D5.1: Open Data Sets for AI Data Driven Networking

Revision: v1.0

Work package	WP 5
Task	Task 5.1
Due date	31/01/2024
Submission date	31/01/2024
Deliverable lead	HSP
Version	V1.0
Authors	Enrique Marin (HSP), Nuria Trujillo (HSP), Marius Corici (FHG), Benjamin Barth (DLR)
Reviewers	Justin Tallon (SRS)
Abstract	The deliverable reports the datasets planned to be used in the course of WP5 and the related applications/scenarios for which their use is considered relevant.
Keywords	Database, satellite, AI, Networking, Simulator, Emulator

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	18/01/2024	First draft "Satellite Network Dataset" section	Enrique Marín (HSP) & Nuria Trujillo (HSP)
V0.2	31/01/2024	Consolidated section "Introduction" and "Large Scale Satellite System	Marius Corici (FhG) & Benjamin Barth (DLR)

WWW.5G-STARDUST.EU

Co-funded by the European Union 6GSNS

Grant Agreement No.: 101096573
Call: HORIZON-JU-SNS-2022

Topic: HORIZON-JU-SNS-2022-STREAM-A-01-02
Type of action: HORIZON-JU-RIA

		Simulation/Emulation"	
V0.3	02/02/2024	QA review from SRS	Justin Tallon (SRS)
V1.0D	02/02/2024	Draft version implementing the changes proposed in the QA version	Enrique Marín (HSP) & Nuria Trujillo (HSP)
V1.0.F	02/02/2024	Final approved version for the EC portal	Tomaso de Cola (DLR)

DISCLAIMER

Co-funded by the European Union 6G SNS

5G-STARDUST (*Satellite and Terrestrial Access for Distributed, Ubiquitous, and Smart Telecommunications*) project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096573. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

COPYRIGHT NOTICE

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	**R**	
Dissemination Level		
PU	*Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)*	✓
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

EXECUTIVE SUMMARY

This report has been prepared in the framework of WP5 (Networking Operations) and specifically within the scope of Task 5.1 “Multi-Connectivity Solutions” and Task 5.2 “AI Data Driven Network Management and Slicing”. Deliverable D5.1 describes the data sets that will contribute to train ML-based algorithms with close to real operational data from a satellite mission.

The data sets have been divided into two parts, i.e., real GSO satellite network dataset and large-scale satellite system simulation/emulation dataset. The document details the features and the format of the data collected in both networks. The data set detailed in Section 2 is obtained from a satellite network operated by HISPASAT that targets America and Europe. The inputs provided by the satellite operator are related to broadband services for fixed clients (residential broadband) and for mobile communications (maritime communications, in this case). The dataset detailed in Section 3 is obtained from the large-scale simulator/emulator developed by Fraunhofer FOKUS Institute in European Space Agency (ESA) funded projects.

This deliverable provides a brief overview of how the data can be applied within work package 5; the algorithms will be presented within the technical deliverables. Therefore, this deliverable will support the development of network-management related optimizations based on the AI knowledge of the following work package tasks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

5G-STARDUST	5G-Satellite and Terrestrial Access for Distributed, Ubiquitous, and Smart Telecommunications
AI/ML	Artificial Intelligence/Machine Learning
AMZ-5	Amazonas-5 Satellite
CN0	Carrier-to-noise density ratio
CSV	Comma-Separated Values
GEO	Geosynchronous Earth Orbit
H30W-6	Hispasat 30W-6 Satellite
KML	Keyhole Markup Language
KMZ	Keyhole Markup Language Zipped
NTN	Non-Terrestrial Networks
PDF	Portable Document Format
SINR	Signal-to-Interference-Plus-Noise Ratio
VBA	Visual Basic for Applications
WP	Work Package

1 INTRODUCTION

1.1 SCOPE

In this deliverable, the data sets which are used as a basis for artificial intelligence (AI) and machine learning (ML) algorithms are presented. The main focus of AI/ML within work package 5 is on the networking aspects, i.e. the higher layers of the communication system, and how it can be used to improve the network performance. Traffic flows and service usage is of high interest as input for this but also user location within the network to predict network usage and flows to optimize the overall performance of the system. Main task processing this data will be T5.2 AI-driven network management and slicing, but it is generally of high value for the overall work package and can also be considered for other tasks, e.g. in T5.1 about multi-connectivity. The described data will be publicly provided in the ongoing course of the project.

1.2 DOCUMENT DISTRIBUTION

The data has been gathered from two satellites operated by HISPASAT, which are oriented toward America and Europe and large-scale satellite system simulation/emulation provided by Fraunhofer. More detailed information is included in the following sections, which are organized as follows:

Section 2 for the Satellite network dataset includes all information related to the radio parameters for the satellite inputs studied in the 5G-STARDUST project. This section is divided into four different sub-sections.

- Information about the service provided: an overview on the services selected as inputs for the development of the 5G-STARDUST AI Data-Driven Network Management and Slicing.

- Geographical information: general information about the satellites, oriented to America and Europe, analyzed to extract the inputs for the project.

- Format, Data type and time granularity for the broadband traffic: including all the information corresponds for the analysis of the inputs provided by Hispasat as satellite operator.

- Traffic type: a section oriented for the traffic inputs provided by Hispasat, general overview about the data files and information provided and finally, the dataset consolidation.

- Purpose of this dataset: contains an overview of the next steps for the development of the AI Data Driven Network Management and Slicing and how these inputs, in coordination with WP4, are useful for its implementation.

Section 3 for the large-scale Satellite System Simulation Emulation dataset provides an overview of the large-scale simulator/emulator developed by Fraunhofer FOKUS Institute in European Space Agency (ESA) funded projects. This innovative solution is pivotal in generating a comprehensive dataset, indispensable for diverse machine learning applications in the satellite field.

- Motivation including an overview why this simulator/emulator is useful for the project

- Large Scale Satellite Simulator/Emulator environment: Including an overview about the abstract calculations, the simulator, and the emulator.
- Extracted Data types: Including the traffic information provided by the simulator/emulator and the data format type.
- Purpose of this dataset: contains an overview of the next steps for the development of the AI-Data Driven Network Management and Slicing and possibly also to other tasks of WP5 and how these inputs, in coordination with WP4, are useful for its implementation.

2 SATELLITE NETWORK DATASET (HSP)

2.1 INFORMATION ABOUT THE SERVICE PROVIDED

The services selected for the elaboration of the satellite network dataset are aligned with the selection of scenarios for the proof-of-concept carried out during 5G-STARDUST WP2. For this purpose, and although the 5G-NTN technology is still in its definition and development, it was defined that the inputs provided by the satellite operator, Hispasat in this case, would be related to the real satellite broadband services for fixed clients (residential broadband) and for mobile communications (maritime communications, in this case).

In order to create a robust and interesting dataset for the project, the satellite operator has defined anonymous clients with a huge number of end users for its elaboration. In this case, the number of clients for residential broadband scenario is more than 500 connected users and for the maritime communications is around 60 ships connected. More detailed information is included in the proposed dataset.

The following sections define the geographical information about both indicated services, the satellites used with general information, format, type of data and temporal granularity of the inputs exported from the satellite operator's internal platforms and the type of traffic provided.

2.2 GEOGRAPHICAL INFORMATION

In order to provide real information on satellite traffic, we have selected satellites according to the services they offer. Thus, residential broadband and mobility services are not offered by the same satellite or in the same geographical area.

However, both services are transversal in that the inputs come from real values of GEO satellites operated by Hispasat.

In the tables below lists the most relevant characteristics of the satellites that have been selected to extract the inputs for the 5G-STARDUST dataset.

2.2.1 Residential Broadband

The residential broadband scenario comes from real fixed broadband traffic provided in Mexico by the Amazonas 5 Hispasat Satellite [1] .The following Table 1 summarize the general characteristics for the above-mentioned satellite.

Table 1: Amazonas 5 general characteristics [1]

Information	Value
Satellite Name	Amazonas 5
Orbital Position	61º West

Launch Date	2017
Transponders	24 Ku Band + 34 Ka band spots
Bandwidth	36 MHz in Ku band + 225 MHz in Ka Band* *Reference value, could vary between different spots
Antennas Number	5
Satellite Type	Transparent (no onboard processing)
Payload Power	10 kW
Available Power	11,5 kW

The AMZ-5 satellite footprint is oriented to America with linear polarization. In this example, the operational transponders that have selected for the datasets are related to Ka spots. The figure below shows the general Ka and Ku beams footprint.

Figure 1: AMZ-5 general America Footprint (Ka (left picture) and Ku (right picture) beams) [1]

2.2.2 Mobility Communications

The mobility scenario comes from real broadband traffic in Europe by the H30W-6 Hispasat Satellite. The following Table 2 summarize the general characteristics for the above-mentioned satellite.

Table 2: H30W-6 general characteristics [2]

Information	Value
Satellite Name	Hispasat 30W-6
Orbital Position	30º West
Launch Date	2018
Transponders	40 Ku Band + 7 Ka band + 10C Band spots

Bandwidth	36 MHz-72MHz in Ku and C band
Antennas Number	5
Satellite Type	Transparent (no onboard processing)
Payload Power	10 kW
Available Power	11,5 kW

The satellite footprint its oriented toward Europe and America with linear polarization. In this case, the operational transponders that we have selected for the datasets are related to Ku spots.

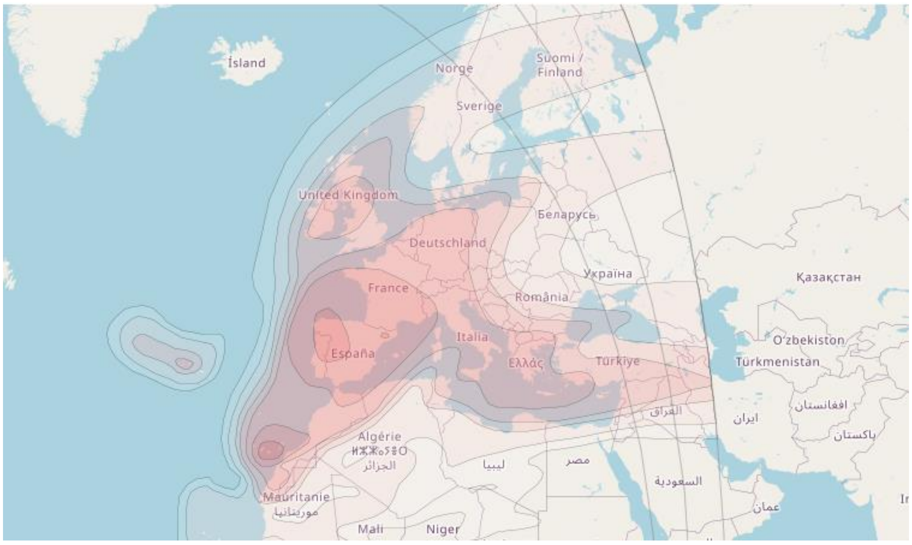


Figure 2: H30W-6 general Europe Footprint (Ku beams) [2]

Hispasat offers broadband traffic services for maritime communications through the above-mentioned satellite. Since the ships are in motion, and Hispasat only deals with the maintenance of the satellite, its operation and commissioning, but not with its commercialization, we cannot estimate in real time the exact position of the ships.

2.3 FORMAT, DATA TYPE AND TIME GRANULARITY FOR THE BROADBAND TRAFFIC

The inputs extracted from Hispasat's internal tools are in Comma-Separated Values (CSV) format. These inputs are then processed and merged using Visual Basic for Applications (VBA) in Excel, creating tables with identical formatting between the different services and satellites. Finally, the 5G-STARDUST satellite network dataset format is Microsoft Excel macro-enabled worksheet (.xlsm).

The data type involves all the information related to the traffic pattern. The traffic has been extracted per beam and per service in Megabits per second (Mbps). Moreover, for the Residential broadband all beams traffic pattern, the protocols (YouTube traffic, Google traffic, Microsoft Teams, etc) have been also included in the 5G-STARDUST dataset, where the 10 most used protocols are quantified. However, it is important to note that the amount of traffic per protocol, including only the 10 most used protocols, does not necessarily correspond to the total traffic pattern.

The time granularity depends on the sample chosen if it is weekly or monthly. For the weekly sample, the time granularity is 5 minutes (2016 samples) and for the monthly sample the time granularity is 1 hour (745 samples).

Table 3: Data Traffic format, type and time granularity

Data Traffic	Value
Extracted Format	CSV
Dataset Format	Excel
Data Type	Traffic pattern (per beam and per whole service)
Time Granularity	Weekly: 5 min Monthly: 1 hour

2.4 TRAFFIC TYPE

As defined above, the broadband traffic included in the dataset is related to the traffic pattern for each service, residential broadband, and maritime communications. Although both are related to GEO Hispasat satellites, depending on the satellite, the customer and the service, the exported file may vary significantly. In this case, the inputs extracted from Hispasat's internal tools refer to the traffic observed at the satellite reception, after transporting the traffic from the end user to Hispasat's satellite communications platform. This metric allows the satellite operator to analyse the current state of the non-terrestrial networks.

2.4.1 Residential Broadband traffic type

As has been described before, Hispasat provides residential broadband in Mexico with Amazonas 5 satellite. In this case, Amazonas-5 Ka band footprint covers all the Mexico area with the beams 34,36,37,38,39,40 [1]. The end users are provisioned for each beam according to the position, available capacity per beam and signal quality following the requirements specified in the customer agreement.

These are the inputs extracted and provided by Hispasat relative to traffic information in comma-separated values (CSV) format:

- 2023.10 Traffic Beam34 1 Month Mx.csv
- 2023.10 Traffic Beam34 1 Week Mx.csv
- 2023.10 Traffic Beam36 1 Month Mx.csv
- 2023.10 Traffic Beam36 1 Week Mx.csv
- 2023.10 Traffic Beam37 1 Month Mx.csv
- 2023.10 Traffic Beam37 1 Week Mx.csv
- 2023.10 Traffic Beam38 1 Month Mx.csv
- 2023.10 Traffic Beam38 1 Week Mx.csv
- 2023.10 Traffic Beam39 1 Month Mx.csv
- 2023.10 Traffic Beam39 1 Week Mx.csv
- 2023.10 Traffic Beam40 1 Month Mx.csv
- 2023.10 Traffic Beam40 1 Week Mx.csv

The following inputs extracted and provided by Hispasat are related to the traffic pattern and protocol consumption involving all the beams indicated above. These inputs provide a global view of the residential broadband service provided by the Amazonas-5 satellite in Mexico. The inputs are provided in comma separated value (CSV) format.

- 2023.10 Traffic Pattern 1 Week Mx.csv
- 2023.10 Traffic Pattern 1 Month Mx.csv
- 2023.10 Protocol Consumption 1 Week Mx.csv
- 2023.10 Protocol Consumption 1 Month Mx.csv

Finally, the following inputs extracted and provided by Hispasat refer to an example of traffic consumed by a customer in the selected periods. This allows the 5G-STARDUST consortium to approximate the average traffic consumed by each active customer. The inputs are provided in comma-separated values (CSV) format.

- 2023.10 Traffic Example 1 Terminal 1 Month Mx.csv

- 2023.10 Traffic Example 1 Terminal 1 Week Mx.csv

These are the inputs extracted and provided by satellite operator relative to real SINR & CN0 for a clients located in Beam 36, that will be compared with theoretical values included in the link budget, in comma-separated values (CSV) format:

- 2023.10 CNo Terminal Edge Beam 36.csv
- 2023.10 CNo Terminal Centre Beam 36.csv
- 2023.10 SNR Terminal Edge Beam 36.csv
- 2023.10 SNR Terminal Centre Beam 36.csv

2.4.2 Mobility Communications traffic type

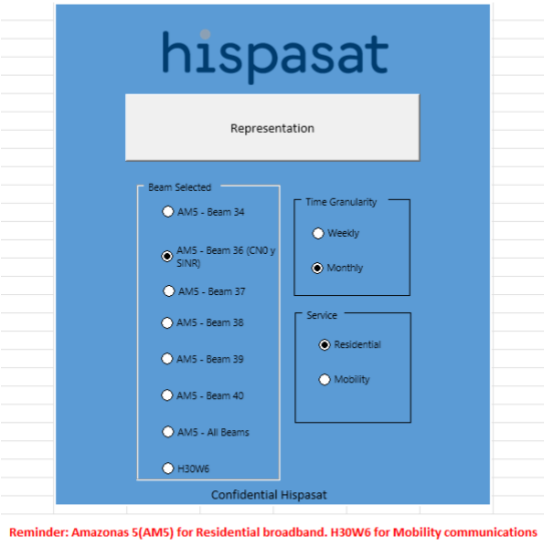
For the mobility communications, Maritime in this case, Hispasat tools are not the same as for residential broadband because the platform, the satellite used, and the final client are different. Because of that, the traffic pattern is exported for a particular client, which may be provisioned by H30W-6 several beams [2], and the SNR information from a particular end-user, that will be compared with theoretical values included in the link budget.

These are the inputs provided by Hispasat relative to maritime communications, in comma-separated values (CSV) format:

- 2023.10 Europe Maritime Traffic
- SNR 1 Terminal Example

2.4.3 Dataset consolidation

In order to harmonize the dataset proposed by the satellite operator to the 5G-STARDUST consortium, Hispasat has prepared a Microsoft Excel macro-enabled worksheet (.xlsm) as indicated in the section 2.4, where the information is processed and represented in the formats previously confirmed with the project consortium. This tool has been developed in order to present a more user-friendly dataset, and to assist in the process of developing the AI/ML and RRM 5G-STARDUST technologies. The figure below illustrates the methodology for this Microsoft Excel macro-enabled worksheet where the consortium could choose the service, time granularity and satellite beam, and the worksheet prints the choices in Excel table format, with additional information and some graphics if relevant. The development of this book is to simplify the usage of the inputs provided to the 5G-STARDUST consortium.



Fecha	Return	Forward
10/01/2023 0:00	3.073 Mbps	14.450 Mbps
10/01/2023 1:00	3.224 Mbps	8.490 Mbps
10/01/2023 2:00	4.239 Mbps	5.325 Mbps
10/01/2023 3:00	3.199 Mbps	4.151 Mbps
10/01/2023 4:00	3.646 Mbps	5.840 Mbps
10/01/2023 5:00	3.466 Mbps	5.610 Mbps
10/01/2023 6:00	3.637 Mbps	8.086 Mbps
10/01/2023 7:00	5.189 Mbps	12.006 Mbps
10/01/2023 8:00	3.794 Mbps	17.240 Mbps
10/01/2023 9:00	3.571 Mbps	18.556 Mbps
10/01/2023 10:00	2.334 Mbps	18.795 Mbps
10/01/2023 11:00	2.250 Mbps	17.261 Mbps
10/01/2023 12:00	2.478 Mbps	18.135 Mbps
10/01/2023 13:00	2.166 Mbps	15.272 Mbps
10/01/2023 14:00	2.411 Mbps	15.253 Mbps
10/01/2023 15:00	2.373 Mbps	16.285 Mbps
10/01/2023 16:00	1.334 Mbps	18.856 Mbps
10/01/2023 17:00	1.827 Mbps	17.406 Mbps
10/01/2023 18:00	3.554 Mbps	17.953 Mbps
10/01/2023 19:00	1.390 Mbps	20.604 Mbps
10/01/2023 20:00	1.853 Mbps	20.797 Mbps
10/01/2023 21:00	1.397 Mbps	15.914 Mbps
10/01/2023 22:00	0.951 Mbps	14.952 Mbps
10/01/2023 23:00	0.674 Mbps	11.601 Mbps
10/02/2023 0:00	0.294 Mbps	7.725 Mbps
10/02/2023 1:00	0.329 Mbps	5.650 Mbps
10/02/2023 2:00	1.434 Mbps	3.313 Mbps
10/02/2023 3:00	0.345 Mbps	3.419 Mbps
10/02/2023 4:00	1.039 Mbps	6.199 Mbps

Average CN0 in the Beam center	74,75
Average CN0 in the Edge of Coverage	69,75
Average SINR in the Beam center	11,12737944
Average SINR in the Edge of Coverage	9,520941208

Figure 3: Hispasat Traffic Dataset

2.5 PURPOSE OF THIS DATASET

The dataset provided by HISPASAT enables the modelling of different networking aspects. In T5.2, the goal will be to develop AI/ML solutions capable of managing different service types with different requirements in order to deploy Network Slices in an optimized way. To this aim the idea would be to exploit the traffic data to obtain information about users' service usage that is requested to satellites in order to have a realistic Network Slicing scenario. Despite the fact that the data has been collected using a GEO system and in 5G-STARDUST we are also addressing LEOs, we expect that the user needs and their usage of services can be applied for these scenarios, too. The usage of some low latency service, e.g. online gaming, might increase in reality.

The data shared contains the traffic patterns for residential broadband and maritime scenarios. In the case of residential broadband, the traffic demands in the adjacent beams will give an idea about the usage trends in the given area. The share of traffic for each service type is also available, which has different SLA requirements. This information can be used for load balancing between the terrestrial and non-terrestrial links in multi-connectivity setups. This is also useful in identifying the service types that can be handed over to the terrestrial network in case of congestion. Approaches based on AI/ML load prediction can be used for this hand over. In situations where the entire traffic of the user has to be handed over to the terrestrial network, the users with the least impacted SLA while switching from NTN can be hand over based on this approach.

The traffic pattern for the maritime scenario is useful for evaluating mobility scenarios. The use of AI/ML based approach can be suitable to predict the position of the users. This prediction approach will also give an indication of the traffic demand in the near future. The vessels which will be in the coverage of the terrestrial network can also be identified using this data. In areas where the vessel would be in the coverage region of the terrestrial network, the traffic demand on the satellite beam can be evaluated to identify whether the terrestrial capacity can be leveraged to provide extra capacity for the vessels. If the offloading of the traffic to the terrestrial network facilitates the satellite to allocate additional capacity to the users which can be served only by the satellite, the vessel can be completely served by the terrestrial network. In scenarios where both the links are available, the traffic demands in both the networks can be evaluated to divide/offload the traffic.

3 LARGE SCALE SATELLITE SYSTEM SIMULATION/EMULATION DATA SET

3.1 MOTIVATION

In the rapidly evolving domain of satellite mega-constellations, a critical limitation for advancing machine learning and system optimization is the scarcity of real-world training data. This scarcity stems from the relatively undeveloped state of mega-constellation operations. To bridge this gap, Fraunhofer FOKUS Institute in European Space Agency (ESA) funded projects, has developed a large-scale simulator/emulator. This innovative solution is pivotal in generating a comprehensive dataset, indispensable for diverse machine learning applications in the satellite field.

Although the simulator/emulator relies on artificially generated inputs, such as the theoretically computed orbits of satellites or hypothetical system loads, the resulting data is of significant practical relevance. It includes a significant number of otherwise impossible to compute details and side effects, like the mobility patterns of satellites and the influence of various routing protocols on local congestion issues within the network. This data is essential in understanding and predicting the operational dynamics of mega-constellations.

To answer this limitation, we have incorporated a basic set of data acquired from this simulator/emulator into our current deliverable. This inclusion serves a dual purpose: firstly, as a foundational placeholder for the kind of internal data we anticipate from operational satellite system mega-constellations and low orbit networks; and secondly, as a robust dataset for researchers and practitioners to delve into system optimization and machine learning challenges specific to this cutting-edge domain of satellite communications.

3.2 LARGE SCALE SATELLITE SYSTEM SIMULATION/EMULATION ENVIRONMENT

The Fraunhofer FOKUS simulation/emulation environment offers a range of functionalities for exploring key parameters in NTN networks. This testbed comprises three main components:

1. **Abstract Calculations:** This initial component processes system specifications, calculating the basic parameters of the NTN and TN links in the system. It generates links with specific establishment and tear-down patterns, representing the abstract data introduced in the system and through this being the theoretical input differing from a real system acquired data. The abstract calculations are highly simplified compared to complete systems, considering perfect orbits for the satellites and that the user and feeder link connections don't have any attenuation. More accurate abstract calculations can be added in case there is a need.
2. **Simulator Environment**: The abstract calculations are used in a discrete event simulator (DES) that leverages previously generated link calculations to simulate a network of nodes. The focus here is on simulating forwarding protocols as expected from Task 5.3. Within this DES framework, routing decisions and forwarding actions are monitored and can be recorded for external analysis. This allows for the determination of end-to-end (E2E) link characteristics over time. All elements within this simulation are streamlined to their essential functions to minimize simulation overhead and ultimately to provide the possibility to execute multiple consecutive simulations.

3. **Emulation Environment:** This component utilizes the E2E link characteristics generated in the previous stage to recreate the environment of a single system or multiple system components. This allows for the embedding and testing of a specific Payload Under Test in a comprehensive realistic environment. The Emulation Environment includes an end-to-end link emulator which enables the E2E link characteristics to connect multiple Payloads Under Test. Control over these payloads is managed by the “Control Center”, which handles configurations for forwarding and routing behaviour and acquires the node usage information in the form of CPU and memory usage as well as information on the momentary data traffic enabling the acquisition of congestion related information.

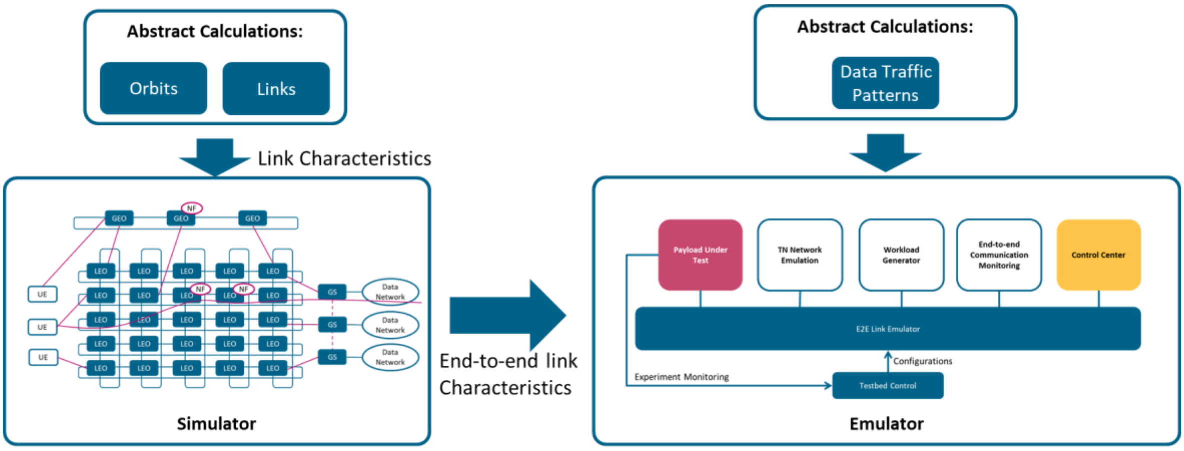


Figure 4 – Fraunhofer FOKUS Simulation / Emulation Environment

3.2.1 Abstract Calculations

The mega-constellation system comprises numerous satellites, each requiring unique parametrization. As such, the system's specification represents a compilation of definitions for every satellite in the network. This process includes several steps. Initially, the positions of all system participants are modelled in 3D space, factoring in their orientation. Consequently, both the position and orientation of a satellite significantly influence the visibility between the satellites and their terminals. The data regarding position and orientation is employed to determine the geometric visibility among all terminals within the system.

Subsequently, this list of geometrically feasible links undergoes further analysis by a link planning block. This phase incorporates additional information like the duration of link availability and the satellites' movement directions, thus enhancing the link data with these added characteristics. However, factors such as weather conditions for space-ground links and link cost are excluded to simplify the abstract data computation.

This procedure is repeated at various discrete time intervals, offering insights into the system's behaviour over time. It's important to note, though, that this model does not include certain effects found in a real operational system, such as solar storms and lunar gravity, as well as the impacts of weather and link cost considerations.

3.2.2 Simulator

The primary goal of the simulator is to accurately represent the mega-constellation NTN system and provide the first set of system level data. At the heart of this simulation lies a discrete event simulator which is instrumental in studying and analysing system behaviour over extended periods, with the simulation results being captured in real-time.

For a comprehensive modelling of the NTN system, we utilize the selected links list from an earlier step. This list serves as the core of the model, establishing the connections among all system participants. Essentially, the list of selected links fulfils a dual purpose: constructing the physical domain within the simulation (determining which interface connects to which) and informing the Network Control Center. The Network Control Center, designed to oversee all network participants, is thus enabled to operate the interconnected nodes effectively.

This simulation process ultimately yields end-to-end (E2E) link characteristics, vital for subsequent stages of the overall testbed design. These E2E link characteristics, encompassing all significant E2E links and their subsections, are generated and exported based on the format of the data from the next section. When examining a process involving multiple stages, it is crucial to process and store all segments of the E2E link that are segmented by physically modelled participants. These physically modelled participants are defined within the overall simulation scenario to minimize simulation overhead and provide a generalized perspective of the simulated system.

This data will be used in Task 5.1 for the multi-link characteristics as well as in Task 5.3 for the assessment of SDN-based network control.

3.2.3 Emulator

The emulator aims to merge the physical domain with insights from preceding calculation and simulation phases. Its main goal is to forge a mixed environment, where physical devices are seamlessly integrated with emulated system components. This unique setup is designed to reflect real-life scenarios, utilizing data from simulation-generated links to configure the virtual segments of the emulated system. A pivotal aspect of the emulator is its ability to interact with real devices in real time, enhancing the study of how physical devices harmonize with system components and to acquire node related monitoring data.

To ensure effective integration with physical devices under test and the simulated physical and network domain, the emulation of the links connected to these physical devices is crucial. This responsibility is handled by the “Testbed Control” and “E2E Link Emulator”, which leverage simulation outcomes to replicate the environment of the device or payload under test. To profile the data plane, a traffic generator is incorporated, ensuring thorough measurement analysis.

The hardware testbed also has the capability to integrate payloads into a small 5G network. This is accomplished by incorporating the 5G core network within the network emulation, with certain 5G network functions possibly executed on the payload through the satellite’s virtualization environment. This arrangement facilitates the deployment of network services and allows for the investigation of their effects on end-to-end (E2E) system behaviour. Given the testbed's versatile design, it can support various scenarios, providing a flexible environment to assess the potential and possibilities of network function placement.

From the emulator, data related to the usage of the payload is acquired, this including the data related to the CPU and memory usage as well as the data exchanges on the data plane, enabling the understanding and the evaluation of the specific optimization algorithms especially in the direction of end-to-end QoS characteristics as required by Task 5.1 and Task 5.4.

3.3 EXTRACTED DATA TYPES

The Fraunhofer simulator/emulator environment is providing the following types of data characteristics to be further used to develop and later assess the different optimizations developed across WP5.

Abstract link characteristics are exported in the form of a JSON file.


```
{
 "edges": [
   {
     "characteristics": {
       "capacity": 400000000000,
       "distance": 4100000
     },
     "destination_id": 2,
     "destination_interface": 3,
     "edge_id": 18,
     "source_id": 1,
     "source_interface": 4
   }

 ]

}
```


The main information of the abstract link characteristics are the link capacity and the distance. The distance will be used based on the specific type of link used (i.e. Rf or optic) to determine the specific link delay. In addition to this, at a later stage, a packet drop rate may be added during the 5G-Stardust developments.

End-to-end link characteristics are exported in the form of a Comma-Separated-Values (CSV) file including the following information:

- Bandwidth – the capacity of the end-to-end link based on the different link segments it is composed of.
- Delay – the aggregated delay of the link segments into an end-to-end link delay
- Packet loss – based on the packet drop rate of the different links, an end-to-end packet drop rate could be computed. This is not part of the initial collected data.

Emulation environment data is exported in the form of a Comma-Separated-Values (CSV) from the monitoring system of the emulation environment and consists of:

- End-to-end application-level data flow characteristics: bandwidth, delay and packet loss
- Payload under test monitoring: CPU and memory usage, packets processed per second.

Please note that the simulator/emulator environment is still in the development until mid-2024 and may suffer minimal changes in the acquired data characteristics. This would not impact the developments of the 5G Stardust WP5 as the data is needed at a later stage when the different new concepts are developed to be able to develop their proof of concept and to validate them.

3.4 PURPOSE OF THE DATA SET

Given that the data set originates from a simulation/emulation environment, it lacks numerous real-world characteristics. Therefore, the applicability of this data set is inherently restricted, a factor that must be taken into account with due consideration. It's important to recognize that the abstract input data does not translate directly and immediately into the output data. Instead, there's a notable variation through the environment, akin to what one would expect in actual real-world deployments. The data set will be used by the following WP5 tasks:

- Multi-connectivity – the characteristics of multiple end-to-end links will be used to be able to assess end-to-end connections (Task 5.1)
- SDN-based network control – the end-to-end link characteristics as well as the node resource usage will be used to determine a basic routing solution suitable for the target mega-constellation (Task 5.3)
- QoS assessment – the end-to-end 5G NTN QoS will be assessed using the emulation environment (Task 5.4)

4 REFERENCES

[1] Hispasat, "Satellite Characteristics AMAZONAS 5," [Online]. Available: https://www.hispasat.com/contenidos/web/0/192-satelite-amazonas-5-de-hispasat-3.pdf.

[2] Hispasat, "Satellite Characteristics H30W-6," [Online]. Available: https://www.hispasat.com/contenidos/web/0/194-h30w-6.pdf.