On-Device Cognitive Spectrum Allocation for Coexisting URLLC and eMBB Users in 5G Systems

Shu-Feng Cheng, *Member, IEEE*, Li-Chun Wang, *Fellow, IEEE*, Chien-Hwa Hwang,
Ju-Ya Chen, and Li-Yu Cheng

***Abstract*—Ultra Reliable and Low Latency Communications (URLLC) play a key role in 5G vertical markets, but pose many technical challenges especially when sharing the spectrum with Enhanced Mobile Broadband (eMBB) customers. This study aims to overcome the spectrum inefficiency issue of fully separate (FS) approach and the contention issue of the fully overlap (FO) approach. We present a user-initiated probability elastic resource (UPER) approach by dynamically adjusting the probability of using the shared spectrum for eMBB and URLLC traffic based on the current success and failure status of packet transmission status. The probabilities of successful transmission are derived for UPER, FS, and FO and partially overlap (PO) sharing spectrum approaches. We find that the successful transmission probability of UPER approach is 28% and 46% higher than FS and FO approaches, respectively. We further evaluate the reliability and throughput performance of URLLC and eMBB. When the URLLC packet load is low, the UPER method can almost achieve the best performance of the FS method. When the URLLC packet load is high, we show that UPER can improve the reliability performance up to 54% compared with other methods.**

***Index Terms*—Spectrum management, ultra reliable and low latency communications, coexisting systems.**

I. INTRODUCTION

ULTRA-RELIABLE and low latency communications (URLLC) of the fifth generation (5G) wireless communications aim for providing time-critical machine-to-machine or human-to-machine vertical applications, such as factory automation, vehicular communications, and augmented reality, etc., [1]–[4]. Most URLLC services require 99.999% reliability performance within 1 msec latency in the data plane [5], [6]. Remote monitoring of patient's the vital signals is an important human-to-machine URLLC applications use case. On the other hand, intelligent transportation systems is an important machine-to-human URLLC use case [7], where vehicles obtain important information from neighboring vehicles or roadside facilities to enhance the security of self-driving cars.

However, serving URLLC together with enhanced mobile broadband (eMBB) and machine type communications (mMTC) users need to design the advanced spectrum management to overcome the packet collision issue. In general, the service requests for time-critical URLLC applications happen much less frequent than the non-emergency eMBB and mMTC applications. Reserving too many exclusive spectrum for URLLC will result in the waste of resources, while reserving insufficient spectrum will cause packet collision. As such, multiplexing URLLC, eMBB and mMTC packets in a sharing spectrum environment is a good strategy from the viewpoint of spectrum efficiency, but needs to overcome the packet collision issue [8], [9]. Because 5G eMBB services are more popular than mMTC services, we focus on the multiplexing issue of URLLC and eMBB packets [10]–[12].

To support time-critical applications with the latency less than 1 msec, URLLC adopts the grant-free transmission protocol in the medium access control (MAC) layer [13]–[15]. Grant-free transmission protocol can simplify the handshake process between the device and the base station to the two-way handshake. The conventional grant-based requires the four-way handshake process [16]. Therefore, grant-free transmission MAC protocol can control the access delay within one msec with much higher probability than the grant-based transmission protocol. Nevertheless, grant-free transmission may face serious packet collisions. When URLLC share the same spectrum with eMBB, the packet collision issue is worsen.

In the literature, power control [13], [17] and spectrum management [18]–[23] are the two main technical directions to solve the packet collision issue in a spectrum sharing environment for URLLC and eMBB. However, using different power levels to assist multiple access may degrade the performance of the URLLC system because high eMBB signals still interferes with URLLC [24]. From the spectrum management aspect for multiplexing URLLC and eMBB, we need to resolve the optimal dynamic spectrum allocation for different service request rate from URLLC and eMBB.

Therefore, this study develops a dynamic spectrum allocation method that can reduce the packet collisions of multiplexing URLLC and eMBB packets based on grant-free MAC protocol in order to maintain low latency and high reliability for URLLC customers. This paper makes the following contributions.

Manuscript received October 3, 2019; revised March 3, 2020 and May 27, 2020; accepted June 28, 2020. Date of publication July 8, 2020; date of current version March 8, 2021. This work is jointly sponsored by the National Chiao Tung University and Ministry of Education (MOE), the Ministry of Science and Technology (MOST) and Pervasive Artificial Intelligence Research (PAIR) Labs, and MediaTek-NCTU Research Lab under the projects "Higher Education Sprout," MOST 108-2634-F-009-006, and MTKC-2013-1414, respectively. The associate editor coordinating the review of this article and approving it for publication was Y. Zeng. *(Corresponding author: Li-Chun Wang.)*

Shu-Feng Cheng, Li-Chun Wang, and Li-Yu Cheng are with the Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: lichun@cc.nctu.edu.tw).

Chien-Hwa Hwang and Ju-Ya Chen are with the Department of Information Technology, MediaTek Inc., Hsinchu 300, Taiwan.

Digital Object Identifier 10.1109/TCCN.2020.3007890

- Propose a novel user-initial probability elastic resource (UPER) for multiplexing URLLC and eMBB traffic, which can flexibly adjust the probability of using dedicated spectrum based on the current packet collision status. It will be shown that UPER, an on-device data-driven cognitive spectrum allocation, can improve the reliability performance up to 54% compared with other methods when the URLLC packet load is high. When the URLLC packet load is low, the reliability of the proposed UPER can reach 99.45%, which approaches to the FS method and is 6.74% better than the PO method.
- Derive the closed-form expressions for the successful transmission probability in terms of the number of URLLC and eMBB packets based on grant-free MAC protocol under four different spectrum allocation methods, including fully separate (FS), fully overlap (FO), partially overlap (PO) and UPER. Such analytical models can help verify the feasibility of UPER and provide quantitative performance comparison with other considered spectrum allocation methods.
- A simple but effective load control mechanism is proposed for the co-existing eMBB and URLLC network, which is shown to reduce packet collisions and respectively improve the reliability performance of the FO, PO and UPER methods by 52%, 16% and 2% compared to the case without load control.
- The URLLC throughput with the packet length of 50 bytes is 56.25% higher than that of 32 bytes. That is, the throughput performance of URLLC is proportional to the packet length in our considered case.

The rest of the paper is organized as follows: Related works are discussed in Section II. We describe our system model in Section III. Section IV introduces the *Dedicated, Shared and Hybrid* spectrum allocation approaches considered in this paper. Then, the transition probability analysis of network model is shown in Section V. We demonstrate some numerical results with different traffic loads for comparison in Section VII. Finally, we conclude this paper in Section VIII.

II. RELATED WORKS

A. Spectrum Allocation Methods

In the literature, some spectrum allocation methods were proposed to reduce the collision of data packets in URLLC and eMBB multiplexing systems [18]–[23]. The spectrum allocation methods can be classified into three types as shown in Fig. 1.

- *Dedicated spectrum allocation [18]–[20]:* Basically, retaining some exclusive spectrum for URLLC users can improve the URLLC performance. In [18], the dedicated spectrum was allocated for periodic URLLC packets to improve the reliability performance. However, it was mentioned in [19], [20] that the dedicated spectrum allocation is inefficient for URLLC in the URLLC and eMBB coexistence system because URLLC traffic is sporadic. Preserving too much spectrum will result in wasted spectrum.

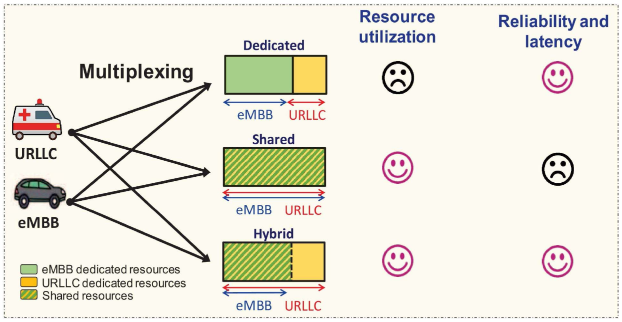


Fig. 1. Spectrum Sharing Approach - *Dedicated, Shared and Hybrid*.

- *Shared spectrum allocation [21], [22], [25], [26]:* The URLLC users and eMBB users who share the whole spectrum can have better spectrum efficiency, while causing more packets collisions.
- *Hybrid spectrum allocation [23]:* To overcome the shortcomings of dedicated and shared spectrum allocation methods, a hybrid approach was proposed to periodically allocate the dedicated and the shared spectrum. In this way, the advantages of high reliability of the dedicated spectrum allocation method and the high spectrum utilization of the sharing method can be preserved.

Nevertheless, the quantitative performance comparison of dedicated, shared and hybrid, spectrum allocation approaches for multiplexing URLLC and eMBB systems is rare in the literature.

B. Load Control

In the following, we introduce some related works of the load control mechanism to reduce the congestion of the access channel. Extended access barring (EAB) [27] and access class barring (ACB) [28], [29] are two load control mechanisms proposed by 3GPP. In the former mechanism, different access classes are assigned to the devices based on their service requirements of the device to relieve the load on the network. In this way, some non-emergency services will be temporarily blocked to improve access efficiency. In the latter mechanism, the base station periodically broadcasts an access probability factor. Each device selects a random value between [0, 1) before accessing the radio resources. When the selected value is less than the access probability factor, the device is allowed to transmit the data. Otherwise, the device will restart the uplink packet transmission after a random backoff time. The authors of [30] derived the optimal access probability factor of the ACB mechanism in the uplink transmission interval to calculate the random access channel capacity within average time period (i.e., the expected number of nodes successfully transmitting in an average period of time). Additionally, an analytical model was proposed in [31] to determine the expected delay for dynamically changing the ACB access probability factor.

III. SYSTEM MODEL

A. Assumptions

We consider a scenario where N_u URLLC devices coexist with N_e eMBB devices. The arrival processes of URLLC

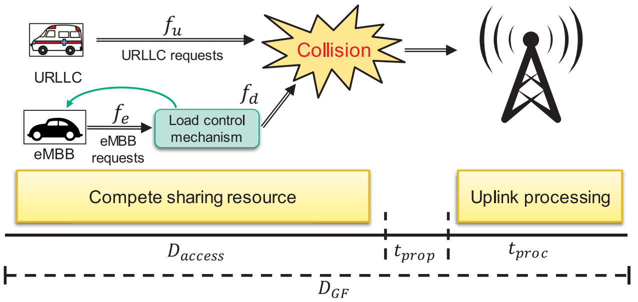


Fig. 2. System model of the considered URLLC and eMBB coexistence with load control mechanism.

and eMBB packets are modeled as Poisson processes with arrival rates λ_u and λ_e, respectively. Suppose that f_u URLLC requests will contend with f_e eMBB requests for the spectrum resource blocks. Both URLLC users and eMBB users adopt grant-free transmission [13], [32]. To ease the analysis, we define a grant-free opportunity (GFO) as the total time of single grant free transmission until a URLLC device either receives an acknowledgment (ACK) from a base station or its timer expires. If no packet collision occurs in the GFO, the base station will successfully receive the uplink data. Otherwise, the conflicting data will be retransmitted in the next GFO.

B. Load Control Mechanism

To reduce the data packet conflicts between URLLC and eMBB users, we exploit a load control mechanism in the system model as shown in Fig. 2. The role of the load control mechanism is to manage the number of eMBB devices accessing the network. We use the ACB mechanism because all the eMBB users are regarded as the same category in our system model. Before sending the data packet, the eMBB request randomly selects a value between 0 and 1. When the selected value is less than the predefined access probability, the packets will be allowed to transmit. Otherwise, the backlog of requests continue to wait and re-execute load control in the next GFO. Therefore, the number of eMBB requests changes from f_e to f_d after using load control. Finally, f_u URLLC requests and f_d eMBB requests will compete for spectrum resources based on the different spectrum sharing methods.

When we regulate the number of eMBB requests competing for spectrum resources, the successful transmission probability of URLLC request will increase. Therefore, we combine the design of load control mechanism (e.g., access class barring [29]) with the spectrum allocation schemes to mitigate the packet collisions in the coexisting system of URLLC and eMBB.

C. Network Model

The Markov network model of URLLC and eMBB device is shown in Fig. 3. The two states of Markov chain are defined as follows.

- *WAIT state:* The device is waiting for uplink data generation.

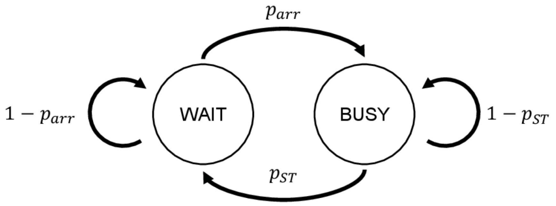


Fig. 3. The diagram of two-state Markov network model.

- *BUSY state:* The device is trying to send data packets to the base station and competing for spectrum resources with other devices.

Note that p_{arr} is the transition probability when the device status changes from WAIT to BUSY. Similarly, p_{ST} is the probability of the opposite state transition. The device maintains the same state in the WAIT and BUSY states with the probability $1 - p_{arr}$ and $1 - p_{ST}$, respectively. By using the Markov network model, we will derive the transition probabilities of the existing and the proposed spectrum allocation strategies in Sections V and VI, respectively.

D. Delay Model

The overall communication delay from the transmitter to the receiver is shown in Fig. 2. We define the time interval between a device generating a packet and receiving an feedback ACK as grant-free transmission delay (Φ), which can be written as

$$\Phi = D_{access} + 2t_{prop} + t_{proc}, \tag{1}$$

where D_{access} is the delay caused by the channel access and spectrum resource competition before packet transmissions. t_{prop} is the round-trip propagation time between the device and the base station, and t_{proc} is the processing time of the data packet received at the base station.

In grant-free setting, D_{access} is only affected by the time to compete for the spectrum resource because users transmit the data directly. Besides, we assume that t_{prop} is negligible [30]. t_{proc} is assumed to be 4 transmission time intervals (TTIs) for data decoding at the base station [32]. The TTI length (L_{TTI}) is defined as

$$L_{TTI} = N_{symbol} * L_{symbol}, \tag{2}$$

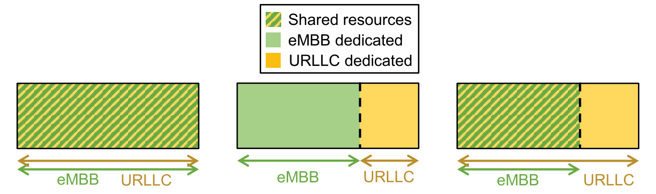
where N_{symbol} is the number of symbols and L_{symbol} is the corresponding symbol length. The number of symbol per TTI can be set to 2, 4 or 8, etc. However, we use a GFO of 0.25 msec in our simulation. The corresponding symbol length is 16.7 μsec. Therefore, the constant t_{proc} is very short compared to the length of GFO.

E. Performance Metrics

In our system-level simulations, two performance metrics of URLLC are defined in the following. The reliability is as a function of latency requirement, while the throughput is as a function of packet length and latency requirement.

- *Reliability:*

$$\mathcal{R}_u = \frac{\sum_{i=1}^{f_u} I\left(\Phi_u^{(i)} \le \tau_u\right)}{f_u} \tag{3}$$



(a) Fully overlap (FO) scheme. (b) Fully separate (FS) scheme [33]. (c) Partially overlap (PO) scheme [33].

Fig. 4. Illustrations of existing allocation schemes.

- *Throughput:*

$$\mathcal{M}_u=\sum_{i=1}^{f_u}\frac{\sigma_u^{(i)}}{\Phi_u^{(i)}}I\left(\Phi_u^{(i)}\le\tau_u\right) \tag{4}$$

where indicator function is denoted by $I(\cdot)$ and takes the value 1 or 0 on the condition that $\Phi_u^{(i)}\le\tau_u$ is true or not. τ_u is the latency requirement of URLLC. σ_i is the packet length of URLLC request i and $\Phi_u^{(i)}$ is the corresponding transmission delay. When we replace f_u with f_d and replace the subscript u with e, the performance metrics are applicable to eMBB.

IV. RESOURCE ALLOCATION APPROACHES

In URLLC and eMBB multiplexing, the spectrum allocation is important to reduce the collision probability of grant-free URLLC packets. To improve the latency performance of URLLC, some spectrum resources are reserved for URLLC transmission. In the following, we will introduce three existing network-centric spectrum allocation schemes, including dedicated, shared, and hybrid approaches. We further propose a user-centric hybrid scheme that can flexibly use spectrum resources according to the designed strategy.

A. Existing Resource Allocation Schemes [33]

As shown in Fig. 4, the network-centric spectrum allocation methods of dedicated, shared and hybrid are defined as fully separate (FS), fully overlap (FO) and partially overlap (PO). In the FO method, eMBB devices and URLLC devices share the same entire spectrum resources together. In the FS method, the entire spectrum resource is divided into two dedicated parts, which are exclusive used for URLLC devices and eMBB devices, respectively. The difference between the PO and the FS is that one part of the spectrum in the PO method is only reserved for URLLC devices. Another part of the spectrum is shared by URLLC devices and eMBB devices at the same time.

B. Proposed User-Initiated Probability Elastic Reservation (UPER) Scheme

Algorithm 1 Dynamic Update of the Proposed User-Initiated Probability Elastic Reservation (UPER) Scheme

```
initialize: WAIT state with $P_r(0) \leftarrow 1$;
if Packet arrives then
    Transit from WAIT state to BUSY state
    repeat
        Perform grant-free transmission
        if Receive ACK then
            Transit from BUSY state to WAIT state return
        else
            if Use the dedicated resources then
                $P_r(t+1) \leftarrow \left(P_r(t)+\frac{r}{\mathcal{S}+r}\right)/2$
            else
                $P_r(t+1) \leftarrow (P_r(t)+1)/2$
            end if
        end if
        Check stabilization
        if $|P_r(t+1)-P_r(t)| \le \epsilon$ then
            $P_r(t+1) \leftarrow P_r(t)$
        else
            Update $P_r(t+1)$
        end if
    until latency requirement is reached
    Transit from BUSY state to WAIT state
end if
```

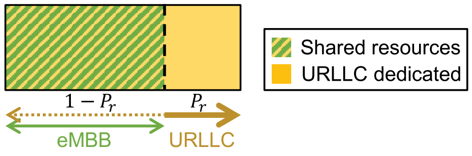


Fig. 5. Illustration of proposed user-initiated probability elastic reservation (UPER) scheme.

However, the network-centric spectrum allocation method may not be feasible for URLLC users with stringent latency. Therefore, we proposed a user-initiated probability elastic reservation (UPER) scheme to further improve the latency performance of URLLC when coexisting with eMBB users. The proposed UPER scheme is a hybrid spectrum allocation approach as shown in Fig. 5. The entire spectrum resources $(\mathcal{S}+r)$ are divided into two parts. Resources $\mathcal{S}$ is shared by eMBB devices and URLLC devices. On the other hand, resource r is proprietary to URLLC devices. To implement the user-centric approach, we design a resource decision parameter to determine that URLLC devices will transmit the data on the shared or the reserved spectrum resources. The decision parameter P_r is a selection probability of the reserved resources when the URLLC packet performs a grant-free transmission. In other words, the shared part of the spectrum can be selected with a probability of $1-P_r$. When P_r is equal to 1 and $\frac{r}{\mathcal{S}}$, the proposed UPER scheme can become the FS scheme and the PO scheme, respectively. Therefore, the UPER scheme has the advantages of these two schemes by utilizing the spectrum resources flexibly.

Here, we will introduce the policy of the proposed UPER method. The dynamic update algorithm of the proposed method is shown in Algorithm 1. The URLLC users can dynamically update P_r based on the status of the network congestion at each GFO until receiving an ACK from the base station. When a collision occurs and the last resources selected for uplink transmission is on the dedicated spectrum, the value

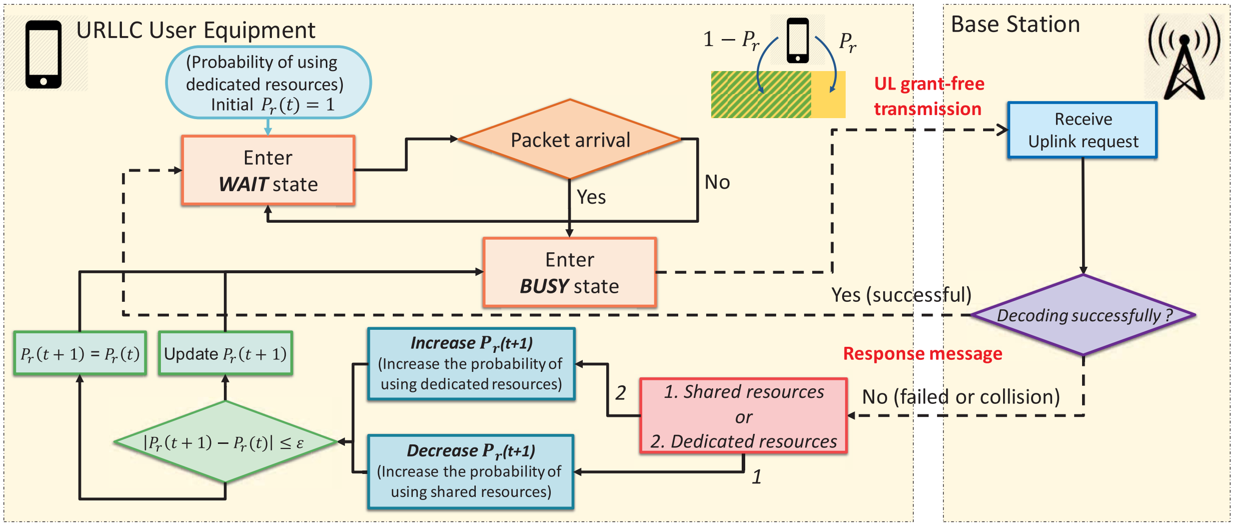


Fig. 6. Dynamic update flow chart of the proposed UPER scheme.

of P_r will be decreased to alleviate the congestion problem of URLLC packers in the next transmission. In contrast, the probability of using proprietary spectrum for URLLC packets will be increased to reduce the contention with eMBB packets. To stabilize the algorithm, we define $|P_r(t+1)-P_r(t)| \leq \epsilon$ as a condition. Note that ϵ is the difference of *Pr* between time t and time $t+1$. If ϵ is very small, P_r will not be updated. Otherwise, we update the new P_r to adjust the probability of using the dedicated spectrum.

In Fig. 6, we illustrate a flow chart of the proposed algorithm between the URLLC user and the base station. The value of P_r is initially set to 1 when the URLLC device is idle and no data is being transmitted. If the packet arrives, the status of the device will change from WAIT state to BUSY state. Then, the URLLC device performs grant-free transmission with $P_r = 1$ in the next GFO. According to the response message from the base station, the device will re-enter the waiting state after the successful transmission or adjust the value of P_r after the failed transmission. When the packet conflicts or fails, P_r will be dynamically updated by the aforementioned policy.

V. PROBABILISTIC ANALYSIS FOR EXISTING RESOURCE ALLOCATION SCHEMES

In this section, we first calculate the approximate number of competing request to explain the motivation for using load control. Then, we derive the transition probabilities of Markov network model as mentioned in Section III.

A. The Approximate Number of Competing Request

In [30], the typical probability that radio access network (RAN) can successfully transmit at each time slot can be written as

$$p_{ST}(n) = \left(1 - \frac{1}{\mathcal{Q}}\right)^{n-1}, \tag{5}$$

where n is the number of competing requests and $\mathcal{Q}$ is the number of total spectrum resources for competition. The average number of requests successfully served by the base station at each GFO is

$$N_{ST}(n) = n\left(1 - \frac{1}{\mathcal{Q}}\right)^{n-1}. \tag{6}$$

When $\mathcal{Q}$ is large, we calculate that

$$n^* = \lim_{\mathcal{Q}\to\infty} \frac{1}{\log\left(\frac{\mathcal{Q}}{\mathcal{Q}-1}\right)} \cong \mathcal{Q}. \tag{7}$$

Hence, we know that the approximate number of the competing requests is same as the total number of available spectrum resources. Based on the above derivation, f_d can be written as

$$f_d = \min(f_e, \mathcal{Q}). \tag{8}$$

The approximation result motivates the use of a load control mechanism to limit the number of eMBB requests that compete for the spectrum resources. By substituting (7) for *n* in (6), we have

$$N_{ST}(\mathcal{Q}) = \mathcal{Q}\left(1 - \frac{1}{\mathcal{Q}}\right)^{\mathcal{Q}-1} \simeq \frac{\mathcal{Q}}{e} + \frac{1}{2e}. \tag{9}$$

B. Probability of Request Arrival

The probability of the request arrival is derived under the condition that at least one request arrives in each GFO. As shown in Fig. 2, the device will transit from WAIT to BUSY state when a request arrives. Hence, the transition probability p_{arr} is obtained in Lemma 1.

Lemma 1 (Transition Probability From WAIT State to BUSY State): If the packet arrivals of devices follow Poisson process with arrival rate λ, the probability mass function (PMF) of Poisson distribution is *P*(.). The transmission probability from WAIT state to BUSY state p_{arr} can be written as

$$p_{arr} = 1 - e^{-\lambda}. \tag{10}$$

The PMF of the Poisson random variable k is

$$P(X=k)=\frac{e^{-\lambda}\lambda^{k}}{k!},\tag{11}$$

where k is the number of the packet arrivals. The probability of at least one request arrives at a certain time interval can be expressed as

$$p_{arr}=1-P(0).\tag{12}$$

By replacing (12) with (11), we have

$$\begin{aligned}p_{arr}&=1-\frac{e^{-\lambda}\lambda^{0}}{0!}\\&=1-e^{-\lambda}.\end{aligned}\tag{13}$$

C. Probability of Successful Transmission

In the following, we will analyze the probability of successful transmission for the three existing spectrum allocation schemes as shown in Fig. 4. Note that the number of URLLC requests and the number of eMBB requests are f_u and f_e in uplink grant-free transmission, respectively. After we adopt the load control mechanism, the number of eMBB requests changes from f_e to f_d.

- *Fully Overlap (FO) Scheme:* In FO scheme, both URLLC UEs and eMBB UEs share all spectrum resources ($\mathcal{S}+r$) to perform grant-free transmission. The probability that URLLC and eMBB requests can successfully perform radio access can be expressed as

$$p_{ST}^{(u)}=p_{ST}^{(e)}=\left(1-\frac{1}{\mathcal{S}+r}\right)^{f_u+f_d-1},\tag{14}$$

  where $p_{ST}^{(u)}$ and $p_{ST}^{(e)}$ are the probability of successful transmission for the URLLC UE user and the eMBB user, respectively. In this case, URLLC requests and the eMBB requests have the same priority. Thus, the probability of successful transmissions for two services have the same formula.
- *Fully Separate (FS) Scheme:* In FS scheme, the whole spectrum resources are divided into two parts to different service priorities. One part of the spectrum resource defined as $\mathcal{S}$ is for eMBB users. The other part of the spectrum resources defined as *r* is for URLLC users. That is, the URLLC users and the eMBB users each have their exclusive spectrum to transmit their packets and do not affect each other. The successful transmission probability of URLLC devices can be expressed as

$$p_{ST}^{(u)}=\left(1-\frac{1}{r}\right)^{f_u-1}.\tag{15}$$

  Similarly, the successful transmission probability of eMBB devices can be expressed as

$$p_{ST}^{(e)}=\left(1-\frac{1}{\mathcal{S}}\right)^{f_d-1}.\tag{16}$$

  Obviously, the probability of successful transmission for URLLC traffic is only affected by the number of URLLC requests in BUSY state. When the amount of the spectrum resources reserved for URLLC increases, the successful transmission probability of the URLLC request increases.
- *Partially Overlap (PO) Scheme:* The PO scheme is a hybrid spectrum resource allocation approach that combines the benefits of the FO and the FS schemes. URLLC users can use the entire radio spectrum resources $\mathcal{S}+r$. However, the eMBB UEs exploit only a portion of the total resources defined as $\mathcal{S}$ to transmit the packets. The successful transmission probability of URLLC devices can be expressed as

$$\begin{aligned}p_{ST}^{(u)}=&\left(1-\frac{1}{\mathcal{S}+r}\right)^{f_u-1}\\&\times\left[\frac{\mathcal{S}}{\mathcal{S}+r}\left(1-\frac{1}{\mathcal{S}}\right)^{f_d}+\frac{r}{\mathcal{S}+r}\right].\end{aligned}\tag{17}$$

  The successful transmission probability of eMBB devices can be expressed as

$$p_{ST}^{(e)}=\left(1-\frac{1}{\mathcal{S}}\right)^{f_d-1}\left(1-\frac{1}{\mathcal{S}+r}\right)^{f_u}.\tag{18}$$

  Note that in the FO scheme, because the number of eMBB uplink requests is greater than the number of URLLC uplink requests, the probability of successful transmission for the URLLC UE is mainly affected by the number of eMBB requests. Therefore, some spectrum resources are provided exclusively for URLLC requests to avoid packet collisions with eMBB requests. In this way, only certain URLLC requests will conflict with eMBB requests, thereby improving the reliability of URLLC.

VI. PROBABILITY ANALYSIS OF PROPOSED UPER SCHEME

To provide timely URLLC services, we propose a technique for dynamically adapting the probability on the reserved spectrum resources of URLLC users. In the UPER scheme, spectrum resource allocation is same as the PO scheme. We design a resource selection parameter P_r in the proposed method to avoid the severe packet collisions. The parameter P_r will dynamically change the probability of using the dedicated spectrum for URLLC user. To support the sporadic traffic of URLLC, the user-centric P_r can be updated based on the proposed algorithm mentioned in Section IV. The probability of successful transmission for URLLC devices can be expressed as

$$\begin{aligned}p_{ST}^{(u)}=&(1-P_r)\left\{\left(1-\frac{1}{\mathcal{S}}\right)^{f_d}\left(1-\frac{1-P_r}{\mathcal{S}}\right)^{f_u-1}\right\}\\&+P_r\left(1-\frac{P_r}{r}\right)^{f_u-1}.\end{aligned}\tag{19}$$

Similarly, the probability of successful transmission for eMBB devices can be expressed as

$$p_{ST}^{(e)}=\left(1-\frac{1}{\mathcal{S}}\right)^{f_d-1}\left(1-\frac{1-P_r}{\mathcal{S}}\right)^{f_u}.\tag{20}$$

Proof: To analyze the probability of successful transmission, we first need to obtain the non-collision probability of both URLLC and eMBB packets. Under the condition that a given URLLC user uses the shared spectrum, the probability that at least one user among other URLLC users does not select the same resource is represented as

$$p_{ns}^{(u)} = 1 - \frac{1-P_r}{\mathcal{S}}. \tag{21}$$

Given that a certain URLLC user uses the dedicated spectrum, the probability that at least one of other URLLC UEs does not select the same resource is calculated as

$$p_{nr}^{(u)} = 1 - \frac{P_r}{r}. \tag{22}$$

On the other hand, eMBB users share the shared spectrum with the URLLC users. Therefore, only under the condition that the given URLLC UEs use the shared spectrum resources, we calculate the probability that at least one of eMBB users does not select the specific shared resource is expressed as

$$p_{ns}^{(e)} = 1 - \frac{1}{\mathcal{S}}. \tag{23}$$

The probability of successful transmission for the URLLC user ($p_{ST}^{(u)}$) is equal to the sum of the conditional probabilities for a specific URLLC user transmitting the packet on the shared and dedicated spectrum resource. Hence, we can obtain $p_{ST}^{(u)}$ as follows.

$$\begin{aligned} p_{ST}^{(u)} = (1-P_r)\left(\left(p_{ns}^{(e)}\right)^{f_d}\left(p_{ns}^{(u)}\right)^{f_u-1}\right) \\ + P_r\left(p_{nr}^{(u)}\right)^{f_u-1}. \end{aligned} \tag{24}$$

By substituting (21), (22) and (23) with (24), we have

$$\begin{aligned} p_{ST}^{(u)} = (1-P_r)\left(\left(1-\frac{1}{\mathcal{S}}\right)^{f_d}\left(1-\frac{1-P_r}{\mathcal{S}}\right)^{f_u-1}\right) \\ + P_r\left(1-\frac{P_r}{r}\right)^{f_u-1}. \end{aligned}$$

Besides, the probability of successful transmission for the eMBB user $p_{ST}^{(e)}$ can be written as

$$p_{ST}^{(e)} = \left(p_{ns}^{(e)}\right)^{f_d-1}\left(p_{ns}^{(u)}\right)^{f_u}. \tag{25}$$

We use (21) and (23) instead of (25) to obtain

$$p_{ST}^{(e)} = \left(1-\frac{1}{\mathcal{S}}\right)^{f_d-1}\left(1-\frac{1-P_r}{\mathcal{S}}\right)^{f_u}. \tag{26}$$

The packet collision probabilities of URLLC and eMBB can be respectively written as

$$p_c^{(u)} = 1 - p_{ST}^{(u)}, \tag{27}$$

and

$$p_c^{(e)} = 1 - p_{ST}^{(e)}. \tag{28}$$

■

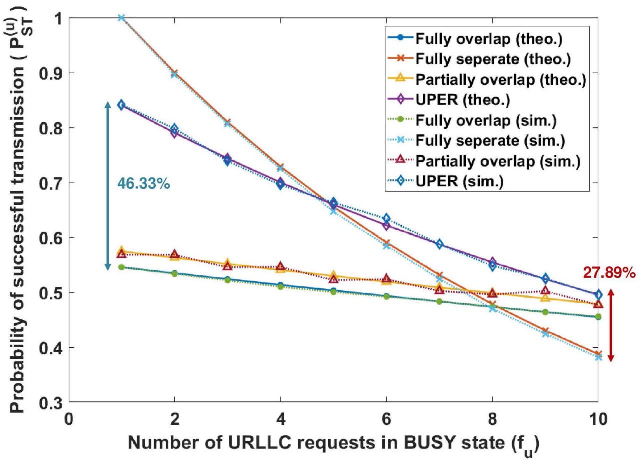


Fig. 7. Probability of successful transmission for URLLC UE with various number of URLLC in FO scheme, FS scheme, PO scheme, and the proposed UPER scheme in the case of $r = 10$, $\mathcal{S} = 40$ and $P_r = 0.7$.

VII. NUMERICAL RESULTS

Here, we first show the theoretical results in Section VII-A. Then, we discuss the system-level simulations in Sections VII-B to VII-D to evaluate a variety of spectrum allocation methods. In our simulations, we show the average reliability and throughput performance of URLLC users and eMBB users for various spectrum allocation methods. We compare the performance impact of using and not using load control. We further show that when the URLLC packet load is high, the different reserved widths of the dedicated spectrum can cause changes in reliability performance.

A. Analytical Results

In the following, we show the analytical results derived in Section V and VI. For ease of analysis, we assume that the size of the spectrum resources is the same. We divide the entire spectrum into 50 segments in the frequency domain. The amount of the reserved spectrum and the shared are $r = 10$ and $\mathcal{S} = 40$, respectively. Fig. 7 shows the successful transmission probability of URLLC ($p_{ST}^{(u)}$) for different number of URLLC requests (f_u) when $P_r = 0.7$ and $f_d = 30$. Compared with the PO scheme, serious data packet conflicts will occur in the FO scheme because URLLC users and eMBB users share the entire spectrum resources at the same time. When the number of URLLC packets competing for spectrum resources is small, the probability of successful transmission of the FS scheme is higher than that of the PO scheme. This is because URLLC user only sends data packets in dedicated resources, and the packets will not conflict with eMBB. On the contrary, when the number of URLLC data packets increases, the dedicated spectrum resources are insufficient to accommodate URLLC requests, resulting in serious collisions. However, the above problem can be solved by the proposed UPER. The UPER method has the better $p_{ST}^{(u)}$ than 27.89% of the FS method, especially when the URLLC traffic load is high. On the other hand, the successful transmission probability of the proposed scheme is 27.89% higher than that of the FS scheme when

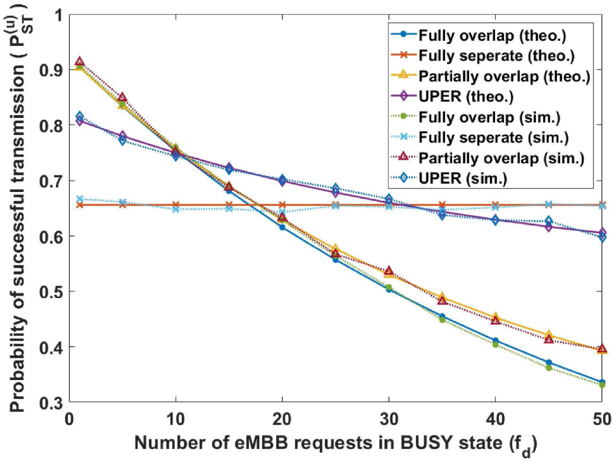
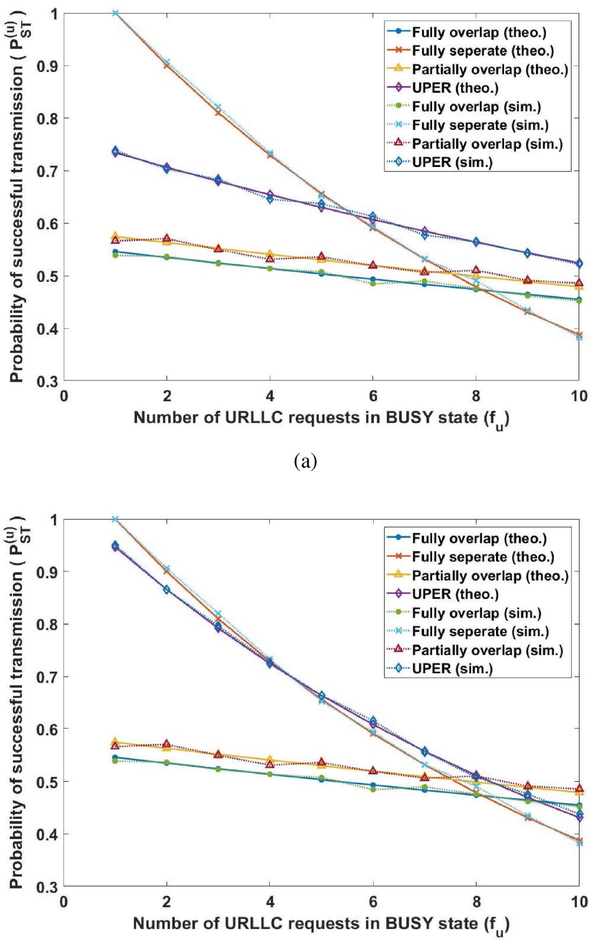


Fig. 8. Probability of successful transmission for URLLC UE with various number of eMBB in FO scheme, FS scheme, PO scheme, and the proposed UPER scheme in the case of $r = 10$, $\mathcal{S} = 40$ and $P_r = 0.7$.



(a)

(b)

Fig. 9. The impact of P_r on the probability of successful transmission for URLLC UE with various number of URLLC in FO scheme, FS scheme, PO scheme, and the proposed UPER scheme in the case of $r = 10$, $\mathcal{S} = 40$: (a) $P_r = 0.5$; (b) $P_r = 0.9$.

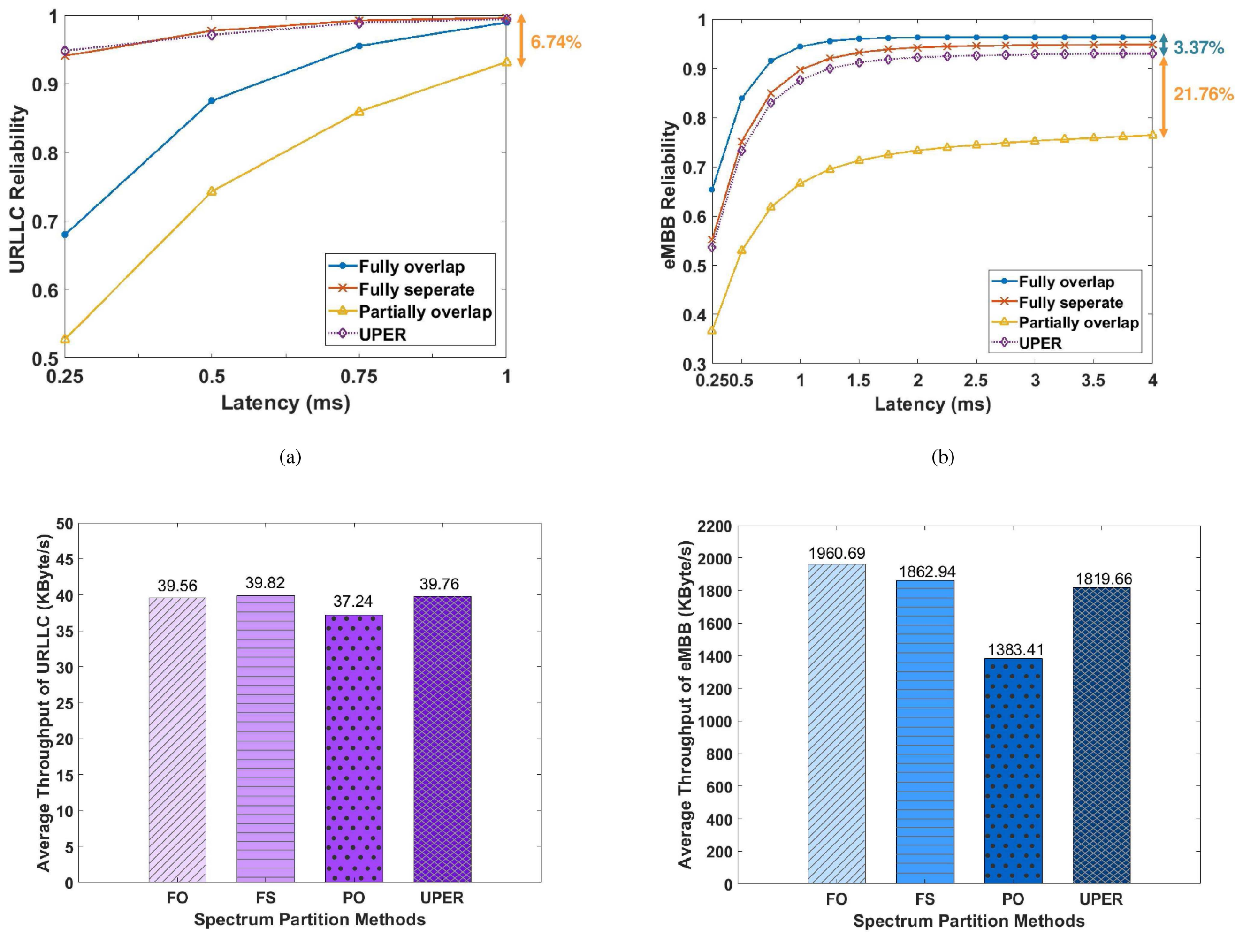
URLLC traffic load is low. If $P_r < 0.7$, the curve of $p_{ST}^{(u)}$ is closer to the PO method as shown in Fig. 9(a). URLLC users tend to send packets on the shared spectrum resources. If $P_r > 0.7$, the curve of $p_{ST}^{(u)}$ of the UPER is approximate to the FS method as shown in Fig. 9(b), thereby reducing the collisions with the packets requested by eMBB.

In Fig. 8, we show the change of $p_{ST}^{(u)}$ under different number of eMBB requests (f_d) when $P_r = 0.7$ and $f_u = 5$. When f_d increases, the $p_{ST}^{(u)}$ in the FS method does not change. The reason is that the URLLC users have their own the dedicated spectrum without being affected by the number of eMBB requests. In the FO, PO and UPER scheme, $p_{ST}^{(u)}$ gradually decreases as f_d increases because URLLC users share all or part of the spectrum resources with eMBB users. For the PO scheme, the URLLC user will not share the entire spectrum with the eMBB user. The URLLC UEs will not share the entire spectrum in the PO scheme with eMBB UEs. Thus, the $p_{ST}^{(u)}$ of the PO scheme is higher than that of the FO scheme in the case of high eMBB traffic load (e.g., f_e is greater than 30). Compared to the PO and the FO methods, the proposed UPER has the designed the resource selection parameter, which can effectively reduce the impact of f_d on the successful transmission probability of URLLC user.

B. System Performance of URLLC and eMBB

Next, we will consider the average reliability and throughput performance of URLLC users and eMBB users within 1 msec and 4 msec latency requirements in system-level simulations, respectively. The system-level simulation parameters are listed in TABLE I. The numbers of URLLC users and eMBB users are $N_u = 10$ and $N_e = 10$, respectively. The packet arrival rates of URLLC users (λ_u) and eMBB users (λ_e) are set to different values because URLLC packets arrive less frequently that eMBB packets. We consider that the corresponding packet sizes of URLLC and eMBB are 32 and 160 bytes, respectively. The slot length of GFO is defined in Section III. We assume that all the users transmit their packets to the base station by grant-free transmission. When no packet collision occurs in the GFO, the base station will successfully receive the uplink data. Otherwise, the data will be retransmitted in the next GFO. For URLLC with the 1 msec latency requirement, the packets can be transmitted up to four times, including the first transmission and three retransmission. Similarly, the packets can be transmitted up to 16 times for eMBB with the 4 msec delay requirement.

Fig. 10 shows the average reliability and average throughput of URLLC and eMBB, respectively. We consider that the time-critical URLLC applications happen much less frequently than eMBB in the case of $\lambda_e = 8$ and $\lambda_u = 0.5$. The average reliability of URLLC within latency requirement of 1 msec is shown in Fig. 10(a). We observe that the FS method is superior to other methods. The reason is that URLLC users send data packets on the dedicated spectrum without conflicting with eMBB packets. When the URLLC traffic load is highly low, the FS method has the better performance than other spectrum allocation methods. The reliability of the FO method is better than that of the PO method because URLLC users compete with eMBB for resources in the entire spectrum rather than in the shared spectrum. Thus, the probability that URLLC



(a) (b)

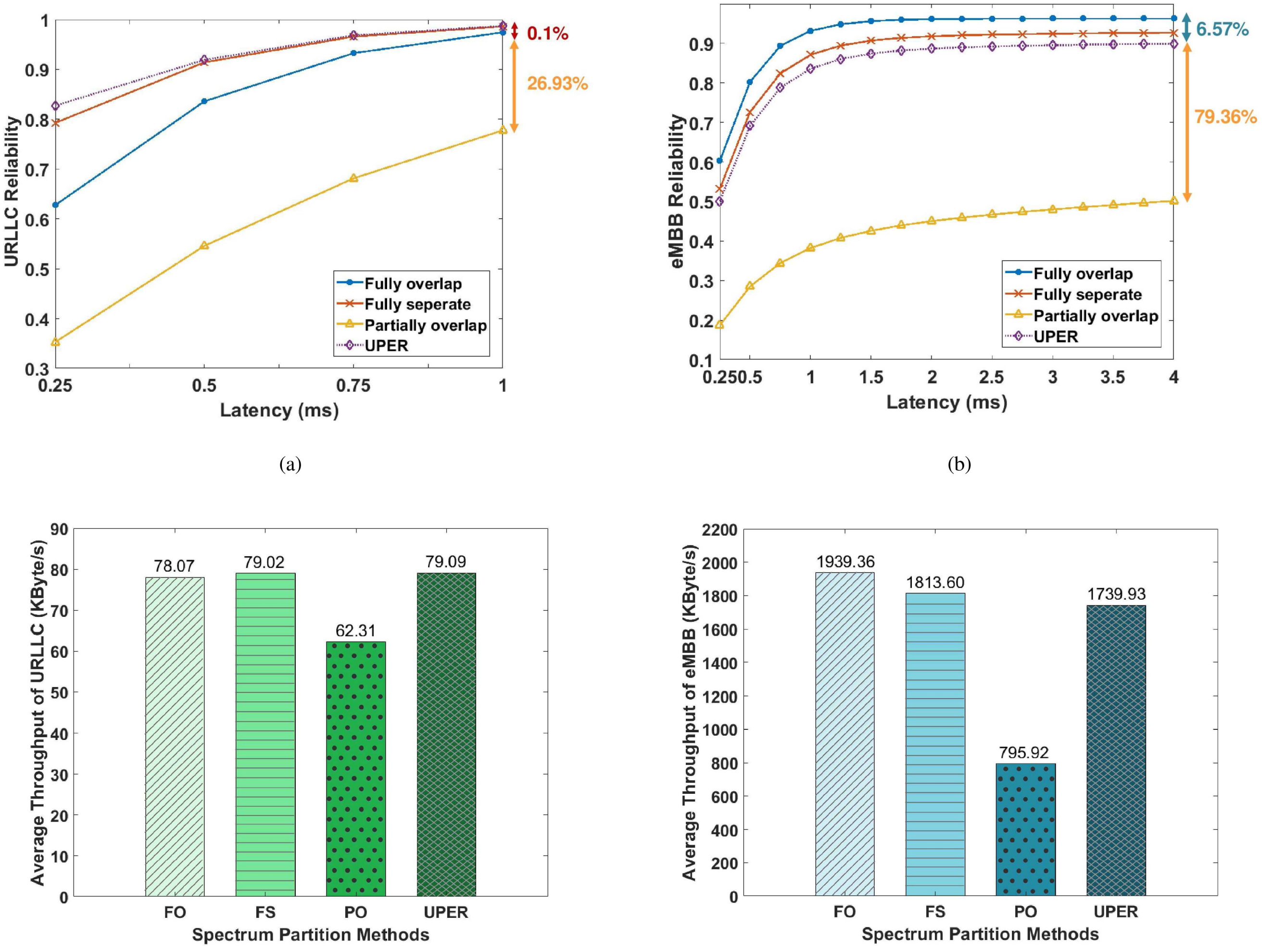
(c) (d)

Fig. 10. System performance comparisons between URLLC and eMBB with load control in the case of $\lambda_e = 8$ and $\lambda_u = 0.5$: (a) URLLC reliability with latency requirement of 1 msec; (b) eMBB reliability with latency requirement of 4 msec; (c) URLLC throughput with latency requirement of 1 msec; (d) eMBB throughput with latency requirement of 4 msec.

packets collide with eMBB packets is small. Nevertheless, the reliability of the proposed UPER is 99.45%, which is slightly worse than the FS method, and is 6.74% higher than the PO method with the worst reliability. The corresponding average throughput of URLLC is shown in Fig. 10(c). Due to the low URLLC traffic load, the average throughput of different spectrum allocation methods is not significantly different.

The average reliability of eMBB within latency requirement of 4 msec is shown in Fig. 10(b). The FO is superior to other methods because eMBB users use the entire spectrum to transmit data packets and are not prone to collide with other URLLC and eMBB users. In comparison, the average throughput of eMBB in our method is only 3.37% lower than the FO method and 21.76% higher than the PO with the worst reliability. The corresponding average throughput of eMBB is shown in Fig. 10(d). We observe that although the average throughput of the proposed method is only higher than that of the PO method, it is similar to the FO and the FS methods.

Fig. 10 shows the average reliability and average throughput of URLLC and eMBB, respectively. We consider that the time-critical URLLC applications happen much less frequently than eMBB in the case of $\lambda_e = 8$ and $\lambda_u = 0.5$. On the other hand, we show the reliability and throughput performance when eMBB and URLLC packets arrive at a rate of $\lambda_e = 8$ and $\lambda_u = 1$, respectively. In Fig. 11, the trend of the curve is similar to the trend in Fig. 10. When the packet arrival rate of URLLC increases, the reliability and throughput performance of URLLC and eMBB decreases in all methods. In terms of the URLLC reliability, the performance of the proposed UPER is 98.73%, which is almost the same as the FS method, and superior to other methods. At this time, the eMBB reliability of the UPER is 6.57% lower than the PO method and 79.36% higher than the PO method. The performance degradation of PO method is greater than that in the Fig. 10. The reason is that the probability of collision between eMBB packets sent in the dedicated spectrum and URLLC packets is higher than the probability of collision between eMBB packets sent in the entire spectrum and URLLC packets.



(a) (b)

(c) (d)

Fig. 11. System performance comparisons between URLLC and eMBB with load control in the case of $\lambda_e = 8$ and $\lambda_u = 1$: (a) URLLC reliability with latency requirement of 1 msec; (b) eMBB reliability with latency requirement of 4 msec; (c) URLLC throughput with latency requirement of 1 msec; (d) eMBB throughput with latency requirement of 4 msec.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
N_u	10
N_e	10
λ_e	8 (packets/msec)
λ_u	{0.5, 1, 3} (packets/msec)
σ_u	32 bytes
σ_e	160 bytes
GFO	0.25 msec

*GFO: time unit of grant-free transmission time.

TABLE II
IMPROVEMENT RATIO OF URLLC RELIABILITY AFTER USING LOAD CONTROL IN THE CASE OF $\lambda_e = 8$ AND $\lambda_u = 1$

Spectrum Partition Methods	Improvement Ratio
UPER	2.39 %
Fully overlap (FO)	51.75 %
Partially overlap (PO)	16.44 %
Fully separate (FS)	0 %

C. Comparison of URLLC Reliability With and Without Load Control Mechanism

Fig. 12 shows the URLLC reliability performance when eMBB has no load control mechanism under the conditions of $\lambda_e = 8$ and $\lambda_u = 1$. Compared with the performance of Fig. 11(a), the reliability performance of URLLC under uncontrolled eMBB traffic load is worse than that under controlled eMBB traffic load. TABLE II lists the improvement ratio of URLLC reliability after we use load control mechanism for eMBB traffic. We found that the use of the load control has the greatest impact on the reliability of the FO method. This is because most eMBB users who transmit data packets in the whole spectrum will have serious conflicts with URLLC or eMBB, when the number of eMBB access networks is not limited. Therefore, if we do not control a large number of eMBB accessing networks, URLLC performance will decrease, especially when the number of URLLC packets increases.

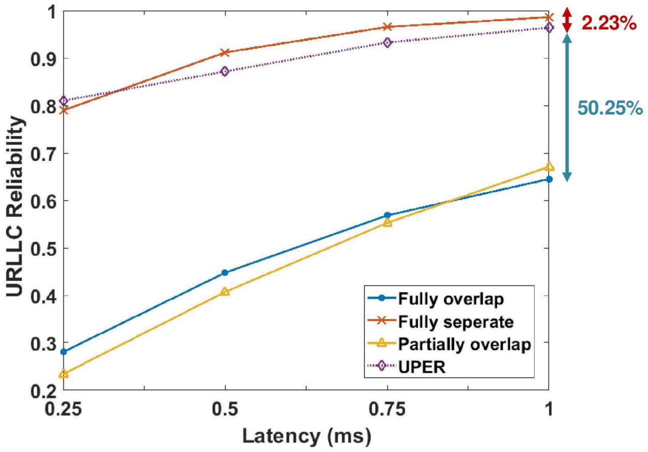


Fig. 12. URLLC reliability performance without eMBB load control mechanism in the case of $\lambda_e = 8$ and $\lambda_u = 1$.

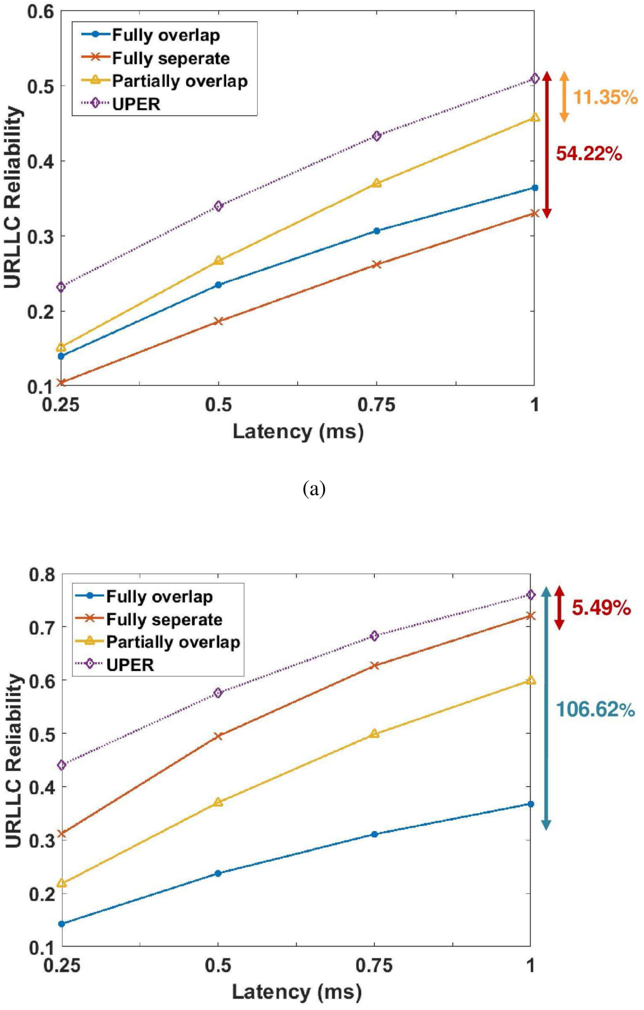
TABLE III
THE RELIABILITY IMPROVEMENT RATE OF THE DEDICATED SPECTRUM INCREMENT IS LISTED IN THE CASE OF $\lambda_e = 8$ AND $\lambda_u = 3$

Spectrum Partition Methods	Improvement Ratio
Fully separate (FS)	118.39 %
UPER	49.39 %
Partially overlap (PO)	31.15 %
Fully overlap (FO)	1.09 %

D. Performance Variation for High URLLC Packet Load

As shown in Fig. 13, we present the changes in the reliability performance of URLLC under different reserved widths of the dedicated spectrum when eMBB traffic has a load control. We consider the case where URLLC packet load is high, for example, $\lambda_e = 8$ and $\lambda_u = 3$. In Fig. 13(a), we show the URLLC reliability performance when the reserved widths of the dedicated spectrum is 10. We observe that the URLLC reliability in the FS method is the worst. Reserving insufficient dedicated spectrum cause URLLC users to have severe packet collisions on the dedicated spectrum, thereby greatly degrading the performance of URLLC. However, the reliability performance of the proposed UPER method is 54.22% higher compared with the FS method and even better than other methods. Comparing Figs. 13(a) and 13(b), the reliability performance of all the methods is improved by increasing the reserved widths of the dedicated spectrum from $r = 10$ to $r = 15$. The reliability improvement ratio of the dedicated spectrum increment is listed in TABLE III. When the amount of dedicated spectrum resources increases, the reliability of URLLC in the FS method is significantly improved. In addition, the UPER method also outperform than other spectrum partition methods.

In Fig. 14, we show the URLLC throughput as per the impact of packet length. The URLLC throughput with packet length of 50 bytes is 56.25% higher than that of 32 bytes. That is, the throughput performance of URLLC is proportional to the packet length in our considered case. From the figure, one can also observe that the performance of the proposed UPER is superior to other spectrum allocation methods, including the FS, FO and PO.



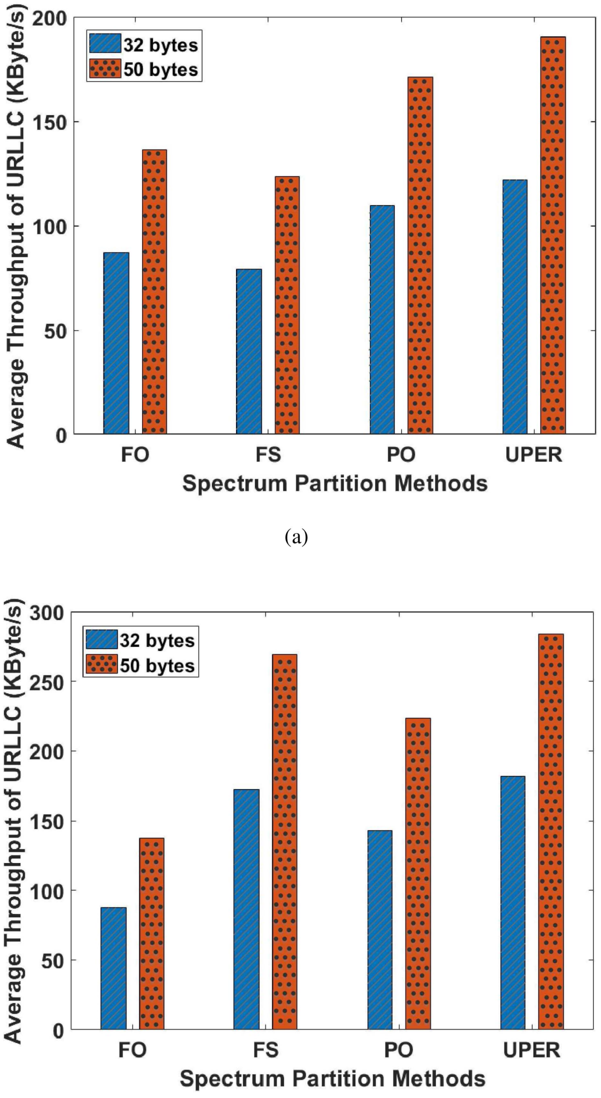
(a)

(b)

Fig. 13. Comparisons of reliability performance with eMBB load control mechanism under different dedicated spectrum width in the case of $\lambda_e = 8$ and $\lambda_u = 3$: (a) The dedicated spectrum resource $r = 10$; (b) The dedicated spectrum resource $r = 15$.

VIII. CONCLUSION

In this paper, we proposed an on-device data-driven spectrum allocation algorithm, called the user-initial probability elastic resource (UPER) scheme, to dynamically adjust the probability of selecting shared and dedicated spectrum for URLLC based on the packet collision status coexisting with eMBB customers. The feasibility of UPER was evaluated first analytically according to the derived closed-form expressions for the successful transmission probability of the UPER, fully separated (FS), fully overlapped (FO), and partially overlapped (PO) four spectrum allocation methods considered in this paper. Further, the proposed UPER spectrum allocation method was compared to other three considered approaches by extensive simulations under different conditions, including URLLC traffic load, eMBB load control and URLLC packet length. We observed the following phenomena:



(a)

(b)

Fig. 14. The impacts of URLLC packet length on throughput performance with eMBB load control mechanism under different dedicated spectrum width in the case of $\lambda_e = 8$ and $\lambda_u = 3$: (a) The dedicated spectrum resource $r = 10$; (b) The dedicated spectrum resource $r = 15$.

- The reliability of the UPER method performed the best, and is 54%, 32% and 11% higher than the FS, PO and FO spectrum allocation methods when the URLLC traffic load is high.
- The UPER method can almost reach the reliability and throughput performance as the FS method when the URLLC traffic load is low.
- The proposed method with the load control mechanism can respectively improve the reliability performance of the FO, PO and UPER methods by 52%, 16% and 2% compared to the case without load control.
- The URLLC throughput with the packet length of 50 bytes is 56% higher than that of 32 bytes. That is, the throughput performance of URLLC is proportional to the packet length in our considered case.

To sum up, we believe that the proposed on-device data-driven UPER spectrum management framework with load control mechanism can satisfy the stringent latency and reliability requirement for URLLC in a sharing spectrum environment with eMBB customers.

In the future, many interesting research topics can be extended from the current study. First, how to design an method to quickly calculate probability of using dedicated spectrum resources is an important and interesting research topic. Secondly, it would be worthwhile further investigating the optimal bandwidth of the reserved exclusive spectrum for URLLC by using machine learning techniques based on the long-term network status. Last, the system parameters affecting the performance of the URLLC/eMBB/mMTC coexisting system are from different protocol layers, including the modulation and coding scheme in the physical layer and the number of requested resource blocks, packet length and different access protocols in the MAC layer. Hence, designing the optimal URLLC/eMBB/mMTC coexistence system from a PHY/MAC cross-layer perspective is still an open issue.

REFERENCES

[1] A. Osseiran *et al.*, “Scenarios for 5G mobile and wireless communications: The vision of the METIS project,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[2] P. Popovski, “Ultra-reliable communication in 5G wireless systems,” in *Proc. IEEE Int. Conf. 5G Ubiquitous Connectivity*, 2014, pp. 146–151.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[4] N. A. Johansson, Y.-P. E. Wang, E. Eriksson, and M. Hessler, “Radio Access for ultra-reliable and low-latency 5G communications,” in *Proc. IEEE Int. Conf. Commun. Workshop*, 2015, pp. 1184–1189.

[5] V. Hytönen, Z. Li, B. Soret, and V. Nurmela, “Coordinated multi-cell resource allocation for 5G ultra-reliable low latency communications,” in *Proc. IEEE Eur. Conf. Netw. Commun.*, 2017, pp. 1–5.

[6] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, “MAC layer enhancements for ultra-reliable low-latency communications in cellular networks,” in *Proc. IEEE Int. Conf. Commun. Workshops*, 2017, pp. 1005–1010.

[7] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai. (2017). *A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions*. [Online]. Available: https://arxiv.org/abs/1708.02562

[8] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim. (2017). *Introduction to Ultra Reliable and Low Latency Communications in 5G*. [Online]. Available: https://arxiv.org/abs/1704.05565

[9] “Study on RAN improvements for machine-type communications,” 3GPP, Sophia Antipolis, France, Rep. TR37.868, Sep. 2011.

[10] R. Kassab, O. Simeone, P. Popovski, and T. Islam, “Non-orthogonal multiplexing of ultra-reliable and broadband services in fog-radio architectures,” *IEEE Access*, vol. 7, pp. 13035–13049, 2019.

[11] A. A. Esswie and K. I. Pedersen, “Opportunistic spatial preemptive scheduling for URLLC and eMBB coexistence in multi-user 5G networks,” *IEEE Access*, vol. 6, pp. 38451–38463, 2018.

[12] A. Bairagi, M. S. Munir, M. Alsenwi, N. H. Tran, and C. S. Hong, “A matching based coexistence mechanism between eMBB and URLLC in 5G wireless networks,” in *Proc. 34th ACM/SIGAPP Symp. Appl. Comput.*, New York, NY, USA, 2019, pp. 2377–2384.

[13] A. T. Abebe and C. G. Kang, “Comprehensive grant-free random access for massive & low latency communication,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–6.

[14] G. Wunder, P. Jung, and M. Ramadan, “Compressive random access using a common overloaded control channel,” in *Proc. IEEE Global Commun. Conf. Workshops*, 2015, pp. 1–6.

[15] R. B. Abreu *et al.*, “On the multiplexing of broadband traffic and grant-free ultra-reliable communication in uplink,” in *Proc. IEEE Veh. Technol. Conf.*, 2019, pp. 1–6.

[16] "Basic grant-free transmission procedure for URLLC," Nokia, Alcatel-Lucent Shanghai Bell, Reno, NV, USA, 3GPP RAN1 #87, R1-1612250, Nov. 2016.

[17] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, Jun. 2019.

[18] Y. Han, S. E. Elayoubi, A. Galindo-Serranoa, V. S. Varma, and M. Messai, "Periodic radio resource allocation to meet latency and reliability requirements in 5G networks," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–6.

[19] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5G ultra-reliable and low-latency systems design," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–5.

[20] "Issues on URLLC resource allocation," 3GPP, ZTE, Spokane, WA, USA, 3GPP TSG RAN WG1 Meeting #95, R1-1812392, Nov. 2018.

[21] "On UL multiplexing of URLLC and eMBB transmissions," 3GPP, Huawei, HiSilicon, Athens, Greece, 3GPP TSG RAN WG1 Meeting #88, R1-1701666, Feb. 2017.

[22] "LTE URLLC and eMBB multiplexing," 3GPP, Huawei, HiSilicon, Athens, Greece, 3GPP TSG RAN WG1 Meeting #92, R1-1801390, Feb. 2018.

[23] Z. Zhou, R. Ratasuk, N. Mangalvedhe, and A. Ghosh, "Resource allocation for uplink grant-free ultra-reliable and low latency communications," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 2465–2577.

[24] "eMBB and URLLC dynamic multiplexing and preemption indication on the uplink," 3GPP, Qualcomm Incorporat., Sanya, China, 3GPP TSG RAN WG1 Meeting #92b, R1-1804820, Feb. 2018.

[25] L.-C. Wang and C.-W. Wang, "Spectrum management techniques with QoS provisioning in cognitive radio networks," in *Proc. IEEE 5th Int. Symp. Wireless Pervasive Comput.*, 2010, pp. 116–121.

[26] L.-C. Wang, C.-W. Wang, and K.-T. Feng, "A queueing-theoretical framework for QoS-enhanced spectrum management in cognitive radio networks," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 18–26, Dec. 2011.

[27] R.-G. Cheng, J. Chen, D.-W. Chen, and C.-H. Wei, "Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 2956–2968, Jun. 2015.

[28] A. Larmo and R. Susitaival, "RAN overload control for machine type communications in LTE," in *Proc. IEEE Global Commun. Conf. Workshops*, 2012, pp. 1626–1631.

[29] "Access class barring and overload protection (ACBOP)," 3GPP, Sophia Antipolis, France, Rep. TR23.898, Apr. 2005.

[30] M. Koseoglu, "Lower bounds on the LTE-A average random access delay under massive M2M arrivals," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2104–2115, May 2016.

[31] S. Duan, V. Shah-Mansouri, and V. W. Wong, "Dynamic access class barring for M2M communications in LTE networks," in *Proc. IEEE Global Commun. Conf. Workshops*, 2013, pp. 4747–4752.

[32] "UL grant free transmission for eMBB," CATT, Reno, NV, USA, 3GPP TSG RAN WG1 Meeting #87, R1-1611399, Nov. 2016.

[33] K.-D. Lee, S. Kim, and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," in *Proc. IEEE Global Commun. Conf. Workshops*, 2011, pp. 373–377.

Shu-Feng Cheng (Member, IEEE) received the B.S. degree in electrical engineering from National Taiwan Normal University, Taipei, Taiwan, in 2014. She is currently pursuing the Ph.D. degree in the electrical and computer engineering with National Chiao Tung University, Hsinchu, Taiwan. Her current research interests include yield modeling for multiband RF-interconnected through-silicon-via, scheduling design for narrowband Internet of Things, and cross-layer analysis for ultra-reliable and ultra-low latency and enhanced mobile broadband multiplexing in 5G wireless communications.

Li-Chun Wang (Fellow, IEEE) received the B.S. degree in electrical engineering from National Chiao Tung University, Taiwan, in 1986, the M.S. degree in electrical engineering from National Taiwan University in 1988, and the Ms.Sci. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1995 and 1996, respectively.

From 1990 to 1992, he was with the Telecommunications Laboratories, Ministry of Transportations and Communications, Taiwan. In 1995, he was affiliated with the Bell Northern Research of Northern Telecom, Inc., Richardson, TX, USA. From 1996 to 2000, he was with AT&T Laboratories, where he was a Senior Technical Staff Member with the Wireless Communications Research Department. In August 2000, he joined National Chiao Tung University. He is currently a Chair Professor with the Department of Electrical and Computer Engineering and is jointly appointed by the Department of Computer Science. He has published over 90 journal papers and 180 conference papers, 19 U.S. patents, and co-edited a book *Key Technologies for 5G Wireless Systems* (Cambridge, 2016). His recent research interests are focused on cross-layer optimization and data-driven learning techniques for 5G ultra-reliable and ultra-low latency communications, edge computing, unmanned aerial vehicle communications networks, and AI-empowered mobile networks. He won the 1997 IEEE Jack Neubauer Best Paper Award in 1997, the Distinguished Research Award of Ministry of Science and Technology, Taiwan, in 2012 and 2018, and the IEEE Communications Society Asia–Pacific Board Best Paper Award in 2015. He was elected to the IEEE Fellow for his contributions in cellular architectures and radio resource management in wireless networks.

Chien-Hwa Hwang received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, USA. He is currently with MediaTek working on LTE and 5G system design and advanced receivers. His research interests include statistical signal processing, MIMO communication, and interference management of cellular systems.

Ju-Ya Chen received the B.S., M.S., and Ph.D. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan. He is currently with MediaTek working on 5G system design and product development. His research interests include statistical signal processing and communications theory.

Li-Yu Cheng received the M.S. degree from National Chiao Tung University, Taiwan, in 2018. His research interests are in the field of mobile network and 5G URLLC.