IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 October 2024, accepted 15 November 2024, date of publication 25 November 2024, date of current version 4 December 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3505921*

RESEARCH ARTICLE

# A Reliable Multicast Protocol for Streaming Services in Non-Terrestrial Networks

**ENPING ZHOU[1], (Student Member, IEEE), MAKOTO KOBAYASHI[2], (Member, IEEE), TAKUYA FUJIHASHI[1], (Member, IEEE), MD. ABDUL ALIM[3,4], (Member, IEEE), SHUNSUKE SARUWATARI[1], (Member, IEEE), MASAHIRO NISHI[2], (Member, IEEE), AND TAKASHI WATANABE[1], (Life Member, IEEE)**

[1]Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan
[2]Graduate School of Information Sciences, Hiroshima City University, Hiroshima 731-3194, Japan
[3]Department of Electronic Systems, Aalborg Universitet, 9220 Aalborg, Denmark
[4]Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

Corresponding author: Enping Zhou (zhou.enping@ist.osaka-u.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP19H01101; and in part by the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.

**ABSTRACT** With the increasing popularity of extended reality (XR) applications in sixth-generation (6G) communication systems, the demand for low-latency and reliable global communication is increasing. Non-terrestrial networks (NTNs) have been explored as potential solutions to meet these requirements. However, NTNs present unique challenges such as long propagation delays, satellite movement, and handovers, which make media access control (MAC) protocols a critical topic. This paper proposes an NTN reliable multicast (NTN-RM) protocol designed for networks comprising multiple low earth orbit (LEO) satellites, high-altitude platforms (HAPs), and user equipment (UE). NTN-RM employs multibeam cellularization to efficiently cover the satellite area by dividing it into independent cells and uses HAP overhearing along with non-orthogonal multiple access (NOMA) to minimize retransmission delays. Additionally, a queue control scheme was developed to ensure seamless NOMA retransmissions, and a beam angle adjustment and handover method was designed to effectively manage satellite handovers. Simulation results based on 3rd Generation Partnership Project (3GPP) releases demonstrate that NTN-RM reduces the mean latency by approximately 23% and decreases jitter by approximately 75.8% compared to the conventional LEO direct transmission method. Additionally, NTN-RM improves reliability by approximately 3.2–8.5% compared to recent reliable multicast protocols and exhibits robust network capacity as the number of UE increases from $10^2$ to $10^8$. These results suggest that NTN-RM is a more suitable approach for reliable multicast in NTNs. Based on the latency results, a LEO satellite density greater than $5 \times 10^{-6}$ LEO/$km^2$ is recommended for multicast applications in NTNs.

**INDEX TERMS** Non-terrestrial network, reliable multicast, multibeam, streaming service, NOMA.

## I. INTRODUCTION

The sixth-generation (6G) mobile communication system promises to revolutionize wireless communication with ultra-high-speed, ultra-low-latency, and ultra-reliable connectivity. It offers extensive coverage and supports high-density devices. Key applications of 6G include worldwide streaming services, such as extended reality (XR) applications, which demand high data rates, low latency, high reliability, and global connectivity [1].

Non-terrestrial networks (NTNs) have emerged as a promising solution in the 6G landscape for achieving low-latency and seamless global coverage. Recently, there has been growing interest in both low Earth orbit (LEO) satellites and high-altitude platforms (HAPs) from industry and academia. Companies such as SpaceX, OneWeb, Telesat, Amazon, and Iridium Next are dedicated to developing LEO satellite networks [2]. Additionally, several companies

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

have focused on the research and development of HAPs, including the Zephyr Platform by Airbus, Stratobus Platform by Thales, and Solar HAPS by HAPSMobile [3]. The relationship between LEO satellites and HAPs is complementary rather than competitive [3]. LEO satellite networks provide global coverage, whereas HAPs offer low-latency and stable services.

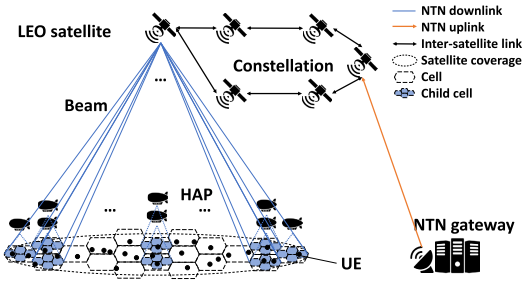



**FIGURE 1.** The overview of the non-terrestrial network.

The NTN considered in this study, illustrated in Figure 1, comprises LEO satellites at altitudes between 300 and 1,200 km, HAPs at altitudes between 20 and 50 km, and NTN gateways and user equipment (UE) on the ground. In this NTN, content from the NTN gateway can be delivered to any satellite in the constellation via inter-satellite links (ISLs). The last-hop satellite uses multiple beams to serve cells in its coverage, with HAPs to form a hierarchical space-air-ground network. The LEO constellation ensures global coverage and enables the NTN to serve any location on the Earth. Additionally, optical ISLs facilitate low-latency communication over long distances because of faster signal propagation in space compared to optical fibers, making NTNs a viable solution for worldwide XR applications.

Based on the type of XR application, the NTN delivers streaming content to different groups of cells on the ground. For instance:

- **Worldwide Augmented Reality (AR) Applications**: For applications such as live matches displayed in real-time across different regions, the NTN delivers content only to the cells that request it.
- **Worldwide Virtual Reality (VR) Applications**: For applications such as metaverse, where users interact with the virtual environment or with each other, the NTN delivers the same content to all cells, as all users should be contained in the same virtual world.
- **Worldwide Mixed Reality (MR) Applications**: For applications such as Pokémon GO, where both users and virtual items interact with the real world, the NTN delivers different content to each cell, as each cell has a different physical environment.

Notably, although XR content depends on user viewpoints, transmitting only the current view can cause a drop in quality when users change viewpoints. To address this, a common strategy is to multicast the entire scene at a lower quality while unicasting high-quality streams for the current viewpoint. This ensures that the content is pre-loaded and available without delay, highlighting the importance of multicast in XR.

In this study, we focused on latency-sensitive and reliability-sensitive XR multicast applications, such as a worldwide multiplayer XR gaming event. In such events, many users interact in a shared virtual environment, are divided into teams, and play in real-time. The game could be a large-scale battle royale or futuristic sports match, where players must react quickly to events. Additionally, millions of spectators can join XR, watching and interacting with the environment without affecting gameplay.

A typical solution to multicast content for multiple UE within a terrestrial network is to use reliable multicast protocols. However, the long distance between LEO satellites and the UE introduces significant packet loss and retransmission delays. The increase in latency and jitter can lead to playback stalls during streaming services, significantly degrading the quality of experience (QoE). Furthermore, LEO satellites may serve a substantial number of UE, making uplink feedback a critical concern. Although some studies have focused on the uplink of Internet of Things (IoT) devices in NTNs [4], [5], [6], [7], [8], limited attention has been paid to uplinks in multiuser (MU) NTNs. The primary distinction between IoT and MU lies in the frequency of uplink messages, with IoT devices typically generating sparse uplink messages, whereas MU devices transmit uplink feedback more frequently and in a more concentrated manner.

This paper proposes an NTN reliable multicast (NTN-RM) protocol aimed at delivering reliable multicast streaming services using an NTN for worldwide XR applications. The proposed protocol leverages LEO constellations and multiple HAPs to achieve a low-latency, reliable global service for multiple UE. Specifically, this approach adopts a multibeam-based hierarchical cellularization method to efficiently serve large areas covered by LEO satellites. Additionally, HAPs assist with UE feedback uplinks and allow packet repair through overhearing and non-orthogonal multiple access (NOMA) retransmission techniques. This approach also introduces LEO satellite queue control based on queue monitoring to ensure compliance with NOMA constraints. Furthermore, since LEO satellites are in constant motion around the Earth, the UE on the ground must periodically switch access points. A beam angle adjustment and handover scheme based on movement prediction and negative acknowledgment (NAK) feedback is designed to facilitate seamless handover between neighboring satellites and ensure uninterrupted services. The effectiveness of the proposed approach was evaluated through simulations employing parameters based on the 3rd Generation Partnership Project (3GPP) releases [9], [10], [11].

The key contributions of this paper are as follows.

- The NTN-RM protocol is proposed as a novel solution for providing reliable streaming multicast services in a system that includes multiple LEO satellites, HAPs, and UE within the NTN.

- The NTN-RM protocol integrates HAP overhearing and NOMA-based retransmission to minimize the latency during retransmission. It effectively employs HAPs to alleviate the uplink feedback load on LEO satellites, resulting in reduced processing requirements for LEO satellites within the NTNs.
- To ensure NOMA transmission in NTNs, this study introduces a novel LEO satellite queue control scheme that leverages the streaming feature, eliminating the need for synchronization between HAPs and LEO satellites by maintaining continuous streaming transmission from LEO satellites.
- The movement of LEO satellites and handovers between satellites were also investigated. A solution that utilizes ISL to enable LEO satellites to exchange information regarding cells located at the edge of coverage was proposed, ensuring seamless handovers in NTNs multicast.
- A system-level simulation was conducted to evaluate the performance of an NTN comprising multiple satellites, HAPs, and UE. The simulation is based on 3GPP releases and investigates the impact of various parameters on network performance.

The remainder of this paper is organized as follows. In Section II, we present related work in the areas of existing reliable multicast protocols, NOMA-based protocols for NTNs, MU-multiple-input multiple-output (MU-MIMO) networks, NTNs with LEO satellites and HAPs, and NTNs for multicast. Section III describes the system model used in this study, including the system overview, channel model, fading effects, signal transmission, and LEO satellite movement.

Section IV provides a detailed description of the proposed NTN-RM protocol, including an overview, the multibeam-based hierarchical cellularization, NAK suppression-based uplink offloading utilizing HAPs, NOMA-based local repair utilizing HAPs, LEO satellite queue control based on queue monitoring, and beam angle adjustment and handover based on movement prediction and NAK feedback mechanisms. Section V outlines the evaluation methods and indicators, simulation environment, and parameters, and presents the results and analysis.

Finally, Section VI summarizes the paper and provides the conclusions.

## II. RELATED WORK

### A. EXISTING RELIABLE MULTICAST PROTOCOLS

Existing reliable multicast protocols were not initially proposed for NTNs; therefore, they often encounter issues when applied to NTNs, or fail to take advantage of the hierarchical architecture of NTNs.

The reliable and efficient multicast protocol (REMP) was initially proposed for IEEE 802.11 wireless local area networks (WLANs) [12]. It assumes that source multicast data are sent to multiple receivers in a WLAN, and source repair is performed to ensure reliability. When applying REMP to NTNs, HAPs are not used because there is no local repair scheme. Only LEO satellite repairs are allowed in the REMP, resulting in a high retransmission delay owing to the long propagation delay between the LEO satellite and UE.

Pragmatic general multicast (PGM) is a reliable multicast protocol designed for wired networks [13]. It multicasts data from a single source to multiple receivers through hierarchical forward routers, some of which may have designated local repairs (DLRs). When packet loss is detected, repairs are generated by either the source or the DLR for retransmission. In NTNs, the LEO satellite can be regarded as the source, whereas the HAPs can be considered as a combination of forward routers and DLRs. However, HAPs may be busy relaying packets when an NAK arrives, causing queuing delays as repair packets wait in the queue until the current transmission ends.

Similar to the PGM, the NAK-oriented reliable multicast (NORM) protocol relies on the NAK mechanism [14]. The key distinction is that the NORM does not assume local repair. Consequently, the application of NORM to NTNs is expected to yield effects similar to those of REMP.

The reliable multimedia multicast (RMM) protocol, proposed in [15], was designed for video multicast applications. It takes into account the Group of Pictures (GOP) structure in modern video compression standards such as MPEG-4, which consists of three frame types: I-frames, P-frames, and B-frames. In essence, RMM limits retransmissions to I-frame packets because of their critical role in maintaining video quality. The loss of an I-frame can severely degrade the video, whereas missing P- or B-frames has less impact. Although this approach reduces latency, it can result in the loss of some packets.

The network coding-based medium access control protocol (NC-MAC) [16] enhances multicast reliability in vehicle-to-everything (V2X) networks by combining retransmission, NAK mechanism, and network coding-based retransmission. In the context of Non-Terrestrial Networks (NTNs), a LEO satellite sends each new packet twice in NC-MAC. When a UE experiences packet loss, it sends a NAK to the satellite. The satellite then generates and transmits two linear combinations of the retransmitted packets and a new packet, as described in [16]:

$$\begin{aligned} C_1 &= \alpha_{11}p_1 + \alpha_{12}p_2 + \alpha_{13}p_3 \\ C_2 &= \alpha_{21}p_1 + \alpha_{22}p_2 + \alpha_{23}p_3 \end{aligned} \tag{1}$$

where $C_1$ and $C_2$ are two linear combinations, $p_1, p_2$ represent two different retransmitted packets, and $p_3$ represents a new packet. The coefficients $\alpha_{ij} \in GF(2^k)$; $i = 1, 2, j = 1, 2, 3$. If UE receives both $C_1$ and $C_2$ and has one of the retransmitted packets, it can successfully recover the lost packet and decode the new packet $p_3$. Otherwise, if the UE is unable to decode the new packet due to missing $C_1$, $C_2$, or $p_1, p_2$, it sends a NAK for $p3$. Although this method improves reliability, it may still result in packet loss when the linear combinations cannot be decoded.

### B. NOMA-BASED PROTOCOLS FOR NTNS

Several NOMA-based media access control (MAC) protocols for NTNs have been studied [17], [18], [19].

In [17], a NOMA-based integrated terrestrial-satellite network was proposed, where the terrestrial network downlinks NOMA to users and satellites serve users not covered by the terrestrial network. Therefore, the protocol proposed in [17] cannot be applied to scenarios without a terrestrial network.

The research in [18] investigated NOMA downlinks in unmanned aerial vehicle (UAV) networks, where UAVs are considered movable devices equipped with dual-diversity receivers, and the ground station (GS) employs NOMA to transmit data. However, previous studies [17], [18] have not fully exploited the hierarchical architecture of NTNs, ignoring the signal power difference caused by altitude differences that provide natural conditions for NOMA transmission.

The HAP-to-UAV NOMA downlink protocol in [19] utilizes the altitude difference between the HAP and UAV for the NOMA downlink, but it only considers simple transmissions and does not support multicast or packet repair.

### C. MU-MIMO NETWORKS

Several studies have investigated the use of MU-MIMO in 5G networks [20], [21], [22], which can serve as a reference for NTNs with multiple UE. For instance, [21] proposed a distributed MIMO network for enhanced interference coordination and radio resource management. In [22], a distributed MIMO network improved the sum link capacity by assigning an additional fractional band to cell-edge users based on the inter-cell interference condition. The advantage of a distributed MIMO network over a collocated MIMO network suggests a need for HAPs in NTNs.

### D. NTNs WITH LEO SATELLITES AND HAPs

Some studies [3], [23], [24], [25], [26] have researched NTNs with LEO satellites and HAPs, but none of them have focused on reliable multicast in NTNs.

In [3], HAPs and LEO satellites were considered jointly to achieve massive access and data backhaul for remote area users. A common aspect between their research and ours is the investigation of the MU uplink considering both satellites and HAPs. The difference lies in their focus on uplinks for massive access and data backhaul, whereas we concentrate on streaming services for the downlink, with the uplink primarily serving as feedback. Paper [23] proposed the use of HAPs as distributed parameter servers to facilitate federated learning in NTNs. Paper [24] researches downlink throughput optimization, and papers [25], [26] focused on resource allocation in NTNs with LEO satellites and HAPs.

### E. NTNS FOR MULTICAST

Furthermore, some studies investigated the use of NTNs to provide multicast services [27], [28], [29], [30].

The research by [27] proposed the adoption of a reliable multicast protocol for satellite networks called the satellite reliable multicast transport protocol (Sat-RMTP). Sat-RMTP relies on forward error correction (FEC) instead of retransmission to ensure reliability, thereby eliminating the long retransmission delay problem in NTNs. However, the evaluations in Section V indicate that FEC alone is insufficient to satisfy the 99.99% reliability requirement for 5G applications [31].

In [28], a dynamic multicast/broadcast single frequency network (MBSFN) beam area formation (D-MBAF) algorithm was proposed that dynamically groups beams into dedicated MBSFN beam areas (MBAs) to deliver video content to all interested NTN terminals in a multibeam geostationary earth orbit (GEO) satellite NTN system.

In [29], quality-of-experience (QoE)-assured live streaming via satellite backhaul in 5G networks was investigated. These two studies also used multibeams to support multicast in NTNs, but they assumed a GEO satellite as the transmitter and did not consider other devices, such as LEO satellites and HAPs in NTNs.

The authors previously discussed NOMA retransmission performed by HAPs in NTNs [30], focusing on a system with one satellite, one HAP, and multiple UE. The current study significantly extends that work to a more complex system involving multiple satellites, HAPs, and UE, introduces novel approaches, and provides a more comprehensive evaluation.

## III. SYSTEM MODEL

### A. OVERVIEW

This study considered a cellular NTN consisting of multiple LEO satellites, HAPs, and UE, as shown in Figure 1. The LEO satellites are deployed at an altitude of 600 km, and the HAPs are deployed at an altitude of 20 km, according to the standards in [9] and [10]. The elevation angle range of the LEO satellites was set to 30°–90°.

A seamless global LEO satellite network is assumed, meaning that when one satellite moves out of service, another satellite in the same orbit takes over to maintain continuous coverage. The coverage area of the LEO satellite was calculated based on its altitude and elevation. LEO satellites are assumed capable of transmitting multiple beams for multicast transmission. The footprint diameter of one LEO satellite beam was calculated to be 50 km using the 3 dB beamwidth obtained from [9]. The cells are assumed to have the same area as the beam footprints, with each cell served by a single LEO satellite beam. With the LEO coverage area $S_{LEO}$ and beam footprint area $S_{beam}$, the required number of beams in this study was calculated as $S_{LEO}/S_{beam}$. Therefore, each LEO satellite generates up to 1,171 beams to serve these cells, which is practical as mentioned in [32].

The UE is distributed within the LEO satellite coverage area following either a uniform or normal distribution. The HAP deployment strategy prioritizes cells with a higher concentration of UE. Each HAP is assumed to serve one cell and can form multiple beams with a footprint diameter

of 5 km for each beam, as suggested in [9]. Consequently, each HAP requires 100 beams to cover the entire cell, which is achievable in reality as demonstrated in a previous study [33]. Therefore, a cell contains 100 child cells if it is served by a HAP. A detailed description of this hierarchical cellularization structure is provided in Section IV-B.

### B. CHANNEL MODEL

#### 1) SIMPLE CHANNEL

A simple channel is one that only needs to consider the transmit power, transmit antenna gain, receive antenna gain, path loss, fading, and noise. Examples include the LEO-to-UE direct downlink channel and LEO-to-HAP direct downlink channel.

Let the transmission power be denoted by $P_t$. The effective isotropic radiated power (EIRP) is calculated as follows:

$$\mathrm{EIRP[dBW]} = P_t[\mathrm{dB}] + G_t[\mathrm{dBi}] \tag{2}$$

where $G_t$ is the gain of the transmission antenna. The antenna gain-to-noise-temperature is given by

$$\begin{aligned}&\frac{G}{T}\,[\mathrm{dB/K}]\\&= G_r[\mathrm{dBi}] - N_f[\mathrm{dB}]\\&\quad - 10\lg\left(T_0[\mathrm{K}] + (T_a[\mathrm{K}] - T_0[\mathrm{K}])\,10^{-0.1\,N_f[\mathrm{dB}]}\right)\end{aligned} \tag{3}$$

where $G_r$ is the gain of the receiver antenna, $N_f$ is the noise figure, $T_0$ is the ambient temperature, and $T_a$ is the antenna temperature. The free-space path loss can be calculated as

$$PL_{FS}[\mathrm{dB}] = 20\lg(d[\mathrm{m}]) + 20\lg(f[\mathrm{Hz}]) - 147.55 \tag{4}$$

where $d$ is the distance between the sender and receiver and $f$ is the carrier frequency of the signal. The carrier-to-noise ratio (CNR) can be obtained using the formula in the 3GPP release17 [9]:

$$\begin{aligned}\mathrm{CNR[dB]} = \mathrm{EIRP[dBW]} &+ \frac{G}{T}\,[\mathrm{dB/K}] - k\,[\mathrm{dBW/K/Hz}]\\&- PL_{FS}[\mathrm{dB}] - PL_A[\mathrm{dB}] - PL_{SM}[\mathrm{dB}]\\&- PL_{SL}[\mathrm{dB}] - PL_{AD}[\mathrm{dB}] - B[\mathrm{dBHz}]\end{aligned} \tag{5}$$

where $k$ is the Boltzmann constant, $PL_{FS}$ denotes the free-space path loss, $PL_A$ denotes the atmospheric path loss, $PL_{SM}$ denotes the shadowing margin, $PL_{SL}$ denotes the scintillation loss, $PL_{AD}$ is the additional loss, and $B$ is the channel bandwidth. It is worth noting that because we do not know the specific fading before signal transmission, we used the shadowing margin $PL_{SM}$ to reserve a reasonable CNR in advance.

#### 2) NOMA DOWNLINK CHANNEL

In the proposed approach, HAPs perform NOMA to retransmit packets simultaneously with LEO satellite transmissions when packet repair is required. This implies that HAPs and LEO satellites use the same frequency and transmit data simultaneously. However, the HAP and LEO satellite signals have different received powers owing to the difference in the downlink distance. The LEO signal is attenuated over a much longer distance and arrives at the UE with a significantly lower power than the HAP signal. To decode the signals from the HAPs and LEO satellites, the UE performs successive interference cancellation (SIC) to distinguish the signals using the power difference. Specifically, the UE first decodes the HAP signal, considering the LEO signal as interference, because the HAP signal has greater power. It then subtracts the decoded HAP signal from the superimposed signal and decodes the LEO signal. Therefore, the LEO downlink channel can be expressed using (5), whereas the NOMA downlink channel cannot be calculated using (5).

The received power of the HAP signal $P_{rH}$ or the received power of the LEO signal $P_{rL}$ can be calculated as follows:

$$\begin{aligned}&P_r[\mathrm{dBW}]\\&= \mathrm{EIRP[dBW]} + G_r[\mathrm{dBi}] - PL_{FS}[\mathrm{dB}]\\&\quad - PL_A[\mathrm{dB}] - PL_{SM}[\mathrm{dB}] - PL_{SL}[dB] - PL_{AD}[\mathrm{dB}]\end{aligned} \tag{6}$$

The noise at UE can be calculated as

$$\begin{aligned}&N[\mathrm{dBW}]\\&= B[\mathrm{dBHz}] + k[\mathrm{dBW/K/Hz}] + N_f[\mathrm{dB}]\\&\quad + 10\lg\left(T_0[\mathrm{K}] + (T_a[\,\mathrm{K}] - T_0[\mathrm{K}])\,10^{-0.1\,N_f[\mathrm{dB}]}\right)\end{aligned} \tag{7}$$

The CNR of HAP-to-UE NOMA downlink channel is given by

$$\mathrm{CNR[dB]} = P_{rH}[\mathrm{dBW}] - 10\lg\left(10^{\frac{N[\mathrm{dBW}]}{10}} + 10^{\frac{P_{rL}[\mathrm{dBW}]}{10}}\right) \tag{8}$$

#### 3) OVERHEARING CHANNEL

In NTN-RM, the HAP obtains data packets by overhearing LEO signals. The transmit power of the LEO-to-UE signal is significantly greater than that of the LEO-to-HAP signal because the UE is at a longer distance from the LEO and has a smaller receiver gain. Equation (5) can still be applied to the overhearing channel; however, the values should be changed accordingly. Therefore, for the overhearing channel, the CNR can be calculated as

$$\begin{aligned}\mathrm{CNR[dB]} = \mathrm{EIRP_U[dBW]} &+ \left[\frac{G}{T}\right]_H\,[\mathrm{dB/K}] - B_U[\mathrm{dBHz}]\\&- k[\mathrm{dBW/K/Hz}] - PL_{FS_H}[\mathrm{dB}] - PL_{A_H}[\mathrm{dB}]\\&- PL_{SM_H}[\mathrm{dB}] - PL_{SL_H}[\mathrm{dB}] - PL_{AD_H}[\mathrm{dB}]\end{aligned} \tag{9}$$

where $\mathrm{EIRP_U}$ and $B_U$ of the overhearing channel are the EIRP and bandwidth of the LEO-to-UE simple channel, respectively. However, $\frac{G}{T}$ and various path losses should be calculated using the LEO-to-HAP channel information.

#### 4) MODULATION

The default modulation scheme employed in the simulations was QPSK, as described in [34]. The BER in QPSK

modulation is defined by [35]:

$$\text{BER} = \frac{1}{2} \times \text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{10}$$

where $E_b$ and $N_0$ are the energy per bit and the noise power density, respectively, and the function erfc() is defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \tag{11}$$

In addition, $\frac{E_b}{N_0}$ can be calculated by CNR based on the following formula:

$$\text{CNR[dB]} = 10\log_{10}\left(\frac{E_b}{N_0}\right) + 10\log_{10}\left(\frac{R}{B}\right) \tag{12}$$

where R is the data rate and B is the channel bandwidth.

### C. SHADOW FADING

This study considered shadow fading and clutter loss in a large-scale model. The line-of-sight (LOS) probability for these scenarios, dependent on the elevation angle, was obtained from Table 1 [10]. The smallest elevation angle considered in this study was 30°, corresponding to the furthest cell with an elevation angle of 30° from the LEO satellite. Consequently, cells in the coverage area of a LEO satellite may have elevation angles between 30° and 90°.

**TABLE 1. LOS probability.**

| Elevation | Dense urban scenario | Urban scenario | Suburban and Rural scenario |
|---|---|---|---|
| 30° | 39.8% | 49.3% | 91.9% |
| 40° | 46.8% | 61.3% | 92.9% |
| 50° | 53.7% | 72.6% | 93.5% |
| 60° | 61.2% | 80.5% | 94.0% |
| 70° | 73.8% | 91.9% | 94.9% |
| 80° | 82.0% | 96.8% | 95.2% |
| 90° | 98.1% | 99.2% | 99.8% |

**TABLE 2. Shadow fading [dB] and clutter loss (CL) [dB].**

| Elevation | Dense urban scenario | | | Urban scenario | | | Suburban and Rural scenario | | |
|---|---|---|---|---|---|---|---|---|---|
| | LOS | NLOS | | LOS | NLOS | | LOS | NLOS | |
| | $\sigma_{SF}$ | | $CL$ | $\sigma_{SF}$ | | $CL$ | $\sigma_{SF}$ | | $CL$ |
| 30° | 2.9 | 12.4 | 29.0 | 4 | 6 | 29.0 | 1.14 | 8.78 | 18.42 |
| 40° | 3.0 | 11.7 | 27.7 | 4 | 6 | 27.7 | 0.92 | 10.25 | 18.28 |
| 50° | 3.1 | 10.6 | 26.8 | 4 | 6 | 26.8 | 1.42 | 10.56 | 18.63 |
| 60° | 2.7 | 10.5 | 26.2 | 4 | 6 | 26.2 | 1.56 | 10.74 | 17.68 |
| 70° | 2.5 | 10.1 | 25.8 | 4 | 6 | 25.8 | 0.85 | 10.17 | 16.50 |
| 80° | 2.3 | 9.2 | 25.5 | 4 | 6 | 25.5 | 0.72 | 11.52 | 16.30 |
| 90° | 1.2 | 9.2 | 25.5 | 4 | 6 | 25.5 | 0.72 | 11.52 | 16.30 |

When the UE is in the LOS condition, the clutter loss is negligible and should be set to 0 dB in the basic path-loss model. Shadow fading is modeled by a log-normal distribution, which, when expressed in decibel units, is a zero-mean normal distribution with a standard deviation $\sigma_{SF}^2$. The values of $\sigma_{SF}$[dB] are listed in Table 2 at the reference elevation angles for the dense urban, urban, suburban, and rural scenarios defined by 3GPP [10]. The UE should use the reference angle nearest to its true elevation angle, and select the LOS probability and shadow fading accordingly. The shadow fading follows $PL_{SF} \sim \log N(0, \sigma_{SF}^2)$. In the LOS scenario, we used real-time $PL_{SF}$ to replace $PL_{SM}$ in (5), (6), and (9) to account for the real-time shadow fading effects. In the NLOS scenario, we used real-time $PL_{SF} + CL$ to replace $PL_{SM}$ in (5), (6), and (9) to account for both real-time shadow fading and clutter loss at the UE.

### D. SIGNAL TRANSMISSION

The bit rate is limited by modulation according to Hartley's law:

$$R \leq 2B\log_2(M) \tag{13}$$

where $M$ is the number of distinct messages in one symbol and can be obtained from $M$-array modulation ($M = 2$ in BPSK and $M = 4$ in QPSK).

Assuming that the data packet size is $s_{\text{data}}$, the data packet error rate $PER_{\text{data}}$ is given as

$$PER_{\text{data}} = 1 - (1 - \text{BER})^{s_{\text{data}}} \tag{14}$$

when FEC is not considered. The adoption of the FEC is discussed in Section VI.

### E. LEO SATELLITE MOVEMENT

In our system, UE and HAPs have fixed locations. However, LEO satellites must move around the Earth owing to physical limitations, with a velocity of approximately 7.9 km/s. Therefore, the distance and channel state between a specific cell and LEO satellite change over time. To measure the influence of the LEO satellite movement, we used time-varying functions to express the distance $D$ between a cell and the LEO satellite.

$$D = \begin{cases} \sqrt{(x_0 - x_c + vt)^2 + (y_0 - y_c)^2 + (z_0 - z_c)^2}, \\ \qquad\qquad ((t - x_c v) \mod T) \leq \dfrac{T}{2}; \\ \sqrt{(x_0 - x_c - vt + vT)^2 + (y_0 - y_c)^2 + (z_0 - z_c)^2}, \\ \qquad\qquad ((t - x_c v) \mod T) > \dfrac{T}{2} \end{cases} \tag{15}$$

where $(x_0, y_0, z_0)$ is the initial location of the LEO satellite, the orbit along which the LEO satellite moves is assumed to be the $y = 0$ line, $(x_c, y_c, z_c)$ is the location of the cell, $v$ is the velocity of the LEO satellite, $t$ is time, and mod denotes the modulo operation. Additionally, $T$ is the service period of a cell when a LEO satellite moves through it, which can be calculated from the satellite speed, cell size, and minimum elevation angle.

From (15) and Figure 2, it can be observed that the distance changes periodically with time. The real-time elevation of the cell and the beam angle of the satellite can be approximately calculated by:

$$Elevation = \sin^{-1}((z_0 - z_c)/D) \tag{16}$$

$$Angle = \cos^{-1}((z_0 - z_c)/D) \tag{17}$$

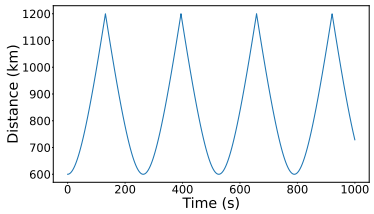



**FIGURE 2.** The distance between a cell and LEO satellites as a function of time (when the elevation is 30°).

Based on the elevation angle, the real-time BER can be obtained using (10), as listed in Table 2.

## IV. NTN-RM PROTOCOL

### A. OVERVIEW

The network architecture considered in this study comprises multiple LEO satellites, HAPs, and UE, as illustrated in Figure 1. A detailed description of the system model is provided in Section III. The primary goal of the proposed work is to achieve low-latency and reliable communication downlinks for worldwide streaming services using NTNs. Additionally, the NTN-RM protocol is designed to be capable of coexisting with unicast traffic delivery. Further details are provided in Section IV-C2.

Algorithm 1 presents an overview of the proposed NTN-RM protocol. The inputs for Algorithm 1 represent the key entities involved in the protocol. In our setup, LEO refers to a low earth orbit satellite responsible for receiving content from the source via the satellite network and delivering it to the designated service area. HAP[] denotes the set of HAPs that assist with uplink offloading and localized retransmissions. UE[] refers to the UE within the service area, while Cell[] represents the set of all cells into which the service area is divided.

Specifically, the coverage area is divided into location-fixed cells based on the beam footprint of LEO satellites (Initialize, Algorithm 1). The proposed system utilizes multibeam technology, employed by LEO satellites and HAPs, to provide services to both the cells and child cells within them (line 2, Algorithm 1). We assume that the UE and HAPs have means to access the NTN. Before transmission, the HAPs estimate the channel state to decide whether to provide services (line 3, Algorithm 1). The criteria for this judgment are discussed in the evaluation section V-C7. Once connected to the NTN, the UE generates requests and sends them to the respective NTN devices (LEO satellites or HAPs) (lines 4–11, Algorithm 1). The HAPs collect and forward requests to the corresponding LEO satellite (line 12, Algorithm 1). Upon receiving the requests, the LEO satellite retrieves the requested data from the server through the NTN and multicasts it into cells that require it (lines 15–17, Algorithm 1). The UE receives the packet and checks for bit errors. If there are errors, the UE considers the packet lost and sends a NAK to request retransmission (lines 18–20, Algorithm 1). Within each cell, a HAP-based local repair scheme was employed to reduce the retransmission delays

**Algorithm 1** NTN-RM Protocol

**Input:** LEO, HAP[], UE[], Cell[]
**Initialize:** The served area is divided into Cell[] with unique cell sequences, and all *ue* ∈ UE[] are assigned to a *cell* ∈ Cell[]. Some *cell*s are also served by *hap* ∈ HAP[]
1: **Before the Transmission**
2: LEO and HAP[] perform *Multibeam-based Hierarchical Cellularization* to serve all *cell* ∈ Cell[]
3: HAP[] estimate the channel state to decide whether to provide service
4: UE[] generate content requests
5: **for** *ue* ∈ UE[] **do**
6: **if** *ue* is served by a *hap* **then**
7: *ue* sends the request to the *hap*
8: **else**
9: *ue* sends the request to LEO
10: **end if**
11: **end for**
12: HAP-based Uplink Offloading is performed
13: LEO and HAP[] keep track of which cells require content
14: **During the Transmission**
15: **while** there are packets to be transmitted **do**
16: **for** *cell* ∈ Cell[] that requires content **do**
17: LEO transmits content directly to *cell*, performing the LEO Satellite Queue Control in Algorithm 2 and Beam Angle Adjustment scheme in Algorithm 4
18: **for** all *ue* assigned to current *cell* **do**
19: **if** the packet is lost **then**
20: *ue* sends a NAK according to the results of lines 5–11
21: **if** *cell* is served by a *hap* **then**
22: HAP-based Uplink Offloading and Local Repair is performed to handle retransmission
23: **else**
24: LEO retransmits the corresponding packet
25: **end if**
26: **else**
27: **continue**
28: **end if**
29: **end for**
30: **if** a *cell* is getting out of the coverage of LEO **then**
31: LEO performs the Handover mechanism in Algorithm 4
32: **end if**
33: **end for**
34: **end while**
35: **End**

(line 22, Algorithm 1). The LEO satellite queue control scheme was utilized to support the NOMA transmission. Finally, NTN-RM defines a beam angle adjustment and handover scheme to enable service continuity during the movement of LEO satellites and supports handovers between them (lines 30–31, Algorithm 1).

### B. MULTIBEAM-BASED HIERARCHICAL CELLULARIZATION

In this study, cells on Earth's surface were stationary and had unique sequence numbers assigned for identification purposes. The coverage area of each cell corresponds to the footprint of the LEO satellite, implying that one LEO satellite beam can serve one cell. However, this implies that LEO satellites face challenges in providing more specific services

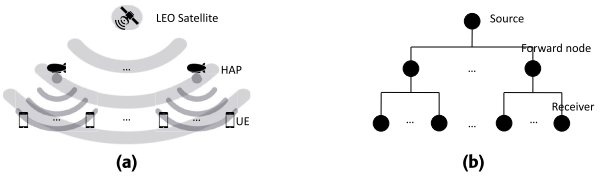



**FIGURE 3.** (a) is the network architecture of NTNs, (b) is the traditional network architecture.

based on UE requirements within a cell because of physical limitations.

In large-scale areas, such as LEO satellite cells, the UE can experience significantly different channel states. After transmission, some UE in the cell may successfully receive the packet, while others may lose it. However, a LEO beam cannot differentiate between them, and cannot provide them with different content during subsequent transmissions.

To address this issue, NTN-RM utilizes HAPs as edge servers to provide detailed services to the LEO satellite cells. Specifically, we prioritize deploying HAPs in cells with a larger number of UE, ensuring that the service area of each HAP precisely matches that of the corresponding LEO satellite cell. This enables dividing LEO cells with HAPs into child cells, each with unique local sequences assigned by the HAPs. Moreover, the basic service unit in these areas changes from a LEO satellite cell to a HAP child cell, increasing the probability of the UE receiving the required services. The beams are independent of each other, rendering the cells or child cells independent of each other. Consequently, retransmissions occurring in other cells do not affect transmissions in the current cell. The coverage areas of HAPs (cells) were assumed to be non-overlapping, and the handover mechanism described in Section IV-E ensures that only one satellite serves each cell at any given time. Consequently, interference between multiple HAPs, beams, or satellites was not considered in this study.

The HAP devices assumed in this study are either balloon-based or aircraft-based, circling above a specific region and maintaining a consistent service area after deployment. Additionally, it is worth noting that although the protocol assumes that all cells are served by a HAP, appropriate solutions (lines 5–11, Algorithm 1) have been incorporated to address cases where this condition is not met. In some scenarios, a sufficient number of HAPs may not be available to serve every cell. In such cases, UE in unserved cells sends requests directly to the LEO satellite, which then handles both transmission and retransmission.

The network architecture of a hierarchical cellular NTN differs from that of a traditional network, as shown in Figure 3. In this study, the satellites can simultaneously transmit to the HAPs and UE, and packet loss at the HAPs does not affect the UE. This feature facilitates multicast transmission in the NTNs. The transmission process is detailed in Section IV-C.

### C. HAP-BASED UPLINK OFFLOADING AND LOCAL REPAIR

#### 1) UPLINK OFFLOADING BASED ON NAK SUPPRESSION

Before transmission, the UE initiates random access (RA) to send requests to corresponding HAPs. Specifically, when a UE has a message to transmit, it waits for a random duration, determined by the anticipated number of UE requiring NAK transmission and the UE uplink transmission delay, before sending it. Once HAPs receive requests from the UE, they use the RA to send requests to the LEO satellite. Duplicate requests received by the HAPs were ignored. If every cell is served by a HAP, the number of uplink messages to the LEO satellites is limited to the number of HAPs. This limitation greatly improves the NTN's performance, as having too many uplink messages can lead to collisions with the LEO satellite, especially with a large number of UE. In addition, the transmission power of a UE is typically constrained by its handheld nature, resulting in a lower transmission success rate for direct uplinks to LEO satellites. However, HAPs have a higher transmit power than the UE, and thus enjoy a higher uplink success rate.

During transmission, there is also uplink feedback where the UE or HAPs may send NAKs to the LEO satellite when they lose a packet. Unlike the random generation of uplink requests before transmissions, during transmissions, the UE may send NAKs at similar time points because the LEO satellites multicast to all UE simultaneously. Consequently, the direct-to-LEO uplink from the UE is more susceptible to collisions during the transmission. The NTN-RM addresses this problem by allowing HAPs to receive NAKs and perform retransmissions. In each cell, the number of UE is much lower, making it easier for HAPs to successfully receive NAKs. Additionally, if all repair requests are met by HAPs, there is no need to send NAKs to the LEO satellites, thereby reducing uplink traffic to the LEO satellites. If the HAPs do not have the corresponding packets, they will forward the NAKs to the LEO. During retransmission or after sending NAKs to the LEO, HAPs will ignore duplicate NAKs from other UE. Once the retransmission is complete, HAPs will remove the corresponding NAKs from their memory.

An example is shown in Figure 4, where UE 1, 2, and 3 are served by the same HAP. Before transmission, all UE generates requests randomly and sends them to the HAP. After receiving the first request from UE 1, the HAP transmits the request to the LEO satellite and disregards identical requests. As a result, only one uplink request is transmitted to the LEO satellite, even though there are multiple UE.

In summary, the HAP-based uplink offloading scheme significantly reduces uplink traffic, alleviates the uplink processing burden on LEO satellites, and improves the uplink success rate by transmitting uplink signals at higher power levels.

#### 2) LOCAL REPAIR BASED ON NOMA

During transmission, LEO satellites multicast the requested data to cells that require it within the coverage. In each cell, when a UE loses a packet, it sends a NAK to the HAP,

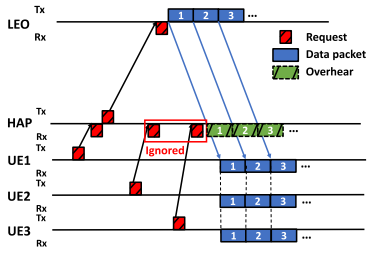



**FIGURE 4.** HAP-based uplink offloading in NTN-RM (UE 1,2, and 3 are served by the same HAP).

including a unique cell sequence number and the lost packet number. HAPs overhear satellite transmissions, store data for potential repair requirements, and use NOMA to retransmit the lost packets. Additionally, HAPs determine an appropriate storage duration for each packet based on the round-trip time (RTT) to free memory efficiently. If a NAK is received, the storage timer is reset. In particular, as shown in Figure 5, UE 1 and UE 2 are in the same LEO satellite cell, but different HAP child cells. The LEO satellite transmits the same content to all the UE. When UE 2 loses a packet, it sends a NAK to the HAP. The HAP performs NOMA to retransmit the requested packet simultaneously with the LEO satellite transmission, using the same frequency. The retransmission delay is small compared to the direct retransmission from the LEO satellite, and both the transmission and retransmission data can be concurrently received by UE 2.

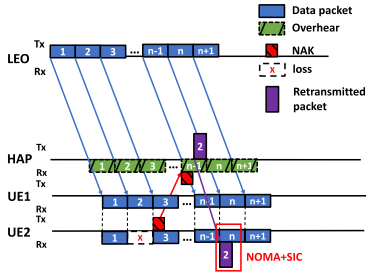



**FIGURE 5.** HAP-based local repair in NTN-RM (UE 1 and 2 are in the same cell but in different child cells).

NOMA offers significant advantages by enhancing spectrum efficiency, which directly translates into increased network capacity—a critical factor for NTNs. Without NOMA, NTNs must allocate separate frequencies for LEO and HAP downlink communication within a single application. In contrast, NOMA enables the use of a single frequency for both LEO and HAP downlink communication within the same application, freeing up frequencies for other applications and optimizing overall network efficiency. Additionally, without NOMA, HAPs would need to manage both new packet transmissions and retransmissions, significantly increasing the risk of queue congestion, particularly in high-traffic scenarios. NOMA addresses these challenges by removing the need for HAPs to relay packets from LEO to UE, thereby reducing congestion and improving overall system performance under heavy network loads.

This paragraph explains how interference between satellites and HAPs is managed. As illustrated in Figure 6, the received power of the HAP signal at the UE is significantly higher than that of the LEO signal, primarily due to the shorter downlink distance from the HAP to the UE. Consequently, the UE can perform Successive Interference Cancellation (SIC) to decode the LEO and HAP signals based on their power differential. The UE first decodes the stronger HAP signal, treating the weaker LEO signal as acceptable background interference. This approach works because the power differential ensures that the LEO signal does not significantly affect the decoding of the HAP signal. After successfully decoding the HAP signal, the UE subtracts it from the received superimposed signal, allowing for the subsequent decoding of the LEO signal without interference from the HAP. Additionally, as shown in Figure 5, only UE

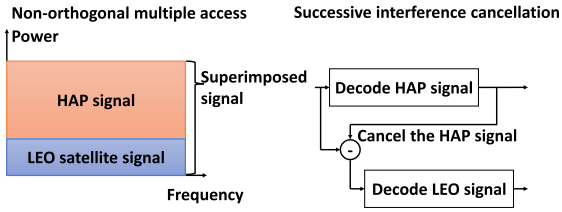



**FIGURE 6.** NOMA and SIC in NTN-RM.

2 receives the retransmitted packet. Without the HAP, all UE in this cell would receive the retransmitted packet from the LEO satellite without NOMA transmission, which would have a detrimental impact on the transmission process.

In summary, the HAP-based local repair scheme reduces the retransmission latency, alleviates the retransmission processing burden on LEO satellites, and enhances the transmission efficiency using NOMA and SIC techniques. Additionally, NTN-RM can support joint multicast and unicast transmissions via NTNs using NOMA. By utilizing different power levels for retransmission signals from HAPs and for unicast and multicast signals from LEO satellites, UE can effectively separate the signals using SIC. Moreover, recent research [36] has demonstrated that Rate-Splitting Multiple Access (RSMA) efficiently manages joint multicast and unicast traffic in LEO satellite systems. Our NTN-RM protocol could integrate RSMA for joint downlink multicast and unicast transmissions from LEO satellites, while maintaining the role of HAPs to optimize both traffic management and spectral efficiency. This can be achieved by applying NOMA between the satellite and HAPs and performing RSMA within the satellite. Under these conditions, the UE would first decode the HAP signal, which has the highest power level, and then use RSMA decoding to extract the multicast and unicast signals from the satellite.

### D. LEO SATELLITE QUEUE CONTROL BASED ON QUEUE MONITORING

To facilitate the effective implementation of NOMA in the considered system, it is crucial to employ a LEO satellite queue control algorithm. This is necessary because the server

was assumed to transmit 4K streaming to the NTN at a bit rate ranging from 20 to 60 Mbps [37], and the downlink channel capacity of the LEO satellite was calculated to be larger than this range. If the LEO satellite immediately transmits the signal at the maximum bit rate upon receiving data from the server, the packets from the LEO satellite would become intermittent. Similarly, the packets from HAPs would also be intermittent based on the NAK request of the UE, potentially leading to the asynchronous arrival of packets from both LEO satellites and HAPs at the UE. For NOMA and SIC to function properly, the UE must receive superimposed signals when the HAPs transmit signals and be able to identify whether the received signal is superimposed and when the superimposed signal concludes.

A typical approach for synchronizing LEO satellites and HAPs poses challenges, particularly when LEO satellites are in motion and the links are dynamic. To avoid this issue, the proposed approach ensures the continuity of the LEO signal by limiting the data rate of the idle beam. Because the UE consistently receives the LEO signal, it must be superimposed whenever the HAP signal arrives. A HAP signal possesses greater power because it experiences less path loss and fading. Consequently, when the UE receives a strong signal, it can be determined that it has received a superimposed signal, thereby enabling the utilization of SIC to decode the signal.

A pseudocode describing the functionality of the proposed LEO satellite queue-control algorithm is presented in Algorithm 2. This algorithm ensures the desired behavior and continuity of the LEO signal, thereby facilitating effective NOMA and SIC operations.

The algorithm takes the current_time as input and retrieves information such as Capacity (the maximum bitrate of the LEO downlink channel), Default_Bitrate (the bitrate of the 4K streaming), trans_end_time (the transmission end time of the last packet), and other packet queue details. The first step of the algorithm is to check whether the packet queue is empty at the current time. If it is empty, the algorithm returns zero and terminates (lines 1–3, Algorithm 2). However, if there are packets in the queue, it implies that the LEO satellite has data to transmit. The satellite selects the first packet in the queue for transmission and checks whether it is currently being transmitted or not. The algorithm checks whether the transmitted packet is the last packet in the queue. If other packets are still waiting, the satellite transmits the current packet at the maximum bitrate; otherwise, it transmits the packet at the same bitrate as the 4K streaming (lines 4–12, Algorithm 2). If the current packet has already been transmitted and its transmission is complete at the current time, the algorithm removes the packet and checks whether there are any other packets in the queue (lines 13–16, Algorithm 2). If additional packets are present, the previous process is repeated (lines 17–25, Algorithm 2). However, if the transmission of the current packet has not yet finished, the algorithm returns zero and proceeds with the ongoing transmission (line 28, Algorithm 2). Notably, each beam has its own queue because the transmission is independent of the beams.

**Algorithm 2** LEO Satellite Queue Control Algorithm

**Initialize:** Capacity, Default_Bitrate, trans_end_time
**Initialize:** packet_queue[], current_packet, current_bitrate
**Input:** current_time
1: satellite monitors the receive queue to estimate the source data rate and determine the Default_Bitrate
2: **if** packet_queue[] is empty **then**
3: **Output:** 0
4: **end if**
5: current_packet = packet_queue[0]
6: **if** current_packet has not been processed **then**
7: **if** the length of packet_queue[] is larger than 1 **then**
8: current_bitrate = Capacity
9: **else**
10: current_bitrate = Default_Bitrate
11: **end if**
12: trans_end_time = Processing()
13: **Output:** current_packet
14: **else if** current_packet has been processed and the trans_end_time is smaller than current_time **then**
15: Remove current_packet from packet_queue[]
16: **if** packet_queue[] is empty **then**
17: **Output:** 0
18: **else**
19: current_packet = packet_queue[0]
20: **if** the length of packet_queue[] is larger than 1 **then**
21: current_bitrate = Capacity
22: **else**
23: current_bitrate = Default_Bitrate
24: **end if**
25: trans_end_time = Processing()
26: **Output:** current_packet
27: **end if**
28: **else**
29: **Output:** 0
30: **end if**

Additionally, in practical scenarios with an adaptive data rate source, the satellite would monitor a specific time window to count the number of packets received from the source. By knowing the packet size, the satellite can calculate the data rate from the previous time window and use this as the default transmission rate. This can be expressed by the following equation:

$$\text{transmission\_rate} = \frac{\text{number\_of\_packets} \times \text{packet\_size}}{\text{window\_size}} \tag{18}$$

**Algorithm 3** Processing()

**Initialize:** packet_size
**Input:** current_time, current_bitrate
1: transmission_delay = packet_size/current_bitrate
2: trans_end_time = current_time + transmission_delay
3: **Output:** trans_end_time

The Processing() function, as defined in Algorithm 3, is responsible for calculating the transmission delay and

transmission end time of the current packet (line 11, Algorithm 2 and lines 1–3, Algorithm 3).

### E. BEAM ANGLE ADJUSTMENT AND HANDOVER BASED ON MOVEMENT PREDICTION AND NAK FEEDBACK

Unlike GEO satellites, LEO satellites orbit the Earth at altitudes ranging from 300 to 1200 km, which means that they cannot remain stationary relative to the ground. This introduces new challenges in utilizing LEO satellites for multicast services. Specifically, the constant change in the position of the satellite relative to the ground cell necessitates real-time adjustments of the beam angles. Moreover, communication interruptions may occur when a ground cell moves from one satellite's coverage area to another.

To address these challenges, the proposed beam angle adjustment and handover scheme estimates and adapts the beam angle variation of LEO satellites based on their velocity and orbit direction, as described in Section III-E. Additionally, satellites receive location feedback through NAKs sent from the ground. Given the close arrangement and unique identification of the ground cells, a NAK from any cell in any direction can provide angle corrections for all satellite beams.

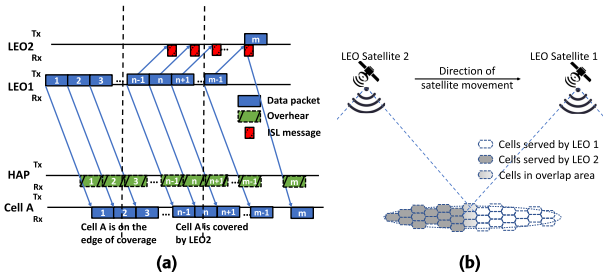



**FIGURE 7. Handover in NTN-RM. (a) is an example of the Algorithm 4, (b) is an example of the network scenario.**

**Algorithm 4** Beam Angle Adjustment and Handover Algorithm

**Initialize:** $(x_0, y_0, z_0)$, $(x_c, y_c, z_c)$, $v$
**Initialize:** max_elevation, angle_threshold
**Input:** current_time
1: calculate distance using formula (15)
2: calculate beam_angle using formula (17)
3: **while** Receive NAK **do**
4: Get the receive angle and cell sequence
5: Update beam angle for the corresponding cell
6: correct beam_angle for this cell
7: **end while**
8: **if** beam_angle i smaller than max_elevation **then**
9: **Output:** beam_angle
10: **else if** beam_angle is larger than max_elevation but smaller than angle_threshold **then**
11: Start sharing transmission information through ISL with the next satellite
12: **Output:** beam_angle
13: **else**
14: Handover service responsibilities to the next satellite
15: Stop serving the current cell by the current satellite
16: **end if**

Furthermore, the use of ISLs facilitates handover procedures for LEO satellites, as shown in Figure 7(a). When the cells move out of coverage, the satellite detects them based on the beam angle and starts sharing transmission information with its neighboring satellite via the ISLs. This information includes the sequence numbers of the corresponding cells and currently transmitted packets. As shown in Figure 7(b), when cells are within the overlapping area of two satellites, they have two possible links to the NTN. However, only the link with the old satellite is active at this stage. Meanwhile, the spare satellite acquires relevant content in preparation for future transmission. Once the cells are entirely covered by the new satellite, the old satellite detects that the beam angle exceeds the threshold, severs the links, discontinues service, and the new satellite takes over based on the information received from the old satellite. Algorithm 4 provides a pseudocode detailing the beam-angle adjustment and handover processes.

## V. EVALUATION

### A. EVALUATION METHODS AND INDICATORS

In this section, we compare the proposed approach with the existing reliable multicast protocols. Notably, these existing protocols were initially designed for terrestrial networks and may not be directly applicable to NTNs. Therefore, we retained the core features of these protocols and envisioned their potential applications in NTNs as benchmarks for comparison.

The proposed approach was evaluated and compared with seven benchmarks using Monte Carlo simulations. The evaluation methods are presented in Table 3.

1) **Benchmark 0**: This is a single-beam LEO direct transmission and retransmission method used to evaluate the effectiveness of the multibeam scheme. However, it is difficult to present the results in the figure along with other methods. The findings are discussed in detail in Section V-C4.
2) **Benchmark 1**: This protocol does not consider the utilization of HAPs and relies solely on multibeam direct LEO transmission and retransmission. This represents the scenario of applying REMP [12] or NORM [14] to NTNs.
3) **Benchmark 2**: This protocol includes HAPs but assumes that HAPs relay packets from LEO and handle possible retransmissions without overhearing LEO signals. This represents the scenario of applying PGM [13] to NTNs.
4) **Benchmark 3**: This pertains to the proposed approach without a handover scheme and was intentionally employed to evaluate the effectiveness of Algorithm 4. In this benchmark, when a cell moves to the edge of a LEO satellite's coverage area, it must send a request for every new packet to avoid interference caused by being served by two LEO satellites simultaneously.

5) **Benchmark 4**: This represents the Sat-RMTP [27] protocol. It utilizes FEC(20,12) to transmit packets. If a packet cannot be recovered at the UE, it is permanently lost in this protocol. Moreover, the check bits are expected to have a negative influence on transmission delay and throughput.
6) **Benchmark 5**: This benchmark illustrates the application of the RMM [15] protocol in NTNs, which allows the retransmission of only I-frame packets.
7) **Benchmark 6**: This benchmark represents the NC-MAC [16] protocol in NTNs, which improves multicast reliability through repeat transmissions and network coding-based retransmissions. Under normal transmission conditions, the satellite sends each new packet twice. For retransmissions, the satellite sends two independent linear combinations of the new packet and two retransmitted packets.

**TABLE 3. Evaluation methods.**

| | Multi-beam | Retrans-mission | HAP | Overhear & NOMA | Handover method (Algorithm 4) |
|---|---|---|---|---|---|
| Benchmark 0 - single-beam satellite | × | ✓ | × | × | ✓ |
| Benchmark 1 - REMP [12] or NORM [14] | ✓ | ✓ | × | × | ✓ |
| Benchmark 2 - PGM [13] | ✓ | ✓ | ✓ | × | ✓ |
| Benchmark 3 - NTN-RM w/o Algorithm4 | ✓ | ✓ | ✓ | ✓ | × |
| Benchmark 4- SatRMTP [27] | ✓ | × | × | × | ✓ |
| Benchmark 5- RMM [15] | ✓ | ✓ | × | × | ✓ |
| Benchmark 6- NC-MAC [16] | ✓ | × | × | × | ✓ |
| Proposed Work - NTN-RM | ✓ | ✓ | ✓ | ✓ | ✓ |

The methods are evaluated using the following quality of experience indicators:

- **Mean latency** represents the average delay from LEO satellites to UE for all packets received by all UE.
- **Reliability** is defined as the ratio of the number of successfully received packets to the total number of transmitted packets.
- **Throughput** refers to the bit rate of packet delivery through NTNs.
- **Jitter** represents the fluency of multimedia received by the UE, calculated from the average value of a set of randomly selected UE.

Furthermore, certain related works cannot be directly compared with the proposed work because of the differences in their core features and assumptions regarding NTNs. For instance, [28], [29] assumed a GEO satellite-based NTN multicast application, where the propagation delay and coverage characteristics of GEO satellites are significantly different from those of LEO satellites. Therefore, it would be challenging to directly compare the proposed work with these works, as they are designed for different types of satellite systems.

### B. SIMULATION ENVIRONMENT AND PARAMETERS

The simulation was repeated several times to achieve a 95% confidence interval. A region with an area equal to the coverage of a LEO satellite was assumed. LEO satellites traverse this area and obtain the required content from the source at a bitrate of 60 Mbps [37]. The simulation did not consider the propagation delay and packet loss before the LEO satellite received the data. Upon receiving the data, the LEO satellites downlink it using the protocols listed in Table 3. As the cells can be considered independent of each other, we simulated each cell separately and aggregated their results to obtain the final outcomes. However, we also considered interactions between cells, such as NAK collisions, which can cause failed retransmissions. We used random process simulations to capture their effects. The simulation time was the period of one satellite handover. For example, it was approximately 263 seconds for an elevation angle of 30° and 87.7 seconds for a default elevation angle of 60°.

#### 1) NAK COLLISION RATE

Assuming that the waiting time for each receiver is independently and identically distributed according to an exponential distribution with mean $\mu$, the probability density function of the waiting time is given by

$$f(t) = (1/\mu)e^{-t/\mu}, t \geq 0 \tag{19}$$

where $\mu$ is the mean waiting time. The probability that a UE sends a NAK within a time interval of length $T$ starting from the detection of the error is given by the cumulative distribution function (CDF) of the exponential distribution

$$F(T) = 1 - e^{-T/\mu}, T \geq 0 \tag{20}$$

Therefore, the probability that a UE sends a NAK after time $t_1$ and before time $t_2$ is

$$\begin{aligned} P(t_1 \leq T < t_2) &= F(t_2) - F(t_1) \\ &= e^{-t_1/\mu} - e^{-t_2/\mu}, t_2 \geq t_1 \geq 0 \end{aligned} \tag{21}$$

The waiting time is random and independent across receivers, so the distribution of the time until the first UE sends a NAK is the minimum of the waiting times of all UE. This minimum waiting time is also exponentially distributed with mean $\mu$ but with a different parameter. Specifically, the parameter of the minimum waiting time distribution is $\mu/m$, where $m$ is the number of UE. Therefore, the cumulative distribution function of the minimum waiting time is

$$G(T) = 1 - e^{-mT/\mu}, T \geq 0 \tag{22}$$

The probability that the first UE sends a message after time $t$ is

$$P(T \geq t) = 1 - G(t) = e^{-mt/\mu}, t \geq 0 \tag{23}$$

Let $T_i$ be the waiting time for the $i$-th UE, and let $N_i$ be the indicator variable that is 1 if the $i$-th UE sends a NAK and 0 otherwise. Then, the probability that the $i$-th UE sends a message at time $T$ is

$$P(N_i = 1, T_i = T) = f(T)(1 - F(T))^{i-1} F(T)^{m-i} \quad (24)$$

The probability that two or more UE send a NAK at the same time can be calculated as the sum of the probabilities that each pair or a larger group of UE send a message at the same time:

$$P(collision) = \sum_{i=2}^{m} \sum_{j=1}^{i-1} P(Ni = 1, Nj = 1, |T_i - T_j| < T_x) \quad (25)$$

where $T_x$ is the transmission time of a NAK.

**TABLE 4. Simulation parameters.**

| | LEO | | HAP | | UE | |
|---|---|---|---|---|---|---|
| Altitude | 600 km | | 20 km | | 0 km | |
| Device density | refer to elevation | | 0 or 1 for each LEO cell (1 as default) | | $10^2 \sim 10^8$ ($10^4$ as default) | |
| Number of beams | up to 1171 | | 100 | | / | |
| Frequency [9] | To UE | To HAP | To UE | To LEO | To LEO | To HAP |
| | 2GHz | 20GHz | 2GHz | 30GHz | 2GHz | 2GHz |
| Bandwidth [9] | 30MHz | 400MHz | 30MHz | 400MHz | 360kHz | 360kHz |
| EIRP [9] | 48.8dBW | 30dBW | 46.2dBW | 46.2dBW | -7dBW | -7dBW |
| $G/T$ [9] | N/A | N/A | N/A | 13dB/K | 1.1dB/K | N/A |
| Rx Antenna gain [9] | 0dBi | 39.7dBi | 0dBi | N/A (already included in $G/T$) | | 28.1dBi |
| Noise figure [9] | 7dB | 1.2dB | 7dB | N/A (already included in $G/T$) | | 1.2dB |
| Antenna temperature [9] | 290K | 150K | 290K | N/A (already included in $G/T$) | | 150K |
| Ambient temperature | 298K | 244K | 298K | N/A (already included in $G/T$) | | 244K |
| Atmospheric loss [9] | 0.1dB | 0.5dB | 0dB | 0.5dB | 0.1dB | 0dB |
| Shadow fading [10] | $LogN(0, \sigma_{SF}^2)$ | | | | | |
| Scintillation loss [9] | 2.2dB | 0.3dB | 0dB | 0.3dB | 2.2dB | 0dB |
| Elevation angle | 30°/ 40°/ 50°/ 60°(default) / 70°/ 80° | | | | | |
| Satellite number | required number for global coverage (corresponding to the elevation): 258 / 493 / 948 / 1948 (default) / 4829 / 20417 | | | | | |
| UE distribution | uniform distribution (default) / normal distribution | | | | | |
| Scenario [9] | rural (default) / urban / dense urban | | | | | |
| the Boltzmann constant $k$ = -228.6 dBW/K/Hz | | | | | | |
| data packet size: 500/1,000(default)/1,500 bytes | | | | | | |
| control packet size: 32+11=43 bit | | | | | | |

#### 2) ENVIRONMENT

The simulation parameters for the system were set based on the specifications of the 3GPP releases [9], [10], [11], as presented in Table 4 and Section III-C. The distances between the devices were calculated based on their altitudes. The effective isotropic radiated power (EIRP) refers to the hypothetical power that an isotropic antenna would need to radiate to achieve the same signal strength as the actual source antenna in the direction of the antenna's strongest beam. The shadow fading model used is discussed in Section III-C. The atmospheric loss represents rain fading, while scintillation loss models ionospheric scintillation fading. The data packet sizes were assumed to be 500, 1,000, and 1,500 bytes, while the control packet size was set to 43 bits, with 32 bits reserved for the data sequence and 11 bits for the cell sequence. These parameter values were selected based on those reported in relevant literature to ensure consistency and accuracy in the simulations. However, due to limited information regarding the link characteristics of HAPs at the time of this research, data from similar devices, such as very small aperture terminals (VSAT), were used as a reference.

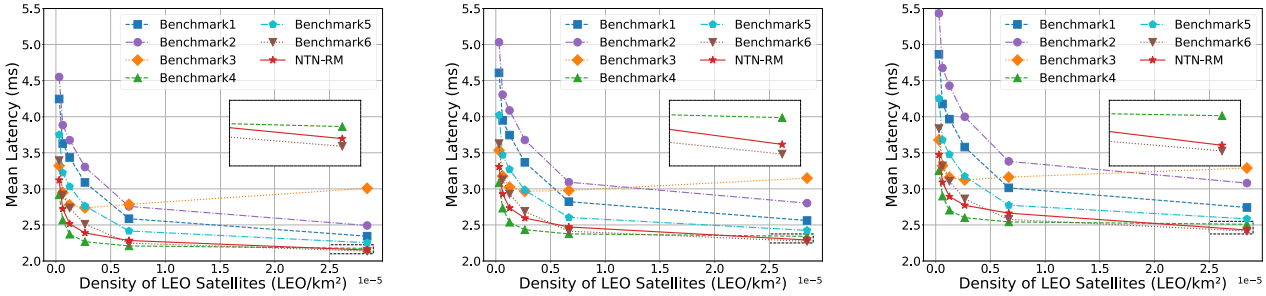
### C. RESULTS AND ANALYSIS

#### 1) MEAN LATENCY WITH VARYING SATELLITE DENSITY

The mean latency was calculated as the average delay for all packets received by the UE. The delay was measured from the time the LEO satellites initiated packet reception to the time when the packets reached the UE. In Figure 8, the six satellite densities correspond to the elevation angles of 30, 40, 50, 60, 70, and 80 degrees, respectively. The results demonstrate that the mean latency is influenced by the packet size, primarily due to transmission delays. It can also be observed that the effect of packet size is mainly on the absolute value, with a slight impact on the relative relationship among different methods. These trends are also evident in subsequent evaluations. Therefore, in the following results, we present data using a 1000-byte packet size.

In Figure 8(b), for methods except benchmark 3, a higher satellite density results in a lower mean latency. This is primarily due to the impact of propagation delay and fading, which are influenced by the average distance between the satellites and UE. However, the mean latency of benchmark 3 initially decreases and then increases. This pattern is mainly attributed to the increasing frequency of handovers with an increase in satellite density. Benchmark 3 lacks the method shown in Algorithm 4 to handle handovers, which contributes to the observed behavior.

Among the results, benchmark 4 achieved a low mean latency because the lost packets in this method cannot be recovered at the UE, thus, the impact of packet loss is not reflected in the mean latency for this method. However, NTN-RM outperformed benchmark 4 when the satellite density reached its maximum, as the number of retransmissions in NTN-RM decreased while the transmission delay for benchmark 4 remained constant. Additionally, benchmark 6 also demonstrated a low mean latency under a high satellite density. This is because repeated transmissions significantly reduce the need for retransmissions when the satellite downlink channel quality improves with higher satellite density. Benchmark 1 exhibited a large mean latency due to the long retransmission delay from the satellite. Benchmark 5 showed a lower latency compared to benchmark 1, as it only retransmits I-frame packets. Benchmark 2 exhibited a larger mean latency than benchmark 1, even though it used HAPs to relay packets and handle local retransmissions. This




(a) data packet size: 500 bytes (b) data packet size: 1,000 bytes (c) data packet size: 1,500 bytes

**FIGURE 8.** Mean latency with varying LEO satellite density.

is because the CNR and success rate of satellite-to-HAP links are low in the default environment, which are influenced by the Rx antenna gain of the HAP. Additional results are presented in the following evaluation, where the receiver gains are varied to provide an objective assessment.

In summary, under the default conditions, NTN-RM outperformed benchmark 1 by approximately 23%, benchmark 2 by 29.4%, benchmark 3 by 11.9%, benchmark 5 by 13%, and benchmark 6 by 3.5% in terms of mean latency. NTN-RM was only slightly worse than benchmark 4, lagging by about 6.6%. However, evaluations of reliability and throughput showed that NTN-RM surpassed benchmark 4 in these areas. The 95% confidence intervals for Benchmarks 10–6 and NTN-RM were on the order of $1 \times 10^{-3}$, $2 \times 10^{-3}$, $7 \times 10^{-4}$, $1 \times 10^{-4}$, $2 \times 10^{-3}$, $3 \times 10^{-3}$, and $6 \times 10^{-4}$ ms, respectively. Based on the results, a LEO satellite density greater than $5 \times 10^{-6}$ LEO/$km^2$ is recommended for multicast applications.

### 2) RELIABILITY WITH VARYING SATELLITE DENSITY

Figure 9 shows the reliability across varying satellite densities. Benchmarks 1, 2, and 3 show results similar to those of the NTN-RM. Benchmark 6 had a packet arrival rate below 99% when the satellite density was less than $5 \times 10^{-6}$ LEO/$km^2$, due to packet loss caused by the failure to decode combination packets in the NC-MAC protocol. Benchmark 4's packet arrival rate dropped below 99% when the satellite density fell below $2.5 \times 10^{-5}$ LEO/$km^2$, highlighting its approach of sacrificing reliability for lower latency. Benchmark 5 was unable to achieve a 99% packet arrival rate under any satellite density, as it does not retransmit P- and B-frames. Considering that the reliability requirement for 5G is 99.99% [31], benchmarks 4, 5, and 6 are unsuitable for latency-sensitive and reliability-sensitive XR applications, which is the focus of this study. However, they may be appropriate for reliability-insensitive applications, as they do not require HAPs and have lower deployment costs. Additionally, satellite direct downlink methods such as benchmark 1 could be a cost-effective solution for latency-insensitive but reliability-sensitive applications.

As the satellite density increases from 0 to $5 \times 10^{-6}$ LEO/$km^2$, the packet arrival rates of benchmarks 4, 5, and 6 initially increase, then fall, and finally increase again. This phenomenon occurs because, when the elevation angle shifts from 30Â° to 60Â°, the satellite downlink channel experiences fluctuations, as shown in Table 1 and Table 2. However, these fluctuations are not significant enough to impact the packet arrival rates of benchmarks 1, 2, 3, or NTN-RM. The 95% confidence intervals for benchmarks 1–6 and NTN-RM were on the order of $4 \times 10^{-6}$, $7 \times 10^{-6}$, $7 \times 10^{-7}$, $2 \times 10^{-4}$, $1 \times 10^{-4}$, $2 \times 10^{-4}$, and $1 \times 10^{-6}$, respectively. Furthermore, the differences between the methods are specific to the default environment (as detailed in Table 4). A larger number of UE, increased background traffic, or denser scenarios can amplify these differences. As illustrated in Figures 12, 15, and 19, NTNs with HAPs demonstrate better performance under such conditions.

### 3) THROUGHPUT WITH VARYING SATELLITE DENSITY

Figure 10 lists the throughput across varying satellite densities. Benchmarks 1, 2, and 3 yielded results similar to those of the NTN-RM. The throughput of benchmark 4 was significantly affected by FEC encoding. While using more check bits could help improve the reliability of benchmark 4, it would lead to a lower throughput. The 95% confidence intervals for benchmarks 1–6 and NTN-RM were on the order of $2 \times 10^{-4}$, $4 \times 10^{-4}$, $4 \times 10^{-5}$, $9 \times 10^{-3}$, $1 \times 10^{-2}$, $1 \times 10^{-2}$, and $1 \times 10^{-4}$ Mbps, respectively.

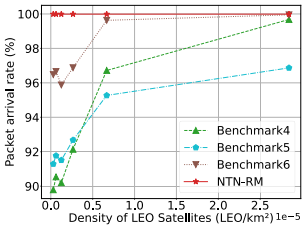



**FIGURE 9.** Packet arrive rate with varying LEO satellite density.

### 4) MEAN LATENCY AND RELIABILITY WITH VARYING NUMBERS OF UE

Figure 11 depicts the mean latency as the number of UE varies. Since benchmark 4's performance is theoretically and practically unaffected by the number of UE, its results are not shown in the figure. According to the results, the mean latency of benchmark 1 starts to increase when the number

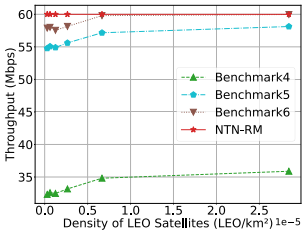



**FIGURE 10.** Throughput with varying LEO satellite density.

of UE exceeds ten thousand. This increase occurs because, as the number of UE increases, packet loss becomes more frequent, leading to more retransmission requests and causing transmit queue congestion at the satellite. The latency tends toward infinity when the number of UE exceeds two hundred thousand. For the same reason, benchmark 5 experiences a rise in mean latency after the number of UE surpasses one hundred thousand, and it approaches infinity as the number of UE nears one million.

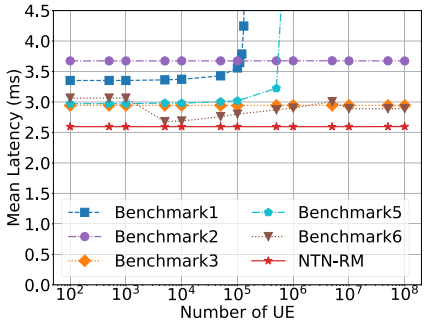



**FIGURE 11.** Mean latency with varying number of UE.

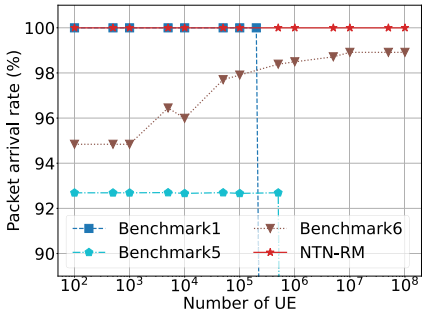



**FIGURE 12.** Packet arrival rate with varying number of UE.

In contrast, benchmarks 2 and 3, and the proposed approach did not experience significant changes in mean latency as the number of UE increased to one hundred million. These results demonstrate the potential of HAPs in enhancing the network capacity of NTNs. For benchmark 6, the mean latency initially decreases as the number of UE increases. This is because combination packets require two different retransmitted packets, and with fewer UE, a retransmitted packet may wait longer for the other packet. As the number of UE continues to rise, the mean latency then increases due to a higher frequency of NAK collisions between cells, which leads to longer retransmission delays.

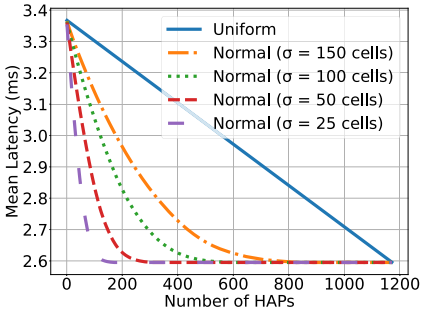



**FIGURE 13.** Mean latency of NTN-RM with varying number of HAPs.

However, once the number of UE reaches a point where all cells require retransmission at each time slot, the NAK collision rate stabilizes, and the mean latency decreases again as the interval between retransmitted packets becomes the dominant factor. Finally, the mean latency stabilizes, as each packet is now being retransmitted. The 95% confidence intervals for benchmarks 1–6 and NTN-RM were on the order of $2\times10^{-3}$, $2\times10^{-3}$, $7\times10^{-4}$, $4\times10^{-5}$, $2\times10^{-3}$, $3\times10^{-3}$, and $7\times10^{-4}$ ms, respectively.

Additionally, the mean latency of benchmark 0 was approximately 8,945 ms under the default conditions. This indicates that heavy congestion occurs when a LEO satellite attempts to serve a large number of UE using a single beam.

Moreover, we evaluated the packet arrival rate under varying numbers of UE, as shown in Figure 12. The packet arrival rate for Benchmarks 2, 3, and NTN-RM remains consistently high at 99.99%, regardless of the number of UE. In contrast, for Benchmarks 1 and 5, the packet arrival rate drops to zero when the latency becomes infinite due to network congestion. For benchmark 6, the packet arrival rate increases from 94.84% to 98.71% as the number of UE grows, stabilizing at 98.91% when the number of UE exceeds $10^7$. This phenomenon occurs because, with a smaller number of UE, there is a higher chance that two combined retransmitted packets originate from the same UE, leading to decoding failure. As the number of UE increases, each packet is encoded twice with its preceding and succeeding packets, improving the success rate of decoding and consequently increasing the packet arrival rate. The 95% confidence intervals for benchmarks 1–6 and NTN-RM were on the order of $1\times10^{-6}$, $3\times10^{-6}$, $5\times10^{-7}$, $3\times10^{-4}$, $2\times10^{-4}$, $2\times10^{-4}$, and $3\times10^{-7}$, respectively.

### 5) MEAN LATENCY AND RELIABILITY WITH VARYING NUMBER OF HAPs

Figure 10 illustrates the mean latency of NTN-RM as the number of HAPs varies. We evaluated the performance of NTN-RM under different UE distributions: a uniform distribution and a normal distribution with varying standard deviations ($\sigma$). In this study, the maximum number of HAPs is achieved when each cell is served by one HAP. If the number of HAPs is less than this maximum, some cells will be served solely by the LEO satellites. Our HAP deployment

strategy prioritizes deploying HAPs in cells with a higher concentration of UE.

The results show that, when the UE is uniformly distributed, the mean latency decreases linearly with the increase in the number of HAPs. In the case of normally distributed UE, depending on the degree of aggregation, the mean latency reaches an acceptably low level when the number of HAPs reaches 100, 300, 500, and 800 for standard deviations ($\sigma$) of 25, 50, 100, and 150 cells, respectively.

Moreover, we also evaluated the packet arrival rate under varying numbers of UE to assess its impact on reliability. The results show that the packet arrival rate for NTN-RM remains consistently high at 99.99%, regardless of the number of HAPs. Combined with the conclusion from the previous section, if the number of UE exceeds $10^5$, LEO direct transmission cells will experience a decrease in reliability. In such cases, the total packet arrival rate increases as the number of HAPs increases. Additionally, the 0.8 ms difference between scenarios with 1200 HAPs and 0 HAPs is a specific result observed under the default environment (as detailed in Table 4). Fewer LEO satellites, a larger number of UE, or denser scenarios can amplify this difference, with the latter two factors having a particularly significant impact. As illustrated in Figures 11 and 18, a large number of UE or a dense scenario can render NTNs without HAPs nonfunctional due to network congestion.

### 6) MEAN LATENCY AND RELIABILITY WITH VARYING BACKGROUND TRAFFIC

The mean latency with varying background traffic is shown in Figure 14. The mean latency for benchmarks 2, 3, and NTN-RM tends to become infinite when the background traffic exceeds 60 Mbps, as the network capacity is surpassed and congestion occurs.

Between 0 and 50 Mbps, benchmark 2's mean latency increases, because the background traffic extends the waiting time in the queue. However, the mean latency of benchmark 3 and NTN-RM remains relatively unchanged before 50 Mbps, as they both employ the LEO satellite queue control method. The background traffic keeps the satellite queue busy, preventing the need to suppress the data rate and thereby reducing transmission delays for the required packets.

Benchmark 1 shows infinite latency when the background traffic exceeds 20 Mbps, as it lacks HAPs for retransmission offloading, making it more susceptible to congestion. Similarly, benchmark 5 experiences infinite latency when the background traffic surpasses 40 Mbps, although the strategy of retransmitting only I-frame packets reduces the load to some extent.

Benchmark 4, which does not involve any retransmissions, reaches a Nyquist capacity of 120 Mbps (60 Mbps for the required content and 60 Mbps for background traffic). In contrast, benchmark 6, which already utilizes all available capacity for repeat transmissions, cannot tolerate any additional background traffic. The 95% confidence intervals for benchmarks 1–6 and NTN-RM were on the order of $3 \times 10^{-3}$, $3 \times 10^{-3}$, $6 \times 10^{-4}$, $3 \times 10^{-5}$, $2 \times 10^{-3}$, $3 \times 10^{-1}$, and $6 \times 10^{-4}$ ms, respectively.

The packet arrival rates with varying background traffic are shown in Figure 15. The packet arrival rate drops to zero when network congestion occurs. The packet arrival rates for benchmarks 2 and 3 show similar results to those of NTN-RM. These results align with the observed latency trends. The 95% confidence intervals for benchmarks 1–6 and NTN-RM were on the order of $9 \times 10^{-7}$, $3 \times 10^{-6}$, $3 \times 10^{-7}$, $3 \times 10^{-4}$, $1 \times 10^{-4}$, $2 \times 10^{-4}$, and $5 \times 10^{-7}$, respectively.

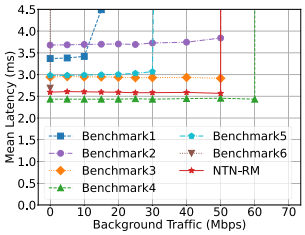



**FIGURE 14.** Mean latency with varying background traffic.

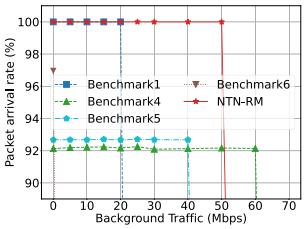
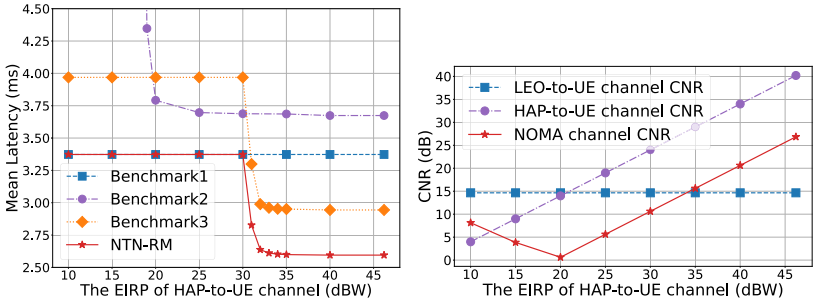



**FIGURE 15.** Packet arrive rate with varying background traffic.

### 7) MEAN LATENCY WITH VARYING EIRP OF HAP

The following sections focus on evaluating the impact of HAPs. Since benchmarks 4, 5, and 6 do not incorporate the use of HAPs, they are not included in the following results. However, benchmark 1 is presented as a reference for comparison. The mean latency and channel CNR with varying EIRP for the HAP-to-UE channel are shown in Figure 16. The CNR of the NOMA channel, influenced by the CNR difference between the HAP-to-UE and LEO-to-UE channels, decreases as the HAP-to-UE channel's EIRP rises from 10 dBW to 20 dBW due to a diminishing difference. Beyond 20 dBW, as this difference increased, the CNR of the NOMA channel increased.

As shown in Figure 16(b), when the EIRP of the HAP-to-UE channel changes from 35 dBW to 30 dBW, the CNR of the NOMA channel decreases below 15 dB. This results in a nonlinear increase in the mean latency for both benchmark 3 and the proposed approach, as depicted in Figure 16(a). This increase brings the mean latency values closer to or even higher than those of benchmark 1. This is because the overhear and NOMA schemes employed in benchmark 3 and NTN-RM were ineffective when the CNR of the NOMA channel was low. Consequently, the proposed approach switches to LEO direct transmission and




(a) Mean latency (b) Channel CNR (w/o fading)

**FIGURE 16.** Mean latency and channel CNR with varying EIRP of HAP-to-UE channel.

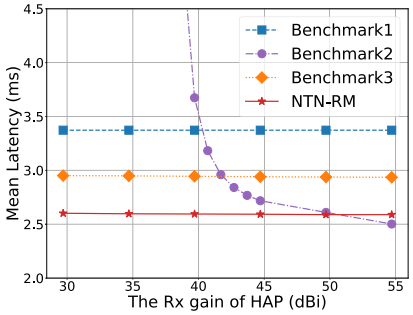



**FIGURE 17.** Mean latency with varying Rx gain of HAPs.

behaves similarly to benchmark 1, whereas benchmark 3 also resorts to LEO direct transmission and can be considered as benchmark 1 without the handover scheme, resulting in a higher mean latency compared to that of benchmark 1 when the EIRP is lower than 30 dBW. For benchmark 2, a nonlinear increase in the mean latency began when the EIRP fell below 25 dBW. This is because, in benchmark 2, the HAP signal remains unaffected by the LEO satellite signal at the UE, allowing it to maintain a higher CNR and a lower requirement for the EIRP of HAP. In summary, the NTN-RM requires a higher EIRP for the HAP-to-UE channel to achieve optimal performance.

### 8) MEAN LATENCY WITH VARYING RX GAINS OF HAPS

Figure 12 illustrates the mean latency with varying receiver (Rx) gains of the LEO-to-HAP channel.

When the Rx gain ranges from 39.7 dBi to less, the mean latency of benchmark 2 increased rapidly. This is because the HAPs in benchmark 2 had a low success rate in receiving packets from the LEO satellites under these conditions, resulting in numerous retransmission requests. Moreover, the mean latency of benchmark 2 exhibited a nonlinear decrease with an increase in the Rx gain of HAPs beyond 39.7 dBi. When the Rx gain exceeds 54.7 dBi, benchmark 2 outperforms the proposed approach. This is because under the conditions of sufficiently high transmit power and Rx gain of the HAPs, the relay process in benchmark 2 tends to function optimally. However, the performances of benchmark 3 and the proposed approach are also influenced by the LEO-to-UE channel, resulting in benchmark 2 outperforming the NTN-RM.

In contrast, the performances of benchmark 3 and the proposed approach remained relatively stable with varying Rx gain of the HAP. This is because HAPs receive packets by overhearing them and rely on direct transmission links from the LEO satellites to the UE. The LEO-to-UE channel has a higher transmit power than the LEO-to-HAP channel, reducing the Rx gain requirements for the HAPs to effectively receive the LEO signal. In summary, the proposed approach has a lower requirement for the Rx gain of HAPs than benchmark 2. When the Rx gain is sufficiently high, benchmark 2 may provide a better performance. However, a lower Rx gain for benchmark 2 carries certain risks.

Additionally, it is important to note that the reference values of EIRP as 46.2 dBW and Rx gain as 39.7 dBi are based on similar equipment in the 3GPP release [9]. The numerical adjustments made in these evaluations are solely for evaluation purposes and do not consider the feasibility of physical layer implementation. Hence, further discussion and experiments considering the physical layer are necessary for future work.

### 9) MEAN LATENCY AND JITTER WITH DIFFERENT SCENARIOS

The mean latencies under different scenarios are shown in Figure 18. These scenarios, defined in [10], affect the results by influencing LOS probability, fading factor, and clutter loss.

The results show that benchmark 1 is only viable in rural scenarios, as its mean latency becomes infinite in both urban and dense urban environments. The poor LEO-to-UE channel conditions in these environments make it difficult for LEO satellites to efficiently manage a large number of retransmission requests. For the same reason, the performance of Benchmarks 4, 5, and 6 is significantly affected in urban and dense urban scenarios. Benchmark 2 also exhibited high mean latency in these environments. Simply increasing the LEO transmit power does not resolve the issue in benchmark 2, as it introduces substantial interference to the HAP signal at the UE. In contrast, NTN-RM consistently maintained a mean latency of less than 4 ms across all three scenarios, demonstrating its robust performance within the NTN framework.

Furthermore, to evaluate the streaming data transmission performance of each method, we calculated the jitter to reflect the smoothness of the multimedia display, which is defined as the standard deviation of packet latency in this study. A random subset of the UE was selected after the simulation, and the latency of all the received packets was recorded. The jitter was then calculated for each method and the mean value was used to represent the jitter of the corresponding method, as shown in Figure 19.

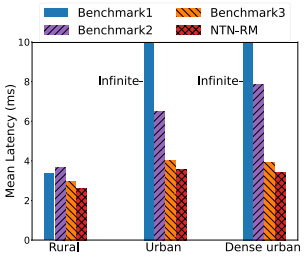



**FIGURE 18.** Mean Latency with different scenarios.

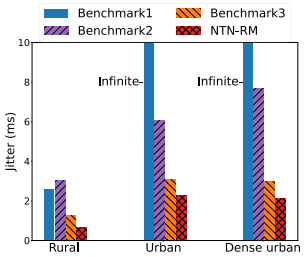



**FIGURE 19.** Jitter with different scenarios.

The results revealed that in rural areas, the proposed approach achieved 75.8% lower jitter than benchmark 1, 79.1% lower jitter than benchmark 2, and 49.3% lower jitter than benchmark 3. This difference can be attributed to the enhanced efficiency of the HAPs in NTN-RM in terms of obtaining transmission content from LEO satellites, handling retransmissions, and facilitating uplink feedback. In urban and dense urban areas, benchmarks 1 and 2 exhibited significant jitter for the reasons mentioned in Section V-C9. In contrast, NTN-RM maintained lower jitter in all scenarios.

### 10) LATENCY DISTRIBUTION

To further analyze the simulation results and verify the conclusions drawn in the previous section, we examined the distribution of packet latency for the different methods, as shown in Figure 20 and 21.

Considering the overall view shown in Figure 20(a), it is evident that benchmark 3 and NTN-RM exhibit relatively concentrated latency. However, benchmarks 1 and 2 have a higher number of packets arriving at a higher latency, particularly in urban and dense urban scenarios. Figure 20(b) shows a zoomed-in view of the specific latency range between 0 and 15 ms for all lines. Within the range of 2–3 ms, packets are either transmitted once and received successfully or undergo retransmission by HAPs without queue congestion. All methods show a nearly linear increase within this range, primarily due to the approximately linear change in the distance between the cells and LEO satellites, as governed by (15), resulting from the movement of LEO satellites.

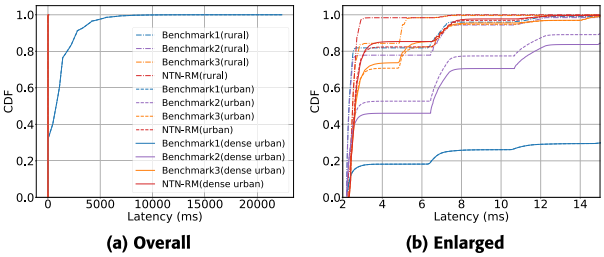



**FIGURE 20.** Cumulative distribution of the latency. (a) is for the overall view, and (b) is for a partial view of latency between 0 and 15 milliseconds.

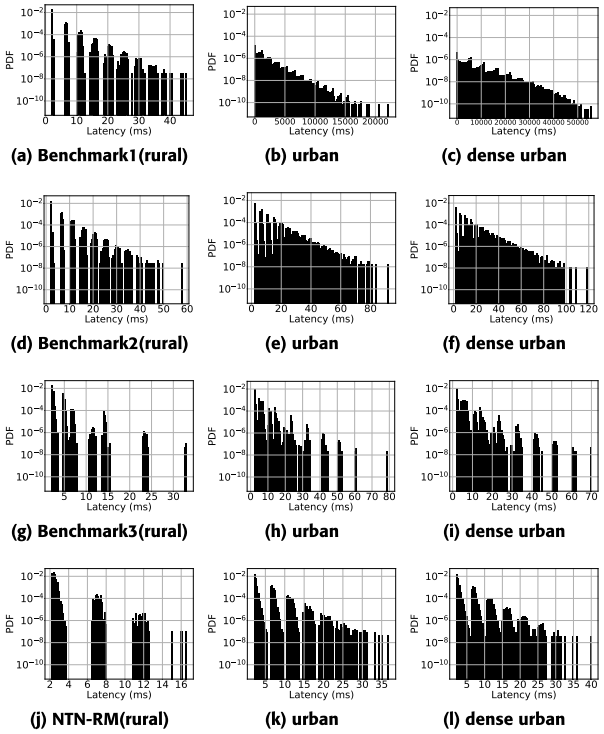



**FIGURE 21.** Probability density of the latency.

Beyond the 4 ms latency range, a gap appears until the latency exceeds 6 ms, which is more pronounced in the figures depicting rural scenarios in Figure 21. This gap arises because of the significant increase in latency caused by LEO satellite retransmission. For benchmark 3, the latency range of 5–6 ms arises because of the absence of a handover method when cells are at the edge of satellite coverage. The number of packets decreased as the latency increased, displaying an overall logarithmic relationship and a step-like distribution. This phenomenon is attributed to the packets being retransmitted by satellites one or more times. The first step occurs at approximately 6–8 ms and primarily comprises the latency of packets that have been retransmitted once by satellites, followed by subsequent steps corresponding to packets that have been retransmitted two or more times.

Furthermore, as shown in Figure 20(b), LEO retransmission with a latency of at least 6 ms occurs in benchmark 1 and 2, or benchmark 3 (urban/dense urban) and NTN-RM (urban/dense urban), whereas it occurs with at least 8 ms in benchmark 3 (rural) and NTN-RM (rural). This observation indicates that LEO retransmission only occurs when there is a significant distance between the cells and LEO satellites in benchmark 3 and NTN-RM in the rural scenario.

In Figure 21, considering the urban and dense urban scenarios, it is evident that the number of steps, representing the number of satellite retransmissions, increases significantly for all methods. This signifies that poor channel conditions make it more challenging for HAPs and UE to obtain data packets, thereby increasing the number of satellite retransmissions. Additionally, it can be observed that for benchmarks 1 and 2, not only does the number of steps increase, but the width of each step also increases. This indicates that the average retransmission delay increases due to queue congestion, as numerous packets waste time waiting in the queue or undergoing many retransmissions. These observations support the conclusions presented in Section V-C9.

In the case of benchmark 1, it is clear that a higher number of retransmissions is required as the number of UE increases. The distribution appears to be connected rather than discrete due to three factors. First, the propagation delay varies as satellites move. Second, when a retransmitted packet enters the satellite queue, the satellite may be transmitting other packets, resulting in an approximately random shift in the delay. Finally, uplink random access faces a higher collision rate when a large number of UE send NAK to the LEO satellite, leading to an additional noticeable random shift in the delay.

Comparing Figures 21(j) and 21(a), it can be observed that the presence of HAP significantly reduces the number of LEO retransmissions for all packets. However, it still lags far behind the proposed approach. This disparity can be attributed to two reasons. First, as mentioned in the analysis of benchmark 2, HAPs may be occupied by relaying packets from LEO satellites, which can increase the latency due to a busy HAP queue. Second, the packet loss rate of the HAP in benchmark 2 is higher, leading to more satellite retransmissions. This is because, in the proposed approach, the LEO satellite sends signals directly to the UE using 2GHz with higher transmit power, and the HAP experiences a lower bit error rate when overhearing.

### 11) COMPLEXITY FOR EACH PACKET TRANSMISSION

The complexity for each packet transmission in benchmarks 1–6 and NTN-RM is outlined in Table 5. In benchmark 0, for each packet transmission, the satellite transmits the packet once and retransmits it $k$ times, resulting in a complexity of $O(1+k)$ for both the satellite and UE.

For all methods except benchmark 0, the satellite's complexity for cellularization and beam angle adjustment is approximately $O(n)$, where $n$ represents the number of cells. During transmission, the packet transmission to each cell contributes to a complexity of $O(n)$ for these methods.

In NTN-RM, if HAPs are present, they assist the satellite with retransmissions. Considering retransmission, the total complexity of the LEO satellite in NTN-RM is $O(n+(n-m)k)$, where $k$ is the average number of retransmissions per packet, $m$ is the number of HAPs (and $m$ does not exceed $n$). For the UE in NTN-RM, SIC must be performed, where the UE decodes and cancels the HAP signal from the received signal during retransmission. Therefore, the complexity for the UE in NTN-RM is $O(1+2k)$.

The satellites in benchmark 2 have the same complexity as NTN-RM since benchmark 2 also assumes that HAPs assist with retransmissions. For Benchmarks 1, 5, and 6, the satellites handle retransmissions directly, giving them a base complexity of $O(n+nk)$. However, Benchmarks 5 and 6 employ unique retransmission methods that influence their overall complexity. In benchmark 4, no retransmissions occur, and the satellite uses the FEC. Therefore, the complexity of the satellite and UE in benchmark 4 depends on the specific FEC scheme employed.

**TABLE 5. Complexity for each packet transmission.**

| | satellite | UE |
|---|---|---|
| Benchmark 0 | $O(1+k)$ | $O(1+k)$ |
| Benchmark 1 | $O(n+nk)$ | $O(1+k)$ |
| Benchmark 2 | $O(n+(n-m)k)$ | $O(1+k)$ |
| Benchmark 3 | $O(n+(n-m)k)$ | $O(1+2k)$ |
| Benchmark 4 | $O(nf)$ | $O(f)$ |
| Benchmark 5 | $O(n+nki)$ | $O(1+ki)$ |
| Benchmark 6 | $O(n+nkc)$ | $O(1+kc)$ |
| NTN-RM | $O(n+(n-m)k)$ | $O(1+2k)$ |
| n: number of cells; m: number of HAPs; m≤n;<br>k: average number of retransmissions per packet;<br>(k may change a lot depends on the channel condition)<br>i: the percentage of I-frame packets;<br>f: the complexity of FEC; c: the complexity of network coding | | |

### 12) COMPUTATIONAL COMPLEXITY OF NTN-RM

In addition, we include a detailed computational complexity analysis of the NTN-RM protocol. We also noted the potential processing delays based on our simulation, which was conducted using a personal computer with an 11th Gen Intel(R) Core(TM) i5-1135G7 CPU. This information is summarized in Table 6.

Before the transmission, the LEO satellite and HAPs utilize multibeam technology to establish service to the cells, with a complexity of $O(n_{cell})$, where $n_{cell}$ represents the number of cells. Since the maximum number of cells in this study is 1,171, the potential processing delay is less than 2 ms. Additionally, HAPs need to estimate the downlink channel to determine whether NOMA technology can be used, with a complexity of $O(n_{UE})$, where $n_{UE}$ represents the number of UE in the cell. The processing delay for this operation ranges from 1 to 200 ms, depending on the number of UE.

During the transmission, the LEO satellite performs queue control, as shown in Algorithm 2, to determine the transmission data rate. The satellite checks the number of packets in the queue, with a complexity of $O(l)$, where $l$ represents the

queue length. In NTN-RM, most retransmissions are handled by HAPs, minimizing the number of packets in the LEO queue, resulting in a processing delay of less than 0.1 ms in our simulation. Additionally, the LEO adjusts the beam angle during transmission, with a complexity of $O(n_{cell})$, and the processing delay for this is less than 2 ms. If the transmission angle exceeds a threshold, the LEO initiates a handover mechanism to share transmission information with the next satellite, with a complexity of $O(1)$ and a processing delay of less than 0.1 ms.

For HAPs, when they receive a NAK from a UE, they check if the NAK has already been sent by another UE. If so, they ignore duplicate NAKs, which is the uplink offloading function, with a complexity of $O(m_{NAK})$, where $m_{NAK}$ is the number of recorded NAKs in the current HAP memory. If the NAK is new, it is recorded, and the HAP performs local repair by retransmitting the corresponding packet. During local repair, HAPs check if the required packet is in their storage, with a complexity of $O(m_{packet})$, where $m_{packet}$ is the number of stored packets in the current HAP memory. In our simulation, due to memory management mechanisms and the high success rate of HAP-to-UE transmissions, the processing delays for uplink offloading and local repair are both under 0.1 ms.

**TABLE 6. Computational complexity of NTN-RM.**

| Operation | Complexity | Potential processing delay |
|---|---|---|
| **Before the transmission** | | |
| Cellularization | $O(n_{cell})$ | $< 2ms$ |
| Channel state estimation | $O(n_{UE})$ | $1 - 200ms$ |
| **During the transmission** | | |
| LEO satellite queue control | $O(l)$ | $< 0.1ms$ |
| Beam angle adjustment | $O(n_{cell})$ | $< 2ms$ |
| Uplink offloading | $O(m_{NAK})$ | $< 0.1ms$ |
| Local repair | $O(m_{packet})$ | $< 0.1ms$ |
| Handover mechanism | $O(1)$ | $< 0.1ms$ |
| $n_{cell}$: the number of cells | | |
| $n_{UE}$: the number of UE in a HAP service area | | |
| $l$: the length of the LEO queue | | |
| $m_{NAK}$: the number of recorded NAKs in the current HAP memory | | |
| $m_{packet}$: the number of stored packets in the current HAP memory | | |

## VI. CONCLUSION

In this study, we investigate the use of NTNs to support worldwide XR services, such as a worldwide multiplayer XR gaming event mentioned in Section I, and discuss the limitations of existing reliable multicast protocols. Furthermore, we propose the NTN-RM protocol, which aims to achieve lower mean latency while ensuring reliability in a network consisting of multiple LEO satellites, HAPs, and UE. This is accomplished by leveraging multibeam cellularization, HAP-based uplink offloading and local repair, a LEO satellite queue control scheme, and a beam angle adjustment and handover method.

Through simulations, we compared the system-level performance of the NTN-RM against seven benchmarks: single-beam LEO satellite direct transmission, multibeam LEO satellite direct transmission (REMP or NORM methods), HAP relay and retransmission (PGM method), NTN-RM without the handover approach, LEO direct transmission with FEC (SatRMTP method), LEO direct transmission with limited retransmission (RMM method), and repeat transmission with network coding retransmission (NC-MAC method). The results demonstrate that the NTN-RM:

1) Improves the mean latency of reliable multicast transmission by at least 3.5% to 29.4% compared with most benchmarks,
2) Maintains a stable 99.99% packet arrival rate, regardless of the number of satellites and UE,
3) Increases the network capacity of the NTN,
4) Requires higher HAP EIRP but lower HAP Rx gain compared to the HAP relay and retransmission method (benchmark 2),
5) Reduces the impact of LEO satellite movement and handover on streaming transmission,
6) Results in fewer packets retransmitted by LEO satellites, leading to 49.3% to 79.1% lower jitter than the benchmarks,
7) Performs well in urban and dense urban scenarios compared to the benchmarks.

Moreover, according to the latency results in Section V-C1, it is recommended that LEO constellations have a satellite density larger than $5 \times 10^{-6}$ LEO/$km^2$ to support multicast applications in NTNs. Furthermore, the integration of FEC into NTN-RM while retaining retransmission is expected to decrease the packet loss rate, thereby alleviating the demand for retransmission. However, this is expected to significantly reduce the network throughput based on the throughput results presented in Section V-C3. Given that the latency is already sufficiently low and the reliability is sufficiently high with the assistance of HAPs in NTN-RM, the authors believe that incorporating FEC is unnecessary, as it would reduce system throughput.

For real-world deployment feasibility, HAPs must be able to remain stationed above a specific area for extended periods. Experiments conducted by the companies mentioned in the introduction section have already made progress toward achieving this. The transmit power required for HAPs in this study is lower than that of satellites, indicating that the current technology is sufficient to meet the necessary antenna requirements. In terms of energy consumption, it would be beneficial for HAPs to use renewable energy sources, such as solar power, to enhance long-term sustainability. The main concern with HAP deployment is cost, as it may seem that a large number of HAPs are required under satellite coverage. However, although satellites provide global coverage, HAPs can be selectively deployed in areas where they are most needed. Considering that land accounts for only 30% of the Earth's surface, and not all land areas are inhabited, the actual number of required HAPs is lower than expected. To further reduce deployment costs, HAPs can be prioritized in dense urban and urban areas where satellite direct downlinks are less effective, as demonstrated by our evaluation.

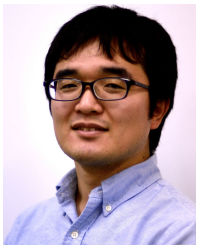
The proposed approach is compatible with any LEO satellite-based NTN structure, regardless of the number of

HAPs or UE. Theoretically, HAPs can be replaced with UAVs. Future work could consider the power consumption of LEO satellites or Doppler shifts, extend the idea of considering HAPs and UAVs together, and integrate NTNs with terrestrial networks. Furthermore, the effect of the handover mechanism on retransmission and protocols that account for this effect can be explored. Moreover, since security and privacy are critical in NTNs involving satellites and HAPs, future research could focus on addressing potential challenges such as eavesdropping, spoofing, and denial-of-service (DoS) attacks in NTNs.

## REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.
[2] G. Giambene, S. Kota, and P. Pillai, "Satellite-5G integration: A network perspective," *IEEE Netw.*, vol. 32, no. 5, pp. 25–31, Sep. 2018.
[3] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, "Joint HAP access and LEO satellite backhaul in 6G: Matching game-based approaches," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1147–1159, Apr. 2021.
[4] C. Amatetti, R. Campana, A. Georganaki, and A. Vanelli-Coralli, "Neural network based non orthogonal random access for 6G NTN-IoT," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2022, pp. 1389–1394.
[5] C. Amatetti, M. Conti, A. Guidotti, and A. Vanelli-Coralli, "NB-IoT random access procedure via NTN: System level performances," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 2381–2386.
[6] T. Kellermann, R. P. Centelles, D. Camps-Mur, R. Ferrús, M. Guadalupi, and A. C. Augé, "Novel architecture for cellular IoT in future non-terrestrial networks: Store and forward adaptations for enabling discontinuous feeder link operation," *IEEE Access*, vol. 10, pp. 68922–68936, 2022.
[7] H. Chougrani, S. Kisseleff, W. A. Martins, and S. Chatzinotas, "NB-IoT random access for nonterrestrial networks: Preamble detection and uplink synchronization," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 14913–14927, Aug. 2022.
[8] B. Manzoor, A. Al-Hourani, and B. A. Homssi, "Improving IoT-over-satellite connectivity using frame repetition technique," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 736–740, Apr. 2022.
[9] *Study on Narrow-Band Internet of Things (NB-IoT)/Enhanced Machine Type Communication (eMTC) Support for Non-Terrestrial Networks (NTN) (Release 17)*, Standard Specification: TR 36.763, 3GPP, 2021.
[10] *Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 17)*, Standard TR 38.901 3GPP, 2022.
[11] *Study on Satellite Access Phase 3 (Release 19)*, Standard TR 22.865 3GPP, 2022.
[12] W.-S. Lim, D.-W. Kim, and Y.-J. Suh, "Design of efficient multicast protocol for IEEE 802.11n WLANs and cross-layer optimization for scalable video streaming," *IEEE Trans. Mobile Comput.*, vol. 11, no. 5, pp. 780–792, May 2012.
[13] J. Gemmell, T. Montgomery, T. Speakman, and J. Crowcroft, "The PGM reliable multicast protocol," *IEEE Netw.*, vol. 17, no. 1, pp. 16–22, Jan. 2003.
[14] J. P. Macker, C. Bormann, M. J. Handley, and B. Adamson, *NACK-Oriented Reliable Multicast (NORM) Transport Protocol*, document RFC 5740, Nov. 2009.
[15] M. K. Afzal, B.-S. Kim, and S. W. Kim, "Efficient and reliable MPEG-4 multicast MAC protocol for wireless networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1026–1035, Mar. 2015.
[16] H. Mosavat-Jahromi, Y. Li, L. Cai, and L. Lu, "NC-MAC: Network coding-based distributed MAC protocol for reliable beacon broadcasting in V2X," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
[17] X. Zhu, C. Jiang, L. Kuang, N. Ge, and J. Lu, "Non-orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2253–2267, Oct. 2017.
[18] T. Z. H. Ernest, A. S. Madhukumar, R. P. Sirigina, and A. K. Krishna, "NOMA-aided UAV communications over correlated Rician shadowed fading channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 3103–3116, 2020.
[19] G. Zhang, Z. Han, H. Xin, X. Wei, and B. Di, "Distortion minimization for multimedia transmission in NOMA HAP-UAV integrated aerial access networks," *Chin. J. Aeronaut.*, vol. 35, pp. 81–94, Sep. 2022.
[20] F. Adachi, R. Takahashi, H. Matsuo, S. Xia, C. Ge, and Q. Chen, "Cellular distributed MU-MIMO communication system-architecture design and interference countermeasure," IEICE, Tokyo, Japan, Tech. Rep. 121.391, Mar. 2022, pp. 96–101.
[21] F. Adachi, R. Takahashi, and H. Matsuo, "Enhanced interference coordination and radio resource management for 5G advanced ultra-dense RAN," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.
[22] R. Takahashi, H. Matsuo, and F. Adachi, "Performance of enhanced interference coordination using multi-layered clustering for 5G advanced ultra-dense RAN," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–5.
[23] M. Elmahallawy, T. Luo, and K. Ramadan, "Communication-efficient federated learning for LEO constellations integrated with HAPs using hybrid NOMA-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1097–1114, May 2024.
[24] Z. Jia, M. Sheng, J. Li, and Z. Han, "Toward data collection and transmission in 6G space–air–ground integrated networks: Cooperative HAP and LEO satellite schemes," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 10516–10528, Jul. 2022.
[25] Y. Li, N. Deng, and W. Zhou, "A hierarchical approach to resource allocation in extensible multi-layer LEO-MSS," *IEEE Access*, vol. 8, pp. 18522–18537, 2020.
[26] S. Wang, Y. Li, Q. Wang, M. Su, and W. Zhou, "Dynamic downlink resource allocation based on imperfect estimation in LEO-HAP cognitive system," in *Proc. 11th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2019, pp. 1–6.
[27] T. Asfour-Block and A. Serhrouchni, "Reliable multicast over satellite networks," in *Proc. 10th IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2005, pp. 698–703.
[28] F. Rinaldi, A. Tropeano, S. Pizzi, A. Molinaro, and G. Araniti, "Dynamic MBSFN beam area formation in 6G multibeam non-terrestrial networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 5, pp. 3760–3774, Oct. 2022.
[29] C. Ge, N. Wang, I. Selinis, J. Cahill, M. Kavanagh, K. Liolis, C. Politis, J. Nunes, B. Evans, Y. Rahulan, N. Nouvel, M. Boutin, J. Desmauts, F. Arnal, S. Watts, and G. Poziopoulou, "QoE-assured live streaming via satellite backhaul in 5G networks," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 381–391, Jun. 2019.
[30] E. Zhou, M. Kobayashi, T. Fujihashi, M. A. Alim, S. Saruwatari, M. Nishi, and T. Watanabe, "Reliable multicast streaming for hierarchical non-terrestrial network," in *Proc. 9th Int. Conf. Wireless Netw. Mobile Commun. (WINCOM)*, Oct. 2022, pp. 1–6.
[31] R. Ali, Y. B. Zikria, A. K. Bashir, S. Garg, and H. S. Kim, "URLLC for 5G and beyond: Requirements, enabling incumbent technologies and network intelligence," *IEEE Access*, vol. 9, pp. 67064–67095, 2021.
[32] F. Vidal, H. Legay, G. Goussetis, T. Ströber, and J.-D. Gayrard, "Benchmark of MEO multibeam satellite adaptive antenna and payload architectures for broadband systems," in *Proc. 10th Adv. Satell. Multimedia Syst. Conf. 16th Signal Process. Space Commun. Workshop (ASMS/SPSC)*, Oct. 2020, pp. 1–8.
[33] A. Zbiciak and T. Markiewicz, "A new extraordinary means of appeal in the Polish criminal procedure: The basic principles of a fair trial and a complaint against a cassatory judgment," *Access Justice Eastern Eur.*, vol. 6, no. 2, pp. 25–42, Mar. 2023.
[34] O. Belce, "Comparison of advanced modulation schemes for LEO satellite downlink communications," in *Proc. Int. Conf. Recent Adv. Space Technol. (RAST)*, Istanbul, Turkey, 2003, pp. 432–437.
[35] T. Hanada, K. Fujisaki, and M. Tateiba, "Average bit error rate for satellite downlink communications in Ka-band under atmospheric turbulence given by Gaussian model," in *Proc. Asia–Pacific Microw. Conf.*, Dec. 2009, pp. 1092–1095.
[36] Z. Li, S. Wang, S. Han, and C. Li, "Non-orthogonal broadcast and unicast joint transmission for multibeam satellite system," *IEEE Trans. Broadcast.*, vol. 69, no. 3, pp. 647–660, Sep. 2023, doi: 10.1109/TBC.2023.3267239.
[37] J. Li, Z. Li, and D. Li, "Performance comparison of H.264 and H.265 encoders for 4K video sequences," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Oct. 2016, pp. 531–536.

**ENPING ZHOU** (Student Member, IEEE) received the B.E. degree from Tianjin University, China, in 2020, and the M.E. degree from Osaka University, Japan, in 2024, where he is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology. His research interests include wireless networks and non-terrestrial networks, especially medium access control and routing.



**MAKOTO KOBAYASHI** (Member, IEEE) received the M.E. and Ph.D. degrees from Osaka University, in 2016 and 2019, respectively. He was an Assistant Professor with Hiroshima City University, Japan, from 2019 to 2024. He has been a Lecturer with Hiroshima City University, since 2024. His research interest includes wireless networks and communications. In particular, he has a significant interest in the MAC protocol to support diversified requirements and services. He was a Research Fellow (DC2) of Japan Society for the Promotion of Science, from 2017 to 2019. He is a member of IEEE ComSoc, IEEE VTS, IPSJ, and IEICE.

**TAKUYA FUJIHASHI** (Member, IEEE) received the B.E. and M.S. degrees from Shizuoka University, Japan, in 2012 and 2013, respectively, and the Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan, in 2016. From 2014 to 2015, he was an Intern with Mitsubishi Electric Research Laboratories (MERL), Electronics and Communications Group. He is currently an Assistant Professor with the Graduate School of Information Science and Technology, Osaka University. His research interests include video compression and communications, with a focus on multi-view video coding and streaming. From 2014 to 2016, he was a Research Fellow (DC1) of Japan Society for the Promotion of Science. He was a Research Fellow (PD) of Japan Society for the Promotion of Science, in 2016. He was selected as one of the Best Paper Candidates in the IEEE International Conference on Multimedia and Expo (ICME) 2012.

**MD. ABDUL ALIM** (Member, IEEE) received the B.Sc. degree in electronics and communication engineering (ECE) from Khulna University, Bangladesh, in 2003, the M.Eng. degree in ECE from Chongqing University, China, in 2013, and the Ph.D. degree from Osaka University, Japan, in 2018. He is currently a Faculty Member with the Electronics and Communication Engineering Discipline, Khulna University. He is on study leave to conduct postdoctoral research with Aalborg University, Denmark. He has professional experience in cellular wireless communications as an Executive Engineer with SIEMENS Bangladesh. His research interests include MAC design for wireless communications, wireless power transmissions and estimations, interference mitigation in Wi-Fi, and future mobile communication.

**SHUNSUKE SARUWATARI** (Member, IEEE) received the B.E. degree from The University of Electro-Communications, Japan, in 2002, and the M.S. and Ph.D. degrees from The University of Tokyo, Japan, in 2004 and 2007, respectively. In 2007, he was a Visiting Researcher with Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana–Champaign. From 2008 to 2011, he was a Research Associate with the Research Center for Advanced Science and Technology, The University of Tokyo. From 2012 to 2015, he was a Tenure-Track Assistant Professor with the Graduate School of Informatics, Shizuoka University, Japan. He is currently an Associate Professor with the Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include the areas of wireless networks, sensor networks, and systems software. He is a member of ACM, IPSJ, and IEICE.

**MASAHIRO NISHI** (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communications engineering from Osaka University, in 1995, 1997, and 1999, respectively. He joined Hiroshima City University as a Research Associate, in 1999. In 2005, he became an Associate Professor with Hiroshima City University and he was appointed as a Professor with Hiroshima City University in 2016. He has been doing research on radio sciences, radio wave propagation, and wireless network systems. He is a member of IEICE and IPSJ.

**TAKASHI WATANABE** (Life Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Osaka University, Japan, in 1982, 1984, and 1987, respectively. He joined the Faculty of Engineering, Tokushima University, in 1987, and moved to the Faculty of Engineering, Shizuoka University, in 1990. He was a Visiting Researcher with the University of California at Irvine, from 1995 to 1996. He has been a Professor with the Graduate School of Information Science and Technology, Osaka University, since 2013. His research interests include mobile networking, ad hoc sensor networks, the Internet of Things/M2M networks, and intelligent transport systems, especially medium access control and routing. He is a member of IPSJ and IEICE. He has served on program committees for many networking conferences, such as the IEEE, ACM, IPSJ, and IEICE.