

# Heterogeneous Traffic Offloading in Space-Air-Ground Integrated Networks

LIANG ZHANG (Student Member, IEEE), WIEM ABDERRAHIM (Member, IEEE), and BASEM SHIHADA (Senior Member, IEEE)

Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia (e-mail: {liang.zhang, wiem.abderrahim, basem.shihada}@kaust.edu.sa).

Corresponding author: Basem Shihada (e-mail: basem.shihada@kaust.edu.sa).

**ABSTRACT** While the 5-th generation of communication networks (5G) is taking its first deployment steps, high doubts still concern the fulfilment of the stringent requirements of its slices. To meet these complex requirements, robust access networks should support the current 5G air interface. In this regard, space and air networks, namely satellites and unmanned aerial vehicles (UAVs) are expected to play a key role due to their wide coverage and scalable deployment respectively. Currently, the integration of these platforms in the terrestrial networks is weak though. Therefore, we suggest bridging this gap by designing a heterogeneous traffic offloading approach in the space-air-ground integrated network (SAGIN). Our innovative offloading approach covers the co-existing requirements of two heterogeneous slices of 5G by offloading smartly the traffic to the appropriate segment of SAGIN. Specifically, the ultra-reliable low-latency communications (URLLC) traffic is offloaded to the UAV link and to the terrestrial link to satisfy its stringent requirements in terms of latency. However, the enhanced mobile broadband (eMBB) traffic is offloaded to the UAV link, to the terrestrial link and to the satellite link because it is less sensitive to delay but needs high data rates. Our offloading approach boosts the network's availability and reduces the latency experienced in SAGIN through an efficient resource allocation and an optimized design of the UAVs trajectory. Our findings highlight the key role that the concrete integration between SAGIN segments plays to achieve a better quality of service (QoS) for different slices with heterogeneous requirements.

**INDEX TERMS** Enhanced mobile broadband (eMBB), space-air-ground integrated network (SAGIN), traffic offloading, ultra-reliable low latency communication (URLLC), unmanned aerial vehicle (UAV).

## I. INTRODUCTION

THE fifth generation of communication networks (5G) and beyond is envisioned to address various slices with heterogeneous requirements [1]. Among these slices, ultra-reliable low latency communication (URLLC) is receiving tremendous attention in academia and industry because it enables new applications and emerging services such as autonomous factory, inter-vehicular communications and e-health. URLLC addresses the intermittent transmissions that impose stringent requirements in terms of reliability (the acceptable dropping rate is less than  $10^{-5}$ ) and low latency (1ms) [2], [3]. These requirements make URLLC a challenging implementation scenario though, especially since 5G should simultaneously answer different needs imposed by the other slices. For instance, the enhanced mobile broadband (eMBB) slice of 5G, which addresses the stable connections with large payloads over extended time intervals, demands

high data rates but is less sensitive to reliability [1].

To face these challenges, various technologies such as Terahertz communications, optical wireless communications, free-space optical communication, and dynamic network slicing should support the 5G networks [4]. Among these technologies, space-air-ground integrated network (SAGIN) is expected to stand out as a key player [4], [5]. SAGIN consists of three network segments: the space segment which includes the satellites, the air segment which includes the unmanned aerial vehicle (UAV) and the ground segment which includes the terrestrial communication networks [5]–[7]. These segments can cooperate and complement each other through their own privileges [6], [7]. For instance, satellites have the privileges of wide coverage and wide bandwidth. However, UAVs have the privileges of scalable deployment, mobility, low latency and reliability [8], [9]. Hence, space and air segments can support the terrestrial

network to increase the users' connectivity and hence improve the service continuity, especially in under-served and rural areas. Moreover, space and air segments can support the terrestrial network to increase the backhaul capacity and hence improve the network's availability, especially in dense areas where a large volume of data is exchanged constantly.

Therefore, we observe that a concrete and self-adaptive integration of space and air segments in the terrestrial network could help fulfill the complex and heterogeneous slices of 5G while reducing considerably the service costs [5], [7]. The main integration approach in the literature is traffic offloading in SAGIN [9]–[14]. For instance, in [9], a sensor offloading framework was developed to design the UAVs trajectories while maximizing the collected data from different sensors and considering the limited energy on-board of UAVs. In [10], an offloading scheme, which schedules the offloaded tasks of the edge users between the ground base stations and the UAVs, was presented. The authors investigated the role of UAVs to improve the average throughput and the spectral efficiency during tasks offloading. In [11], the authors optimized the ground base stations and UAVs locations in MEC. An online offloading algorithm was developed to minimize the energy consumed by the user equipment (UE) through appropriate user association and resource allocation. In [12], service provisioning to the Internet of things (IoT) devices was facilitated by using a collaborative offloading scheme between UAVs and edge servers in MEC. The offloading scheme reduces the service delay experienced by IoT devices and the energy consumed by UAVs. In [13], a traffic offloading scheme in integrated satellite-terrestrial networks (ISTN) was proposed to maximize the number of 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 [14], a SAGIN architecture was studied where the air network nodes served as flying edge servers in the air segment. However, the low earth orbit (LEO) satellites in the space segment connected the IoT devices with the cloud servers. The proposed architecture helps the remote IoT applications to decide on the typical offloading location, namely locally, at the air segment, or at the space segment. Although huge efforts were invested in the topic of traffic offloading in SAGIN, few shortcomings can be spotlighted. Indeed, most of the current research studies separately air platforms [9]–[12], space platforms [13] and terrestrial platforms [15]. Moreover, energy efficiency is the main focus point in the literature [11], [12], [14], [16] and the UAV trajectory is usually predefined in SAGIN [14]. However, the high traffic assignment and the dynamic network condition and uncertainties in 5G were widely neglected. Besides, 5G slices, namely eMBB and URLLC, and their specific and heterogeneous requirements in terms of delay and reliability were largely overlooked during offloading. Indeed, offloading the optimal amount of each traffic type while meeting its needs is challenging. Specifically, scanty offloading results in high dropping ratio

for eMBB, whereas excessive offloading increases the load of SAGIN segments and leads to supplementary delays for URLLC. Hence, it is crucial to design wisely traffic offloading in SAGIN and to assign adaptively and efficiently the necessary communication resources.

To overcome the previously discussed limitations, we propose in this paper a traffic offloading approach in SAGIN, which hinges on concrete cooperation between satellites, UAVs, and terrestrial networks. Our offloading approach steers smartly the traffic towards the suitable network segment through an efficient resource allocation and an optimized design of the UAVs' trajectories. Specifically, the URLLC traffic is offloaded to the UAV segment and to the terrestrial segment to satisfy its stringent requirements in terms of latency. However, the eMBB traffic is offloaded to the UAV segment, to the terrestrial segment and to the satellite segment because it is less sensitive to delay but needs rather high data rates. The amount of the traffic offloaded to the different segments of SAGIN is intelligently programmed. Additionally, the trajectory of the UAVs and the resource allocation are wisely designed to improve the availability and reduce the latency in SAGIN. To the best of our knowledge, our paper is the first to consider offloading different traffic types and to answer their respective requirements in SAGIN. We propose an action-refined deep reinforcement learning (DRL) approach to adjust the offloading strategy by dynamically observing the information of the system environment. Specifically, we apply the deep deterministic policy gradient (DDPG) algorithm, which embeds the actor-critic method [17]. The main contributions of this paper can be summarized as follows:

- We propose an innovative offloading approach in SAGIN that covers various aspects of the quality of service (QoS) required by heterogeneous traffic types and imposed by the use cases of 5G. Specifically, our offloading approach answers the challenging and co-existing requirements imposed by eMBB slice and URLLC slice respectively in terms of high data rates, high reliability and low latency.
- We take advantage of the concrete cooperation between SAGIN segments by smartly offloading the traffic to the appropriate segments of the SAGIN system while optimizing the resource allocation and efficiently designing the trajectory of the UAVs. We formulate our sequential offloading approach as a multi-objective optimization problem and we solve this complex and dynamic problem by applying the action-refined DRL approach.
- Our findings highlight that the proposed offloading approach significantly enhances the network's availability and the experienced delay and underline the pivotal role that space and air segments play to boost the QoS in terrestrial networks.

The remainder of this paper is organized as follows: In section II, we describe our system model. In Section II, the problem of the UAVs trajectory design and traffic offloading

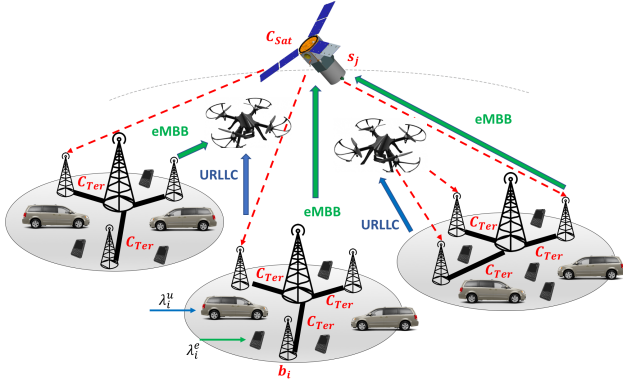


Fig. 1: Heterogeneous traffic offloading in SAGIN

is formulated. In section IV, we present our DRL algorithm to solve this problem. In section V, we evaluate our proposed approach and we analyse the obtained results. Section VI concludes the paper.

**Notations:** Lower case boldface letters denote vectors while upper case boldface letters denote matrices. The transpose of matrix  $\mathbf{A}$  is denoted by  $\mathbf{A}^T$ .  $\|\cdot\|$  denotes the Euclidean norm.  $\mathbb{E}(\cdot)$  denotes the expectation over the time sequence.

## II. SYSTEM MODEL

We consider a space-air-ground integrated network (SAGIN) architecture that consists of a satellite,  $N$  UAV,  $W$  macro base stations, and  $V$  micro base stations. The network architecture is depicted in Fig.1. The satellite is characterized by a link capacity per beam  $C_{sat}$  and a visibility period  $\Delta t$ . During  $\Delta t$ , the satellite is fixed. Let us decompose  $\Delta t$  into  $T$  time slots of duration  $\delta t$  such that  $\Delta t = T \times \delta t$ . The UAVs are in motion during  $\Delta t$ . The location set of the UAVs is denoted as  $\mathbf{L}[t] = \{\ell_1[t], \dots, \ell_n[t], \dots, \ell_N[t]\}$ , where  $\ell_n[t] = (x_n[t], y_n[t], z_n[t])^T$  represents the coordinates of the  $n$ -th UAV. The UAV's altitude  $z_n$  is fixed. Each UAV  $n$  is characterized by the data rate  $R_{v,n}^{UAV}$  of its respective link with the micro base station  $v$ . The macro base station  $w \in \{1, \dots, W\}$  serves  $v_w$  micro base stations and each micro base station  $v \in \{1, \dots, V_w\}$  serves  $I_v$  users. The terrestrial backhaul between the micro base station and the macro base station is characterized with a link capacity  $C_v^{ter}$ . The serving region of each micro base station is confined as a small cell. The users  $I_v$  can generate two types of traffic; namely eMBB and URLLC; which are sent to the micro base stations  $v \in \{1, \dots, V_w\}$ . The generated traffic can be offloaded by the micro base stations  $v \in \{1, \dots, V_w\}$  to the macro base station  $w \in \{1, \dots, W\}$ , to UAV  $n \in \{1, \dots, N\}$  and to the satellite based on the allocated link capacities in the different segments of SAGIN. Therefore, we study the traffic model and the allocated data-rates in the following sub-sections:

### A. TRAFFIC MODEL

In each small cell, a user  $U_i$  sends either eMBB traffic or URLLC traffic to its associated micro base station. We define  $U_e[t] = \{U_1^e, \dots, U_i^e, \dots, U_{N_e[t]}^e\}$  as the set of  $N_e[t]$  eMBB users and  $U_u[t] = \{U_1^u, \dots, U_i^u, \dots, U_{N_u[t]}^u\}$  as the set of  $N_u[t]$  URLLC users during time slot  $t$ . The users are randomly positioned and their traffic is independent in each time slot  $t$ . The total users set is defined as  $U[t] = U_u[t] \cup U_e[t]$ , such that  $U[t] = \{U_1, \dots, U_i, \dots, U_{I_v[t]}\}$  and  $I_v[t]$  is the total number of the users served by micro base station  $v$ . In each time slot  $t \in \{1, \dots, T\}$ , the eMBB flows outgoing from all eMBB users to micro base station  $v$  follow a Poisson process with an arrival rate  $\lambda_v^e[t]$  because eMBB traffic is generated by applications that exchange large payloads over an extended time interval [18]. Therefore, the inter-arrival time can be modeled with an exponential distribution whose events occur continuously and independently. However, the URLLC flows outgoing from all URLLC users to micro base station  $v$  follow a Pareto distribution with arrival rate  $\lambda_v^u[t]$  because URLLC traffic is generated by applications with intermittent transmissions that exchange more or less important payload during short periods of time [18]. Therefore, it can be modeled with a power-law distribution namely a Pareto distribution [18]–[22]. Flow sizes are independently and identically distributed with mean  $\frac{1}{\mu_e}$  for eMBB slice and  $\frac{1}{\mu_u}$  for URLLC slice, such that  $\frac{1}{\mu_e}$  and  $\frac{1}{\mu_u}$  represent the flow size for both traffic types, respectively.

The total load of URLLC traffic and eMBB traffic in the  $v$ -th micro base station are denoted by  $L_v^e[t]$  and  $L_v^u[t]$  respectively and are assessed as [19], [20], [23]:

$$L_v^u[t] = \sum_{i=1}^{N_u[t]} f_i^v[t] \log_2 \left( 1 + \frac{P_i^v[t] g_i^v[t]}{N_0 f_i^v[t] + \sum_{v' \neq v} P_i^{v'}[t] g_i^{v'}[t]} \right), \quad (1)$$

and

$$L_v^e[t] = \sum_{i=1}^{N_e[t]} b_i^v[t] \log_2 \left( 1 + \frac{P_i^v[t] g_i^v[t]}{N_0 b_i^v[t] + \sum_{v' \neq v} P_i^{v'}[t] g_i^{v'}[t]} \right), \quad (2)$$

where  $b_i^v[t]$  denotes the frequency resources allocated to eMBB user  $i$  by micro base station  $v$  in time slot  $t$ ,  $f_i^v[t]$  denotes the frequency resource allocated by micro base station  $v$  to URLLC user  $i$  in time slot  $t$ ,  $N_0$  is the noise power density,  $P_i^v[t]$  is the downlink transmit power of micro base station  $v$ ,  $g_i^v[t]$  is the channel gain between micro base station  $v$  and user  $i$  in time slot  $t$  and is determined based on the Rayleigh channel model and  $\sum_{v' \neq v} P_i^{v'}[t] g_i^{v'}[t]$  is the interference caused by the other micro base stations. In our system, we assume that the macro base stations allocate orthogonal channels to the micro base stations. Therefore, the interference between the micro base stations is negligible ( $\sum_{v' \neq v} P_i^{v'}[t] g_i^{v'}[t] \approx 0$ ). Therefore, the total load of URLLC traffic and eMBB traffic in the  $v$ th micro base station

$L_v^e[t]$  and  $L_v^u[t]$  can be simplified as follows:

$$L_v^u[t] = \sum_{i=1}^{N_u[t]} f_i^v[t] \log_2 \left( 1 + \frac{P_i^v[t] g_i^v[t]}{N_0 f_i^v[t]} \right), \quad (3)$$

and

$$L_v^e[t] = \sum_{i=1}^{N_e[t]} b_i^v[t] \log_2 \left( 1 + \frac{P_i^v[t] g_i^v[t]}{N_0 b_i^v[t]} \right). \quad (4)$$

### B. UAV DOWN-LINK DATA RATE

We consider the wireless communication between the hovering UAV  $n$  and micro base station  $v$ . Both line-of-sight (LoS) propagation and non-line-of-sight (N-LoS) propagation are considered in the module of average path-loss. The LoS probability is formulated as [24]:

$$p_{v,n}^{\text{LoS}}[t] = \frac{1}{1 + b_1 e^{-b_2 \left( \arctan \left( \frac{h_n(\ell_n[t])}{d_{v,n}(\ell_n[t])} \right) - b_1 \right)}}, \quad (5)$$

where  $h_n$  and  $d_{v,n}$  denote the height of the  $n$ -th UAV and the horizontal distance between the  $n$ -th UAV and the  $v$ -th macro base station respectively;  $b_1$  and  $b_2$  are the S-curve parameters determined by the chosen environment, e.g. urban, sub-urban or dense urban. The signal propagated from the UAV first goes through the free space and then through the urban environment. Therefore, the signal path loss mainly consists of two parts: the free space path-loss (FSPL)  $PL_{\text{FSPL}}$  and the excessive path-loss  $PL_{\text{ur}}$ . Based on the models proposed in [24] and [25], the expression of the total path-loss is given by:

$$\begin{aligned} PL_{v,n}[t] &= PL_{\text{FSPL}}[t] + PL_{\text{ur}}[t] \\ &= 20 \log \left( \frac{4\pi f_c d_{v,n}(\ell_n[t])}{c} \right) + p_{v,n}^{\text{LoS}}[t] \xi_{\text{LoS}} + \\ &\quad (1 - p_{v,n}^{\text{LoS}}[t]) \xi_{\text{NLoS}}, \end{aligned} \quad (6)$$

where  $\xi_{\text{LoS}}$  and  $\xi_{\text{NLoS}}$  represent respectively the additional path-loss corresponding to the LoS and N-LoS transmission,  $c$  is the speed of the light and  $f_c$  is the carrier frequency.

The channel gain of the links between the  $n$ -th UAV and the  $v$ -th micro base station is given by [24]:

$$g_{v,n}^{\text{UAV}}[t] = 10^{-PL_{v,n}[t]/10}. \quad (7)$$

Accordingly, the signal-to-interference-plus-noise ratio (SINR) of the link between the  $n$ -th UAV and the  $v$ -th micro base station is given by:

$$\Gamma_{v,n}[t] = \frac{P_{v,n}[t] g_{v,n}^{\text{UAV}}[t]}{N_0 b_{v,n}^{\text{UAV}}[t] + I_{v,n}[t]}, \quad (8)$$

where  $P_{v,n}[t]$  is the power consumed due to the signal transmission between the UAV  $n$  and the micro base station  $v$ ;  $b_{v,n}^{\text{UAV}}[t]$  represents the bandwidth allocated by UAV  $n$  for the  $v$ th micro base station at time slot  $t$  and  $I_{v,n}$  is the total interference experienced by micro base station  $v$  given by:

$$I_{v,n}[t] = \sum_{n' \neq n} P_{v,n'}[t] g_{v,n'}^{\text{UAV}}[t] + \sum_{v' \neq v} P_{v',n}[t] g_{v',n}^{\text{UAV}}[t]. \quad (9)$$

To provide efficient services and avoid overlap, the UAVs are usually placed far enough from each other such that their mutual interference can be overlooked (i.e.  $\sum_{n' \neq n} P_{v,n'}[t] g_{v,n'}^{\text{UAV}}[t] \approx 0$ ). Furthermore, we assume that the UAVs assign orthogonal channels to the micro base stations and the co-channel interference becomes negligible (i.e.  $\sum_{v' \neq v} P_{v',n}[t] g_{v',n}^{\text{UAV}}[t] \approx 0$ ) since the existing techniques such as cell planning, frequency reuse, and beam-forming are capable of significantly mitigating the interference [26]. We note also that the macro base station, the UAVs and the satellite operate in different frequency bands. Therefore, their mutual interference can be neglected, and we consider the signal-to-noise ratio (SNR) in SAGIN.

To satisfy the dynamic QoS requirements and to achieve an energy-efficient communication, the serving process by the UAV should be well designed. Specifically, the communication with the micro base station is set up only if the SNR of the link between the  $n$ -th UAV and the  $v$ -th micro base station is higher than the predefined SNR threshold  $\Gamma_{\text{th}}$  for a given QoS, such that  $\Gamma_{v,n} \geq \Gamma_{\text{th}}$ .

According to Shannon capacity bound and after neglecting the interference effects, the instant data rate served by the  $n$ -th UAV is given by:

$$\begin{aligned} R_{v,n}^{\text{UAV}}[t] &= b_{v,n}^{\text{UAV}}[t] \log_2 (1 + \Gamma_{v,n}[t]) \\ &= b_{v,n}^{\text{UAV}}[t] \log_2 \left( 1 + \frac{P_{v,n}[t] g_{v,n}^{\text{UAV}}[t]}{N_0 b_{v,n}^{\text{UAV}}[t]} \right). \end{aligned} \quad (10)$$

### C. SATELLITE DOWN-LINK DATA RATE

The achievable data rate  $R_v^{\text{sat}}[t]$  between the  $v$ th micro base station and the satellite at time slot  $t$  is expressed as [13], [23]:

$$R_v^{\text{sat}}[t] = b_v^{\text{sat}}[t] \log_2 \left( 1 + \frac{P_v^{\text{sat}}[t] g_v^{\text{sat}}[t]}{N_0^{\text{sat}} b_v^{\text{sat}}[t]} \right), \quad (11)$$

where  $b_v^{\text{sat}}[t]$  and  $P_v^{\text{sat}}[t]$  are the bandwidth and power allocated by the satellite to the  $v$ -th micro base station at time slot  $t$ ;  $g_v^{\text{sat}}[t]$  is the channel gain between the satellite and the  $v$ -th micro base station at time slot  $t$  that mainly depends on the path loss due to the transmission distance in satellite links such that  $g_v^{\text{sat}}[t] \propto \frac{1}{d_{v,s}^\eta[t]}$  where  $d_{v,s}[t]$  is the distance between the satellite and the  $v$ -th micro base station and  $\eta$  is the path-loss exponent [27]. We should note that  $R_v^{\text{sat}} \leq C^{\text{sat}}$ , where  $C^{\text{sat}}$  is a metric fixed by the satellite manufacturer.

### III. OFFLOADING STRATEGY IN SAGIN

Our goal is to design an offloading strategy in SAGIN that jointly schedules the UAVs trajectory and allocates efficiently the resources in order to improve the network availability and to minimize the latency experienced by the URLLC traffic. According to our offloading strategy, URLLC traffic is offloaded to the UAV link and to the terrestrial link to satisfy its stringent requirements in terms of latency. However, eMBB traffic is offloaded to the UAV link, to the terrestrial link and to the satellite link because it is less sensitive to delay but

needs high data rates. We note that the traffic generated by the users and sent to the micro base stations  $v \in \{1, \dots, V_w\}$  as expressed in (3) and (4) is offloaded to the different segments of SAGIN based on the terrestrial link capacity  $C_v^{\text{ter}}$  and the available data rates developed in (10) and (11). In this section, we define and formulate our offloading problem to meet the aforementioned requirements and constraints.

### A. OFFLOADED TRAFFIC IN SAGIN

We define  $\alpha_v[t] = (\alpha_v^{\text{ter}}[t], (\alpha_v^{\text{UAV}}[t])^T, \alpha_v^{\text{sat}}[t])^T$  as the offloading vector of eMBB traffic from the  $v$ -th micro base station at time slot  $t$ , where  $\alpha_v^{\text{ter}}[t]$ ,  $\alpha_v^{\text{UAV}}[t]$ , and  $\alpha_v^{\text{sat}}[t]$  correspond to the offloading proportions of the eMBB traffic to the terrestrial network, the UAVs, and the satellite respectively. Similarly, we define  $\beta_v[t] = (\beta_v^{\text{ter}}[t], (\beta_v^{\text{UAV}}[t])^T)^T$  as the offloading vector of URLLC traffic from the  $v$ -th micro base station at time slot  $t$ , where  $\beta_v^{\text{ter}}[t]$  and  $\beta_v^{\text{UAV}}[t]$  correspond to the offloaded proportions of the URLLC traffic to the terrestrial network and to the UAVs, respectively. Specifically,  $\alpha_v^{\text{UAV}}[t] = (\alpha_{v,1}^{\text{UAV}}[t], \dots, \alpha_{v,n}^{\text{UAV}}[t], \dots, \alpha_{v,N}^{\text{UAV}}[t])^T$  and  $\beta_v^{\text{UAV}}[t] = (\beta_{v,1}^{\text{UAV}}[t], \dots, \beta_{v,n}^{\text{UAV}}[t], \dots, \beta_{v,N}^{\text{UAV}}[t])^T$ , such that  $\alpha_{v,n}^{\text{UAV}}[t]$  and  $\beta_{v,n}^{\text{UAV}}[t]$  represent the proportions of the eMBB traffic and the URLLC traffic offloaded by the  $v$ -th micro base station to the  $n$ -th UAV, respectively. Accordingly, the traffic that can be offloaded by micro base station  $v$  to the satellite can be expressed as:

$$L_v^{\text{sat}}[t] = \min \left\{ \alpha_v^{\text{sat}}[t] L_v^{\text{e}}[t], R_v^{\text{sat}}[t] \right\}, \quad (12)$$

The traffic that can be offloaded by micro base station  $v$  to UAV  $n$  can be expressed as:

$$L_{v,n}^{\text{UAV}}[t] = a_{v,n}^{\text{UAV}}[t] \min \left\{ (\alpha_{v,n}^{\text{UAV}}[t] L_v^{\text{e}}[t] + \beta_{v,n}^{\text{UAV}}[t] L_v^{\text{u}}[t]), R_{v,n}^{\text{UAV}}[t] \right\}, \quad (13)$$

where  $a_{v,n}^{\text{UAV}}[t]$  is a binary variable that controls the establishment of the link with UAV  $n$  and is defined as:

$$a_{v,n}^{\text{UAV}}[t] = \begin{cases} 1, & \text{if } \Gamma_{v,n}[t] \geq \Gamma_{\text{th}}, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

The traffic that can be offloaded by micro base station  $v$  to the macro base station can be expressed as:

$$L_v^{\text{ter}}[t] = \min \left\{ (\alpha_v^{\text{ter}}[t] L_v^{\text{e}}[t] + \beta_v^{\text{ter}}[t] L_v^{\text{u}}[t]), C_v^{\text{ter}} \right\}, \quad (15)$$

In a nutshell, the total traffic offloaded by micro base station  $v$  at time slot  $t$  is given by:

$$f_v^{\text{load}}[t] = L_v^{\text{sat}}[t] + \sum_{n=1}^N L_{v,n}^{\text{UAV}}[t] + L_v^{\text{ter}}[t]. \quad (16)$$

### B. DROPPED TRAFFIC IN SAGIN

If the traffic supposed to be offloaded to the network exceeds the channel capacity, the eMBB flows are preferentially dropped to satisfy the reliability requirement of the URLLC traffic because URLLC has a stringent requirement in terms

of reliability contrarily to eMBB. In this regard, the dropped traffic per micro base station  $v$  at time slot  $t$  in SAGIN is calculated as:

$$f_v^{\text{drop}}[t] = \lambda_v^{\text{e}}[t] + \lambda_v^{\text{u}}[t] - f_v^{\text{load}}[t], \quad (17)$$

We note that  $f_v^{\text{drop}}[t] \geq 0$  since the offloaded traffic is always less than the incoming traffic.

### C. URLLC DELAY IN SAGIN

In this section, we study the mean delay  $f_v^{\text{delay}}$  experienced by the URLLC packets offloaded by micro base station  $v$  in time slot  $t$  over the terrestrial link and the UAV link [23], [28], [29]. We assume that the scheduling discipline is first come first serve (FCFS). For micro cell  $v$ , the waiting time  $D_k$  of packet  $k$ , whether it is eMBB or URLLC, scheduled after  $K$  packets during time slot  $t$  over link  $\kappa \in \{\text{ter}, \text{UAV}\}$  is given by:

$$D_{v,\kappa}^k[t] = r_{v,\kappa}^k[t] + \sum_{j=k-K}^{k-1} X_{v,\kappa}^j[t], \quad (18)$$

where  $X_{v,\kappa}^j$  is the service time of the  $j$ -th packet that arrived before the  $k$ -th packet, and  $r_{v,\kappa}^k$  is the residual service time. Using Little's formula, we obtain the average waiting time  $\bar{D}_v^\kappa$  of any packet  $k$  as:

$$\bar{D}_v^\kappa[t] = \frac{\bar{r}_v^\kappa[t]}{1 - \rho_v^\kappa[t]}, \quad \forall \kappa \in \{\text{ter}, \text{UAV}\}, \quad (19)$$

where  $\rho_v^\kappa[t]$  is the load of the micro cell  $v$  and is expressed in both links as follows:

$$\rho_v^\kappa[t] = \begin{cases} \frac{L_v^{\text{ter}}[t]}{C_v^{\text{ter}}} & \text{over the terrestrial link } (\kappa = \text{ter}) \\ \frac{L_{v,n}^{\text{UAV}}[t]}{R_{v,n}^{\text{UAV}}[t]} & \text{over the UAV link } (\kappa = \text{UAV}) \end{cases} \quad (20)$$

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

$$\bar{r}_v^\kappa[t] = \frac{\rho_v^\kappa[t]}{2} \frac{\overline{(X_v^\kappa)^2}[t]}{\bar{X}_v^\kappa[t]}. \quad (21)$$

In order to evaluate the mean service time  $\bar{X}_v^\kappa$ , we approximate our general distribution to an exponential distribution with a mean service rate  $\mu_e$  that depends on the eMBB packet length. Therefore, the first moment of the residual service time can be simplified, based on (21), as:

$$\bar{r}_v^\kappa[t] = \frac{\rho_v^\kappa[t]}{2} \frac{\frac{2}{\mu_e^2}}{\frac{1}{\mu_e}} = \frac{\rho_v^\kappa[t]}{\mu_e}. \quad (22)$$

Accordingly, the mean delay experienced by URLLC packets over both links namely UAV and terrestrial is given by:

$$f_v^{\text{delay}}[t] = \frac{\sum_{\kappa \in \{\text{ter}, \text{UAV}\}} \bar{D}_v^\kappa[t]}{\sum_{n=1}^N a_{v,n}^{\text{UAV}}[t] + 1} = \frac{\sum_{\kappa \in \{\text{ter}, \text{UAV}\}} \frac{\rho_v^\kappa[t]}{\mu_e (1 - \rho_v^\kappa[t])}}{\sum_{n=1}^N a_{v,n}^{\text{UAV}}[t] + 1}, \quad (23)$$

#### D. PROBLEM FORMULATION

We aim to adjust the association between the micro base stations with the satellite, the UAVs and the macro base stations respectively by determining:

- 1) the offloaded eMBB and URLLC traffic vectors denoted by  $\alpha_v[t]$  and  $\beta_v[t]$ , where  $v \in \{1, \dots, V\}$ ,  $t \in \{1, \dots, T\}$ .
- 2) the bandwidth resources allocated to the micro base stations denoted by  $\mathbf{b}_v[t] = \{b_v^{\text{sat}}[t], b_{v,n}^{\text{UAV}}[t]\}$ , where  $v \in \{1, \dots, V\}$ ,  $n \in \{1, \dots, N\}$ . We focus on the allocation of the bandwidth particularly because the frequency bands are limited resources.
- 3) the trajectory of UAVs denoted by  $\mathbf{L}[t]$ .

The problem can be posed as follows:

$$\begin{aligned}
 & \min_{\pi\{\alpha_v[t], \{\beta_v[t]\}, \{\mathbf{b}_v[t]\}, \{\mathbf{L}[t]\}\}} \left\{ \mathbb{E}(\sum_v f_v^{\text{drop}}[t]), \mathbb{E}(\sum_v f_v^{\text{delay}}[t]) \right\} \\
 \text{s.t. } & C1: Pr\left(\beta_v^{\text{ter}}[t]L_v^u[t] + \beta_v^{\text{UAV}}[t]L_v^u[t] < \lambda_v^u[t]\right) \leq \epsilon, \\
 & C2: \alpha_v^{\text{ter}}[t] + \alpha_v^{\text{UAV}}[t] + \alpha_v^{\text{sat}}[t] \leq 1, \\
 & C3: \beta_v^{\text{ter}}[t] + \beta_v^{\text{UAV}}[t] \leq 1, \\
 & C4: \left\{ \sum_{v=1}^{v_n} b_{v,n}^{\text{UAV}}[t], \sum_{v=1}^V b_v^{\text{sat}}[t] \right\} \leq \{b_n^{\text{UAV}}, b^{\text{sat}}\}, \\
 & C5: \|\ell_n[t] - \ell_n[t+1]\| \leq v_{\max}\delta t.
 \end{aligned} \tag{24}$$

Problem (24) is formulated as a constrained multi-objective optimization problem with constraints C1-C5. C1 guarantees the reliable offloading of URLLC traffic, where  $\epsilon$  is a small positive constant ( $\epsilon \ll 1$ ) that keeps outage probability of URLLC traffic below a negligible threshold. Meanwhile, C2 and C3 ensure that the offloaded eMBB and URLLC traffic are respectively less than the total traffic. C4 guarantees that the bandwidth allocated to the micro base stations do not surpass the total available resources in the air network  $b_n^{\text{UAV}}$  and the satellite network  $b^{\text{sat}}$  respectively. Finally, C5 ensures that the next location of the UAV is within the maximum moving range during  $\delta t$  where  $v_{\max}$  is the maximum speed of the UAVs.

We can simplify constraint C1 by using  $F_D(\cdot)$ , the cumulative distribution function (CDF) of  $\lambda_v^u[t]$ , as follows:

$$\beta_v^{\text{ter}}[t]L_v^u[t] + \beta_v^{\text{UAV}}[t]L_v^u[t] \geq F_D^{-1}(1 - \epsilon). \tag{25}$$

Since  $\lambda_v^u$  follows a Pareto distribution with parameters  $a[t]$  and  $x_m[t]$ , we express (25) as:

$$(\beta_v^{\text{ter}}[t]L_v^u[t] + \beta_v^{\text{UAV}}[t]L_v^u[t])^{a[t]} \geq \frac{x_m[t]^{a[t]}}{\epsilon}, \tag{26}$$

where  $a[t]$  is a constant that determines the distribution shape and  $x_m[t]$  is the positive scale distribution parameter.

It is challenging to solve the optimization problem formulated in (24) by using standard optimization tools, since the instantaneous traffic types and amounts incoming to the different base stations cannot be determined beforehand.

Moreover, the channel condition related to the resource allocation is not foreseeable and cannot be easily predicted. Therefore, we have recourse to DRL to learn the network and its dynamics and to handle our multi-objective problem with its various constraints imposed by SAGIN [8].

#### IV. DEEP REINFORCEMENT LEARNING FOR TRAFFIC OFFLOADING IN SAGIN

In this section, we adopt a DRL-based method to solve problem (24). Therefore, we define our Markov Decision Process and we design our algorithm based on an constrained DDPG approach.

##### A. MOTIVATIONS TO ADOPT DRL

Compared with the conventional analysis and the optimization-based approaches, which usually assume a static or perfectly characterized network model, the reinforcement learning (RL)-based-approach we proposed can refine the offloading strategy by dynamically observing the information from the system environment. We opted for reinforcement learning because the complete model of the environment's evolution is unknown in wireless networks. For instance, in the scope of this paper, the sequential load of eMBB traffic and the URLLC traffic at each time slot, the instantaneous channel path-loss, and the resource distribution are unknown to the agent in advance.

Specifically, we use DRL because it can efficiently handle the high-dimensions status in the network system we proposed and circumvent the high-computation and storage costs of the tabular-based RL. For tabular-based RL, we need to store and traverse the whole table to update the Q-value. The space complexity of the tabular-based RL algorithm is  $\mathcal{O}(S \cdot A \cdot H)$ , and the time complexity of the proposed schemes is  $\mathcal{O}(K \cdot H)$  [30], where  $S$  is the number of status of the observation space  $\mathcal{S}$ ,  $A$  represents the number of the status of the action space  $\mathcal{A}$ ,  $K$  is the number of training episodes, and  $H$  is the number of steps in each episode. For continue action space, the number of the status is infinite. Therefore, the tabular-based RL algorithm can not converge. On the other hand, for the DRL algorithm, we store only the parameters of the neural networks. The space complexity is  $\mathcal{O}(\sum_{f=1}^F n_f)$  and the time complexity is  $\mathcal{O}(\sum_{f=1}^F n_f \cdot n_{f-1})$  [31], where  $n_f$  is the number of neural units in fully-connected layer  $f$ , and  $F$  is the total layer of the network. With the instantaneous observations as input, the well trained deep neural network (DNN) model predicts the optimized UAV steps and the efficient resource allocation. The actions of the system considered in this paper are extracted from the continuous action space, which can not be determined by the conventional DRL. Therefore, we apply DDPG algorithm, which is an advanced DRL algorithm that embeds the actor-critic method and combines the value based-RL with the policy based-RL [17]. The DDPG algorithm is extended from the deterministic policy gradient (DPG) algorithm [32] and takes advantage of experience replay and slow-learning target networks from deep Q-network (DQN) [33].

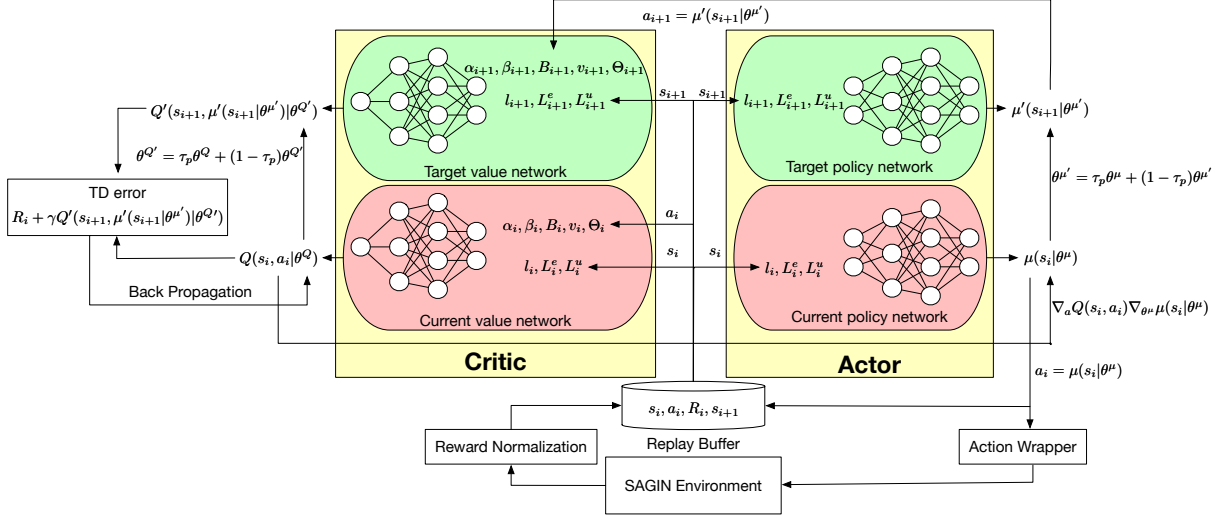


Fig. 2: DDPG-based heterogeneous traffic offloading approach

## B. MARKOV DECISION PROCESS FORMULATION

The first step is to reformulate the proposed optimization problem (24) as a Markov decision process (MDP). Specifically, in each time slot, the DRL agent operated on the central server collects the information from the SAGIN environment and makes a decision based on the observations. We define the MDP with 4 elements tuple  $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$ , where  $\mathcal{S}$  is the observation space,  $\mathcal{A}$  represents the action space,  $\mathcal{T}$  is the transition matrix and  $\mathcal{R}$  is the possible reward set that can be obtained from the SAGIN environment. The definition of each element in the tuple is described as follows:

- **Observation space:**  $\mathcal{S} \triangleq \{\ell_n[t], L_v^u[t], L_v^e[t]\}$ . The observation space includes continuous variables, such as the location of the UAVs  $\{\ell_n[t]\}$ , the load set of eMBB traffic  $\{L_v^u[t]\}$ , and the load set of URLLC traffic  $\{L_v^e[t]\}$ .
- **Action space:**  $\mathcal{A} \triangleq \{\alpha_v[t], \beta_v[t], b_v[t], M_n[t]\}$  where  $M_n[t]$  denotes the moving action of the  $n$ -th UAV.  $M_n[t]$  consists of the navigation speed  $v_n[t]$ ,  $v_n[t] \in [0, v_{\max}]$  and the rotation angle  $\Theta_n[t]$ ,  $\Theta_n[t] \in [0, 2\pi]$ ,  $v_n[t]$  and  $\Theta_n[t]$  are continuous variables. The moving action  $M_n[t]$  is defined such that the constraint C5 in problem (24) is satisfied.
- **Transition matrix:**  $\mathcal{T} \triangleq \mathcal{S} \times \mathcal{A} \times \mathcal{S}$ . The value of the transition matrix is binary and is determined by the constraints (C2-C5) in problem (24) such that:

$$T(s[t], a[t], s[t+1]) = \begin{cases} 1, & \text{if (C2-C5) are satisfied,} \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

In this regard, the transition matrix is sparse and most of the action exploration are wasted. To accelerate the training process, we develop a wrapper adhered to the SAGIN environment. The actions are compressed beforehand with linear normalization and soft-max nor-

malization as follows:

$$\alpha_v^k[t] \leftarrow \frac{\alpha_v^k[t]}{\sum_{k \in (\text{ter}, \text{UAV}, \text{sat})} \alpha_v^k[t]}, \quad (28)$$

$$\beta_v^k[t] \leftarrow \frac{\beta_v^k[t]}{\sum_{k \in (\text{ter}, \text{UAV})} \beta_v^k[t]}, \quad (29)$$

$$K_v[t] \leftarrow \frac{e^{K_v[t]}}{\sum_{v \in V} e^{K_v[t]}}, \quad K_v[t] \in \{B_n^{\text{UAV}}[t], B^{\text{sat}}[t]\}. \quad (30)$$

- **Reward Set  $\mathcal{R}$ :** We design the reward function to jointly minimize the traffic dropping and the network latency while guaranteeing the availability of URLLC traffic. We note that minimizing the dropped traffic is equivalent to maximizing the offloaded traffic for a certain amount of traffic at each time slot. Therefore, and to avoid a persistent negative reward, we involve the offloaded traffic  $f_v^{\text{load}}[t]$  rather than the dropped traffic  $f_v^{\text{drop}}[t]$  in the definition of the reward function. Hence, the value of the reward function in time slot  $t$  is defined as:

$$R[t] = \omega_1 \sum_v f_v^{\text{load}}[t] - \omega_2 \sum_v f_v^{\text{delay}}[t] - \omega_3 a_{\text{av}}^u[t]. \quad (31)$$

The non-linear constraint C1 of problem (24) is considered in the reward function  $R[t]$  by adding a sufficient penalty through  $a_{\text{av}}^u[t]$ , which is the indicator of the reliable offloading of URLLC traffic. If constraint C1 in problem (24) is satisfied,  $a_{\text{av}}^u[t]$  is equal to zero. Otherwise,  $a_{\text{av}}^u[t]$  is equal to 1. By considering  $\omega_3 \gg \omega_1$  and  $\omega_3 \gg \omega_2$ , the DRL agent is assigned a big penalty if the reliability is not guaranteed. We note that the offloading function  $f_{\text{load}}[t]$  and the latency function  $f_{\text{delay}}[t]$  are normalized for a joint optimization, stable training and faster convergence. We make the trade-off between the offloading and the latency, by tuning the parameters  $w_1$  and  $w_2$ . In general, the total offloaded traffic increases at

the expense of latency and vice versa. If we care more about the total offloaded traffic, then  $w_1$  is set to be slightly larger than  $w_2$ . However, if the network is more sensitive to latency, then  $w_2$  is larger than  $w_1$ .

### C. CONSTRAINED DDPG ALGORITHM DESIGN

To design the sequential strategy, we optimize the long-term accumulative reward by evaluating the state-action value function, which is given by:

$$Q_\pi[t](s, a) = \mathbb{E} \left( \sum_{T_i=1}^{T-t} \gamma^{T_n} R[t + T_n] \middle| s, a, \pi \right), \quad (32)$$

where  $0 < \gamma < 1$  is the discount factor.

The DDPG algorithm is implemented with an actor-critic approach. The actor policy network  $\mu(s|\theta^\mu)$  specifies the action  $a$  given the state  $s$  currently occupied by the agent, where  $\theta^\mu$  shows the weights of the actor network. The critic value network  $Q(s, a|\theta^Q)$  specifies the temporal difference (TD) error to criticize the actions made by the actor, where  $\theta^Q$  represents the weights of the critic network. The network architecture is depicted in Fig. 2 and a detailed description of the training process is presented in Algorithm 1. To get a stable training, we used DNN rather than the table-based value estimation method because the observation space is highly-dimensional and the action space is continuous. We conducted also the experience replay technique because it provides buffer for mini-batch sampling. Moreover, we imported the target models by using different networks for both critic and actor training processes to avoid the harmful correlations.

During the training process, the critic value network  $Q(s, a|\theta^Q)$ , the actor policy network  $\mu(s|\theta^\mu)$ , the environment status  $S$  and the replay buffer are first randomly initialized (Line 2). Then, the target networks  $Q'(s, a|\theta^{Q'})$  and  $\mu'(s|\theta^{\mu'})$  are initialized by reproducing the weights of the networks  $Q(s, a|\theta^Q)$  and  $\mu(s|\theta^\mu)$  such that (Line 3):

$$\theta^{Q'} \leftarrow \theta^Q, \quad (33)$$

$$\theta^{\mu'} \leftarrow \theta^\mu. \quad (34)$$

The action is selected from a continuous action space by:

$$a[t] = \mu'(s[t]|\theta^{\mu'}) + G[t], \quad (35)$$

where  $G[t]$  is a random Gaussian noise added to balance the exploitation of the optimized action and the exploration of the environment (Line 5).

To equivalently reformulate problem (24) as MDP and solve the problem, we propose a constrained DDPG RL algorithm. We develop the action wrapper to re-scale the elements of actions by (28), (29), (30). The traffic is offloaded to different types of networks and the bandwidth resource is assigned to different macro base stations, while the sum of the proportions equals to one. In this manner, the constraints (C2-C4) of problem (24) can be satisfied (Line 6). If the offloaded traffic exceeds the capacity of the network, we preferentially drop the eMBB traffic due to the stringent reliability requirement

of URLLC traffic (Lines 6-8). Another borderline case occurs when the SNR of the link between the UAV and the micro BS is lower than the SNR threshold. The link should not be set up and the traffic should not be offloaded through that given link, such as  $R_{v,n}^{\text{UAV}} = 0$ . In order to satisfy the QoS, the traffic should be served by other links or network segments (Lines 9-11).

By executing the action  $a[t]$ , the agent observes the new state  $s[t+1]$  from the SAGIN environment and obtains the reward  $R[t]$  (Line 12). The interaction information is stored as a tuple  $(s[t], a[t], R[t], s[t+1])$  in the replay buffer for further sampling (Line 13). We train the model by using batch normalization. The samples of each mini-batch are randomly selected in the replay buffer (Line 14). The adaptive moment estimation (Adam) algorithm [34] is the deployed optimizer for both critic network and actor network.

In each mini-batch, the critic value network can be updated by minimizing the loss (Line 15):

$$L = \frac{1}{I} \sum_{i=1}^I (y_i - Q(s_i, a_i|\theta^Q))^2, \quad (36)$$

where

$$y_i = R_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'}), \quad (37)$$

is the target value network, and  $I$  is the batch size. Index  $i \in I$  describes the sample selected from the replay buffer. The weight of the value network is given by:

$$\begin{aligned} \theta^Q &= \theta^Q - \alpha_Q \nabla_{\theta^Q} L \\ &= \theta^Q - \frac{\alpha_Q}{I} \sum_{i=1}^I 2(y_i - Q(s_i, a_i|\theta^Q)) \nabla_{\theta^Q} Q(s_i, a_i|\theta^Q), \end{aligned} \quad (38)$$

where  $\alpha_Q$  is the actor learning rate.

The actor policy network is updated by maximizing the expected accumulative reward (Line 16):

$$J(\theta^\mu) = \mathbb{E}[Q(s_i, a_i|\theta^\mu)]. \quad (39)$$

Consequently, the weight of the policy network is updated as:

$$\begin{aligned} \theta^\mu &= \theta^\mu + \alpha_\mu \nabla_{\theta^\mu} J(\theta^\mu) \\ &= \theta^\mu + \frac{\alpha_\mu}{I} \sum_{i=1}^I \nabla_a Q(s_i, a_i) \nabla_{\theta^\mu} \mu(s_i|\theta^\mu). \end{aligned} \quad (40)$$

Equation (40) is proved by [32]. The target value network and the target policy network are fixed for several steps and then are updated as (Line 17):

$$\theta^{Q'} = \tau_p \theta^Q + (1 - \tau_p) \theta^{Q'}, \quad (41)$$

$$\theta^{\mu'} = \tau_p \theta^\mu + (1 - \tau_p) \theta^{\mu'}, \quad (42)$$

where  $\tau_p \ll 1$  is the update coefficient, which describes how much the target networks are updated by the current networks.

The training process is operated until convergence, which is validated in Section V. Once well trained, the model can be installed in the UAVs embedded with computational

**Algorithm:** Constrained DDPG RL Algorithm**Input :**  $\{L_v^c[t], L_v^u[t]\}, \epsilon$ **Output:**  $\theta^Q, \theta^\mu$ **repeat**

Initialize the state  $s$ , the critic value network  $Q(s, a|\theta^Q)$ , the actor policy network  $\mu(s, a|\theta^\mu)$ , and the replay buffer

Initialize the target network  $Q(s, a|\theta^{Q'})$  and  $\mu(s, a|\theta^{\mu'})$ :  $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$

**for**  $t < T$  **do**

Select action  $a[t] = \mu(s[t]|\theta^\mu)$  from  $\mathcal{A}$

Compress the action elements by Eq. (28),

(29), (30) to satisfy the constraints C2-C4.

**if** the load of traffic exceeds the capacity **then**

└ drop preferentially the eMBB traffic

**if**  $\Gamma_{v,n}[t] < \Gamma_{th}$  **then**

└ set  $R_{v,n}^{UAV} = 0$

Obtain the reward  $R[t]$  and observation

$s[t+1]$

Store the transition  $(s[t], a[t], r[t], s[t+1])$  in the replay buffer.

Sample randomly  $I$  transitions

$(s_i, a_i, r_i, s_{i+1})$  from the replay buffer constituting a mini-batch

Update the critic value network by minimizing the loss:  $L = \frac{1}{I} \sum_{i=1}^I (y_i - Q(s_i, a_i|\theta^Q))^2$   
 $\theta^Q = \theta^Q - \alpha_Q \nabla_{\theta^Q} L$

Update the actor policy network by maximizing the expected accumulative reward:  $J(\theta^\mu) = E[Q(s_i, a_i|\theta^\mu)]$   
 $\theta^\mu = \theta^\mu + \alpha_\mu \nabla_{\theta^\mu} J(\theta^\mu)$

Update the target networks:

$\theta^{Q'} = \tau_p \theta^Q + (1 - \tau_p) \theta^{Q'}$

$\theta^{\mu'} = \tau_p \theta^\mu + (1 - \tau_p) \theta^{\mu'}$

**until** convergence;

TABLE 1: System Parameters

| Parameter                                         | Numerical Value |
|---------------------------------------------------|-----------------|
| Macro BS Bandwidth ( $B^w$ )                      | 100 (MHz)       |
| Resource Block Bandwidth ( $B^{RB}$ )             | 0.18 (MHz)      |
| Noise power density ( $N_0$ )                     | -174 (dbm/Hz)   |
| Micro BS transmission power ( $W_v$ )             | 20 (W)          |
| URLLC Tolerable Outage Probability ( $\epsilon$ ) | $10^{-6}$       |
| eMBB packet size ( $\frac{1}{\mu_c}$ ) [37]       | 100 (b)         |
| S-curve parameters ( $b_1, b_2$ )                 | 9.6, 0.16       |
| Pass-loss for N-LoS transmission ( $\xi_{NLoS}$ ) | 20 (dB)         |
| Pass-loss for LoS transmission ( $\xi_{LoS}$ )    | 1 (dB)          |
| Maximum speed of the UAVs ( $v_{max}$ )           | 1 (km/min)      |
| Satellite bandwidth ( $b^{sat}$ ) [36]            | 60 (MHz)        |

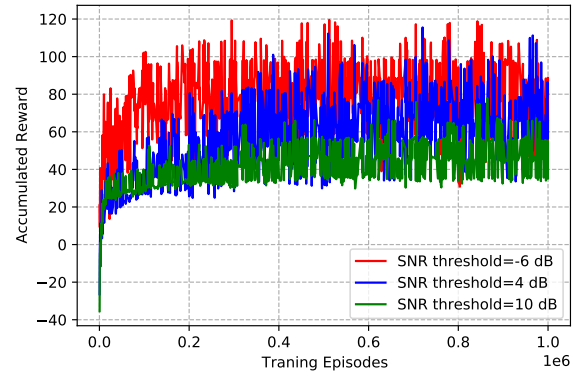


Fig. 3: Accumulated Reward over Training Episodes.

system. Given the observation from the environment, the time to process the inputs and generate the output actions is negligible. Consequently, the execution of the model can be regarded as a real-time implementation.

## V. RESULTS AND ANALYSIS

In this section, we evaluate the performance of our system in terms of latency and availability. For our simulations, we consider two macro base stations separated by a distance  $d = 40$  km. We deploy 11 micro base stations according to a Point Poisson Process with a density  $\lambda_p = 0.004$ . Initially, the UAVs are randomly positioned. We run our simulations with Python 3.7 and Torch 1.3 on a supercomputer server with Linux 3.10 operating system, NVIDIA TITAN, 1 GPU and 6 CPUs for each task. A generic RL environment for SAGIN is developed based on open AI gym framework [35]. For each neural network, we adopted a multi-layer perception structure with two layers and 64 neurons in each layer. The

commonly used Ornstein Uhlenbeck noise is used as action noise, where the mean is set as zero and the variance equals to 0.5. We trained the proposed DRL-based method for 1M episodes, each of which has ten epochs (i.e. time slots). After training, we test the algorithm for a period of  $T = 10$  epochs. We set the reward function weights as  $\omega_1 = 1$ ,  $\omega_2 = 1$  and  $\omega_3 = -10$ . The critic learning rate and the actor learning rate are set as  $10^{-3}$  and  $10^{-4}$ , respectively. The discount factor is set as 0.95. The simulations parameters are detailed in Table 1. We note that the used carrier signal to noise power mentioned in Table 1 is relative to the Telesat LEO satellites constellation [36].

We start first by examining the behaviour of our algorithm over different epochs. Fig.3 shows the accumulated reward over time for different SNR thresholds of the UAVs links. We remark that the accumulated reward increases monotonically over time until 200000 epochs. This observation is due to the fact that the more we train the model, the better the agent learns the environment until it converges after 200000 epochs. We also remark that the reward increases when the SNR threshold of the UAV links decreases. This observation is due to the fact that when the SNR threshold is lower, more traffic is accepted into the network over the UAV links. Consequently, our offloading approach offloads more traffic to the backhaul and hence obtains a higher reward as

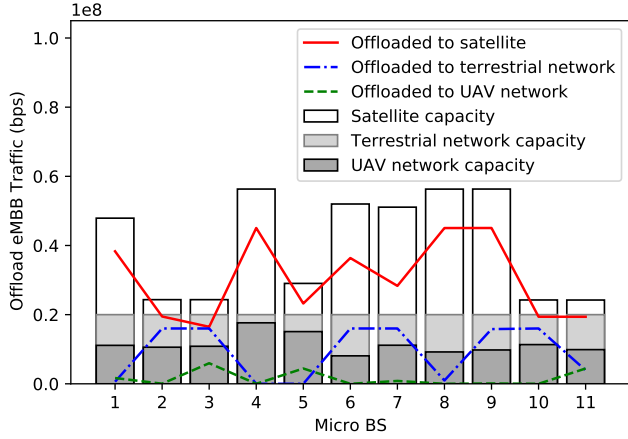


Fig. 4: Distribution of the eMBB Offloaded Traffic over 3 Links.

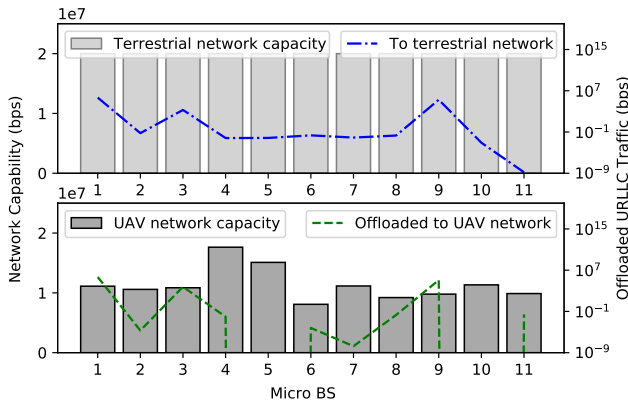


Fig. 5: Distribution of the URLLC Offloaded Traffic over 2 Links.

anticipated by equation (31).

Then, we evaluate the performance of our offloading approach through three simulations, according to the variation of the time, the SNR threshold of the UAV links  $\Gamma_{th}$  and the number of UAVs. The optimization problem presented in (24) is a non-convex and non-linear programming problem, which is hard to solve in general. Obtaining an optimal solution even for a single time slot will yield prohibitive time complexity. Furthermore, the time and computational resource consumption are not affordable by implementing greedy or DQN algorithm since both the state and the action need to be traversed in continuous spaces. Therefore, we compare our results to two benchmarks:

- 1) *Random offloading approach*: the first benchmark is a random offloading approach where the offloading proportions and the resource allocation parameters are randomly determined. The comparison to this first benchmark underlines the key role of DRL to achieve a better network performance.
- 2) *Integrated satellite terrestrial network (ISTN) offloading approach* [23]: the second benchmark is an ISTN offloading approach, where all the base stations are in the ground covered by one satellite. For a fair comparison, the ground base stations in ISTN have an equivalent capacity equal to the sum

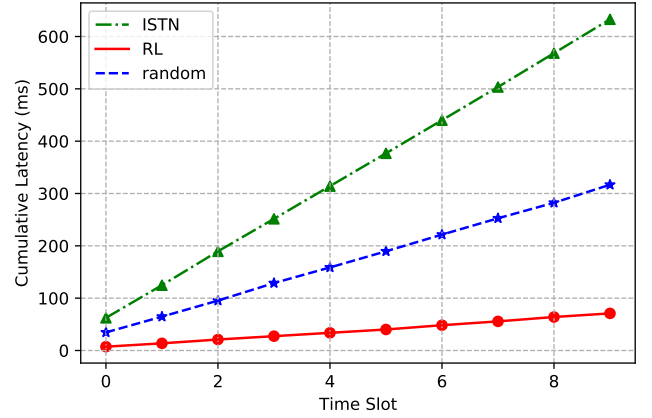


Fig. 6: Cumulative URLLC Latency as function of Time.

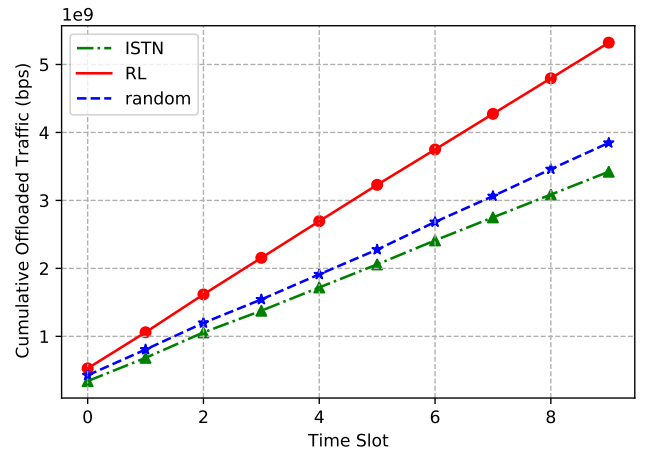


Fig. 7: Cumulative Offloaded Traffic as function of Time.

of the capacities of the ground base stations and the data rates of the UAVs in SAGIN. In the ISTN offloading approach, the authors offload the URLLC traffic to the terrestrial backhaul only and offload the eMBB traffic to the satellite backhaul and to the terrestrial backhaul [23]. The comparison to this second benchmark highlights the importance of UAVs mobility to improve the network's QoS. Indeed, ISTN has only ground base stations. However, SAGIN has ground base stations and air base stations, which are the UAVs.

#### A. VARIATION OF TIME

In the first simulation, we examine the behaviour of our offloading approach through 10 time slots. We fix the number of UAVs at 3 and the SNR threshold for the UAV links at  $\Gamma_{th} = 0.1$ .

First, we compare the used capacities by our offloading approach for both traffic types respectively in the three links, namely, satellite, UAVs and terrestrial links. Because eMBB is generated in huge volumes (in the order of  $10^7$  bps) compared to URLLC (in the order of  $10^4$  bps), we can see in Fig.4 and Fig.5 that most of the available capacity in the UAVs links and in the terrestrial links is taken by the eMBB

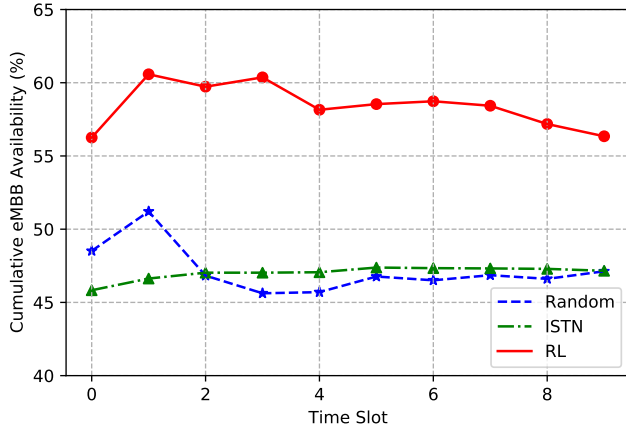


Fig. 8: Cumulative eMBB Availability as function of Time.

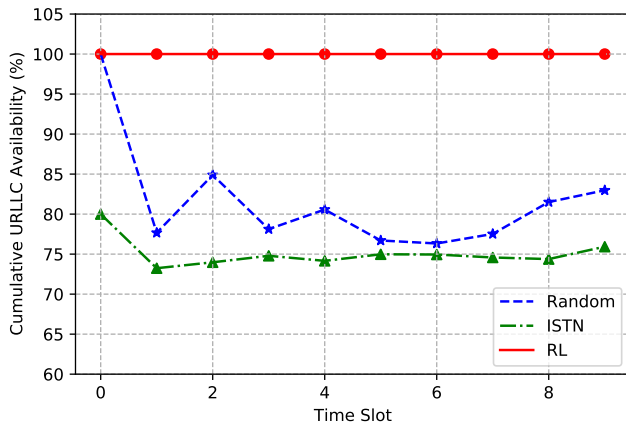


Fig. 9: Cumulative URLLC Availability as function of Time.

traffic. We highlight also the pivotal role of the satellite to deliver the eMBB traffic to the backhaul as illustrated in Fig.4. Indeed, approximately 68% of the total eMBB traffic is offloaded to the satellite against 30% to the terrestrial links and 1% to UAV links. As far as URLLC is concerned, the offloaded traffic is almost equally offloaded to the terrestrial links and to the UAV links as illustrated in Fig.5.

Second, we study the variation of the URLLC delay experienced in the terrestrial link and in the UAV links. Based on the results depicted in Fig.6, we observe that our offloading approach reduces substantially the total latency experienced by the URLLC packets over 10 time slots compared to both benchmarks.

Then, we study the total amount of the offloaded traffic in the three links; namely satellite, UAV and terrestrial. As illustrated in Fig.7, our offloading approach boosts more importantly the total offloaded traffic to the backhaul compared to both benchmarks, especially for higher traffic amounts (i.e. at the last time slots). This proves its capability to operate decently in dense networks.

Afterwards, we study the network availability for both traffic types. Therefore, we evaluate the percentage of the

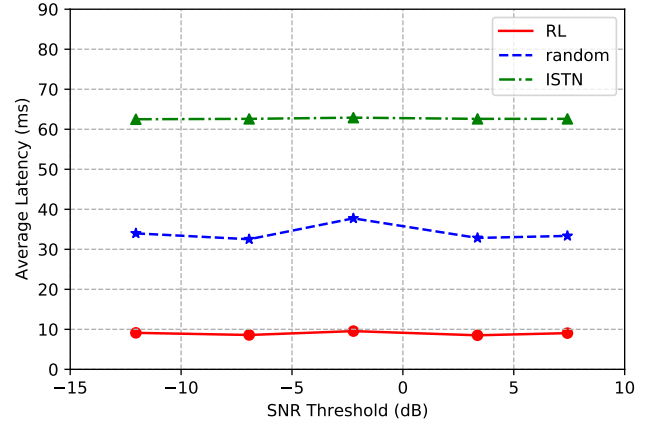


Fig. 10: Average URLLC Latency as function of UAV SNR Threshold.

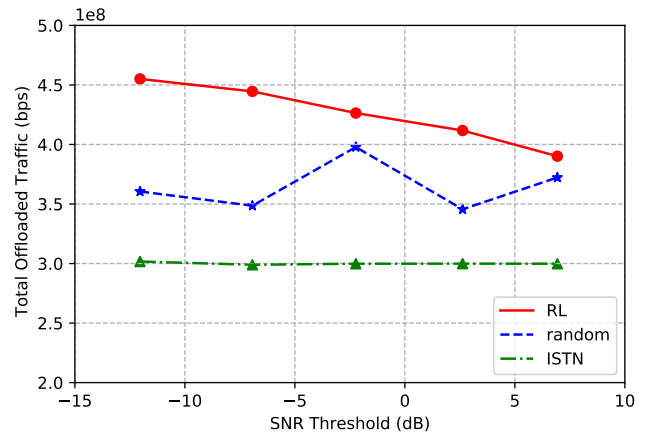


Fig. 11: Total Offloaded Traffic as function of UAV SNR Threshold.

successfully offloaded traffic out of the totally sent traffic.

In the results presented in Fig.8 and Fig.9, we notice that our offloading approach surpasses both benchmarks in terms of the offered availability to both eMBB and URLLC traffic. Interestingly, our offloading approach succeeds to achieve an availability rate of 100% over all the time slots for the URLLC traffic. This result fulfills perfectly the requirements of this latter slice.

## B. VARIATION OF THE SNR THRESHOLD OF UAV LINKS

In the second simulation, we investigate the influence of the UAV link quality on the performance of our system. We fix the number of UAVs at 3 and we vary the SNR threshold for the UAV links  $\Gamma_{th}$  between -15 dB and 10 dB. We study first the variation of the URLLC delay experienced in the terrestrial link and in the UAV links. The results depicted in Fig.10 show that our offloading approach surpasses distinctly both benchmarks. We note also that the variation of  $\Gamma_{th}$  does not alter significantly the URLLC latency for all three models. This observation is due to the fact that the decrease of the offloaded traffic for increasing SNR thresholds  $\Gamma_{th}$  is

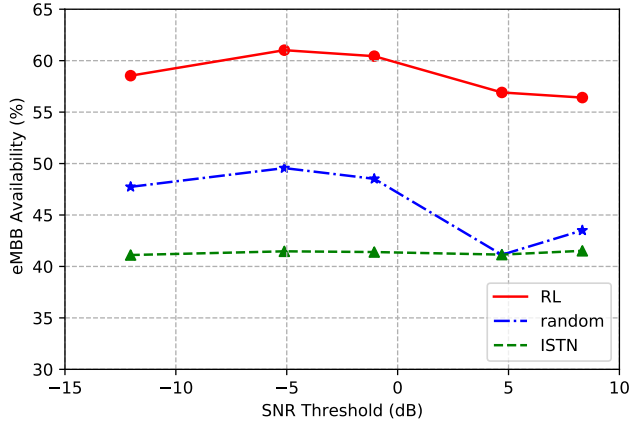


Fig. 12: Mean eMBB Availability as function of UAV SNR Threshold.

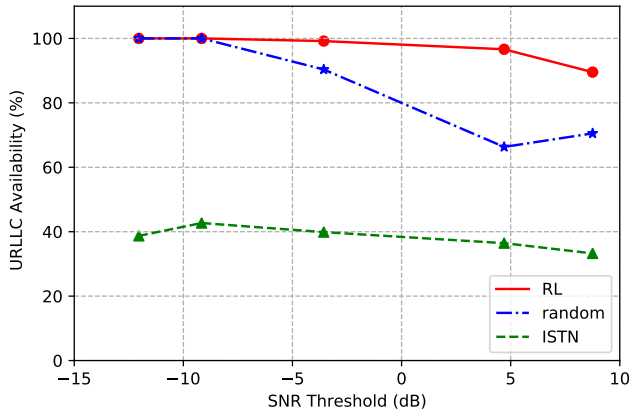


Fig. 13: Mean URLLC Availability as function of UAV SNR Threshold.

not important enough to impact the latency.

Then, we study the total amount of the offloaded traffic to the three links (i.e. satellite, UAV and terrestrial) as depicted in Fig.11. As anticipated, we notice that the total offloaded traffic decreases when  $\Gamma_{th}$  increases for our offloading approach. This observation is due to the fact that less traffic is accepted in the UAVs links when  $\Gamma_{th}$  increases. We remark also that our offloading approach has the highest offloaded traffic even for the lowest  $\Gamma_{th}$  values (i.e. the UAV links are mostly established). This observation emphasizes the validity of our model in dense networks.

Afterwards, we study the network availability for both traffic types. On the one hand, the results presented in Fig.12 show that our offloading approach ameliorates the eMBB availability, especially compared to ISTN (with 20% on average) and to the random approach also (around 15% on average). It is noteworthy that eMBB availability is not strongly affected by the SNR threshold in ISTN. On the other hand, the results presented in Fig.13 underline that our offloading approach has advantage over both benchmarks in terms of URLLC availability, especially for the lowest

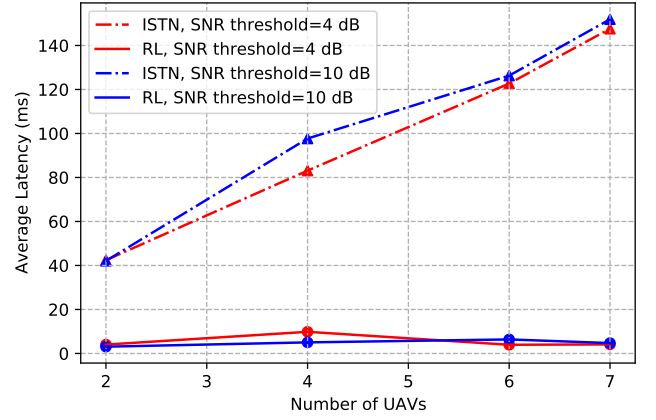


Fig. 14: Average URLLC Latency as function of the Number of UAVs.

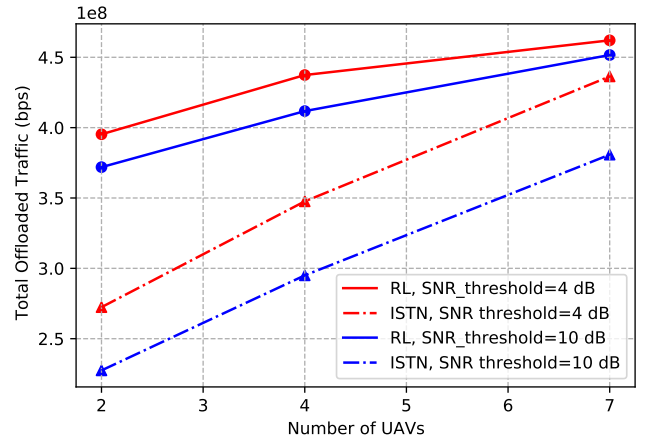


Fig. 15: Total Offloaded Traffic as function of the Number of UAVs.

$\Gamma_{th}$  where availability is stable at 100%. This observation is due to the fact that even a mediocre signal quality can establish communication links with UAVs. These results are also in line with the total offloaded traffic results depicted in Fig.7, since higher availability levels are achieved when  $\Gamma_{th}$  decreases.

### C. VARIATION OF THE NUMBER OF UAVS

In the third simulation, we inspect how the UAVs number impacts our system performance. We vary the number of the present UAVs between 2 and 7 and we consider the SNR threshold of UAVs links as a parameter rated at 4 dB and 10 dB respectively. We study first the variation of the URLLC delay experienced in the terrestrial link and in the UAV links. Based on the results depicted in Fig.14, we observe that our offloading approach helps decrease significantly the URLLC latency compared to ISTN. This observation reveals the key role that the UAVs mobility plays to enhance the experienced latency in the network. We notice also that the URLLC latency in ISTN is seriously affected by the number of the present UAVs because more traffic is offloaded to

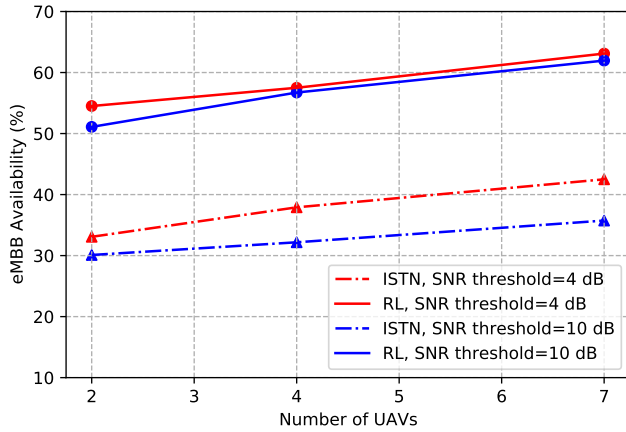


Fig. 16: Mean eMBB Availability as function of the Number of UAVs.

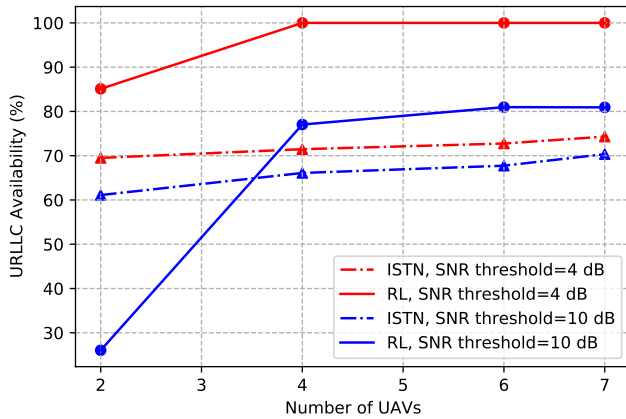


Fig. 17: Mean URLLC Availability as function of the Number of UAVs.

the backhaul links that are all static. Then, we study the total offloaded traffic in the three links (i.e. satellite, UAV and terrestrial). As illustrated in Fig.15, we remark that our offloading approach outperforms ISTN specially when a low number of UAVs are entailed. We note also that the SNR threshold parameter influence more clearly the results for ISTN than our offloading approach. This observation is due to the benefits of the UAV trajectory design adopted in our DRL algorithm, which places the UAVs in the most loaded areas and maintains a close total offloaded traffic for different SNR thresholds.

Afterwards, we study the network availability for both traffic types. As depicted in Fig.16, we remark that our offloading approach improves considerably the eMBB availability compared to ISTN for different SNR threshold of the UAV links. We also notice that the eMBB availability increases with the number of UAVs because more traffic can be offloaded to the backhaul, which is endorsed by the capacities of the added UAVs. As illustrated in Fig.17, our offloading approach outperforms ISTN in terms of URLLC availability for different SNR threshold of the UAV links. Particularly, we

note that URLLC needs are not met by ISTN for both SNR thresholds where the availability is less than 80%. However, our DRL algorithm meets the URLLC requirements once 4 UAVs or more are present in the network, and achieves a 100% rate exclusively for  $\Gamma_{th} = 4$  dB.

## VI. CONCLUSION

In this paper, we proposed a heterogeneous traffic offloading approach in SAGIN to meet the various requirements of 5G slices in terms of high data rates, delay and reliability. According to this approach, the eMBB traffic is offloaded to the satellite and to UAVs to satisfy its need for high throughput. However, the URLLC traffic is only offloaded to UAVs to satisfy its need for ultra-low latency. Our results stressed the importance of the integration between satellite, UAVs and terrestrial network to fulfill a better QoS for different slices and traffic types. Precisely, our offloading approach succeeded to enhance substantially the availability and the latency experienced respectively by eMBB and URLLC traffic compared to ISTN and spotlighted hence the power of UAVs mobility. Our results stressed also the importance of DRL as a key tool to learn more easily the network dynamics than the optimization approaches and to allocate efficiently the available resources. For future work, we plan to investigate the energy consumption of partial observable multi-UAV system in this network dynamic context with heterogeneous traffic types and requirements. Moreover, we will focus on the massive machine type communications (mMTC) traffic in addition to eMBB and URLLC to cover simultaneously and efficiently the conflicting needs of the 5G slices.

## REFERENCES

- [1] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, Sep. 2018.
- [2] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.
- [3] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5G URLLC: Design challenges and system concepts," in *Proc. IEEE ISWCS*, Lisbon, Portugal, Aug. 2018, pp. 1–6.
- [4] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, Jul. 2020.
- [5] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 4, pp. 2714–2741, May 2018.
- [6] N. Kato, Z. M. Fadlullah, F. Tang, B. Mao, S. Tani, A. Okamura, and J. Liu, "Optimizing space-air-ground integrated networks by artificial intelligence," *IEEE Wirel. Commun.*, vol. 26, no. 4, pp. 140–147, Aug. 2019.
- [7] N. Cheng, W. Quan, W. Shi, H. Wu, Q. Ye, H. Zhou, W. Zhuang, X. S. Shen, and B. Bai, "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wirel. Commun.*, vol. 27, no. 1, pp. 178–185, Feb. 2020.
- [8] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2059–2070, Sep. 2018.
- [9] J. Cui, Z. Ding, Y. Deng, A. Nallanathan, and L. Hanzo, "Adaptive UAV-trajectory optimization under quality of service constraints: a model-free solution," *IEEE Access*, vol. 8, pp. 112 253–112 265, Jun. 2020.

- [10] H. Wu, Z. Wei, Y. Hou, N. Zhang, and X. Tao, "Cell-edge user offloading via flying UAV in non-uniform heterogeneous cellular networks," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 4, pp. 2411–2426, Apr. 2020.
- [11] F. Jiang, K. Wang, L. Dong, C. Pan, W. Xu, and K. Yang, "Deep learning based joint resource scheduling algorithms for hybrid MEC networks," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6252–6265, Nov. 2019.
- [12] Z. Yu, Y. Gong, S. Gong, and Y. Guo, "Joint task offloading and resource allocation in UAV-enabled mobile edge computing," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3147–3159, Apr. 2020.
- [13] 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 Trans. Wirel. Commun.*, vol. 18, no. 1, pp. 47–62, Jan. 2018.
- [14] 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 J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.
- [15] I. AlQerm and B. Shihada, "Energy efficient traffic offloading in multi-tier heterogeneous 5G networks using intuitive online reinforcement learning," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 691–702, Sep. 2019.
- [16] L. Zhang, A. Celik, S. Dang, and B. Shihada, "Energy-efficient trajectory optimization for uav-assisted iot networks," *IEEE Transactions on Mobile Computing*, 2021.
- [17] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, Sep. 2015.
- [18] 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*, vol. 37, no. 4, pp. 881–895, 2019.
- [19] A. Anand, G. De Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," pp. 1970–1978, 2018.
- [20] 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*, pp. 1271–1273, 2018.
- [21] 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)*. IEEE, 2018, pp. 1–7.
- [22] 3rd Generation Partnership Project (3GPP), "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," (3GPP), Technical Report (TR) 45.820, 03 2015, version 0.3.0. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2719>
- [23] W. Abderrahim, O. Amin, M.-S. Alouini, and B. Shihada, "Latency-aware offloading in integrated satellite terrestrial networks," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 490–500, Apr. 2020.
- [24] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
- [25] R. I. Bor-Yaliniz, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–5.
- [26] W. Nam, D. Bai, J. Lee, and I. Kang, "Advanced interference management for 5g cellular networks," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 52–60, 2014.
- [27] B. Zhao, G. Ren, X. Dong, and H. Zhang, "Spatial group based optimal uplink power control for random access in satellite networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7354–7365, 2020.
- [28] J. Li and Y. Han, "Optimal resource allocation for packet delay minimization in multi-layer UAV networks," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 580–583, Mar. 2017.
- [29] L. Kleinrock, *Queueing systems, volume 2: Computer applications*. New York: Wiley, Apr. 1976, vol. 66.
- [30] C. Jin, Z. Allen-Zhu, S. Bubeck, and M. I. Jordan, "Is Q-learning provably efficient?" *arXiv preprint arXiv:1807.03765*, 2018.
- [31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [32] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. ACM ICML*, Beijing, China, Jun. 2014, pp. 387–395.
- [33] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al.,

- "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, Dec. 2014.
- [35] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, Jun. 2016.
- [36] 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," *Elsevier Acta Astronaut.*, vol. 159, pp. 123–135, Jun. 2019.
- [37] 3rd Generation Partnership Project (3GPP), "5G : Study on Scenarios and Requirements for Next Generation Access Technologies," (3GPP), Technical Report (TR) 38.913, May 2017, version 14.2.0.



LIANG ZHANG ((S'19) received the B.Sc. degree in physics from University of Science and Technology Beijing, China, in 2016, the M.Sc. degree in electrical engineering from the King Abdullah University of Science and Technology, Saudi Arabia, in 2018. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the King Abdullah University of Science and Technology. Her research interests include flying network, Internet of Things, deep learning and reinforcement learning.



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, edge computing, network virtualization, software defined networking and recently satellite communications and machine learning.



BASEM SHIHADA (SM'12) received the Ph.D. degree in computer science from the University of Waterloo, Waterloo, ON, Canada. He is an Associate and a Founding Professor with the Computer, Electrical and Mathematical Sciences Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. In 2009, he was appointed as a Visiting Faculty Member with the Department of Computer Science, Stanford University, Stanford, CA, USA. His current research interests include 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.

...