IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 12, 2021, accepted December 22, 2021, date of publication February 7, 2022, date of current version March 2, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149765

Feedback Control for QoS-Aware Radio Resource Allocation in Adaptive RAN

HARUHISA HIRAYAMA, YU TSUKAMOTO, AND HIROYUKI SHINBO
KDDI Research Inc., Saitama 356-8502, Japan

Corresponding author: Haruhisa Hirayama (ha-hirayama@kddi.com)

This work was supported by the Ministry of Internal Affairs and Communications, Japan, under Grant JPJ000254.

ABSTRACT In order to meet the quality requirements of various communication services in the advanced 5G era around 2025, the authors have proposed an Adaptive radio access network (RAN) that changes the placement of virtualized base station functions according to the services. In Adaptive RAN, each base station function has its own scheduler, and multiple schedulers share common radio resources. Therefore, in order to guarantee communication quality for more users, it is necessary to control the allocation of the radio resources required for scheduling to each base station function with neither excess nor deficiency. This allocation control is performed based on the information collected from the base station functions by the RAN intelligent controller (RIC). If a large amount of information is used for control, allocation can be performed with higher accuracy, but the collection of information overuses the limited bandwidth of the network. Therefore, it is necessary to reduce the information volume required for control while at the same time guaranteeing communication quality. In this paper, we propose a method that guarantees communication quality while reducing the information volume required for control by combining a required resource estimation control in a long cycle and a resource allocation modification control based on feedback information related to communication quality in a short cycle. We evaluated the impact of the proposed method on communication quality compared to conventional methods through simulation, and verified that the proposed method can guarantee better communication quality than conventional methods while suppressing the increase in information volume.

INDEX TERMS 5G, SLA assurance, radio resource allocation, RAN slicing, RAN virtualization.

I. INTRODUCTION

The 5th generation mobile communication system (5G system) has been launched in numerous countries since around 2020. In the 5G system, various communication services such as 4K video streaming, autonomous driving, and internet of things (IoT) are supported in the same mobile system. These services are categorized into enhanced mobile broadband (eMBB), ultra-reliable and low-latency communication (URLLC), and massive machine-type communication (mMTC) [1] according to their communication characteristics. Around 2025, there will be services with higher performance requirements. On the one hand, the requirements will be enhanced to achieve a higher peak data rate, higher connection density and lower latency. On the other hand, services will need to support a combination of the multiple communication characteristics of eMBB, URLLC, and mMTC.

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

To meet the quality requirements of various more advanced services provided for individual UEs, we have proposed an Adaptive radio access network (RAN) [2], [3], in which virtualized base station functions are flexibly deployed to different sites based on the service requirements.

Adaptive RAN consists of multiple virtualized central units (vCUs) and virtualized distributed units (vDUs) to flexibly provide a suitable configuration for the services. Multiple vDUs are connected to a single radio unit (RU). A vDU has a radio scheduler, which is in charge of allocating physical resource blocks (PRBs) to the user equipment (UE). An RU provides communication services to UEs using a frequency bandwidth corresponding to PRBs. This means that the total number of PRBs used for scheduling is determined by the frequency bandwidth of the RU. Since multiple vDUs are connected to a single RU, the PRBs corresponding to the frequency bandwidth of the RU need to be allocated to vDUs to secure the PRBs required for scheduling in each vDU. In order to meet the requirements of more UEs given the limited PRBs, the PRBs required for scheduling should be allocated to vDUs without any excess.

As a related study to PRB allocation to vDUs, PRB allocation to schedulers in virtualized RAN [4] is studied. It can be categorized into two types based on the control cycle: short-cycle control [5], [6] and long-cycle control [7], [8]. However, in the PRB allocation in both cycles, there are the following problems. **Problem 1**: In a short cycle, a lot of information is collected for the PRB allocation, and the quality requirements can be satisfied well. However, the limited bandwidth of the network being wasted by collecting a lot of information is a problem. **Problem 2**: If the information collection cycle is too long, the quality requirements not being satisfied due to allocation errors is a problem.

To address these problems, we propose a method for PRB allocation to vDUs that satisfies the quality requirements while suppressing the data volume required for information collection. The proposed method is achieved by collecting feedback information related to communication quality in a short cycle. Since the feedback information has less data volume than the information used in the conventional short-cycle control, the required data volume can be suppressed. Therefore, the problem 1 is solved. Moreover, the feedback information is more informative, so the proposed method can satisfy the quality requirements even with a small data volume, thus solving the problem 2. We also implement simulations to analyze the performance of the proposed method. The simulation results show that the proposed method can better satisfy the quality requirements than the conventional methods [5], [7] with small data volume.

The rest of this paper is organized as follows. In section II, the system model of Adaptive RAN and problems for radio resource allocation are described. In section III, we propose a resource allocation method that satisfies the quality requirements while suppressing the data volume required for information collection. In section IV, we evaluate the performance achieved by the proposed method compared to conventional methods. In section V, we conclude the paper.

II. RADIO RESOURCE ALLOCATION IN ADAPTIVE RAN

A. SYSTEM MODEL

The Adaptive RAN architecture is shown in Fig. 1. Adaptive RAN is composed of a RAN intelligent controller (RIC) responsible for controlling RAN, vCUs responsible for header compression, encryption of user data, and processing of the control signals, vDUs responsible for scheduling and modulation/demodulation, and RUs responsible for transmission of radio waves. A vCU and a vDU can be deployed at either a local office or an antenna site. Achievable communication performance, such as bandwidth required for transport links, communication delay, and ability to support intercell coordination depends on placement. Placing vCUs and vDUs flexibly according to the quality requirements of the services and traffic demand enables a base station configuration to be provided that is suitable for the services concerned.

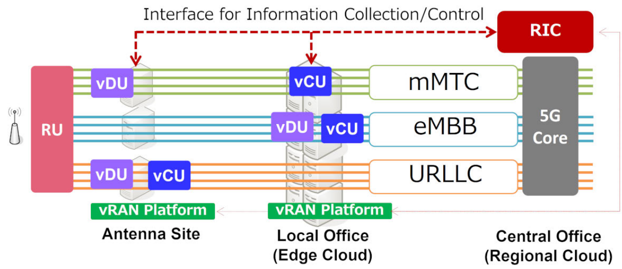


FIGURE 1. Adaptive RAN architecture.

For example, for eMBB services, throughput of a UE at the cell edge is improved by utilizing intercell coordination by placing a vDU at the local office. For URLLC services, latency can be lowered by placing both a vCU and a vDU at the antenna site and utilizing mobile edge computing (MEC). For mMTC services, computing resources can be used efficiently due to statistical multiplexing achieved by placing a vCU at the local office and aggregating the processing of more UEs. Multiple vCUs and vDUs placed according to the service characteristics are connected to a common RU.

B. PROBLEMS IN RADIO RESOURCE ALLOCATION

In Adaptive RAN, the vDU that has a scheduling function responsible for allocating PRBs to UEs is split into multiple vDUs, so common PRBs corresponding to the frequency bandwidth of the RU are shared among multiple vDUs. Therefore, PRBs are allocated to UEs in two steps: allocation of PRBs to each vDU, and scheduling for UEs using the PRBs allocated to each vDU. The number of PRBs required by a vDU for scheduling to UEs varies depending on the radio quality and traffic volume of the UEs under the vDU. In order to meet the requirements of more UEs within the limited PRBs, the PRBs required for scheduling should be allocated to vDUs without any unnecessary excess.

In allocating PRBs to vDUs, the RIC collects performance information from vDUs and decides how many PRBs will be allocated to each vDU. As described above, the number of PRBs decided on is sent to vDUs as a control command. Collecting the performance information and sending the control command is done via the interface between vDUs and the RIC. As a related study to PRB allocation to vDUs in Adaptive RAN, PRB allocation to schedulers in two-level scheduling [4] for virtualized RAN is proposed. In two-level scheduling, since the scheduler is split, scheduling is performed in two steps: allocation of PRBs to each scheduler, and scheduling for UEs using the PRBs allocated to each scheduler. Many studies have investigated PRB allocation to schedulers [5]–[8]. These methods are categorized into two types based on the control cycle: short-cycle control, in which the control cycle closely approximates the transmission time interval (TTI), and long-cycle control, in which the control cycle is 1 second or longer. Each of the two conventional methods have problems such as an increase in the information volume required for control and deterioration in communication quality due to allocation errors.

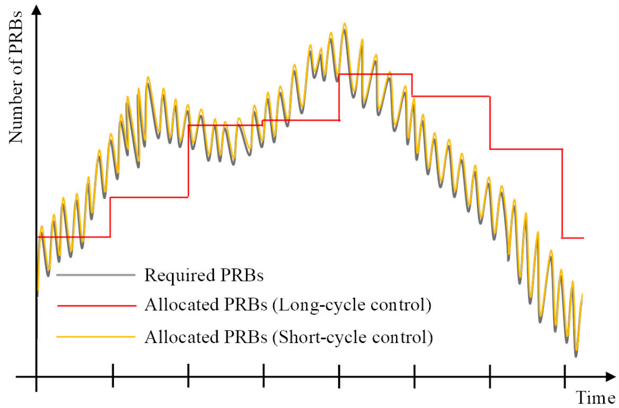


FIGURE 2. Image of time variation in the number of allocated PRBs and required PRBs for two allocation methods.

Fig. 2 shows an image of the time variation in the number of allocated PRBs and required PRBs for short-cycle control and long-cycle control. Since short-cycle control as proposed in [5], [6] collects information such as radio quality and traffic volume over a short cycle, the number of PRBs required by the vDU can be estimated with a high degree of accuracy as shown in Fig.2. However, in Adaptive RAN, since the RIC and vDUs are placed in separate locations, a huge amount of information needs to be sent via the interface between them. In a real environment, a large number of vDUs are accommodated under a single RIC, hence, collecting the information in short-cycle control has the problem of the bandwidth of the interface becoming tight. The long-cycle control proposed in [7], [8] collects information such as radio quality and traffic volume of the UEs in a long cycle to allocate PRBs to schedulers. Since this method needs just a small amount of information due to the long control cycle, the amount of information collected via the interface cannot become a problem. Since this method is used for control in a long cycle, the number of PRBs required by each vDU on average can be estimated and allocated to vDUs as shown in Fig.2. However, since the number of PRBs required by a vDU can change instantaneously due to changes in the radio environment and traffic volume, allocation errors such as insufficient or surplus PRBs occur. Since a limited number of PRBs are distributed to multiple vDUs, if there is a surplus of PRBs in one vDU, there can be a shortfall in the PRBs allocated to other vDUs. If the number of PRBs made available to the vDU is insufficient due to these allocation errors, PRBs required by the UEs cannot be sufficiently allocated by the vDU in scheduling to the UEs. Therefore, there is a problem that the communication quality of the services accommodated in the vDU deteriorates.

III. PROPOSED RESOURCE ALLOCATION METHOD

In this section, we propose a method of allocating PRBs to vDUs that balances suppression of data volume for information collection with assurance of communication quality.

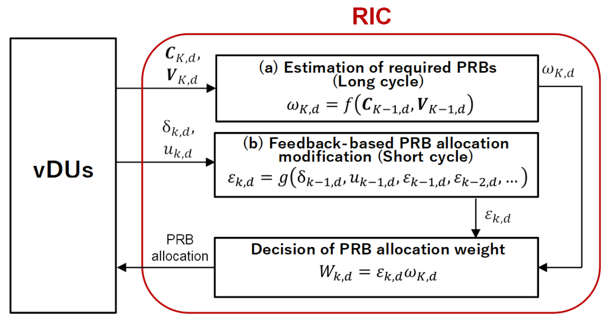


FIGURE 3. Diagram of proposed resource allocation method.

A. OVERVIEW OF PROPOSED RADIO RESOURCE ALLOCATION

A block diagram of the proposed PRB allocation method is shown in Fig. 3. In the proposed method, to decide the allocation weights for vDUs, we combine two functions with different time scales: (a) Estimation of required PRBs, which is equivalent to the conventional long-cycle control, and (b) Feedback-based PRB allocation modification in short cycle.

As the operation sequence, first, $f(\boldsymbol{C}_{K-1,d}, \boldsymbol{V}_{K-1,d})$ in (a), a function for estimating the average number of PRBs required by a vDU, is calculated based on a channel quality vector $\boldsymbol{C}_{K-1,d}$ and a traffic volume vector $\boldsymbol{V}_{K-1,d}$. These vectors consist of values for the UEs u under vDU d in time step $K-1$ as expressed in (1) and (2). Note that K is the long-cycle time step for (a), while k is the short-cycle time step for (b). In this paper, we use channel quality indicator (CQI) values as a measure of channel quality, and $c_{K-1,d,u}$ represents the spectrum efficiency [bps/PRB] corresponding to the CQI value to simplify the notation in the equation. $\omega_{K,d}$, the number of PRBs required by a vDU, is calculated in a long cycle based on the spectrum efficiency for each CQI value [9] and traffic volume for the UEs under vDU d as shown in (3). Here, N_d is the number of UEs under vDU d. Second, in (b), a correction factor $\varepsilon_{k,d}$ for modifying the allocation according to the instantaneous estimation error for $\omega_{K,d}$ that occurred in (a) is calculated in a short cycle. Finally, the PRB allocation weight $W_{k,d}$ to the vDU is determined as the product of the number of PRBs $\omega_{K,d}$ estimated in (a) and the correction factor $\varepsilon_{k,d}$ in (b), and allocation is performed. The controls in (a) and (b) are periodically performed in different cycles. Each time $\omega_{K,d}$ or $\varepsilon_{k,d}$, which is the output of each control, is updated, $W_{k,d}$ is determined and PRBs are allocated to vDUs.

In the proposed method, the communication quality is guaranteed by adjusting the correction factor based on the degree to which quality in each vDU has deteriorated, and modifying the PRB allocation among vDUs. Specifically, correction factor $\varepsilon_{k,d}$ is calculated by the function g using the past values of quality deterioration rate $\delta_{k,d}$, PRB usage rate $u_{k,d}$ and $\varepsilon_{k,d}$. Note that g is a function whose formula changes depending on the conditions as explained in section III-B.

Since quality deterioration occurs when the number of PRBs allocated to a vDU is insufficient, $\varepsilon_{k,d}$ is increased in proportion to the quality deterioration rate. On the other hand, when the PRB usage rate of a vDU is low, since the limited number of PRBs available are being unnecessarily allocated to the vDU, $\varepsilon_{k,d}$ is decreased in proportion to the percentage of unused PRBs. By changing the correction factors in this way, PRBs can be transferred from a vDU that has unused PRBs to a vDU that would otherwise be short of PRBs and result in quality deterioration. Therefore, the proposed method can solve the problem of quality deterioration caused in the conventional long-cycle control. Moreover, the proposed method is effective in suppressing the information volume. Since the vectors $\boldsymbol{C}_{K-1,d}$ and $\boldsymbol{V}_{K-1,d}$ consist of values for the UEs under vDU d, their information volume increases according to the number of UEs. In the conventional short-cycle control, the RIC needs to collect these vectors in the short cycle, so the required information volume becomes huge. On the other hand, $\delta_{k,d}$ and $u_{k,d}$ are per-vDU values, not per-UE values, so the information volume required for the proposed method is small even when they are collected in the short cycle. Therefore, the proposed method can solve the problem regarding information volume caused in the conventional short-cycle control.

$$\boldsymbol{C}_{K-1,d} = \begin{bmatrix} c_{K-1,d,1} & \dots & c_{K-1,d,u} & \dots & c_{K-1,d,N_d} \end{bmatrix} \tag{1}$$

$$\boldsymbol{V}_{K-1,d} = \begin{bmatrix} v_{K-1,d,1} & \dots & v_{K-1,d,u} & \dots & v_{K-1,d,N_d} \end{bmatrix} \tag{2}$$

$$\omega_{K,d} = f(\boldsymbol{C}_{K-1,d}, \boldsymbol{V}_{K-1,d}) = \sum_{u=1}^{N_d} \left\lceil \frac{v_{K-1,d,u}}{c_{K-1,d,u}} \right\rceil \tag{3}$$

B. DECIDING THE CORRECTION FACTOR BASED ON FEEDBACK INFORMATION

In the proposed method described in III-A, since the degree to which the quality requirements are satisfied depends on how the correction factor $\varepsilon_{k,d}$ is determined, the determination of $\varepsilon_{k,d}$ corresponding to function g in Fig. 3 is explained in detail in this section. Although feedback control is widely used in control theory, the proposed method is unique in that it designs the control according to the conditions such as the degree of quality deterioration and resource availability in order to satisfy the quality requirements efficiently.

The likelihood of quality deterioration due to allocation error depends on the quality requirements and traffic fluctuations for the services accommodated in the vDU. For example, in a vDU that accommodates services with strict delay requirements, when the vDU is momentarily short of PRBs, the delay requirements cannot be satisfied, thus causing quality deterioration. However, in a vDU that accommodates services with loose delay requirements, since buffered data only needs to be sent within the required delay, quality deterioration does not occur even when there is a momentary shortage of PRBs. In addition, since the degree to which the number of PRBs required by a vDU fluctuates depends on the traffic load of the UEs and the number of UEs actively sending traffic under the vDU, the likelihood of quality deterioration changes over time. In this way, the likelihood of quality deterioration due to PRB allocation errors varies from vDU to vDU and from time to time, so the correction factor $\varepsilon_{k,d}$ should be determined periodically for each vDU. The correction factor $\varepsilon_{k,d}$ for each vDU is determined periodically for the three cases (1)-(3). In 1), when quality deterioration occurs only in some vDUs, PRBs are transferred from other vDUs. 2) and 3) are exceptional cases that 1) is not applicable to, and PRBs are rebalanced among vDUs when quality deterioration does not occur in all vDUs or when quality deterioration occurs in all vDUs.

1) IF $\delta_{k,d} \leq \delta_{\text{th}}$ IN SOME vDUs AND $\delta_{k,d} > \delta_{\text{th}}$ IN OTHER vDUs

$\varepsilon_{k,d}$ is calculated by (4). u_{th} is the threshold value at which quality deterioration is considered to occur in the vDU. Depending on the values of $u_{k,d}$ and $\delta_{k,d}$, $\varepsilon_{k,d}$ is further calculated by one of (4a) to (4c). (4a) means that for a certain vDU if the PRB usage rate is higher than the threshold u_{th} and the quality deterioration rate is lower than the threshold δ_{th}, the PRB allocation is considered to be correct, and the correction factor of one time step before is carried over to the next time step. (4b) means that for a vDU for which the PRB usage rate is higher than the threshold value and the quality deterioration rate is higher than the threshold value, the correction factor is increased in proportion to the quality deterioration rate in order to provide a PRB margin. Here, $\sigma^{\text{inc},d}$ is a coefficient that increases $\varepsilon_{k,d}$ according to $\delta_{k,d}$. (4c) means that for a vDU for which the PRB usage rate is lower than the threshold value, the correction factor is decreased according to the percentage of PRBs unused by the vDU, and the excess allocation is reduced. Here, $\sigma^{\text{dec},d}$ is a coefficient that decreases $\varepsilon_{k,d}$ according to the percentage of unused PRBs $1 - u_{k,d}$. $E_{k,d}$ is the moving average of $\varepsilon_{k,d}$ and calculated by (5) using the number of sampled time steps n. By repeating (4a) to (4c), the correction factor approaches a value that does not make excess PRBs and does not cause quality deterioration. As a result, the allocation to the vDU that had surplus PRBs will be reduced, and the allocation to the vDU that caused quality deterioration will be increased. Therefore, the overall communication quality is better satisfied within the limited number of PRBs available.

$$\varepsilon_{k,d} = \begin{cases} \varepsilon_{k-1,d} \quad \text{if } u_{k,d} > u_{\text{th}} \,\&\, \delta_{k,d} < \delta_{\text{th}} & \text{(4a)} \\ \left(1 + \sigma^{\text{inc},d}\delta_{k,d}\right) E_{k,d} \quad \text{if } u_{k,d} > u_{\text{th}} \,\&\, \delta_{k,d} > \delta_{\text{th}} & \text{(4b)} \\ \left\{1 - \sigma^{\text{dec},d}\left(1 - u_{k,d}\right)\right\} E_{k,d} \quad \text{if } u_{k,d} < u_{\text{th}} & \text{(4c)} \end{cases} \tag{4}$$

$$E_{k,d} = \sum_{i=1}^{n} \frac{\varepsilon_{k-i,d}}{n} \tag{5}$$

2) IF $\delta_{k,d} \leq \delta_{\text{th}}$ IN ALL vDUs

$\varepsilon_{k,d}$ is calculated by (6). This equation means that when no quality deterioration occurs in all vDUs, $\varepsilon_{k,d}$ is changed so that the difference in PRB utilization among vDUs becomes smaller. This prevents over allocation of PRBs to only a portion of vDUs. Here, $\hat{u}_k$ is the average PRB usage rate of all vDUs, and $\sigma^{\text{bal\_u},d}$ is a coefficient that adjusts $\varepsilon_{k,d}$ according to the difference between the PRB usage rate of the vDU $u_{k,d}$ and average $\hat{u}_k$.

$$\varepsilon_{k,d} = \left\{1 + \sigma^{\text{bal\_u},d} \left(u_{k,d} - \hat{u}_k\right)\right\} E_{k,d} \tag{6}$$

3) IF $\delta_{k,d} > \delta_{\text{th}}$ IN ALL vDUs

$\varepsilon_{k,d}$ is calculated by (7). This equation means that when the PRB usage rate is above the threshold for all vDUs, $\varepsilon_{k,d}$ is changed so that the difference in the quality deterioration rate among vDUs becomes smaller. This prevents the quality deterioration being concentrated on certain vDUs. Here, $\hat{\delta}_k$ is the average quality deterioration rate of all vDUs, and $\sigma^{\text{bal\_d},d}$ is a coefficient that adjusts $\varepsilon_{k,d}$ according to the difference between the quality deterioration of the vDU $\delta_{k,d}$ and average $\hat{\delta}_k$.

$$\varepsilon_{k,d} = \left\{1 + \sigma^{\text{bal\_d},d} \left(\delta_{k,d} - \hat{\delta}_k\right)\right\} E_{k,d} \tag{7}$$

$\sigma^{\text{inc},d}$, $\sigma^{\text{dec},d}$, $\sigma^{\text{bal\_u},d}$ and $\sigma^{\text{bal\_d},d}$ are coefficients that determine the responsiveness of the PRB allocation, that is, the degree to which $\varepsilon_{k,d}$ is raised or lowered depending on the quality deterioration rate and PRB usage rate for each vDU. If the control cycle is short enough compared to the traffic variation, and feedback-based PRB allocation modification is performed frequently, there will be no difference in the responsiveness of the PRB allocation to variations in the PRBs required by vDUs, unless the coefficient is an extremely large value such that $\varepsilon_{k,d}$ changes several tens of percent in one time step or an extremely small value where $\varepsilon_{k,d}$ hardly changes. Therefore, in this paper, we set all of the coefficients to 0.1 to search for the appropriate value for $\varepsilon_{k,d}$ in several control cycles. Moreover, the time window $T_{\text{usg},d}$ for calculating the PRB usage rate $u_{k,d}$ is important because it determines how well PRB allocation is in sync with the PRB usage rate. Since when there are excess PRBs in a vDU, it is better to transfer the surplus PRBs to other vDUs immediately, $T_{\text{usg},d}$ is set to the control cycle (i.e. 5 ms). In addition to these parameters, the time window $T_{\text{pf},d}$ for calculating the quality deterioration rate $\delta_{k,d}$ is important because it determines how well PRB allocation is in sync with the quality deterioration rate. Since an appropriate value for $T_{\text{pf},d}$ depends on the required delay of the services accommodated in each vDU, how to determine this value is described in IV-C. Values of the parameters used in the evaluation are shown in Table 1.

TABLE 1. Value of parameters for evaluation.

$\sigma^{\text{inc},d}$	$\sigma^{\text{dec},d}$	$\sigma^{\text{bal\_u},d}$	$\sigma^{\text{bal\_d},d}$	u_{th}
0.1	0.1	0.1	0.1	0.9
$T_{\text{pf},1}$	$T_{\text{pf},2}$	$T_{\text{pf},3}$	$T_{\text{usg},d}$	δ_{th}
1.0	0.1	0.1	0.005	0.01

IV. PERFORMANCE EVALUATION

In this section, we evaluate the communication quality achieved by the proposed method and the conventional short-cycle control [5] and long-cycle control [7] by simulation to validate the improvement in the performance achieved by the proposed method.

A. EVALUATION CONDITIONS

This section describes the specifications and traffic of each service used in the evaluation. 3GPP has investigated future services, and their rough requirements such as throughput, delay, packet size, and arrival rate have been studied [10]. We chose 6 representative services with different characteristics to be used in the simulation. According to [10], we define 6 services as shown in Table 2. Note that we set throughput equivalent to HD video transmission (i.e. 4.5 Mbps [11]) for services that include video data transmission. This is because the system capacity for simulation is not enough to accommodate many users that use ultra-high-definition video services (e.g. 8K video). Traffic is generated based on the interrupted poison process (IPP) model [12] to emulate the traffic variation. In addition, by randomly switching each UE's traffic on and off between 0 and 120 seconds, variations in the number of active UEs are simulated. The target average throughput for each service is set to the value shown in Table 2. The number of UEs associated with each service is decided so that the average of the total number of PRBs required by all the UEs fits within the system upper limit. Each service is associated with a service ID from S1 to S6. Each vDU accommodates 2 services. Simulation parameters for the radio environment and PRB allocation are shown in Table 3. To simulate a fading channel, we use 7 hexagonal cells with a radius of 200 m. The UEs are placed only in the coverage area of the center cell. The UEs move according to a random walk model within a range such that the CQI value cannot be 0. The speed of the UEs is set to 30 km/h. The channel quality is calculated including pathloss and shadowing in the 3GPP UMi model [13], and Rayleigh fading. The UEs are associated with a service (e.g. S1 to S6) one by one. The system upper limit for the number of PRBs is 273. Each vDU has its queue and scheduler. Packets generated in a higher layer are stored in the queue. For scheduling for UEs in each vDU using the PRBs allocated by each allocation method, we use a proportional fairness algorithm that considers throughput and delay to decide scheduling weights for the UEs [14]. The packets are divided into transport blocks (TBs) based on the transport block size (TBS) set by the scheduler. For the TBs, a block error occurs with a probability of 10%. When a block error occurs for a TB, the TB is retransmitted on a priority basis.

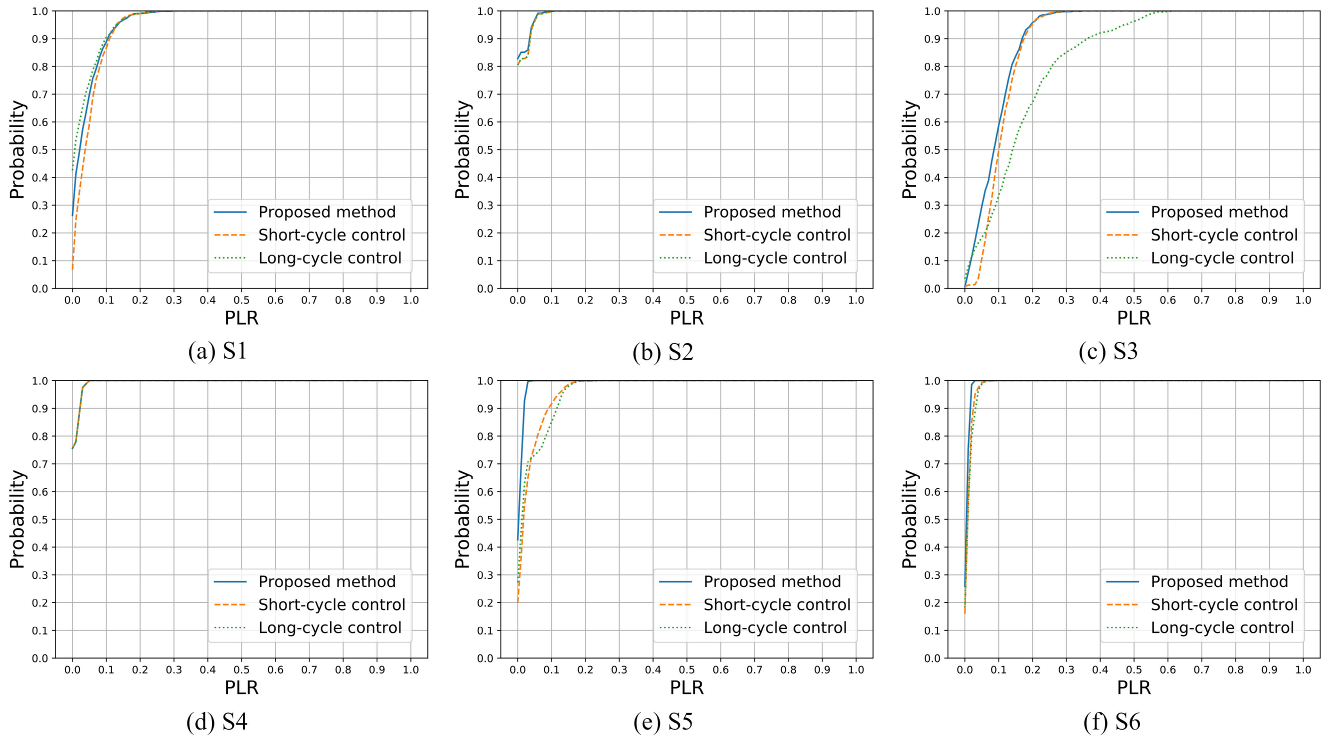


FIGURE 4. CDF for PLR for different PRB allocation methods.

TABLE 2. Service model for simulation.

ID	vDU	Service Type	Service Name	Throughput [Mbps]	Delay [ms]	Packet Size [bytes]	Arrival Rate [1/s]	Number of UEs
1	1	eMBB	HD Video Streaming	4.5	1000	1500		16
2	1	eMBB	Web Browsing	0.22	1000	1500		50
3	2	eMBB/URLLC	Connected Car	4.5	6	1500		16
4	2	URLLC	Factory Automation (FA)		6	50	50	50
5	3	mMTC	Process Automation (PA)		100	100	10	100
6	3	mMTC	IoT		100	50	10	200

TABLE 3. Simulation parameters for radio environment and PRB allocation.

Number of vDUs	3
Number of UEs	432
Simulation time	1800 s
Periodicity of PRB allocation	5 ms
Periodicity of required PRB estimation	1 s
Cellular layout	Hexagonal grid, 7 cells
Antenna pattern	Omnidirectional
Cell radius	200 m
Pathloss, Shadowing	3GPP UMi
Fading model	Rayleigh fading
UE speed	30 km/h
Center frequency	3.6 GHz
Bandwidth	100 MHz
Sub-career spacing	30 kHz
Transmit power	11 dBm/PRB
Number of PRBs	273

B. EVALUATION RESULTS

In the evaluation, if the retention time of the buffered packet exceeds the required delay, the quality requirement is not satisfied. Therefore, the packet is set to be discarded when the retention time of the packet exceeds the required delay. Under this condition, the packet loss ratio (PLR) becomes equal to the quality deterioration rate, so we use the PLR as the evaluation index. Moreover, only for S1 to S3, which have throughput requirements, we also use the throughput as the evaluation index. In addition, if the total number of PRBs required by all vDUs is much smaller than the upper limit of the number of PRBs in the system, there will be unused PRBs, so quality deterioration will not occur in any vDUs regardless of the allocation method. Therefore, the communication quality is compared only for the period when the total number of PRBs required by all vDUs is 95% or more of the upper limit, since there will not be excess PRBs during the period and control is required.

First, we compare the methods in terms of the PLR. Fig. 4 shows the cumulative distribution function (CDF) of the PLR for each service when the proposed method and the conventional methods, long-cycle control [5] and short-cycle control [7], are used. From Fig. 4 (b), (d), and (f), there is no difference in PLR for S2, S4 and S6 between the 3 methods. This is because S2, S4, and S6 are accommodated in vDU1, 2, and 3, respectively, but the quality requirements are looser and deterioration is less likely to occur compared to S1, S3, and S5, respectively. Therefore, regardless of the method used, the PLR is low enough so that there is no difference in the PLR for S2, S4 and S6. Next, from Fig. 4 (a) for S1, the proposed method has a slightly lower PLR than short-cycle control and a slightly higher PLR than long-cycle control. From Fig. 4 (c) for S3, the proposed method has a slightly lower PLR than that of short-cycle control and significantly lower than that of long-cycle control. The average PLR of the proposed method, short-cycle control, and long-cycle control is 9.1%, 10.1%, and 17.0%, respectively. That is, the PLR of the proposed method is improved by about 1.0 point compared to short-cycle control and about 7.9 points compared to long-cycle control. From Fig. 4 (e) for S5, the PLR of the proposed method is lower than that of short-cycle control and long-cycle control. The average PLR of the proposed method, short-cycle control, and long-cycle control is 0.7%, 3.2%, and 3.5%, respectively. That is the PLR of the proposed method is improved by about 2.5 points compared to short-cycle control and about 2.8 points compared to long-cycle control. Fig. 5 shows the CDF of the average PLR of all services. From Fig. 5, the average PLR of all services for the proposed method is lower than that of short-cycle control and long-cycle control. The average PLR of the proposed method, short-cycle control, and long-cycle control is 2.5%, 3.5%, and 4.3%, respectively. That is the PLR of the proposed method is improved by about 1.0 point compared to short-cycle control and about 1.8 points compared to long-cycle control.

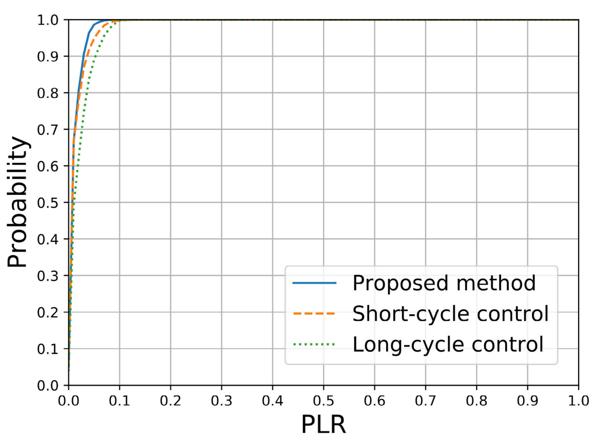


FIGURE 5. CDF for average PLR of all services.

Second, we compare the methods in terms of throughput. Fig. 6 shows the CDF of throughput for S1, S2, and S3, which have the throughput requirements. From Fig.6 (a)-(c), the throughput has the same trend as that of the PLR. This is because throughput decreases when packets are discarded, and they have a strong correlation. Since there is almost no difference in throughput among the methods in S1 and S2, we focus on only S3. From Fig. 6 (c) for S3, the average throughput of the proposed method, short-cycle control, and long-cycle control is 4.12 Mbps, 4.05 Mbps, and 3.75 Mbps, respectively. That is, the throughput of the proposed method is improved by about 0.07 Mbps compared to short-cycle control and about 0.37 Mbps compared to long-cycle control.

TABLE 4. Parameters used for data volume calculation.

Number of vDUs	3
Periodicity of short-cycle control	5 ms
Periodicity of long-cycle control	1 s
Data volume for CQI	4 bits
Data volume for an unquantized value	64 bits

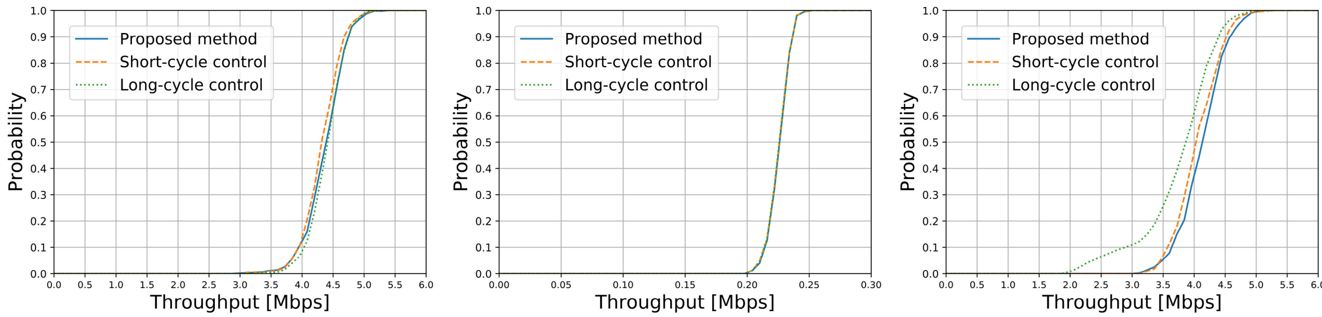
Third, we compare the information volume for each method. Fig.7 shows the information volume for the number of UEs when using the proposed method and the two conventional methods. In calculating the information volume, the parameters shown in Table 4 are used. While CQI is a quantized value in 4 bits, traffic volume, quality deterioration rate, and PRB usage rate are not quantized, so their data volume is set to 64 bits corresponding to IEEE 754 double-precision binary floating-point numbers used in JSON, etc. We assume that the per-UE CQI and traffic volume are collected by the RIC with a period of 5 ms and 1 s for the conventional short-cycle control and long-cycle control, respectively. For the proposed method, in addition to information required for the conventional long-cycle control, quality deterioration rate and PRB usage rate are collected from each vDU by the RIC with a period of 5 ms. The data volume is calculated under these conditions. From Fig. 7 (a) and (b), the proposed method can significantly reduce the information volume compared to conventional short-cycle control. Compared to the long-cycle control, the proposed method slightly increases the information volume. This is because the proposed method uses the information for quality deterioration rate and PRB usage rate collected in a short cycle in addition to the information used in long-cycle control. The increase in data volume due to this additional information is 76.7 kbps regardless of the number of UEs. For example, the information volume in the proposed method can be reduced to about 1.1% when the number of UEs is 1000 and about 0.56% when the number of UEs is 10000, compared to conventional short-cycle control.

The above results show that the communication quality can be improved compared to short-cycle control and with only a slight increase in the information volume for long-cycle control. Thus, it has been verified that the proposed method guarantees communication quality while requiring only a small information volume.

C. CONSIDERATION

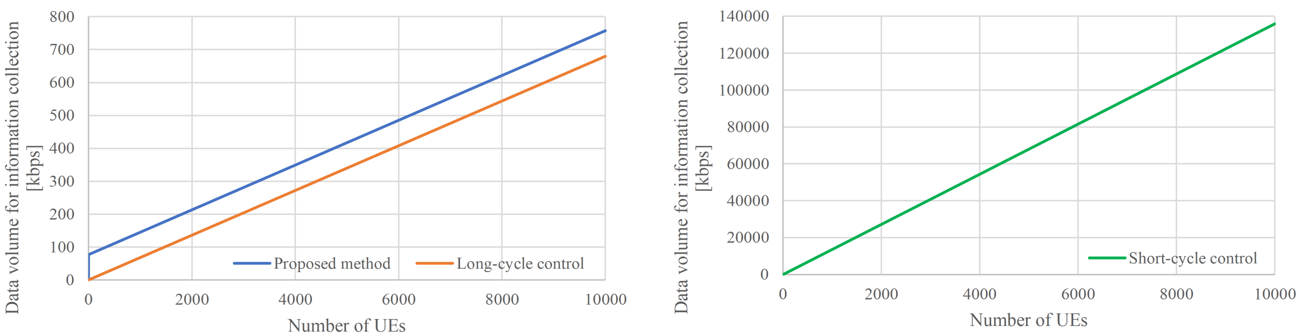
1) CONSIDERATION OF COMMUNICATION QUALITY IMPROVEMENT ACHIEVED BY THE PROPOSED METHOD

First, the difference in communication quality when compared to long-cycle control is considered. Since the PLR and throughput have the same trend, we consider only PLR as



(a) S1 (b) S2 (c) S3

FIGURE 6. CDF for throughput for different PRB allocation methods.



(a) Proposed method and Long-cycle control (b) Short-cycle control

FIGURE 7. Required data volume for each PRB allocation method.

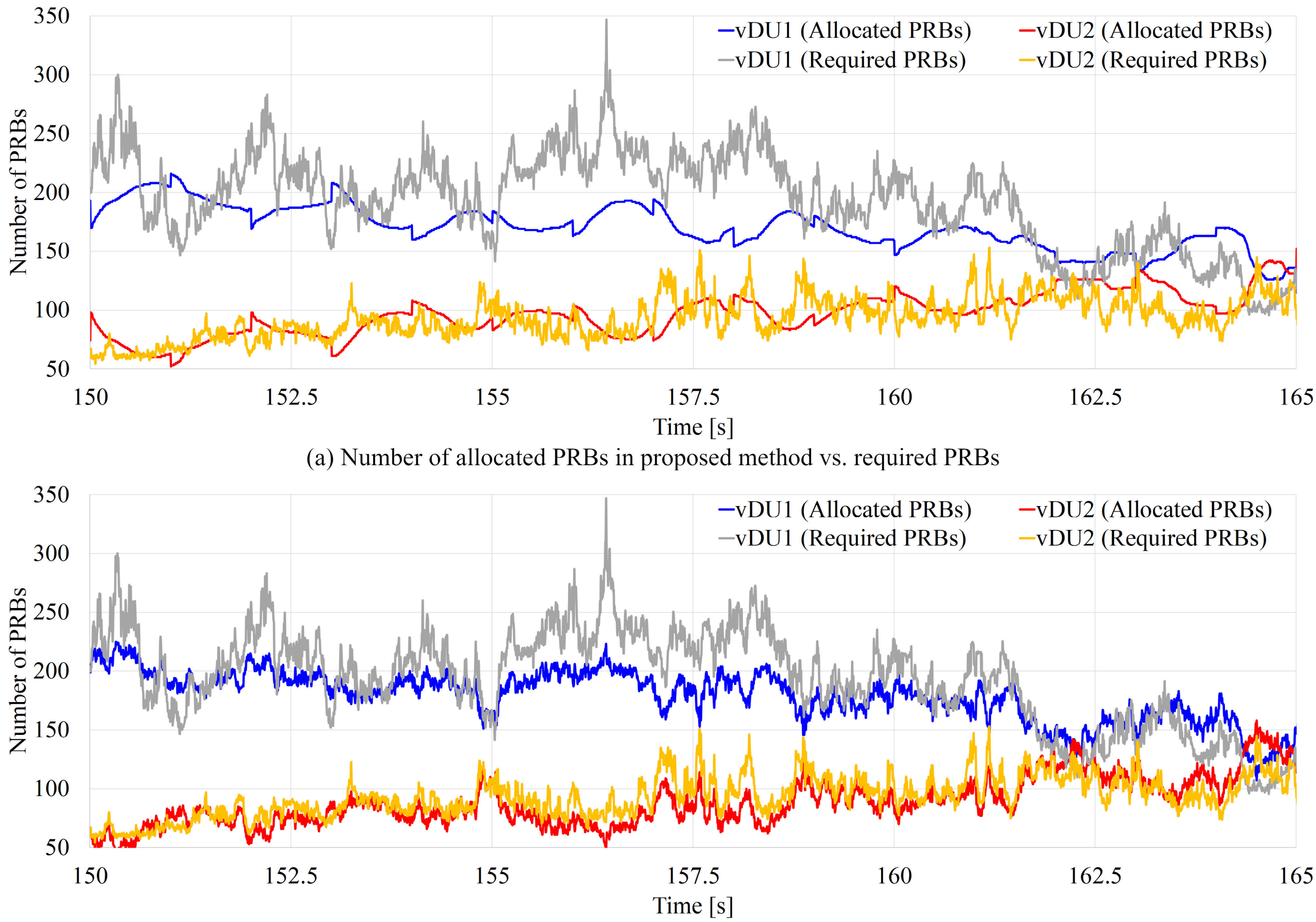
the communication quality. In long-cycle control, more PRBs than actually required are often allocated to vDU1 due to allocation error, so the quality of S1 is considered to be slightly better than the proposed method on average. In addition, in long-cycle control, allocation error occurs in all vDUs. By comparison, in the proposed method, PRBs are more often allocated to vDU2, which accommodates low-latency services, in response to the occurrence of instantaneous quality deterioration due to feedback-based PRB allocation modification. Therefore, the proposed method provides more quality improvement in the order of S3 and S5, which have shorter delay requirements.

Second, to identify the reason for the difference in the communication quality between short-cycle control and the proposed method, we compare the time variation of the number of allocated PRBs and the required PRBs. Fig. 8 shows the time variation of the number of allocated PRBs and required PRBs over a certain period for vDU1 and vDU2 with the proposed method and with short-cycle control.

From Fig. 8 (b), in short-cycle control, quality deterioration continues to occur during the period when the overall PRBs are insufficient (153-162 s), because the situation where the number of allocated PRBs falls short of the number of required PRBs continues in vDU2 where the delay requirement is strict. On the other hand, from Fig. 8 (a), in the proposed method, although quality deterioration occurs in vDU2 where the delay requirement is strict, the shortfall in the number of allocated PRBs does not continue because the allocation weight is increased according to the deterioration rate. In this way, the quality deterioration of S3 in vDU2 is suppressed. In this case, the lack of PRBs in vDU1 continues due to the increased allocation weight of vDU2. However, if the packet retention time does not exceed 1 s, which is the required delay for S1 and S2, quality deterioration does not occur. When the packet retention time exceeds 1 s, the allocation weight of vDU1 is increased according to the quality deterioration rate of vDU1, so that the increase in quality deterioration due to allocating too many PRBs does not occur.

2) CONSIDERATION OF INFORMATION VOLUME REDUCTION ACHIEVED BY THE PROPOSED METHOD

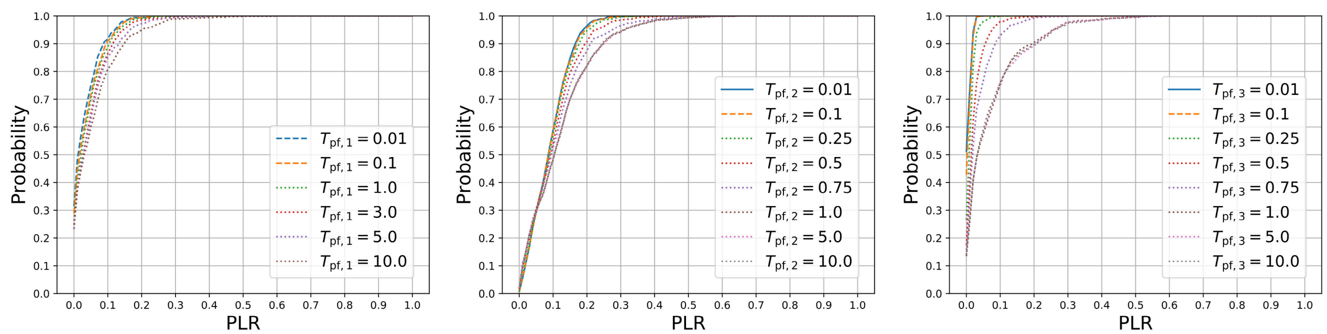
In conventional long-cycle control and short-cycle control, because radio quality information and traffic information are collected from each UE, the information volume increases in proportion to the number of UEs. Since the information collection cycle and the information volume per UE are inversely proportional, the slope of short-cycle control is 200 times greater than that of long-cycle control as shown in Fig. 7 (a) and (b). Therefore, in short-cycle control, the information volume increases significantly as the number of



(a) Number of allocated PRBs in proposed method vs. required PRBs

(b) Number of allocated PRBs in short-cycle control vs. required PRBs

FIGURE 8. **Time variation of the number of allocated PRBs and required PRBs for vDU1 and vDU2.**



(a) S1 (b) S3 (c) S5

FIGURE 9. **CDF for PLR for different $T_{\mathrm{pf},d}$ values.**

UEs increases. On the other hand, in the proposed method, information is collected only from three vDUs regardless of the number of UEs, thus reducing the increase in the information volume even though the information is collected in the same cycle as the conventional short-cycle control.

3) CONSIDERATION OF THE VALUES OF THE PARAMETER IN THE PROPOSED METHOD

In order to determine the value of the parameter $T_{\mathrm{pf},d}$ in the proposed method described in III-B, the influence of its value on communication quality is discussed.

$T_{pf,d}$ is the time window for calculating the quality deterioration rate $\delta_{k,d}$, and it is the parameter that determines the responsiveness of PRB allocation to quality deterioration. Since the delay requirements and traffic patterns of the services accommodated in each vDU are different, the value of $T_{pf,d}$ set for each vDU has an impact on communication quality. To determine the value of $T_{pf,d}$ for each vDU, we compared the communication quality by changing the value of $T_{pf,d}$. Fig. 9 shows the CDF of the PLR of the services belonging to each vDU for different values of $T_{pf,d}$. From Fig. 9 (a)-(c), different PLR trends are observed for each vDU for different values of $T_{pf,d}$. From Fig 9 (a), the quality deterioration of S1 belonging to vDU1 increases when $T_{pf,d}$ is longer than 1 s, which is the delay requirement of S1. From Fig. 9 (b), the quality deterioration of S3 belonging to vDU2 increases when $T_{pf,d}$ is longer than 0.25 s. From Fig. 9 (c), the quality deterioration of S5 belonging to vDU3 increases when $T_{pf,d}$ is longer than 0.1 s.

The reasons for the above PLR trends for each vDU are explained in the following. Since the services accommodated in vDU1 and vDU3 have less stringent delay requirements (i.e. 1 s and 0.1 s, respectively), the data buffering time is less likely to exceed the delay requirement, thus decreasing quality deterioration due to allocation error. However, when $T_{pf,d}$ is set longer than the required delay, the quality deterioration increases. This is because even when the buffering time exceeds the required delay and quality deterioration is occurring, the impact of the recent deterioration on the quality deterioration rate is small, so sufficient PRBs are not allocated for sending data in the buffer, thus increasing the level of quality deterioration. On the other hand, when $T_{pf,d}$ is set shorter than the required delay, there is almost no difference in PLR. This is because the recent deterioration has a major impact on the quality deterioration rate, and the buffered data can be sent faster, but the quality deterioration rate remains the same as long as it is sent within the required delay. Therefore, for vDU1 and vDU3, $T_{pf,d}$ needs to be set to a value equal to or less than the required delay. In this evaluation, we set $T_{pf,1} = 1.0$ and $T_{pf,3} = 0.1$. The services under vDU2 have a strict delay requirement of 6 ms, and an instantaneous shortage of PRBs in the vDU results in immediate quality deterioration. However, by increasing the weight assigned to the vDU through the proposed method, the deterioration caused by allocation errors can be suppressed. In the proposed method, the correction factor set based on the quality deterioration rate is judged to be sufficient for the instantaneous variation in the required number of PRBs, and its value varies. In this case, it is enough to determine whether the PRB margin is sufficient as a trend, rather than whether the individual PRBs for each time step are sufficient. As can be seen from the change in the number of required PRBs in Fig. 8 (a), for an interval of about 0.1 s, although the required number of PRBs fluctuates due to fast fading of the channel, it is possible to determine whether or not the correction factor is appropriate. On the other hand, in intervals exceeding 0.25 s, the trend in the number of required PRBs changes significantly due to shadowing, path loss, and traffic fluctuations. Therefore, if $T_{pf,d}$ is set above 0.25 s, it is not possible to appropriately determine whether the margin is sufficient as a trend. Therefore, we set $T_{pf,2} = 0.1$ in this evaluation.

V. CONCLUSION

We proposed a PRB allocation method for Adaptive RAN that can improve the communication quality while reducing the information volume, and evaluated the impact of the proposed method on communication quality. In the proposed method, the PRBs are allocated to the vDUs that accommodate low-latency services and are prone to quality deterioration, thereby reducing quality deterioration.

In the simulation, the communication quality of the proposed method is compared with that of the conventional control methods. The simulation results show that the proposed method reduced the quality deterioration for low latency and high capacity services. In other services, the communication quality hardly deteriorates, so the overall quality is improved by the proposed method.

In terms of the information volume, we confirmed that the proposed method significantly reduces the information volume compared to conventional short-cycle control.

The results show that the proposed method is effective in guaranteeing the communication quality with less information needing to be collected, because the communication quality is better satisfied with a slight increase in the information volume from long-cycle control than with short-cycle control.

In this paper, we evaluated the proposed method for a certain set of services within a certain cell. In future work, we plan to extend this method to further improve the communication quality by considering multiple cells, and to evaluate it with various sets of services. We will also conduct evaluations on a large scale, assuming a real environment.

ACKNOWLEDGMENT

A part of this work was conducted under "R&D for further advancement of the 5th generation mobile communication system, R&D of technology for high reliability management of advanced 5G network to various requirements of communication services" commissioned by Research and Development for Expansion of Radio Wave Resources of the Ministry of Internal Affairs and Communications in Japan (JPJ000254).

REFERENCES

[1] *IMT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, document ITU-R M.2083-0 Sep. 2015.

[2] H. Shinbo, T. Murakami, Y. Tsukamoto, and Y. Kishi, "R&D of technology for high reliability management in advanced 5G network to meet the various requirements of different communication services," in *Proc. IEEE VTS 17th Asia Pacific Wireless Commun. Symp. (APWCS)*, Aug. 2021, pp. 1–5.

[3] Y. Tsukamoto, H. Hirayama, S. I. Moon, S. Nanba, and H. Shinbo, "Feedback control for adaptive function placement in uncertain traffic changes on an advanced 5G system," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–6.

[4] A. Ksentini and N. Nikaein, "Toward enforcing network slicing on RAN: Flexibility and resources abstraction," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 102–108, Jun. 2017.

[5] L. Li, N. Deng, W. Ren, B. Kou, W. Zhou, and S. Yu, "Multi-service resource allocation in future network with wireless virtualization," *IEEE Access*, vol. 6, pp. 53854–53868, 2018.

[6] A. Ksentini, P. A. Frangoudis, P. C. Amogh, and N. Nikaein, "Providing low latency guarantees for slicing-ready 5G systems via two-level MAC scheduling," *IEEE Netw.*, vol. 32, no. 6, pp. 116–123, Nov. 2018.

[7] S. Bakri, P. A. Frangoudis, and A. Ksentini, "Dynamic slicing of RAN resources for heterogeneous coexisting 5G services," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[8] J. Perez-Romero, I. Vila, O. Sallent, B. Blanco, A. Sanchoyerto, R. Solozabal, and F. Liberal, "Supporting mission critical services through radio access network slicing," in *Proc. Int. Conf. Inf. Commun. Technol. Disaster Manage. (ICT-DM)*, Dec. 2019, pp. 1–8.

[9] *Physical Layer Procedures for Data*, document TS 38.214 V16.6.0, 3GPP, Jun. 2021.

[10] *Study on New Services and Markets Technology Enablers*, document TR 22.891 V1.2.0, 3GPP, Nov. 2015.

[11] *YouTube Help*. Accessed: Feb. 4, 2022. [Online]. Available: https://support.google.com/youtube/answer/1722171

[12] C. Baugh, *4IPP traffic model for IEEE 802.16. 3. IEEE802*, IEEE 802 LAN/MAN Standards Committee (Task Group: IEEE 802.16 Broadband Wireless Access Working Group), 2000. [Online]. Available: https://www.ieee802.org/16/tg3/contrib/802163c-00_51.pdf

[13] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document TR 38.901 V16.1.0, 3GPP, Dec. 2019.

[14] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, 2nd Quart., 2013.

HARUHISA HIRAYAMA received the B.E. degree in information and communication engineering from Tokyo Metropolitan University, Tokyo, Japan, in 2015, and the M.E. degree in electrical and electronic engineering from the Tokyo Institute of Technology, Tokyo, in 2017. He joined KDDI Corporation, in 2017, and has been with KDDI Research Inc. His research interests include radio access networks and networks slicing.

YU TSUKAMOTO received the A.E. degree in electronic control system engineering from the National Institute of Technology, Numazu College, Shizuoka, Japan, in 2012, the B.E. degree in electrical engineering from the Kyoto Institute of Technology, Kyoto, Japan, in 2014, and the M.E. degree in electrical engineering from Kyoto University, Kyoto, in 2016. He joined KDDI Corporation, Tokyo, Japan, in 2016. Since 2017, he has been with KDDI Research Inc., Saitama, Japan, where he has studied radio access networks technologies.

HIROYUKI SHINBO received the B.S. degree in electro information communication from The University of Electro-Communications, Tokyo, Japan, in 1987, and the M.S. degree from the Graduate School of Information and System, The University of Electro-Communications, in 1990. He joined KDD Corporation (now KDDI Corporation), in 1999, and currently works with the KDD Research and Development Laboratory (now KDDI Research Inc.). He was seconded to work with the Advanced Telecommunications Research Institute International, from 2013 to 2016. Currently, he is the Senior Manager of the Advanced Radio Application Laboratory, KDDI Research Inc. His research interests include beyond 5G/6G systems (especially radio access networks), TCP/IP, flying base stations, networks operation systems, and space communications.