Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Thesis in ICT for Internet and Multimedia

Simulation Study, Design and Implementation of 5G NR Layer-2 Protocols for Non-Terrestrial Networks

Master Candidate

Francesco Rossato

Student ID 2082121

Supervisor

Professor Michele Zorzi

University of Padova

Co-supervisor

Professor Marco Giordani

University of Padova

Dr. Matteo Pagin

University of Padova

Graduation Date 03-09-2024
Academic Year 2023/2024

A chi mi vuole bene.

Abstract

The advent of 5G technology has revolutionized the world of mobile communication, offering high throughput, low latency and high reliability. As the demand for connectivity continues to grow, non-terrestrial networks (NTNs) have been identified as crucial solutions for future 5-6G deployments, complementing terrestrial networks with the aim of increasing resiliency and providing connectivity in remote areas. The characteristics of NTNs, such as the long distances involved, the high propagation delay and the larger cell footprint, introduce new and more critical complexities than in terrestrial networks, which require a thorough re-definition and re-design of most of the protocol stack.

After a careful literature review, in this thesis we identify key characteristics and constraints of NTNs, highlighting strengths and limitations of legacy 3GPP NR protocols in the NTN scenario. In light of this, we propose and implement new 5G NR Layer-2 protocols specifically tailored to NTN, focusing in particular on retransmission management via Hybrid Automatic Repeat reQuest and scheduling. The proposed solutions have been implemented in the ns-3 end-to-end full-stack simulator by extending the ns3-ntn module, which guarantees the realism and accuracy of the results.

This work resulted in a major improvement of the ns-3 ntn module, whose implementation of the scheduler is now capable of tolerating scenarios with high propagation delays typical of NTNs. The numerical results obtained from the simulations show a measurable improvement in terms of reliability and throughput. Problems such as expiring timers, inflated buffer status reports, and unnecessary retransmissions were documented and addressed.

We then focused the attention to the implementation of the HARQ protocol, that initially caused the throughput to drop significantly. We argued that its use is still beneficial in contexts of low SNR, while implementing modifications that increased the maximum achievable throughput.

Contents

List of Figures

List of Acronyms

TN terrestrial network

NTN non-terrestrial network

HAP high-altitude platform

UAV unmanned aerial vehicle

MEO medium earth orbit

HARQ hybrid automatic repeat request

ARQ automatic repeat request

NR new radio

LEO low earth orbit

GEO geosynchronous equatorial orbit

UE user equipment

gNB gNodeB

3GPP third generation partnership project

MAC media access control

RTT round-trip time

ACK acknowledgement

TB transport block

TBS total block size

SNR signal-to-noise ratio

NACK negative acknowledgement

ISL inter-satellite links

FEC forward error correction

RV redundancy version

SR scheduling request

BSR buffer status report

PDCP packet data convergence protocol

SDU service data unit

LOS line of sight

vSAT very small aperture terminal

MTU maximum transmission unit

PDR packet delivery ratio

CQI channel quality indicators

OFDM orthogonal frequency division multiplexing

1 Introduction

1.1 General introduction

The third generation partnership project (3GPP)[1], the standardization body for the development of protocols for mobile communication networks, recently put a great emphasis on the importance of the integration of different access technologies along with the existing terrestrial mobile telecommunication infrastructure [1]. This choice was made to accomodate for the ever-increasing number of Internet-connected devices and the need for wide-area coverage as well as ensuring service availability, continuity and scalability [2].

Specifically, the envisioned future for mobile communications, starting with the already established 5G new radio (NR) and further expansions with the new sixth-generation cellular networks (6G), foresees the integration of a new non-terrestrial component. The latest 3GPP releases (Rel. 17 and Rel. 18) require 5G and 6G networks to be able to provide non-terrestrial satellite access complementing the already existing terrestrial access technologies. This because NTNs have the right characteristics to satisfy the needs previously expressed, overcoming the limitations of the current terrestrial infrastructure, as detailed in the following sections [3], [4].

To give a proper definition, the term NTN refers to a category of networks where at least one link is routed via an aerial-borne or space-borne vehicle

[1] `3gpp.org`

such as high-altitude platforms (HAPs), unmanned aerial vehicles (UAVs), or telecommunication satellites.

1.2 Limitations of terrestrial networks

In order to explain the motivations behind the choice of expanding the current terrestrial infrastructure, the following sections present a few scenarios where the current terrestrial infrastructure fails to provide an adequate service due to some of their intrinsic limitations, and NTNs can help to provide a better coverage.

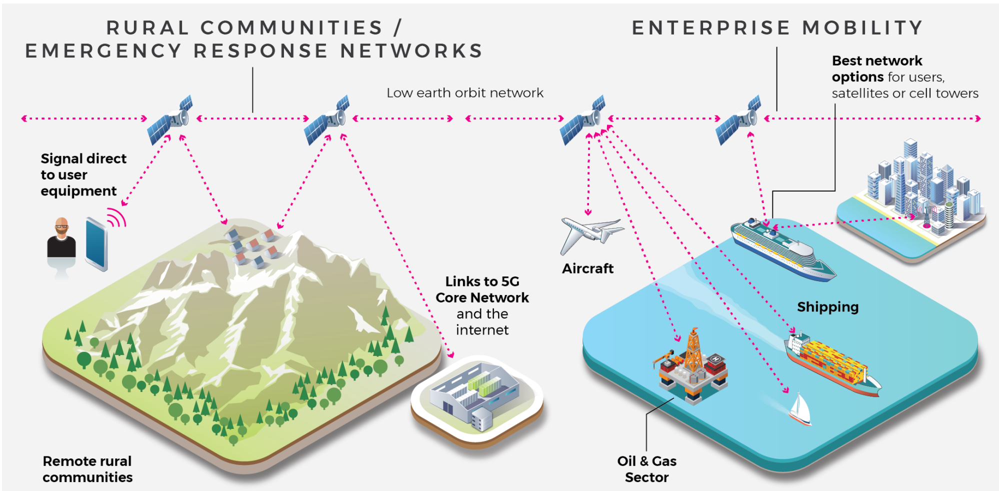

Figure 1.1: NTN use cases, courtesy of telecominfraproject.com

1.2.1 Remote places

While terrestrial networks (TNs) make the well-established foundation of today's mobile communication infrastructure, their own nature poses some intrinsic limitations to their deployment in certain scenarios, especially in contexts of rural and remote areas. Conditions such as harsh terrain and geographical impediments like mountain ranges and large bodies of water act as natural barriers to the deployment of terrestrial infrastructure [5]. Moreover, the need for ground equipments such as base stations requires the presence of an already established and reliable power grid, further driving up the costs that telecommunication companies would have to sustain to bring connectivity in isolated areas without such amenities.

Population density The population density in remote and rural places is typically much lower than in the urban areas, and users are distributed in vast areas of land. Because of this reason, the investment that has to be made on a per-user basis in order to provide coverage would be much higher compared to a more urbanized scenario, where potential users are more densely distributed. The already high infrastructure cost will therefore have a hard time generating any profit, making this kind of market even more unattractive to private investors and further limiting the possibility for the people living there to access the internet, a resource which is becoming increasingly more important as time goes by [6].

In those scenarios, the large coverage area of NTNs can help providing a cheaper alternative to traditional access solutions that no longer requires the ubiquitous presence of costly ground infrastructure. In this way, telecommunication companies are able to serve many users with the same satellite even if those users are loceted in isolated places, contributing to depreciate the necessary investment to reach each one of them.

Opportunities As studied and documented in [7], the issue of an inadequate broadband coverage in rural regions is an enormous challenge. Providing connectivity to the half of the world population living in rural or underprivileged areas requires a colossal effort, but it would also be a unique opportunity to kick-start the economy of currently underdeveloped countries. Access to the Internet would provide the population a possibility to progress on the educational, environmental, business and health planes, promoting a more fair access to information and alleviating the problem of digital divide between different parts of the world.

1.2.2 Emergency communication and additional capacity

Another limitation of the current terrestrial infrastructure is the lack of redundancy and robustness against natural disasters. Extreme events such as earthquakes, hurricanes, fires and floods, but also deliberate behaviors such as targeted attacks by terrorist organizations and sabotages can disrupt the connectivity, leading to outages that can potentially last for a long period of time. This in turn can cause significant economical damage and, in emergency situations, hinder the already difficult rescue efforts, potentially even leading to loss of lives [5].

The simple installation of additional base stations is not a viable solution because they would all share the same weaknesses, and extreme weather events such as tornados would easily be able to render all the additional equipment useless. The cost of essentially doubling the existing access network to make it more resilient would be enormous.

In this scenario, NTNs can act as a redundant access methodology to decrease the downtimes of terrestrial infrastructure, providing emergency communication services as well as additional capacity when required.

Fig. 1.2, courtesy of [8], shows a scenario where terrestrial and non-terrestrial access technologies are used transparently to access the network. It depicts a user equipment (UE) that can use both the non-terrestrial link provided by the satellite or the terrestrial one. In this case, the satellite is connected to a ground gateway where the gNodeB (gNB) is located, therefore only rerouting the signal coming from the UE.

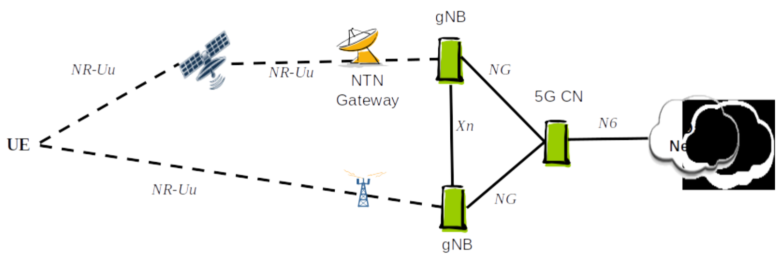

Figure 1.2: Transparent use of multiple radio access technology [8]

1.2.3 Long distances and sensors

Remote equipment, offshore plants and distribution grids will also benefit from the research carried out in this field, since providing terrestrial connectivity in those scenarios is a very challenging task due to the physically long distances between end nodes and the data collectors (generally a gateway or a base station). The installation of an underwater optical fiber link to serve a single endpoint, such as an offshore oil rig or power plant located in the open ocean, would bear a disproportional cost compared to the functions required, and providing maintenance would be another challenging and expensive task. The deployment of NTNs would be the easiest way to provide connectivity on a global scale, therefore allowing internet access in isolated places without the need for an expensive, dedicated connection.

Consider now the problem of connecting a number of sensors sparsely distributed in a large area. When distances are large, solutions may be either the densification of radio stations or the use of custom made hardware. However, those approaches have their downsides and are not always feasible, requiring costly, ad-hoc solutions tailored for each specific scenario. In this case, the vast coverage area of NTNs will undoubtedly be useful to provide connection to and from the sensors [9].

Other scenarios where NTNs can become useful in overcoming the limitation of their terrestrial counterpart are well described in [5] and [10].

1.3 Thesis objectives and structure

The main objective of this thesis is to use and extend a full-stack end-to-end network simulator to: (1) evaluate the problems of the current 5G NR protocol stack for an NTN satellite scenario characterized by high propagation delays, and (2) design, implement and evaluate original protocol solutions for NTN, focusing on the role of retransmissions via HARQ and scheduling. We aim at adapting the current implementation of the ns-3 network simulator software to be able to work in a non-terrestrial scenario with a high propagation delay, as well as adapting the layer-2 protocols to work under such conditions. Then, a performance evaluation will be carried out to study the effects of NTN on the protocols' behavior, identifying any erroneous behavior and proposing innovative solutions tailored to the characteristics of NTNs.

Despite 3GPP interest in this evolving field, the literature on this topic is still scarce. The effects of high propagation delays on the NR protocol stack have not been properly simulated and documented, since the tools for network simulations have little to no support for 3GPP-like NTNs.

This thesis will approach the problem by implementing a non-terrestrial communication scenario in the ns-3 simulator, creating a testing framework that is currently not available as part of the original codebase of the software, modifying and rewriting the code for the protocols that misbehave, as well as optimizing their behaviors via the analysis of numerous simulations. This practical approach will provide a continuous feedback, enforcing a bottom-up attitude, meaning that the first layers of the protocol stack that will experience a rework will be the lowest ones, gradually moving up the ladder until we have a full working stack.

At the end of this thesis, important modifications were made to the NR protocol stack implementation to be able to work in non-terrestrial scenarios. Numerous problems linked to the scheduler were solved, i.e., makig it correctly account for propagation delays by scheduling resources with a longer advance compared to the terrestrial implementation, reworking the BSR mechanism of the UE to improve the efficiency of the communication, and changing the reordering timers to avoid discarding correctly received packets.

HARQ protocol was reworked as well by evaluating the performance of increasing the number of concurrent processes per-user, as well as proposing a more aggressive version and evaluating its behavior. Since 3GPP proposed to completely disable the protocol, simulations were conducted comparing scenarios with and without HARQ, and it was argued that disabling the protocol can be beneficial only in high SNR scenarios, while low SNR ones can still benefit from its usage to improve reliability.

After the above introduction, Chapter 2 consists of a deep dive into the field of non-terrestrial networks, resulting from a comprehensive review of the current state of the art. The three main categories of telecommunication satellites are described, and their characteristics evaluated. Finally, an overview of the possible payload design choices is presented, discussing advantages and disadvantages, as well as a brief description of the potential of multilayered networks.

Chapter 3 is focused on the state of the art of network simulation softwares and ns-3, the employed one, highlighting the non-terrestrial channel model used as well as its implementation, and detailing the reasons behind its choice.

After the initial presentation and characterization of the used tools, Chapter 4 is about the design and implementation of an extension to the network simulator ns-3 for simulating the NR stack in NTNs. We present the simulated scenario, describing the characteristics and the parameters of all the involved devices. The problems encountered whilst reworking the NR stack are presented, many of them involving the scheduler, and the implemented solutions are carefully explained.

In Chapter 4, the scheduler was modified to schedule resources more in advance to account for the additional delay introduced by the satellites, and to expect a non-immediate UE response, a behavior that was not considered in the standard ns-3 implementation for terrestrial-like scenarios.

A problem regarding the expiration of BSRs also arose, leading to wasted resources since the UE essentially flooded the base station with unnecessary requests. A modification was implemented to allow a UE to wait for at least a full round-trip time before trying to send a new request.

The standard implementation of the UE MAC protocol is configured in such a way that each incoming packet triggers a scheduling request for the full transmission buffer, regardless of previous requests that may still be in transit, therefore wasting capacity. The implementation of a less aggressive mechanism managed to increase the efficiency.

Then, moving up in the ladder of protocol layers, the problems related to the operation of HARQ in a non-terrestrial scenario are presented in Chapter 5, discussing the effects of propagation delay as well as the impact of different values of SNR, linked to the satellite orbiting altitude and UE antenna. Some new HARQ solutions for NTN are proposed and evaluated using the ns-3 network simulator.

The maximum number of concurrent HARQ processes that NR implementation enforces is found to be a major limiting factor for the throughput, therefore its increase has been evaluated, and, while this managed to increase the achievable throughput, its effects were limited when compared with the additional complexity that such approach would require.

Simulations performed without HARQ showed that better results could be obtained, but this solution only works in conditions with high SNR, therefore a more aggressive version of HARQ is proposed, and implemented and evaluated in ns-3, aimed at lowering the impact of the limited amount of concurrent processes.

Finally, the conclusions of this thesis are presented in Chapter 6, highlighting how scheduler and HARQ are affected by the characteristics of a non-terrestrial link, summarizing all the obtained results as well as the proposed solutions, and providing some paths that can be explored in future studies.

2

Non-terrestrial networks

This chapter describes the main charateristics of NTNs, i.e., networks where at least one link endpoint is either an aerial or space platform. Specifically, the remainder of this chapter highlights the differences between types of satellites, their advantages and disadvantages as well as the possible choices for payload types. This thesis focuses on the implementation of 5G NR in a non-terrestrial scenario, therefore NR terminology is used.

2.1 Satellite types

Satellites are divided in three main different categories, depending on their orbiting altitude: geosynchronous equatorial orbit (GEO), medium earth orbit (MEO), and low earth orbit (LEO) satellites. Each category exhibits its own characteristics, presenting some upsides and downsides as described in the remainder of this section. Fig. 2.1 illustrates the different orbiting altitudes as well as estimates of the corresponding coverage areas.

2.1.1 GEO satellites

Orbiting at an altitude of 35.786Km, to an observer placed on the Earth surface GEO satellites appear stationary, since their orbiting period is the same as the Earth rotational period.

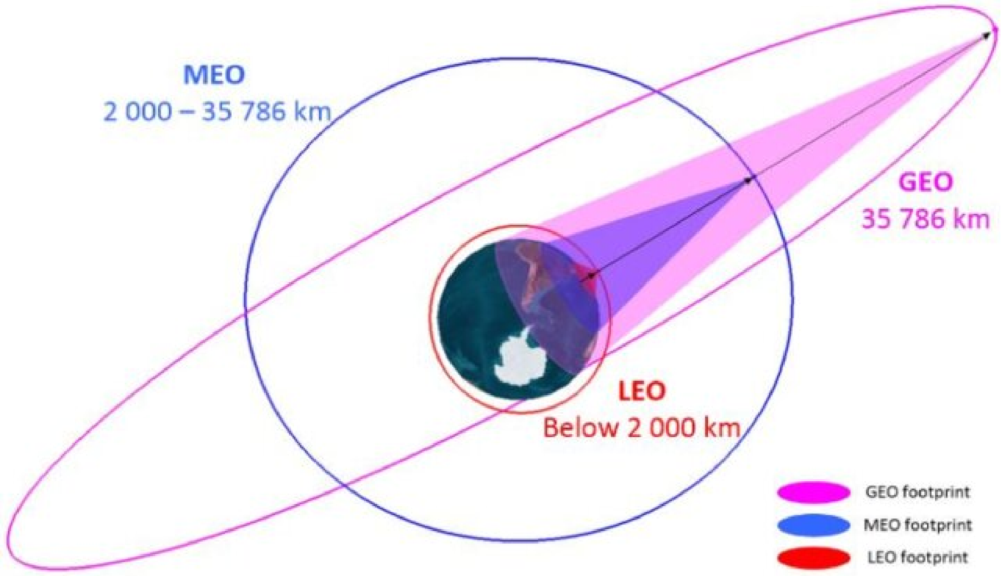

Figure 2.1: Altitude and approximate coverage areas for satellites at different orbits [11]

Advantages Since GEO satellites are geostationary, continuous coverage to a designated area can be provided using as little as a single satellite, while the use of non-GEO satellites requires, in general, the deployment of a constellation which is both more complex and more expensive.

Geostationary satellites also vastly simplify the problem of the ground equipment having to be able to track the satellite. Since the satellite position is always known, once the position of the UE is established, the relative position of the satellite can be easily estimated.

As shown in Fig. 2.1, the high altitude of GEO satellites creates a large cell footprint. Overall, while the deployment cost of a single GEO satellite is higher than both MEO and LEO ones, the cost per coverage area is lower, and an almost full coverage of the terrestrial globe can be achieved using as little as three equally spaced satellites [12].

Disadvantages The disadvantages of GEO satellites are mainly caused by their large distance with respect to UEs placed on the Earth surface: the transmission power and the antenna gain have to be high enough to overcome the greater propagation losses. Moreover, the propagation delay of the signal increases the baseline latency by approximately 120ms. This means that if the UE sends a request to a server at $t = 0$ through a GEO link, the packet will be received by the destination node at least $t = 240$ms. The response will then finally reach the UE after at least 480ms since the initial transmission. These calculations do not factor in any delay related to medium access requests, packet transmission times and processing delays, which would further increase the overall latency.

In addition to the positive aspects previously discussed, the large cell footprint also brings some downsides with it. Due to the vast coverage area, a single satellite will be required to serve a massive number of users. As a consequence, the total available capacity will have to be shared between a bigger number of equipments, and the throughput experienced by each of them will be reduced. Preliminary solutions to overcome this issue comprise, for instance, the use of beamforming to divide the covered area in smaller cells, and the employment of higher frequency bands towards Ku, K and Ka as depicted in Fig. 2.2 [13]. Moreover, the high number of users also leads to a large rate of initial access requests, with the possibility of channel saturation as described in [8].

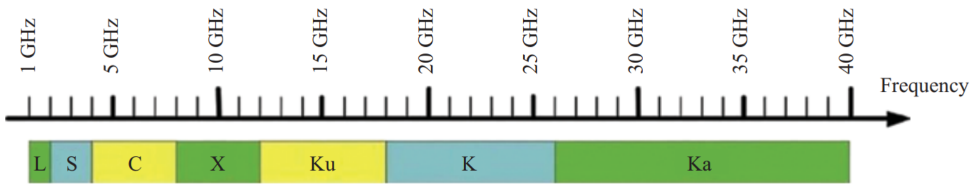

Figure 2.2: Satellite spectrum bands allocation [13]

2.1.2 MEO satellites

MEO orbits sit between the LEO and GEO counterparts. Accordingly, all satellites orbiting between 2000 Km and 35.786 Km are considered as MEO. This vast orbital space is mainly used by navigation systems such as GALILEO [12].

While the propagation delay can vary a lot depending on the altitude, in general it is larger than that of LEO and smaller than that of GEO. The same point can be made regarding the cell size and number of served users.

Similarly to LEO satellites, MEO ones do require a constellation in order to provide continuous coverage over a designated area, since they are not geostationary. Nevertheless, the number of satellites needed for providing continuous coverage is smaller than that required when using LEO satellites, since the former can serve a larger area.

Their many downsides in addition to the lack of any substantial benefit over their competitors besides for the need for a smaller constellation makes them less than ideal candidates for applications in NTNs.

2.1.3 LEO satellites

Orbiting below the threshold altitude of 2.000Km, LEO satellites represent the most promising solution in the realm of NTNs, mainly because they can offer really compelling throughput and propagation delay. However, the latter comes at the cost of numerous disadvantages which are peculiar to satellites deployed at this altitude.

Advantages The low altitude entails a shorter propagation delay, between 2 and 6 ms. Furthermore, the smaller coverage area of each satellite with respect to MEO and GEO means that the total number of users that need to be served is smaller. This also enables the use of high frequency bands and the constraints of high antenna gains are less stringent compared to GEO satellites, since the experienced path loss is much smaller. In turn, this leads to a higher throughput, more suited to satisfy the requirements of modern days broadband connectivity, as detailed in [14].

The cost per deployed satellite is significantly smaller than that of GEO and MEO satellites, and multiple deployments within a single launch are possible, further reducing the deployment costs.

Disadvantages Given that LEO satellites are not geostationary, a large constellation is needed to provide continuous service, driving the deployment costs significantly up. As an example of how vast those constellations can become, Fig. 2.3 depicts the LEO satellites employed by Starlink[1], counting about 4.808 units in service at the time of writing[2].

Since low orbiting satellites remain view of the UEs only for a short period of time, with an average in-view duration of just 13 minutes as calculated in [15], all the connected users are expected to be handed over to the next available satellite

[1]Starlink is a satellite internet constellation operated by Starlink Services, LLC, a wholly-owned subsidiary of American aerospace company SpaceX.

[2]Source: `satellitemap.space`

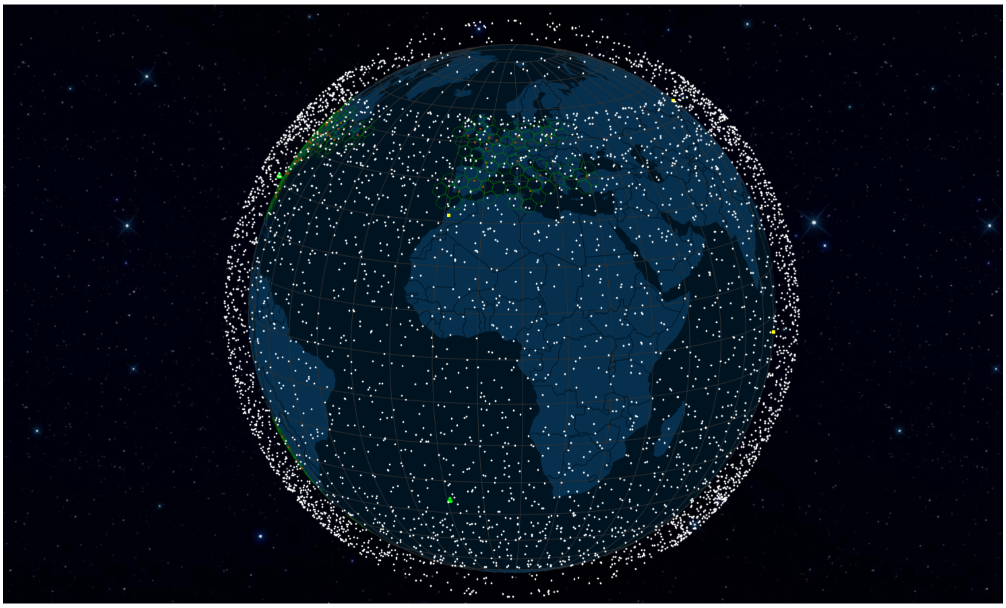

Figure 2.3: Starlink constellation as of July 2024. Source: `satellitemap.space`

within this time window. Such behavior would create a noticeable protocol overhead, consuming available channel capacity and potentially adding more latency. However, the predictable nature of this phenomenon might allow for a partial automation without requiring explicit control messages to be exchanged.

The small coverage area means that more terrestrial gateways have to be deployed, since each satellite can only communicate with the ground via the terrestrial gateways that fall within its view. A different solution to the densification of gateways is the use of inter-satellite links (ISL), i.e., high-bandwidth links between different satellites of the constellation. The latter are used to connect satellites that do not have gateways in sight to ones that are connected to a gateway, allowing traffic to be routed to the ground via additional hops. Inter-satellite links have also to be implemented if coverage over the oceans is required. Ultimately, this further increases the high constellation deployment costs.

An additional downside affecting all the non-geostationary satellites involves their speed relative to the user equipment located on the ground. A LEO satellite moves with a speed of approximately 7.8 km/s [16], thus exhibiting a non-negligible Doppler shift. This has to be compensated for, and preliminary solutions in this sense can only be applied to UEs which feature GNSS capabilities, which is not always a reasonable assumption [14], [17]. Moreover, the dependency of the network on a third-party service in order to operate properly adds a point of failure which is outside the control of the network operators. Indeed, a disruption in GPS service would halt the network capabilities.

The aforementioned advantages make LEO satellites the most promising choice in the field of NTNs.

2.1.4 Multilayered networks

The aforementioned orbits are not to be considered mutually exclusive. In fact, the multitude of possibilities that their combinations offer opens to the possibility of various different scenario in which the space, air, and ground layers are orchestrated to improve quality of service. An example of such multi-layer network is depicted in Fig. 2.4, showcasing a highly sophisticated non-terrestrial multilayered network scenario where different access technologies are used.

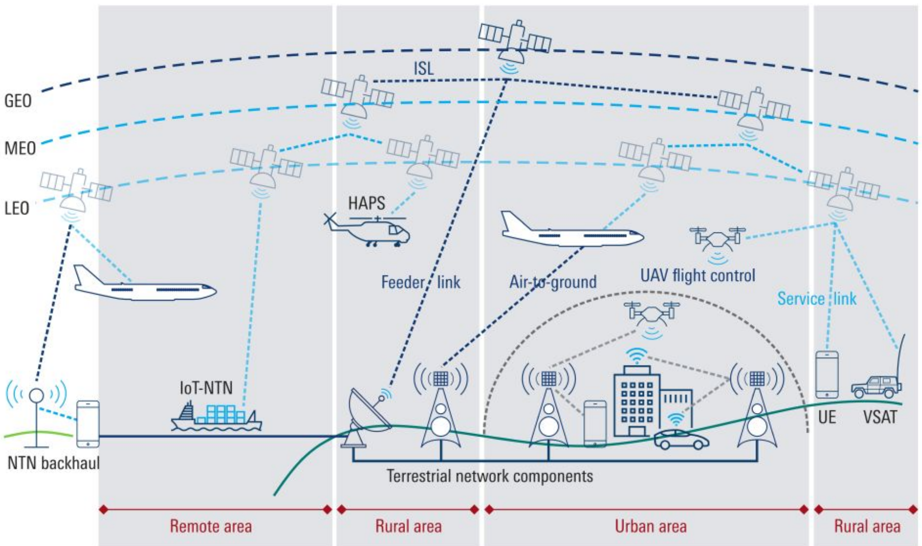

Figure 2.4: Complex multilayered NTN scenario [18]

For instance, [10] argues that that the use of HAPs as relays between the ground segment of the network and the upper GEO satellite links can deliver up to six times the capacity, and better overall outage probability, than point-to-point GEO transmissions.

2.2 Types of payloads

When implementing a NTN, an important choice to be made is the type of payload to use. There are two main categories: bent-pipe payload and on-board gNB. Each one has its own benefits as briefly described below.

2.2.1 Bent-pipe payload

This is the simplest approach, where the role of the satellite consists only of repeating the signal received from the UE on the ground towards the terrestrial gateway. Configurations such as the one depicted in Fig. 2.5 go by the name of bent pipe payloads and are characterized by the presence of a terrestrial g-NodeB, while the satellite has the sole purpose of providing a transparent link to the UE.

While such solutions are by far the simplest in terms of payload complexity, the main drawback is the even longer experienced latency. Since communications between users served by the same satellite would also have to be routed through the terrestrial gateway, the overall latency increases by at least two times the propagation delay. This solution also poses strict bandwidth requirements on the feeder link, since all the traffic must necessarily pass through it. More complex solutions are able to route at least part of the inbound traffic autonomously, without routing everything back to earth.

Figure 2.5: 5G NR NTN architecture for access network based on satellites with bent pipe payload [8]

2.2.2 On-board g-NodeB

A slightly more sophisticated approach foresees the installation of g-NodeB capabilities directly onto the satellite payload, as depicted in Figure 2.6. This has the benefit of reducing the experienced latency, as well as reducing the utilization of the feeder link. Certain protocols, designed to terminate at the gNB, can in this case reach their designated endpoint without necessitating to be routed back to the ground gateway.

Figure 2.6: 5G NR NTN architecture for access network based on satellites with on-board gNB payload [8]

2.3 Commercial solutions

Legacy solutions Commercial solutions, initially concerning satellite-based phone calls and successively evolved to provide Internet access, have been available dating back to 2003, when the first Internet satellite was launched [19]. However, the majority of this legacy infrastructure makes use of GEO satellites, therefore presenting all the limitations discussed in section 2.1, i.e., they offer limited throughput and large delays. Such constraints render this technology not suitable for the needs of modern internet connections standards.

Recent developments Commercial solutions which make use of LEO satellites are starting to appear, and they are experiencing moderate commerical success. Notable examples are Starlink and Eutelsat OneWeb. However, these solutions make use of proprietary protocols. Until now, no internationally agreed standard has been defined. This is where most of the research is currently being conducted and where the business world is posing its attention due to the characteristics of LEO satellites that makes them the most suited to provide high speed satellite internet access.

3

E2E simulation of 5G networks

As part of its contributions, this thesis introduces a preliminary end-to-end simulator for 5G NR NTN deployments. This simulator builds upon the ns3-mmwave simulation module, which accurately models the 5G NR protocol stack, and has been recently extended with the 3GPP channel model for NTNs [20]. This chapter describes this baseline, which has been used as the starting point for designing and developing an end-to-end simulator for 5G NR NTN deployments.

3.1 Discrete events network simulators

The network simulators scenario consists of two main categories: discrete events and real time simulators. The former option is the most performing one, since it allows for the processing of complex events even with modest hardware. Real time simulators require high-end hardware, otherwise the processing of complex events can slow down the whole simulation.

In discrete events simulators, each operation to be performed is associated to an event, and in turn, each event is associated with a set of instructions and its execution time. The simulation proceeds by processing and executing events, stepping from one to the next, as the simulation time passes. From the point of view of the simulation, each event is executed in zero time, since the time is stopped while executing a single event, and its course resumes only when transitioning between events scheduled at different times. If no events are

Tool	ns-3	OMNET++	SWANS	NetSim	QualNet
Interface	C++, Python	C++, NED	Java	C, Java, .NET	Parsec
License	Free	Academic	Free	Paid	Paid
Parallelism	No	No	Yes	No	Yes
OS	Linux, FreeBSD, MacOS Windows	Linux, MacOS, Windows	Linux, MacOS, Windows	Windows	Linux, MacOS, Windows, Unix
Mobility support	Yes	No	Yes	Yes	Yes
GUI	Limited	Yes	Yes	Yes	Yes

Table 3.1: Network simulation software comparison

scheduled to execute for a certain period of time, the simulation immediately transitions to the next scheduled one. This kind of behavior is what discerns discrete events simulators from their counterparts, real-time simulators. As the simulation unfolds, events are sequentially executed, each of the former possibly scheduling new additional events. As an example, the event of a packet being transmitted in a network may generate the corresponding reception event after a set propagation delay [21].

The most popular simulation softwares are OMNeT++ [22], SWANS [23], NetSim by TetCos [24], QualNet, ns-3, and its predecessor ns-2 [25]. Their main characteristics are summarized in Table 3.1, from [21], which posed the accent on the suitability of ns-3 for research purposes, highlighting its success amongst the scientific community.

3.2 NS-3

Figure 3.1: ns-3 logo nsnam.org

Being a modular, extensible, community-supported, full-stack network software simulation tool based on a discrete event approach, the ns-3 simulator was the software of choice to conduct the testing campaigns. It is an open source project licensed under the GNU GPLv2 license, thus allowing users to freely re-use the software and to introduce modifications to the codebase [26].

The community Being specifically targeted for the academic world, and for research, as well as being open-source, ns-3 sees a thriving community of developers and researchers, with an active forum[1], a well maintained documentation and a lot of independent lectures, tutorials and articles.

Modularity Ns-3 exhibits a modular philosophy. The base version provides the core of the simulator with a plethora of functionality, but lots of expansion modules have been coded by researchers and developers. The following sections present the ns3-ntn and ns3-mmwave modules, that were used in this thesis in order to provide an initial support to base our work on.

3.3 Preliminary support for NTN NR simulations

The ns3-mmwave module provides support for the simulation of 5G NR stack, while [20] provides a standard implementation of a non-terrestrial channel model, but the support for NR stack in NTN is still missing. These two modules are presented in this section, then further extended to obtain a ntn module capable of simulating the NR stack in NTN.

3.3.1 Channel model

The 3GPP channel modeling framework, described in TR 38.901 [27], which supports frequencies up to 100 GHz, statistically simulates both large and small scale fading parameters. Four different reference scenarios are specified: rural macro, urban macro, urban micro, and indoor hotspot. Furthermore, each scenario is subdivided in line of sight (LOS) or NLOS condition.

The remainder of this chapter briefly describes the 3GPP TR 38.901 channel model, with a focus on its extensions for modeling NTN scenarios. For a comprehensive description of the baseline model, we instead refer the interested reader to [28].

[1]Link to google group about ns-3 groups.google.com/g/ns-3-users

Free space path loss

The free space path loss is the major attenuation component in non-terrestrial links due to their length, and can be calculated as the ratio between the received power and the transmitted power with the well-known Friis formula

$$\frac{P_r}{P_t} = D_t D_r \left(\frac{\lambda}{4\pi d} \right)^2 \tag{3.1}$$

Where D_t and D_r denote the directivities of the transmitting and receiving antennas, λ is the wavelength and d the distance.

Atmospheric attenuation

In addition to the free-space path loss that characterizes the majority of wireless communication systems, atmospheric absorption also plays an important role in attenuating certain frequency bands of the signal, as described in Figure 3.2. The peaks at 60 and 120 GHz are due to the resonance with molecular oxygen, while the peak at 180 GHz and the small hump at 25 GHz are due to the absorption from water vapor [29].

The nature of this phenomenon makes it susceptible to variations as the humidity rate varies, and different altitudes also lead to different absorption values.

Shadowing

Shadowing is a variation in the signal amplitude which stems from the latter reflection and scattering by surrounding objects, therefore arriving at the receiving antenna from many different paths. This causes multiple copies of the signal to be received, each copy having its own attenuation and phase, since the propagation paths and their distance are, in general, different. This effect can cause rapid fluctuations in the signal amplitude, due to either constructive or destructive interference.

Additional factors

Other factors causing additional attenuation are the presence of rain, cloudy conditions, the presence of fog, and different meteorological parameters. These factors are described in [30].

Furthermore, the different scenarios that the UE can experience also have a major impact on its communication capabilities. Such scenarios are divided by 3GPP in four main categories: Dense urban, Urban, Suburban and Rural. In each scenario, the probability of having direct LOS with a satellite differs, since the density of obstacles such as buildings varies depending on the situation.

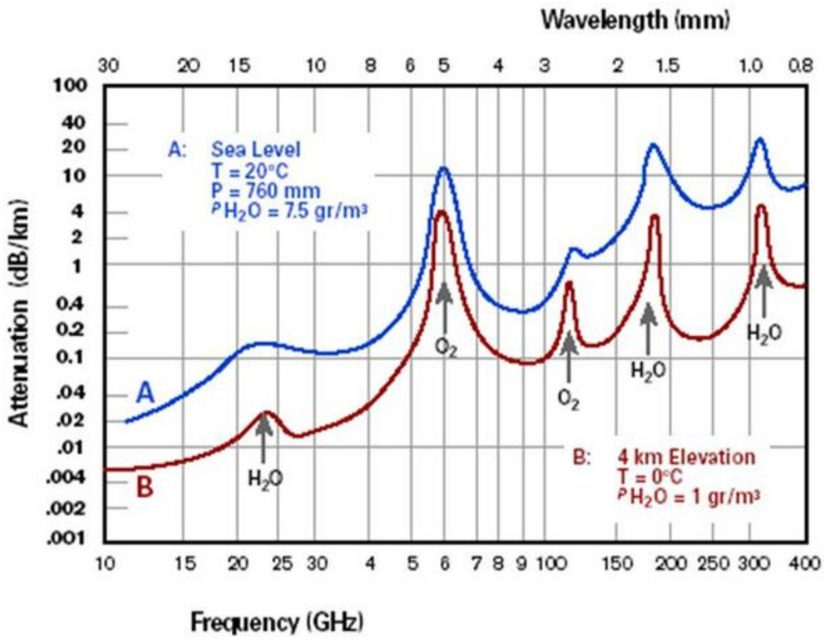

Figure 3.2: Atmospheric absorption in dB/km, from [29]

3.3.2 NS-3 channel model implementation

A ns-3 module to simulate a non-terrestrial channel model has already been implemented, and is available open-source[2]. This model is based on the 3GPP specifications as detailed in the standard TR 38.811 [8], paving the way for simulating wireless channels in space.

The implementation of such channel model required the modification and the creation of some ns-3 classes. This work is extensively described in [20], while a brief overview is hereby reported.

[2]gitlab.com/mattiasandri

Modified classes

- `ThreeGppChannelModel`: different parameters were introduced in order for this class to be able to correctly characterize the non-terrestrial use case. The large number of NTN-related parameters made necessary the use of data structures such as maps to store them, and the mobility model of both the satellite and the UE have been integrated in the computation of the small scale parameters returned by the class.

- `GeographicPositions`: class tasked with the various conversions between different coordinates systems such as longitude, latitude and altitude, the Geocentric Cartesian system (also called Earth Centric Earth Fixed or ECEF), and the local tangent plane coordinate system expressing the position in North, East and Up coordinates [31]. All those three systems are depicted in Fig. 3.3.

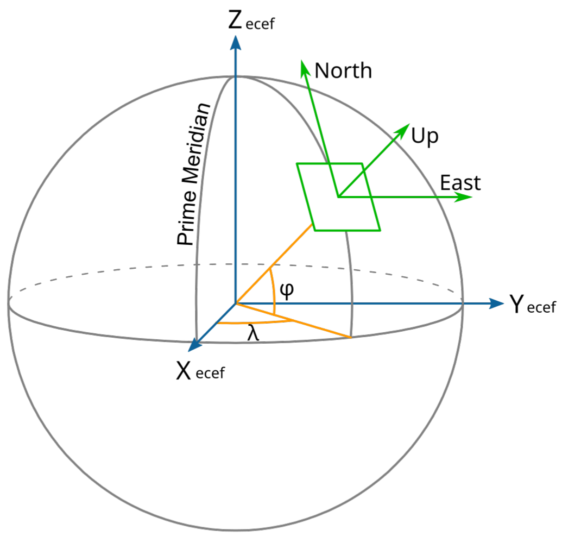

Figure 3.3: Showcase of different coordinates systems [31]

New classes

- `ThreeGppNTNScenarioChannelConditionModel`: the main task of this class is to store the channel state and condition. Four new classes were developed to store the four possible scenarios described by 3GPP:
  - Dense urban,

  - Urban,
  - Suburban,
  - Rural

- `ThreeGppNTNScenarioPropagationLossModel`: four different scenarios are implemented in as many classes. Such classes are tasked with the computation of the total path loss, which includes contributions from the standard free space path loss, atmospheric absorption, scintillation, fading and clutter loss.

- `GeocentricConstantPositionMobilityModel`: allows for the use of terrestrial coordinates to position UEs on the Earth surface. Conversions amongst different systems are done using the `GeographicPositions` class described above.

- `CircularApertureAntennaModel`: an exact implementation of the parabolic antenna model, making use of a new and efficient C++ function for the computation of Bessel functions required when considering the radiation pattern of aperture antennas [32]. Differently from the default ns-3 implementation, this newer version does not introduce approximations.

3.4 mmWave module

The Ns-3-mmWave module, created by the NYU Wireless Research Centre and the SIGNET Research Group at University of Padova, enables the end-to-end simulation of standard-compliant 5G NR cellular networks. 3GPP-specified statistical channel models are implemented, with modular and highly customisable classes [33].

After an initial study of the mmWave module, its implementation in a non-terrestrial scenario was devised, and numerous modifications were implemented to its classes in order to overcome all the problems that initially caused it to malfunction. These modifications are detailed in the following chapters, and represent an importnat original contribution of this thesis that was not previously treated by the literature.

In particular, we modified the classes of the radio stacks of both the base station and UE (`MmWaveEnbNetDevice` and `MmWaveUeNetDevice`), as well as the classes implementing the MAC layer and scheduler, namely `MmWaveEnbMac`, `MmWaveUeMac`, and `MmWaveMacScheduler`.

3.5 Use in this work

The ns3-mmWave module enables the simulation of the 5G NR protocol stack, while the ns3-ntn module [20] provides a 3GPP-like channel model for NTNs. However, prior to the work done in this thesis, the simulation of a joint scenario of NR stack in a non-terrestrial scenario was not possible, since the network stack was not designed to accomodate for the propagation delays exhibited by NTN links. The two modules were not able to operate in conjunction. Our work made non-terrestrial NR simulation possible.

4

E2E simulation of NR in NTN scenario

Given the absence of a tool to properly simulate the behavior of the NR stack in a non-terrestrial scenario, the first objective of this thesis is to design and implement some modifications to the protocol stack in order to make it work with long propagation delays. The current codebase of ns3-mmWave and ns3-ntn modules only provides separate implementations of the NR stack and the non-terrestrial channel model.

The objective of obtaining a base support that permitted NTN NR simulation was achieved by devising a NTN test scenario where we implemented the NR stack, launching simulations and examining the results. Each unexpected behavior was documented, analyzed and solved proposing original solutions. Such work was necessary since the current state of the art regarding network simulators is still lacking proper support for the scenario we meant to investigate.

After describing the network topology and scenario that we implemented in the simulator, this chapter focuses on the problems that hampered the ability to correctly send and receive packets. Since many of them involved the scheduler, a brief introduction about the scheduler and its working principles is presented, then each implemented solution is detailed.

4.1 Implemented scenario

This section aims at describing the reference scenario that was implemented in the ns-3 simulator in order to test the NR protocol suite in a non-terrestrial communication setting.

4.1.1 Network topology

To conduct the various simulation campaigns, ensuring the reproducibility and the comparability of the obtained results, a reference network topology has been made, and the parameters specified by 3GPP have been used.

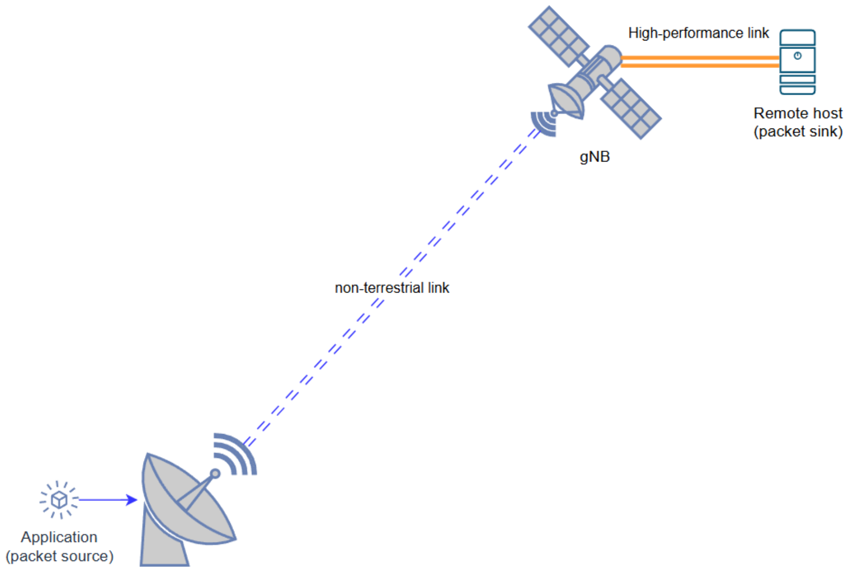

Figure 4.1: Network simulation scenario

The simple network setup is depicted in Fig. 4.1 and it consists of the following elements:

- **Packet source**: application installed on the user equipment that generates packets with a specified periodicity. Both the generation rate and the packets' size can be varied by acting on their respective parameters. All the other variables that can be controlled are listed in the table 4.1.
- **UE antenna**: the transmission of data is performed by mean of a very small aperture terminal (vSAT) antenna placed at the UE side. The main parameters of such antenna are found in the table 4.2.

- **Non-terrestrial link**: link connecting the UE placed on the ground with the gNB. This wireless link is characterized its propagation delay, bandwidth and frequency. However, since the aim of this work is to study the effects of propagation delay on the protocol suite, the bandwidth and frequency remained constant across all the simulations to better isolate the variables that could potentially cause problems. The list of parameters can be found in Table 4.4.

- **g-NodeB**: the adopted approach is to incorporate the g-NodeB into the satellite payload, therefore adopting the configuration described in section 2.2.2. This decision was made since the high one-way propagation delay is enough to cause some of the involved protocols to start malfunctioning. Adopting the bent-pipe configuration described in 2.2.1 would have resulted in effectively doubling the delay between UE and gNB. The satellite aperture antenna parameters follow the ones specified in the scenario named "10 DL" described in [17], and are listed in the table 4.3.

- **High performance link**: link connecting the g-NodeB to the packet sink. This is part of 5G core network, and it shall not be causing any additional problems, since that would be out of the scope of this work. This link was therefore meant to be as close as possible to an ideal one, with a capacity of 100Gb/s, a maximum transmission unit (MTU) of 1500B and a delay of a single microsecond.

- **Remote host**: packet sink representing the destination node of all the packets generated at the UE.

4.2 5G scheduler

As the name suggests, the main task of the scheduler is to allocate resources to the various connected users in the form of transmission and reception opportunities. In the NR standard, the scheduling is dictated by the network and the UE has to follow the provided indications [34].

In the NR standard, the scheduling process is similar to LTE, with the addition of some new factors and flexibility options.

Parameter name	Description	Default
enableNagle	whether to enable Nagle's algorithm	false
enableHarq	whether to enable HARQ protocol	false
numHarq	number of concurrent HARQ processes per UE	16
harqTimeout	timeout for HARQ processes	100 ms
rrcIdeal	use ideal RRC	false
tcpMinRto	minimum TCP RTO	200 ms
tcpBufSize	TCP buffer size	13107200 B
ipv4FrExpTimeout	IPv4 fragment expiration timeout	200 ms
perr	target error probability for PHY packets	0.1
transportPrtcl	whether to use UDP or TCP	UDP
numPackets	max number of packets to be sent	2000
appStartTimeSec	application start time	0.5 s
appStopTimeSec	application stop time	5.5 s
simStopTimeSec	simulation stop time	6 s
packetSizeBytes	application packets size	200 B

Table 4.1: Application and UE configuration parameters

Parameter name	Description	Default
vsatAntennaGain	gain of the vSat antenna	39.7 dB
vsatAntennaDiameter	diameter of the vSat antenna	0.6 dB
vsatAntennaNoiseFigure	noise figure of the vSat antenna	1.2 dB

Table 4.2: UE antenna parameters

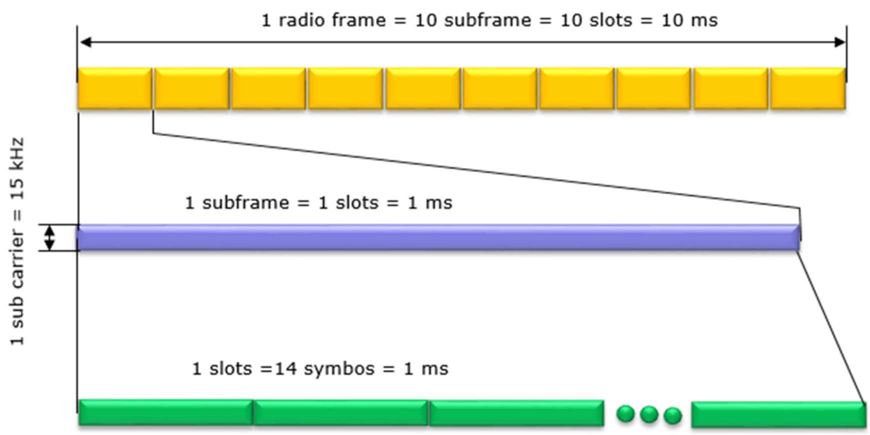

Figure 4.2: Radio frame structure for $\mu = 0$ [35]

4.2.1 Frame structure

The scheduler divides time using four different entities: frames, subframes, slots and symbols. 5G NR frame structure is defined in 3GPP TS 38.211 [36].

Parameter name	Description	Default
satEIRPDensity	EIRP density of the satellite antenna	40 dBW/MHz
satAntennaGain	gain of the satellite antenna	58.5 dB
satAntennaDiameter	diameter of the satellite antenna	5 m
distance	orbiting altitude of the satellite	-1 km

Table 4.3: Satellite antenna parameters

Parameter name	Description	Default
propDelay	propagation delay	6 ms
frequency	carrier frequency	20E9 Hz
bandwidth	link bandwidth	400E6 Hz

Table 4.4: Non-terrestrial link parameters

Frames always have a duration of 10 ms, while subframes are 1 ms long. Slots and symbols have variable lengths instead, based on the value of a parameter called numerology and represented with the letter μ. Regardless of its time duration, each slot always contain 14 orthogonal frequency division multiplexing (OFDM) symbols. [37].

The μ parameter can assume integer values in the range $[0, 4]$, and its value affects the subcarrier spacing and bandwidth, as well as the slot and the symbol durations. Table 4.5 contains the values for each supported value of μ [37].

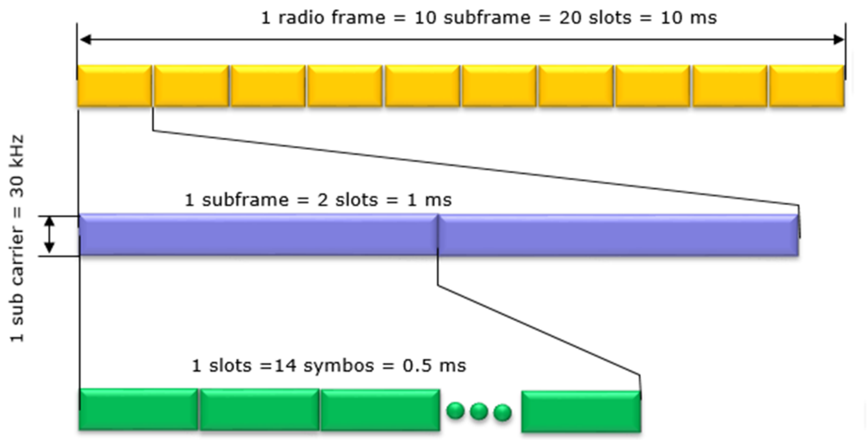

Figure 4.3: Radio frame structure for $\mu = 1$ [35]

Figures 4.2 and 4.3 show the difference in radio frame configuration for $\mu = 0$ and $\mu = 1$. In the first case, each subframe has exactly one slot, while in the second case each subframe contains two slots, the subcarrier bandwidth is

μ	**Subcarrier spacing** $\Delta f = 2^{\mu} \times 15$**kHz**	**Resource block bandwidth** $\Delta f \times 12$	**Slots per frame**	**Slot duration**	**Symbol duration**
0	15 kHz	180 kHz	10	1 ms	71.43 μs
1	30 kHz	360 kHz	20	500 μs	35.71 μs
2	60 kHz	720 kHz	40	250 μs	17.86 μs
3	120 kHz	1.44 MHz	80	125 μs	8.93 μs
4	240 kHz	2.88 MHz	160	62.5 μs	4.46 μs

Table 4.5: Supported numerologies

doubled and the duration of each symbol halved.

4.2.2 Dedicated configuration

Differently from LTE, where a predefined pattern was in place when allocating downlink and uplink slots in a radio frame, in NR this is done much more flexibly using a plethora of parameters such as the periodicity of UL and DL transmissions, the number of consecutive DL and UL slots and symbols at the beginning of each pattern and more [34].

In LTE, once a subframe is flagged as either UL or DL, all of its symbols are used for the allocated purpose, while NR enables symbol-level scheduling, meaning that we don't need to use every symbol within a slot, and each slot can be further divided in groups of symbols reserved for UL and DL. The term slot format indicates how each symbol within a slot is used [35].

4.2.3 Propagation delay and differential delay

The important key concept is that the scheduler shall account for the propagation delay. While in terrestrial communications a guard period of six symbols when switching from downlink communications to uplink communications is sufficient to account for the delay in a few km-radius cell, and timing advance commands can account for the different propagation delays experienced by users located in the center of the cell and users located in the cell edge [38], the same cannot be said for non-terrestrial use cases, where the distances that come into play are much longer, therefore delays are higher.

Consider the case depicted in Figure 4.4 from [39]. For a LEO satellite orbiting at 600Km, cell diameter of 82Km and elevation angle of 45°, the maximum differential round-trip delay of UE2 with respect to UE1 is of 370μs [39]. On the other hand, a 3km-radius cell experiences a maximum round-trip differential delay of $2 * \text{radius}/c \approx 10\mu s$

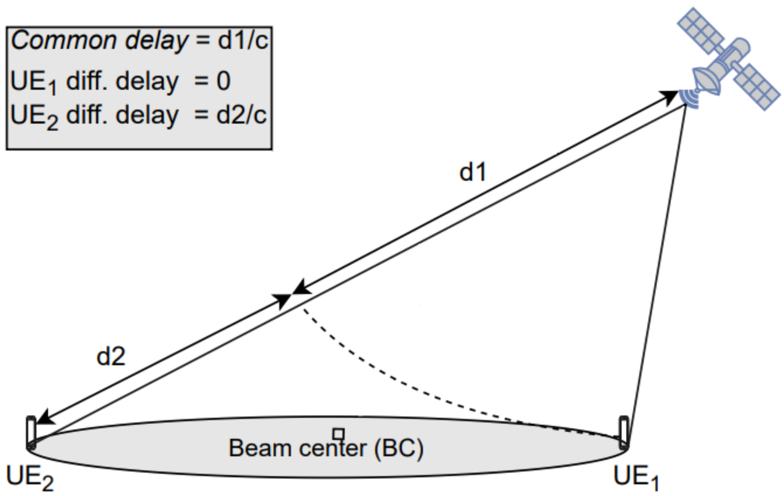

Figure 4.4: Differential delay for NTN, courtesy of [39].

4.3 Accounting for propagation delay in scheduling

4.3.1 Problem description

The first encountered problem while implementing a non-terrestrial communication scenario in the simulator was the inability of the scheduler to account for the propagation delay when allocating radio resources to the connected user equipment on the ground.

The implementation of the 5G scheduler in ns-3 is designed to allocate resources less than a single subframe in advance, and since each subframe has a duration time of 1ms, the resource grant was already expired by the time it was able to reach the UE, since it was referring to a past subframe.

Example Consider a scenario with a propagation delay τ_p of 6ms, corresponding to a satellite at an altitude of 1800 km. The UE sends a request for uplink resources at time t_0 since it has some data to send. In the standard scheduler implementation, assuming an ideal transmission, the gNB would receive such request at $t_0 + \tau_p = 6ms$. Provided that other transmissions have not been scheduled yet, the gNB grants the UE the possibility to transmit in the following subframe, which will start after 1ms at $t_0 + \tau_p + t_{SF} = 7ms$. The grant that allows the UE to transmit at $t = 7ms$ is ideally transmitted by the gNB as soon as it receives the request, therefore at $t = 6ms$. However, this grant will reach the UE only after another propagation delay, at $t_0 + 2\tau_p = 12ms$, when the transmission opportunity would already have expired.

4.3.2 Implemented solution

The implemented solution assumes that the scheduler has knowledge of an estimation of the propagation delay, that can be obtained through the use of probe signals and basing on previously measured values. This is a reasonable assumption since systems such as GPS already rely on a precise estimation of the delay between the user on the ground and the satellite.

The scheduling then proceeds as normal with the only difference being that the information regarding the propagation delay is used to postpone the allocated symbols.

Example Consider a scenario with propagation delay τ_p of 5 slots (roughly 1,2ms). The new implementation of the scheduler accounts for the propagation delay by allocating the first available slot after τ_p, so the time for the grant to reach the UE is accounted for, and the gNB marks the slot after $2\tau_p$ as reserved.

The last part of reserving a different slot is not as immediate. However, this mechanism needs to be in place because of the behavior depicted in Figure 4.5 and hereby detailed. In this scenario, the numerology μ is 2, therefore each subframe contains 4 slots.

- UE sends the scheduling request to gNB at frame 1, subframe 0, slot 0.
- The scheduling request (SR) reaches the gNB after $1\tau_p$ of 5 slots and the UE is scheduled to transmit at frame 1, subframe 2, slot 2 since there is some noticeable propagation delay.
- The SR reaches the UE at frame 1, subframe 2, slot 2, and the UE can transmit right away.

- The packet reaches the gNB after another τ_p, hence the base station needs to know that it cannot schedule other transmissions to take place in this slot, otherwise interference and collisions may arise.

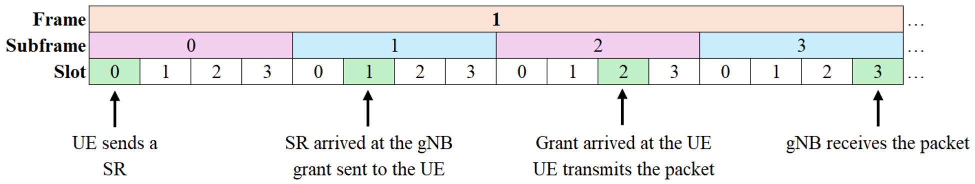

Figure 4.5: Difference between allocated slot and gNB reception

ns-3 implementation This solution has been implemented as a modification to the classes `MmWaveFlexTtiMacScheduler` and `MmWaveEnbPhy`, that now account for delayed responses of previously scheduled transmissions when allocating new slots.

The classes `MmWaveUePhy` and `MmWavePhy` have also been modified to allow access to the propagation delay parameter of the physical channel.

4.4 BSR timer

After implementing the solution discussed in the previous point in the ns-3 network simulator, allowing the communication to take place, some other irregularities were found regarding the periodic BSR timer, as described in the following.

4.4.1 Problem description

Reduced latency Figure 4.6 shows a rather peculiar trend regarding latency. The expected result was a linear increase in latency with a slope of 3, where every packet arrived at the destination after three times the propagation delays. The reasoning behind this expectation was that every packet generated by the application should have triggered a scheduling request received by the gNB after τ_p, a grant to be emitted requiring another τ_p to reach the UE, and finally the transmission to be received by the gNB after the third propagation delay (with the assumption of ideal transmission).

The observed pattern did not match any of the expectations. The output obtained from the simulation campaign showed a nonlinear saw tooth behavior presenting values that consistently stayed under the expected $3\tau_p$ threshold value, shown in green in Fig. 4.6.

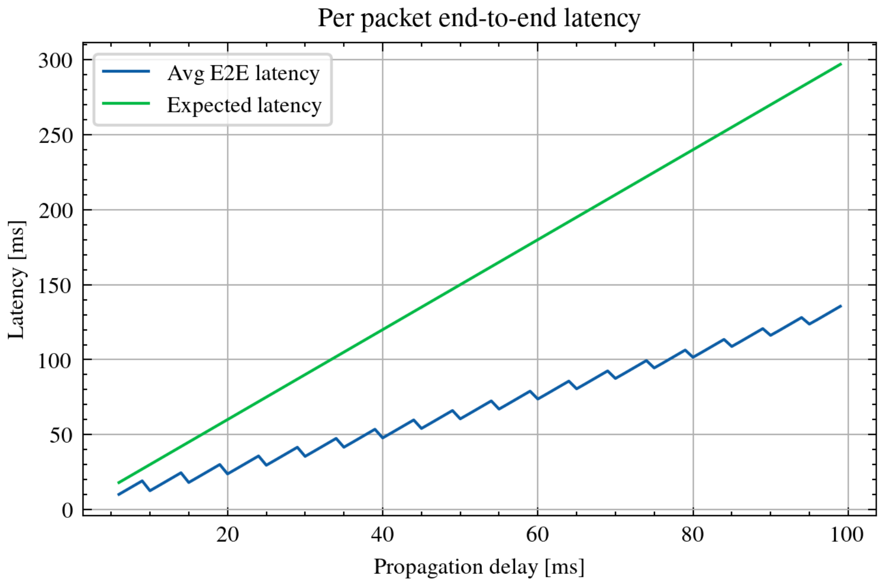

Figure 4.6: E2E latency vs. propagation delay with periodic BSR

Periodic buffer status reports Further investigation on the subject permitted the identification of numerous BSRs sent from the UE to the gNB at regular intervals of 10ms, even when no new packets were produced by the application.

BSRs are messages that the UE uses to communicate to the gNB the status of its transmission buffer. Every new packet arriving to the transmission buffer of the UE changes its status and a new BSR is sent, containing information such as the number of queued packets and their size. Upon reception, the gNB can grant the UE some resources to transmit part or all of the contents of its buffer. Each BSR is therefore interpreted as a SR.

This behavior happens because the implementation of 5G media access control (MAC) layer includes a periodic BSR that the UE sends to the gNB as long as the transmission buffer contains some pending data. The details of this mechanism are documented in Section 5.4.5 of the standard [40], where it is also stated that the default interval between those periodic BSR is of 10ms.

This characteristic was meant to act as a safeguard against lost scheduling requests, and does not account for the propagation delay at all. The UE therefore continues to send additional BSRs at regular 10ms intervals even though the previous ones were not lost, but still travelling towards the base station because of a (very long) in-fligh propagation delay.

Figure 4.7 illustrates this behavior. The arrival at the UE's application buffer of packet p1 at time $t = 0$ triggers the transmission of a first SR. Since the hereby depicted scenario has a propagation delay of 20ms, the SR will arrive at the g-NodeB at time 20ms. However, the BSR timer expires at $t = 10$ms by default, so during a single round-trip time the BSR is repeated four times. The green arrows marked with a capital G denote all the grants issued to the UE. Only one of the four depicted will actually be used by the UE for the intended packet, effectively wasting three out of four transmission opportunities that could have been allocated to other users waiting to transmit.

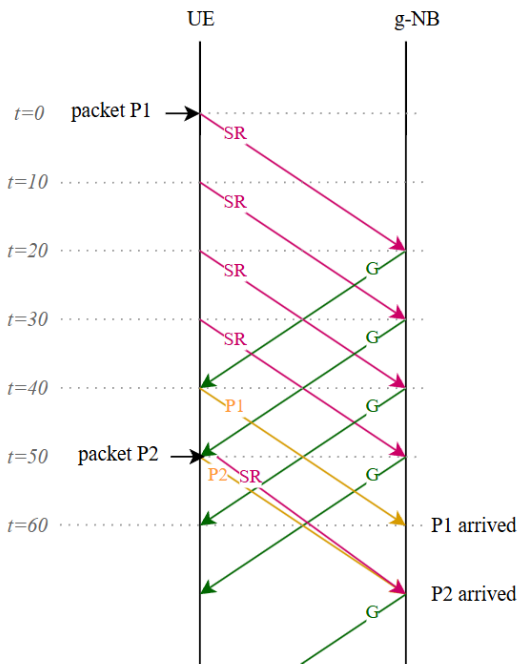

Figure 4.7: Packet diagram with periodic BSR

The reduction in latency was observed because newly arrived packets could, in certain situations, make use of the resource grants that were meant for the previous packets but have yet to arrive at the UE. Consider the arrival of another packet (P2) at theUE's application buffer. in figure 4.7. It triggers another SR, but a grant (originally intended for and trigger by P1) is received immediately after, therefore allowing for an immediate transmission of P2, which will experience a latency of only a single propagation delay. This instantaneous grant is nothing but the result of the previous SR storm caused by the UE itself.

While this can present some beneficial aspects such as the reduced overall latency, it originates from a behavior that is outside the original design intentions. Moreover, this approach is particularly greedy, leading to wasted capacity.

4.4.2 Implemented solution

After the successful identification and characterization of the problem, the solution consisted of the implementation of a simple adaptive algorithm that would dynamically set the BSR timer to a value of twice the propagation delay plus a constant of 4ms that accounts for the processing times as well as any possible delay. The constant was chosen accounting for the characteristics of the link, the packets transmission times as well as processing times.

ns-3 implementation Modifications were made to the `LteRlcUm` class to expose the BSR timer parameter using the ns-3 attribute framework, therefore permitting its dynamic adjustment based on the propagation delay.

This allowed the simulation manager software to test different values of the BSR timer, as well as a more standardized way of setting its value without changes to the source code. Figure 4.8 shows the result of our implementation. Latency now closely follows the expected results, only presenting a marginal increase due to the non-null transmission and processing times.

Area of potential improvement The implementation of the aforementioned solution, however, while working as intended, had the expected but unwelcomed effect of increasing the overall latency of the system, since all the additional grants required by previous packets that newly generated ones could exploit were no longer available.

This unexpected problematic behavior of wasted transmission opportunities accidentally also unveiled the potential of preemptively transmitting additional scheduling requests, so that future packets could be sent with a lower latency.

If implemented correctly, the UE could adopt a predictive approach towards its traffic needs for the near future basing its estimates on the type of application, and send the gNB scheduling requests for packets that have yet to be generated.

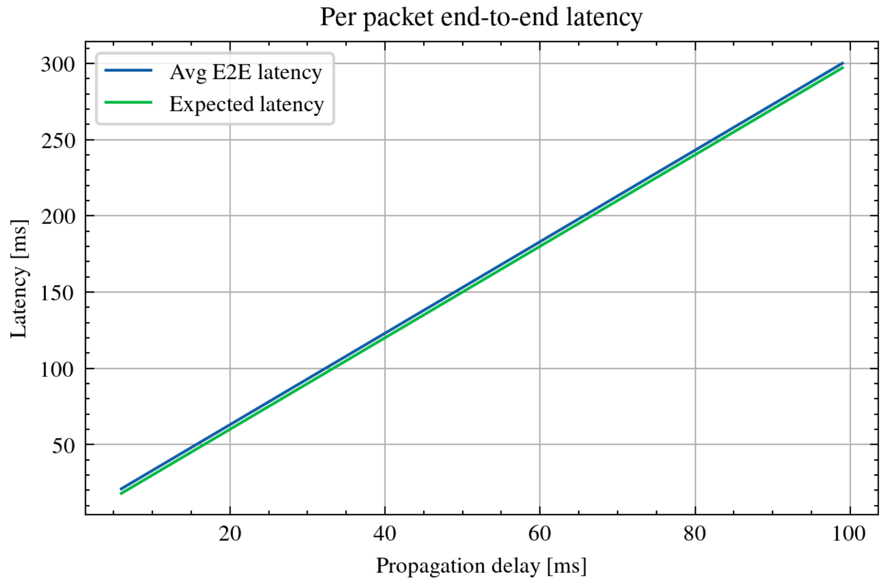

Figure 4.8: E2E latency vs. propagation delay with implemented solution

4.5 Inflated BSR

4.5.1 Problem description

Another problem occurred when the interval between the packets generated by the application, i.e. the packets interarrival time, was smaller than a single round-trip time.

The arrival of each packet automatically triggers the transmission of a scheduling request by the UE. However, each request is made for the whole transmission buffer. This causes a problem when many packets arrive before the gNB has the chance to respond with the appropriate resource grant.

Referring to Figure 4.9, we can see that the arrival of P1 in the transmission buffer of the UE triggers the transmission of a SR for a single packet. The second packet P2 arriving after 10ms triggers another request, however this second request is cumulatively made for the full buffer length, regardless of the fact that the first SR is still pending. This is denoted by the "SRx2" writing. Packets P3 and P4 behave in the same way, triggering requests for the allocation of three and four additional packets respectively.

Finally, the grants start arriving at the UE. The first 1-packet grant allows for the transmission of P1, while the second grant, requested for two packets, allows for the transmission of packets P2 and P3. The third grant, which corresponds to a previous request for 3 packets and could therefore allow the transmission of 3 packets, can now only be used by P4, wasting 2/3 of its potential capacity.

The last grant, that would allow the UE to transmit 4 consecutive packets, is completely wasted since the send buffer is now empty.

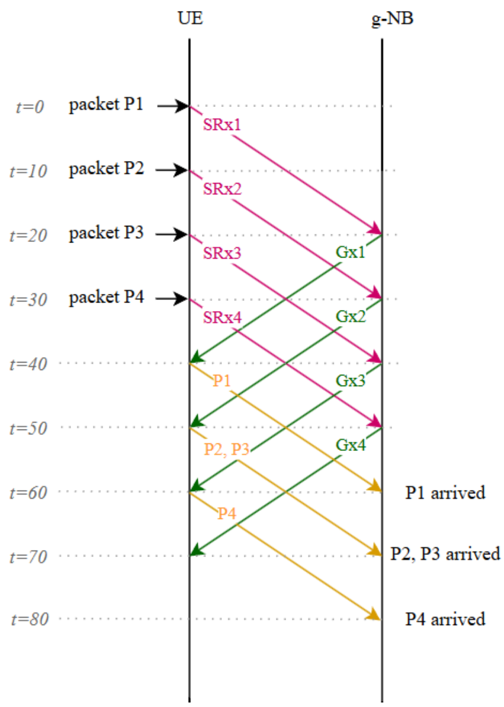

Figure 4.9: Packet diagram for interarrival times smaller than propagation delay

Repercussions The described behavior caused the physical throughput to be significantly higher than the application source rate, since the physical throughput metric counts all the capacity assigned to the UE, even though part of it might be just padding. Figure 4.10 shows the observed physical throughput for a constant source rate of 160Kb/s. The area between the two lines indicates that a lot of radio resources are being wasted to transmit few data. All the unused space in the grants that the UE receives is filled with padding, that counts as physical throughput but is useless at the higher levels. The growing trend when the propagation delay increases happens because the SR and grants take more time to arrive at their destinations, and the UE keeps sending inflated BSRs for more time, further contributing to waste more resources.

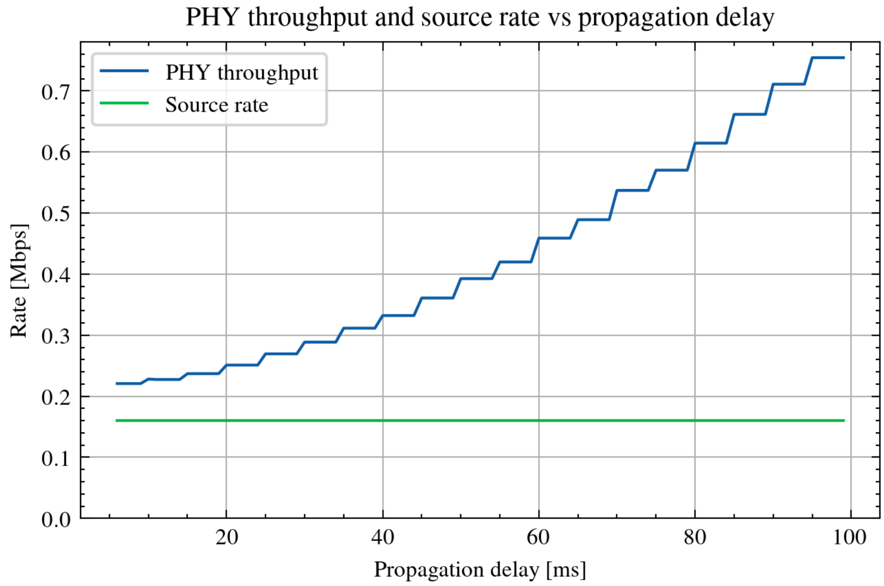

Figure 4.10: Physical throughput and application source rate vs. propagation delay with inflated requests

4.5.2 Implemented solution

Various different approaches could be undertaken in order to tackle this problem. Nonetheless, a comprehensive study detailing upsides and downsides of each of them is outside the scope of this thesis.

The implemented solution consists of a modification to the SR algorithm that limits the requests to only account for the newly received data, therefore disabling the cumulative behavior. If a packet is already in the buffer when a new one arrives, the scheduling request only asks the size of the second packet to be allocated.

The effectiveness of the implemented solution can be seen in Figure 4.11, showing that the physical layer throughput remains constant as the propagation delay increases, as reasonably expected since the source rate is constant throughout the simulation.

Finally, Figure 4.12 shows the behavior of the physical throughput as the application source rate increases. Note that the application rate is always smaller than the physical throughput. This happens because the transport block (TB) that the gNB assigns to the UE have dimensions that can only be integer multiples of the size of an OFDM symbol, therefore each TB can be slightly larger than the actual application-level packet. Furthermore,physical throughput also

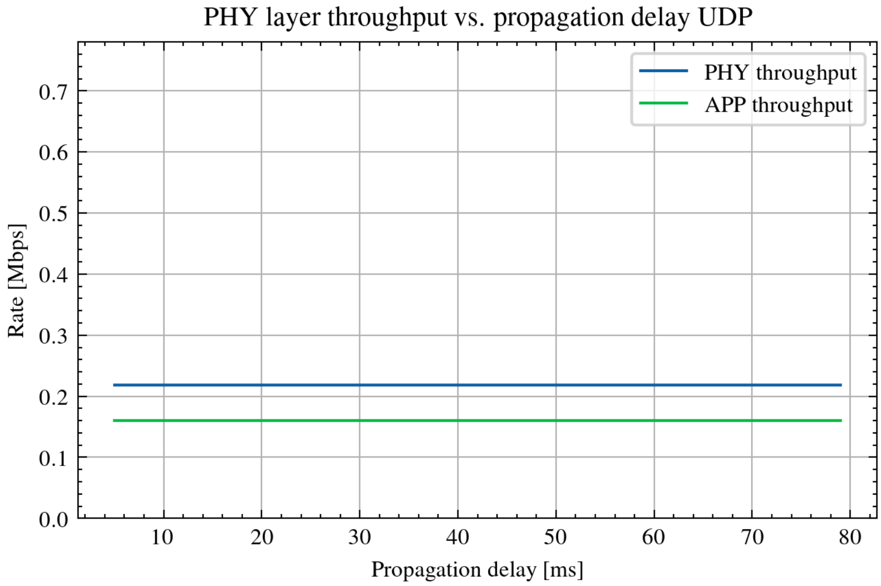

Figure 4.11: PHY throughput vs. propagation delay with implemented solution

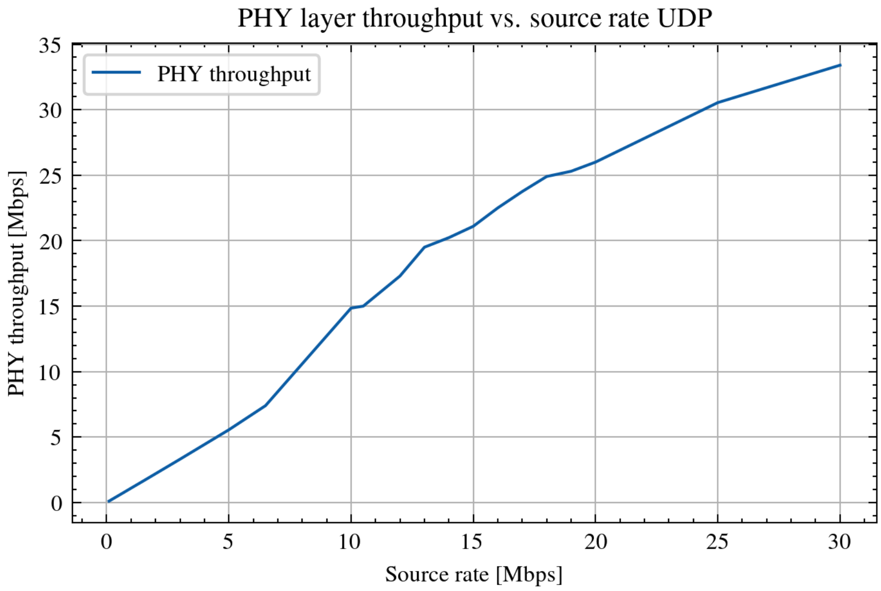

Figure 4.12: Physical throughput vs. source rate

accounts for protocol overheads, otherwise discarded in the computation of the application source rate.

ns-3 implementation The method responsible for the cumulative BSRs is `DoReportBufferStatus()` in the `LteRlcUm` class, therefore it was modified to only account for the newly arrived data when sending a SR.

4.6 Reordering timer

4.6.1 Problem description

Fragmentation The packets interarrival times are not necessarily multiples of the propagation delay, and the resources that the g-NodeB grants to the UE are usually a few bytes larger than the packet size. This happens because, even though 5G scheduling can be done on a much finer granularity than 4G scheduling, the base station still cannot grant values that are smaller than a single symbol, so the general rule is to allocate the minimum number of symbols whose cumulative size is greater or equal to the requested allowance.

This misalignment between packet size and transmission opportunities allows for packets to be split in smaller pieces. If two packets of 200B each are waiting in the buffer to be transmitted, and a resource grant for 220B arrives at the UE, the first packet can be completely transmitted, but it would be wasteful not to use also the 20 additional bytes that were granted, so the second packet is split and only its first bytes are transmitted. The remaining ones will wait for the next opportunity.

PDCP The detailed implementation of the packet data convergence protocol (PDCP) layer can be found in the 3GPP technical specification [41], while the diagram describing its functionality has been reported in Figure 4.13. This layer offers reordering services by employing a sequence number, integrity protection and ciphering if the packet in question is associated to a PDCP service data unit (SDU), and duplication and routing when operating split bearers.

The observed problem, as the propagation delay increases, is related to an expiring timer when reassembling the packets at the PDCP layer. The so-called reordering timer (t-reordering), which despite the name also affects the recomposition of fragmented packets, is not configured for the use in a non-terrestrial scenario, and it regularly expires due to the long propagation delay before the second half of the packet managed to arrive, leading to the discard of otherwise good packets.

This re-ordering functionality of PDCP layer is in place to ensure a sequential delivery of packets to the upper layers. In case of missing packets, the approach is to wait until either the packets arrive or the reordering timer expires [42]. It was observed that this timer also started when the PDCP layer was waiting for

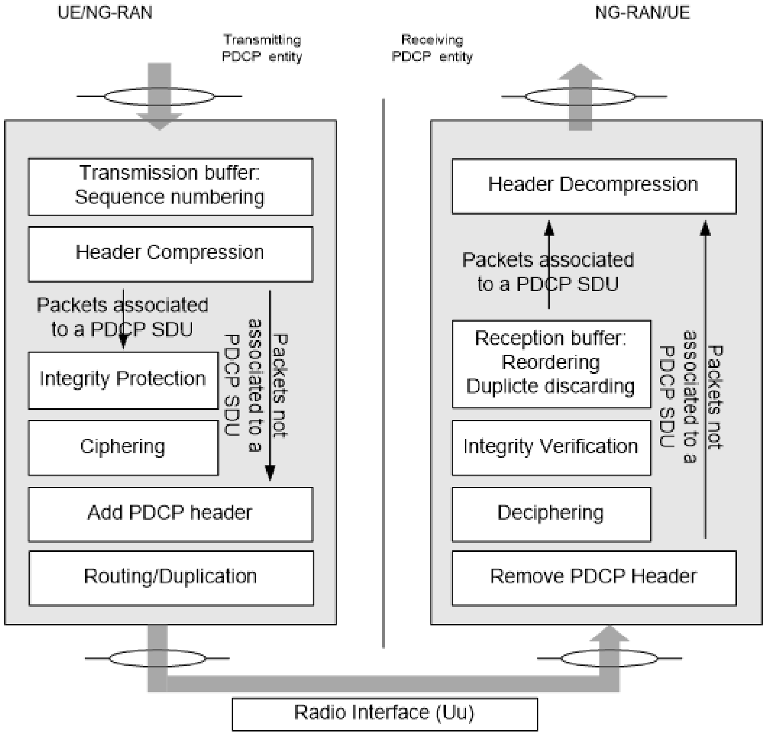

Figure 4.13: PDCP diagram [41]

the arrival of the second part of a fragmented packet. However, due to the high propagation delay, the second half of the packet did not arrive before the timer expiration, leading to the packet being discarded.

Should the reordering timer be set too high, it may cause additional latency, and, if set too low, it might cause many packets to be discarded, as in the case that was hereby simulated.

4.6.2 Implemented solution

Since the use of timers always bears trade-offs depending on their value, in this case either being a higher latency or a higher packet discard ratio, there can not be a one-fits-all approach: the differences in propagation delays between satellites orbiting at different altitudes are so drastic that the implementation of a single default value would lead to suboptimal performances in every possible scenario.

The class `LteRlcUm` was therefore modified to expose the `ReorderingTimer` parameter, allowing to easily set its value using the ns-3 attribute framework, similarly to the approach undertaken while solving the previous problem. This in turn permitted to set the timer to the value of the propagation delay plus a small constant accounting for processing delays for each simulator run.

5 Optimizing NR HARQ for NTNs

The previous chapters describe how we developed a working simulator for NR stack in a non-terrestrial scenario, the encountered problems and the proposed solutions. This chapter focuses on the improvement of NR NTN performances, in particular regarding the HARQ protocol, that has been identified by 3GPP as prone to failure due to the high propagation delays of NTNs. Our aim is to verify its current status and performances in NTNs, and optimize its behavior by designing and implementing some original modifications.

After a brief overview of HARQ key concepts, the problems with HARQ that emerged during the simulation campaing are presented, critically evaluated and followed by a detailed explanation of the implemented solutions.

5.1 HARQ overview

The main objective of telecommunication networks is the transfer of information between different actors. Modern systems aim at an efficient usage of the available resources, while trying to meet all the necessary requirements of the application that is generating the data to be transferred.

The HARQ protocol hereby described helps achieve a more efficient use of the available resources when errors occur, using an intelligent system of retransmissions that, in turn, reduces the error rate at the expense of a higher latency.

The main building blocks of HARQ are automatic repeat request (ARQ) and forward error correction (FEC). The role of ARQ is to automatically request the retransmission of the whole packet when the receiver detects the presence of errors, or when packets are not received at all, while FEC is tasked to correct such errors using redundancy bits added to the packet by the transmitter. The joint operation of those two protocols makes the foundation of HARQ, currently in use in all the most popular network standards such as 4G, Wi-Fi and 5G [43]. HARQ peculiarity resides in the fact that it avoids the retransmission of the whole packet in case of errors, preferring to send additional redundant information to help the decoding process.

5.2 Concurrent processes limit

One of the problems of HARQ highlighted by the 3GPP technical report [8] in the NTN scenario regards the maximum number of concurrent HARQ processes.

5.2.1 Problem description

The details of HARQ protocol implementation in the 5G NR standard is extensively treated in many publications such as [44]. However, for the purpose of understanding what is a HARQ process and how it affects the throughput in a non-terrestrial scenario, we only give a brief overview in the following paragraphs.

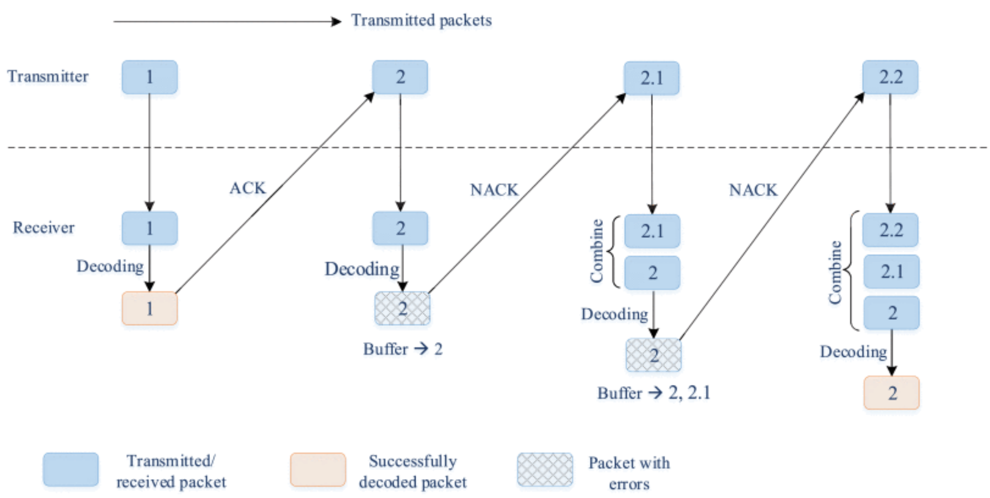

Figure 5.1: HARQ retransmission diagram [44]

HARQ working principle

Fig. 5.1 illustrates how HARQ processes work. Upon successful reception, depicted in the first column of Fig. 5.1, an acknowledgement (ACK) is sent back, triggering the transmission of the successive TB, which is represented in the second column. This behavior is the normal state in which transmissions are received correctly, and it keeps repeating itself until errors are detected.

Should the receiver detect errors in the received TB, represented by the greyed packet, a negative acknowledgement (NACK) is sent back to the sender, which in turn proceeds to send some additional redundancy bits. Note that the sender does not repeat the whole TB. The receiver now proceeds to decode the transfer block using all the information that it has received so far.

This is the most important feature of HARQ protocol: it does not discard the packets affected by errors, since they can be at least used to recover some information. The erroneous packets are stored in buffers and used for joint decoding [45].

If the redundancy bits that the sender just transmitted are still not enough to allow for a correct decoding of the transfer block, or if another error is detected, a retransmission is triggered. This is shown in the last column of Fig. 5.1. Note that all the information previously received, and not only the last part, is jointly used to decode packet 2.

Finally, if even this retransmission is affected by errors and the combination of the information received so far is not enough to complete the decoding of the TB, additional information is sent. After this fourth interaction, no further attempts are made to correct the packet [46].

The various transmissions and retransmissions being made by the protocol are called redundancy version (RV), and are numbered from 0 to 3. Their order can vary depending on the implementation and configuration. Figure 5.2 shows another example where the redundancy versions are ordered as 0, 3, 2.

Stop and wait

The presented behavior means that HARQ is a stop-and-wait kind of protocol, since it is designed to wait for the arrival of previous packet's ACK before

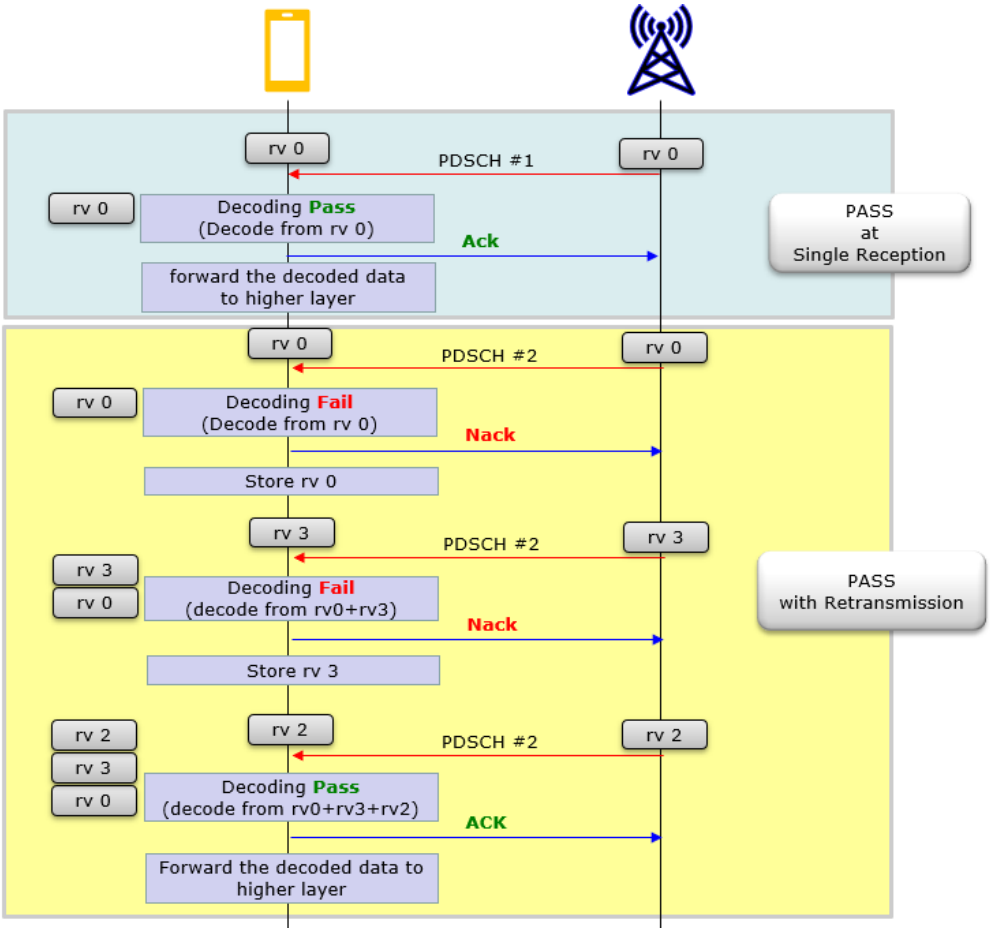

Figure 5.2: HARQ diagram with different RVs[46]

sending the new one. Figures 5.2 and 5.1 highlight this pattern. While this enforces the delivery of ordered packets, it also brings the downside of severely underutilized channel capacity, wasting resources that could potentially be used for transmission instead [47].

This limitation is overcome by the introduction of multiple concurrent processes.

Processes

A HARQ process starts when a TB is passed to the HARQ entity and finishes when the ACK relative to that same TB is received by the sender. After the ACK is correctly received, the next TB starts being processed. Considering a link with propagation delay τ_p, the minimum active time for a process is therefore $2\tau_p$, i.e., the time for the TB to arrive to the destination plus the time for the acknowledgement to travel back to the sender.

The 5G NR standard allows the base station to configure a certain number of concurrent HARQ processes to be assigned to each connected user, ranging from a minimum of 1 (even though the default is 8) to a maximum of 16 [45], [48]. In this way it is possible to partially limit the effects of the stop-and-wait behavior, allowing for a better throughput.

Application in NTNs

Since the propagation delay of NTNs is orders of magnitude larger than their terrestrial counterpart, the current maximum number of HARQ processes, which was optimized for terrestrial transmissions, leads to possible throughput degradation in the NTN domain, as detailed in the following toy example.

Example Consider a scenario where each process tries to send a TB every $2\tau_p$, which is the maximum rate at which transfer blocks can be sent under the condition of waiting for the acknowledgement to arrive before starting a new transmission. The other endpoint is a LEO satellite orbiting at 2.000Km, therefore $\tau_p \approx$ 6ms. Assuming the best possible conditions with no need for retransmissions and assuming 16 concurrent HARQ processes, the total send rate is of 16 transfer blocks every 12ms. In order to target a throughput of 50Mbps, the block size must therefore be of at least

$$\frac{\textit{target throughput} \times 2\tau_p}{\textit{number of processes}} = 37,5Kb$$

Doing the same calculation for a terrestrial scenario with the gNB placed at a distance of 600m from the UE, we obtain that the minimum total block size (TBS) must be of just $12b$. Both calculations do not factor in overheads, control information, channel access requests and processing delays, but are helpful to give an idea of the disproportion in place between the two conditions.

While the necessary block size for the NTN case is technically possible to achieve even with the older 4G technology, it necessitates a high signal-to-noise ratio (SNR) to work properly. This is because the larger is the block, the higher is the probability of encountering at least one error, therefore more retransmissions would become necessary. This constraint becomes even more conservative in the non-terrestrial case, where the propagation delay is much higher, and retransmissions can quickly become a lot more costly [49].

5.3 Possible solutions

5.3.1 Increasing the number of processes

The easiest solution would be to increase the number of maximum concurrent HARQ processes. Since each process works with a stop-and-wait behavior, the long round-trip time (RTT) can easily keep all the 16 maximum concurrent processes waiting. While all the processes are waiting their respective ACKs, the channel remains unused. Increasing the maximum number of concurrent HARQ processes can improve the throughput since the UE can spend less time waiting.

However, this comes with some caveats mainly regarding the higher computational capabilities required and higher power consumption, that can quickly become problematic in battery-operated equipment such as smartphones and /or sensors. Each process also requires the presence of a buffer at both the receiver side and the sender side, so additional memory resources are required at the gNB, too, which may not be feasible in some application scenarios with strict end user constraints.

5.3.2 Aggressive HARQ

A more sophisticated approach could involve the design of an aggressive version of the HARQ protocol, where each process is allowed to send multiple packets before receiving an acknowledgement e.g., implementing Sliding-Window HARQ. Since there already are multiple concurrent processes, each ACK packet must already contain a field specifying the number of processes it belongs to, and the information identifying the specific packet to be acknowledged within a process could be encoded inside this field.

5.3.3 Disable HARQ

Lastly, the option of disabling HARQ completely and rely solely on ARQ retransmissions has been proposed by 3GPP itself [50]. This, however, would come with a performance penalty in terms of communication accuracy and reliability, since satellite links typically suffer from more severe conditions than terrestrial ones, and [51] demonstrated that a version of HARQ specifically designed for NTNs would be beneficial.

5.4 Implemented solution - More processes

5.4.1 Testing current implementation

To test the practical effect that the concurrent HARQ processes' limitation has on the achievable throughput, a simulation campaign was conducted where the HARQ protocol was firstly disabled and successively enabled, therefore obtaining results for the two different scenarios. The maximum number of concurrent processes with HARQ enabled was set to 16, that is the maximum that a gNB can allocate to each UE according to the 5G NR specifications.

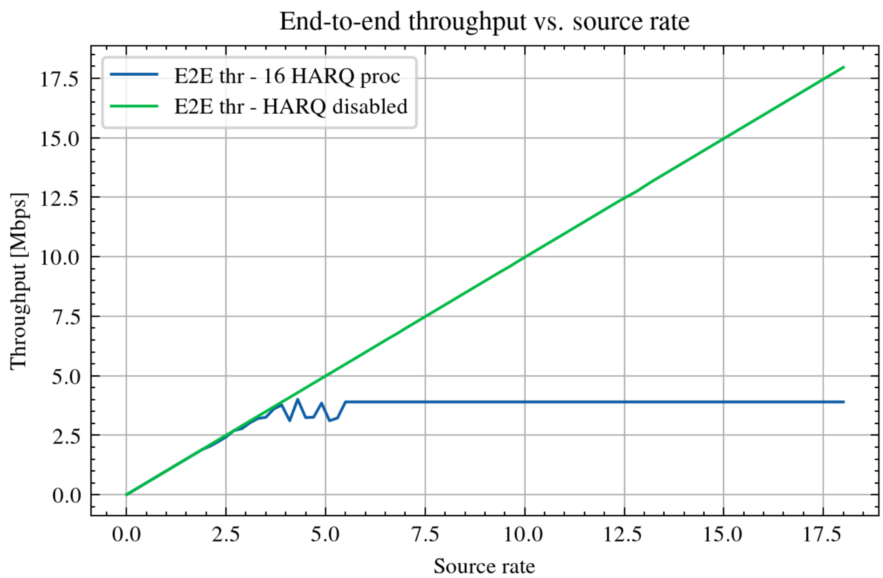

Figure 5.3: End-to-end throughput comparison with and without HARQ, τ_p = 6ms.

Comparing the obtained results confirmed that employing HARQ caused a noticeable negative impact on the throughput, as illustrated in Fig. 5.3. The throughput without HARQ perfectly matches the source rate, since the SNR values of this simulation are high enough to correctly deliver almost the totality of the packets.

On the other hand, the throughput with HARQ enabled reaches its plateau as the source rate crosses the 4Mb/s threshold, settling at this value. HARQ struggles to keep up with the increasing source rate since the limited number of processes starts to cause packets to be dropped whenever a retransmission is needed. As the source rate increases, the HARQ protocol is completely overwhelmed, as all the 16 processes are always busy, and the arriving packets that do not find an available process are promptly discarded directly at the sender. This is not a retransmission problem, rather a design problem.

Impact of propagation delay Figure 5.4 shows the application-level throughput as the source rate increases, this time while keeping the number of HARQ processes fixed to 16 but varying the propagation delay. Simulations were made with $\tau_p \in [6, 10, 20]$ms. We can see that the maximum achievable throughput is smaller as the propagation delay increases, since each HARQ process needs to wait more time for ACKs to arrive, dropping newer packets in the meantime.

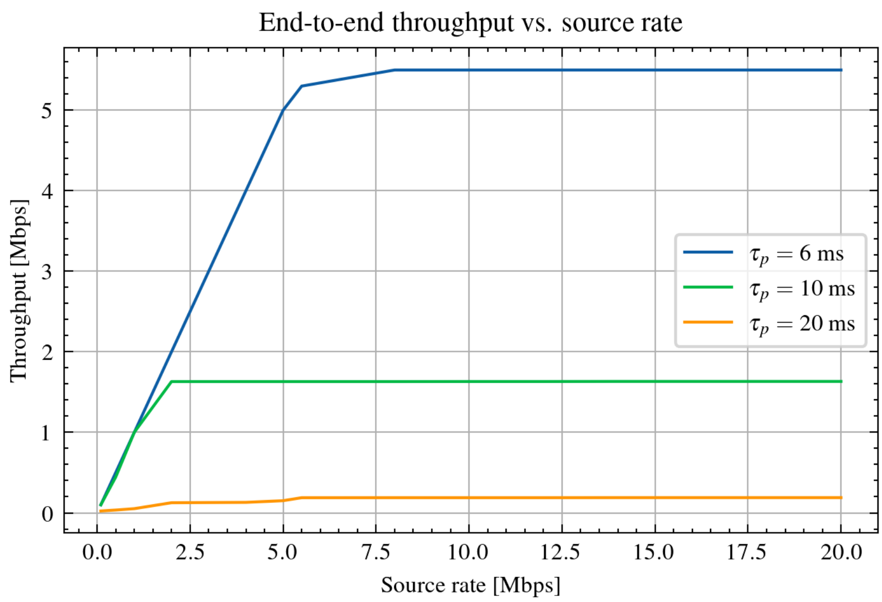

Figure 5.4: End-to-end throughput with different propagation delays, 16 HARQ processes

5.4.2 NS-3 IMPLEMENTATION

Having verified that HARQ does indeed limit the achievable throughput, a modification was implemented to the protocol suite to manually force a higher number of concurrent processes.

This was done by modifying the `NumHarqProcess` parameter of the class `MmWavePhyMacCommon`, making it accessible and simulating various scenarios with different number of processes.

However, a flaw in the `MmWaveFlexTtiMacScheduler` class in the ns-3 implementation of HARQ made the whole simulator crash whenever a new outbound packet found that all the processes were already occupied. If n processes were available, the simulator tried to access the $n+1$th process, causing an exception. The implemented solution consisted of a safeguard that checked the number of available processes, eventually discarding the packet.

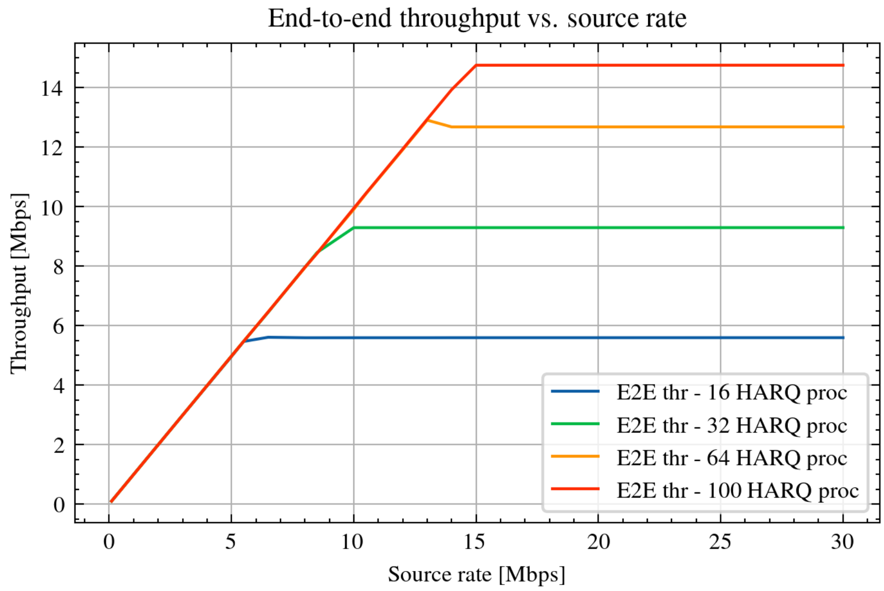

Figure 5.5: End-to-end throughput comparison with different number of concurrent HARQ processes, $\tau_p = 6$ ms.

5.4.3 Results

The results shown in figure 5.5 make clear that increasing the parameter regulating the number of HARQ processes allows for higher throughput. Each line in the plot was obtained with a different number of allowed processes, starting from 16, the maximum possible value in the 5G NR standard, then 32, 64 and ultimately 100. No further increases have been evaluated, since those would currently be unrealistic to achieve.

Increasing the allowed number of HARQ processes pushed the maximum achievable throughput up to around 14 Mb/s when considering 100 parallel processes. Going past such threshold causes packets to start being dropped, and the throughput to saturate, since there are no more available processes to send the traffic generated by the application.

However, constraints regarding the required computational power and the increased energy and memory consumption shall also be taken into careful consideration, since raising the number of concurrent processes from 16 to 100 is a big step, bringing a consistent increase in both the receiver and transmitter complexity with it.

We can conclude that, while certainly being a way to improve the performances of HARQ in non-terrestrial scenarios, the increase in processes alone cannot solve the problem of HARQ limiting the throughput, therefore other ways have to be found, and future studies should also consider alternative proposals.

5.5 Implemented solution - Disabling HARQ

Figure 5.3 could suggest that completely disabling the HARQ protocol could be a viable solution, since the end to end throughput of the simulation with HARQ switched off always matched the source rate, meaning that all the packets generated by the application were correctly received by the remote host.

However, the reason of this behavior is the high SNR that was forced in the simulated scenario with the choice of using high gain antennas. This design choice is not fully realistic, espcially considering direct-to-handle communication. Such choice was made because this thesis is focused on the study of the propagation delay, and the high gain of aperture antennas allowed to limit the impact of some other problems caused by low SNR. Having a communication channel with high SNR allowed for fewer variables to impact the simulations' results, in turn enabling the study of propagation delay alone.

In order to test the performances without HARQ enabled, a new simulation campaign was conducted factoring in the problems related to poor SNR, such as the presence of errors in the received packets and the need for retransmissions. The gain of the UE's antenna was lowered in order to appreciate the difference between the case with HARQ enabled and the case with HARQ disabled. In conditions of low SNR, e.g., when the distance between the terrestial end node and the satellite (and so the propagation delay) increases, which can be plausible in a non-terrestrial communication scenario, the packet delivery ratio (PDR) and therefore the throughput can drop considerably, impacting the link reliability.

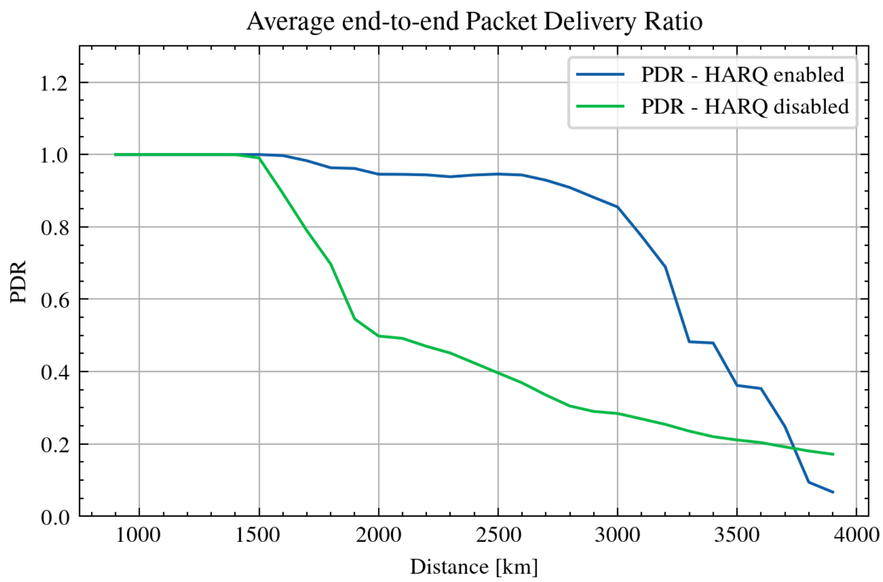

Figure 5.6: Packet delivery ratio with HARQ on and off, 16 concurrent HARQ processes

Simulations have shown that disabling HARQ is a good solution only when both UE and gNB experience good channel conditions with high SNR, while poor channel conditions benefit from having HARQ enabled. Figure 5.6 shows the packet delivery ratio as the orbiting altitude of the satellite increases, thus reducing the SNR. Initially, both scenarios correctly deliver all the packets, but, as the link quality decreases, the use of HARQ protocol proves to be beneficial for the reliability of the transmission, since it manages to correctly deliver more packets.

A hybrid approach can also be thought, where channel quality indicators (CQI) can be used to assess the state of the channel and decide whether to enable HARQ, therefore limiting the throughput in exchange for a higher reliability, or disabling it, should the SNR be high enough.

5.6 Implemented solution - Aggressive HARQ

In the previous section we disabled HARQ and compared the link reliability against the same scenario with HARQ enabled. Results showed that HARQ is still beneficial in cases of low SNR, therefore a modification was designed and implemented to try optimizing the protocol's behavior when dealing with high propagation delays. Now each HARQ process is allowed to send two packets at once without having to wait for the ACK to arrive. Once the ACK arrives, the process can send two more packets and so on.

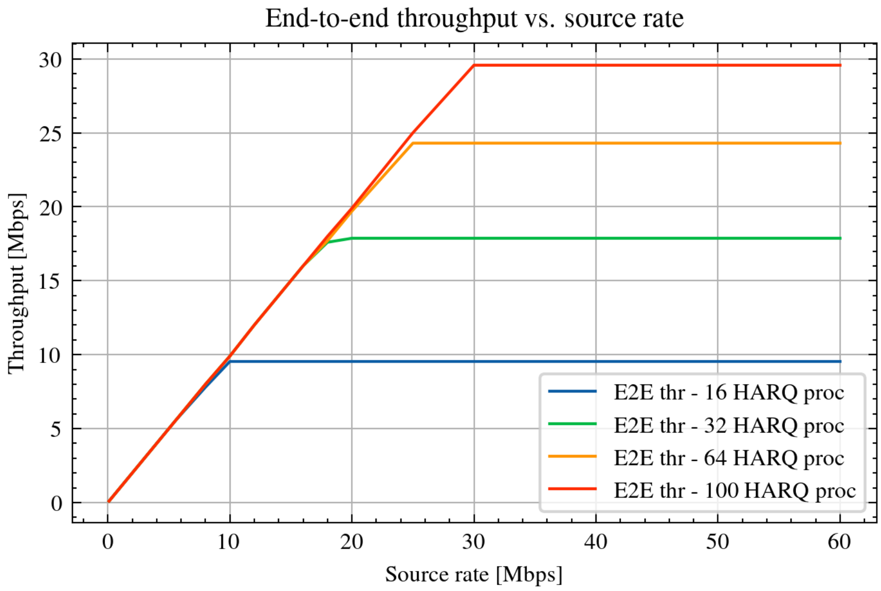

Figure 5.7: End-to-end throughput comparison with aggressive HARQ.

Figure 5.7 shows the results of the aggressive HARQ approach. Each line in the plot represents a different number of maximum concurrent HARQ processes allocated to each user.

As expected, for each number of concurrent HARQ processes, the throughput is roughly doubled with respect to the standard HARQ implementation of Figure 5.5, while the overall trend is similar.

Drawbacks This approach could be further pushed to accomodate for even more packets per process, allowing for a higher throughput to be achieved. However, we expect to incur in a penalty in terms of additional overheads caused by both control messages and packets retransmissions. Furthermore, additional resources are required, both in terms of computational capabilities and memory.

6
Conclusions and Future Works

This work has provided a comprehensive study on the impact that the high propagation delay experienced in NTNs has in layer-2 protocols. The current codebase of the ns-3 network simulator has been properly extended to allow for end-to-end simulation of the NR stack in a NTN scenario, since the ns-3 implementation of NR technology was not designed for high propagation delays. A lot of work went therefore into the design, implementation, and testing of a base support for the simulation of NTNs.

It was shown that propagation delay plays a crucial role in the performances of NR in NTNs, affecting both the efficiency and reliability of data transmission.

Once the simulator was ready, focus has been placed on the optimization of the HARQ protocol in NTNs, investigating its shortcomings when delaing with large propagation delays as well as proposing some original solutions to increase its performances.

6.1 Results

6.1.1 Simulator redesign

Since the state of the art in network simulation tools was not ready for complete NR NTN simulations, presenting numerous problems especially at the scheduler, a redesign of certain behaviors was performed.

Scheduling with propagation delay

This study has found (section 4.3) that the current scheduling with short advance (i.e. scheduling only a slot in advance: at time t_0 the schedule allocates the resources of time t_0 + 1ms) does not work in NTN because the propagation delay itself is longer than the delay between the act of scheduling and the time for which the scheduling happens.

The proposed solution uses information on the propagation delay to adjust the advance of the scheduling process.

BSR timer

Another flaw discovered in the implementation of current NR protocols in NTN involves the periodic BSR that the UE keeps sending every 10ms (section 4.4).

While this approach presents some (undesired) benefits, it is not the intended behavior, therefore the current implementation, albeit working, is far from an ideal choice.

The proposed solution consists of adapting the BSR timer accordingly to the propagation delay, never allowing it to assume values smaller than a round-trip time.

Inflated BSR

If the interval between the packets generated by the application is smaller than the RTT, each BSR will report the whole size of the transmission buffer, even though other SR relative to packets already present in the buffer may be still in-flight. While leading to an overall lower latency, this vastly increases the amount of resources being wasted, since the UE will find itself with many unnecessary transmission grants (Section 4.5).

The proposed solution is to limit the scheduling requests that are triggered every time a packet arrives into the transmission buffer to the new data only.

Reordering timer

The conducted simulation campaigns have found that anytime a packet is fragmented and sent across multiple frames, a reordering and recomposition timer is activated at the gNB side, which, however, is too short with respect to the large propagation delay in NTNs, expiring before all the pieces of the packet have the chance to arrive (Section 4.6).

The implemented solution was to extend such timer in order to account for the propagation delay.

6.1.2 HARQ

Concurrent processes

This work confirmed that the 16 maximum concurrent HARQ processes allowed per user heavily limits the achievable throughput (section 5.2).

Different solutions have been proposed, including completely disabling the protocol as well as increasing the number of allowed concurrent processes.

Both these solutions have also been implemented in the simulator, evaluating scenarios with no HARQ, then 16, 32, 64 and 100 maximum concurrent processes.

It was found that increasing the number of processes allows for considerably higher throughputs, while disabling HARQ is only feasible in conditions of high SNR.

Since having HARQ enabled still proved to be helpful in a scenario of low SNR, a more aggressive version of HARQ was designed, where each process was allowed to send two packets instead of one before waiting for the ACK. This solution managed to roughly double the achievable throughput with respect to the standard HARQ implementation.

6.2 Future work

The preparation of this work required many simulations to be run and evaluated, and as expected some design problems emerged. With a thoughtful approach, each unexpected behavior was investigated and solutions were devised, implemented and tested.

While effort was made to provide more than a single solution, striving to look at the problems from different perspectives to find more than a single way to approach it, some possibilities still require a deeper study, and some observations were made whenever it was felt that a point might benefit from additional work.

This section describes some possible future paths that might have the potential to improve the solutions proposed in this work, as well as different approaches which performances are still to be evaluated.

The current behavior of waiting for packets to arrive at the transmission buffer of the UE before transmitting the SR to the gNB harshly impacts the experienced latency, since in the best-case scenario it at least adds a round-trip time of delay. While this is not a problem in terrestrial networks because base stations are relatively close to the UEs that are serving, in NTNs the added delay is noticeable.

A predictive algorithm capable of visualizing patterns in the UE-generated traffic, forecasting its behavior in the immediate future and preemptively sending SR so that new packets will be able to be transmitted right away without additional delays would be of invaluable help in reducing the overall latency of the link.

Timers often represent a trade-off between higher performances and a more robust network. This is the reason behind the proposal of a dynamic approach when setting the values for BSR periodic requests and reordering timer. Since the delay of NTNs can experience large variations depending on the satellite orbit, configurable timers shall be preferred instead of using fixed values.

Regarding HARQ, a dynamic way of enabling and disabling it on the fly based on channel quality indicators could be beneficial, since it was shown that high-SNR scenarios performed better with HARQ disabled, while conditions of low SNR benefitted from having it enabled.

It is clear that further research is needed to develop more effective strategies for managing propagation delay in NTNs. This includes exploring new technologies, improving existing methodologies, and devising innovative network architectures.

Ultimately, this work aims to pave the way for future research and practical applications in the field of NTNs, emphasizing the need for continuous innovation and adaptation to meet the evolving demands of modern communication systems.

The findings of this thesis not only contribute to the existing body of knowledge on NTNs but also pave the way for future research in this evolving field.

References

[1] 3GPP, "Tr 21.917 "release 17 description; summary of rel-17 work items" v17.0.1," Tech. Rep., January 2023.

[2] F. Rinaldi, H.-L. Maattanen, J. Torsner, *et al.*, "Non-terrestrial networks in 5g and beyond: A survey," *IEEE Access*, vol. 8, pp. 165 178–165 200, 2020. DOI: 10.1109/ACCESS.2020.3022981.

[3] M. M. Saad, M. A. Tariq, M. T. R. Khan, and D. Kim, "Non-terrestrial networks: An overview of 3gpp release 17 & 18," *IEEE Internet of Things Magazine*, vol. 7, no. 1, pp. 20–26, 2024. DOI: 10.1109/IOTM.001.2300154.

[4] H.-L. Maattanen, B. Hofstrom, S. Euler, *et al.*, "5g nr communication over geo or leo satellite systems: 3gpp ran higher layer standardization aspects," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9014090.

[5] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6g era: Challenges and opportunities," *IEEE Network*, vol. 35, no. 2, pp. 244–251, 2021. DOI: 10.1109/MNET.011.2000493.

[6] A. Chaoub, M. Giordani, B. Lall, *et al.*, "6g for bridging the digital divide: Wireless connectivity to remote areas," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 160–168, 2022. DOI: 10.1109/MWC.001.2100137.

[7] E. Yaacoub and M.-S. Alouini, "A key 6g challenge and opportunity—connecting the base of the pyramid: A survey on rural connectivity," *Proceedings of the IEEE*, vol. 108, no. 4, pp. 533–582, 2020. DOI: 10.1109/JPROC.2020.2976703.

[8] 3gpp, "Tr 38.811 "study on new radio (nr) to support non-terrestrial networks" v15.4.0," Tech. Rep., October 2020.

[9] D. Wang, A. Traspadini, M. Giordani, M.-S. Alouini, and M. Zorzi, "On the performance of non-terrestrial networks to support the internet of things," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*, 2022, pp. 881–887. DOI: 10.1109/IEEECONF56349.2022.10052102.

[10] D. Wang, M. Giordani, M.-S. Alouini, and M. Zorzi, "The potential of multilayered hierarchical nonterrestrial networks for 6g: A comparative analysis among networking architectures," *IEEE Vehicular Technology Magazine*, vol. 16, no. 3, pp. 99–107, 2021. DOI: 10.1109/MVT.2021.3085168.

[11] M. Höyhtyä, S. Boumard, A. Yastrebova, P. Jarvensivu, M. Kiviranta, and A. Anttonen, "Sustainable satellite communications in the 6g era: A european view for multi-layer systems and space safety," *IEEE Access*, vol. PP, pp. 1–1, Jan. 2022, Licensed under CC-BY 4.0 public license. DOI: 10.1109/ACCESS.2022.3206862.

[12] ESA, *Types of orbits*, esa.int, Accessed 2024-07-15.

[13] B. R. Elbert, "Commercial communications satellites in the post-2020 era," in *Advances in Communications Satellite Systems. Proceedings of the 37th International Communications Satellite Systems Conference (ICSSC-2019)*, 2019, pp. 1–16. DOI: 10.1049/cp.2019.1235.

[14] M. Giordani and M. Zorzi, "Satellite communication at millimeter waves: A key enabler of the 6g era," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 383–388. DOI: 10.1109/ICNC47757.2020.9049651.

[15] I. Morgan-Jones and P. Loskot, *Regional coverage analysis of leo satellites with kepler orbits*, 2019. arXiv: 1910.10704 [physics.space-ph]. [Online]. Available: https://arxiv.org/abs/1910.10704.

[16] A. May, *Low earth orbit: Definition, theory and facts*, rohde-schwarz.com, Accessed 2024-07-20.

[17] 3GPP, "Tr 38.821 "solutions for nr to support non-terrestrial networks (ntn)" v16.2.0," Tech. Rep., April 2023.

[18] Rohde-Schwarz, *Connecting the world with 5g ntn*, rohde-schwarz.com, Accessed 2024-07-20.

[19] S. Daily, *First internet ready satellite launched*, [Online; accessed 15-August-2024], 2003. [Online]. Available: https://www.spacedaily.com/news/satellite-biz-03zza.html.

[20] M. Sandri, M. Pagin, M. Giordani, and M. Zorzi, "Implementation of a channel model for non-terrestrial networks in ns-3," in *Proceedings of the 2023 Workshop on ns-3*, ser. WNS3 2023, ACM, Jun. 2023. DOI: 10.1145/3592149.3592158. [Online]. Available: http://dx.doi.org/10.1145/3592149.3592158.

[21] L. Campanile, M. Gribaudo, M. Iacono, F. Marulli, and M. Mastroianni, "Computer network simulation with ns-3: A systematic literature review," *Electronics*, vol. 9, no. 2, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9020272. [Online]. Available: https://www.mdpi.com/2079-9292/9/2/272.

[22] *Omnet++*, [Online; accessed 15-August-2024]. [Online]. Available: https://omnetpp.org/.

[23] *Swans*, [Online; accessed 15-August-2024]. [Online]. Available: http://jist.ece.cornell.edu/.

[24] *Netsim*, [Online; accessed 15-August-2024]. [Online]. Available: https://www.tetcos.com/.

[25] *Ns-3*, [Online; accessed 15-August-2024]. [Online]. Available: https://www.nsnam.org/.

[26] *Ns-3 official website*, nsnam.org, Accessed 2024-07-22.

[27] 3gpp, "Tr 38.901 "5g study on channel model for frequencies from 0.5 to 100 ghz" v16.1.0, release 16," Tech. Rep., November 2020.

[28] T. Zugno, M. Polese, N. Patriciello, B. Bojović, S. Lagen, and M. Zorzi, "Implementation of a spatial channel model for ns-3," in *Proceedings of the 2020 Workshop on Ns-3*, ser. WNS3 '20, Gaithersburg, MD, USA: Association for Computing Machinery, 2020, pp. 49–56, ISBN: 9781450375375. DOI: 10.1145/3389400.3389401. [Online]. Available: https://doi.org/10.1145/3389400.3389401.

[29] A. Ammar, A. Ellaf, A. Zerek, and Y. Jaradat, "The e-band as future candidate for next generation networks," Dec. 2018.

[30] D. D. Morabito, "A comparison of estimates of atmospheric effects on signal propagation using itu models: Initial study results," November 15, 2014. [Online]. Available: https://ipnpr.jpl.nasa.gov/progress_report/42-199/199D.pdf.

[31] W. contributors, *Local tangent plane coordinates - wikipedia, the free encyclopedia*, [Online; accessed 26-July-2024], 2023. [Online]. Available: `https://en.wikipedia.org/w/index.php?title=Local_tangent_plane_coordinates&oldid=1177505970`.

[32] D. J. L. Whitrow, "The radiation patterns of circular apertures," *National Security and ISR Division, Defence Science and Technology Group*, 2018.

[33] M. Mezzavilla, M. Zhang, M. Polese, *et al.*, "End-to-end simulation of 5g mmwave networks," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 3, pp. 2237–2263, 2018. DOI: `10.1109/COMST.2018.2828880`.

[34] ShareTechNote, *5g/nr - scheduling*, sharetechnote.com, Accessed 2024-08-16.

[35] 5. Networks, *"5g nr frame structure"*, 5g-networks.net, Accessed 2024-08-16, 2020.

[36] 3gpp, "Ts 38.211 "5g nr physical channels and modulation v16.2.0, release 16," Tech. Rep., July 2020.

[37] GaussianWaves, *5g nr resource block*, gaussianwaves.com, Accessed 2024-08-16.

[38] GSMA, "5g tdd synchronisation guidelines and recommendations for the coexistence of tdd networks in the 3.5 ghz range," Tech. Rep., April 2020, Accessed 2024-07-21.

[39] H. Chougrani, S. Kisseleff, W. A. Martins, and S. Chatzinotas, "Nb-iot random access for nonterrestrial networks: Preamble detection and uplink synchronization," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14 913–14 927, 2022. DOI: `10.1109/JIOT.2021.3123376`.

[40] 3gpp, "5g; nr; medium access control (mac) protocol specification (3gpp ts 38.321 version 18.1.0 release 18)," Tech. Rep., May 2024.

[41] 3gpp, "5g; nr; packet data convergence protocol (pdcp) specification (3gpp ts 38.323 version 18.1.0 release 18)," Tech. Rep., May 2024.

[42] S. Pandey, R. M. M. K. Eda, and D. Das, "Efficient reordering-reassembly pdcp and rlc window management algorithm in 5g and beyond," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2020, pp. 1–6. DOI: `10.1109/CONECCT50063.2020.9198680`.

[43] 3gpp, "3gpp 38 series of specifications: 5g new radio," Tech. Rep.

[44] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem, and E. Hossain, "Hybrid automatic repeat request (harq) in wireless communications systems and standards: A contemporary survey," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 4, pp. 2711–2752, 2021. DOI: 10.1109/COMST.2021.3094401.

[45] Devopedia, *"5g nr hybrid arq"*, devopedia.org, Version 6, December 19, Accessed 2024-07-14, 2021.

[46] ShareTechNote, *5g/nr - harq*, sharetechnote.com, Accessed 2024-07-14.

[47] 3gpp, "Tr 38.214 "5g nr physical layer procedures for data" v16.5.0, release 16," Tech. Rep., April 2021.

[48] 3gpp, "Tr 38.300 "5g nr overall description stage-2" v15.9.0, release 15," Tech. Rep., April 2021.

[49] ShareTechNote, *4g/lte - phy processing*, sharetechnote.com, Accessed 2024-07-14.

[50] A. Mukherjee, "Hybrid arq schemes," 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:219510991.

[51] A. Vanelli-Coralli, A. Guidotti, T. Foggi, G. Colavolpe, and G. Montorsi, "5g and beyond 5g non-terrestrial networks: Trends and research challenges," in *2020 IEEE 3rd 5G World Forum (5GWF)*, 2020, pp. 163–169. DOI: 10.1109/5GWF49715.2020.9221119.

Acknowledgments

Dedico quest'ultima parte della mia tesi a tutte le persone, professori, familiari, parenti e amici, che mi hanno accompagnato durante questo percorso.

Desidero ringraziare i miei relatori, professori Michele Zorzi e Marco Giordani, per avermi seguito durante il tirocinio e nella fase di stesura della tesi. Un sentito ringraziamento va anche a Matteo Pagin, guru di ns-3, che mi ha aiutato a familiarizzare con questo strumento un po' ostico, contribuendo alla realizzazione di questo elaborato.

Ringrazio la mia famiglia, mamma Katia, papà Rudy e Pietro. Grazie per l'affetto e gli incoraggiamenti che mi avete dato, per i sacrifici che avete fatto, e per tutte le lezioni di vita che in questi 24 anni mi hanno formato come persona. Grazie a tutti i parenti: nonni Paolo e Franca, zie Chiara, Gianky e Patrizia, zii Nicola, Michele, Walter, e cugine, Giulia, Marta, Matilde, Agnese, Maddalena, Margherita per essermi sempre stati a fianco, per tutte le serate passate in allegra compagnia, i pranzi a base di sushi e i tanti bei momenti che abbiamo condiviso.

Ci tengo a ringraziare la nonna Anna e il nonno Armando, che, pur non essendo qui con noi, hanno sempre creduto in me volendomi bene, guidandomi in questo viaggio anche se da lontano. Penso che se fossero qui ora sarebbero felici, e questo mi riempie di gioia.

È il turno dei compagni di viaggio. Grazie a Mattia, che lo è stato sin dai primi giorni con Canazza e Martinazzi fino ad oggi, per concludere insieme quest'avventura. Grazie a Sara per tutte le lunghe chiacchierate che abbiamo fatto, e per aver offerto le meringhe a tutta l'aula. Grazie a Prince, con cui ho condiviso troppe recensioni di pizzerie e ristoranti, e grazie ad Auro e Fede per tutte le sedute di gossip terapeutico.

Agli "amici di Dese". Grazie a Francesco e Bassetto per essere stati compari di mille avventure. Grazie a, Nico, Lollo, Bolzo, Jacopo, Bisio, Luca, Mattia. Grazie a Eleonora, Virginia, Lisa, Anna, Giulia, Martina. Siete tanti ma mi fa

piacere nominarvi tutti. A tutte le volte che ci siamo trovati anche solo per fare due chiacchiere al pub, agli insulti che mi avete riservato per il mio 29 in digital storytelling, a tutte le risate in compagnia, le arrabbiature, le ore piccole a chiacchierare in macchina, per esserci stati a condividere le gioie e dividere le preoccupazioni.

Cesare Pavese scrisse "*Non si ricordano i giorni, si ricordano gli attimi*", e voi me ne avete donati molti, grazie.

Francesco.