IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 10, 2019, accepted May 12, 2019, date of publication May 20, 2019, date of current version June 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917751

Machine Learning Based Flexible Transmission Time Interval Scheduling for eMBB and uRLLC Coexistence Scenario

JINGXUAN ZHANG, (Student Member, IEEE), XIAODONG XU, (Senior Member, IEEE), KANGJIE ZHANG, (Student Member, IEEE), BUFANG ZHANG, (Student Member, IEEE), XIAOFENG TAO, (Senior Member, IEEE), AND PING ZHANG, (Fellow, IEEE)

National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Xiaodong Xu (xuxiaodong@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871045, in part by the 111 Project of China under Grant B16006, and in part by the NOKIA.

ABSTRACT The enhanced Mobile Broadband (eMBB) and ultra-Reliable Low Latency Communications (uRLLC) are the two main scenarios of 5^{th} generation (5G) mobile communication system networks. There is an obvious difference in service requirements between different scenarios. When multi-scenario services coexist in the 5G networks, exploring optimized resource scheduling and allocation strategies become a critical issue. The 5G New Radio (NR) and numerology technologies have been standardized, which lay the foundation for flexible frame structure and adaptive scheduling. In this paper, we propose the self-adaptive flexible transmission time interval (TTI) scheduling (SAFE-TS) strategy in the eMBB and uRLLC coexistence scenario. Machine learning (ML) is applied to achieve flexible TTI scheduling. Moreover, we design the random forest-based ensemble TTI decision algorithm (RF-ETDA) to accomplish the TTI selection for each service. Compared with the existing ML methods, the proposed algorithm has a performance improvement in selecting TTI, especially for the uRLLC services. Then, the TTI selection results will be the basis of system resource scheduling and allocation. The simulation results prove that the proposed SAFE-TS effectively reduce the delay and packet loss rate of the uRLLC services while guaranteeing the eMBB requirements. Therefore, it is highly recommended that flexible TTI scheduling should be applied in the construction of the 5G networks to achieve superior network performances.

INDEX TERMS Flexible TTI scheduling, machine learning, delay, control overhead, eMBB, uRLLC.

I. INTRODUCTION

From the early voice services to current versatile communications services, the mobile communication system has undergone profound changes. Multi-scenario becomes a major feature of 5^{th} generation mobile communication system (5G). For the emerging 5G user cases, such as ultra-Reliable Low Latency Communications (uRLLC), enhanced Mobile Broadband (eMBB) and massive Machine Type Communication (mMTC) scenarios, 5G will confront with increasing requirements and challenges of various services [1]. Moreover, in many cases, different service scenarios coexist in the networks, especially the eMBB and the uRLLC.

The associate editor coordinating the review of this manuscript and approving it for publication was Ning Zhang.

The distinction of service requirements between different scenarios is obvious. The eMBB services, such as Virtual Reality (VR) and 4K video, have high requirements of bandwidth and data rate. The uRLLC services, such as industrial automation and automatic drive, have strict requirements for delay and reliability. When these scenarios coexist in the system, how to compromise the requirements of different scenarios becomes a critical issue. Exploring optimized resource scheduling and allocation strategies is the key to efficient 5G networks [2].

With the emergence of technologies such as multi-connectivity and numerology, the New Radio (NR) of 5G is moving towards flexibility and adaptability. A lot of works have proven the performance advantages of multi-connectivity [3], [4], which provide a possible solution to achieve the data rate requirements of eMBB services.

The multi-numerology technology standardized by the 3rd Generation Partnership Project (3GPP) lays the foundation for the flexible frame structure [5]. In the previous studies of resource allocation issues, the integrated two-dimensional resource allocations of time and power are not rarely seen. However, with the emergence of new scenarios in 5G networks, the existing granularity of resource allocation cannot meet the growing requirements of 5G services. A promising solution is to explore the resource scheduling with the flexibility of time and frequency domains.

Fortunately, these issues might be solved in the field of Machine Learning (ML), which provides brand prospects and possibilities compared with traditional methods. Problems difficult to model and to solve are the potential application directions for ML in 5G networks [6].

Therefore, our work considers eMBB and uRLLC coexisting scenario in 5G networks, and explores innovative joint scheduling strategies based on ML. The contributions of this paper are listed as follows:

(1) We propose a flexible TTI scheduling strategy to satisfy the service requirements in eMBB and uRLLC coexistence scenario. The proposed Self-adaptive Flexible TTI scheduling (SAFE-TS) strategy has obvious performance advantages compared with traditional fixed-length TTI scheduling.

(2) Based on ML, we design the Random Forest based Ensemble TTI Decision Algorithm (RF-ETDA) to implement the flexible TTI scheduling. The length of the TTI is selected according to the four features of the Base Station (BS) and channel conditions when different services arrive.

(3) The flexible TTI scheduling has obvious advantages for uRLLC services. Compared with fixed-length TTI scheduling, SAFE-TS improves the delay performance of uRLLC services by 45.64% on average, and improves the PLR performance by 59.17% on average while ensuring the data rate requirements of eMBB. The eMBB services do not have typical classification characteristics, and most of them select larger TTI.

The remaining of this paper is arranged as follows. In Section II, the related works about flexible TTI scheduling and ML are introduced. The system model and the definition of delay are given in Section III. The strategies of SAFE-TS and RF-ETDA are proposed in Section IV. In Section V, the performance evaluation and simulation results are provided. Finally, there are concluding remarks.

II. RELATED WORKS

From the Long-Term Evolution (LTE) to 5G, the flexible frame structure has always been appreciated [7] [8]. Flexible TTI scheduling is necessary to meet 5G new scenarios, especially for uRLLC services. Many works have proved the benefits of flexible frame structure and dynamic scheduling for critical services. The authors of [9] explored the impact of different situations on TTI selection. They proposed that support for scheduling with different TTI sizes is important for uRLLC services. To describe the influence of the control overhead of shorten TTI, a model based on the flow arrival rate and TTI length is designed in [10]. The authors of [11] investigated the optimizing resource allocation with flexible numerology in both frequency and time domains, and illuminated advantages for capacity enhancement and satisfaction of uRLLC service requirements. However, in the existing work, the factors affecting the TTI selection and the specific selection scheme are not given. The corresponding flexible TTI scheduling algorithm is rarely proposed.

In addition, ML has been widely applied in many aspects of wireless communications in recent years. The authors in [12] focused on the self-organizing networks, and provided future research directions and solutions using intelligent algorithms. With the constraints of the data rate and transmission power, a K-means algorithm to maximize the rate in millimeter wave non-orthogonal multiple access systems is developed in [13]. The authors of [14] proposed a novel ML architecture for short packet communications, in which the received signals are clustered by unsupervised learning, and cluster-symbol mapping is accomplished by known labels. To the best of our knowledge, there are no studies focus on applying ML methods to flexible TTI scheduling.

III. SYSTEM MODEL

The research scenario is the coexistence of multi-user and multi-service in a multi-tier heterogeneous 5G networks, where two types of services, eMBB and uRLLC are considered. As shown in Fig.1, there are several macro BSs and pico BSs. Some services are shown as examples in Fig.1, such as VR, industrial automation, etc. In order to meet the strict requirements of uRLLC, higher priority is usually assigned to uRLLC services [15].

The multi-connectivity technology is also applied in this work. The services can be connected to different BSs for transmission simultaneously, especially for eMBB services.

A. DELAY MODEL

The delay D for a user scheduled in the downlink can be expressed as:

$$D = d_Q + d_{bsp} + d_{FA} + d_{Tx} + d_{mtp} \quad (1)$$

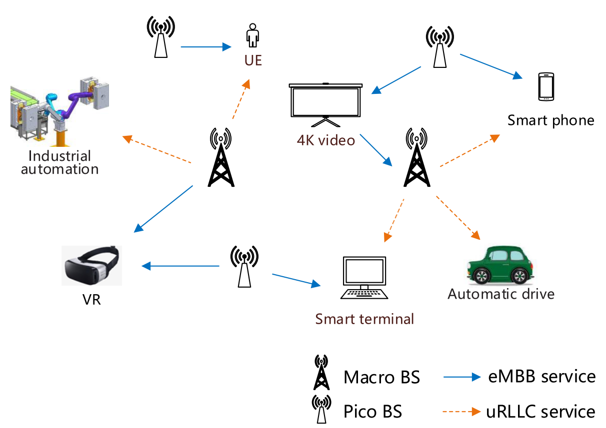

FIGURE 1. Downlink network deployments with eMBB and uRLLC coexisting scenarios.

where d_Q denotes the queuing delay, which is the time that the packet waits in the buffer queue to be transmitted. d_Q will increase with traffic load conditions. When the packet arrives and the resources are ready to be scheduled, the system must wait to the beginning of the next TTI to transmit the packet. This part of the delay is defined as the frame alignment delay, denoted by d_{FA}. d_{FA} is distributed between 0 and the TTI size [16]. d_{Tx} represents the transmission delay, which is the number of TTIs spent to transmit the packet. It depends on the channel conditions, packet sizes and available resources, etc. The d_{bsp} and d_{mtp} represent the processing delay of the BS and mobile terminal respectively, which depends on the processing capability of the device itself. In the LTE system, the downlink processing delay is about few milliseconds [17]. With the development of technologies, these two parts of the delay will be much smaller in 5G [1], which will be ignored in this work.

B. 5G NR AND NUMEROLOGY

3GPP standardizes the 5G NR technology. During the study of NR technology, many proposals on physical and higher layer protocols, radio frequency related issues were investigated [5]. Feasibility and capability of these new technologies have been proven. The NR access technology has a flexible Orthogonal Frequency Division Multiplexing (OFDM) framework, flexible and wide range of bandwidths, multiple deployment options, greater spectrum utilization, and multiple numerologies within one carrier [18].

To maintain the backward compatibility with LTE, the length of one frame is 10 ms, including 10 subframes. The number of slots in each subframe is decided by the numerology configuration μ, where a slot is composed of 14 OFDM symbols. Therefore, the OFDM symbol length (without Cyclic Prefix) is $1\ (= 14 \times 2\mu)$ ms. Different NR numerology parameters determine different sizes of the Resource Block (RB). Multiple OFDM numerologies are supported by 3GPP as Table 1, where μ and the cyclic prefix for a bandwidth part are given by the higher-layer parameters [5].

In the LTE and previous mobile networks infrastructure design, the TTI is almost a fixed value as 1ms, which means $\mu = 0$. With the increase of μ, the bandwidth becomes larger and the TTI becomes shorter. TTI is equal to the length of a slot in numerical value. Without loss of generality, we choose

TABLE 1. Numerologies in 5G NR.

Different size of μ	0	1	2	3	4
Subframes in one frame	10	10	10	10	10
Slots in one subframe	1	2	4	8	16
The length of one slot (ms)	1	0.5	0.25	0.125	0.0625
OFDM symbols in one slot	14	14	14	14	14
Subcarrier spacing (kHz)	15	30	60	120	240
RB size in frequency domain (MHz)	0.18	0.36	0.72	1.44	2.88
RB size in time domain (ms)	1	0.5	0.25	0.125	0.0625

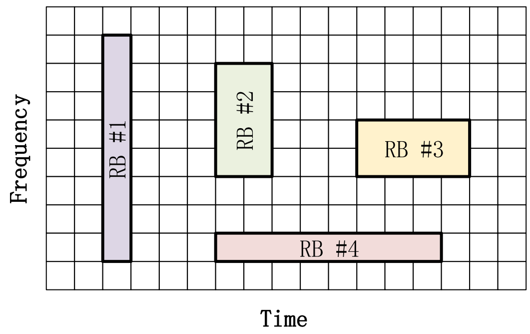

FIGURE 2. Different RB structure with different TTI.

the common value of μ, which equals to 0,1,2,3, and it is typical to analyze the performances [9], [11]. The RB is defined as the minimum granularity of resource allocation, consisting of one slot in the time domain and 12 subcarriers in the frequency domain. Because of the different sizes of slots and subcarriers shown in Table 1, the structure of RB varies with the TTI. The parameters of TTI length and the frequency domain size are closely related and jointly determine the structure of the RB. We use the parameters of TTI to represent these differences, and the changes and impacts of frequency domain are also fully considered.

In Fig.2, the smallest grid in the time axis represents 0.125ms, and the frequency axis represents 180kHz per grid. Each grid can only be assigned to one user. Fig.2 shows the comparison of RB in four cases. RB #1 - RB #4 represent RB with 0.125ms, 0.25ms, 0.5ms, 1ms TTI respectively, and different RB structures are distinguished by different colors.

Fig.3 is a simple illustration of resource allocation with flexible TTI selection. We use the same color as Fig.2 to denote the corresponding RB structure and TTI size. The entire area in Fig.3 represents all available resources. When a service arrives, the TTI is selected according to some characteristics of the channel and BSs at that time. Services with different TTIs share resources such as frequency bands, power in the channel. Just like tiling, different services use existing resources as efficiently as possible. At each opportunity of

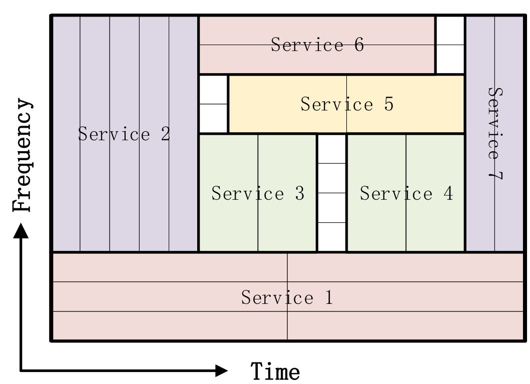

FIGURE 3. An illustration of resource allocation with flexible TTI selection.

scheduling, each service can be assigned certain number of tiles, which providing a high degree of flexibility in bandwidth and TTI length. Due to the size of services and TTIs, there may be a small portion of remaining resources. Corresponding scheduling algorithms applying flexible TTI are considered to reduce resource surplus and optimize resource utilization.

IV. SELF-ADAPTIVE FLEXIBLE TTI SCHEDULING

The flexible frame structure and variable TTI sizes provide the possibility for flexible scheduling. In this paper, we propose the SAFE-TS method to satisfy the requirements of different services and improve the system performances. In the scheduling process, the choice of TTI is the most important issue. The system is flexibly scheduled based on the TTI of each service. The choosing of TTI length involves many influence factors, and it is difficult to be quantified as a formula expression. It is not suitable to apply conventional optimization methods to solve this problem. Hence, we apply the ML methods based on large dataset to research the flexible TTI scheduling instead of complex theoretical methods.

A. CLASSIFICATION FEATURES FOR ML METHODS

There are many factors that influence the choice of TTI, and none of them have a simple threshold to make judgment and decision. We consider the obvious features and use ML to solve this problem.

1) FEATURE 1: RATIO OF EMBB AND URLLC SERVICES

The ratio is defined as the current proportion of RBs occupied by the two types of services in the selected BS when the service arrives. As aforementioned, because of the higher priority of uRLLC services, the ratio of the two types of services should also be considered. When the traffic load conditions of the BSs are similar, especially when there are no available resources, the uRLLC services can occupy the eMBB resources to satisfy strict requirements. For uRLLC services, this ratio reflects the busyness degree of BSs. Hence, it is feasible to use the proportion of the two types of services as a feature.

2) FEATURE 2: TRAFFIC LOAD

When a service arrives at the BS, the size of different TTIs affects the delay of the service. When the traffic load is light, it is better to choose the shorter TTI. However, as the traffic load increases, the delay will increase with the TTI. The main cause of this phenomenon is the impact of queuing delay. With the increase of traffic load, the advantages of small TTI due to transmission delay and frame alignment delay are already less than the impact of queuing delay [10]. The influence of queuing delay is ascending, and it becomes the most dominant factor in the five parts of the delay component as shown in Formula (1). Larger TTI reduces the probability of queuing by increasing spectral efficiency. Therefore, it is feasible to take the traffic load as a feature that affects the TTI selection.

3) FEATURE 3: CONTROL OVERHEAD

Short TTI reduces over-the-air transmission delay at the expense of increased control signaling overhead, and reduces the spectral efficiency. We use a flexible frame structure of 5G NR, and the Control Channel (CCH) is included in the resources allocated to the service. The CCH contains scheduling grants and related information for decoding, etc. Control Channel Element (CCE) is the resource granularity of CCH, which consists of a certain number of 36 Resource Elements (REs). According to 3GPP standards, Table 2 shows the relationship between control overhead and Signal to Interference plus Noise Ratio (SINR) [19]. The value of SINR is calculated based on relevant parameters in the simulation. When the channel quality becomes poor, it is obvious that the control signaling overhead will be larger. On each scheduling opportunity, corresponding amount of resources must be occupied to transmit control signaling.

TABLE 2. Control overhead.

SINR (dB)	Control overhead (REs)
$(-\infty, -2.2)$	$8 \times 36 = 288$
$[-2.2, 0.2)$	$4 \times 36 = 144$
$[0.2, 4.2)$	$2 \times 36 = 72$
$[4.2, +\infty)$	$1 \times 36 = 36$

4) FEATURE 4: PACKET MAGNITUDE

The impact of different packet sizes on delay performance is taken into account in this work. With the increase of data packets, the delay of that service is also increasing due to the impact of transmission time [20]. In addition, due to the increase of packet magnitude, the impact of control overhead on delay is negligible. This also indirectly affects the delay and the choice of TTI. On this basis, in order to fully explore the influence factors, the packet size of the same service type is also diverse.

B. DATA GENERATION AND COLLECTION

Due to the novelty of the scenario and the complexity of the problem, there is no suitable public dataset available for us to research. Lacking of appropriate and adequate dataset is a bottleneck of applicating ML in communication networks. A promising solution is to obtain the dataset through simulation [21]–[23]. We use system-level simulation to generate and collect data. Each set of data contains the four eigenvalues mentioned above, along with the corresponding TTI size. We collect data for the eMBB and uRLLC services separately. The data is simulated over a long period of time on a simulation platform close to the real situation, and meets the relevant requirements in the 3GPP standards [24], [25]. These datasets can also be accessible to other peers and provide reference for related researches. After the SAFE-TS method applied, the actual datasets can be easily collected at the BS side, which will have stronger accuracy and persuasiveness than the simulation datasets.

The system model is described in section III, and specific simulation parameters are given in Table 3. We consider a

TABLE 3. Simulation parameters.

Parameter	Value
BS deployment	12 pico BSs with each macro BS
Number of UE	180 in each macro BS cover range
UE distribution	Evenly distributed
Bandwidth of BS	Macro BS: 20MB
Bandwidth of BS	Pico BS: 100MB
Power of macro BS	46dBm
Power of pico BS	30dBm
Power of noise	-95dBm
Pathloss of macro BS	$128.1 + 37.6 \log d(km)$ dB
Pathloss of pico BS	$140.7 + 36.7 \log d(km)$ dB
Traffic model	Poisson arrival process
Basic scheduling principle	FCFS with priority for uRLLC services
Service arrival interval (ms)	Variable: 10; 30; 50
TTI size (ms)	Variable: 0.125; 0.25; 0.5; 1
Packet of eMBB (MB)	Variable: 1; 2; 3; 4; 5
Packet of uRLLC (KB)	Variable: 6.4; 12.8; 19.2; 25.6; 32

system including three sectors in each macro BS. There are one pico cluster containing four pico BSs in each sector. Two thirds of the UE are distributed in the pico BS area evenly. The other UEs are in the macro BS coverage areas. The service arrival processes follow the Poisson distribution [10], and the SINR is calculated during the simulation process. Among the common scheduling algorithms, we choose the First Come First Served (FCFS) as the basic scheduling principle [9]. The specific parameters are based on 3GPP TR 36.814 [24] and TR 36.872 [25].

The arrival of eMBB and uRLLC services follow the Poisson process, and the different conditions of traffic load can be achieved by adjusting the service arrival interval. After selecting the BS which the service is served, the four eigenvalues can be determined. The ratio of the two services can be statistically obtained. The traffic load can be measured by the ratio of available resources. The control overhead is determined by the SINR, which is calculated by simulation. The packet size is randomly selected among the five sizes. For eMBB services, features of each available BSs are considered because of the multi-connectivity.

During the simulation, different scheduling conditions based on each size of TTI are simulated and compared separately when the four features are determined. The one with the best performance is selected as the choice of the service. That is, for the uRLLC services, the TTI length which achieves the lowest service delay is implemented, and for the eMBB services, the target is data rate. Each loop is randomly focused on one service, the feature values and TTI are recorded as a set of data, and two types of services are recorded separately. Multiple sets of data constitute the dataset.

C. RF-ETDA BASED TTI SELECTION

According to above analysis, the issue is a multi-class classification problem. The features are not independent absolutely with each other and contain both continuous and discrete variables. Therefore, we design the RF-ETDA and choose the following three methods for comparison and analysis.

1) RANDOM FOREST BASED ENSEMBLE TTI DECISION ALGORITHM

In order to achieve higher classification accuracy, we design the RF-ETDA based on the k-Nearest Neighbor (kNN) and the Random Forest (RF) methods. Ensemble learning accomplishes learning tasks by building and combining multiple learners. Several weak classifiers are combined to obtain a strong classifier with superior classification performance. RF itself is an ensemble learning method that integrates multiple decision trees. RF can process high-dimensional data, has strong anti-interference ability, and is suitable for solving classification problems.

We train RF-ETDA models for eMBB and uRLLC services respectively. The eMBB services mostly choose larger TTI, on the contrary, uRLLC services prefer shorter TTI. For uRLLC services, 0.25ms, 0.5ms and 1ms are small categories, and for eMBB services, the categories except 1ms are small categories. The uneven distribution of dataset categories affects the accuracy of algorithm classification. Therefore, we optimize the RF algorithm and combine the kNN method to implement classification.

The specific algorithm procedure is shown in Algorithm 1. When the number of samples in each category is similar, we define that the dataset is balanced. The RF uses bootstrap to extract multiple samples from the original dataset to form multiple new sub-datasets. Each sub-dataset selects certain appropriate features, and uses the weak classifier decision tree to train the samples. Multiple classifiers vote to determine the output. The RF-ETDA algorithm is more suitable for network requirements and has better classification performance.

Algorithm 1 RF-ETDA Procedure

Input: Training set: $D = (\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \ldots, (\boldsymbol{x_m}, y_m)$
Number of trees: T

1: **for** each sample $(\boldsymbol{x_i}, y_i)$ belongs to small category
2: Select a sample $(\hat{\boldsymbol{x}}_i, \hat{y}_i)$ from the k-nearest neighbors of $(\boldsymbol{x_i}, y_i)$ in the category randomly
3: Generate a new sample $(\boldsymbol{x_{new}}, y_{new})$:
$\boldsymbol{x_{new}} = \boldsymbol{x_i} + (\hat{\boldsymbol{x}}_i - \boldsymbol{x_i}) \times \delta_i$, $y_{new} = y_i$
random number $\delta_i \in [0, 1]$.
4: Add the new sample to the original dataset.
5: **until** the dataset is balanced.
6: **for** $t = 1, 2, \ldots, T$
7: Take samples randomly from D, and form new sub-datasets D_t containing k samples.
8: Select certain eigenvalues, and train the decision tree model G_t based on D_t.
9: **end for**
10: Vote for the T weak classifiers.

Output: The final result is the category with the most votes.

2) COMPARISON METHODS

a: NEURAL NETWORK (NN)

The NN classification uses a feed-forward neural network, which consists of an input layer, an output layer, and several hidden layers. The input features are weighted and passed to the hidden layer by the input layer neurons, and are sorted by the output layer through a series of calculations. Backpropagation algorithm is applied to determine the weight between neurons. After several times attempts and adjustments to the parameters, a compromise between complexity and accuracy is achieved.

b: SUPPORT VECTOR MACHINES (SVM)

The SVM classifier was originally only applied to the binary classification. In order to meet the requirements of our problem, multiple SVM classifiers are hierarchically constructed with multi-class classifiers. The classification principle of SVM is to find the best boundary, which makes the classification result more accurate. Multi-dimensional transformation of data is achieved by nonlinear mapping to make the optimal boundary a hyperplane [26]. SVM classifier has a strong ability of generalization, but the disadvantage is that when the dataset has a large scale, the time cost is relatively high.

c: RANDOM FOREST (RF)

In order to better compare the performance of the proposed RF-ETDA, we also chose the Random Forest for comparison.

D. SCHEDULING WITH SAFE-TS

After applying the ML method for TTI selection, we propose the downlink SAFE-TS algorithm. For each RB, we judge whether the service is completed in each different scheduling period. The loop is accomplished every 0.125ms, but related processes are performed only when the RB cycle is completed, such as updating the queue and releasing resources.

When the new service arrives, the network first selects the BSs to access according to the maximum Reference Signal Receiving Power (RSRP) criterion. For the eMBB services, we take multi-connectivity into account, that is, several BSs with larger RSRP are taken as alternatives. After judging the service type, the size of TTI are decided by RF-ETDA according to the channel and BS conditions. If the service type is uRLLC, when the BS available resources cannot meet the service requirements, it can occupy the resources of eMBB services in the same BS. It is important to record the TTI size corresponding to each RB, which is the basis for flexible TTI scheduling. We choose FCFS as the basic scheduling principle, and the uRLLC services have higher priority. The specific scheduling algorithm is shown in Algorithm 2.

Algorithm 2 Downlink SAFE-TS Procedure

```
for each 0.125ms
    for each RB
        if the period of RB is accomplished
            Determine whether the service is
            completed, record data, update queues of
            RB, and release resources
        end if
    end for
end for
for each service arrival
    Select the BSs according to the max-RSRP criterion
    Select TTI using ML based on channel and BS
    conditions
    if it is uRLLC service
        if the idle resources of the BS can satisfy the
        requirements of the service
            Occupy the idle resources of the BS
        else if the resources of eMBB in the BS can
        satisfy the requirements of the service
            Occupy the resources of eMBB in the
            same BS
        else the service enters the queue
        end if
    else it is eMBB service
        if the idle resources of alternative BSs can
        satisfy the requirements of the service
            Occupy the resources of one or
            more alternative BSs
        else the service enters the queue
        end if
    end if
    Record the TTI corresponding to each RB
end for
```

TABLE 4. Accuracy of different ML methods.

ML Methods	Accuracy of uRLLC	Accuracy of eMBB
NN	86.97%	98.88%
SVM	84.85%	98.81%
RF	87.72%	98.95%
RF-ETDA	95.93%	98.93%

V. PERFORMANCE EVALUATION

A. COMPARISON OF ML METHODS

We performed simulation analyses on the ML methods mentioned above. The datasets of the eMBB and uRLLC services are composed of 10000 sets of data respectively. We use supervised learning, and randomly select 80% of the data in the dataset as the training set and the others as the test set. To reduce the impact of randomness, we repeat this process 20 times and take the average of the results. We train the learning model with the training set and evaluate with the test set. When the characteristic values of the test data is input, and the output classification result is consistent with the known result, the classification is correct. The classification accuracy is defined as the ratio of the number of correct classifications to the total number of test set. After adjusting some parameters, the classification accuracy is shown in Table 4.

As shown in Table 4, for the uRLLC services, the proposed RF-ETDA method has the highest classification accuracy, and it improves performance compared to other methods. The SVM method has lowest classification accuracy.

For the eMBB services, the classification performances of four methods are very close. By observing the dataset, we find that only a very small part of the eMBB services do not select the TTI of 1 ms, and the proportion of this part is about 1%. This is also the reason why the classification accuracy of four methods is about 99%. The dataset shows that 1ms TTI is more suitable for eMBB services. The classification accuracy of our proposed RF-ETDA does not improved obviously. For uRLLC services with distinct classification features, a balanced dataset can improve the performance of classifier. Since the eMBB services do not have typical classification features, this process does not lead to an increase in classifier performance.

We also analyze the reasons for choosing the fixed-length TTI of 1ms for eMBB services. Due to the eMBB services aim at high data rate, while the advantages of short TTI are mainly reflected in the delay. The frequent scheduling caused by short TTI does not significantly contribute to the improvement of data rate. Different with uRLLC, eMBB has larger packet sizes, and transmission time of service is longer. Larger packet size means that eMBB services are not sensitive to the impact of control overhead. Besides that, due to the multi-connectivity, they have alternative BSs to choose from, the influence of traffic load is relatively small. Since the eigenvalue has little effect on the data rate, the eMBB services do not have typical classification features. This conclusion will also be further proved in the simulation of Section V-B.

Therefore, in the following work, we apply the proposed RF-ETDA for uRLLC services to select TTI size, while eMBB services adopt fixed-length TTI of 1ms.

B. SCHEDULING WITH SAFE-TS

After comparing the performance of ML, we simulate the proposed SAFE-TS to demonstrate the advantages of proposed strategy. The settings of the simulation environment are consistent with the dataset generation process.

We first analyze the relevant performances of uRLLC services. When the uRLLC service arrives at the BS, the TTI will be determined based on the model trained by the RF-ETDA. Different from the scheduling in LTE, since each service has different TTIs and the basic granularity of resource allocation is RB, there may be different scheduling slots in one BS. We use different slots for each RB to perform queue processing and resource scheduling respectively. Different conditions of traffic load have an impact on the choice of TTI. Figures 4–6 is a comparison of the uRLLC services delay in different traffic load conditions. We achieve different load conditions by modulating the service arrival interval, and take a portion of the picture in the range of 0–2ms.

From any of the figures we can conclude that as the TTI size increases, the delay tends to increase substantially, which demonstrates the benefits of small TTI for uRLLC services. In most cases, 0.125ms has better delay performance. However, there are some cases that the delay of 0.125ms is significantly increased. This change is especially noticeable under heavy traffic load condition. As the traffic load increases, the ratio of 0.125ms fixed-length TTI scheduling delay exceeds 1ms increases, and the proportion varies from 3% to 8% and then to 12%. In the case of large traffic load, certain proportion delay of 0.125ms even exceeds 2ms, which can be concluded that a small TTI is an unfavorable choice in that case. Our proposed SAFE-TS selects the suitable TTI under different circumstances.

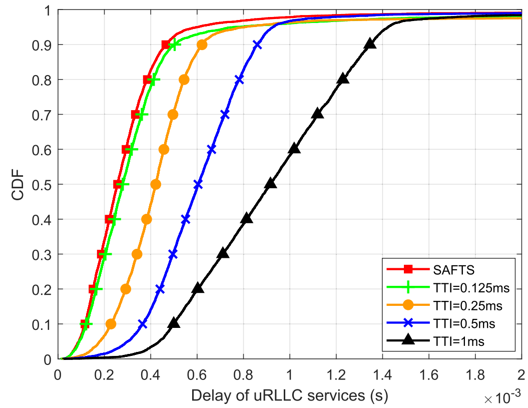

FIGURE 4. The delay of uRLLC services in light traffic load condition.

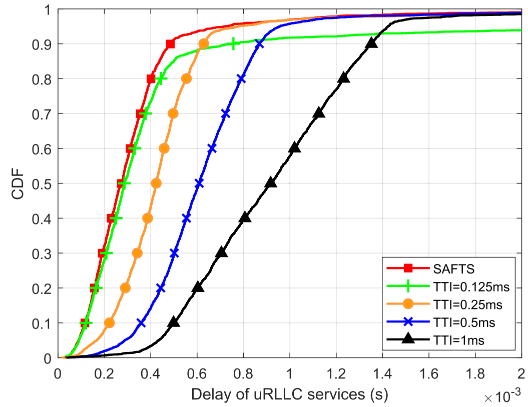

FIGURE 5. The delay of uRLLC services in medium traffic load condition.

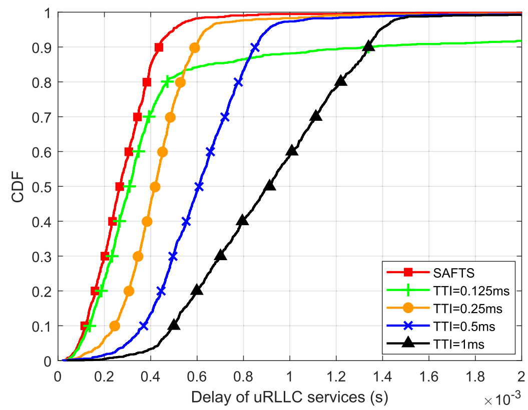

FIGURE 6. The delay of uRLLC services in heavy traffic load condition.

TABLE 5. Delay performance gain of SAFE-TS compared with fixed-length TTI scheduling methods.

Gain / Scheduling / Traffic load	TTI= 0.125ms	TTI= 0.25ms	TTI= 0.5ms	TTI= 1ms
Light traffic load	8.29%	36.37%	55.87%	70.73%
Medium traffic load	20.31%	33.09%	53.83%	69.13%
Heavy traffic load	40.07%	34.70%	55.25%	70.07%

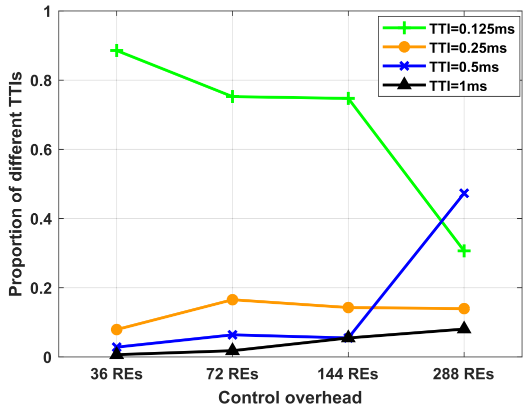

FIGURE 7. Control overhead impact on TTI selection.

Comparing the three graphs, we can notice that the SAFE-TS has the best delay performance in any case. The delay performance gain of SAFE-TS compared with fixed-length TTI scheduling methods are shown in Table 5. From Table 5, we can calculate that the delay performance of SAFE-TS is increased by 45.64% on average.

Fig.7 shows the impact of control overhead on TTI selection. Through the statistics of the dataset, the figure compares the different selection ratios of the four TTIs in proposed SAFE-TS when the RE size is different. We performed this set of simulations under medium traffic load conditions. The ratio of 0.125ms decreases from 90% at 36 REs to 30% at 288 REs. Meanwhile, the proportion of larger TTIs increases, especially in 0.5ms. When the control overhead is 288 REs, the advantage of 0.5ms is obvious.

As the control overhead increases due to different SINR conditions, the proportion of larger TTIs is increases. This is because on each scheduling opportunity, corresponding resources should be allocated to transmit control signaling. Frequent scheduling caused by short TTI leads to large control overhead and low spectral efficiency. However, the increase trend of 1ms TTI is not obvious, which is because the control overhead advantage brought by TTI size is not enough to balance the influence of frame alignment delay and transmission delay.

Table 6 compares the probability of packet loss occurring under different traffic load conditions and different scheduling methods. The uRLLC services have strict requirements of delay and reliability. We define that packet loss occurs when the delay of uRLLC services exceed 1ms. The PLR is an important indicator to measure the reliability of uRLLC services. When the traffic load situation is determined,

TABLE 6. The PLR of uRLLC services.

PLR / Scheduling / Traffic load	SAFE-TS	TTI= 0.125ms	TTI= 0.25ms	TTI= 0.5ms	TTI= 1ms
Light traffic load	1.39%	1.61%	2.53%	3.24%	46.11%
Medium traffic load	1.63%	4.00%	3.60%	2.54%	39.11%
Heavy traffic load	1.45%	9.20%	3.44%	1.67%	39.16%

TABLE 7. PLR performance gain of SAFE-TS compared with fixed-length TTI scheduling methods.

Gain / Scheduling / Traffic load	TTI= 0.125ms	TTI= 0.25ms	TTI= 0.5ms	TTI= 1ms
Light traffic load	13.67%	45.06%	57.10%	96.99%
Medium traffic load	59.25%	54.72%	35.83%	95.83%
Heavy traffic load	84.24%	57.85%	13.17%	96.30%

the SAFE-TS has the lowest PLR. As the traffic load increases, the PLR of short TTI increases, while that of large TTI decreases. The performance advantages of large TTI begin to manifest. When the TTI is 1ms, the value of PLR is about 40%, because of the frame alignment delay. In the simulation, frame alignment delay is a random value between 0 and TTI size. The impact of this part leads to a high probability of the total delay exceeds 1ms. Table 6 demonstrates the impact of traffic load on TTI selection again and testifies the flexibility of our proposed scheduling method.

The PLR performance gain of SAFE-TS compared with fixed-length TTI scheduling methods are shown in Table 7. In all cases, SAFE-TS has apparent PLR performance advantage. Compared with 0.125ms fixed-length TTI scheduling, the gain of flexible scheduling increases as the traffic load increases. The variation trend is reversed in 0.5ms scheduling. In the case of light traffic load, most of the services in the flexible scheduling select a shorter TTI, and the SAFE-TS has less performance gain than the shorter fixed-length TTI scheduling. As the load increases, the advantage of larger TTI becomes more obvious. This is consistent with the aforementioned analysis. The SAFE-TS has obvious gain than 1ms scheduling because the 1ms TTI brings large delay to the uRLLC services, which leads to poor PLR performance. By calculating the data in Table 7, the PLR performance of proposed SAFE-TS is increased by 59.17% on average.

As shown in Table 8, we compare the TTI selection ratio of our proposed SAFE-TS method under different traffic load conditions. The 0.125ms is the choice of most services under light traffic load conditions. The reasons why other services do not select short TTI are the influence of control overhead and packet size. With the traffic load increases, the proportion of short TTIs gradually decreases, and the advantages of large TTIs become apparent. However, under different traffic load conditions, short TTI is always favored. This is consistent with the previous analysis.

In the SAFE-TS, the eMBB services use a fixed length TTI as mentioned before. As shown in Fig.8, larger TTI scheduling has larger data rates. In order to make multiple services better coordinate resources, we limit the data rate in the range

TABLE 8. TTI selection ratios of uRLLC services in SAFE-TS.

TTI ratios / TTI length / Traffic load	0.125ms	0.25ms	0.5ms	1ms
Light traffic load	86.93%	9.56%	2.74%	0.77%
Medium traffic load	77.89%	14.46%	4.59%	3.06%
Heavy traffic load	71.38%	13.98%	7.65%	6.99%

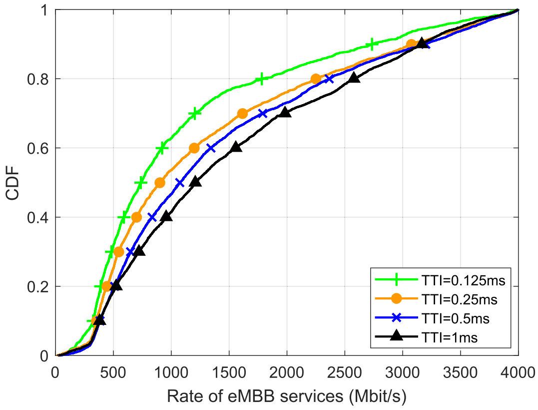

FIGURE 8. Data rates of eMBB services.

of $0-4000$ Mbit/s, which makes it possible to support eMBB services such as VR. Since eMBB services have large packet sizes, the transmission time is relatively long. Because it is not sensitive to delay, frequent scheduling will not bring performance advantages, but will lead to greater overhead. According to the conclusion of Fig.8, it is not necessary for eMBB services to adopt flexible TTI scheduling, but to adopt the 1ms fixed-length TTI.

VI. CONCLUSION

In this paper, we propose the SAFE-TS strategy in eMBB and uRLLC coexistence scenario to meet the service requirements. Appropriate eigenvalues are chosen and the RF-ETDA is designed to implement the TTI selection for each service. RF-EDTA has better performance than existing ML algorithms. The resource scheduling and allocation are performed according to the different TTIs of services. The simulation results show that the proposed SAFE-TS has obvious advantages compared with the fixed-length TTI. The delay performance of the uRLLC services improves by 45.64% on average, and the PLR performance improves by 59.17% on average. The eMBB services do not have typical classification characteristics, and prefer the largest TTI. Since there are no suitable datasets available for reference, we simulated the relevant scene to collect data. Our datasets will be accessible to other peers and provide reference for related researches. In future applications, the dataset will be easily collected on the BS side, and the ML method will become more convincing.

REFERENCES

[1] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[2] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5G radio access networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1656–1686, 3rd Quart., 2016.

[3] X. Ba and Y. Wang, "Load-aware cell select scheme for multi-connectivity in intra-frequency 5G ultra dense network," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 354–357, 2019.

[4] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. S. Filho, and G. Fettweis, "How reliable and capable is multi-connectivity?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1506–1520, Feb. 2019.

[5] *Technical Specification Group Radio Access Network; NR; Physical Channels and Modulation*, document 38.211, Version 15.1.0, 3GPP, Mar. 2018.

[6] X. You, C. Zhang, X. Tan, S. Jin, and H. Wu, "AI for 5G: Research directions and paradigms," *Sci. China Inf. Sci.*, vol. 62, no. 2, p. 21301, Feb. 2019.

[7] P. Thakur, A. Kumar, S. Pandit, G. Singh, and S. N. Satashia, "Advanced frame structures for hybrid spectrum access strategy in cognitive radio communication systems," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 410–413, Feb. 2017.

[8] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, "Frame structure design and analysis for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1508–1522, Mar. 2017.

[9] G. Pocovi, K. I. Pedersen, B. Soret, M. Lauridsen, and P. Mogensen, "On the impact of multi-user traffic dynamics on low latency communications," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2016, pp. 204–208.

[10] Z. Zhang, Y. Gao, Y. Liu, and Z. Li, "Performance evaluation of shortened transmission time interval in LTE networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–5.

[11] L. You, Q. Liao, N. Pappas, and D. Yuan, "Resource optimization with flexible numerology and frame structure for heterogeneous services," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2579–2582, Dec. 2018.

[12] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.

[13] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, "Unsupervised machine learning-based user clustering in millimeter-wave-NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7425–7440, Nov. 2018.

[14] Q. Zhang, P. P. Liang, Y.-D. Huang, Y. Pei, and Y.-C. Liang, "Label-assisted transmission for short packet communications: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8846–8859, Sep. 2018.

[15] 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.

[16] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint link adaptation and scheduling for 5G ultra-reliable low-latency communications," *IEEE Access*, vol. 6, pp. 28912–28922, 2018.

[17] S. Ahmadi, *LTE-Advanced: A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies*. New York, NY, USA: Academic, 2013.

[18] A. Yazar and H. Arslan, "A flexibility metric and optimization methods for mixed numerologies in 5G and beyond," *IEEE Access*, vol. 6, pp. 3755–3764, 2018.

[19] D. Laselva, F. Capozzi, F. Frederiksen, K. I. Pedersen, J. Wigard, and I. Z. Kovács, "On the impact of realistic control channel constraints on QoS provisioning in UTRAN LTE," in *Proc. IEEE 70th Veh. Technol. Conf. Fall*, Sep. 2009, pp. 1–5.

[20] L. Wang *et al.*, "Study on flexible TTI scheduling for LAA systems," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2018, pp. 1–5.

[21] E. Kurniawan, P. H. Tan, S. Sun, and Y. Wang, "Machine learning-based channel-type identification for IEEE 802.11 ac link adaptation," in *Proc. 24th Asia–Pacific Conf. Commun. (APCC)*, Nov. 2018, pp. 51–56.

[22] C. Antón-Haro and X. Mestre, "Learning and data-driven beam selection for mmWave communications: An angle of arrival-based approach," *IEEE Access*, vol. 7, pp. 20404–20415, 2019.

[23] Y. Liu, C. He, X. Li, C. Zhang, and C. Tian, "Power allocation schemes based on machine learning for distributed antenna systems," *IEEE Access*, vol. 7, pp. 20577–20584, 2019.

[24] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, document 36.814, 3GPP, Version 9.0.0, Mar. 2010.

[25] *Small Cell Enhancements for E-UTRA and E-UTRAN—Physical Layer Aspects*, document 36.872, 3GPP, Version 12.1.0, Dec. 2013.
[26] J. Pérez-Romero, O. Sallent, R. Ferrús, and R. Agustí, "Knowledge-based 5g radio access network planning and optimization," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Sep. 2016, pp. 359–365.

JINGXUAN ZHANG (S'18) received the B.S. degree in communication engineering from the Beijing Information Science and Technology University, Beijing, China, in 2017. She is currently pursuing the Ph.D. degree in information and communication engineering with the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. Her research interests include mobile communication, machine learning, and resource allocation in 5G networks.

XIAODONG XU (S'06–M'07–SM'18) received the B.S. degree in information and communication engineering and the master's degree in communication and information system from Shandong University, in 2001 and 2004, respectively, and the Ph.D. degree in circuit and system from the Beijing University of Posts and Telecommunications (BUPT), in 2007, where he is currently a Professor. He has coauthored nine books and more than 120 journal and conference papers. He is also the inventor or co-inventor of 39 granted patents. His research interests include moving networks, D2D communications, mobile edge computing, and caching. He is an Associate Editor of IEEE Access.

KANGJIE ZHANG (S'18) received the B.S. degree in physics from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include mobile communication, distributed computing, and resource allocation in 5G networks.

BUFANG ZHANG (S'19) received the B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2017, where she is currently pursuing the master's degree in information and communication engineering. Her research interests include mobile communication, multi-connectivity, and resource allocation in 5G networks.

XIAOFENG TAO (S'99–A'02–M'03–SM'13) received the B.S. degree in electrical engineering from Xian Jiaotong University, Xi'an, China, in 1993, and the M.S.E.E. and Ph.D. degrees in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2002, respectively. He was a Visiting Professor with Stanford University, Stanford, CA, USA, from 2010 to 2011. He was the Chief Architect with the Chinese National FUTURE Fourth-Generation (4G) TDD Working Group, from 2003 to 2006 and established the 4G TDD CoMP Trial Network, in 2006. He is currently a Professor with the BUPT. He is a fellow of the Institution of Engineering and Technology. He is the inventor or co-inventor of 50 patents and the author or coauthor of 120 papers in 4G and beyond 4G. He is currently involved in fifth-generation networking technology and mobile network security.

PING ZHANG (M'07–SM'15–F'18) received the Ph.D. degree from Beijing University of Posts and Telecommunications, in 1990, where he is currently a Professor. He has published eight books and more than 400 papers, and he holds approximately 170 patents. His current research interests include mobile communications, ubiquitous networking, and service provisioning, especially in the key techniques of the 5G systems. He is an Executive Associate Editor-in-Chief of information sciences on the Chinese Science Bulletin, a member of next-generation broadband wireless communication network in National Science and Technology Major Project committee, a member of the 5th Advisory Committee of the National Natural Science Foundation of China (NSFC), the Chief Scientist of the 973 National Basic Research Program of China, a member of the Ministry of Science and Technology (MOST) 863 Program Expert Team, and a member of MOST IMT-Advanced 5G Expert Team.