IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 November 2023, accepted 23 November 2023, date of publication 28 November 2023,
date of current version 4 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3337552

RESEARCH ARTICLE

Traffic Shaping With Optical-Wireless Cooperative Control to Reduce Uplink Jitter Due to Radio Transmission

KENJI MIYAMOTO, YOSHIHITO SAKAI, TATSUYA SHIMADA, AND TOMOAKI YOSHIDA, (Member, IEEE)
NTT Access Network Service Systems Laboratories, NTT Corporation, Yokosuka, Kanagawa 239-0847, Japan
Corresponding author: Kenji Miyamoto (kenji.miyamoto@ntt.com)

ABSTRACT Beyond 5th generation (5G) and 6th generation (6G) mobile systems require not only low latency but also low jitter for deterministic services such as remote operation. Since remote operation requires high-definition video streaming in the uplink, the uplink transmission performance will become more important. However, current new radio (NR) mobile systems cause jitter in the uplink traffic of the mobile backhaul (MBH) network because of the radio transmission schemes such as downlink/uplink separation of time division duplexing (TDD) and medium access control (MAC) service data unit (SDU) concatenation to compose transport blocks. Thus, we propose MBH uplink jitter reduction techniques with optical-wireless cooperative control. Our proposed techniques execute traffic shaping at a router in the MBH network to reduce uplink jitter. The shaping rate that reduces jitter while minimizing latency increase is calculated with mobile scheduling information forwarded from a next generation node B (gNB) via an extended cooperative transport interface (eCTI). The eCTI provides an interface between gNBs and network equipment to achieve optical-wireless cooperative control. In this paper, we experimentally validate our proposed jitter reduction techniques through a combination of simulations for radio transmission and experiments with a shaper prototype we developed. The results regarding end-to-end (E2E) jitter performance indicate that the proposed techniques reduce the maximum jitter by 88%, from 1.1 ms to 137 μs.

INDEX TERMS Deterministic network, jitter, mobile backhaul, radio access network, traffic shaping.

I. INTRODUCTION

The development of beyond 5th generation (5G) and 6th generation (6G) mobile systems will have extreme requirements [1], [2]. These requirements will include not only low latency but also low jitter for deterministic services. The currently discussed jitter requirement is less than 1 ms [3], [4]. We expect that the end to end (E2E) uplink performance will become more important in use cases, which require real-time video streaming with wireless communications, such as remote operation, vehicle to everything (V2X), and extended reality (XR) [5], [6], [7]. For example, remote infrastructure inspection and repair using drones requires high-definition video streaming in the uplink. In such a use case, high jitter in the network between drones and remote operators would degrade the video quality and lead to critical errors in remote operation. In this context, uplink jitter in mobile systems has often been focused and analyzed in previous studies [8], [9], [10]. Since the jitter in the network is not known when deploying video application server, the receiver buffer in the video application server is usually set to a large volume to absorb jitter in worst cases. This is not cost-efficient, and the required buffer size for worst-case jitter will be larger when high-definition video streaming with a large data volume is required [11], [12]. By proactively reducing network jitter prior to video streaming data input to the video application server, it becomes possible to reduce the receiver buffer size and its deployment cost. Therefore, it is imperative to fundamentally reduce jitter in the network for safe and cost-efficient remote operation.

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

However, in the uplink transmission of the current 5G radio access network (RAN), jitter occurs in the output of user equipment (UE) and remains in the output of the next generation node B (gNB) and mobile backhaul (MBH) network behind the gNB. This is because 5G new radio (NR) transmission uses schemes such as downlink/uplink separation of time division duplexing (TDD) [13] and medium access control (MAC) service data unit (SDU) concatenation to compose transport blocks [14]. These radio transmission schemes often keep uplink data waiting until they can be transmitted. This wait time leads to non-uniform uplink data transmission intervals in the UE output, even if application traffic, such as video streaming, periodically arises with uniform transmission intervals inside the UE. Thus, these non-uniform intervals appear as jitter and degrade the video quality of the uplink transmission. Therefore, we need to address this uplink jitter due to radio transmission for ensuring E2E jitter performance from the UE to the video application server via the gNB, MBH network equipment such as the routers, and the core node such as a user plane function (UPF) in the 5G core network (5GC). The same can be said for downlink jitter, but we focus on uplink jitter in this paper because we believe that the traffic volume of downlink control signals for remote operation is very small compared with the volume of uplink video streaming traffic. Therefore, the jitter of such traffic should be easily absorbed in the UE buffer. Another cause of jitter not mentioned above is retransmission control in NR radio transmission. Although retransmission significantly increases latency and jitter, packet duplication and repetition techniques [15], [16] to enable ultra-reliable and low-latency communications (URLLC) solve this problem by avoiding retransmission itself.

Various techniques to reduce jitter have been proposed for both wired and wireless networks. For wired networks, typical technologies to reduce jitter include IEEE 802.1 time-sensitive networking (TSN) [17], which consists of core techniques such as time synchronization and scheduling in Ethernet networks. Because TSN was designed to reduce jitter due to traffic congestion in wired optical networks, it is difficult to apply it for reducing jitter due to radio transmission schemes. For wireless networks, grant-free scheduling is standardized in 5G NR as a technique to support URLLC [14]. In grant-free scheduling, the dedicated radio resources are periodically pre-assigned to UE, and the UE transmits uplink data by using them without requesting to the gNB. Although this technique prevents latency variation due to radio resource scheduling and reduces jitter, dedicated resource assignment leads to inefficient resource utilization.

In this paper, we propose MBH uplink jitter reduction techniques with optical-wireless cooperative control. These techniques use the interface called an extended cooperative transport interface (eCTI). The cooperative transport interface (CTI) is an interface between gNBs and network equipment, which was standardized in the open RAN (O-RAN) alliance, an organization defining specification for open and interoperable RAN [18]. The CTI is used for optical-wireless cooperative control between RAN and optical access systems, such as the passive optical network (PON), to achieve low-latency mobile fronthaul (MFH) transmission. Optical-wireless cooperative control by using the CTI is an important study item in O-RAN working group 4 (WG4) for open fronthaul interfaces [19]. The eCTI extends the applicability of the CTI by supporting mobile midhaul (MMH) and MBH, in addition to the current CTI support for MFH [20]. The proposed jitter reduction techniques execute traffic shaping at a router in the MBH network to reduce uplink jitter. Traffic shaping is a general bandwidth management method that buffers certain packets and limits the data rate to a constant shaping rate, which aligns non-uniform packet intervals. While traffic shaping is a general technique and effective in reducing jitter, the latency increase caused by buffering packets is inevitable. Therefore, the most important point for reducing jitter by shaping is determining the appropriate shaping rate that reduces jitter while minimizing latency increase. The proposed techniques calculate such a shaping rate by estimating the video streaming traffic rate with mobile scheduling information that is forwarded from a gNB via the eCTI. Our previous study [21] evaluated the E2E jitter performance of the proposed techniques and we confirmed that they reduced jitter by 93% with a latency penalty of less than 1 ms. However, the evaluation involved full simulation of both radio transmission and the shaping at a router, and the evaluation condition was limited to the ideal TDD configuration in which the uplink transmission opportunities were not restricted. To confirm the feasibility of the shaping with optical-wireless cooperative control, it is necessary to experimentally evaluate jitter performance using actual equipment that receives mobile scheduling information, calculates the shaping rate, and executes shaping. In terms of radio transmission, we need to clarify the impact of TDD configuration on uplink jitter. Therefore, as the extension of our previous study [21], we evaluated the E2E jitter performance of the proposed jitter reduction techniques through a combination of simulations for radio transmission and experiments with a shaper prototype we developed. For radio transmission, we varied the TDD configuration including when the uplink transmission opportunities are restricted. Our contributions in this paper are summarized as follows.

- We confirmed the feasibility of shaping rate calculation with mobile scheduling information and traffic shaping execution to MBH uplink data by implementing those functions on programmable system on chip (SoC) as a shaper prototype.
- We show the impact of TDD configuration on uplink jitter for radio transmission and clarify that the restricted uplink transmission opportunities by TDD have such a detrimental impact that the uplink jitter exceeds 1 ms.
- We experimentally demonstrate that our proposed techniques achieve the required jitter performance of less than 1 ms at a constant application traffic rate.

The rest of this paper is organized as follows. Section II discusses related works on jitter reduction techniques. Section III explains our assumed system model and E2E protocol stacks. Section IV describes our proposed uplink jitter reduction techniques. Section V presents experimental results to demonstrate the jitter reduction performance of our proposed techniques. Finally, Section VI concludes the paper.

II. RELATED WORK

Time aware shaper (TAS) [22] and frame preemption [23] of TSN techniques are candidates for providing a deterministic network. TAS provides the scheduling of high and low priority Ethernet frames in fixed-length time windows. This technique helps guarantee bounded latency for time-critical frames with higher priority. Frame preemption allows higher priority Ethernet frames to interrupt the transmission of lower priority Ethernet frames. This can be useful for time-critical applications that require low latency for high-priority frames. However, these techniques were designed to reduce the jitter caused by traffic congestion in wired networks [24], [25]. Their application to reducing jitter caused by radio transmission schemes is still challenging.

To support the deterministic service with E2E low latency and low jitter for wired and wireless networks, the integration of 5G with TSN has been standardized [26], [27]. In the system architecture, the 5G system (5GS) including RAN and 5GC acts as a virtual layer 2 (L2) Ethernet bridge to be integrated seamlessly with the TSN network. This 5GS bridge supports time synchronization and TSN functions to provide bounded delay and no traffic collisions. The integration of 5G and TSN can offer low latency and low jitter across wired and wireless networks [28], [29], but managing large-scale traffic scheduling, bandwidth allocation, and time synchronization are still challenging.

Enhanced radio resource scheduling techniques have been proposed to reduce jitter within the RAN [30], [31], [32]. These techniques reduce jitter due to radio transmission by adjusting the timing of data transmission from the UE or that of data forwarding to the upper layer after data reception at the gNB. However, they require complex algorithms in the UE, which requires simple implementation in terms of processing load and battery capacity.

Jitter reduction techniques with transport layer protocols, such as quick user datagram protocol (UDP) internet connections (QUIC) have been proposed for the approaches of the layer above internet protocol (IP) [33], [34]. By utilizing and improving QUIC's characteristic of packet loss recovery with independent stream control, these techniques minimize retransmission latency and associated jitter. They are also effective in reducing jitter due to retransmission, but it is difficult to reduce jitter due to radio transmission schemes such as TDD.

Our proposed jitter reduction techniques directly reduce jitter due to radio transmission by shaping at a router in the entrance of the MBH network, so it can be combined with TSN techniques to reduce jitter due to traffic congestion within wired networks. The proposed techniques apply optical-wireless cooperative control, but no time synchronization is required, and the required processing to reduce jitter is only shaping rate calculation and shaping execution using it. Thus, it is relatively easy to deploy. The shaping of the proposed techniques only requires the forwarding of already standardized mobile scheduling information output from the gNB via the eCTI and does not require additional overheads. Furthermore, the proposed technique is also capable of reducing jitter that cannot be addressed in the upper layer. Therefore, our proposed techniques complement all the important previous studies.

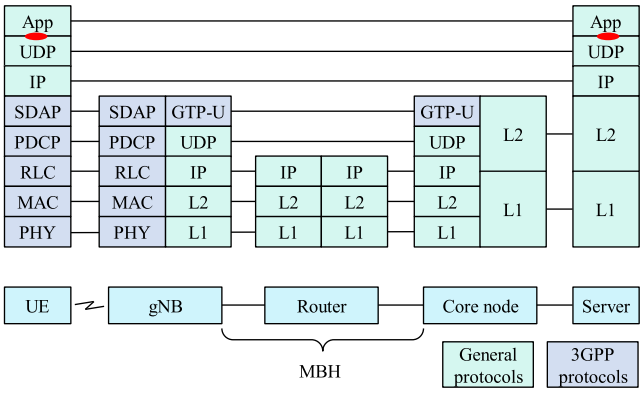


FIGURE 1. System model and protocol stacks for uplink transmission.

III. SYSTEM MODEL

Fig. 1 shows the system model and protocol stacks for uplink transmission, including a UE, gNB, router, core node, and server. For simplicity, we deal with one of each of these network elements. In uplink transmission, traffic arises at the application (App) layer inside the UE. We assume a video streaming use case and select the UDP as the transport layer protocol. The UDP and IP layer procedures are carried out first. The UE then executes the NR protocol layer procedures defined by the 3rd generation partnership project (3GPP), i.e., procedures for the service data adaptation protocol (SDAP), packet data convergence protocol (PDCP), radio link control (RLC), MAC and physical (PHY) layers [35]. After the NR protocol reception procedures, the gNB encapsulates the uplink data with the general packet radio service (GPRS) tunneling protocol user plane (GTP-U) [36]. The gNB then transmits the encapsulated data to the MBH network by using general layer 1 (L1) and L2 protocols after the second UDP and IP procedures. In the MBH network, the router forwards the uplink data to the core node via the IP, L2, and L1 protocols. The core node decapsulates and forwards the uplink data to the server after the GTP-U and UDP procedures corresponding to those at the gNB. Finally, the server receives the uplink data and executes the App layer procedure.

We define the uplink latency as the time to deliver an App layer data packet from the UDP ingress point at the UE to the UDP egress point at the server, as indicated by the red points in Fig. 1. Packet jitter is defined as the difference between

consecutive data packet latencies [37]. Let L_i be the latency of the i-th data packet, then J_i is expressed as

$$J_i = |L_i - L_{i-1}| \tag{1}$$

The average jitter J is expressed as

$$J = E[|L_i - L_{i-1}|] \tag{2}$$

Although jitter usually occurs in wired networks because of traffic congestion, it also occurs in the radio transmission of mobile systems. A major cause for this is TDD transmission, in which downlink and uplink data are transmitted alternately in the time domain. As uplink data are not transmitted during downlink transmission, waiting uplink data accumulate at the UE then transmitted as burst traffic at the next uplink transmission opportunity. Another cause for jitter in mobile systems is SDU concatenation in the UE MAC layer. The radio transmission data size is determined as a transport block size (TBS) by the MAC scheduler. If the TBS is much larger than the packet size from the upper layers, the UE gathers multiple packets to compose a transport block. The packets accumulated at the beginning in the MAC layer wait until the transport block is composed. The waiting uplink data accumulate at the UE then transmitted as burst traffic, as in the case of jitter due to TDD.

IV. UPLINK JITTER REDUCTION TECHNIQUES

Fig. 2 shows the configuration of the proposed jitter reduction techniques. The gNB is connected to the controller via the eCTI, and the router is controlled by the controller. We mainly assume that the gNB, router, and controller are located at the same location. The router can also accommodate multiple gNBs, and the controller can also be connected to multiple gNBs. As with the architecture defined in O-RAN [18], the gNB has an eCTI client, which is a process that forwards mobile scheduling information, while the controller has an eCTI server, which is a process that receives mobile scheduling information. The controller calculates the shaping rate by using the mobile scheduling information forwarded from the gNB via the eCTI. After the controller calculates the shaping rate to be set in the router, it then applies traffic shaping to the router. The router executes shaping on the uplink data output from the gNB, which reduces the packet jitter due to uplink radio transmission. In the proposed techniques, we assume the UE transmits uplink data with a constant traffic rate.

Fig. 3 shows the sequence of the shaping procedures. In NR mobile systems, user data and mobile scheduling information are transmitted at the time indicated by the numbers of radio frames, subframes, and slots. A radio frame has a length of 10 ms and consists of 10 subframes of 1-ms length. A subframe consists of multiple slots, and the slot length depends on a radio subcarrier spacing configuration called a numerology [38]. We define the numbers of the radio frames, subframes, and slots as N_{RF}, N_{SF}, and N_{SL}, respectively. When a UE starts a new uplink transmission, it sends a scheduling request (SR) message [14] to the gNB. This message requests uplink radio resources for the first

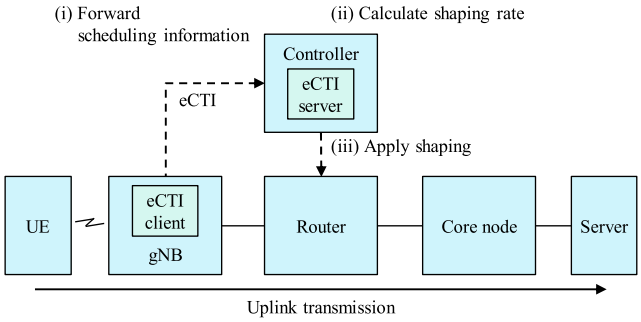


FIGURE 2. Configuration of the proposed jitter reduction techniques.

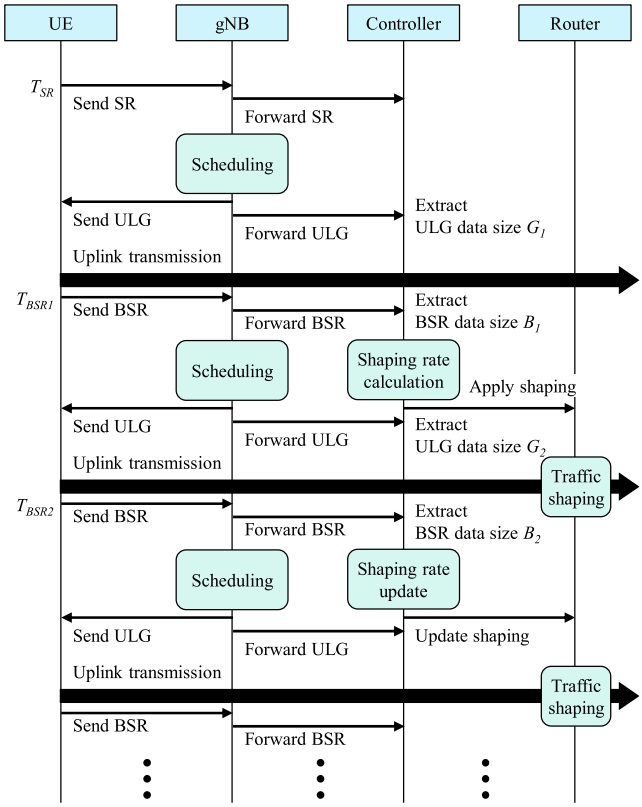


FIGURE 3. Sequence of shaping procedures.

transmission. An SR message is just a flag and does not indicate the explicit data volume to transmit. Using the eCTI, the gNB forwards the SR message to the controller, along with its radio frame number N_{RF-SR}, subframe number N_{SF-SR}, and slot number N_{SL-SR} at the time of sending the SR message. Upon receiving the SR message, the controller calculates a millisecond value of time when the SR message was sent, T_{SR} from N_{RF-SR}, N_{SF-SR}, and N_{SL-SR} as

$$T_{SR} = 10N_{RF-SR} + N_{SF-SR} + T_{SL}N_{SL-SR} \tag{3}$$

where T_{SL} is the slot length. The gNB executes MAC scheduling in reaction to receiving the SR message and sends an uplink grant (ULG) message [14] to the UE. This message allows the UE to transmit user data with an explicit data

volume. The gNB also forwards the ULG message to the controller via the eCTI.

Algorithm 1 Shaping Rate Calculation

Input: N, T_{SL}
Output: R
Receive SR message
if SR flag = true
  $R = 0$
  Extract N_{RF-SR}, N_{SF-SR}, and N_{SL-SR}
  Calculate $T_{SR} = 10N_{RF-SR} + N_{SF\_SR} + T_{SL}N_{SL-SR}$
  for $n = 1$ to N
    Receive ULG message
    Extract G_n
    Receive BSR message
    Extract $N_{RF-BSRn}$, $N_{SF-BSRn}$ and $N_{SL-BSRn}$
    Calculate $T_{BSRn} = 10N_{RF-BSRn} + N_{SF-BSRn} + T_{SL}N_{SL-BSRn}$
    Extract B_n
    Calculate $R = \frac{\sum_{n=1}^{N} G_n + B_n}{T_{BSRn} - (T_{SR} - T_{SLL})}$
  end for
end if
return R

The controller holds G_1, which is the granted uplink data size for the next uplink transmission, extracted from this first ULG message for shaping rate calculation. The UE transmits the first uplink data in accordance with the received ULG message and simultaneously sends a buffer status report (BSR) message [14] to the gNB. This message provides the gNB with information on the size of the uplink data remaining inside the UE buffer. At this time, the gNB also forwards the BSR message and its radio frame number, subframe number, and slot number at the time of sending this first BSR message, i.e., $N_{RF-BSR1}$, $N_{SF-BSR1}$, and $N_{SL-BSR1}$, respectively, to the controller via the eCTI. Upon receiving the first BSR message, the controller extracts B_1, which is the uplink data size remaining inside the UE buffer, from it. It then calculates a millisecond value of time when the BSR massage was sent, T_{BSR1} from $N_{RF-BSR1}$, $N_{SF-BSR1}$, and $N_{SL-BSR1}$ as

$$T_{BSR1} = 10N_{RF-BSR1} + N_{SF-BSR1} + T_{SL}N_{SL-BSR1} \quad (4)$$

Next, the controller calculates the shaping rate R as follows:

$$R = \frac{G_1 + B_1}{T_{BSR1} - (T_{SR} - T_{SL})} \quad (5)$$

The numerator in (5) represents the sum of the data size that has been sent and the data size that remained in the UE buffer when it sent the BSR message. The denominator represents the time interval from when traffic arose inside the UE to when the UE sent the latest BSR message. Note that the term (T_{SR} - T_{SL}) indicates that the traffic arose one slot before sending the SR message. Thus, (5) indicates the average traffic rate of the uplink data at the time when the UE sent the first BSR message. Assuming a constant traffic rate for an application, such as video streaming, the value of R in (5) is well-estimated original application traffic rate, thus is an appropriate shaping rate value to reduce jitter. This is because the jitter is eliminated by recovering packet intervals disrupted due to radio transmission to the original ideal condition output from the App layer. The most important point of the proposed techniques is that it estimates the original application traffic rate through optical-wireless cooperative control and uses it as the shaping rate. In addition, the calculation of (5) is very simple, and thus its simplicity for implementation to calculate appropriate shaping rate is an advantage of the proposed techniques. After calculating R, the controller applies shaping to the router, which reduces the jitter of the uplink data in the router output. Note that shaping is not executed for the first uplink transmission because the UE has not yet sent a BSR message. The controller does not have the UE buffer data size after receiving the SR message, thus cannot calculate R. This has little practical impact on continuous video streaming traffic because the first uplink transmission is only in preparation for the main transmission, and its data size is very small.

The controller can update R by using the second ULG and BSR messages forwarded from the gNB. Specifically, it extracts the uplink data sizes G_2 and B_2 from those messages. After receiving the radio frame number $N_{RF-BSR2}$, subframe number $N_{SF-BSR2}$, and slot number $N_{SL-BSR2}$ at the timing of the second BSR message, the controller calculates a millisecond timing value T_{BSR2} with $N_{RF-BSR2}$, $N_{SF-BSR2}$, and $N_{SL-BSR2}$ as follows:

$$T_{BSR2} = 10N_{RF-BSR2} + N_{SF-BSR2} + T_{SL}N_{SL-BSR2} \quad (6)$$

The updated shaping rate R' is then calculated as

$$R' = \frac{G_1 + G_2 + B_2}{T_{BSR2} - (T_{SR} - T_{SL})} \quad (7)$$

The data size B_1 is not used in (7), because R' calculation only requires the data size B_2 in the latest BSR message and all of the past sent data sizes, G_1 and G_2. This is because the data size, which remained in the UE buffer in the past, does not need to be counted to estimate the original application traffic rate. The shaping rate can thus be updated N times by using successive ULG and BSR messages each time of receiving the BSR message, i.e., each time of uplink transmission. Algorithm 1 details the shaping rate calculation procedures, where we use n as a variable to indicate the n-th shaping rate calculation. The input parameters that need to be determined in advance are the number of shaping rate calculations N and the slot length T_{SL}, which is determined by NR numerology. The UE stops the uplink transmission when its buffer becomes empty, but a new SR message is sent to the gNB and forwarded to the controller when the UE starts its next new uplink transmission. At that time, the controller initializes the shaping rate and repeats the shaping rate calculation procedure explained above. In terms of UE mobility and wireless environment changes, even if the

wireless throughput decreases due to them, G_1 and G_2 in (5) and (7) decrease while B_1 and B_2 in (5) and (7) increase. Therefore, the numerator in (5) and (7) becomes a constant value even when the UE moves or wireless environment changes, so there is little impact on shaping rate calculation to estimate the original application traffic rate. Since the gNB only needs to forward SR, ULG, and BSR messages via eCTI throughout the sequence in Fig. 3, no overhead specific to the proposed techniques occurs for mobile systems. Additional time synchronization is also not required because the shaping rate is calculated in relative time using N_{RF}, N_{SF}, and N_{SL}.

V. EXPERIMENTAL EVALUATION

We evaluated the E2E jitter performance of the proposed techniques through experiments involving simulations. To clarify the impact of traffic shaping on latency, we also evaluated E2E latency. We developed a shaper prototype to confirm the feasibility of the proposed techniques using actual equipment.

A. EXPERIMENTAL SETUP

Fig. 4 shows the experimental setup, which comprised a traffic generator, the shaper prototype, and a traffic analyzer. Offline simulation with the ns-3 [39] was used to emulate NR radio transmission. Because the ns-3 outputs MBH uplink data as packet capture files, we created MBH uplink data accordingly. The ns-3 also outputs the mobile scheduling information logs of SR, ULG, and BSR messages; thus, we also created mobile scheduling information on the basis of these logs. The UE, gNB, and traffic generator in Fig. 4 correspond to the UE and gNB in Fig. 2, which is the configuration of the proposed techniques. The shaper prototype, which corresponds to the controller and router in Fig. 2, used Xilinx Zynq-7000 programmable SoC widely used in communication systems and implemented the functions of shaping rate calculation and traffic shaping execution. For receiving mobile scheduling information in the shaper prototype via the eCTI, we defined three types of eCTI frames corresponding to SR, ULG, and BSR messages. The eCTI frame is a L2 Ethernet frame forwarded from gNBs and we independently specified where the mobile scheduling information was stored within the eCTI frames. The traffic generator of Spirent TestCenter output the created MBH uplink data via 10 gigabit Ethernet (GbE) and the created eCTI frames via 1 GbE. After the shaper prototype received them, it calculated the shaping rate by using the mobile scheduling information and applied shaping to the MBH uplink data. The shaping aligned non-uniform packet intervals of input MBH uplink data by buffering packets with packet intervals shorter than the interval determined by the calculated shaping rate. The traffic analyzer of Synesis Distributed received the MBH uplink data via 10 GbE and measured the latency of each packet of those data. The traffic analyzer corresponds to the core node and server in Fig.2, and we assume that these are unified and regarded as a traffic end point in the experiments. The latency was measured by calculating the difference between the latency of MBH uplink data passing through the shaper prototype and the latency of back-to-back MBH uplink data directly transmitted to the traffic analyzer. This latency was the transmission time only between the traffic generator and traffic analyzer, i.e., the latency of the wired networks. Thus, we summed the wired transmission latency measured with the traffic analyzer and the radio transmission latency that was output by ns-3 and defined this sum as the E2E latency for the experiments. The E2E jitter was then calculated from the E2E latencies with reference to (1).

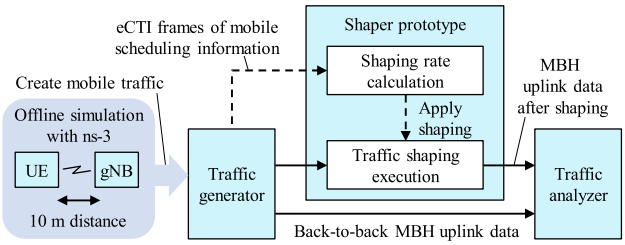


FIGURE 4. Experimental setup.

TABLE 1. Experimental parameters.

Parameters	Values
Center frequency	28 GHz
Bandwidth	400 MHz
Slot length	0.125 ms
Duplex scheme	TDD
Channel model	UMi street canyon
Number of gNB	1
Number of gNB antenna	16
Number of UE	1
Number of UE antenna	1
Distance between UE and gNB	10 m
MBH link bandwidth	10 Gbps
Application data packet size	1250 bytes
Number of transmitted packets	1000 packets

For radio transmission, we used the NR parameters reported by 3GPP [40]. Table 1 shows the experimental parameters. We used the 28-GHz center frequency, 400-MHz bandwidth, and 0.125-ms slot length. To evaluate how large the impact of jitter is due to radio transmission even with a single UE, the setup comprised one UE and one gNB. This condition is realistic if we assume that the proposed techniques are applied in private mobile networks such as remote operation in a narrow coverage area. The UE transmitted uplink data by TDD, and the channel model was urban micro (UMi) street canyon. The UE had one antenna and one layer, while the gNB had 16 antennas. The distance between the UE and gNB was fixed to 10 m, and the modulation and coding scheme (MCS) index [41] was also fixed to a maximum of 27 because of the 10-m fixed distance. The value of 10 m is the minimum distance between the UE and gNB specified in [40]. These parameters were determined to evaluate jitter under the condition with more than 1-Gbps wireless throughput assuming video streaming in the uplink with one layer. In the experiments, a wireless

environment was ideal, and we assumed no packet loss in the radio transmission. We used two configurations to clarify TDD impact on the uplink jitter. All the slot in the first TDD configuration were flexible, and we called this "flexible TDD." A flexible slot is one for which we can select downlink or uplink at the symbol level [13]. As there was no downlink user traffic in the experiments, we used almost all the symbols for uplink transmission. The second TDD configuration was called "configured TDD." Specifically, it was configured as "DDDFUUDDDD," where "D," "F," and "U" denote downlink, flexible, and uplink slots, respectively. This configuration is widely used for coexistence with legacy TDD-long term evolution (LTE). A practical TDD configuration must be subject to this configuration to avoid inter-cell interference, even if there is no downlink user traffic [42]. We used flexible TDD to evaluate the jitter impact of MAC SDU concatenation to compose transport blocks, whereas we used configured TDD to evaluate the combined jitter impact of downlink/uplink separation of TDD and MAC SDU concatenation.

For E2E transmission, we set the application data packet size to 1250 bytes and number of transmitted packets to 1000. The number of 1000 packets is enough to evaluate the jitter because we sufficiently confirmed a periodicity of transmitted data volume in the radio transmission without wireless environment changes. The intervals of generating application data packets inside the UE were varied from 100 to 10 μs, which corresponded to varying the application traffic rate from 100 to 1000 Mbps. We compared the E2E jitter for each application traffic rate with and without shaping at the shaper prototype.

B. SHAPING RATE CALCULATION

In a preliminary experiment, we evaluated the shaping rate calculation performance of the proposed techniques. Fig. 5 shows the results of calculating the shaping rate 10 times with application traffic rates of 100, 500, and 1000 Mbps for (a) flexible TDD and (b) configured TDD. The plotted points represent experimental values, while the lines represent theoretical values. The theoretical values were the original application traffic rate, i.e., appropriate shaping rate, and these values were calculated with overhead added to the 1250-byte application packets for MBH uplink data in accordance with the protocol stacks shown in Fig. 1. The average errors between the experimental and theoretical values were 2.1% in flexible TDD and 1.5% in configured TDD.

These results indicated that the shaping rate calculation in the proposed techniques could estimate the ideal shaping rate well in any cases. We also observed that there was no need to repeatedly calculate the shaping rate, as long as the application traffic rate was constant. We believe there were two reasons for the errors between the experimental and theoretical values. First, the minimum time granularity in the shaping rate calculation was the slot length, as in (5) and (7), yet the application packets arose with a period less than the slot length. Second, the BSR message indicates the uplink data size in the UE buffer as the indexes of the ranges, as defined in [14], whereas we assumed that the uplink data size given by the BSR message was maximum in the indicated range.

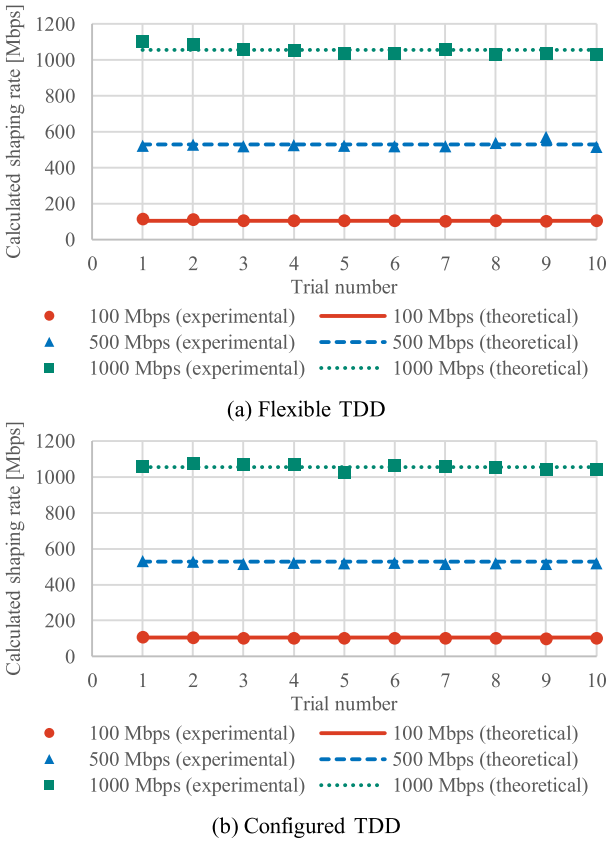
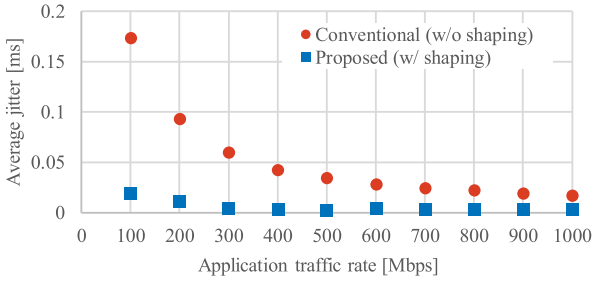


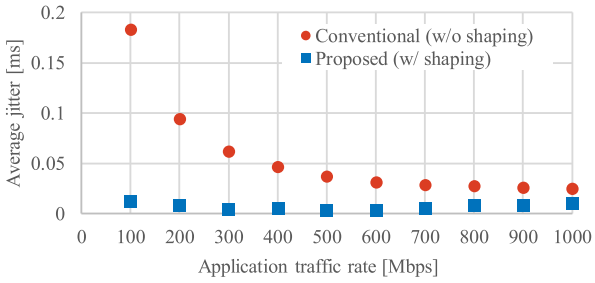
FIGURE 5. Shaping rate calculation performance.

C. E2E JITTER PERFORMANCE

Given the above results, we evaluated the E2E jitter performance with a one-time shaping rate calculation. Figs. 6 and 7 show the experimental results for the average jitter and latency for (a) flexible TDD and (b) configured TDD. As shown in Fig. 6, the average jitter decreased by shaping with the proposed techniques. The maximum jitter reduction rate was 93% for both flexible TDD and configured TDD. Fig. 7 shows that the difference in the average latency with and without shaping was less than 1 ms. The maximum latency penalties of the proposed techniques for flexible TDD and configured TDD were 660 and 669 μs, respectively. The latency for configured TDD in Fig. 7(b) increased as the application traffic rate increased, but the trend in the latency penalties remained the same. Note that the packet loss, i.e., block error rate (BLER) of each transport block was always zero for the radio transmission emulated with ns-3. This was because the received signal to noise ratio (SNR) at the gNB in the uplink transmission was always 34 dB due to 10-m short radio transmission. This sufficiently large SNR was the reason why the BLER was always zero.

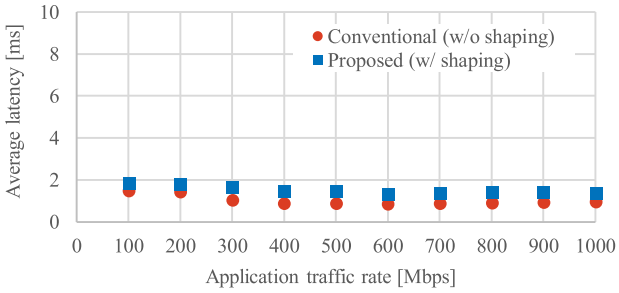


(a) Flexible TDD

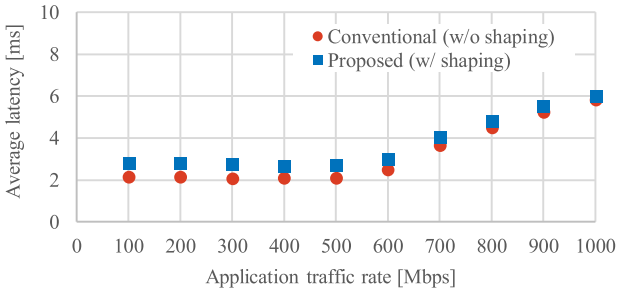


(b) Configured TDD

FIGURE 6. Average jitter performance.



(a) Flexible TDD



(b) Configured TDD

FIGURE 7. Average latency performance.

From analyzing the results in more detail, we determined three discussion points. The first discussion point is the causes of jitter in flexible TDD and configured TDD. The jitter in flexible TDD, shown in Fig. 6(a), was caused by the MAC SDU concatenation. The TBSs indicated by ULG messages were over 10000 bytes in many cases, but the packet size from the App layer was 1250 bytes. Because the TBS was much larger than 1250 bytes, the UE accumulated multiple packets until a transport block was composed, so this wait time caused jitter. Although setting the maximum transmission unit (MTU) size in the UE to a large value, such as 9000 bytes, is one of the solutions for this problem, jitter still occurs because it is necessary to concatenate at least two MAC SDUs to compose a transport block of more than 10000 bytes. The jitter in configured TDD, shown in Fig. 6(b), was caused by both the downlink/uplink separation of TDD and the MAC SDU concatenation. Due to the TDD configuration of "DDDFUUDDDD," theuplinkdataaccumulated at the UE and waited eight slots at most for the next uplink transmissionopportunity,sothis wait time also caused jitter.

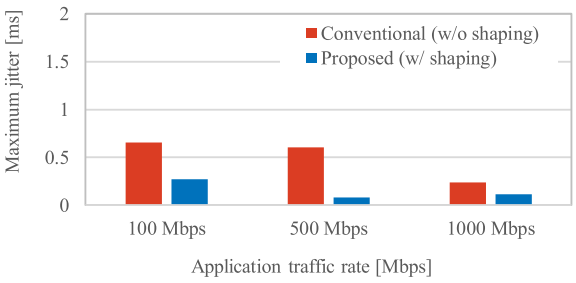
The second discussion point is why the average jitter values were large when the application traffic rates were small in both Figs. 6(a) and (b). The reason is that small transport blocks including only a few 1250-byte packets were transmitted sparsely. When packets waited a long time to be incorporated into a transport block after they arose, they caused high jitter. When the application traffic rate was large, large transport blocks including many 1250-byte packets were transmitted frequently. Thus, almost all packets were incorporated into a transport block that were allowed to be sent just after the packets arose.

The third discussion point is the increased average latency in Fig. 7(b). This was due to a bottleneck in the radio transmission throughput. Under the conditions of the NR parameters used in this study, the theoretical value of the radio transmission throughput was about 2 Gbps in flexible TDD, where almost all symbols in flexible slots were used for uplink transmission. However, the throughput decreased to about 500 Mbps in configured TDD, because the uplink transmission opportunities were restricted in accordance with the TDD configuration of "DDDFUUDDDD." Therefore, the average latency values in Fig. 7(b) increased when the application traffic rate was over 600 Mbps. The uplink data chronically accumulated in the UE for 1000 packets until the uplink transmission was allowed, and this phenomenon appeared as high latency. As shown in Fig. 6(b), the proposed techniques were effective in jitter reduction even in those situations, but the potential latency increases due to the radio transmission bottleneck slightly degraded jitter reduction performance. The same can be said for the performance when we change the radio transmission parameters. For example, if we increase the distance between the UE and gNB, it has the effect of decreasing the MCS index and theoretical value of the radio transmission throughput. It also causes a radio transmission bottleneck and latency increase, even though it is not because of the TDD configuration as shown in the experimental results.

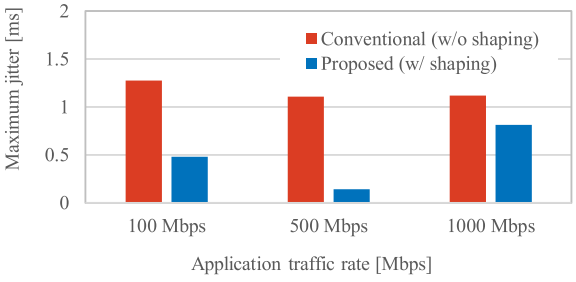
As shown in Figs. 6(a) and (b), there was little difference between the average jitter performance for flexible TDD and configured TDD. Both jitter values with and without shaping were also very small, less than 0.2 ms. This was because almost all the jitter values were very small, except for certain packets that waited a long time because of the jitter causes explained above. However, it is the maximum jitter values that degrade video quality, so it is most important to

evaluate whether the proposed techniques reduce the maximum jitter values of the worst case. Thus, to analyze the worst jitter performance and the impact of TDD on jitter in detail, we compared the maximum jitter values. Fig. 8 shows the experimental results for the maximum jitter with application traffic rates of 100, 500, and 1000 Mbps for (a) flexible TDD and (b) configured TDD. The maximum jitter values were reduced to less than 0.5 ms, except when the application traffic rate was 1000 Mbps for configured TDD, in which case the radio transmission bottleneck degraded jitter reduction performance. The maximum jitter values without shaping in configured TDD were over 1 ms, while those in flexible TDD were hundreds of microseconds. It is the downlink/uplink separation of TDD that caused the difference in the maximum jitter values. As shown in Fig. 8(b), the proposed techniques reduced the larger jitter in configured TDD. The maximum jitter was reduced by 88%, from 1.1 ms to 137 μs, when the application traffic rate was 500 Mbps. The maximum jitter when the application traffic rate was 1000 Mbps largely increased compared with the jitter when the application traffic rate was 500 Mbps. This was also due to a bottleneck in the radio transmission throughput, as mentioned in the discussion of Figs. 6(b) and 7(b). We believe that jitter would also be reduced with an application traffic rate of 1000 Mbps in configured TDD if there was no radio transmission bottleneck. This is because the jitter in flexible TDD without the radio transmission bottleneck, shown in Fig. 8(a), decreased when the application traffic rate was 1000 Mbps.

Overall, these experimental results indicated that the functions of the shaping rate calculation and shaping execution implemented on the shaper prototype worked correctly, thus they validated the proposed techniques for reducing jitter. We also clarified that the restricted uplink transmission opportunities by TDD have such a detrimental impact that the uplink jitter exceeds 1 ms, which is a future jitter requirement mentioned in Section I. The experimental results also indicated that the proposed techniques reduced the maximum jitter to less than 1 ms. We showed that the proposed techniques directly reduced jitter due to radio transmission by only executing shaping using standardized mobile scheduling information without time synchronization. Therefore, the proposed techniques can easily and efficiently reduce jitter compared with current jitter reduction techniques. As long as we execute shaping, a slight increase in latency is inevitable. However, the results of large jitter reduction with less than 1-ms average latency penalties showed that the shaping worked well to largely reduce jitter, even if the packet buffering time was short. We believe that there will be cases in which video quality degradation due to high jitter will become more serious while we have enough latency budget, e.g., remote operation via edge computing in which the distance between the UE and server is short. Hence, the proposed techniques will enable 6G use cases, such as remote operation, V2X, and XR, that require real-time video streaming and extremely low E2E jitter in the uplink. This will need to be demonstrated under more realistic conditions through experiments using testbeds in the future.



(a) Flexible TDD



(b) Configured TDD

FIGURE 8. Maximum jitter performance.

VI. CONCLUSION

We experimentally evaluated the E2E jitter performance of our proposed MBH uplink jitter reduction techniques for future mobile systems that will require low jitter. Experimental results with a shaper prototype and two TDD configurations demonstrated that the proposed techniques reduced the maximum jitter by 88%, from 1.1 ms to 137 μs, which met a future jitter requirement of less than 1 ms.

Future work will include the E2E jitter performance evaluation of multiple traffic flows when we carry out traffic shaping to the traffic flows of multiple gNBs accommodated to a router or the traffic flows of multiple UEs connected to a gNB.

REFERENCES

[1] *White Paper 5G Evolution and 6G*, NTT DOCOMO, Tokyo, Japan, Jan. 2023.
[2] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart., 2023.
[3] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. ElBakoury, "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 88–145, 1st Quart., 2019.
[4] *IOWN GF System and Technology Outlook*, IOWN Global Forum, Wakefield, U.K., Apr. 2021.
[5] A. Stornig, A. Fakhreddine, H. Hellwagner, P. Popovski, and C. Bettstetter, "Video quality and latency for UAV teleoperation over LTE: A study with ns3," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–7.
[6] D. Wang and T. Sun, "Leveraging 5G TSN in V2X communication for cloud vehicle," in *Proc. IEEE Int. Conf. Edge Comput. (EDGE)*, Oct. 2020, pp. 106–110.

[7] W. Chen, Y. Cao, Y. Qin, E. Chen, G. Zhou, and W. Li, "Real-time super-resolution: A new mechanism for XR over 5G-advanced," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2023, pp. 1–6.
[8] M. Sahu, "Delay jitter analysis for uplink traffic in LTE systems," in *Proc. 11th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2019, pp. 504–506.
[9] M. Sahu and A. A. Kherani, "End-to-end delay jitter in LTE uplink: Simple models, empirical validation & applications," in *Proc. Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2020, pp. 221–228.
[10] F. Ronteix-Jacquet, "On radio access network uplink latency and jitter : Measurements and analysis," in *Proc. 33th Int. Teletraffic Congr. (ITC-33)*, Aug. 2021, pp. 1–3.
[11] Y. Xu, Y. Chang, and Z. Liu, "Calculation and analysis of compensation buffer size in multimedia systems," *IEEE Commun. Lett.*, vol. 5, no. 8, pp. 355–357, Aug. 2001.
[12] H. Dahmouni, A. Girard, M. Ouzineb, and B. Sanso, "The impact of jitter on traffic flow optimization in communication networks," *IEEE Trans. Netw. Service Manage.*, vol. 9, no. 3, pp. 279–292, Sep. 2012.
[13] 3GPP, "Physical layer procedures for control," France, Tech. Rep. TS 38.213, V17.6.0, Jun. 2023. [Online]. Available: https://www.3gpp.org/about-us/introducing-3gpp and https://www.3gpp.org/contact-us
[14] 3GPP, "Medium access control (MAC) protocol specification," France, Tech. Rep. TS 38.321, v17.5.0, Jun. 2023. [Online]. Available: https://www.3gpp.org/about-us/introducing-3gpp and https://www.3gpp.org/contact-us
[15] J. Rao and S. Vrzic, "Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis," *IEEE Netw.*, vol. 32, no. 2, pp. 32–40, Mar. 2018.
[16] N. Zhang, H. Zhang, and J.-B. Wang, "Performance analysis of repetition-based grant-free access for URLLC," in *Proc. 7th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2021, pp. 1738–1742.
[17] *IEEE Standard for Local and Metropolitan Area Networks–Timing and Synchronization for Time-Sensitive Applications*, Standard IEEE 802.1 AS-2020, Jun. 2020.
[18] *O-RAN Cooperative Transport Interface Transport Control Plane Specification 4.0*, O-RAN-WG4.CTI-TCP.0-R003-v04.00, O-RAN Alliance, Alfter, Germany, Jun. 2023.
[19] *Control, User and Synchronization Plane Specification*, O-RAN-WG4.CUS.0-R003-v12.00, O-RAN Alliance, Alfter, Germany, Jun. 2023.
[20] *Technical Outlook for Mobile Networks Using IOWN Technology—Advanced Transport Network Technologies for Mobile Network*, IOWN Global Forum, Wakefield, U.K., Apr. 2023.
[21] K. Miyamoto, Y. Sakai, T. Shimada, and T. Yoshida, "Mobile backhaul uplink jitter reduction techniques with optical-wireless cooperative control," in *Proc. 27th OptoElectronics Commun. Conf. (OECC) Int. Conf. Photon. Switching Comput. (PSC)*, Jul. 2022, pp. 1–4.
[22] *IEEE Standard for Local and Metropolitan Area Networks – Bridges and Bridged Networks—Amendment 25: Enhancements for Scheduled Traffic*, Standard IEEE 802.1Qbv-2015, Mar. 2016.
[23] *IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks—Amendment 26: Frame Preemption*, Standard IEEE 802.1Qbu-2016, Aug. 2016.
[24] Y. Huang, S. Wang, B. Wu, T. Huang, and Y. Liu, "TACQ: Enabling zero-jitter for cyclic-queuing and forwarding in time-sensitive networks," in *Proc. ICC IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.
[25] Y. Kawakami, H. Kawata, T. Kubo, N. Yasuhara, S. Yoshihara, and T. Yoshida, "Applying time-aware shaper considering user identifier to service provider network," in *Proc. IEEE 19th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2022, pp. 505–506.
[26] 3GPP, "System architecture for the 5G system (5GS)," France, Tech. Rep., TS23.501, v18.2.2, Jul. 2023. [Online]. Available: https://www.3gpp.org/about-us/introducing-3gpp and https://www.3gpp.org/contact-us
[27] M. K. Atiq, R. Muzaffar, Ó. Seijo, I. Val, and H.-P. Bernhard, "When IEEE 802.11 and 5G meet time-sensitive networking," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 14–36, Dec. 2022.
[28] P. M. Rost and T. Kolding, "Performance of integrated 3GPP 5G and IEEE TSN networks," *IEEE Commun. Standards Mag.*, vol. 6, no. 2, pp. 51–56, Jun. 2022.
[29] Y. Zhang, Q. Xu, M. Li, C. Chen, and X. Guan, "QoS-aware mapping and scheduling for virtual network functions in industrial 5G-TSN network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.
[30] L. Kong, P. Wu, J. Chu, W. Xu, and L. Zhou, "A deterministic communication technique in the 5G-Adv/6G access network systems," in *Proc. 14th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nov. 2022, pp. 955–960.
[31] F. Ronteix-Jacquet, X. Lagrange, I. Hamchaoui, and A. Ferrieux, "Rethinking buffer status estimation to improve radio resource utilization in cellular networks," in *Proc. IEEE 95th Veh. Technol. Conference: (VTC-Spring)*, Jun. 2022, pp. 1–5.
[32] M. C. Lucas-Estañ and J. Gozalvez, "Sensing-based grant-free scheduling for ultra reliable low latency and deterministic beyond 5G networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4171–4183, Apr. 2022.
[33] P. K. Kharat, A. Rege, A. Goel, and M. Kulkarni, "QUIC protocol performance in wireless networks," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2018, pp. 0472–0476.
[34] M. Kanagarathinam, S. F. Hasan, S. Rengan, S. Singh, G. K. Choudhary, N. M. F. Qureshi, and H. Lee, "Enhanced QUIC protocol for transferring time-sensitive data," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2022, pp. 1–6.
[35] 3GPP, "NR and NG-RAN overall description; stage 2," France,Tech. Rep. TS 38.300, v17.5, Jun. 2023. [Online]. Available:https://www.3gpp.org/about-us/introducing-3gpp and https://www.3gpp. org/contact-us
[36] 3GPP, "General packet radio system (GPRS) tunneling Protocol user plane (GTPv1-U)," Tech. Rep. TS 29.281, v18.0, Jun. 2023.
[37] *IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)*, document RFC 3393, IETF, Nov. 2002.
[38] 3GPP, "Physical channels and modulation," Tech. Rep. TS 38.211, v17.5.0, Jun. 2023.
[39] *5G-LENA Project*. Accessed: Nov. 9, 2023. [Online]. Available: https://5g-lena.cttc.es/
[40] 3GPP, "Study on channel model for frequency spectrum above 6 GHz," France, Tech. Rep. TR 38.900, v15.0.0, Jun. 2018. [Online]. Available: https://www.3gpp.org/about-us/introducing-3gpp and https://www.3gpp.org/contact-us
[41] 3GPP, "Physical layer procedures for data," Tech. Rep. TS 38.214, v17.6.0, Jun. 2023.
[42] *5G TDD Uplink White Paper*, NGMN, Frankfurt, Germany, Dec. 2021.

KENJI MIYAMOTO received the B.E. degree in electronic engineering from Doshisha University, Kyoto, Japan, in 2010, and the M.E. and Ph.D. degrees in electrical, electronic and information engineering from Osaka University, Osaka, Japan, in 2012 and 2017, respectively. In 2012, he joined NTT Access Network Service Systems Laboratories, NTT Corporation, Yokosuka, Japan. His research interest includes optical and wireless converged access networks. He is a member of the Institute of Electronics, Information, and Communication Engineers (IEICE), Japan.

YOSHIHITO SAKAI received the B.E. and M.E. degrees from the Tokyo Institute of Technology, Tokyo, Japan, in 2003 and 2005, respectively. In 2005, he joined NTT Access Service Systems Laboratories, NTT Corporation, where he was engaged in research and development on next-generation optical access networks and systems. Since 2020, he has been researching optical networks for mobile systems.

TATSUYA SHIMADA received the B.E. degree in applied physics from Tohoku University, Sendai, Japan, in 1997, and the Ph.D. degree from Hokkaido University, Sapporo, Japan, in 2009. In 1997, he joined the NTT Multimedia Systems Development Center, Chiba, Japan, where he was engaged in the development of ATM optical access systems. In 1999, he moved to NTT Access Network Service Systems Laboratories, NTT Corporation, Chiba. In 2000, he worked on the WDM optical access network and systems. Since 2014, he has been working on an optical network for 5G beyond and 6G mobile systems. He is a member of the Institute of Electronics, Information, and Communication Engineers (IEICE), Japan.

TOMOAKI YOSHIDA (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 1996, 1998, and 2007, respectively. In 1998, he joined the NTT Multimedia Systems Development Center, Chiba, Japan, where he engaged in the development of ATM optical access systems. In 1999, he moved to NTT Access Network Service Systems Laboratories, NTT Corporation, Chiba. Since 2000, he has been engaged in research on next-generation optical access networks and systems. From 2013 to 2015, he was involved in research on WDM/TDM-PON and worked on the standardization of optical access systems, such as NG-PON2. He is a member of the Institute of Electronics, Information, and Communication Engineers (IEICE), Japan.