Enabling Grant-Free URLLC for AoI Minimization in RAN-Coordinated 5G Health Monitoring System

Beom-Su Kim, Byung Hyun Lim, Beomkyu Suh, Sangtae Ha, Ting He, Babar Shah and Ki-Il Kim*

***Abstract*—Age of information (AoI) is used to evaluate the performance of 5G health monitoring systems because stale data can be fatal for patients with serious illness. Recently, grant-free ultra-reliable and low latency communications (URLLC) have shown greater potential of minimizing AoI than conventional grant-based approaches; however, existing grant-free schedulers cannot provide guaranteed performance in 5G health monitoring systems because they involve two fundamental problems in time and frequency domains, namely the joint scheduling problem and physical resource block (PRB) allocation. In this study, we investigate two resource allocation problems for the first time, aiming to enable grant-free URLLC to minimize AoI in 5G health monitoring systems. Specifically, we propose two adaptive solutions based on an open radio access network-coordinated wireless system: 1) a joint scheduling algorithm and 2) an adaptive PRB allocation algorithm. To verify the effectiveness of the proposed solutions, we built a simulation environment similar to a real health monitoring system and captured the performance variations under realistic deployment scenarios.**

***Index Terms*—age of information, low-power sensors, grant-free URLLC, joint scheduling.**

I. Introduction

FIFTH-generation (5G) wireless technology is expected to support a wide range of emerging application services in addition to cellular broadband services. One promising usage scenario for 5G is ultra-reliable and low-latency communication (URLLC). Various applications identified for 5G URLLC require different service-level agreements (SLAs). In particular, strict SLAs are expected in healthcare services that are directly related to human life. The 5G public-private partnership association specified that a delay of 30 ms or less should be guaranteed for remote health monitoring services [1]; however, the key performance indicator for health monitoring services is the freshness of information because stale data can prove to be fatal to a patient's life or meaningless owing to the arrival of newly sampled data. To address this concern, a new concept called the age of information (AoI) was conceived in a previous study [2]. AoI is a metric indicating the freshness of information and is defined as the elapsed time between the generation time of the data sampled at the source node and the present time from the perspective of the destination.

The waiting time within the buffer is primarily responsible for increasing the AoI of the sampled data; hence, the main line of research in regards to AoI is the design of transmission schedulers [3]–[19]. These commonly assume a single-hop wireless system in which N source nodes collect data and forward them to a base station (BS) over a shared wireless channel. Considering the limited channel capacity, the time is slotted and the BS attempts to find an optimal scheduling policy that minimizes the long-term average or peak AoI of the entire system. These schedulers satisfy AoI constraints with a strong theoretical guarantee; however, it is rarely possible to provide guaranteed performance in real-world scenarios because they do not comply with the latest transmission standards (i.e., 5G), nor do they account for time-varying channel conditions.

5G-compliant AoI schedulers have been proposed to overcome these limitations [20]–[24]. The main advantages of an AoI scheduler that conforms to 5G communication standards are twofold: 1) Unlike conventional AoI schedulers that use fixed-size timeslots (i.e., 1 ms), AoI schedulers in a 5G-based wireless system divide time into variable transmission time intervals (TTIs); the duration of a slot varies depending on numerology [25]. Adopting a sub-millisecond TTI allows the scheduler to minimize the waiting time of the sampled data and respond quickly to the time-varying channel conditions. 2) In each slot, a large number of physical resource blocks (PRBs) can be assigned to each source node for uplink transmission; hence, 5G schedulers can achieve a lower AoI by assigning PRBs to multiple source nodes in each slot.

As mentioned above, existing 5G schedulers can greatly minimize the waiting time of sampled data via numerology that fits best with URLLC service; however, they have two fundamental problems that increase the AoI: 1) Real-world applications are heterogeneous, and thus the sampling period, data length, and priority are different among the source nodes. To minimize the AoI, the 5G scheduler must allocate uplink slots that match the sampling start time of each source node; however, existing approaches simply assume a general wireless system in which the sampled data are homogeneous and take one slot for each transmission. 2) Source nodes (i.e., sensors) periodically generate samples of information according to their sampling rates; thus, frequent handshaking for scheduling requests between the source node and BS increases the AoI. Hence, the 5G scheduler should also reduce the control plane latency caused by handshaking procedures. However, existing 5G schedulers assume grant-based uplink scheduling.

Beom-Su Kim, Byung Hyun Lim, Beomkyu Suh and Ki-Il Kim are with the Department of Computer Science and Engineering at Chungnam National University, Korea.
Sangtae Ha is with Computer Science Department at the University of Colorado Boulder, Boulder, US
Ting He is with School of Electrical Engineering and Computer Science at Pennsylvania State University, State College, US
Babar Shah is with College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates
*Corresponding author (kikim@cnu.ac.kr)
Manuscript received April 19, 2005; revised August 26, 2015.

Grant-free AoI schedulers were proposed [26]–[30] to address the above issues. Existing grant-free schemes preconfigure uplink slots that match the sampling start time of each source node; thus, they can greatly reduce the waiting time of sampled data and control plane latency caused by handshaking procedures. In particular, in a health monitoring system where source nodes periodically generate samples of information, grant-free scheduling is important for reducing the waiting time of sampled data. However, it cannot provide guaranteed performance in 5G health monitoring systems because the fundamental problems in the time and frequency domains remain unsolved.

a) Joint scheduling problem: Owing to short-range wireless communications, data packets generated from body sensors pass through the coordinator and BS in a 5G health monitoring system (see Figure 1). To minimize AoI in such a two-hop wireless system, the coordinator should allocate timeslots (in the 2.4-GHz band) that match the traffic generation time of each source node, and the grant-free scheduler should allocate uplink slots as close as possible to the arrival time of each packet at the coordinator. That is, both scheduling policies should be jointly optimized to minimize the AoI for low-power body sensors; however, this has never been explored in any prior work.

b) PRB allocation problem: In the frequency domain, the grant-free scheduler must allocate PRBs to each source node after it arranges uplink slots in the time domain. However, PRB allocation entails a complex decision-making problem owing to the limited number of PRBs within a given bandwidth part (BWP). To minimize the weighted sum of the AoI, existing approaches [31]–[33] preferentially allocate PRBs based on the priority of the source nodes; however, the average AoI may increase owing to retransmissions when PRBs are assigned to source nodes with poor channel quality. In addition, the long-term average AoI may also increase owing to the starvation of lower-priority nodes.

In this study, we investigate the above resource allocation problems to enable grant-free URLLC for AoI minimization in a 5G health monitoring system. Specifically, we propose a joint scheduler based on an open radio access network (O-RAN)-coordinated system [34] in which a RAN intelligent controller (RIC) collects traffic information from all source nodes and manages their transmission schedules. To minimize the weighted sum of the long-term average AoI for all health monitoring traffic in the time domain, the proposed scheduler jointly configures the uplink scheduling policies of the coordinator and BS via a heuristic approach. In the frequency domain, the proposed scheduler transforms the PRB allocation problem into a multi-criteria decision-making (MCDM) problem and then determines the order of PRB allocation.

The main contributions of this study are summarized as follows:

- **Design of relay framework:** In a 5G health monitoring system, source nodes are interconnected with the coordinator (Figure 1); hence, the RIC should communicate with the coordinator to obtain global knowledge of the source nodes. To enable interaction between the coordinator and RIC, we design a relay framework that runs on the coordinator. Using the relay framework, the coordinator forwards internal traffic information to the RIC and receives the corresponding scheduling information.
- **Joint scheduling algorithm:** To minimize the AoI, the proposed joint scheduler aims to assign uplink slots that match the sampling start time of each source node in the time domain via an offline scheduling algorithm in which all required traffic information are available. However, the AoI may increase if the coordinator does not allocate an appropriate number of timeslots (in a 2.4-GHz band) to each source node. In this paper, we address this as a joint scheduling problem. However, the joint scheduling problem is in the domain of combinatorial optimization, which is NP-hard. To solve this problem, we first show that the joint scheduling problem is NP-hard and then find a near-optimal solution via a heuristic approach.
- **Adaptive PRB allocation algorithm:** After the proposed scheduler arranges uplink slots, it must assign PRBs to each source node in the frequency domain. However, PRB allocation involves a complex decision-making problem owing to the limited number of PRBs for each slot. To address this problem, we propose a new metric based on the MCDM method [35]. Using this metric, the proposed scheduler can make a flexible decision according to changes in the network conditions without significantly changing the result of offline scheduling.
- **Practical simulation:** Unlike information-theoretic AoI studies, which are based on numerical evaluations, this study conducts practical simulations in real-world scenarios using NS-3. To build a realistic health monitoring system, we added certain principles on the proposed scheduling scheme based on the IEEE 802.15.6 and 5G communication standards. We built a simulation environment similar to a real health monitoring system and captured the performance variations under various network scenarios.

The remainder of this study is organized as follows. Section II presents a review of the relevant literature. In Section III, we describe the system model. In Section IV, we present the problem statement and our approach. In Section V, we introduce a relay framework that enables the practical deployment of the proposed solution. Section VI describes the proposed scheme. We present the experimental setup and results of a performance evaluation conducted demonstrate the efficacy of our approach in Section VII. Our conclusions are summarized in Section VIII, along with some suggested avenues for further research.

II. Literature Review

In the previous section, we discussed the motivation for introducing grant-free AoI schedulers and their limitations; however, existing AoI studies employing a single-hop wireless model also have the same potential problem (i.e., the joint scheduling problem) in a two-hop setting. To highlight the motivation for this study, in this section, we describe some existing optimization-oriented AoI schedulers and consider some recent research.

TABLE I: List of symbols

$G_i(t)$	Data Generation time at source node i
$A_i(t)$	AoI for source node i at time t
$\overline{A}$	Weighted sum of long-term average AoI
P	Sampling period
L	Data length
W	Data Priority
d	Deadline
C	Scheduling space
q_t	Sampling time at time t
r_t	Reception time of data at time t
e_{delay}	Expected delay
e_{queue}	Queuing delay
e_{prop}	Propagation delay
e_{tx}	Transmission delay
σ	Standard deviation
V	Timeslot allocation vector
V_{max}	Maximum number of timeslots
Q	Priority queue
c	Iteration counter
X	Pairwise comparison matrix
$\overline{X}$	Normalized pairwise comparison matrix
v	Eigen vector

Conventional AoI schedulers commonly assume a single-hop wireless system in which a source node collects and forwards data to the BS via a shared wireless channel. Specifically, time is equally slotted and the BS determines a scheduling policy that can minimize the AoI with a strong theoretical guarantee. For example, several works [3], [4], [20], [21], [26]–[30] have explored methods to minimize the long-term average AoI of an entire system. Specifically, they defined the average AoI over time t as an objective function and then found an optimal scheduling policy through iterations to minimize the objective. Their mechanism provides a lower bound for the objective function, which is used as a benchmark for developing an AoI scheduler. To minimize the AoI of critical data, several other works [5]–[7], [23] have defined the weighted sum of the long-term average AoI over time t as an objective function. Other authors [8]–[11] have also investigated a new metric called peak AoI, which provides information about the maximum value of AoI. Meanwhile, [12], [24] proposed a new metric called the age of incorrect information to minimize the freshness of abnormal information and provide useful data for monitoring. This metric extends the traditional concept of updating new information to updating informative data.

In contrast, the authors [13]–[18] proposed an AoI scheduler that meets the hard deadlines of each source node. They assumed that each source has an AoI requirement (i.e., a deadline) and that data past the deadline is be dropped. In their objective function, the deadline is used as an upper bound of AoI. In other words, they aimed to design a feasible scheduler that ensures an AoI less than the deadline for all source nodes. The authors of [19] proposed an AoI scheduler that allowed occasional violations of AoI deadlines. Their approach assumed an AoI threshold (i.e., a deadline) and a violation tolerance rate for each source node. Their mechanism aimed to find a scheduler that reduces the violation rate to less than the tolerance rate for all source nodes.

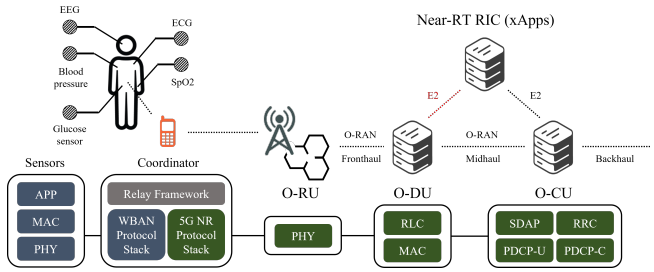


Fig. 1: An illustration of a 5G-compliant health monitoring system.

Recently, a new line of research on AoI, which jointly considers AoI minimization and other constraints has been explored [36]–[38]. For example, the authors of [36] proposed a polynomial-time algorithm designed to meet AoI constraints with a strong theoretical guarantee regarding the required bandwidth. Similarly, the approach presented in [37] aimed to minimize the AoI while guaranteeing power and outage constraints. In this scheme, the rate and powers are updated by using the majorization-minimization principle and geometric programming approximation. The joint scheduling problem was investigated in [38] to optimize the peak AoI at the network edge with directional chargers. Their mechanism derived the theoretical bounds of the peak AoI with respect to the charging latency.

III. System Model

Figure 1 illustrates the envisioned RAN-coordinated 5G health monitoring system. To support on-body communication between the sensor node and coordinator, we constructed a wireless body area network (WBAN) based on the IEEE 802.15.6 standard [39]. A WBAN comprises one coordinator and a set of heterogeneous sensor nodes denoted by $S = \{1, 2, ..., |S|\}$. The coordinator and sensor nodes are interconnected via a one-hop star topology. The sensor nodes are attached to different body parts, and the sampled data are forwarded to the coordinator. In WBANs, the coordinator is responsible for allocating radio resources (in the 2.4-GHz band) to sensor nodes for uplink transmissions, as well as for forwarding the received data to a BS. It must be noted that the BS sends the received data packets to a monitoring server via a 5G core; however, this is not covered in this study.

We jointly consider the uplink scheduling of a BS and set of WBAN coordinators (i.e., WBAN users), denoted by $U = \{1, 2, ..., |U|\}$. The BS allocates PRBs to each WBAN coordinator for uplink transmission; however, it is controlled by a near- real-time RIC (Near-RT RIC). As shown in Figure 1, the Near-RT RIC manages radio resources within the distributed unit (DU) to ensure that the scheduler meets SLAs for URLLC traffic via a slice-specific application (xApp). The proposed joint scheduler (i.e., xApp) running on the near-RT RIC extracts URLLC traffic from the centralized unit

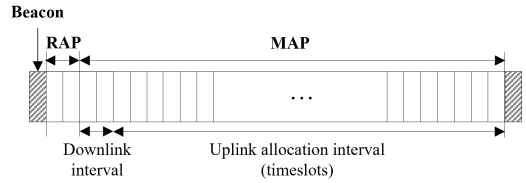


Fig. 2: IEEE 802.15.6-based superframe structure.

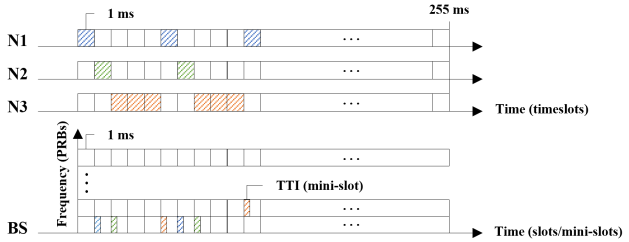


Fig. 3: An illustration of joint scheduling.

(CU) over the E2 interface and controls the RAN elements and resources. In particular, the proposed scheduler manages the uplink scheduling policies of the coordinator and BS to minimize the AoI. A detailed description of the uplink transmission model for each WBAN and 5G new radio (NR) is provided in the following subsection.

A. IEEE 802.15.6-Based Uplink Transmission

This study follows the IEEE 802.15.6 standard, which is a wireless standard for supporting on-body communication. The coordinator adopts beacon mode with superframes as the access mode. In the superframe, the coordinator can arrange the random access phase (RAP) and managed access phase (MAP). During RAP, all sensor nodes can obtain random access channels. MAP is used to reserve uplink timeslots for each source node; hence, we configure the superframe structure, as shown in Figure 2. In MAP, time is equally divided into timeslots of 1 ms each, and the coordinator can allocate one or more consecutive timeslots to each sensor node by considering the size of the sampled data and physical layer (PHY) capabilities. It should be noted that the IEEE 802.15.6 standard limits the maximum superframe length to 255 ms [39], and thus the maximum scheduling cycle is fixed at 255 ms. Upon receiving the beacon frame (i.e., scheduling information is included in the beacon frame), each sensor node sends the sampled data to the coordinator at the scheduled uplink interval. Then, the coordinator forwards the data to the BS.

B. Grant-Free Uplink Transmission

The 5G NR standard introduces the use of grant-free scheduling to eliminate all delays caused by the handshaking procedure in grant-based schemes. With grant-free scheduling, the BS can reserve radio resources for a dedicated terminal device. However, this approach is only suitable for periodic traffic. In this regard, grant-free scheduling is adequate for health monitoring traffic because the transmission period of each WBAN can be planned by the Near-RT RIC. In this study, grant-free scheduling was adopted; thus, the proposed scheduler at the Near-RT RIC preconfigures uplink slots for each WBAN. The internal information of each coordinator (i.e., WBAN) required for grant-free scheduling is forwarded to the proposed scheduler via the E2 interface.

IV. Problem Statement

AoI is defined as the elapsed time between the generation time of the data sampled at source node i denoted by $G_i(t)$ and present time t from the perspective of the coordinator and BS. Thus, the AoI for the source node i can be defined as:

$$A_i(t) = t - G_i(t). \tag{1}$$

On the coordinator side, the time is equally divided into slots of 1 ms each, whereas on the BS side, the time is slotted depending on numerology. For clarity of terminology, we separate the terms "timeslot" and "slot" as the units of time used by the coordinator and BS, respectively. The value of $A_i(t)$ is incremented by $n = 1, 2, 3, \cdots$ if the timeslot for source node i is not reserved by the coordinator at time $t = 0, 1, 2, \cdots$. Thus, $A_i(t)$ at the coordinator can be calculated as:

$$A_i(t+n) = \begin{cases} n, & \text{if timeslots (from } t \text{ to } t+n) = i, \\ A_i(t) + n, & \text{otherwise.} \end{cases} \tag{2}$$

In addition, if the data sampled at source node i are not transmitted to the BS immediately after arriving at the coordinator at time t, $A_i(t)$ at the BS increases by one at time $(t+n)$. Thus, $A_i(t)$ at the BS can be calculated as:

$$A_i(t+1) = \begin{cases} A_i(t) + TTI_i(t), & \text{if } i \in \text{slot } (t), \\ A_i(t) + 1, & \text{otherwise,} \end{cases} \tag{3}$$

where $TTI_i(t)$ denotes the duration of the slots or mini-slots allocated to source node i in slot (t). The long-term average of $A_i(t)$ at the BS is defined as:

$$\overline{A}_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} A_i(t). \tag{4}$$

In WBANs, life-critical data should be prioritized over general data; thus, we aim to minimize the weighted sum of the long-term average AoI for all health monitoring traffic at the BS. The weighted sum of the long-term average AoI for all source nodes $i \in S$ belonging to coordinator $j \in U$ is defined as:

$$\overline{A} = \sum_{j=1}^{U} \sum_{i=1}^{S} W_{ji} \overline{A}_{ji}, \tag{5}$$

where W_{ji} denotes the priority of source node i belonging to coordinator j.

A. Our Approach: Joint Scheduling

Each source node periodically generates samples of information according to the type of application implemented. Owing to the heterogeneity of medical applications, source nodes have different sampling periods, data lengths, and data priorities. We denote the sampling period, data length, and data priority for the source node i belonging to coordinator j as P_{ji}, L_{ji}, and W_{ji}, respectively. Units P_{ji} and L_{ji} are timeslots, and W_{ji} is a constant value ranging from 1 to 8 [39].

Given P_{ji}, L_{ji}, and W_{ji}, we aim to jointly configure the uplink scheduling policies of a BS and set of coordinators (this is referred to as joint scheduling in this work) to minimize $\overline{A}$ (Eq. 5). For example, as shown in Figure 3, each source node (i.e., N1, N2, and N3) has its own sampling period and data length (in units of timeslots). To match the sampling and transmission times of each source node, the proposed scheduler arranges uplink timeslots in the superframe, considering P_{ji}, L_{ji}, and W_{ji}, and then assigns a slot or mini-slot as close as possible to the scheduled timeslot of each source node.

B. NP-Hardness of Joint Scheduling

Minimizing $\overline{A}$ in a two-hop setting requires solving the joint scheduling problem; however, it lies in the domain of combinatorial optimization, which is NP-hard. This means that we must search for the space of all possible combinations of timeslot and slot placements to find a global optimum solution. Moreover, the search space increases exponentially with the number of source nodes and coordinators. Hence, determining an optimal solution in polynomial time is extremely challenging.

Theorem 1. *The joint scheduling problem is NP-hard.*

Proof. We are given a scheduling cycle C and n tasks $< P_1, L_1, W_1 >, ..., < P_n, L_n, W_n >$ for each group j, where each task is determined by a period $P_i \leq C$, length $L_i \leq P_i$, and priority W_i. Task i generates a job (i.e., a sample) of length L_i at each P_i (starting at time 0), which must be scheduled (by the coordinator) before the next job is generated, denoted by $A_i(t) < d_i$, where d_i is the deadline of task i. Subsequently, the scheduled job is scheduled once more (by the BS) before the next job is released, denoted by $A_i(t) \leq d_i$. To prove that the joint scheduling problem is NP-hard, we show that it can be reconstructed as a knapsack problem. Let us consider a knapsack problem with N items and a knapsack capacity B. Subsequently, we set $S = N$ and $C = B$. The value and weight of item i can be converted into W_i and L_i of a job from task i, respectively. The scheduling problem can then be reconfigured to minimize $\overline{A} = \sum_{i=1}^{S} W_i \overline{A}_i$ by arranging n jobs in C. □

C. Proposed Solution: Heuristic-Oriented Scheduling

To address the joint scheduling problem, the proposed scheduling scheme aims to find an appropriate solution in polynomial time, such as earliest deadline first [40] or rate-monotonic [41] scheduling. Thus, this approach is completely different from previous techniques that satisfy AoI constraints with a strong theoretical guarantee [x-x]. In other words, instead of proving that the proposed scheme is optimal, we simply demonstrate that it produces "good" performance in terms of $\overline{A}$ through the results of extensive simulations. Specifically, we split the joint scheduling problem into two sub-problems: 1) a timeslot allocation problem between the source and coordinator and 2) a grant-free slot allocation problem between the coordinator and BS. We then find a suboptimal solution using a heuristic algorithm.

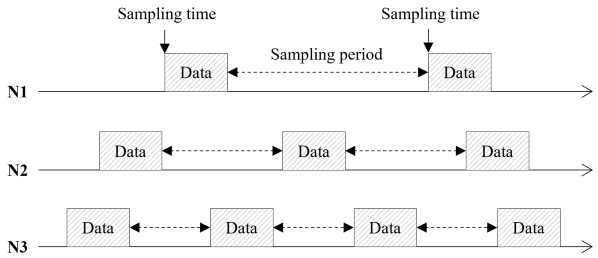


Fig. 4: An example of heterogeneous sampling periods of source nodes.

V. Design of Relay Framework

In this section, we introduce a relay framework that supports the exchange of messages between a user device (i.e., coordinator) and xApp (i.e., joint scheduler). The relay framework operates in addition to the coordinator and bridges the IEEE 802.15.6 and 5G NR protocol stacks (see Figure 1). The main role of the relay framework is two-fold: 1) traffic feature extraction and 2) task offloading. A detailed description of these processes is provided in the following subsection.

A. Traffic Feature Extraction

To enable offline scheduling, the coordinator extracts heterogeneous traffic features of each source node. In this subsection, we propose a specific method for extracting the traffic features (i.e., P_i, L_i, and W_i) for source node i at runtime.

In the IEEE 802.15.6 standard, the medium access control (MAC) layer management entity (MLME) implements the functions of the control plane responsible for resource management. The relay framework interacts with the MLME; that is, the management functions are implemented by the MLME but controlled by the relay framework. In particular, the relay framework is responsible for associating with source nodes, assigning resources, and broadcasting beacon frames.

The IEEE 802.15.6 standard specifies that the coordinator should associate with source nodes to allocate scheduled allocation intervals during an MAP. To request an association, each source node sends a connection request frame during the RAP. The connection request frame includes the length of the sampled data L_i and priority W_i of the source node i. The coordinator (i.e., the relay framework) then returns a connection assignment frame that includes scheduling information at the downlink interval.

During the association phase, the relay framework extracts the traffic features of each source node. When the initial association is established, the relay framework allocates dedicated

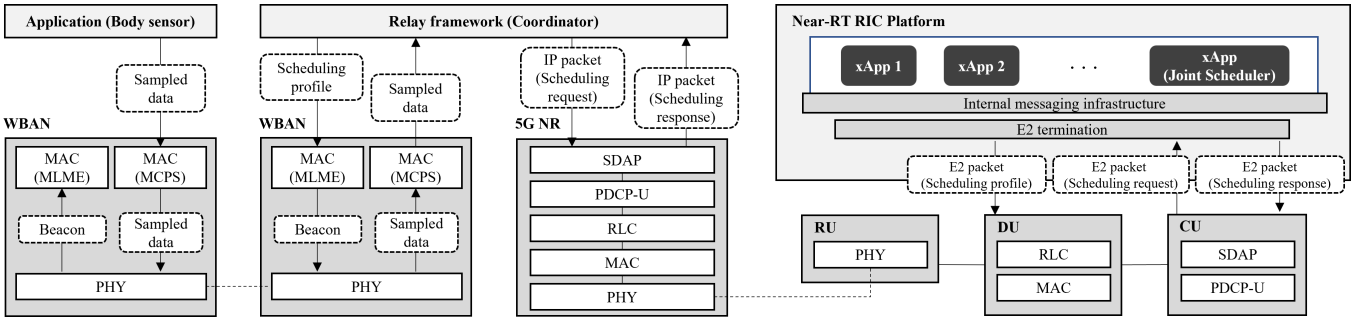


Fig. 5: An illustration of timeslot and grant-free scheduling request/response between relay framework and xApp.

timeslots to the associated node (e.g., N1 in Figure 4) to estimate the sampling period P via a connection assignment frame. In this case, all uplink allocation intervals during the MAP in the superframe are allocated to N1. As soon as the N1 samples are collected, the sampled data can be sent to the coordinator without any queuing delay because all timeslots are dedicated to N1. Upon receiving data from N1, the relay framework calculates the sampling time q at time t as follows:

$$q_t = r_t - e_{delay}, \tag{6}$$

where r_t is the reception time of the data and e_{delay} is the expected delay in transmitting data, which is defined as:

$$e_{delay} = e_{queue} + e_{prop} + e_{tx}, \tag{7}$$

where e_{queue}, e_{prop}, and e_{tx} denote the queuing, propagation, and transmission delays, respectively. Because all timeslots are dedicated to N1 in the association phase, e_{queue} is negligible. The values of the above parameters are presented in [42].

The relay framework can then estimate the sampling period P at time t as follows:

$$P_t = q_t - q_{t-1}. \tag{8}$$

When the sampling period p_t of N1 reaches a steady state, traffic estimation for N1 is completed. Here, a steady state means that the value of p_t does not change over time and is determined as follows:

$$\sigma = \sqrt{\frac{\sum (P - \bar{P})^2}{N}}, \tag{9}$$

where σ and N are the standard deviation and the number of samples of P, respectively. Our previous study [42] showed that at least ten samples were adequate to determine a steady state. Finally, the relay framework determines that the samples of P are in a steady state if σ is less than 0.5.

B. Task Offloading Request

As shown in Figure 5, the relay framework sends a task offloading request (i.e., a grant-free scheduling request) to xApp (i.e., a joint scheduler) in the form of an IP packet over the physical uplink shared channel after extracting the sampling periods for all source nodes in the association phase. The grant-free scheduling request is a user-defined message containing internal traffic information (i.e., P_i, L_i, and W_i for each source node i). Notably, the relay framework preconfigures the IP address of the Near-RT RIC and port number of xApp [34].

Upon receiving the grant-free scheduling request, the CU forwards it to xApp running on the Near-RT RIC via the E2 interface protocol. E2 allows the Near-RT RIC to control the functional operations of the E2 nodes (i.e., DUs and CUs). In the E2 setup phase, a connection between xApp and the E2 node is established via the stream control transmission protocol. The Near-RT RIC collects both the control signals and user-defined messages from the DU/CU periodically or after a predefined trigger event.

The joint scheduler notifies the DU/CU of the timeslot scheduling and grant-free scheduling policies. Each scheduling policy is forwarded to the coordinator through different channels. As shown in Figure 5, upon receiving the grant-free scheduling policy, the DU forwards it to the 5G NR MAC layer on the coordinator over a physical downlink control channel. By contrast, the CU forwards the timeslot scheduling information (i.e., scheduling response) to the relay framework over a physical downlink shared channel.

VI. Proposed Scheme

Recall that we aim to find a suboptimal solution using a heuristic algorithm to solve the joint scheduling problem (i.e., an NP-hard problem). To address this problem, we split the joint scheduling problem into two sub-problems: 1) a timeslot allocation problem (between the source and coordinator) and 2) a grant-free slot allocation problem (between the coordinator and BS). In this section, we describe the proposed scheduling scheme.

A. Scheduling Principles

To build a realistic health monitoring system, we consider the following principles on the proposed scheduling scheme based on the IEEE 802.15.6 and 5G communication standards. **(i)** According to the IEEE 802.15.6 standard, the length of the superframe is limited to 255 ms; thus, we set the timeslot scheduling cycle to 255 ms. The grant-free scheduling cycle is also set to 255 ms. Both the scheduling cycles are synchronized by the relay framework and Near-RT RIC. **(ii)** From the time it receives the beacon frame, each source node

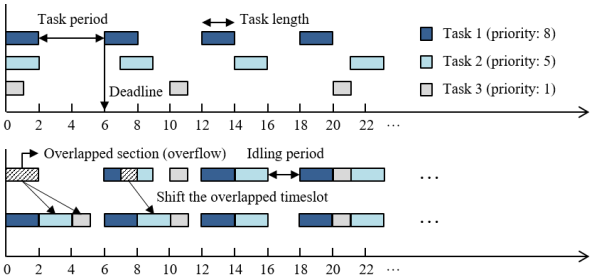


Fig. 6: An example of timeslot scheduling for three tasks.

begins collecting samples according to its sampling period. When a new beacon frame arrives, each source node discards previously sampled data and collects new samples. **(iii)** To guarantee the freshness of information, the data sampled from source node i at time t are discarded unless scheduled until the next sampling period $(t + p_i)$. Thus, the AoI deadlines are determined based on the sampling periods. We set P_i as the upper bound for $A_i(t)$ denoted by d_i. Thus, the scheduler should satisfy the condition $A_i(t) < d_i$. Otherwise, the data packet is dropped. **(iiii)** Recall $A_i(t)$ increases over time if the data sampled at source node i are not transmitted to the BS immediately after arriving at the coordinator at time t. For a data packet from the source node i arriving at coordinator j at time t, the scheduler should allocate a slot or mini-slot to coordinator j within time $n; t \leq n \leq (t + d_i)$.

B. Timeslot Allocation

We transform the timeslot allocation problem into a task-scheduling problem. Given P_i, L_i, W_i for task i, we aim to arrange periodic tasks sequentially by using different priorities in the scheduling space.

To provide a differentiated quality-of-service (QoS) for high-priority nodes, it is obvious that we should minimize the waiting time of the higher-priority tasks over the lower-priority tasks. To achieve this goal, we added the following rules: 1) Higher-priority tasks are preferentially allocated in the scheduling space; that is, a higher-priority task preempts the timeslot of lower-priority tasks. 2) We allocate timeslots to a lower-priority task one or more spaces later than the original sampling period; however, the scheduled timeslot of task i is canceled if it exceeds the deadline (see Principle (iii)).

The basic operation of the proposed timeslot scheduling scheme is illustrated in Figure 6. The higher-priority task (i.e., task 1) preferentially occupies timeslots according to the task period. If the scheduled timeslots overlap (i.e., overflow), the scheduler shifts the overlapped timeslots allocated to the lower-priority task (i.e., task 2) by one until overflow no longer occurs. However, if the scheduled timeslots exceed the deadline, the scheduler cancels the allocation. Once all the periodic tasks are arranged in the scheduling space, idling periods can occur. To prevent the starvation of lower-priority tasks, the scheduler allocates at least one timeslot during the idling period.

Algorithm 1 implements the scheduling principle described above. For each task i within task queue j (line 1), the scheduler preferentially arranges periodic tasks with a higher priority in the given scheduling space (lines 4-9). If an overflow occurs, the scheduler cancels the allocation and shifts the overlapped timeslots by one (lines 10-14). It must be noted that the deadline of task i exceeds the threshold (lines 5-6), and the scheduler cancels the allocation for the current period and moves to the next period of task i (lines 20-21). If the allocation for the source node i is successful, the scheduler allocates grant-free slot allocation (lines 15-18).

Algorithm 1 Joint Scheduling Algorithm

Initialization:

$C \leftarrow 255$ // scheduling space (ms)
$V_j \leftarrow \varnothing$ // timeslot allocation vector of coordinator j
$V_{max} \leftarrow$ maximum number of timeslots within C
P_i: periodicity of task i
L_i: length of task i
W_i: priority of task i
$Q_j \leftarrow < P_i, L_i, W_i >$ // priority queue of coordinator j

Function: Timeslot_Allocation ()

1: **for** each $< P_i, L_i, W_i >$ in Q_j **do**
2: $c \leftarrow 0$ // iteration counter
3: $t \leftarrow 0$ // timeslot index
4: **while** $t < V_{max}$ **do**
5: $d_i \leftarrow t + P_i$ // deadline of task i
6: **for** $k \leftarrow 0, k < L_i$ & $(t - k + L_i) < d_i$ **do**
7: **if** $V[t] = \varnothing$ **then**
8: $V[t] \leftarrow i$
9: $k \leftarrow k + 1$
10: **else**
11: $V[t-1] \leftarrow \varnothing$ // cancel the previous allocation
12: $k \leftarrow 0$
13: **end if**
14: $t \leftarrow t + 1$
15: /* If timeslot allocation is successful */
16: **if** $k = L_i$ **then**
17: **call** Slot_Allocation (j, t, d)
18: **end if**
19: **end for**
20: $c \leftarrow c + 1$
21: $t \leftarrow c * P_i$ // go to the next period of task i
22: **end while**
23: **end for**

C. Grant-Free Slot Allocation

If the timeslot allocation for task i succeeds, the scheduler arranges slots for a grant-free uplink transmission. Notably, the scheduler should assign a slot or mini-slot as close as possible to the scheduled timeslots of task i to minimize the waiting time of the sampled data at the coordinator. To achieve this goal, the scheduler places a slot or mini-slot according to numerology at the next index of the scheduled timeslot for task i (see Figure 3).

Algorithm 2 implements the scheduling principle described above. The reserved slot information is stored in a two-

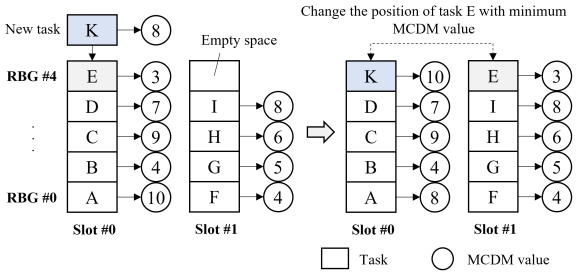


Fig. 7: An example of adaptive PRB allocation.

dimensional array in the time and frequency domains. In the time domain, the scheduler starts assigning slots from the index, following the scheduled timeslot of task i (line 2). The frequency domain is divided into n resource block groups (RBGs) based on the size of the BWP. RBG is a resource-allocation unit in the frequency domain. If there is an available RBG in slot t, the scheduler allocates an RBG to coordinator j connected to the source node i (lines 4-7).

If there are no available RBGs in slot t, the scheduler decides whether the above operation should be repeated in slot $t+1$ or task i should preempt the RBG previously assigned to a task in slot t. To make a flexible decision, the scheduler calculates a new metric using multiple decision parameters that affect the AoI (line 1). This metric is calculated based on the MCDM method, and tasks with larger metric values preempt the RBG (lines 8-13). Figure 7 shows the decision-making process for PRB allocation, which is described in detail in the following section.

D. Adaptive PRB Allocation

As described above, PRB allocation in the frequency domain involves a complex decision-making problem owing to the limited number of RBGs for each slot. To minimize the AoI, the scheduler makes a flexible decision by considering the following decision parameters:

- **QoS class indicator (QCI):** To minimize $\overline{A}$, the scheduler must minimize the waiting time of the higher-priority nodes over the lower-priority nodes. For this, we introduce QCI, which can be defined as the average priority of source nodes belonging to the coordinator.
- **Starvation index (SI):** $\overline{A}$ may increase over time owing to starvation of the lower-priority nodes; thus, we define a new metric called SI, which is incremented by one if a task belonging to coordinator j is not assigned an RBG in the current scheduling cycle.
- **Channel quality indicator (CQI):** If an RBG is assigned to the coordinator with poor channel quality, $\overline{A}$ can significantly increase owing to retransmissions; thus, we use CQI, which is determined when the RIC receives data from the coordinator.

Technically, it is difficult to integrate all the parameters in the decision-making process because they are not related to each other; hence, we require a logical approach to accommodate the decision parameters. In this study, we use an analytic hierarchy process (AHP) [43] to define the relationship

Algorithm 2 Uplink Slot Allocation

Initialization:

$C \leftarrow 255$ // scheduling space (ms)
$T_{max} \leftarrow$ maximum number of slots within C
$B_{max} \leftarrow$ maximum number of RBGs within BWP
$TTI \leftarrow$ initialized by numerology
$B[T_{max}] \leftarrow \varnothing$ // array of PRB allocation vector
QCI_j: QoS class indicator for coordinator j
SI_j: starvation index of coordinator j
CQI_j: channel quality indicator for coordinator j

Function: Slot_Allocation (j, t, d)

1: $r \leftarrow MCDM(QCI_j, CQI_j, SI_j)$
2: **for** $f \leftarrow$ false, t ! = true & $t < d$ **do**
3: **for** $i \leftarrow 0$, $i < B_{max}$ **do**
4: **if** $B[t][i]$ is empty **then**
5: $B[t][i] \leftarrow < j, d, r >$
6: $f \leftarrow$ true
7: **break**
8: **else if** $B[t][i].r < r$ **then**
9: **call** Slot_Allocation ($B[t][i].j$, $t+TTI$, $B[t][i].d$)
10: $B[t][i] \leftarrow < j, d, r >$
11: $f \leftarrow$ true
12: **break**
13: **end if**
14: **end for**
15: $t \leftarrow t + TTI$ // go to the next TTI
16: **end for**
17: /* If PRB allocation is successful */
18: **if** $f = true$ **then**
19: $SI_j \leftarrow 0$ // reset the starvation index
20: **else**
21: $SI_j \leftarrow SI_j + 1$
22: **end if**

between the decision parameters. The AHP requires pairwise comparisons to determine the relative importance of the decision criteria. The results of the pairwise comparisons are represented as a square matrix. For example, X_{ij} represents the relative importance of the factors i to j.

$$X_{ij} = \begin{array}{c} \\ QCI \\ SI \\ CQI \end{array} \begin{array}{c} \begin{array}{ccc} QCI & SI & CQI \end{array} \\ \begin{pmatrix} 1 & 1/2 & 1/3 \\ 2 & 1 & 1/3 \\ 3 & 3 & 1 \end{pmatrix} \end{array} \tag{10}$$

In AHP, the decision-maker should determine the scales of relative importance by considering the network environment. After defining the pairwise comparison matrix, we calculated the normalized eigen vector of matrix X_{ij}. For this purpose, the pairwise comparison matrix was normalized as follows:

$$\overline{X_{ij}} = \frac{X_{ij}}{\sum_{i=1}^{n} X_{ij}} \tag{11}$$

After normalizing the pairwise comparison matrix, the eigen vector of the factor i was computed as follows:

$$v_i = \frac{\sum_{j=1}^{n} \overline{X_{ij}}}{n} \tag{12}$$

where v_i is the relative priority of i among the criteria. Finally, we obtain:

$$v_{X_{ij}} = \begin{matrix} Result \\ \begin{pmatrix} 0.160 \\ 0.251 \\ 0.589 \end{pmatrix} \end{matrix} \tag{13}$$

CQI has the highest priority among the decision criteria, whereas QCI has the lowest priority. After calculating the weight of each criterion, they are combined using a weighted sum equation. To this end, we use the simple additive weighting (SAW) method [44] to obtain the weighted sum of the performance ratings for each alternative. In the SAW method, the decision-maker must evaluate whether each criterion should be maximized or minimized to achieve a high rating. In this study, QCI, SI, and CQI are normalized as follows:

$$n_{ij} = \frac{X_{ij}}{Max(X_{ij})} \tag{14}$$

where n_{ij} represents the normalized value related to criterion X corresponding to row i and column j and $Max(X_{ij})$ denotes the maximum value of X_{ij}. We then calculated the weighted normalization value as:

$$V_{ij} = n_{ij} * v_i \tag{15}$$

where V_{ij}, n_{ij}, and v_i represent the weighted normalization value, normalized value corresponding to row i, column j, and relative weight of criterion i, respectively. Based on V_{ij}, the proposed PRB allocation scheme determines the order of PRB assignments for each task.

E. Time Complexity

The timeslot allocation algorithm is repeated as many times as the total number of tasks i belonging to the task queue j. For each task i, there are d_i iterations for every P_i within the scheduling space C. Thus, the time complexity can be expressed as $O((|n||C||d|)^2)$. However, as d and $C \leq 255$ (see Principles (i) and (iii)), we treat them as constants; thus, we have $O((|n||1||1|)^2) = O(n^2)$. Similarly, the slot-allocation algorithm is repeated as many times as the number of RBGs for each TTI. However, the numbers of RBGs and TTIs vary depending on numerology; therefore, the time complexity can be expressed as $O(n^2)$.

VII. Performance Evaluation

To verify the performance of the proposed solution, we performed comprehensive simulations using ns-O-RAN [45] designed as a pluggable module for ns-3. In this section, the implementation strategies and parameter settings for the simulations are briefly described. Next, we present simulation results and analysis.

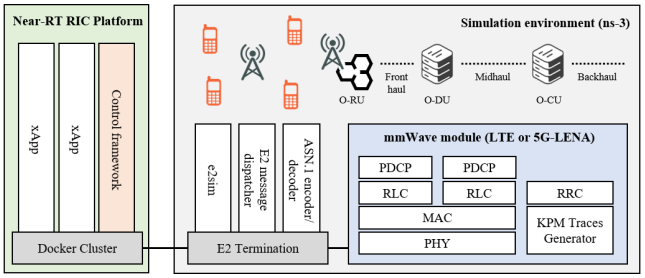


Fig. 8: Illustration of ns-O-RAN architecture.

TABLE II: Node specifications

Type	Application	Priority	Sampling rate
0	Temperature	1	< 10 Kbps
1	Blood pressure	2	< 10 Kbps
2	Insulin pump	3	< 10 Kbps
3	SPO2	4	< 10 Kbps
4	Glucose	5	< 10 Kbps
5	EEG	6	10 Kbps
6	EMG	7	15 Kbps
7	ECG	8	20 Kbps

TABLE III: Simulation parameters for WBAN

Parameter	Value
Frequency band	868 MHz
Modulation	ASK
Data rate	250 Kbps
Access mode	Beacon with superframes
Beacon period	255 ms
Number of source nodes	8
Number of coordinators	100

A. Implementation

The ns-O-RAN simulation platform extends the ns-3 5G `mmWave` module by adding E2 implementation. As shown in Figure 8, the proposed control framework (i.e., xApp) runs on a separate terminal and interacts with the E2 nodes on ns-3 via E2 termination. We embedded the proposed method into ns-O-RAN by simply using E2 termination without additional development effort. The control framework running on the near-RT RIC collects user data from the E2 node (i.e., CU) and controls the function (PRB allocation) of the DU over the E2 interface. Further details are provided in [45]. In this simulator, we implemented the proposed scheme by integrating the 5G NR modules with WBAN modules based on IEEE 802.15.6 developed in our previous study [46].

B. Simulation Parameters

Each coordinator randomly places eight types of medical sensors. The specifications for each source node are listed in Table II. The coordinator and source nodes are interconnected based on the IEEE 802.15.6 standard. The coordinators are randomly deployed in a 1000 m × 1000 m network field and interact with xApp (i.e., a joint scheduler) to report internal traffic information and receive scheduling policies. We ran

TABLE IV: Simulation parameters for 5G NR

Parameter	Value
Frequency band	30 GHz
Duplexing mode	FDD
BWP size for uplink URLLC	10 MHz
Numerology	0
Subcarrier spacing	15 KHz
TTI	1 ms
Access mode	Configured grant
Scheduling interval	1 s

the simulation for 3600 s and then derived $\overline{A}$. The network parameters that affect $\overline{A}$ are set as X-axis variables in the simulation graph. The default values of X-axis parameters are listed in Tables III and IV. The simulation was conducted 100 times with a 95% confidence interval.

C. Simulation Objective

We compared the proposed scheme with two recent 5G-compliant AoI schedulers: 1) a grant-based AoI scheduler [22] and 2) a grant-free AoI scheduler [28]. The purpose of this simulation is two-fold: 1) We determine whether the joint scheduling algorithm can efficiently reduce the waiting time of the sampled data in two-hop settings. 2) We determine whether the PRB allocation algorithm can adaptively adjust the joint scheduling policy according to changes in the network conditions.

To this end, we set up two types of network scenarios. Scenario A assumes an indoor environment with no coordinator movement, and the channel conditions are the same for each TTI. In this scenario, we set the relative importance of the pairwise comparison matrix X_A as follows:

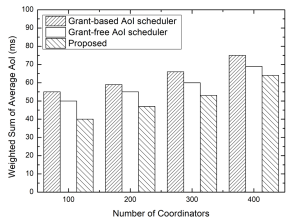
$$X_A = \begin{array}{c} \\ QCI \\ SI \\ CQI \end{array} \begin{array}{c} \begin{array}{ccc} QCI & SI & CQI \end{array} \\ \begin{pmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{pmatrix} \end{array} \tag{16}$$

By contrast, Scenario B assumes an outdoor environment where the coordinators constantly move, and the channel conditions differ for each TTI. In this scenario, we set the importance of CQI to be the highest in the pairwise comparison matrix X_B, which is depicted as follows:

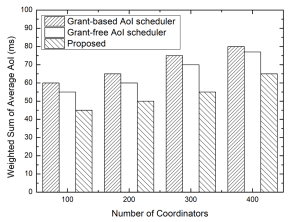
$$X_B = \begin{array}{c} \\ QCI \\ SI \\ CQI \end{array} \begin{array}{c} \begin{array}{ccc} QCI & SI & CQI \end{array} \\ \begin{pmatrix} 1 & 3 & 1/3 \\ 1/3 & 1 & 1/5 \\ 3 & 5 & 1 \end{pmatrix} \end{array} \tag{17}$$

D. Simulation Results

1) Performance Comparison Varying Network Scenarios: In this subsection, we present the analysis of the performance differences between comparable schedulers in Scenarios A and B. As shown in Figures 9, 10, 11, and 12, the performance of the grant-free scheduler is better than that of the grant-based scheduler in both scenarios. This is because the control plane latency is reduced and the waiting time of the sample data is minimized upon reserving uplink resources in advance according to the data generation time.

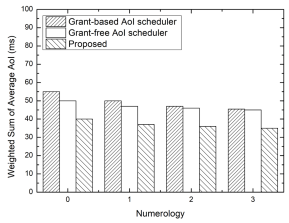


(a) $\overline{A}$ under Scenario A.

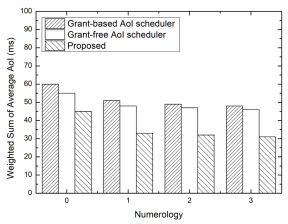


(b) $\overline{A}$ under Scenario B.

Fig. 9: Weighted sum of average AoI: impact of varying number of coordinators.

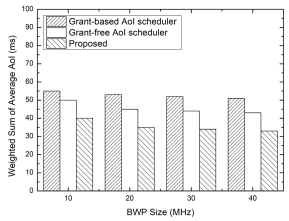


(a) $\overline{A}$ under Scenario A.

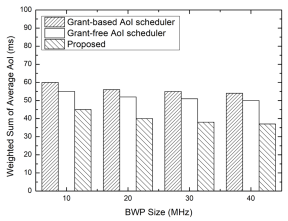


(b) $\overline{A}$ under Scenario B.

Fig. 10: Weighted sum of average AoI: impact of varying numerology.

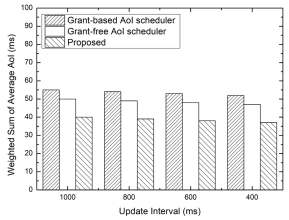


(a) $\overline{A}$ under Scenario A.

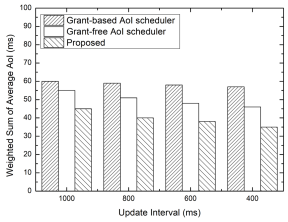


(b) $\overline{A}$ under Scenario B.

Fig. 11: Weighted sum of average AoI: impact of varying BWP size.



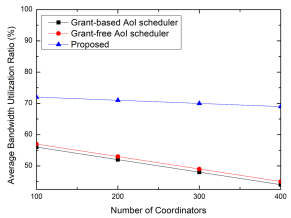
(a) $\overline{A}$ under Scenario A.



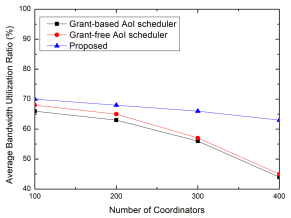
(b) $\overline{A}$ under Scenario B.

Fig. 12: Weighted sum of average AoI: impact of varying update interval.

In particular, it can be confirmed that the performance of the grant-free schedulers employing offline scheduling is superior in Scenario A where channel conditions rarely change. Specifically, the proposed scheme shows the best performance among the comparable schedulers. This is because the waiting time of the sampled data is significantly minimized upon reserving uplink PRBs based on the timing at which the

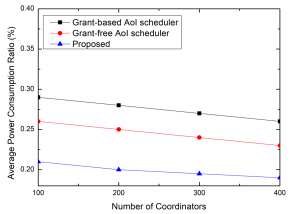


(a) Bandwidth utilization ratio under Scenario A.

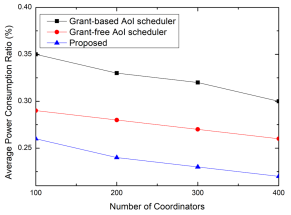


(b) Bandwidth utilization ratio under Scenario B.

Fig. 13: Average bandwidth utilization ratio: impact of varying number of coordinators.



(a) Power consumption ratio under Scenario A.



(b) Power consumption ratio under Scenario B.

Fig. 14: Average power consumption ratio: impact of varying number of coordinators.

sampled data arrive at the coordinator in the two-hop wireless system. Additionally, the proposed PRB allocation scheme preferentially allocates uplink PRBs to higher-priority nodes by increasing the relative importance of QCI in the pairwise comparison matrix (X_A). It also automatically adjusts the order of PRB assignments based on SI values over time to avoid starvation of lower-priority nodes.

However, it can be seen that the performance difference is not large in Scenario B where the channel conditions change frequently. This is because grant-free scheduling makes it difficult to change the results of offline scheduling at runtime. Hence, there is an obvious limitation in that the scheduling results cannot be adjusted in response to the time-varying channel conditions. However, the proposed PRB allocation scheme can minimize the weighted average AoI by providing a flexible trade-off between the decision factors that affect the AoI at runtime. Specifically, in Scenario B, the proposed scheme can skip the PRB allocation for the source node with poor link quality by increasing the relative importance of CQI in the pairwise comparison matrix (X_B). Consequently, the proposed scheme exhibits the best performance by minimizing the AoI, which increases owing to retransmission.

One common problem with existing AoI schedulers is that they assume that each source node has the same sampling period. This assumption significantly affects the average bandwidth utilization ratio because timeslots can be allocated to source nodes that do not have data to send in the buffer. As shown in Figure 13 (a) and (b), the proposed scheme showed the best performance because it allocates timeslots based on the sampling period of each source. However, as the number of nodes increased, the bandwidth utilization ratio of the other schemes gradually increased because the number of unnecessary timeslots allocated to each source decreased.

Figure 14 shows the impact of varying the number of coordinators on the average power consumption ratio. As shown in Figure 14 (a), the proposed method exhibited the best performance among comparable schedulers because it allocates timeslots based on the sampling interval of each source, which reduces energy wastage due to idle listening. In addition, the duty cycle of the source nodes sharing a single channel gradually decreases as the number of coordinators increases, thereby reducing the overall power consumption. In Scenario B (see Figure 14 (b)), the proposed scheme also showed the lowest power consumption ratio because it passes the PRB allocation for a source node with poor link quality, thereby reducing the probability of retransmission.

2) Performance Comparison Varying Decision Variables: In this subsection, we capture the performance variations of comparable schedulers according to the x-axis variables that affect $\overline{A}$. Figure 9 shows the variation in performance as the number of coordinators increases. If the number of coordinators increases in a situation where the number of available PRBs is limited, 5G schedulers determine the priority of PRB allocation in the frequency domain for each TTI. To minimize the weighted average AoI, comparable schedulers allocate uplink PRBs based on the priorities of the source nodes. Consequently, the performance degradation owing to starvation (Figure 9 (a)) and retransmission (Figure 9 (b)) is large. By contrast, the proposed scheme shows the best performance because it adaptively adjusts the order of PRB assignments based on the pairwise comparison matrix.

As the value of numerology increases, the TTI decreases. Thus, comparable schedulers can minimize the waiting time of the sampled data (Figure 10). However, the PRB size also increases as TTI decreases; that is, the number of available PRBs per TTI decreases when the available BWP size is small. As a result, even if the TTI decreases, grant-free scheduling that requires preconfiguration of uplink PRBs has difficulty obtaining significant benefits. In this regard, it can be confirmed that the performance of grant-free schedulers improves as the BWP size increases (Figure 11) because the number of available PRBs increases for each TTI. As shown in Figure 12, it can be confirmed that the performance of the proposed scheme gradually improves as the update period of the scheduling policy becomes shorter. This is because the proposed scheme adaptively changes the resource allocation policy in the time and frequency domains in response to changes in the network conditions within a short period via a heuristic algorithm. By contrast, comparable schedulers offer no significant benefits even if the update interval becomes shorter.

VIII. Conclusion and Future Work

In this study, we have developed a grant-free URLLC to minimize the AoI of low-power sensor devices. To enable grant-free URLLC in two-hop system models such as 5G health monitoring systems, we have proposed two adaptive solutions, including joint scheduling and adaptive PRB allocation based on the O-RAN architecture. The results of an experimental simulation demonstrate that the joint scheduling

scheme can efficiently minimize the waiting time of the sampled data and that the PRB allocation scheme can adaptively adjust the results of joint scheduling according to changes in network conditions. However, the joint scheduling scheme requires strict time synchronization between the coordinator and BS. In a future study, we plan to develop a synchronization algorithm to minimize timing errors. We also plan to develop an AoI scheduler that jointly considers AoI minimization and other constraints, such as required bandwidth and latency.

ACKNOWLEDGMENTS

This work was supported by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean Government (MSIT) (No. 2022-0-01200, Training Key Talents in Industrial Convergence Security) and Zayed University Research Office.

REFERENCES

[1] T. Padmashree and S. S. Nayak, "5g technology for e-health," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2020, pp. 211–216.

[2] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks*. IEEE, 2011, pp. 350–358.

[3] N. Peng, Y. Lin, Y. Zhang, and J. Li, "Aoi-aware joint spectrum and power allocation for internet of vehicles: A trust region policy optimization-based approach," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 19 916–19 927, 2022.

[4] H. Wang, X. Xie, and J. Yang, "Optimizing average age of information in industrial iot systems under delay constraint," *IEEE Transactions on Industrial Informatics*, 2023.

[5] V. Tripathi and E. Modiano, "Optimizing age of information with correlated sources," in *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2022, pp. 41–50.

[6] M. Yi, X. Wang, J. Liu, Y. Zhang, and B. Bai, "Deep reinforcement learning for fresh data collection in uav-assisted iot networks," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2020, pp. 716–721.

[7] W. Lyu, Y. Xiu, S. Yang, P. L. Yeoh, Y. Li, and Z. Zhang, "Weighted sum age of information minimization in wireless networks with aerial irs," *IEEE Transactions on Vehicular Technology*, 2022.

[8] Q. He, D. Yuan, and A. Ephremides, "On optimal link scheduling with min-max peak age of information in wireless systems," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–7.

[9] Q. Chen, Z. Cai, L. Cheng, F. Wang, and H. Gao, "Joint near-optimal age-based data transmission and energy replenishment scheduling at wireless-powered network edge," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 770–779.

[10] Z. Fang, J. Wang, C. Jiang, X. Wang, and Y. Ren, "Average peak age of information in underwater information collection with sleep-scheduling," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 10 132–10 136, 2022.

[11] J. Zhu and J. Gong, "Online scheduling of transmission and processing for aoi minimization with edge computing," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2022, pp. 1–6.

[12] Q. Liu, C. Li, Y. T. Hou, W. Lou, J. H. Reed, and S. Kompella, "Ao 2 i: Minimizing age of outdated information to improve freshness in data collection," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 1359–1368.

[13] N. Lu, B. Ji, and B. Li, "Age-based scheduling: Improving data freshness for wireless real-time traffic," in *Proceedings of the eighteenth ACM international symposium on mobile ad hoc networking and computing*, 2018, pp. 191–200.

[14] C. Li, S. Li, Y. Chen, Y. T. Hou, and W. Lou, "Aoi scheduling with maximum thresholds," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 436–445.

[15] J. Li, Y. Zhou, and H. Chen, "On the age of information for multicast transmission with hard deadlines in iot systems," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2020, pp. 1–6.

[16] D. Sinha and R. Roy, "Deadline-aware scheduling for maximizing information freshness in industrial cyber-physical system," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 381–393, 2020.

[17] Q. Liu, C. Li, Y. T. Hou, W. Lou, and S. Kompella, "Aion: A bandwidth optimized scheduler with aoi guarantee," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–10.

[18] C. Li, Q. Liu, S. Li, Y. Chen, Y. T. Hou, W. Lou, and S. Kompella, "Scheduling with age of information guarantee," *IEEE/ACM Transactions on Networking*, vol. 30, no. 5, pp. 2046–2059, 2022.

[19] C. Li, Q. Liu, S. Li, Y. Chen, Y. T. Hou, and W. Lou, "On scheduling with aoi violation tolerance," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–9.

[20] C. Li, Y. Huang, Y. Chen, B. Jalaian, Y. T. Hou, and W. Lou, "Kronos: A 5g scheduler for aoi minimization under dynamic channel conditions," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 1466–1475.

[21] C. Li, Y. Huang, S. Li, Y. Chen, B. A. Jalaian, Y. T. Hou, W. Lou, J. H. Reed, and S. Kompella, "Minimizing aoi in a 5g-based iot network under varying channel conditions," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14 543–14 558, 2021.

[22] J. Pan, A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing age of information via scheduling over heterogeneous channels," in *Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2021, pp. 111–120.

[23] C.-C. Wu, P. Popovski, Z.-H. Tan, and Č. Stefanović, "Design of aoi-aware 5g uplink scheduler using reinforcement learning," in *2021 IEEE 4th 5G World Forum (5GWF)*. IEEE, 2021, pp. 176–181.

[24] C. Li, Q. Liu, Y. T. Hou, W. Lou, and S. Kornpella, "Aequitas: A uniformly fair 5g scheduler for minimizing outdated information," in *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 2022, pp. 180–187.

[25] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5g: Ran, core network and caching solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3098–3130, 2018.

[26] H. Zhang, Y. Kang, Z. Han, and H. V. Poor, "Aoi minimization for grant-free massive access with short packets using mean-field games," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[27] H. Zhang, Y. Kang, L. Song, Z. Han, and H. V. Poor, "Age of information minimization for grant-free non-orthogonal massive access using mean-field games," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7806–7820, 2021.

[28] S. Saha, V. B. Sukumaran, and C. R. Murthy, "On the minimum average age of information in irsa for grant-free mmtc," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1441–1455, 2021.

[29] B. Yu and Y. Cai, "Age of information in grant-free random access with massive mimo," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1429–1433, 2021.

[30] Y. Huang, J. Jiao, Y. Wang, S. Wu, R. Lu, and Q. Zhang, "G-sc-irsa: Graph-based spatially coupled irsa for age-critical grant-free massive access," *IEEE Internet of Things Journal*, 2022.

[31] A. Pratap, R. Misra, and S. K. Das, "Maximizing fairness for resource allocation in heterogeneous 5g networks," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 603–619, 2019.

[32] A. Karimi, K. I. Pedersen, N. H. Mahmood, G. Pocovi, and P. Mogensen, "Efficient low complexity packet scheduling algorithm for mixed urllc and embb traffic in 5g," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*. IEEE, 2019, pp. 1–6.

[33] M. Setayesh, S. Bahrami, and V. W. Wong, "Joint prb and power allocation for slicing embb and urllc services in 5g c-ran," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[34] W. Azariah, F. A. Bimo, C.-W. Lin, R.-G. Cheng, R. Jana, and N. Nikaein, "A survey on open radio access networks: Challenges, research directions, and open source approaches," *arXiv preprint arXiv:2208.09125*, 2022.

[35] M. Aruldoss, T. M. Lakshmi, and V. P. Venkatesan, "A survey on multi criteria decision making methods and its applications," *American Journal of Information Systems*, vol. 1, no. 1, pp. 31–43, 2013.

[36] Q. Liu, C. Li, Y. T. Hou, W. Lou, J. H. Reed, and S. Kompella, "Aion: A bandwidth conserving scheduler with data freshness guarantee," *IEEE Transactions on Mobile Computing*, 2022.

[37] Z. Shi, H. Wang, Y. Fu, X. Ye, G. Yang, and S. Ma, "Outage performance and aoi minimization of harq-ir-ris aided iot networks," *IEEE Transactions on Communications*, 2023.

[38] Q. Chen, Z. Cai, L. Cheng, F. Wang, and H. Gao, "Joint near-optimal age-based data transmission and energy replenishment scheduling at wireless-powered network edge," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 770–779.

[39] "Ieee standards for local and metropolitan area networks–part 15.6: Wireless body area networks," *IEEE Standard 802.15.6-2012*, pp. 1–271, 2012.

[40] M. Andrews, "Probabilistic end-to-end delay bounds for earliest deadline first scheduling," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064)*, vol. 2. IEEE, 2000, pp. 603–612.

[41] J. Lehoczky, L. Sha, and Y. Ding, "The rate monotonic scheduling algorithm: Exact characterization and average case behavior," in *RTSS*, vol. 89, 1989, pp. 166–171.

[42] B.-S. Kim, B. Shah, and K.-I. Kim, "Adaptive scheduling and power control for multi-objective optimization in ieee 802.15. 6 based personalized wireless body area networks," *IEEE Transactions on Mobile Computing*, 2022.

[43] O. S. Vaidya and S. Kumar, "Analytic hierarchy process: An overview of applications," *European Journal of operational research*, vol. 169, no. 1, pp. 1–29, 2006.

[44] A. Afshari, M. Mojahed, and R. M. Yusuff, "Simple additive weighting approach to personnel selection problem," *International journal of innovation, management and technology*, vol. 1, no. 5, p. 511, 2010.

[45] A. Lacava, M. Polese, R. Sivaraj, R. Soundrarajan, B. S. Bhati, T. Singh, T. Zugno, F. Cuomo, and T. Melodia, "Programmable and customized intelligence for traffic steering in 5g networks using open ran architectures," *arXiv preprint arXiv:2209.14171*, 2022.

[46] B.-S. Kim, T.-E. Sung, and K.-I. Kim, "An ns-3 implementation and experimental performance analysis of ieee 802.15. 6 standard under different deployment scenarios," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, p. 4007, 2020.