

# Latency-Aware Offloading in Integrated Satellite Terrestrial Networks

Wiem Abderrahim <sup>1</sup>, Osama Amin <sup>2</sup>, Mohamed-Slim Alouini <sup>2</sup>, and Basem Shihada <sup>2</sup>

<sup>1</sup>King Abdullah University of Science and Technology (KAUST)





<sup>2</sup>Affiliation not available

October 30, 2023

## Abstract

Next-generation communication networks are expected to integrate newly-used technologies in a smart way to ensure continuous connectivity in rural areas and to alleviate the traffic load in dense regions. The prospective access network in 6G should hinge on satellite systems to take advantage of their wide coverage and high capacity. However, adopting satellites in 6G could be hindered because of the {additional latency introduced}, which is not tolerable by all traffic types. Therefore, we propose a traffic offloading scheme that integrates both the satellite and terrestrial networks to smartly allocate the traffic between them while satisfying different traffic requirements. Specifically, the proposed scheme offloads the Ultra-Reliable Low Latency Communication (URLLC) traffic to the terrestrial backhaul to satisfy its stringent latency requirement. However, it offloads the enhanced Mobile Broadband (eMBB) traffic to the satellite since eMBB needs high data rates but is not always sensitive to delay. Our scheme is shown to reduce the transmission delay of URLLC packets, decrease the number of dropped eMBB packets, and hence improve the network's availability. Our findings highlight that the inter-working between satellite and terrestrial networks is crucial to mitigate the expected high load on the limited terrestrial capacity.

# Latency-Aware Offloading in Integrated Satellite Terrestrial Networks

Wiem Abderrahim , Osama Amin , Mohamed-Slim Alouini  and Basem Shihada 

\*

## Abstract

Next-generation communication networks are expected to integrate newly-used technologies in a smart way to ensure continuous connectivity in rural areas and to alleviate the traffic load in dense regions. The prospective access network in 6G should hinge on satellite systems to take advantage of their wide coverage and high capacity. However, adopting satellites in 6G could be hindered because of the additional latency introduced, which is not tolerable by all traffic types. Therefore, we propose a traffic offloading scheme that integrates both the satellite and terrestrial networks to smartly allocate the traffic between them while satisfying different traffic requirements. Specifically, the proposed scheme offloads the Ultra-Reliable Low Latency Communication (URLLC) traffic to the terrestrial backhaul to satisfy its stringent latency requirement. However, it offloads the enhanced Mobile Broadband (eMBB) traffic to the satellite since eMBB needs high data rates but is not always sensitive to delay. Our scheme is shown to reduce the transmission delay of URLLC packets, decrease the number of dropped eMBB packets, and hence improve the network's availability. Our findings highlight that the inter-working between satellite and terrestrial networks is crucial to mitigate the expected high load on the limited terrestrial capacity.

**Keywords**— Enhanced Mobile Broadband (eMBB), Integrated Satellite Terrestrial Networks (ISTN), Traffic offloading, Ultra-reliable low latency communication (URLLC).

## I Introduction

With the approaching advent of the 5th generation of mobile cellular networks (5G), the next generation of mobile cellular networks (6G) is already required to afford more sophisticated features in terms of coverage and capacity [13]. 6G should not only ensure global connectivity by connecting under-served areas where access to the Internet is limited or absent, but it should also mitigate the limitations in the capacity of terrestrial links, especially with the increasing data demand in dense networks. Of the mix of access technologies in 6G, satellite communications will play a pivotal role due to their wide

---

\*This work was supported in part by the Center of Excellence for NEOM Research at KAUST. The authors are with the Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Makkah Prov., Saudi Arabia (e-mail: wiem.abderrahim, osama.amin, slim.alouini, basem.shihada@kaust.edu.sa)

coverage. Notably, the new generations of low earth orbit (LEO) satellites are gaining a high-visibility in both industry and academia because they guarantee higher capacities (almost 1 Tbps [14, 15]) and lower delays (20-50 milliseconds [15]). In industry, different mega-constellations of LEOs are being manufactured and partially launched into space lately. For example, the aerospace manufacturer SpaceX has an ambitious project to cover the Earth with around 12000 LEO satellites [15]. SpaceX is racing against Amazon and OneWeb, which plan to launch 3236 satellites (Project Kuiper) and 650 satellites, respectively [15, 18]. They all aim to widen the terrestrial network's connectivity and, more crucially to extend its limited capacity [4].

In academia, researchers also are exploring different approaches to extend the limited capacities of terrestrial networks [22]. Recently, several efforts focused on traffic offloading in integrated satellite and terrestrial networks (ISTN) [11, 16, 21, 22, 26, 27]. For instance, the authors of [27] addressed the offloading problem in satellite mobile edge computing (SMEC). They studied and compared different offloading locations, such as the terrestrial station, the LEO satellite, and the terrestrial gateway in SMEC [27]. The terrestrial gateway was the recommended offloading alternative because it guarantees lower complexity in terms of coverage and maintenance. In [26], a double edge computing architecture was considered where the edge servers were deployed in satellite and terrestrial networks. Moreover, a cost matching algorithm was proposed to allocate the suitable edge servers to the offloaded tasks while optimizing the energy consumption and the offloading delay. In [11], a space-air-ground integrated network (SAGIN) architecture was studied where the aerial network nodes served as flying edge servers in the air segment and the LEO satellites in the space segment connected the Internet of things (IoT) devices with the cloud servers. SAGIN architecture helps the remote IoT applications to decide on the typical offloading location, namely locally, at the air segment, or at the space segment. In [16], a traffic offloading scheme in ISTN was proposed to maximize the number of the accommodated users and their sum rate under a dynamic backhaul capacity constraint. The proposed scheme schedules the resources available in both networks to optimize users' association and to increase the dynamic backhaul capacity. In [21], a content offloading/caching scheme in ISTN was proposed. According to this scheme, the macro base station (MBS) controls the traffic offloading by allocating the necessary power and assigning the content blocks to the Access Points and to the satellite. This content can be sent to the end users as unicast through the access points and/or it can be broadcasted through the satellite [21].

In spite of these tremendous efforts in studying traffic offloading in ISTN, we note two main shortcomings in the literature. Firstly, most of the present studies, except [16], disregarded that a concrete cooperation between terrestrial and satellite networks during offloading could yield more effective solutions. This cooperation becomes imperative especially because next generation networks will use the terahertz (THz) band to expand the bandwidth and to satisfy future services [13, 17]. Since, current satellite services use the THz band, future terrestrial networks are expected to use the same band [17]; thus future collaboration schemes between the terrestrial and satellite networks. Yet, both networks coexist, but they do not cooperate to share task management. Secondly, the offloaded traffic is not typically selected, and the traffic requirements are widely overlooked during that. This shortcoming becomes more crucial if we take into consideration the heterogeneous use cases and traffic types defined by 5G.

Specifically, 5G should address the requirements of the eMBB use case, whose services exchange large payloads over an extended time interval and thus there is a need for high throughput and bandwidth links with moderate reliability levels [9, 25]. 5G should also answer the requirements of the URLLC use case, whose services have intermittent transmissions between a low number of connected devices and need high reliability and low latency stringently [9, 25]. To address this issue, several research studied the multiplexing of eMBB services and URLLC services [5–8, 19, 25]. For instance, the authors of [25] exploited the advantage of orthogonal slicing to multiplex eMBB and URLLC services, where they proposed a framework that maximizes the slice requests’ admission and the operator’s revenues. The admission of both slice requests depends on the available resources and the required quality-of-service (QoS) such that the eMBB slice uses the multicast transmission and the URLLC slice uses the unicast transmission. A detailed analysis of non-orthogonal multiplexing of eMBB and URLLC services in a fog-radio access network was conducted in [19]. In this regard, the proposed architecture processed the URLLC traffic in the fog nodes to satisfy its latency constraint. However, the eMBB traffic was processed in the Cloud-RAN to gain high spectral efficiency [19]. Various joint eMBB and URLLC scheduler’s models were investigated in [8], where each scheduler selects and retrieves some resources from the eMBB traffic and allocates them to the URLLC traffic given its stringent latency constraint by using the superposition/puncturing technique. This technique should satisfy the delays tolerable by the URLLC service and mitigate the rate loss of the eMBB service. The puncturing technique was also used in [5–7] to multiplex the eMBB service and the URLLC service. The authors formulate the resources allocation as 2-Dimensions Hopfield neural network and the resources scheduling as an optimization problem with a chance constraint [6]. Unfortunately, all the aforementioned research overlooks the severe capacity limitation from which terrestrial backhauls are suffering. Moreover, it rules out the dense network scenario since a modest number of bases stations and users was set up.

To overcome the previously discussed shortcomings, we propose to manage the multiplexing of the eMBB traffic and the URLLC traffic in ISTN. We advocate that both networks, satellite and terrestrial, cooperate such that each traffic is precisely steered towards the transport network that meets its requirements. To the best of our knowledge, our paper is the first to propose a scheme for traffic offloading in ISTN to satisfy heterogeneous traffic requirements. Our main contribution is to consider the offloading of heterogeneous traffic types in ISTN. Specifically, our scheme offloads the URLLC traffic in the terrestrial backhaul to satisfy its stringent requirement in terms of latency and the eMBB traffic in the satellite backhaul since it needs high data rate but is not very sensitive to delay. Our scheme aims to relieve the constraints on the terrestrial backhaul in the context of dense networks. To this end, we design the proposed joint ISTN system in two steps; first, we optimize the scheduling of the eMBB traffic and the URLLC traffic by using the puncturing technique, while taking into account the terrestrial backhaul capacity limitation. Second, we adjust the association between the satellite and the small base stations (SBS) while prioritizing URLLC traffic over eMBB traffic. We then evaluate the performance of our scheme in three satellite systems: Telsat LEO, OneWeb, and Starlink. As proved by the results, our scheme reduces the transmission delay of the URLLC packets and decreases the number of dropped eMBB packets; hence it improves the availability of the network.

The remainder of this paper is structured as follows : In section II, our system model is introduced. In section III and section IV, our model is formalized in terrestrial networks and in ISTN, respectively. In section V, some performance metrics of our proposed scheme are derived analytically. In section VI, the numerical results are discussed. Finally, the paper is concluded in Section VII.

## II System Model

Our system model consists of a satellite network and a terrestrial network as depicted in Fig. 1. The satellite constellation consists of a set  $\mathcal{S} = \{s_1, s_2, \dots, s_M\}$  of  $M$  LEO satellites. In the scope of this paper, we consider the downlink in the satellite network with a link capacity per beam  $C_{\text{Sat}}$ . The used band in the satellite network is the Ka-band [16]. The terrestrial network consists of a set  $\mathcal{BS} = \{bs_1, bs_2, \dots, bs_N\}$  of  $N$  small base stations in small cells  $i$  that serve each a set  $\mathcal{U}$  of  $U$  UEs. The terrestrial backhaul between the small base station and the macro base station is characterized with a link capacity  $C_{\text{Ter}}$ . The used band in the terrestrial network is the C-band [16].

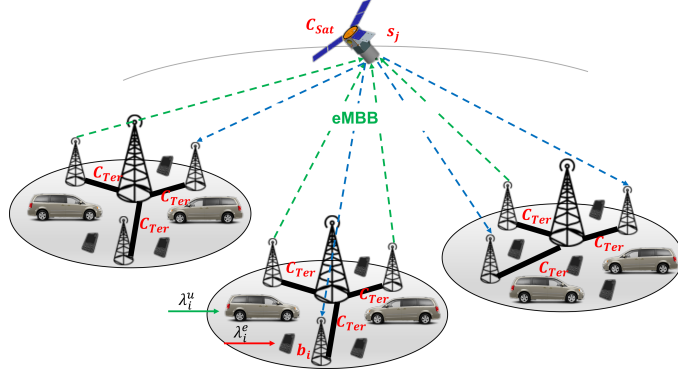


Figure 1: System model of QoS sensitive offloading in ISTN.

### II.A. Traffic Model

Let us assume that the UE belongs to either eMBB slice or URLLC slice such that  $U = U_e \cup U_u$  where  $U_e = \{1, \dots, U_1\}$  is the set of  $U_1$  eMBB users and  $U_u = \{1, \dots, U_2\}$  is the set of  $U_2$  URLLC users; such that  $U_1 \geq U_2$ . The eMBB traffic is generated by applications that exchange large payloads over an extended time interval [25]. Therefore, the inter-arrival time can be modeled with an exponential distribution whose events occur continuously and independently. Accordingly, the eMBB flow requests follow a Poisson process with arrival rate  $\lambda_i^e$ . However, the URLLC traffic is generated by applications with intermittent transmissions that exchange more or less important payload during short periods of time [25]. Therefore, it can be modeled with a power-law distribution namely a Pareto distribution with arrival rate  $\lambda_i^u$  and parameters  $a$  and  $x_m$  [1, 3, 5, 6, 8, 9, 23]. Flow sizes are independently and identically distributed with mean  $\frac{1}{\mu_i^e}$  for eMBB slice and  $\frac{1}{\mu_i^u}$  for URLLC slice, such that  $\frac{1}{\mu_i^e}$  and  $\frac{1}{\mu_i^u}$  represent the packet size for each traffic type, respectively. Given the delay constraints, the small base station sends the URLLC traffic in the terrestrial backhaul and offloads the eMBB traffic to the satellite.

## II.B. Base Station Load

Each small base station  $bs_i$  is characterized by a specific bandwidth  $W_i$  and a downlink transmit power  $P_i$ . We assume that the transmit power of the small base station is the same for all its associated users. The signal-to-noise ratio (SNR) experienced by a user  $v$  served by the small base station  $bs_i$  is given by,

$$\text{SNR}_{iv} = \frac{P_i g_i^v}{N_p + \sum_{i' \in \{1, \dots, N\}/i} P_{i'} g_{i'}^v}, \quad (1)$$

where  $g_{iv}$  is the channel gain between small base station  $bs_i$  and user  $v$ ,  $N_p$  is the noise power of AWGN, and  $\sum_{i' \in \{1, \dots, N\}/i} P_{i'} g_{i'}^v$  is the interference caused by the other small base stations, where we consider it as negligible in the scope of this paper. Indeed, we assume that the proposed system implements the necessary techniques of inter-cell interference coordination to avoid or neglect it (e.g. frequency reuse).

We denote the frequency resources allocated to eMBB user  $v$  in one slot by  $b_v$  and the punctured frequency resources from eMBB user  $v$  to the URLLC traffic due to the hard latency constraints by  $f_v$ . Thus, the SNR experienced by a URLLC user  $v$  is given by

$$\text{SNR}_{iv}^u = \frac{P_i g_i^v}{N_0 f_v}, \quad (2)$$

where  $N_0$  is the noise power density.

On the other hand, the SNR experienced by an eMBB user  $v$  is given by,

$$\text{SNR}_{iv}^e = \frac{P_i g_i^v}{N_0 (b_v - f_v)}. \quad (3)$$

Let  $\gamma := \frac{P_i g_i^v}{N_0}$ , then the maximum data transmission rate of an URLLC user  $v$  is given by [24], [5]:

$$r_{iv}^u = f_v \log_2 \left( 1 + \frac{\gamma}{f_v} \right). \quad (4)$$

The maximum data transmission rate of an eMBB user  $v$  is given by [5, 24]:

$$r_{iv}^e = (b_v - f_v) \log_2 \left( 1 + \frac{\gamma}{b_v - f_v} \right). \quad (5)$$

The total load at the  $i$ -th cell is given by [24],

$$L_i = L_i^u + L_i^e, \quad (6)$$

where  $L_i^u$  and  $L_i^e$  are the load generated by URLLC and eMBB slices, respectively. First, we write  $L_i^u$  as

$$L_i^u = \sum_{v=1}^{U_2} r_{iv}^u = \sum_{v=1}^{U_2} f_v \log_2 \left( 1 + \frac{\gamma}{f_v} \right). \quad (7)$$

On the other hand, we write  $L_i^e$  as

$$L_i^e = \sum_{v=1}^{U_1} r_{iv}^e = \sum_{v=1}^{U_1} (b_v - f_v) \log_2 \left( 1 + \frac{\gamma}{b_v - f_v} \right). \quad (8)$$

### II.C. Admission Control on URLLC Users

The capacity limit on the terrestrial link leads us to define an admission control mechanism on URLLC users such that the URLLC demand exceeding the terrestrial system capacity ( $C_{\text{Ter}}$ ) is blocked [8], thus we have

$$\lambda_i^u \leq C_{\text{Ter}}; \forall i \in \{1, \dots, N\}. \quad (9)$$

### II.D. Downlink Satellite Data Rate

We assume that each small base station can offload its tasks to one satellite at a time.  $\eta_{ij}$  denotes the association of the  $i$ -th small base station, i.e.  $bs_i$ , to the  $j$ -th satellite, i.e.,  $s_j$ , such that

$$\eta_{ij} = \begin{cases} \neq 0 & bs_i \text{ can offload tasks to } s_j \\ 0 & \text{otherwise} \end{cases}. \quad (10)$$

The small base station-satellite association rule can be formulated as

$$\begin{cases} \eta_{ij} \in [0, 1], & \forall i \in \{1, \dots, N\}, j \in \{1, \dots, M\} \\ \sum_{j=1}^M \eta_{ij} = 1, & \forall i \in \{1, \dots, N\} \end{cases} \quad (11)$$

where the backhaul link between small base station  $bs_i$  and satellite  $s_j$  is characterized by a maximum achievable data rate  $R_{ij}$  given by [16]

$$R_{ij} = W_j \log_2 \left( 1 + \frac{C}{N_0 W_j} \right), \quad (12)$$

where  $W_j$  is the bandwidth dedicated to each satellite and shared by its associated small base stations and  $\frac{C}{N_0}$  is the carrier signal power to noise power density of the satellite. We should note that  $R_{ij} \leq C_{\text{Sat}}$ ; where  $\frac{C}{N_0}$  and  $C_{\text{Sat}}$  are fixed by the satellite manufacturer.

In the rest of the paper, we focus on a single satellite of the network and we denote the data rate between the fixed satellite and the small base station  $bs_i$  by  $R_j$ , and the offloaded amount of eMBB traffic between from the satellite and the  $i$ -th small base station, i.e.,  $bs_i$  by  $\beta_i$ .

## III Resource Allocation in the Terrestrial Networks

We start first by scheduling the eMBB traffic and the URLLC traffic using the puncturing technique which superimposes the URLLC traffic upon its arrival on the ongoing eMBB transmission [7, 8]. Our

aim is to characterize the optimal resource allocation for URLLC users while maximizing the data rate of eMBB users and taking into account the limited backhaul capacity. Thus, we formulate the resource allocation problem as follows

$$\max_{f_v} \sum_{v=1}^{U_1} (b_v - f_v) \log_2 \left( 1 + \frac{\gamma}{b_v - f_v} \right) \quad (13a)$$

$$\text{s.t.} \quad \Pr \left( \sum_{v=1}^{U_2} f_v \log_2 \left( 1 + \frac{\gamma}{f_v} \right) < \lambda_i^u \right) \leq \epsilon \quad (13b)$$

$$L_i^u + L_i^e \leq C_{\text{Ter}} \quad \forall i \in \{1, \dots, N\} \quad (13c)$$

$$\sum_{v=1}^{U_2} f_v \leq W_i^{\text{bs}}, \quad (13d)$$

where the constraint (13b) expresses an acceptable outage probability of URLLC traffic, where  $\epsilon$  is a small positive constant,  $\epsilon \ll 1$ , which guarantees low URLLC outage probability below a negligible threshold. Thus, the constraint (13b) guarantees that the dropped URLLC packets during the admission control are below a negligible outage probability threshold  $\epsilon$  to ensure a URLLC packets transmission reliability. The constraint (13c) guarantees that the accepted eMBB load and URLLC load are limited by the terrestrial link capacity, i.e.,  $C_{\text{Ter}}$ . It is worthy emphasizing that extra eMBB packets are dropped in case of overload. The constraint (13d) guarantees that the sum of the allocated frequency resources to the URLLC users is less than the total bandwidth  $W_i^{\text{bs}}$  of the small base station  $bs_i$ .

We can simplify constraint (13b) using the cumulative distribution function (CDF) of  $\lambda_i^u$ ,  $F_D(\cdot)$ , as follows

$$\sum_{v=1}^{U_2} f_v \log_2 \left( 1 + \frac{\gamma}{f_v} \right) \geq F_D^{-1}(1 - \epsilon). \quad (14)$$

Since  $\lambda_i^u$  follows a Pareto distribution with parameters  $a$  and  $x_m$ , we express (14) as

$$\left( \sum_{v=1}^{U_2} f_v \log_2 \left( 1 + \frac{\gamma}{f_v} \right) \right)^a \geq \frac{x_m^a}{\epsilon}, \quad (15)$$

where  $a$  is a constant that determines the distribution shape and  $x_m$  is also a positive scale distribution parameter. In this work, we set  $x_m = 1$  because we assume that the URLLC traffic exhibits self-similarity over time scales of one second and above [12]. In addition, we set  $a = 1$  to reflect a wider variability of the URLLC traffic in the duration of packet generation (ON period) and inter-packets generation (OFF period), which enables us to write (13b) in a convex form. However, the constraint (13c) is a non-convex function in  $f_v$ , which makes the problem (23) non-concave. To deal with this problem, we propose expressing (13c) as a difference of convex (DC) functions. Therefore, we decompose (13c) as the difference between two convex functions  $\mathcal{F}$  and  $\mathcal{G}$

$$\mathcal{F}(f_v) - \mathcal{G}(f_v) \leq C_{\text{Ter}}, \quad (16)$$



such that

$$\mathcal{F}(f_v) = \sum_{v=1}^{U_2} f_v \log_2(f_v + \gamma) + \sum_{v=1}^{U_1} (b_v - f_v) \log_2(b_v - f_v + \gamma) \quad (17)$$

and

$$\mathcal{G}(f_v) = \sum_{v=1}^{U_2} f_v \log_2(f_v) + \sum_{v=1}^{U_1} (b_v - f_v) \log_2(b_v - f_v). \quad (18)$$

First, we prove that  $\mathcal{F}$  and  $\mathcal{G}$  are convex. Indeed,  $\mathcal{F}$  is convex since

$$\mathcal{F}''(f_v) = \sum_{v=1}^{U_2} \frac{2\gamma + f_v}{(\gamma + f_v)^2} + \sum_{v=1}^{U_1} \frac{2b_v + \gamma - f_v}{(\gamma + b_v - f_v)^2} \geq 0 \quad (19)$$

$\forall f_v \in [0, b_v]$ ,  $\mathcal{G}$  is also convex since  $\mathcal{G}(f_v) = \mathcal{F}(f_v)|_{\gamma=0}$ .

Then, we approximate  $\mathcal{G}(f_v)$  using the first order representation of Taylor series at  $f'_v$ , which gives the following lower bound as follows [10],

$$\mathcal{G}(f_v) \leq \mathcal{G}(f'_v) + \frac{\partial \mathcal{G}}{\partial f_v}(f'_v)(f_v - f'_v), \quad (20)$$

where  $\frac{\partial \mathcal{G}}{\partial f_v}$  is the first derivative of  $\mathcal{G}(f_v)$  with respect to  $f_v$ , and is written as

$$\frac{\partial \mathcal{G}}{\partial f_v} = \sum_{v=1}^{U_2} (1 + \log_2(f_v)) - \sum_{v=1}^{U_1} (1 + \log_2(b_v - f_v)). \quad (21)$$

Thus, we can write (20) as follows

$$\mathcal{G}(f_v) \leq \sum_{v=1}^{U_2} (1 + \log_2(f'_v)) - \sum_{v=1}^{U_1} (1 + \log_2(b_v - f'_v)) + \left[ \sum_{v=1}^{U_2} (1 + \log_2(f'_v)) - \sum_{v=1}^{U_1} (1 + \log_2(b_v - f'_v)) \right] (f_v - f'_v). \quad (22)$$

Therefore, a simplified approximated problem of (13) can be expressed by using the aforementioned first-order approximation (22), which yields an upper bound for the left-hand-side (LHS) of the original

constraint (13c). Thus, we can rewrite the resource allocation as follows,

$$\max_{f_v} \sum_{v=1}^{U_1} (b_v - f_v) \log_2 \left( 1 + \frac{\gamma}{b_v - f_v} \right) \quad (23a)$$

$$\text{s.t.} \quad - \sum_{v=1}^{U_2} f_v \log_2 \left( 1 + \frac{\gamma}{f_v} \right) \leq - \frac{x_m}{\epsilon} \quad (23b)$$

$$\begin{aligned} & \sum_{v=1}^{U_2} f_v \log_2 (f_v + \gamma) + \sum_{v=1}^{U_1} (b_v - f_v) \log_2 (b_v - f_v + \gamma) - \\ & \left[ \sum_{v=1}^{U_2} (1 + \log_2 (f'_v)) - \sum_{v=1}^{U_1} (1 + \log_2 (b_v - f'_v)) \right] (f_v - f'_v) - \\ & \sum_{v=1}^{U_2} (1 + \log_2 (f'_v)) + \sum_{v=1}^{U_1} (1 + \log_2 (b_v - f'_v)) \leq C_{\text{Ter}} \end{aligned} \quad (23c)$$

$$\sum_{v=1}^{U_2} f_v \leq W_i^{\text{bs}}. \quad (23d)$$

Thanks to formulating the constraint (13c) in terms of a DC function, and approximating it using first-order Taylor series obtaining (23c), the reduced problem (23) becomes a convex problem. It is worth to emphasize that such an approximation does not violate the problem; however, the proposed bound reduces the feasibility space resulting in a sub-optimal solution. To solve the non-convex problem (13), we use the successive convex approximation (SCA) and solve the approximated convex problem (23) iteratively where  $f'_v$  is updates at each iteration using the previous solution of (23).

---

**Algorithm 1:** Resource Allocation in the Terrestrial Network

---

**Input:**  $[b_v]_{v \in \{1 \dots U_1\}}$ ,  $\epsilon$ ,  $C_{\text{Ter}}$ ,  $\gamma$   
**Initialize:**  $\epsilon_{\text{error}}$ ,  $f'_v$ ;  $\forall v \in \{1 \dots U_1\}$   
**while**  $(|f'_v - f_v^*|_{v \in \{1 \dots U_1\}} \geq \epsilon_{\text{error}})$  **do**  
    Solve problem (23) to get  $[f_v]_{v \in \{1 \dots U_1\}}^*$  ;  
     $f'_v \leftarrow f_v^*$ ;  $\forall v \in \{1 \dots U_1\}$ ;  
**end**

---

Algorithm 1 solves the resource allocation problem of the terrestrial network using the SCA approach. First, we start with an initial arbitrarily values for  $[f'_v]_{v \in \{1 \dots U_1\}}$  and stopping error,  $\epsilon_{\text{error}}$ , for the iterative algorithm. We choose to start with small values of  $f'_v$ , for instance  $f'_v = 0.1b_v$ , since our objective function aims to take as little as possible from the eMBB users resources  $b_v$ . In each iteration of Algorithm 1, we solve the convex optimization problem (23) using the Interior Point algorithm, where  $f'_v$  uses the solution of the previous iteration.

Once the problem (13) is solved, we obtain the optimal scheduling of eMBB users and URLLC users in the terrestrial network. We consider the optimized solution  $[f_v^*]_{v \in \{1 \dots U_1\}}$  in order to estimate the loads  $[L_i^e]^*_{i \in \{1, \dots, N\}}$  and  $[L_i^u]^*_{i \in \{1, \dots, N\}}$  of all small base stations. Using these parameters, we model the inter-working of the satellite with the terrestrial network in the next section.

## IV Resource Allocation in the Integrated Satellite Terrestrial Network

We use the satellite networks that present additional backhaul links for the terrestrial users to fulfill the QoS requirements of user slices (i.e., eMBB and URLLC). We aim to adjust the association between the satellite and  $N$  small base stations by determining  $(\alpha, \beta)$  where  $\alpha = [\alpha_1, \dots, \alpha_i, \dots, \alpha_N]^T$  denotes the bandwidth allocation vector and  $\beta = [\beta_1, \dots, \beta_i, \dots, \beta_N]^T$  denotes the offloading vector. Specifically,  $\alpha_{i \in \{1, \dots, N\}}$  is the bandwidth proportion allocated by the satellite to the small base station  $bs_{i \in \{1, \dots, N\}}$  and  $\beta_{i \in \{1, \dots, N\}}$  is the proportion of the offloaded eMBB traffic  $L_i^{e*}$  from the small base station  $bs_{i \in \{1, \dots, N\}}$  to the satellite. Our scheme takes into account the limitations on backhaul capacities on both terrestrial and satellite networks.

Let  $F_i = \beta_i L_i^{e*}$  denotes the offloaded traffic to the satellite by the small base station  $bs_i$ . Our problem can be posed as follows:

$$\max_{\alpha, \beta} [F_1, F_2, \dots, F_i, \dots, F_N] \quad \forall i \in \{1, \dots, N\} \quad (24a)$$

s.t.

$$F_i \leq \alpha_i W \log_2 \left( 1 + \frac{C}{\alpha_i W N_0} \right) \quad \forall i \in \{1, \dots, N\} \quad (24b)$$

$$\sum_{i=1}^N F_i \leq C_{\text{Sat}} \quad (24c)$$

$$\alpha_i - \alpha_{i'} \leq \frac{1}{\delta(L_i^u, L_{i'}^u)} - 1 \quad \forall i, i' \in \{1, \dots, N\} \quad (24d)$$

$$\beta_i - \beta_{i'} \leq \frac{1}{\delta(L_i^u, L_{i'}^u)} - 1 \quad \forall i, i' \in \{1, \dots, N\} \quad (24e)$$

$$0 \leq \alpha_i \leq 1 \quad \forall i \in \{1, \dots, N\} \quad (24f)$$

$$\sum_{i=1}^N \alpha_i = 1 \quad (24g)$$

$$0 \leq \beta_i \leq 1 \quad \forall i \in \{1, \dots, N\}, \quad (24h)$$

where (24) represent a maximization problem for multi-objective functions with several constraints (24b) – (24h). First, the constraint (24b) ensures that the eMBB traffic offloaded by the base station  $bs_i$  does not exceed the resources  $\alpha_i$  allocated by the satellite. The constraint (24c) reflects the limitation on the satellite backhaul capacity, where the sum of eMBB loads should be less than  $C_{\text{Sat}}$ . However, the constraints (24d) and (24e) are added to prioritize the small base station that has a more important URLLC traffic to offload its eMBB traffic to the satellite backhaul where

$$\delta(L_i^u, L_{i'}^u) = \begin{cases} 1 & \text{if } L_i^u \leq L_{i'}^u \\ 0, & \text{otherwise} \end{cases}. \quad (25)$$

To solve our multi-objective optimization, we use the weighted sum method, where each objective

function  $F_{i \in \{1, \dots, N\}}$  is assigned a weight  $\omega_{i \in \{1, \dots, N\}}$ . One way to fix the weights  $\omega_{i \in \{1, \dots, N\}}$  in our optimization problem (24) is based on the location of the small base station  $bs_{i \in \{1, \dots, N\}}$  within the coverage area of the satellite. The preponderant weights are allocated to the small base stations located at the edge of the coverage area to boost their offloading. These weights can be periodically changed between the  $N$  small base stations based on the visibility window of the satellite. Our problem (24) can be restated as follows

$$\begin{aligned} & \max_{\alpha, \beta} \quad \sum_{i=1}^N \omega_i \beta_i L_i^{e*} \\ & \text{s.t.} \\ & \beta_i L_i^{e*} \leq \alpha_i W_j \log_2 \left( 1 + \frac{P_j g_{ij}}{N_0 \alpha_i W_j} \right) \quad \forall i \in \{1, \dots, N\} \end{aligned} \quad (24b)$$

$$\sum_{i=1}^N \beta_i L_i^{e*} \leq C_{\text{Sat}} \quad (24c)$$

$$(24d) - (24h).$$

Our problem (24) is convex with a linear objective function, a convex constraint (24b) and linear constraints (24c)-(24h). Thus we can solve by any convex optimization tool obtaining the optimal proportion  $\beta_i \in \{1, \dots, N\}$  of the eMBB traffic that should be offloaded by each small base station to the satellite and the optimal proportion  $\alpha_i \in \{1, \dots, N\}$  of the satellite bandwidth that should be allocated to each small base station according to the offloaded traffic. By using the optimal solution  $(\alpha_i, \beta_i)_{i \in \{1, \dots, N\}}$ , the eMBB load  $L_i^{eT}$  that should be transmitted by the small base station  $bs_{i \in \{1, \dots, N\}}$  through the terrestrial backhaul can be computed in ISTN as

$$L_i^{eT} = (1 - \beta_i) L_i^{e*}. \quad (27)$$

This load,  $L_i^{eT}$ , is fundamental to estimate the URLLC delay and the dropped eMBB traffic experienced in the terrestrial backhaul. Accordingly, we can compare the performance of our scheme operating in ISTN to a benchmark scheme operating in the terrestrial network as follows in the next section.

## V Terrestrial Delay Analysis

In this section, we study the URLLC delay experienced in the terrestrial backhaul analytically. This delay depends not only on the URLLC traffic  $L_i^{u*}$  of the small base station  $bs_{i \in \{1, \dots, N\}}$ , but also on the eMBB traffic sent through this backhaul. This latter traffic is given by equation (27); where  $\beta_i$  is the optimal solution of our problem (24) in the case of ISTN. In the case of terrestrial network,  $\beta_i = 0$ .

In this analysis, the terrestrial link of the small base station  $bs_i$  is modeled as M/G/1 queue where :

- the arrival rate of the eMBB packets and URLLC packets  $\lambda_i$  to small base station  $bs_i$  follows a Poisson process given by  $\lambda_i = \lambda_i^e + \lambda_i^u$
- the service time distribution is general.

## V.A. Mean Waiting Time

We assume that the scheduling discipline is First Come First Serve (FCFS). The waiting time  $\mathcal{T}_k$  of packet  $k$  scheduled after  $K$  packets is given by:

$$\mathcal{T}_k = \mathcal{R}_k + \sum_{n=k-K}^{k-1} \mathcal{X}_n, \quad (28)$$

where  $\mathcal{X}_n$  is the service time of packet  $n$  that arrived before packet  $k$  and  $\mathcal{R}_k$  is the residual service time.

Using Little's formula, we obtain the average waiting time  $\overline{\mathcal{T}}_k$  as :

$$\overline{\mathcal{T}}_k = \frac{\overline{\mathcal{R}}_k}{1 - \rho_i}, \quad (29)$$

where  $\rho_i$  is the total load in the small cell  $i$

$$\rho_i = \rho_i^e + \rho_i^u = \frac{\lambda_i^e}{\mu_e} + \frac{\lambda_i^u}{\mu_u}. \quad (30)$$

$\rho_i$  is expressed in both networks as follows

$$\rho_i = \begin{cases} \frac{L_i^{e*} + L_i^{u*}}{C_{\text{Ter}}} & \text{in the terrestrial network} \\ \frac{(1-\beta_i)L_i^{e*} + L_i^{u*}}{C_{\text{Ter}}} & \text{in ISTN} \end{cases} \quad (31)$$

The first moment of the residual service time can be developed as :

$$\overline{\mathcal{R}}_k = \frac{\rho_i}{2} \frac{\overline{\mathcal{X}^2}}{\overline{\mathcal{X}}}. \quad (32)$$

In order to evaluate the mean service time  $\overline{\mathcal{X}}$ , we approximate our general distribution to an exponential distribution with a mean service rate  $\mu_e$ . Therefore, the first moment of the residual service time can be simplified, based on (32), as

$$\overline{\mathcal{R}}_k = \frac{\rho_i}{2} \frac{\frac{2}{\mu_e^2}}{\frac{1}{\mu_e}} = \frac{\rho_i}{\mu_e}. \quad (33)$$

## V.B. Distribution of the Waiting Time

To find the distributions of the time spent in the system and in the queue, we consider the Laplace transform approach. Let  $T$  be the total time spent in the system,  $T_q$  the total waiting time before the service begins and  $X$  the service time. This yields that  $T = Q + X$  such that  $Q$  and  $X$  are independent.

We denote with  $L_T$ ,  $L_Q$  and  $L_B$  the Laplace transforms of  $T$ ,  $T_q$  and  $X$  respectively.

$$L_T(s) = E[e^{-sT}] = E[e^{-s(Q+X)}] = E[e^{-sQ}] \cdot E[e^{-sX}] = L_Q(s) \cdot L_B(s).$$

Hence

$$L_Q(s) = \frac{L_T(s)}{L_B(s)} \quad (34)$$

Based on equation (5.100) in [20],

$$L_T(s) = L_B(s) \frac{s(1 - \rho_i)}{s - \lambda_i + \lambda_i L_B(s)} . \quad (35)$$

Substituting (35) in (34) yields

$$L_Q(s) = \frac{s(1 - \rho_i)}{s - \lambda_i + \lambda_i L_B(s)} . \quad (36)$$

In order to evaluate  $L_Q$ , we approximate our general distribution to an exponential distribution with a mean service rate  $\mu_e$ . Accordingly,

$$L_B(s) = \frac{\mu_e}{\mu_e + s} . \quad (37)$$

Therefore

$$L_Q(s) = (1 - \rho_i) \left( 1 + \frac{\lambda_i}{s + \mu_e - \lambda_i} \right) , \quad (38)$$

and

$$L_T(s) = \frac{\mu_e - \lambda_i}{s + \mu_e - \lambda_i} . \quad (39)$$

Let  $f_Q$  and  $f_T$  denote the Probability Distribution Functions (PDF) of  $Q$  and  $T$  respectively. To determine the waiting time distributions, we apply the inverse Laplace transform on equations (38) and (39). This leads to

$$f_Q(t) = (1 - \rho_i) \left( \delta(t) + e^{(\lambda_i - \mu_e)t} \right) , \quad (40)$$

$$\text{where } \delta(t) = \begin{cases} \infty, & \text{if } t = 0 , \\ 0, & \text{otherwise.} \end{cases}$$

and

$$f_T(t) = (\mu_e - \lambda_i) e^{-(\mu_e - \lambda_i)t} . \quad (41)$$

The CDF of the total waiting time  $F_T$  can finally be obtained as

$$F_T(x) = \int_{-\infty}^x f_T(t) dt = 1 - e^{-(\mu_e - \lambda_i)x}, \quad x \geq 0 . \quad (42)$$

## VI Results and Discussion

In this section, we evaluate our scheme performance in terms of latency and availability. In this regard, we assess the variation of the URLLC delay in the terrestrial backhaul and the dropped volume of eMBB traffic. We conducted our simulations based on the features of three LEO constellations. The simulations parameters are detailed in Table 1 for the terrestrial network and in Table 2 for the satellite network.

First, we prove the convergence of our proposed algorithm for resource allocation in the terrestrial network. Therefore, we plot the eMBB sum rate versus the number of iterations for different scenarios of terrestrial backhaul capacities as shown in Fig.2. We observe a gradual increase of the eMBB sum

Table 1: Terrestrial Network Parameters

Parameter	Numerical Value
Small cell radius (m)	500
Bandwidth (Mhz)	100
RB Bandwidth (Mhz)	0.18
$N_o$ (dbm/Hz)	-174
SBS transmission power (W)	20
URLLC packet size (B) [2]	30
eMBB packet size (B) [2]	100

Table 2: Satellite Features per Downlink Beam [15]

Parameter	Telsat	OneWeb	SpaceX
Bandwidth (Ghz)	0.25	0.25	0.25
Satellite Capacity (Mbps)	558.7	599.4	674.3
$Rx \frac{C}{N_0}$ (db)	9.6	10.5	12

rate with increasing the number of iterations. Then, the eMBB sum rate saturates at a maximum value within 70 iterations.

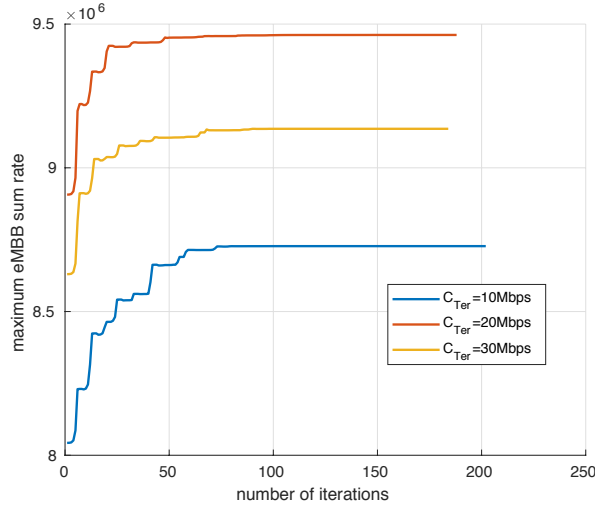


Figure 2: Sum rate vs number of iterations

Then, we evaluate the performance of our scheme according to the variation of the link capacity and of the link load respectively. We compare our results obtained in ISTN to a benchmark that consists of a network with a terrestrial backhaul only. To fairly compare both schemes, the considered benchmark has an equivalent terrestrial backhaul capacity  $C'_{\text{Ter}}$  such that

$$C'_{\text{Ter}} = C_{\text{Ter}} + \frac{C_{\text{Sat}}}{N}. \quad (43)$$

In the first simulation, we investigate the role of satellites integration in terrestrial networks to alleviate the capacity limitation in dense networks. We vary the link capacity  $C_{\text{Ter}}$  between 10 Mbps

and 100 Mbps for a fixed network load  $\rho = 0.8$  that characterizes highly loaded networks. We study first the variation of the URLLC delay experienced in the terrestrial backhaul in both networks.

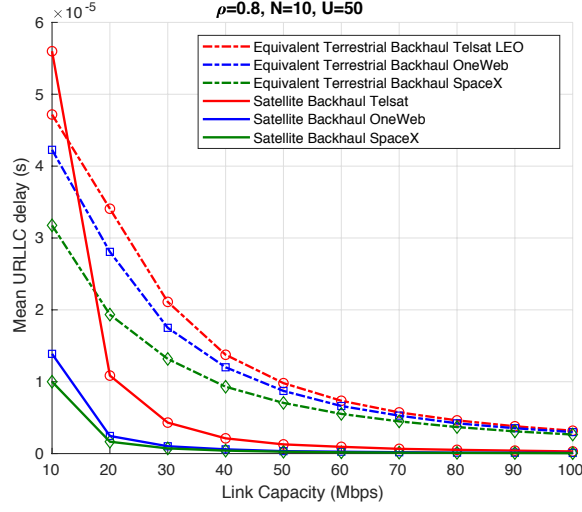


Figure 3: Mean URLLC Delay as function of Link Capacity.

Based on the results depicted in Fig.3, we observe that the delay decreases as expected with larger capacities. We notice also that our scheme, based on satellite and terrestrial network integration, achieves lower delays than the terrestrial benchmark. Interestingly, our scheme outperforms significantly the benchmark for the small link capacities (i.e. less than 30 Mbps). This finding highlights the key role that our scheme plays in dense networks where link capacities are considerably limited.

Then, we study the mean amount of the dropped eMBB data experienced in the terrestrial backhaul in both networks.

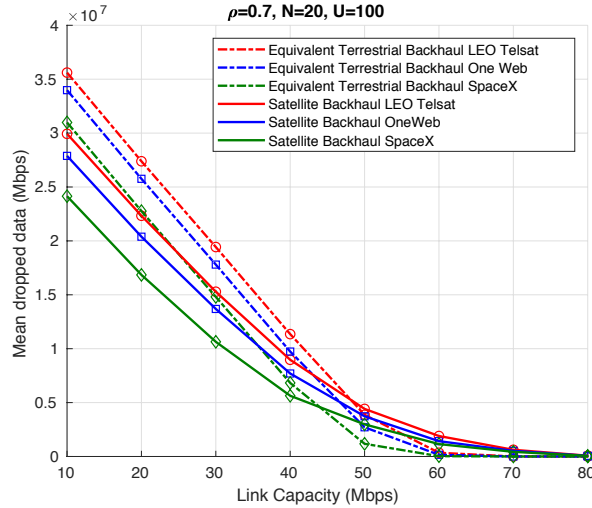


Figure 4: Mean Dropped eMBB Data as function of Link Capacity.

As illustrated in Fig.4, the data dropping significantly diminishes with larger link capacities in both cases. We observe also that our scheme has lower mean dropping compared to the benchmark. We note that our scheme performs better especially for the small link capacities (i.e. less than 30 Mbps).



Afterwards, we study the network availability by assessing the percentage of the successfully transmitted data to the backhaul of the totally sent data.

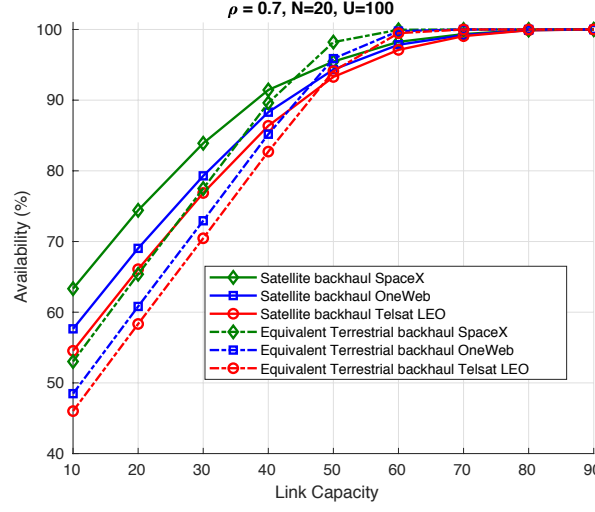


Figure 5: Mean Availability as function of Link Capacity.

The results presented in Fig.5 endorse the previous dropping findings. Precisely, the network availability improves with larger link capacities for both networks. We remark that our scheme enhances the network availability particularly for the small link capacities (i.e. less than 30 Mbps) compared to the terrestrial benchmark. We notice also that the network availability stabilizes at 100% for the large link capacities (i.e. more than 80 Mbps).

These simulation results prove that ISTN outperforms the terrestrial network in terms of delay and availability when the link capacity is increased. The performance improvement of ISTN is remarkable especially in dense networks (link capacities less than 30 Mbps). However, increasing the link capacity beyond a certain limit (e.g. 80 Mbps in our case) is not always efficient.

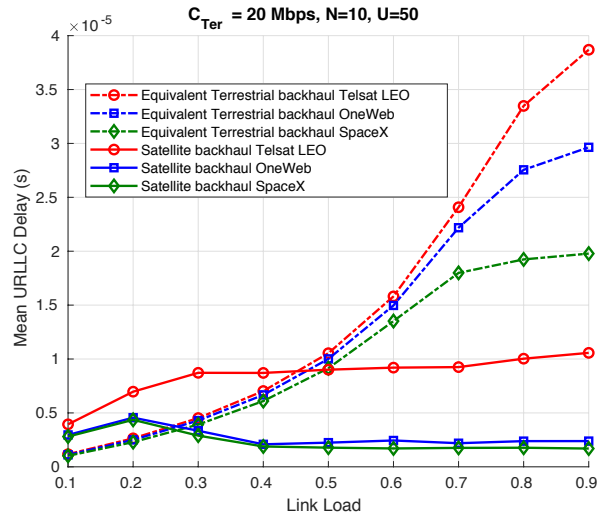


Figure 6: Mean URLLC Delay as function of Link Load.

In the second simulation, we inspect the influence of the network load on both networks performance.

We vary the link load  $\rho$  between 0.1 and 0.9 for a fixed network capacity  $C_{\text{Ter}} = 20$  Mbps that characterizes dense networks. We study first the variation of the URLLC delay experienced in the terrestrial backhaul in both networks.

The results depicted in Fig. 6 show that our scheme surpasses the benchmark starting from a link load equal to 0.5 for Telsat LEO system and 0.3 for OneWeb and SpaceX respectively. The improvement is especially important for  $\rho = 0.9$ . This finding underlines that our scheme helps to reduce dramatically the delays in dense networks.

Then, we study the mean amount of the dropped eMBB data experienced in the terrestrial backhaul in both networks as depicted in Fig. 7 .

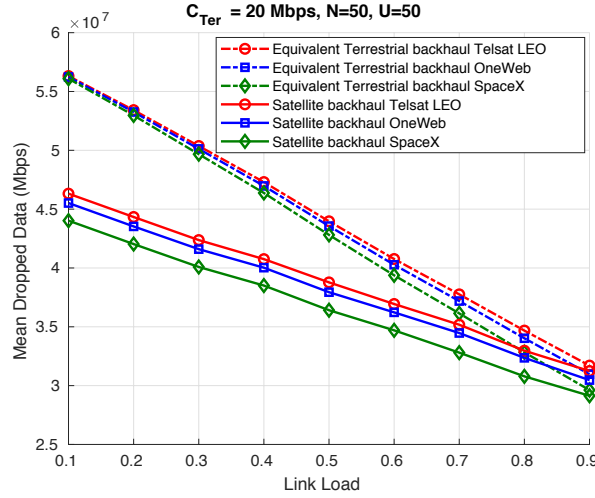


Figure 7: Mean Dropped eMBB Data as function of Link Load.

We notice that the mean dropped eMBB traffic decreases when the network load increases. This observation is due to the fact that a more important traffic volume is accepted in the network when the load is higher. We note also that the mean dropped eMBB traffic is higher for the terrestrial benchmark compared to our scheme mainly for low network loads (i.e. less than 0.3). This observation is due to the fact that more eMBB packets are rejected to maintain a low load in the network.

Afterwards, we study the network availability. The results presented in Fig.8 show that the network availability improves with larger network loads in both cases. We remark that our scheme enhances the network availability particularly for the low loads (i.e. less 0.3) compared to the terrestrial benchmark. This observation is due to the fact that more packets are dropped when the loads are lower.

Both simulation results indicate that a better performance is reached when the constellation features are more robust. Indeed, the best results are obtained when our scheme uses SpaceX parameters, which include a higher capacity and a stronger carrier to noise density (c.f. Table II).

In the third simulation, we evaluate experimentally the CDF of the waiting time to endorse the previous results obtained through the mean values. We compare this experimental CDF to the analytical function derived in equation (42) for both networks by using the features of the constellation Telsat.

As depicted in Fig. 9, we note first the strong agreements between the analytical and the experimental distributions for ISTN and for the terrestrial network. We note also that the CDF curve of ISTN is above

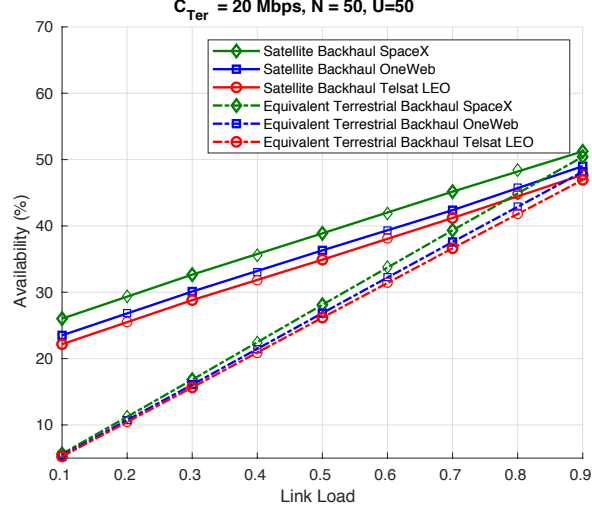


Figure 8: Mean Availability as function of Link Load.

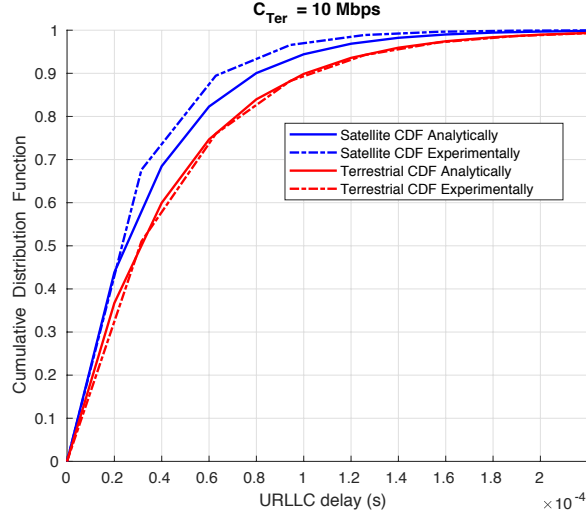


Figure 9: URLLC Delay Cumulative Distribution Function

the CDF curve of the terrestrial network. This observation can be interpreted mathematically as

$$\Pr(D_{\text{ISTN}} < D) \geq \Pr(D_{\text{Ter}} < D) \quad (44)$$

where  $D$ ,  $D_{\text{ISTN}}$  and  $D_{\text{Ter}}$  are respectively a fixed delay value, the delay experienced in ISTN and the delay experienced in the terrestrial network. This finding emphasizes that the integration of the satellite network improves the delay experienced in the terrestrial bakchaul.

## VII Conclusion

In this paper, we proposed an offloading scheme in ISTN to steer the different types of traffics to the transport network that meets their transmission requirements. In particular, we dealt with URLLC and eMBB traffics and offload the former to the terrestrial backhaul, while the latter is offloaded to the satel-

lite network. The proposed scheme used the satellite link efficiently to accommodate high eMBB traffic while prioritizing the URLLC traffic. Our findings indicated that the integration of satellite networks with terrestrial networks is useful to relieve the capacity limitations in terrestrial backhaul. Precisely, our scheme succeeded in decreasing the dropped amount of eMBB traffic significantly in the terrestrial backhaul. Thus, the eMBB applications became able to transmit their large payload successfully to the core network. Moreover, our scheme reduced the transmission delay experienced by the URLLC packets in the terrestrial backhaul to meet the applications' low latency requirements. Hence, it also improves the network availability for broader access to new users while satisfying their different QoS requirements. Our future work will be orientated towards machine learning in ISTN to predict the appropriate offloading times to anticipate packets dropping and enhance network availability.

## References

- [1] 3rd Generation Partnership Project (3GPP). Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT). Technical Report (TR) 45.820, (3GPP), 03 2015. Version 0.3.0.
- [2] 3rd Generation Partnership Project (3GPP). 5G : Study on Scenarios and Requirements for Next Generation Access Technologies. Technical Report (TR) 38.913, (3GPP), 05 2017. Version 14.2.0.
- [3] M. K. Abdel-Aziz, C.-F. Liu, S. Samarakoon, M. Bennis, and W. Saad. Ultra-reliable low-latency vehicular networks: Taming the age of information tail. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2018.
- [4] M. Abo-Zeed, J. B. Din, I. Shayea, and M. Ergen. Survey on land mobile satellite system: Challenges and future research trends. *IEEE Access*, 7:137291–137304, 2019.
- [5] M. Alsenwi and C. S. Hong. Resource scheduling of URLLC/eMBB traffics in 5G new radio: A punctured scheduling approach. *Korean Computer Congress (KCC), South Korea*, pages 1271–1273, 2018.
- [6] M. Alsenwi, S. R. Pandey, Y. K. Tun, K. T. Kim, and C. S. Hong. A chance constrained based formulation for dynamic multiplexing of embb-urllc traffics in 5g new radio. In *2019 International Conference on Information Networking (ICOIN)*, pages 108–113. IEEE, 2019.
- [7] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong. eMBB-URLLC resource slicing: A risk-sensitive approach. *IEEE Communications Letters*, 23(4):740–743, 2019.
- [8] A. Anand, G. De Veciana, and S. Shakkottai. Joint scheduling of URLLC and eMBB traffic in 5G wireless networks. In *IEEE Conference on Computer Communications INFOCOM*, pages 1970–1978, 2018.
- [9] M. Bennis, M. Debbah, and H. V. Poor. Ultra-reliable and low-latency wireless communication: Tail, risk and scale. *arXiv preprint arXiv:1801.01270*, 2018.

- [10] S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [11] N. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen. Space/aerial-assisted computing offloading for IoT applications: A learning-based approach. *IEEE Journal on Selected Areas in Communications*, 37(5):1117–1129, 2019.
- [12] M. E. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on networking*, 5(6):835–846, 1997.
- [13] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini. What should 6g be? *Nature Electron.*, 3(4), Jan. 2020.
- [14] T. De Cola, A. Ginesi, G. Giambene, G. C. Polyzos, V. A. Siris, N. Fotiou, Y. Thomas, et al. Network and protocol architectures for future satellite systems. *Foundations and Trends® in Networking*, 12(1-2):1–161, 2017.
- [15] I. del Portillo, B. G. Cameron, and E. F. Crawley. A technical comparison of three low earth orbit satellite constellation systems to provide global broadband. *Acta Astronautica*, 159:123–135, 2019.
- [16] B. Di, H. Zhang, L. Song, Y. Li, and G. Y. Li. Ultra-dense LEO : Integrating terrestrial-satellite networks into 5G and beyond for data offloading. *IEEE Transactions on Wireless Communications*, 18(1):47–62, 2018.
- [17] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M.-S. Alouini. Terahertz band: The last piece of rf spectrum puzzle for communication systems. *IEEE Open Journal of the Communications Society*, 1:1–32, Jan. 2020.
- [18] M. Harris. Amazon reports collision risk for mega-constellation of kuiper internet satellites, IEEE Spectrum 16-October-2019.
- [19] R. Kassab, O. Simeone, P. Popovski, and T. Islam. Non-orthogonal multiplexing of ultra-reliable and broadband services in fog-radio architectures. *Ieee Access*, 7:13035–13049, 2019.
- [20] L. Kleinrock. *Queueing Systems, Volume 2: Computer Applications*, volume 66. Wiley New York, 1976.
- [21] J. Li, K. Xue, D. S. Wei, J. Liu, and Y. Zhang. Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks. *IEEE Transactions on Wireless Communications*, 2020.
- [22] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato. Space-air-ground integrated network: A survey. *IEEE Communications Surveys & Tutorials*, 20(4):2714–2741, 2018.
- [23] H. Malik, M. M. Alam, Y. Le Moullec, and Q. Ni. Interference-aware radio resource allocation for 5g ultra-reliable low-latency communication. In *2018 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2018.

- [24] N. Sapountzis, T. Spyropoulos, N. Nikaen, and U. Salim. Optimal downlink and uplink user association in backhaul-limited hetnets. In *The 35th Annual IEEE International Conference on Computer Communications INFOCOM*, pages 1–9, 2016.
- [25] J. Tang, B. Shim, and T. Q. Quek. Service multiplexing and revenue maximization in sliced c-ran incorporated with urllc and multicast embb. *IEEE Journal on Selected Areas in Communications*, 37(4):881–895, 2019.
- [26] Y. Wang, J. Zhang, X. Zhang, P. Wang, and L. Liu. A computation offloading strategy in satellite terrestrial networks with double edge computing. In *2018 IEEE International Conference on Communication Systems (ICCS)*, pages 450–455. IEEE, 2018.
- [27] Z. Zhang, W. Zhang, and F.-H. Tseng. Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques. *IEEE Network*, 33(1):70–76, 2019.



**Wiem Abderrahim** (S'14 - M'18) accomplished her undergraduate studies in electrical engineering at the Higher School of Communications of Tunis, Carthage University, Tunisia in 2013. She received her Doctoral Degree in Information and Communication Technologies from the same university in 2017. She worked as a lecturer and then as an adjunct professor at the Higher School of Communications of Tunis between 2014 and 2018. Currently, she is a postdoctoral fellow within King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. Her research interests include cloud computing, network virtualization and recently satellite communications and machine learning.



**Osama Amin** (S'07, M'11, SM'15) received his B.Sc. degree in electrical and electronic engineering from Aswan University, Egypt, in 2000, his M.Sc. degree in electrical and electronic engineering from Assiut University, Egypt, in 2004, and his Ph.D. degree in electrical and computer engineering, University of Waterloo, Canada, in 2010. In June 2012, he joined Assiut University as an assistant professor in the Electrical and Electronics Engineering Department. Currently, he is a research scientist in the CEMSE Division at KAUST, Thuwal, Makkah Province, Saudi Arabia. His general research interests lie in communication systems and signal processing for communications with special emphasis on wireless applications.



**Mohamed Slim Alouini** (F'09) was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas A&M University at Qatar, Education City, Doha, Qatar before joining King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.



**Basem Shihada** (SM'12) is an associate and founding professor in the Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division at King Abdullah University of Science and Technology (KAUST). He obtained his PhD in Computer Science from University of Waterloo. In 2009, he was appointed as visiting faculty in the Department of Computer Science, Stanford University. In 2012, he was elevated to the rank of Senior Member of IEEE. His current research covers a range of topics in energy and resource allocation in wired and wireless networks, software defined networking, internet of things, data networks, smart systems, network security, and cloud/fog computing.