UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

MASTER THESIS IN ICT FOR INTERNET AND MULTIMEDIA

Scheduling resources and managing HARQ retransmissions in Non-Terrestrial Networks

MASTER CANDIDATE

Francesco Rossato

Student ID 2082121

SUPERVISOR

Professor Michele Zorzi

University of Padova

CO-SUPERVISOR

Prof. Marco Giordani

University of Padova

Dr. Matteo Pagin

University of Padova

ACADEMIC YEAR
2023/2024

To my parents
and friends

Abstract

The advent of 5G technology has revolutionized the world of mobile communication, offering higher throughput, lower latency and high reliability. As the demand for connectivity continues to grow, non-terrestrial networks (NTNs) have been identified by the Third generation Partnership Project as crucial 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 complexities not typically encountered in their terrestrial counterpart, and the current protocol stack was not designed with such challenges in mind. Through a comprehensive literature review and the use of ns-3 network simulator, this work identifies key characteristics and constraints of non-terrestrial networks, evaluating existing protocols and highlighting their strengths and limitations in this challenging scenario, before finally delving into the adaptation and optimization of 5G New Radio Layer-2 protocols for NTNs, proposing innovative solutions and paving the way for new applications and services that leverage the full potential of 5G technology beyond traditional terrestrial boundaries.

Contents

List of Figures

List of Acronyms

TN terrestrial networks

NTN non-terrestrial network

NTNs non-terrestrial networks

HAPs high-altitude platforms

HAP high-altitude platform

UAVs unmanned aerial vehicles

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

RACH random access channel

gNB gNodeB

3GPP third generation partnership project

MAC media access control

RTT round-trip time

ACK acknowledgement

TB transport block

TBS total block size

SINR signal plus interference to noise ratio

SNR signal-to-noise ratio

NACK negative acknowledgement

ISL inter-satellite links

FEC forward error correction

RV redundancy version

TDD time division duplexing

SR scheduling request

BSR buffer status report

PDCP packet data convergence protocol

SDU service data unit

LOS line of sight

NLOS non line of sight

vSAT very small aperture terminal

MTU maximum transmission unit

PDR packet delivery ratio

CQI channel quality indicators

1 Introduction

1.1 Structure of this work

The following work is structured as followingly described.

- **Chapter 1** consists of a general introduction to non-terrestrial networks. Their characteristics are hereby described, and some scenarios are provided where their application would be beneficial. The three main categories of telecommunication satellites are presented, and their characteristics evaluated. Finally, an overview of the possible payload design choices are also presented, as well as a brief description of multilayered networks.

- **Chapter 2** describes NS-3, the employed network simulator software, highlighting the NTN channel module used as well as its implementation. Finally, this chapter presents the simulated scenario, describing the characteristics and the parameters of all the involved devices.

- **Chapter 3** is about the scheduler. The 5G scheduler and its operations in time division duplexing are briefly described, then the main design flaws that arose during the simulations in non-terrestrial scenarios are documented, and the implemented solutions are carefully explained. Some additional observations that were not implemented are hereby described as well, possibly providing some potential starting points for future studies.

- **Chapter 4** delves into the problems linked to the operation of the HARQ protocol in a non-terrestrial scenario, discussing the effects of propagation delay as well as the impact of different values of SNR. Some solutions are proposed and evaluated using the NS-3 network simulator.

- **Chapter 5** finally reports the conclusions of this work, highlighting how the scheduler and HARQ are affected by the characteristics of a non-terrestrial link, and providing some cues that can be evaluated in future studies.

1.2 The non-terrestrial era

The third generation partnership project (3GPP)[1], the standardization body tasked with 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].

In the specific, the envisioned future for mobile communications, starting with the already established 5G new radio (NR) and further expanding 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 [30] [22].

To give a proper definition, the term non-terrestrial networks (NTNs) refers to a category of networks where at least one link is routed via an aerial or space-borne vehicle such as high-altitude platforms (HAPs), unmanned aerial vehicles (UAVs) or telecommunication satellites.

1.3 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 current terrestrial infrastructure has some limitations, failing to provide an adequate service, and NTNs can be used to provide a better coverage.

1.3.1 Remote places

While terrestrial networks (TN) 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 rural and remote areas. Conditions such as harsh terrain and geographical impediments act as natural barriers to the deployment of terrestrial infrastructure. Moreover, the need for ground equipments such as base stations and networking gear re-

[1] `3gpp.org`

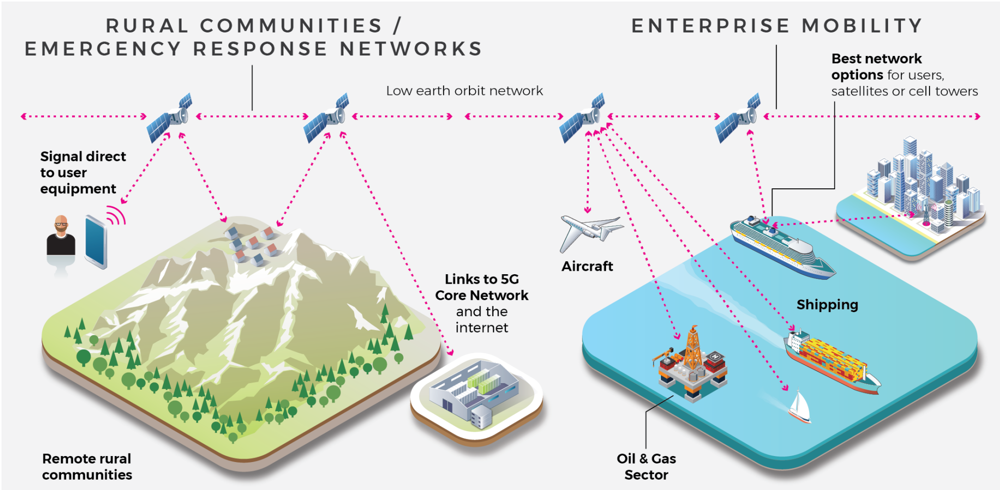

Figure 1.1: NTN use cases telecominfraproject.com

quire the presence of an already established and reliable power grid, further driving up the costs that telecommunication companies would have to sustain.

The population density in remote and rural places is typically much lower than cities, and users are spaced in vast areas of land. Because of this reason, the investment that has to be made in order to provide coverage to a certain number of users would be much higher compared to a more urbanized scenario, where potential users are more densely distributed. The already high infrastructure cost will therefore hardly generate 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.

As studied and documented in [38], 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.

1.3.2 Redundancy 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 last for a long period of time. This in turn can cause significant economical damage and hinder the already difficult rescue efforts, potentially leading to loss of lives.

The simple installation of a greater number of base stations is not a viable solution because an extreme event such as a tornado would easily be able to render all the additional equipment useless, and 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 and also 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.

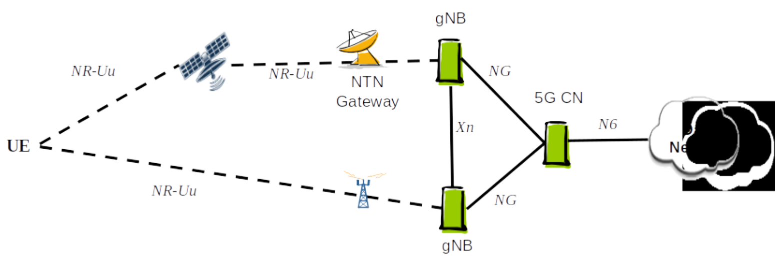

Figure 1.2: Transparent use of multiple radio access technology [8]

1.3.3 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. 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 non-terrestrial networks 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 of a number of sensors placed in a large area. When distances are large, solutions may either be the densification of radio stations or the use of a lower frequency in order to have a less severe propagation loss. However, those approaches have their downsides and are not always feasible, requiring costly, ad-hoc solutions tailored for the specific scenario. In this case, the vast coverage area of NTNs will undoubtedly be useful to provide internet connection [35].

Other scenarios where NTNs can become useful in overcoming the limitation of their terrestrial counterpart are well described in [18] and [36].

1.4 Satellite types

Satellites are divided in three different categories depending on their orbiting altitude: geosynchronous equatorial orbit (GEO), medium earth orbit (MEO), and low earth orbit (LEO) satellites. Each one has its own characteristics, presenting some upsides and downsides as briefly described below. Fig. 1.3 illustrates the different orbiting altitudes as well as the approximate coverage area for each orbit.

1.4.1 GEO satellites

Orbiting at a height 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.

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 would require the deployment of a constellation, which is both more complex and more expensive.

This also vastly simplifies the tracking for the ground equipment. Since the satellite position is always known, once the position of the user equipment (UE) is established, the relative position of the satellite can be easily calculated.

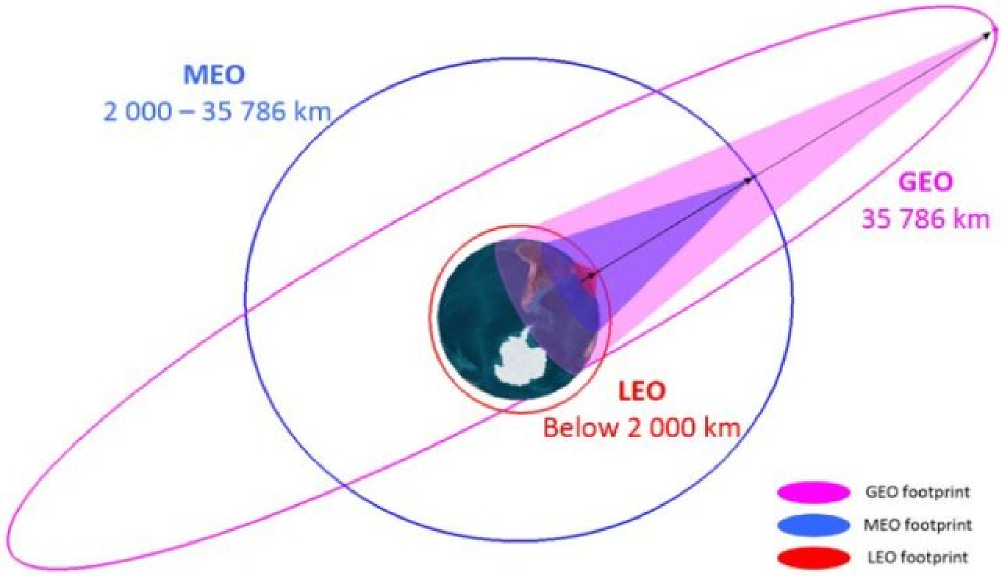

Figure 1.3: Height and approximate coverage areas for satellites at different orbits[21]

As shown in Fig. 1.3, the high altitude of GEO satellites creates a large cell footprint. While the deployment cost of a single GEO satellite is higher than both MEO and LEO ones, the cost per coverage area is overall lower, and an almost full coverage of the terrestrial globe can be achieved using only three equally spaced satellites [17].

Disadvantages The disadvantages of GEO satellites are mainly linked to the large distance with UEs places on the Earth surface: the transmission power and the antenna gain have to be high enough to overcome the greater propagation losses, and the propagation delay of the signal alone adds about 120ms to the overall latency. 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 at $t = 240$ms. The response will then finally reach the UE after at least 480ms from the initial transmission, and these calculations do not factor in any delay related to medium access requests, packet transmission times and processing delays, which would further increased 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 area, a single satellite will be required to serve a massive number of users, so 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.

Solutions to provide a greater capacity have been proposed and are currently in the early stage, such as 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. 1.4 [16]. Moreover, the high number of users leads to a large rate of initial access requests, with the possibility of channel saturation as described in [8].

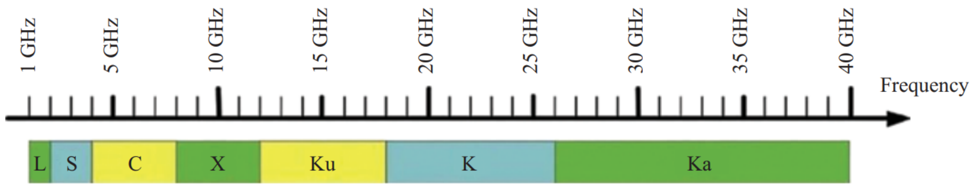

Figure 1.4: Satellite spectrum bands allocation [16]

1.4.2 MEO satellites

MEO orbit is comprised between LEO and GEO, therefore all satellites orbiting between 2000Km and 35.786Km are considered as MEO. This vast orbital space is mainly used by navigation systems such as GALILEO [17].

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

MEO satellites do require a constellation in order to provide continuous coverage over a designated area, since they are not geostationary. The numerology of the constellation, however, is smaller than the one required for LEO satellites, since each platform can serve a larger area.

Their peculiarity of presenting many of the downsides that characterize both GEO and LEO satellites, while being unable to offer any substantial benefit over their competitors besides the need for a smaller constellation, makes them less than ideal candidates for applications in non-terrestrial networks.

1.4.3 LEO satellites

Orbiting below the threshold altitude of 2.000Km, LEO satellites are the most promising solution in the realm of NTNs for a number of reasons hereby discussed.

Advantages The lower altitude entails a shorter propagation delay, on the order of about 6ms, and the smaller coverage area of each satellite depicted in Fig. 1.3 means that the total number of users that need to be served is smaller. This also allows the use of higher frequency bands and the constraints of high antenna gains are less stringent compared to GEO satellites, since the experienced path loss is much smaller. This in turn enables the achievement of higher overall throughputs, more suited to satisfy the requirements of modern days broadband connectivity, as detailed in [19].

Disadvantages The cost per deployed satellite is significantly smaller than GEO and MEO satellites, and multiple deployments within a single launch are possible, further driving the costs down. However, given that LEO satellites are not geostationary, a large constellation is needed to provide a continuous service, driving up the deployment costs significantly. As an example of how vast those constellations can become, Fig. 1.5 depicts the LEO satellites employed by Starlink[2], with 4.808 units in service at the time of writing[3].

Since low orbiting satellites remain in view of the user equipments only for a short period of time, with an average in-view duration of just 13 minutes as calculated in [25], all the connected users are expected to be handed over to the next available satellite 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 data 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: high-bandwidth links between different satellites of the constellation, capable of connecting satellites that do not have gateways in sight to ones that are connected to a gateway, allow-

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

[3]Source: `satellitemap.space`

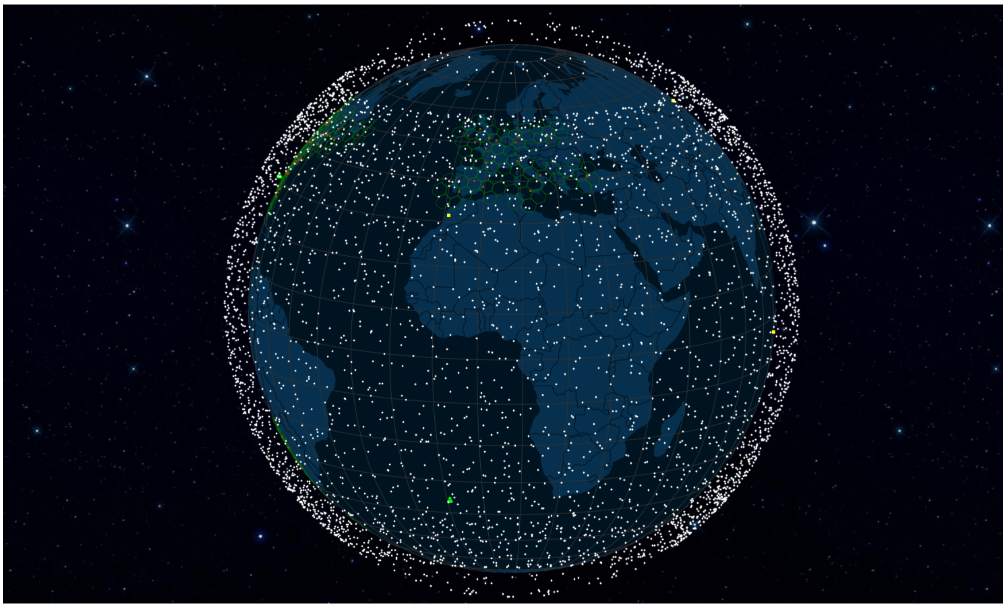

Figure 1.5: Starlink constellation as July 2024. Source: `satellitemap.space`

ing traffic to be routed to the ground with an additional hop. Inter-satellite links have to be implemented if coverage over the oceans is required. This ultimately adds up to the already high constellation deployment costs.

A problem 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 about 7.8Km/s [23], presenting therefore a noticeable Doppler shift. This has to be compensated for, and preliminary solutions in this sense require usage of UEs with GNSS capabilities, which is not always a reasonable assumption [19, 2].

1.4.4 Multilayered networks

The aforementioned solutions are not to be considered mutually exclusive. In fact, the multitude of possibilities that their combinations offer opens to the study of various different scenarios, in which the upsides of the space, air, and ground layers are orchestrated to improve quality of service. Fig. 1.6 showcases a highly sophisticated non-terrestrial multilayered network scenario where different access technologies are used.

In [36] it has been shown that the use of high-altitude platform (HAP) 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 probabil-

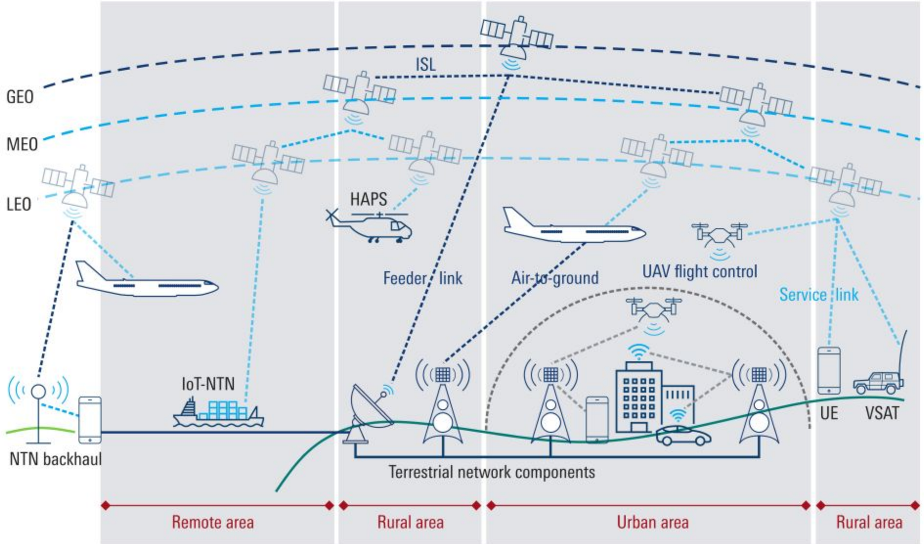

Figure 1.6: Complex multilayered NTN scenario [29]

ity, than point-to-point GEO transmissions.

1.5 Types of payloads

When implementing a non-terrestrial network, an important parameter to be decided is the type of payload to use.

1.5.1 Bent-pipe payload

This is the simplest approach, where the role of the satellite consist only of repeating the signal received from the UE on the ground towards the terrestrial gateway. Configurations such as the one depicted in Fig. 1.7 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 with 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, increasing the latency by at least two times the propagation delay. This scenario also poses strict bandwidth requirements on the feeder link, since all the traffic must necessarily pass through it. Other, smarter, solutions are able to route at least part of the inbound traffic autonomously, without routing everything back to earth.

Figure 1.7: NTN with access network based on satellites with bent pipe payload [8]

1.5.2 On-board g-NodeB

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

Figure 1.8 from [8] shows the architecture of a NTN using access network based on satellites with gNB on board as payload.

Figure 1.8: NTN with access network based on satellites with gNB payload [8]

1.6 Available commercial solutions

Legacy solutions Commercial solutions first concerning satellite-based phone calls and successively evolved to provide internet access have been available since a long time. However, the majority of this legacy infrastructure makes use of GEO satellites, therefore presenting all the limitations discussed in section 1.4, offering limited throughput and large delays.

Recent developments While different commercial solutions using LEO satellites have been successfully deployed so far, with notable examples such as the one presented in section 1.4.3, all of them make use of many proprietary protocols to allow for the communication to take place, and up until now no internationally agreed standard has been defined.

2 The ns-3 network simulator

The study behind this work required a thorough evaluation of a multitude of different non-terrestrial communication scenarios, that in turn required an extensive simulation campaign. Furthermore, the low-level nature of the issues that were expected to be found led to the need for a simulator that allowed access to the protocols' core implementation to be able to modify their behavior if needed. A simple API-level access to some protocols' attributes would have been not sufficient to implement the suggested modifications.

2.1 Description

Figure 2.1: 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, meaning that the condition of being able to have full access to the source code to implement potential modifications is satisfied [27].

Other simulation pieces of software are available, such as OMNET++ (omnetpp.org), SWANS (jist.ece.cornell.edu), NetSim (tetcos.com), QualNet, and

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 2.1: Network simulation software comparison

finally ns-3 predecessor, ns-2. Their main characteristics are summarized in Table 2.1, from [11], which highlighted the suitability of ns-3 for research purposes, highlighting its success amongst the scientific community.

Discrete events simulators

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. At the eyes 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 scheduled to execute for a certain period of time, the simulation immediately transitions to the next scheduled one. This behavior is the main difference with real-time simulators.

As the simulation unfolds, it consumes events, but each executed event may generate new ones. As an example, the event of a packet being transmitted in a network may generate the corresponding reception event after a set propagation delay [11].

The community

Being specifically targeted for the academic world and research purposes, and 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.

2.2 NTN module

2.2.1 Channel model

The 3GPP standardization body considers different possible scenarios when describing non-terrestrial networks. A brief summary of the conditions for each scenario is briefly described in Table 2.2. Moreover, a key factor for non-terrestrial communication is whether the ground terminal is able to view the satellite line of sight (LOS) condition or not.

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. Figure 2.2 from [10] details the behavior of atmospheric absorption in the mm-Wave frequency range. The peaks at 60 and 120GHz are due to the resonance with molecular oxygen, while the peak at 180GHz and the small hump at 25GHz are due to the absorption from water vapor [10].

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 the effect of the signal being reflected and scattered 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

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

having its own attenuation and phase, since the travelled path, and therefore distance, can be different. This behavior can rapidly fluctuate, generating both constructive or destructive interference.

Other 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 [24].

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.

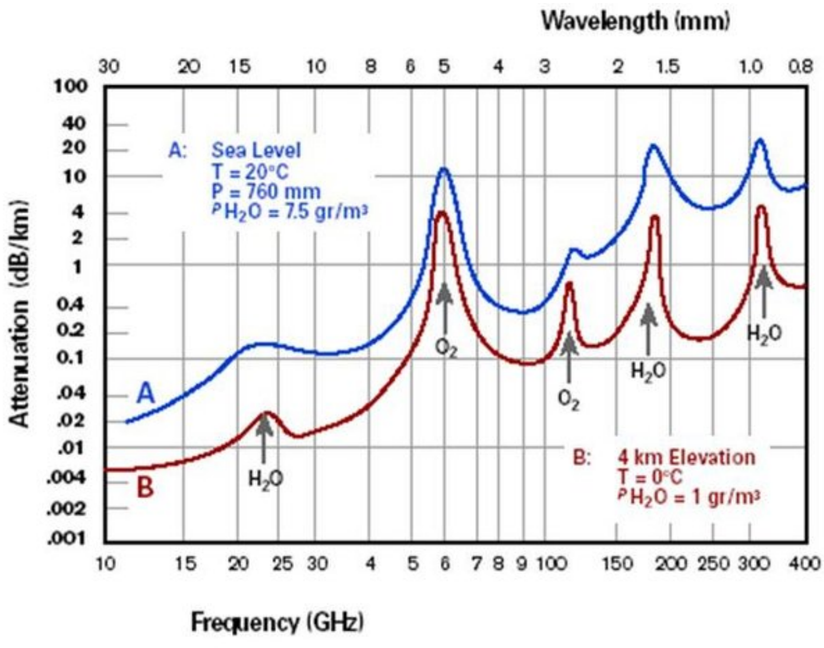

Figure 2.2: Atmospheric absorption in dB per kilometer, from [10]

2.2.2 NS-3 channel model implementation

Work on a NS-3 module to allow for a better integration and simulation of communication in non-terrestrial scenarios are already in progress, and consid-

Case	Orbit	Terminal	Band	Polarization Reuse
1	GEO	VSAT	Ka	Disabled
2	GEO	VSAT	Ka	Enabled
3	GEO	VSAT	Ka	Enabled
4	GEO	Handheld	S	Disabled
5	GEO	Handheld	S	Enabled
6	LEO-600	VSAT	Ka	Disabled
7	LEO-600	VSAT	Ka	Enabled
8	LEO-600	VSAT	Ka	Enabled
9	LEO-600	Handheld	S	Disabled
10	LEO-600	Handheld	S	Enabled
11	LEO-1200	VSAT	Ka	Disabled
12	LEO-1200	VSAT	Ka	Enabled
13	LEO-1200	VSAT	Ka	Enabled
14	LEO-1200	Handheld	S	Disabled
15	LEO-1200	Handheld	S	Enabled

Table 2.2: 3GPP scenarios for NTN

erable effort has been put in the realization of a non-terrestrial channel model[2].

The implementation of the non-terrestrial channel model required the modification and the creation of some ns-3 classes. Such work is extensively described in [31], while a brief overview is hereby reported.

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 [12]. All those three systems are depicted in Fig. 2.3.

[2]The code is available at the following repository: gitlab.com/mattiasandri

New classes

- `ThreeGppNTNScenarioChannelConditionModel`: the main task of this class is to store the channel state and condition. Four new classes were written to store the four possible scenarios described by 3GPP:
  - Dense urban,
  - urban,
  - suburban,
  - rural
- `ThreeGppNTNScenarioPropagationLossModel`: once again, 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`: mainly helps position UEs on the Earth surface in an easier way by allowing them to be input in a more natural coordinate set. Conversions amongst different systems are done using the `GeographicPositions` class described above.
- `CircularApertureAntennaModel`: a more precise implementation of the default ns-3′ parabolic antenna model, this time making use of a newer and more efficient C++ function for the computation of Bessel functions required when considering the radiation pattern of aperture antennas [37].

Use in this work

All the aforementioned implementation is based on the 3GPP specifications as detailed in the standard [8], and the resulting channel model enables full-stack end-to-end simulations considering different NTN scenarios [31].

This work can therefore benefit from an already existing standard implementation of the channel model, a crucial point for its aim of providing a simulation of how the complete NR protocol stack would behave in such a challenging scenario.

2.3 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.

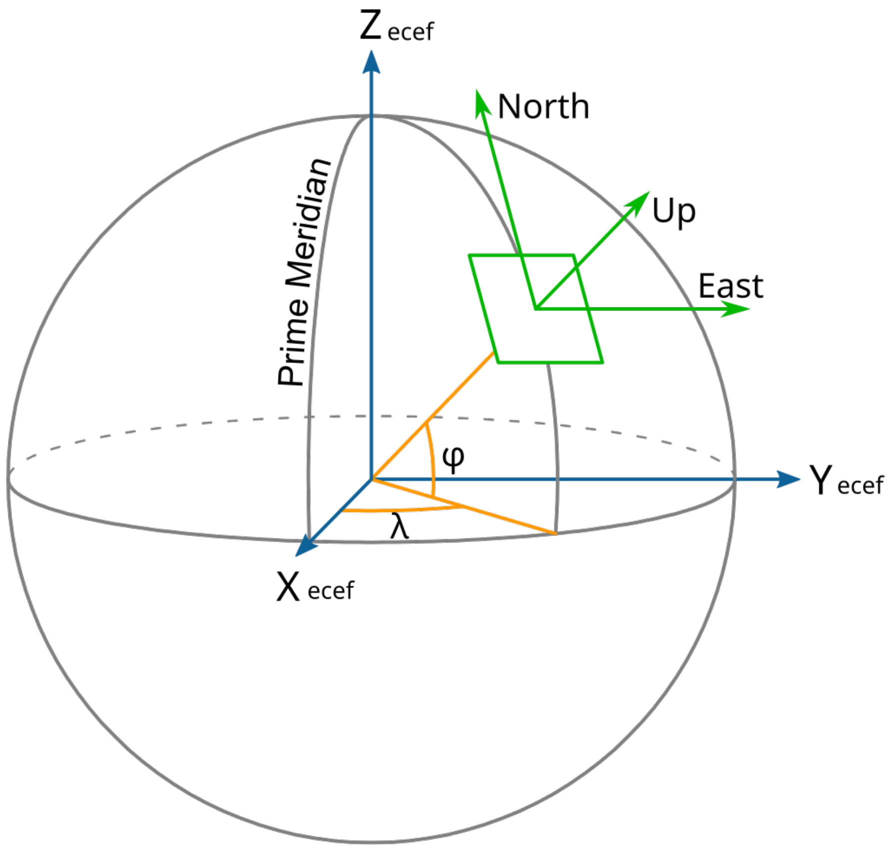

Figure 2.3: Showcase of different coordinates systems [12]

2.3.1 Network topology

The simple network setup is depicted in Fig. 2.4consisted 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 code snippet 2.1
- **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 code snippet 2.2.
- **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 it is the aim of this work to study the effects of propagation delay, on the protocol suite, the other two parameters remained constant across all the simulations. The list of parameters can be found in the code snippet 2.4.

- **g-NodeB**: the adopted approach is to incorporate the g-NodeB into the satellite payload, therefore adopting the configuration described in section 1.5.2. This decision was made since the high one-way propagation delays are enough to cause some involved protocols to start malfunctioning. Adopting the bent-pipe configuration described in 1.5.1 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 [2], and are listed in the code snippet 2.3.

- **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.

```
bool enableNagle = false;  // whether to enable Nagle's algorithm
bool enableHarq = false;   // whether to enable HARQ protocol
uint32_t numHarq = 100;    // max number of concurrent HARQ processes
uint32_t harqTimeout = 10; // timeout for HARQ processes
bool rrcIdeal = false;     // use ideal version of RRC protocol
double tcpMinRto = 200;              // minimum TCP RTO
uint32_t tcpBufSize = 131072 * 100; // TCP buffer size
double ipv4FrExpTimeout = 200;       // IPv4 fragment expiration timeout
double perr = 0.1;                   // target error probability when transmitting PHY-level packets

// Application parameters
std::string transportPrtcl = "UDP"; // Whether to use UDP or TCP
uint32_t numPackets = 2000;          // max number of packets to be sent
double appStartTimeSec = 0.5;        // application start time
double appStopTimeSec = 5.5;         // application stop time
double simStopTimeSec = 6;           // simulation stop time
uint32_t packetSizeBytes = 200;      // application packets' size
```

Code 2.1: Application and UE configuration parameters

```
// UE Parameters
double vsatAntennaGain = 39.7;        // dB
double vsatAntennaDiameter = 0.6;     // meters
```

```
    double vsatAntennaNoiseFigure = 1.2; // dB
```

Code 2.2: UE antenna parameters

```
    // Satellite parameters
    double satEIRPDensity = 40;     // dBW/MHz
    double satAntennaGain = 58.5;   // dB
    double satAntennaDiameter = 5; // meters
    double distance = -1;           // height of the satellite in km
```

Code 2.3: Satellite antenna parameters

```
    uint64_t propDelay = 6;      // propagation delay in ms
    double frequency = 20e9;     // link carrier frequency
    double bandwidth = 400e6;    // link bandwidth
```

Code 2.4: Non-terrestrial link parameters

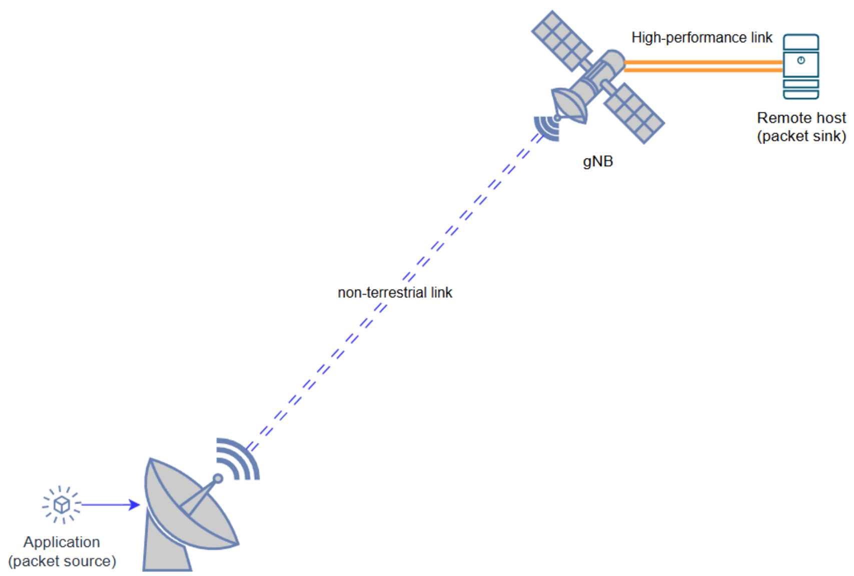

Figure 2.4: Network simulation scenario

3 Scheduling problems

Before starting to tackle the problems that the high propagation delay causes to the HARQ protocol, attention shall be put on the random access procedure and on the scheduler, both happening at a lower level, therefore necessary to be able to receive packets.

If those layers are not in working conditions, the communication will not be able to take place.

3.1 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 mobile communication networks, the scheduling is dictated by the network and the UE has to follow the provided indications.

5G channels

The NR standard comprises the usage of several channels depending on the type of data to be sent. The medium access control layer is a sort of intermediary, using channels provided by the underlying physical layer and providing the so-called logical channels to the upper layers.

A brief overview of such channels can be found in Table 3.1, listing the ones provided by the physical layer, and in Table 3.2, listing the ones provided by the media access control (MAC) layer. Figure 3.1 provides a visual clue at the separation in place between the channels at different levels of the protocol stack, while Figure 3.2 shows how channels are mapped when transitioning between levels.

Transport channel	**Acronym**
Broadcast channel	BCH
Downlink shared channel	DL-SCH
Paging channel	PCH
Uplink shared channel	UL-SCH
Random access channel	RACH
Sidelink broadcast channel	SL-BCH
Sidelink shared channel	SL-SCH

Table 3.1: Transport channels provided by the physical layer

Logical channel	**Acronym**	**Type**
Broadcast control channel	BCCH	control
Paging control channel	PCCH	control
Common control channel	CCCH	control
Dedicated control channel	DCCH	control
Dedicated traffic channel	DTCH	traffic
Sidelink broadcast control channel	SBCCH	control
Sidelink control channel	SCCH	control
Sidelink traffic channel	STCH	traffic

Table 3.2: Logical channels provided by the MAC layer

Different channels are used to convey messages such as the notification of an imminent transmission and the transmission itself, or the uplink scheduling request asked by the UE and the successive grant from the base station [15].

3.1.1 TDD OPERATION

The most complex problem to solve when dealing with propagation delay is certainly the time division duplexing (TDD) mode of operation. In this scenario, each user can communicate only inside its assigned time slots.

Differently from 4G, where a predefined pattern was in place when allocating downlink and uplink allocations in a radio frame, in New Radio 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.

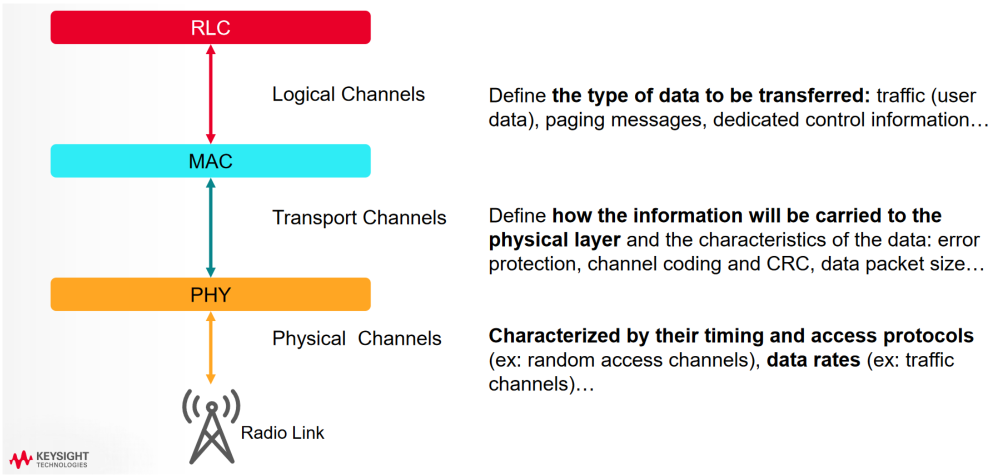

Figure 3.1: Different types of channels used in 5G, courtesy of Keysight technologies

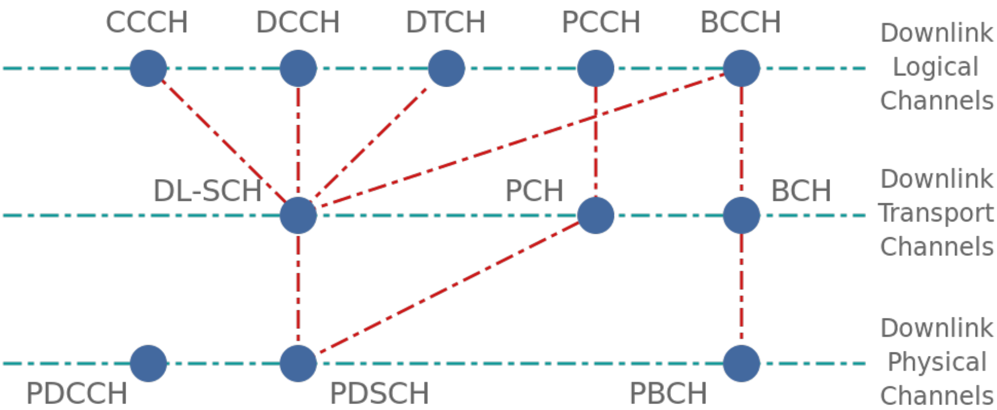

Figure 3.2: Mapping between channels at different levels, courtesy of [14].

The important key concept is that all the scheduler work has to account for the propagation delay. While in terrestrial communications guard periods of six symbols when switching from downlink communications to uplink communications is sufficient to account for the delay in a 32Km-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 [20], the same cannot be said for non-terrestrial use cases, where the distances that come into play are much longer, therefore any delay is orders of magnitude higher.

3.2 Accounting for propagation delay in scheduling

3.2.1 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. The UE sends a request for uplink resources at time t_0 since it has some data to send. In the terrestrial scheduler implementation, the gNB would receive such request at $t_0 + \tau_p = 6ms$ and, provided that other transmissions have not been scheduled yet, grant the UE the possibility to transmit in the following slot, which will start after 1ms at $t_0 + \tau_p + t_{slot} = 7ms$. However, this grant will reach the UE only after another propagation delay, therefore at $t_0 + 2\tau_p = 12ms$, when it will already be too late.

3.2.2 Proposed solution

The implemented solution assumes that the scheduler has knowledge of the propagation delay. 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 the same scenario of the previous example in section 3.2.1. 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 3.3, where a scenario with propagation delay of 5 slots (roughly 1,2ms) is considered:

- 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.

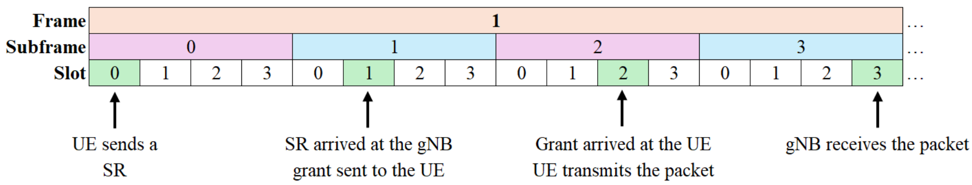

Figure 3.3: Difference between allocated slot and gNB reception

3.3 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 buffer status report timer, as followingly described.

3.3.1 Problem description

Reduced latency Figure 3.4 shows a rather peculiar trend regarding latency. The expectations were of 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 $1\tau_p$, a grant to be emitted requiring another τ_p to reach back to the UE, and finally the transmission to be received by the gNB after the third propagation delay.

The observed pattern did not match any of the expectations. The output plots obtained from the simulation campaign showed a nonlinear saw tooth behavior presenting values that consistently stayed under the expected $3\tau_p$ threshold value (Fig. 3.4).

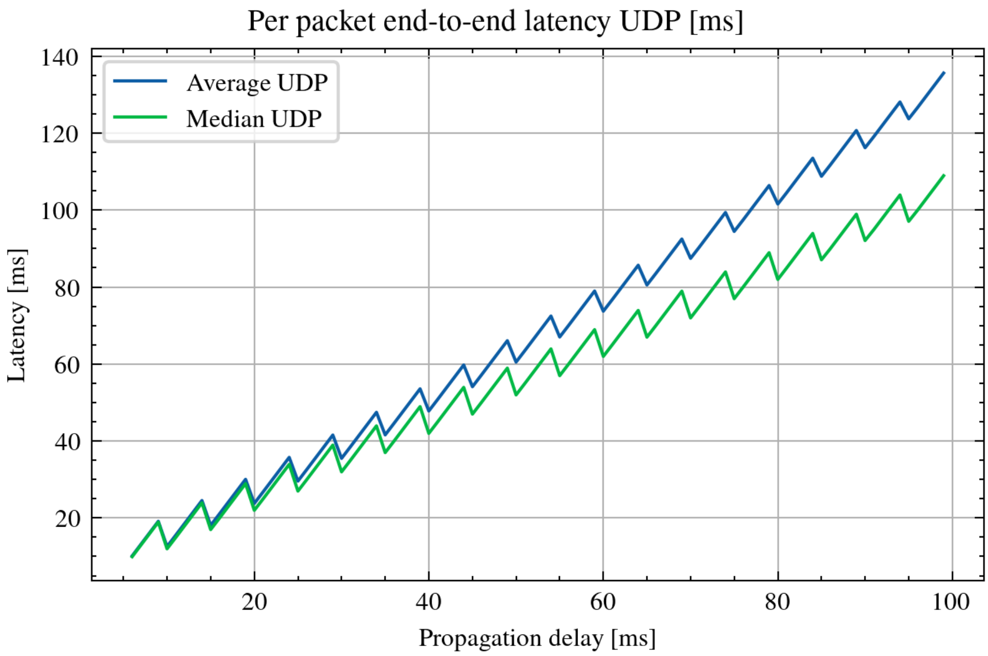

Figure 3.4: E2E latency vs. propagation delay with periodic BSR

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

This behavior happens because the implementation of 5G 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 [4], 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 buffer status reports at regular 10ms intervals even though the previous ones were not lost, but still travelling towards the base station.

Figure 3.5 can provide a visual clue to understand this behavior. The arrival of packet p1 at time zero triggers the transmission of a first SR. Since this scenario has a propagation delay of 20ms, the SR will arrive at the g-NodeB at time 20ms. However, the BSR timer expires in 10ms by default, so during a single round-trip time the SR 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 intended packet, effectively wasting three out of four transmission opportunities that could have been allocated to other users waiting to transmit.

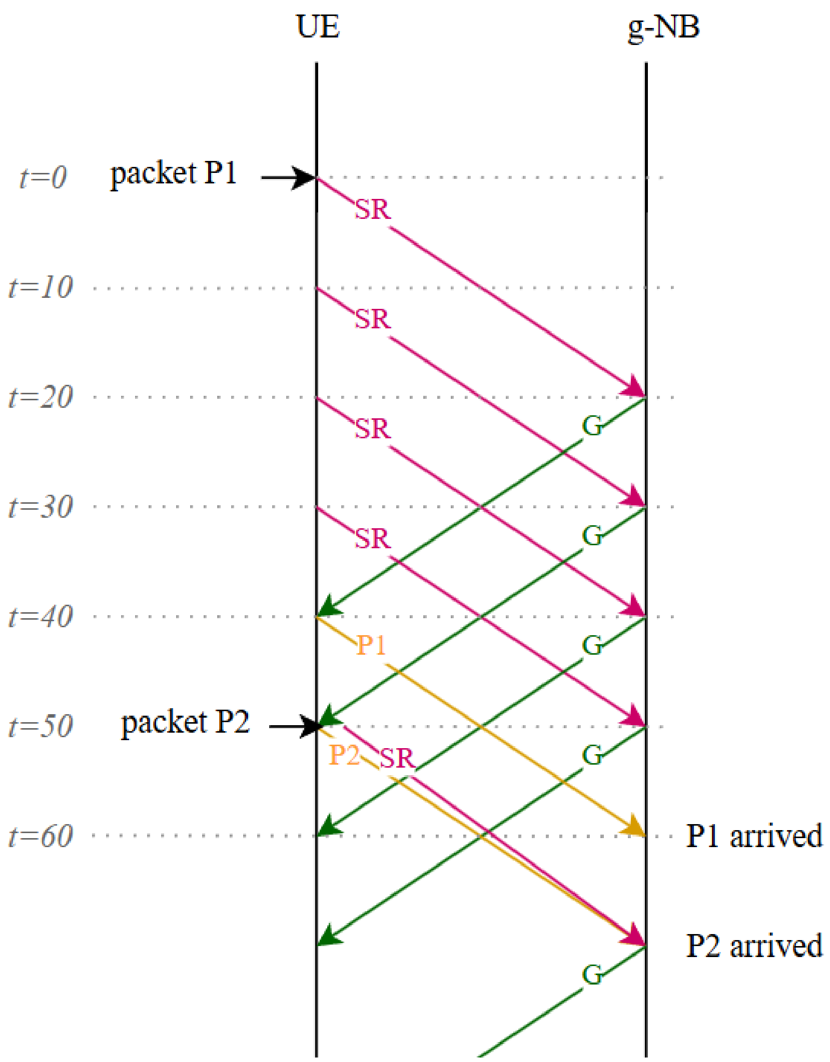

Figure 3.5: Packet diagram with periodic BSR

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 the second packet (P2) in figure 3.5. It triggers another SR, but a grant is received immediately after, therefore allowing for an immediate transmission of the second packet, which will experience a latency of only a single propagation delay. This instantaneous grant is nothing but the result of the previous SR storm originated from 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.

3.3.2 Proposed 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.

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 and 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 depending on the type of application, and send the gNB some scheduling requests for packets that have yet to be generated.

3.4 Inflated BSR

3.4.1 Problem description

Another problem occurred when the interval between the packets generated by the application, i.e. the packets interarrival times, 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 3.6, we can see that the arrival of the first packet 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 writing "SRx2". Packets P3 and P4 behave in the same way, triggering requests for the allocation of three and four additional packets respectively.

Finally, the grants begin 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 found empty.

Repercussions The described behavior caused the physical throughput to be significantly higher than the set application source rate, and it rapidly increased with the propagation delay. Figure 3.7 shows the observed physical throughput for a source rate of 160Kb/s, indicating that a lot of radio resources are being wasted to transmit few data.

3.4.2 Proposed 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 work.

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.

As mentioned before, the implementation of a smarter algorithm, capable of predicting the traffic needs of the UE could be employed to reduce the latency.

3.5 Reordering timer

3.5.1 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 20B are transmitted. The remaining 180B 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 [5], while the diagram describing its functionality has been reported in Figure 3.8. 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 expired before the second half of the packet managed to arrive.

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 behavior is to wait until either the packets arrive or the reordering timer expires [28]. It was observed that this timer also started when the PDCP layer was waiting for the arrival of the second part of a fragmented packet. However, due to the high propagation delay, the arrival of the second half was not possible 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 simulated during this work [28].

3.5.2 Proposed solution

Since the problem is linked to the expiration of a timer, and the propagation delay of NTNs is much larger than the terrestrial case, the implemented solution consisted in the increase of the default timer value to account for the additional delays.

Since the use of timers always bear 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. A much better approach would be to dynamically adapt the reordering timer accounting for the estimated propagation delay, similarly as the proposed solution for the scheduling problems in Section 3.2.2.

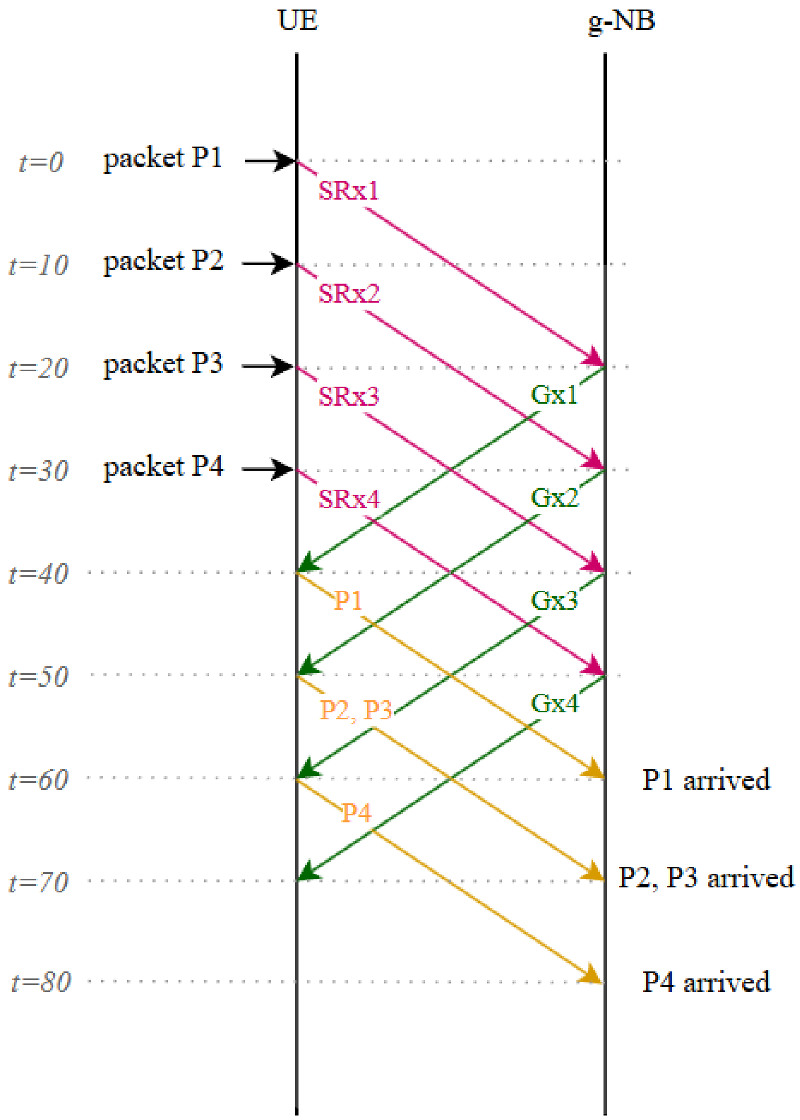

Figure 3.6: Packet diagram for interarrival times smaller than propagation delay

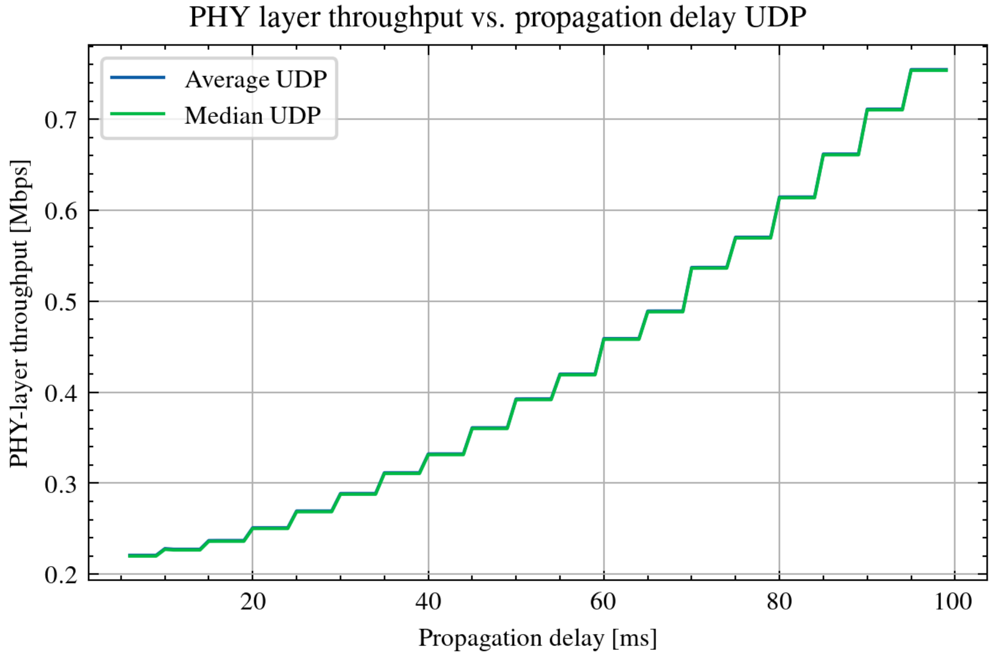

Figure 3.7: Physical throughput vs. propagation delay with periodic BSR

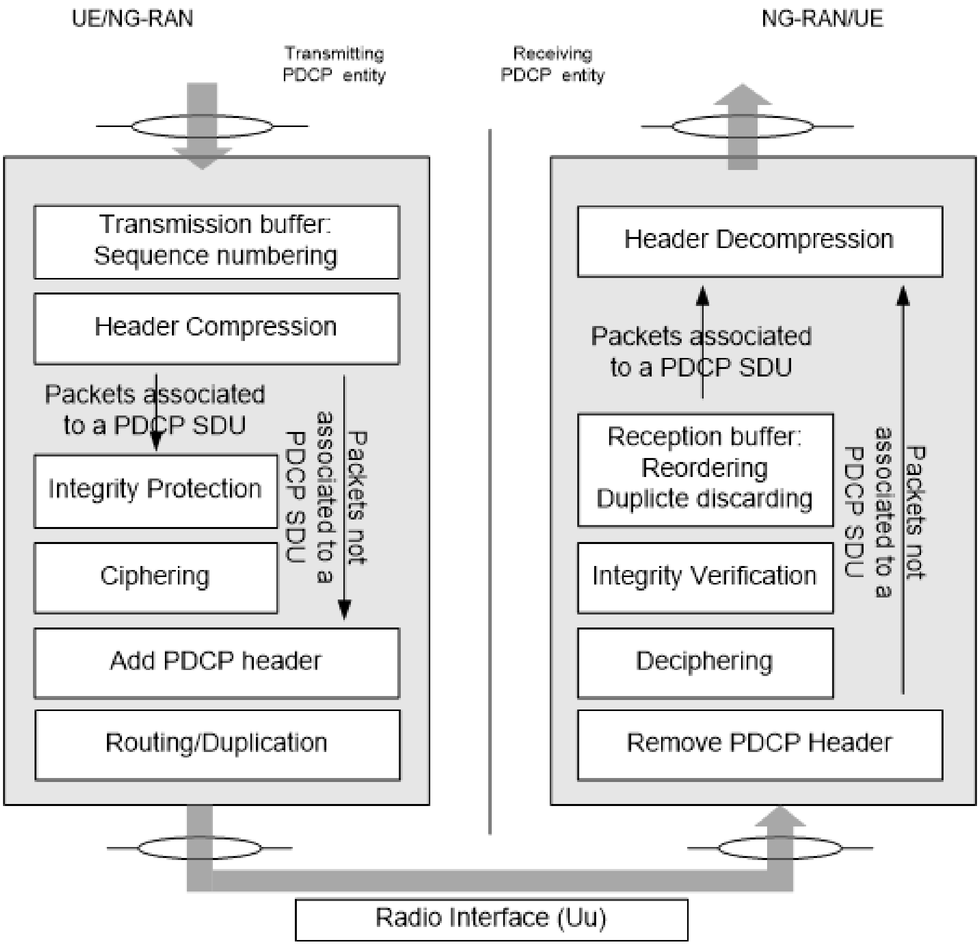

Figure 3.8: PDCP diagram [5]

4

HARQ

4.1 HARQ overview

The main objective of telecommunication technologies 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 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, lowers 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, 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 [3]. 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.

4.2 Concurrent processes limit

One of the problems highlighted by the 3GPP technical report [8] on the matter of non-terrestrial networks regards the maximum number of concurrent HARQ processes.

4.2.1 Problem description

The details of HARQ protocol implementation in the 5G NR standard is extensively treated in many publications such as [9]. However, for the purpose of understanding what is a HARQ process and how it affects the throughput in a non-terrestrial scenario, a brief overview of a few key concepts is enough.

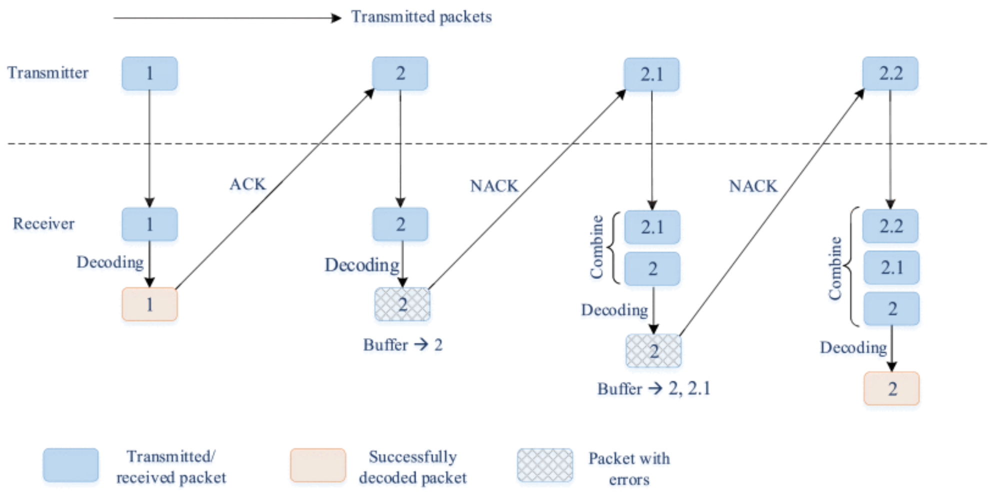

Figure 4.1: HARQ retransmission diagram [9]

HARQ working principle

Fig. 4.1 gives an overview of how HARQ processes work. Upon successful reception, depicted in the first column of Fig. 4.1, an acknowledgement (ACK) is sent back, triggering the transmission of the successive transport block (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 relayed 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 [13].

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. 4.1, where all the packets received so far contribute in correctly decoding 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 [33].

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 4.2 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 sending the new one. While this enforces the delivery of ordered packets, it also brings the downside of severely underutilizing the channel capacity, wasting resources that could potentially be used for transmission instead [6]. Figures 4.2 and 4.1 highlight this pattern.

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

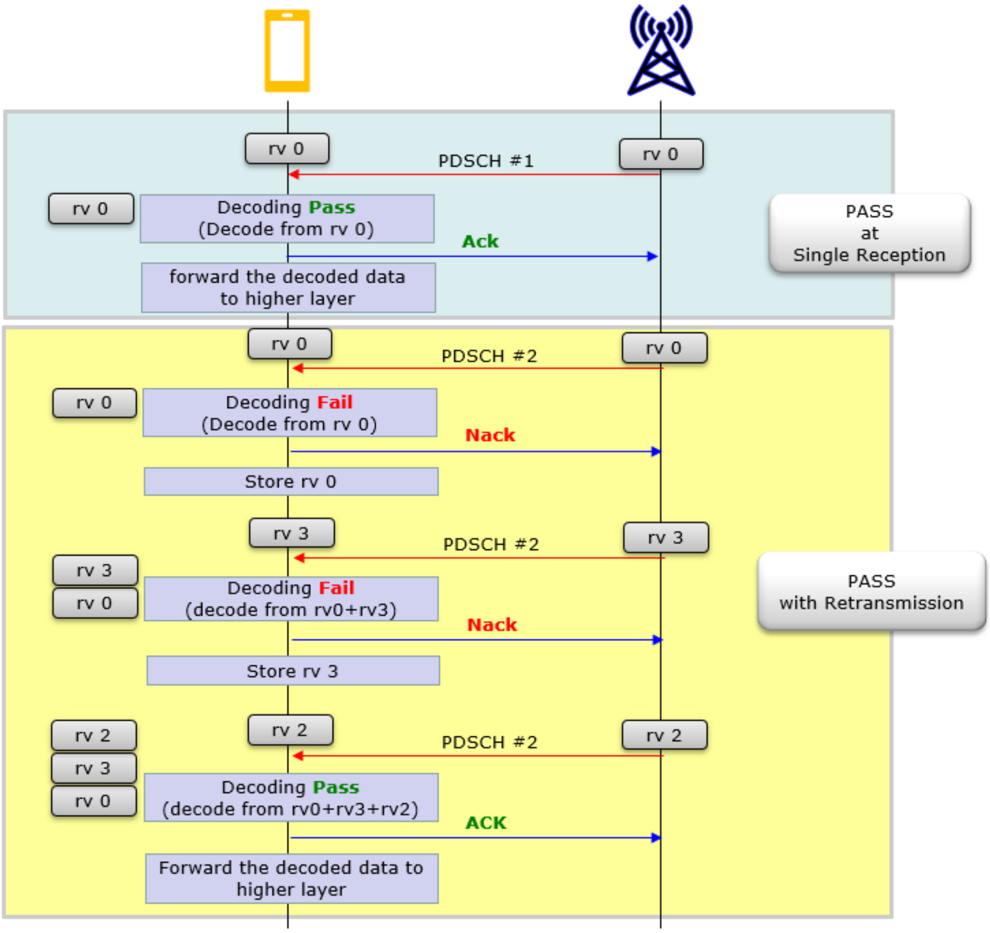

Figure 4.2: HARQ diagram with different RVs[33]

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 transfer block 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 the maximum number of concurrent HARQ processes to be assigned to each connected user, with the default value being of 8 concurrent processes and the maximum 16 [7, 13].

Application in NTNs

Since the propagation delay of NTNs is order of magnitude larger than their terrestrial counterpart, the limited maximum number of concurrent processes lowers the maximum achievable throughput, 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, to a LEO satellite orbiting at 2.000Km, therefore having $\tau_p \approx 6$ms. Assuming the best possible conditions with no need for retransmissions and assuming that the base station grants the UE to the best possible clearance of 16 concurrent 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 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 constraint becomes even more conservative in the non-terrestrial case, since retransmissions add delays in multiples of the propagation delay and can therefore quickly become a lot more costly [32].

4.2.2 Possible solutions

Increasing the processes The easiest solution would be to increase the number of maximum concurrent HARQ processes. 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 equipments such as smartphones. Each process also requires the presence of a buffer on both the receiver side and the sender side, so additional resources are required at the gNB, too.

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. Since there already are multiple concurrent processes, each ACK packet must already contain a field specifying the number of process it belongs to, and the information identifying the specific packet to be acknowledged within a process could be encoded inside this field.

Disable HARQ Lastly, the option of disabling HARQ completely and rely solely on ARQ retransmissions has been proposed by 3GPP itself[26]. This, however, would come with a performance penalty since satellite links typically suffer from more severe conditions than terrestrial ones, and [34] demonstrated that a version of HARQ specifically designed for non-terrestrial networks would be beneficial.

4.2.3 Implemented solution - More processes

Testing current implementation

To test the practical effect that the concurrent processes' limitation have 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.

Comparing the results confirmed that employing HARQ caused a noticeable negative impact on the throughput, as can be appreciated in Fig. 4.3. 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 starts to drop as the source rate crosses the 4Mb/s threshold, to then settle just below 2Mb/s. The uneven behavior is caused by HARQ struggling 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. For a high enough source rate, 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.

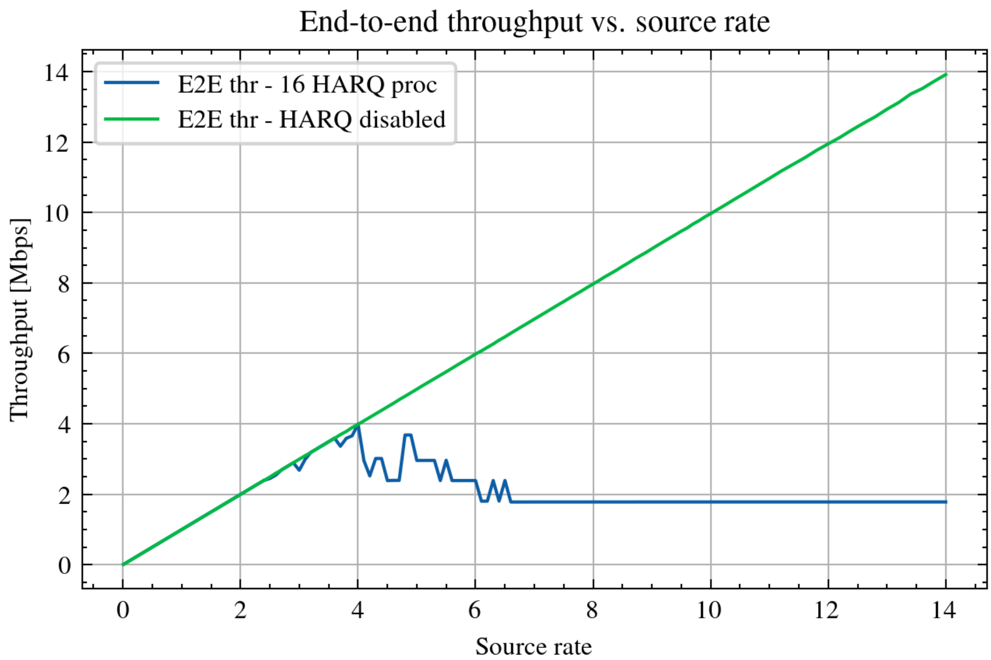

Figure 4.3: End-to-end throughput comparison with and without HARQ, τ_p = 6ms.

Having verified that HARQ does indeed limit the achievable throughput, a modification was made to the protocol suite to manually force a higher number of concurrent processes. The simulation campaign was then re-run and the results are shown in Fig. 4.4

Results

The maximum achievable throughput is now just over 14Mb/s in the scenario with the highest allowance of per-user concurrent processes. Going past such threshold causes packets to start being dropped, and the throughput to saturate, since there are no more available processes.

From figure 4.4 it is clear that increasing the parameter regulating the number of HARQ processes allows for higher speeds. Each line in the plot was obtained

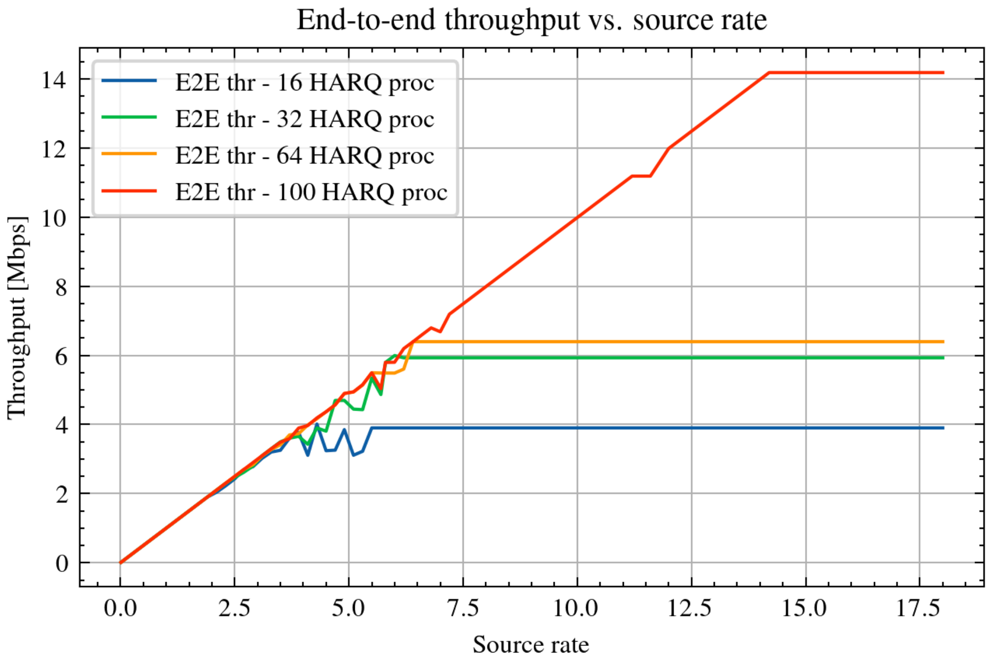

Figure 4.4: End-to-end throughput comparison with different number of concurrent HARQ processes, $\tau_p = 6$ms.

with a different number of allowed processes, starting from 16, the maximum value already in use 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.

This solution, however, also presents some drawbacks, such as the fluctuations when the maximum capacity starts to be approached as the source rate increases. Furthermore, constraints regarding the required computational power and the increased energy consumption shall also be taken into careful consideration, since raising the number of concurrent processes from 16 to 100 is a big modification, bringing a consistent increase in both 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 thought, and future studies should also consider alternative proposals.

4.2.4 Implemented solution - Disable HARQ

Figure 4.3 could suggest that completely disable the HARQ protocol could be a viable solution, since the end to end throughput always matched the source rate. However, the reason is the high SNR that was forced in the simulated scenario by the use of high gain antennas, which use would not be realistic in handheld devices. This approach was undertaken because this work is focused on the study of the propagation delay, and high SNR allowed for fewer variables to impact the simulations' results.

A new simulation campaign has therefore been 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 obtained Figure XXX better depicts the whole picture: in conditions of low SNR, that 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.

Simulations have shown that fully disable HARQ may be a good solution only when both UE and gNB experience optimal channel conditions with high SNR, while poor channel conditions may benefit from having HARQ enabled.

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

Conclusions and Future Works

A	**B**
C	D
E	F
G	H

Table 5.1: Table example

References

[1] 3GPP. *TR 21.917 "Release 17 Description; Summary of Rel-17 Work Items" v17.0.1*. Tech. rep. January 2023.

[2] 3GPP. *TR 38.821 "Solutions for NR to support Non-Terrestrial Networks (NTN)" v16.2.0*. Tech. rep. April 2023.

[3] 3gpp. *3GPP 38 Series of Specifications: 5G New Radio*. Tech. rep.

[4] 3gpp. *5G; NR; Medium Access Control (MAC) protocol specification (3GPP TS 38.321 version 18.1.0 Release 18)*. Tech. rep. May 2024.

[5] 3gpp. *5G; NR; Packet Data Convergence Protocol (PDCP) specification (3GPP TS 38.323 version 18.1.0 Release 18)*. Tech. rep. May 2024.

[6] 3gpp. *TR 38.214 "5G NR Physical layer procedures for data" v16.5.0, Release 16*. Tech. rep. April 2021.

[7] 3gpp. *TR 38.300 "5G NR Overall description Stage-2" v15.9.0, Release 15*. Tech. rep. April 2021.

[8] 3gpp. *TR 38.811 "Study on New Radio (NR) to support non-terrestrial networks" v15.4.0*. Tech. rep. October 2020.

[9] Ashfaq Ahmed et al. "Hybrid Automatic Repeat Request (HARQ) in Wireless Communications Systems and Standards: A Contemporary Survey". In: *IEEE Communications Surveys and Tutorials* 23.4 (2021), pp. 2711–2752. DOI: `10.1109/COMST.2021.3094401`.

[10] Abdussalam Ammar et al. "The E-Band as Future Candidate for Next Generation Networks". In: Dec. 2018.

[11] Lelio Campanile et al. "Computer Network Simulation with ns-3: A Systematic Literature Review". In: *Electronics* 9.2 (2020). ISSN: 2079-9292. DOI: `10.3390/electronics9020272`. URL: `https://www.mdpi.com/2079-9292/9/2/272`.

[12] Wikipedia contributors. *Local tangent plane coordinates - Wikipedia, The Free Encyclopedia*. [Online; accessed 26-July-2024]. 2023. URL: https://en.wikipedia.org/w/index.php?title=Local_tangent_plane_coordinates&oldid=1177505970.

[13] Devopedia. *"5G NR Hybrid ARQ"*. devopedia.org. Version 6, December 19, Accessed 2024-07-14. 2021.

[14] Devopedia. *5G NR Channels*. devopedia.org. Accessed 2024-07-26. April 2021.

[15] Devopedia. *5G NR MAC*. devopedia.org. Accessed 2024-07-21. February 2021.

[16] Bruce 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.

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

[18] Marco Giordani and Michele Zorzi. "Non-Terrestrial Networks in the 6G Era: Challenges and Opportunities". In: *IEEE Network* 35.2 (2021), pp. 244–251. DOI: 10.1109/MNET.011.2000493.

[19] Marco Giordani and Michele 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.

[20] GSMA. *5G TDD Synchronisation Guidelines and Recommendations for the Coexistence of TDD Networks in the 3.5 GHz Range*. Tech. rep. Accessed 2024-07-21. April 2020.

[21] Marko Höyhtyä et al. "Sustainable Satellite Communications in the 6G Era: A European View for Multi-Layer Systems and Space Safety". In: *IEEE Access* PP (Jan. 2022). Licensed under CC-BY 4.0 public license, pp. 1–1. DOI: 10.1109/ACCESS.2022.3206862.

[22] Helka-Liina Maattanen 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.

[23] Andrew May. *Low Earth orbit: Definition, theory and facts*. rohde-schwarz.com. Accessed 2024-07-20.

[24] David D. Morabito. "A Comparison of Estimates of Atmospheric Effects on Signal Propagation Using ITU Models: Initial Study Results". In: (November 15, 2014). URL: https://ipnpr.jpl.nasa.gov/progress_report/42-199/199D.pdf.

[25] Iestyn Morgan-Jones and Pavel Loskot. *Regional Coverage Analysis of LEO Satellites with Kepler Orbits*. 2019. arXiv: 1910.10704 [physics.space-ph]. URL: https://arxiv.org/abs/1910.10704.

[26] Amitav Mukherjee. "Hybrid ARQ Schemes". In: 2020. URL: https://api.semanticscholar.org/CorpusID:219510991.

[27] *ns-3 official website*. nsnam.org. Accessed 2024-07-22.

[28] Surendra Pandey, Raja Moses Manoj Kumar Eda, and Debabrata 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.

[29] Rohde-Schwarz. *Connecting the world with 5G NTN*. rohde-schwarz.com. Accessed 2024-07-20.

[30] Malik Muhammad Saad et al. "Non-Terrestrial Networks: An Overview of 3GPP Release 17 & 18". In: *IEEE Internet of Things Magazine* 7.1 (2024), pp. 20–26. DOI: 10.1109/IOTM.001.2300154.

[31] Mattia Sandri et al. "Implementation of a Channel Model for Non-Terrestrial Networks in ns-3". In: *Proceedings of the 2023 Workshop on ns-3*. WNS3 2023. ACM, June 2023. DOI: 10.1145/3592149.3592158. URL: http://dx.doi.org/10.1145/3592149.3592158.

[32] ShareTechNote. *4G/LTE - PHY processing*. sharetechnote.com. Accessed 2024-07-14.

[33] ShareTechNote. *5G/NR - HARQ*. sharetechnote.com. Accessed 2024-07-14.

[34] Alessandro Vanelli-Coralli et al. "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.

[35] Dengke Wang et al. "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.

[36] Dengke Wang et al. "The Potential of Multilayered Hierarchical Nonterrestrial Networks for 6G: A Comparative Analysis Among Networking Architectures". In: *IEEE Vehicular Technology Magazine* 16.3 (2021), pp. 99–107. DOI: 10.1109/MVT.2021.3085168.

[37] Dr J. L. Whitrow. "The Radiation Patterns of Circular Apertures". In: *National Security and ISR Division, Defence Science and Technology Group* (2018).

[38] Elias Yaacoub and Mohamed-Slim Alouini. "A Key 6G Challenge and Opportunity—Connecting the Base of the Pyramid: A Survey on Rural Connectivity". In: *Proceedings of the IEEE* 108.4 (2020), pp. 533–582. DOI: 10.1109/JPROC.2020.2976703.

Acknowledgments