 14$, $S_{10} = 14$, and $S_{11} = 21$, respectively. Thus, the available subcarrier set for UE$_5$ is $S_5^{sc} \in \{\{0, 1, 2\}, \{3, 4, 5\}, \{6, 7, 8\}, \{9, 10, 11\}\}$. So, for the 4 possible subcarrier sets, the potential resource waste for UE$_5$ is Waste$(5, \{0, 1, 2\}) = \sum_{k'=3\sim11}(\max\{19, 19, 19\}+1\times(8/2)\times1)^+ - \widehat{S}_{k'} = 63$, Waste$(5, \{3, 4, 5\}) = \sum_{k'=0\sim2,6\sim11}(\max\{15, 15, 15\} + 1 \times (8/2) \times 1)^+ - \widehat{S}_{k'} = 17$, Waste$(5, \{6, 7, 8\}) = \sum_{k'=0\sim5,9\sim11}(\max\{16, 16, 16\} + 1 \times (8/2) \times 1)^+ - \widehat{S}_{k'} = 28$, and Waste$(5, \{9, 10, 11\}) = \sum_{k'=0\sim8}(\max\{15, 15, 21\} + 1 \times (8/2) \times 1)^+ - \widehat{S}_{k'} = 87$, which are the doted regions in Figures 4(b), 4(c), 4(d), and 4(e), respectively. Moreover, the subcarrier set of $\{9, 10, 11\}$ in Figure 4(e) is infeasible because it violates the delay deadline of UE$_5$ (i.e., $\max_{k\in\{9,10,11\}}\{\widehat{S}_k\} + (N_5^{RU} \times (N_5^{slot}/2) \times N_5^{rep}) = 21 + 1 \times (8/2) \times 1 = 25 \geq (T_5^{req} + d_5) = 4 + 20 = 24$). Thus, based on (21), the scheme chooses the best feasible subcarrier set $S_5^{sc^*} = \{3, 4, 5\}$ and the start time of UE$_5$'s schedules RUs is $T_5^{sc} = \max_{k\in\{3,4,5\}}\{\widehat{S}_k\} = 15$ because it has the minimal value of Waste$(5, \{3, 4, 5\}) = 17$.

5. Simulation Results

In this section, we develop a simulator in C++ language to verify the efficiency of the proposed scheme (currently, the well-known simulator, such as ns-3 [26], has not supported NB-IoT model in terms of channels mappings and access procedures.). The parameters of the simulation are shown in Table 5. Specifically, the simulator emulates a base station to serve $N = 3000\sim30000$ UEs. Based on the model of *MAR (Mobile Autonomous Reporting)* [8], the uplink requests of UEs arrive randomly with the data size of $D_i = 50\sim200$ bytes according to the Pareto distribution with $\alpha = 2.5$. In addition, the interarrival time of requests includes 30 minutes (5%), 1 hour (15%), 2 hours (40%), and 1 day (40%), respectively.

In this simulation, we compare our scheme *(Ours)* with the standard scheme *(Spec)* [2], Narrowband Link Adaptation scheme *(NBLA)* [14], random scheduling scheme *(Random)*, and round robin scheme *(RR)*. Specifically, Spec scheme chooses the single-tone for UEs with constant repetition number for simplicity (the numbers are 1 and 2 in the

simulation, denoted as Spec(1) and Spec(2)). NBLA can adjust the MCSs and repetition levels iteratively to ensure the transmission quality and delay. Random scheme schedules the UEs in a random order with a random repetition number. RR scheme schedules the UEs in round robin order with the repetition number of 1. Note that the weighting factors of our scheme is $W_1 = 0.5$ and $W_2 = 0.5$.

We consider five performance metrics: (i) *system throughput*: the total number of data bits received successfully by the eNB during the experiment period; (ii) *the number of serving UEs*: the average number of UEs that satisfy QoS and reliability; (iii) *resource consumption*: the frame space allocated to the uplink transmission over the total frame space; (iv) *packet drop rate*: the number of dropped packets due to violating delay constraint over the total number of packets; (v) *energy consumption per UE*: the average consumed energy of each served UE. Note that the scheduling interval is 30 ms and the simulator emulates for 24 hours.

5.1. System Throughput. First, we investigate the effects of number of request UEs on system throughput. As shown in Figure 5(a), we can see that when the number of request UEs increases, the system throughput of all scheme increases fast and then slows down due to system saturation. RR, Spec, and Random perform worse because they cannot satisfy UEs' QoS requirement and reliability. Thus, the total number of data bits received successfully by the eNB is few. Specifically, Spec(2) performs slightly better than Spec(1) because it can meet more UEs' reliability due to applying the larger repetition number. NBLA is better than the above schemes because it can adjust MCSs and repetition levels to satisfy QoS and reliability. Note that our scheme outperforms other schemes because it can flexibly adjust RU type and optimize repetition number to satisfy QoS while ensuring transmission reliability.

We also investigate the effects of distribution of request data size on system throughput. As shown in Figure 5(b), similarly, when the request data size increases, the system throughput of most schemes increases and then slows down due to system saturation. RR and Spec(1) decrease fast because their repetition number is 1 and may not satisfy the successful transmission probability due to larger request size. Our scheme is still the best because it can schedule UEs flexibly while ensuring QoS requirement and reliability.

Then, Figure 5(c) shows the impact of the distribution of delay constraints on system throughout. As can be seen, when the distribution of delay constraints increases, the system throughput of most schemes increases slowly. This is because the longer delay constraint can help more UEs to be satisfied until the frame space is exhausted. RR and Spec(1) increase slowly because they fix repetition number by 1 that would limit the successful transmission probability even when the UEs have longer delay constraint. Note that our scheme has the highest throughput because of its flexible scheduling and appropriate parameter setting.

Finally, we investigate the impact of the distribution of required reliability on system throughout. In Figure 5(d), when the distribution of reliability increases, the system throughput of most schemes decreases. The reason is that more UEs with strict reliability make all schemes harder to
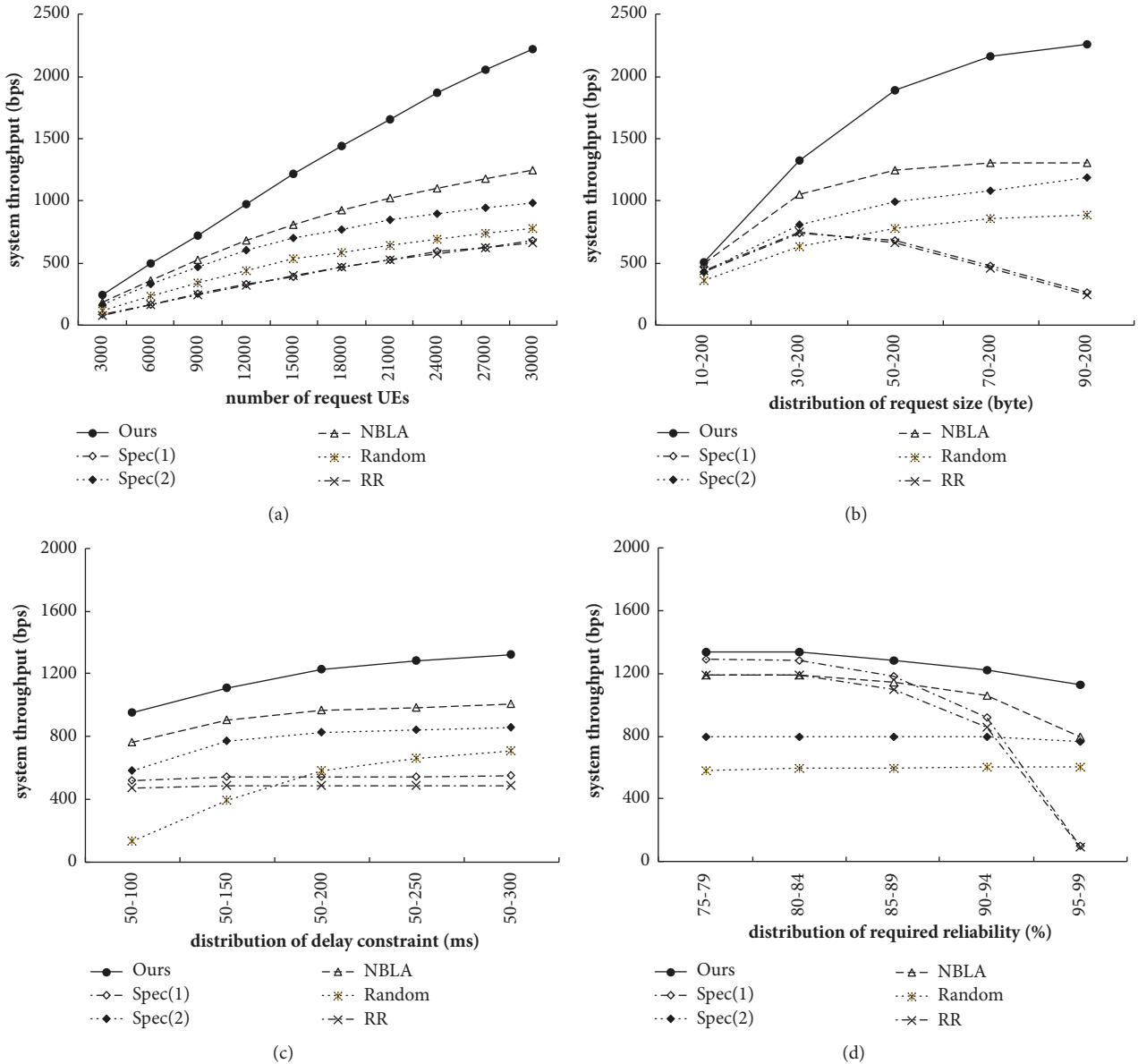
FIGURE 5: Comparisons on the system throughput of all schemes.

serve them due to their higher requirements. RR and Spec(1) decrease dramatically because their repetition number is 1 that could not serve most UEs with higher reliability. Note that our scheme still outperforms others.

5.2. Number of Serving UEs. Then, we investigate the effects of number of request UEs on number of serving UEs. As shown in Figure 6(a), 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 phase 1 and apply the best configuration pair of RU type and MCS to ensure QoS while enhancing the spectrum utilization in phase 2.

We also investigate the effects of distribution of request data size on number of serving UEs. As shown in Figure 6(b), contrarily, when the request data size increases, the number of serving UEs of all schemes decreases because the UEs with larger request size consume more frame space. RR and Spec(1) decrease fast because their repetition number is fixed by 1 that would not satisfy the UEs with larger request size.
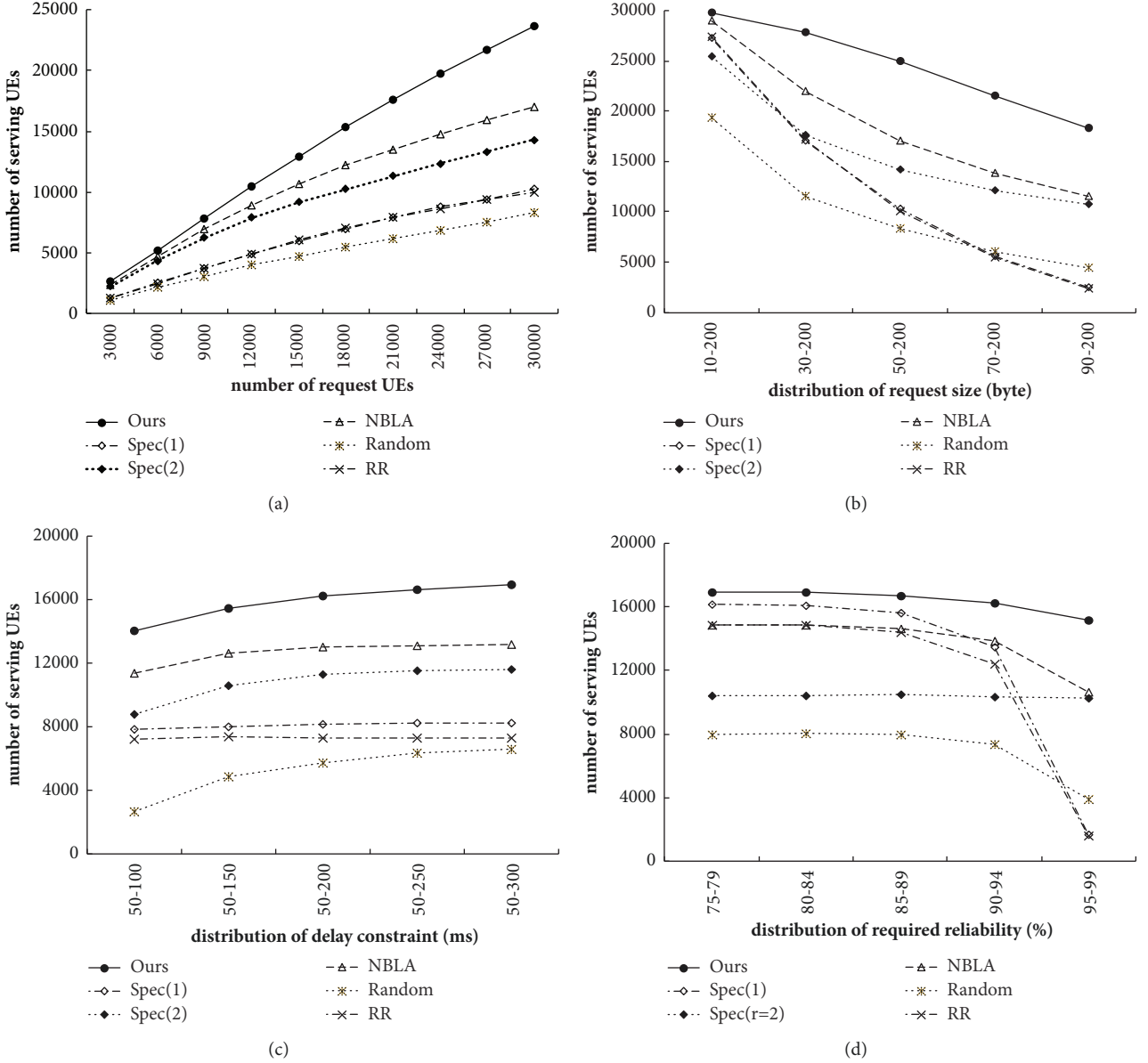
(a)

(b)

(c)

(d)

FIGURE 6: Comparisons on the number of serving UEs of all schemes.

Our scheme is the best because it flexibly schedules UEs to satisfy their delay requirement and reliability.

In Figure 6(c), we observe the impact of the distribution of delay constraints on number of serving UEs. As can be seen, when the distribution of delay constraints increases, the number of serving UEs of all schemes increases slowly. This is because the longer delay constraint can make more UEs tolerate allocation time until the frame space is exhausted. RR and Spec(1) increase very slowly because their fixed repetition number (i.e., 1) would limit the transmission probability although the UEs have longer delay time. Note that our scheme has the highest performance because it can well determine the scheduling setting.

Finally, we investigate the impact of the distribution of required reliability on number of serving UEs in Figure 6(d).

We can see that when the distribution of reliability increases, the number of serving UEs of all schemes decreases. The reason is that more UEs with strict reliability will make schemes hardly serve them due to higher requirements. Note that RR and Spec(1) decrease dramatically because they fix the repetition number by 1 that could not serve the UEs with higher reliability requirements. Also note that our scheme still outperforms others even when the reliability requirements become higher.

5.3. Packet Drop Rate. Then, we investigate the effects of number of request UEs on packet drop rate. As shown in Figure 7(a), we can see that when the number of request UEs increases, the packet drop rate of all schemes increases due to network saturated gradually. Spec, Random, and RR perform
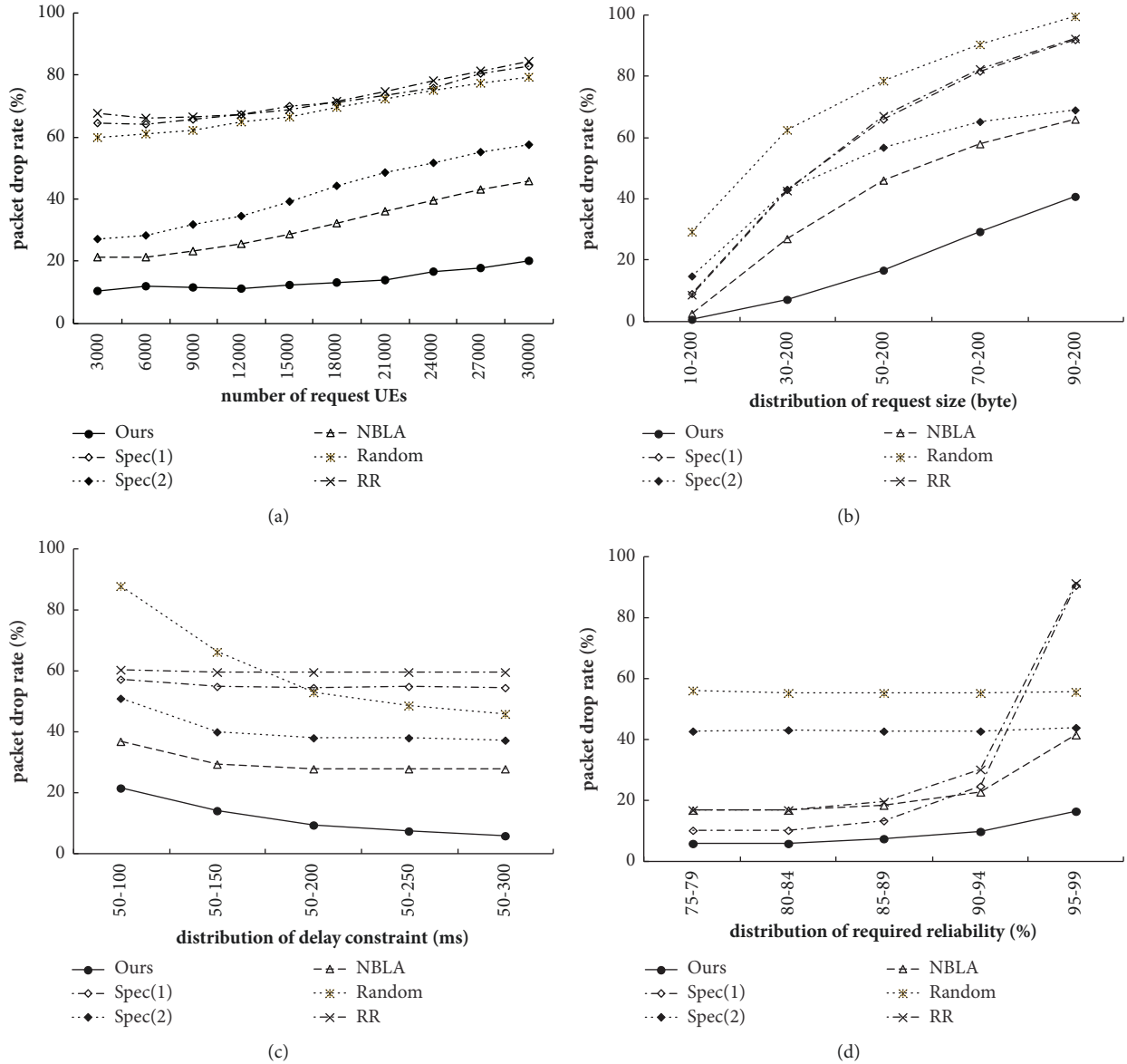
Figure 7: Comparisons on the packet drop rate of all schemes.

worse because they usually schedule UEs with single-tone so the UEs' RU length may exceed their delay constraint. NBLA performs better than above schemes because it can choose better repetition number to mitigate packets being dropped. Our scheme has the lowest packet drop rate because our scheme can flexibly select multitype RUs and minimize the number of repetitions to guarantee the QoS while avoiding packets being dropped.

We also investigate the effects of distribution of request data size on packet drop rate. As shown in Figure 7(b), similarly, when the request data size increases, the packet drop rate of all schemes increases because the larger request size consumes more resources that may exceed the frame space which makes packet being dropped. RR and Spec(1) increase fast because they could not satisfy the UEs with

larger request size, which require higher transmission probability. Our scheme is the best because it can schedule UEs according to their delay and reliability requirements.

Then, in Figure 7(c), we investigate the impact of the distribution of delay constraints on packet drop rate. As can be seen, when the distribution of delay constraints increases, the packet drop rate of all schemes decreases gradually. This is because the longer delay constraint can make more packets tolerate the allocation time before delay expires. RR and Spec(1) decrease very slowly because they fix repetition number by 1 so that it would not help packets being transmitted. Note that our scheme has the lowest packet drop rate because it can well determine the scheduling results.

Finally, we investigate the impact of the distribution of required reliability on packet drop rate in Figure 7(d). As
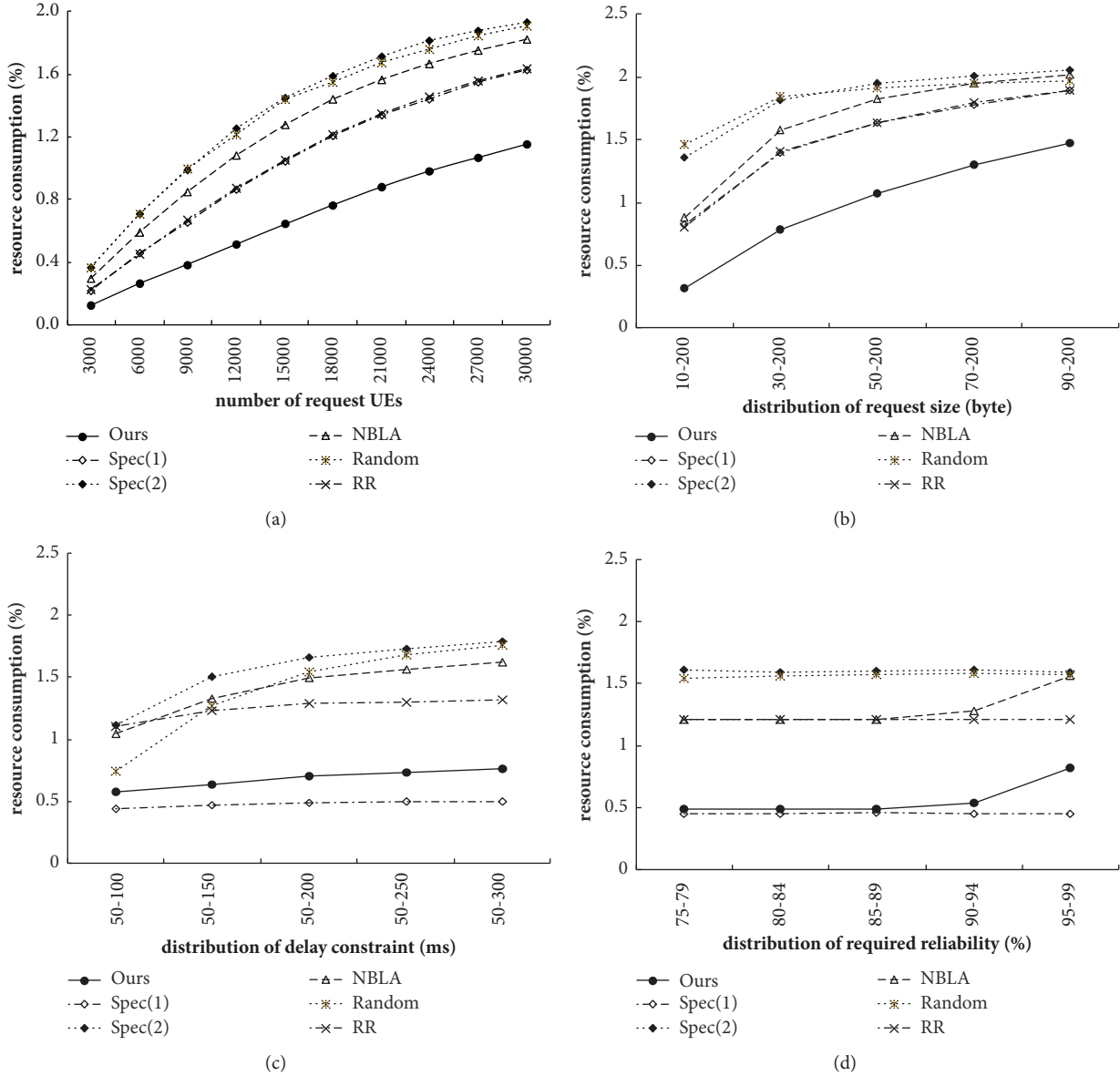
FIGURE 8: Comparisons on the resource consumption of all schemes.

can be seen, when the distribution of reliability increases, the packet drop rate of all schemes increases. Our scheme still outperforms others even when the reliability requirements are up to 95–99%.

5.4. Resource Consumption. Here, we investigate the effects of number of request UEs on resource consumption. As shown in Figure 8(a), we can see that when the number of request UEs increases, the resource consumption of all schemes increases. Spec(2), Random, and NBLA perform worst because they usually schedule the UEs with multiple number of repetitions and thus consume more resources. Spec(1) and RR have lower resource consumption because they prefer to choose the UEs with single repetition to reduce resource. Note that our scheme needs the least spectrum

resource because our scheme exploits a waste function to avoid resource consumption.

Here, we also investigate the effects of distribution of request data size on resource consumption. As shown in Figure 8(b), similarly, when the request data size increases, the resource consumption of most schemes increases slowly. Spec(2), Random, and NBLA waste more resources because they usually schedule the UEs with multiple repetitions and thus consume more resources. Our scheme is still the best because it can schedule UEs flexibly while ensuring QoS requirement and reliability.

Then, in Figure 8(c), we investigate the impact of the distribution of delay constraints on resource consumption. As can be seen, when the distribution of delay constraints increases, the resource consumption of all schemes increases.
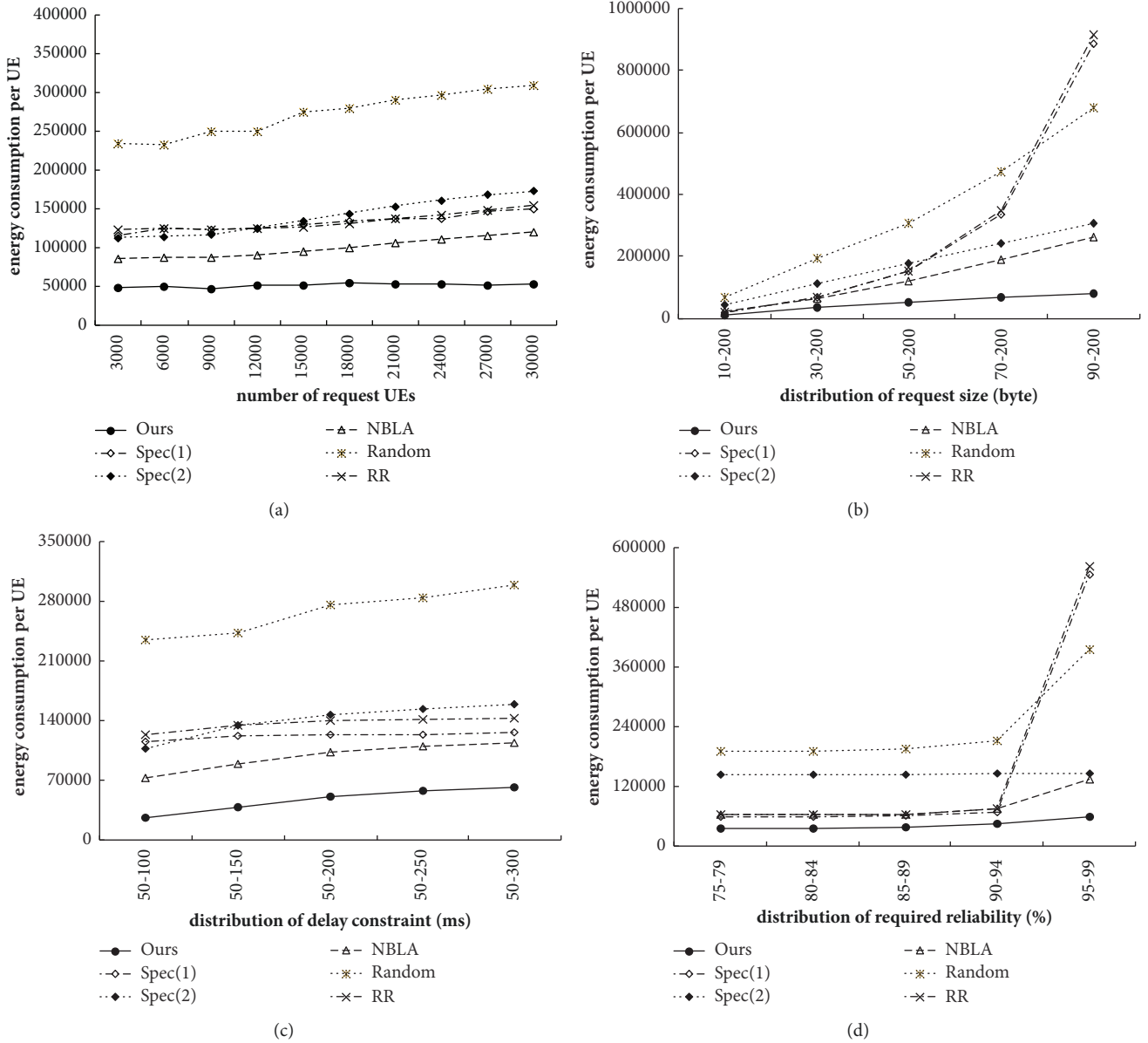
Figure 9: Comparisons on energy consumption per UE of all schemes.

This is because the longer delay constraint can make more UEs satisfied that may consume more resources. Note that although Spec(1) has the lowest resource consumption, it incurs lower system throughput, lower number of serving UEs, and higher packet drop rate, as compared to ours.

Finally, we investigate the impact of the distribution of required reliability on resource consumption, as shown in Figure 8(d). When the distribution of reliability increases, the resource consumption of most schemes increases. Our scheme increases slightly because our scheme can enlarge the repetition number with multiple tone to satisfy the UEs with higher reliability requirement.

5.5. Energy Consumption per UE. Finally, we investigate the effects of the number of request UEs on energy consumption

per UE. As shown in Figure 9(a), 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 chooses 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 phase

1 and leverage the cost ratio to reduce energy consumption in phase 2, thus saving energy more efficiently.

In Figure 9(b), we also investigate the effects of distribution of request data size on energy consumption per UE. As can be seen, when the request data size increases, the energy consumption per UE of all schemes increases. Our scheme incurs the lowest energy consumption because it can determine the minimal transmit power and the corresponding RU setting to save energy.

Then, in Figure 9(c), we observe the impact of the distribution of delay constraints on energy consumption per UE. We can see that when the distribution of delay constraints increases, the energy consumption per UE of all schemes increases. Similarly, our scheme still has the lowest energy consumption because it can calculate the best transmit power and exploit the cost ratio to reduce energy consumption.

Finally, we investigate the impact of the distribution of required reliability on energy consumption per UE. Figure 9(d) shows that when the distribution of reliability increases, the energy consumption of most schemes increases. Our scheme still outperforms all other schemes because it can determine the best scheduling parameters based on the required reliability of UEs.

6. Conclusion

In this paper, we have addressed the problem of energy saving and reliable communications in NB-IoT networks. We first model this problem as an optimization problem and prove it to be NP-complete. Then, we propose an energy-efficient and high-reliable scheme which has two phases. The first phase leverages the strategy of minimal energy cost to choose the default scheduling parameters with least energy consumption. The second phase exploits the weighting score function to balance the emergency and inflexibility of the request transmission and then serve the UEs with least potential resource waste. Simulation results have verified that our scheme can satisfy more UEs with ultra-reliability and QoS requirement while saving their energy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

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

[2] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (E-UTRA)," *Physical Layer Procedures, v14.4.0*, pp. 1–6, 2017.

[3] 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.

[4] E. Mohyeldin, "Minimum Technical Performance Requirements for IMT-2020 radio interface(s)," *ITU-R Workshop on IMT-2020 Terrestrial Radio Interfaces*, pp. 1–12, 2016.

[5] R. Ma, K. H. Teo, S. Shinjo, K. Yamanaka, and P. M. Asbeck, "A GaN PA for 4G LTE-Advanced and 5G: Meeting the telecommunication needs of various vertical sectors including automobiles, robotics, health care, factory automation, agriculture, education, and more," *IEEE Microwave Magazine*, vol. 18, no. 7, pp. 77–85, 2017.

[6] O. Teyeb, G. Wikstr, M. Stattin, T. Cheng, S. Faxér, and H. Do, "Evolving LTE to fit the 5G future," *Ericsson Technology Review*, 2017.

[7] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IoT system for M2M communication," in *Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, WCNC 2016*, pp. 1–5, April 2016.

[8] TR 45.820, "Cellular system support for ultra low complexity and low throughput internet of things," *v2.1.0*, pp. 1–6, 2015.

[9] Y. E. Wang, X. Lin, A. Adhikary et al., "A primer on 3GPP narrowband Internet of Things (NB-IoT)," *IEEE Communications Magazine*, vol. 15, no. 3, pp. 117–123, 2017.

[10] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J. P. Koskinen, "Overview of narrowband IoT in LTE Rel-13," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, pp. 1–7, 2016.

[11] A. D. Zayas and P. Merino, "The 3GPP NB-IoT system architecture for the Internet of Things," in *Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 277–282, Paris, France, May 2017.

[12] N. Mangalvedhe, R. Ratasuk, and A. Ghosh, "NB-IoT deployment study for low power wide area cellular IoT," in *Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6, Valencia, Spain, September 2016.

[13] L. Zhang, A. Ijaz, P. Xiao, and 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.

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

[15] X. Lin, A. Adhikary, and Y. 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.

[16] S. Oh and J. Shin, "An efficient small data transmission scheme in the 3GPP NB-IoT system," *IEEE Communications Letters*, vol. 21, no. 3, pp. 660–663, 2017.

[17] H. Kroll, M. Korb, B. Weber, S. Willi, and Q. Huang, "Maximum-likelihood detection for energy-efficient timing

acquisition in NB-IoT," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1–5, San Francisco, CA, USA, March 2017.

[18] J. Lee and J. Lee, "Prediction-based energy saving mechanism in 3GPP NB-IoT networks," *Sensors*, vol. 17, no. 9, 2017.

[19] A. E. Mostafa, Y. Zhou, and V. W. Wong, "Connectivity maximization for narrowband IoT systems with NOMA," in *IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.

[20] 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA)," *Radio Resource Control (RRC), v15.1.0*, pp. 1–786, March 2018.

[21] G. Tsoukaneri, M. Condoluci, T. Mahmoodi, M. Dohler, and M. K. Marina, "Group communications in Narrowband-IoT: Architecture, procedures, and evaluation," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 1–10, 2018.

[22] H. Malik, H. Pervaiz, M. Mahtab Alam, Y. Le Moullec, A. Kuusik, and M. Ali Imran, "Radio resource management scheme in NB-IoT systems," *IEEE Access*, vol. 6, pp. 15051–15064, 2018.

[23] A. Rico-Alvarino, R. Lopez-Valcarce, C. Mosquera, and R. W. Heath, "FER estimation in a memoryless BSC with variable frame length and unreliable ACK/NAK feedback," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3661–3673, 2017.

[24] M. Jacobsson and C. Rohner, "Estimating packet delivery ratio for arbitrary packet sizes over wireless links," *IEEE Communications Letters*, vol. 19, no. 4, pp. 609–612, 2015.

[25] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*, Springer, 2004.

[26] NS-3 Consortium, "ns-3 network simulator," 2018, https://www.nsnam.org/.