Energy-Efficient Satellite joint Computation and Communication

Master Thesis
submitted to the Faculty of the
Escola Tècnica d'Enginyeria de Telecomunicació de Barcelona
Universitat Politècnica de Catalunya
by

Marc Martinez Gost

In partial fulfillment
of the requirements for the master in
Master's degree in Telecommunications Engineering

Advisor: Ana Isabel Pérez Neira
Barcelona, June 28th, 2021

Abstract

The emerging interest in satellite networks will be a key driver in the path to 6G. The satellite segment must be conceived beyond a mere relay system, where nodes can process data and offload the terrestrial segment. Besides, evidence suggests that energy consumption is among the most important factors for the design of future communication networks. For this motivation, we introduce Sat2C, an energy-efficient algorithm for satellite joint routing, radio resource allocation and task offloading for latency-constrained services. We develop a novel energy model that incorporates the power amplifier subsystem and changes the geometry of the problem. Regarding the routing task, we propose the SHIELD algorithm, based on the submodularity framework and which achieves Pareto-efficient routes. Besides, the RRM problem is formulated as a log-log convex program. The experimental results reveal that Sat2C has low computational complexity, provides routes with low variance in the mean distance and the transmission powers are optimal to ensure energy minimization.

Acknowledgements

I would like to express my special thanks of gratitude to my supervisor, Ana Isabel Pérez-Neira, for offering me this opportunity to learn from her experience and passion. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. You have encouraged me in all the time of my (early) academic research. Also, I am particularly grateful for the technical support given by Israel Leyva-Mayorga in the development of this project.

I would also like to extend my thanks to my family: to my mother, my father and my sister for their encouragement throughout my life and being a part of my getting here. Finally, I wish to thank my fiends, whose support and guidance are with me in whatever I pursue.

Revision History and Approval Record

Revision	Date	Purpose
0	24/05/2021	Document creation
1	25/06/2021	Document revision

DOCUMENT DISTRIBUTION LIST

Name	e-mail
Marc Martinez Gost	marc.martinez.gost@estudiantat.upc.edu
Ana Isabel Pérez Neira	ana.isabel.perez@upc.edu

Written by:		Reviewed and approved by:	
Date	28/06/2021	Date	28/06/2021
Name	Marc Martinez Gost	Name	Ana Isabel Pérez Neira
Position	Project Author	Position	Project Supervisor

Contents

List of Figures

List of Tables

List of Abbreviations

5G	5th Generation mobile network
6G	6th Generation mobile network
AIS	Automatic Identification System
CPU	Central Processing Unit
DL	Downlink
DMC	Disaster Monitoring Constellation
EO	Earth Observation
ESA	European Space Agency
FSO	Free-Space Optics
FSPL	Free-Space Path-Loss
GEO	Geostationary Orbit
GGP	Generalized Geometric Program
GNN	Graph Neural Network
GS	Ground Station
ICT	Information and Communication Technologies
IOL	Inter-Orbital Link
IoT	Internet of Things
ISL	Inter-Satellite Link
LEO	Low Earth Orbit
LoS	Line of Sight
LLCP	Log-Log Convex Program
MEC	Mobile Edge Computing
MEO	Medium Earth Orbit
MIP	Mixed-Integer Program
NGEO	Non-Geostationary
NFV	Network Function Virtualization
NP	Nondeterministic Polynomial
OBO	Output Back-Off
OFDMA	Orthogonal Frequency-Division Multiple Access
OPEX	Operational Expenditure
QoS	Quality of Service
RF	Radio-Frequency
RRM	Radio Resource Management
SAR	Synthetic Aperture Radar
SCPT	Single-Carrier Per Transponder
SDMA	Spatial Division Multiple Access
SNR	Signal-to-Noise Ratio
TCP/IP	Transmission Control Protocol/Internet Protocol
TWTA	Travelling Wave Tube Amplifier
VNF	Virtual Network Function

1 Introduction

The last two decades have seen a growing trend towards space-based Internet services. This is the case of high-tech competitors, such as SpaceX Starlink, OneWeb and Amazon Kuiper, that have deployed mega-constellations of LEO satellites [1]. There is a demanding need to address the standardization of the satellite segment with respect to the ground infrastructure, which will play a pivotal role on the path to 6G.

Of significant interest is the use case of delay-sensitive Earth Observation applications, including emergency communications and real-time surveillance. Opposed to a wired network, a satellite infrastructure provides robustness and more durability. One of the greatest challenges of LEO constellations is the dynamism of the network associated to high speeds. Prior studies have noted the importance of inter-satellite routing and on-board processing as a means to provide the service ubiquitously, with shorter latency, better Quality of Service (QoS) and, ultimately, with the satellite segment becoming less dependent of the terrestrial network.

A key driver in the development of future wireless networks is energy-efficiency, which is motivated by the operational expenditure (OPEX), the development of environmentally sustainable technologies and extending the lifetime of networks. Much of the available literature highlights the relevance of Radio Resource Management (RRM) in an edge computing environment to reduce the energy consumption, even though it ignores the potential satellite on-board processing.

This thesis seeks to explore the fundamental trade-offs between computation and communication in dense LEO satellite networks. We develop Sat2C, a cross-layer algorithm addressing the joint optimization of routing, RRM and task offloading for energy minimization under time delay constraints. The novelty of this study resides in using the computation capabilities of the satellites and not just as a relay system to the data centers.

The remaining part of the thesis proceeds as follows: section 2 gives a brief overview of the state of the art in energy-efficiency, satellite networks for Earth Observation and Edge Computing; section 3 begins by laying out the theoretical dimensions of the research and the proposed system model; section 4 is concerned with the methodology used for this study; section 5 presents the experimental results and section 6 summarizes the conclusions and proposes future lines of research.

1.1 Research contribution

We aim to solve the problem of jointly optimizing the routes, the transmission powers and data offloading decisions in order to reduce the energy consumption of the LEO satellite network and meet latency constraints. This formulation lacks in the current literature, specifically for the emerging field of satellite communications.

Much of the literature is devoted to solving this problem with novel algorithms, though they neglect to using an appropriate model of the energy consumption. Consequently, we

dedicate part of this research to developing a more general power consumption model, in particular, to the effect of the power amplifier module. This is particularly understudied and brings critical implications, as the geometry of the problem changes and the existing solutions for energy minimization may not work anymore.

Since the problem to solve is NP-hard, the methodology we propose to find a suboptimal solution is novel as well. In order to reduce the energy consumption, the Sat2C algorithm encompasses two components: in the former we design the SHIELD algorithm, that tackles the routing procedure via a submodularity framework and the Dijkstra's shortest path algorithm; in the latter, the radio resources and data offloading decisions are reformulated into a convex problem that provides overall minimum energy with respect to the established paths.

The findings of this research have been submitted to the *W11: Evolution of Non-Terrestrial Networks toward 6G* workshop, held within the *2021 IEEE 94th Vehicular Technology Conference: VTC2021-Fall*.

1.2 Gantt diagram

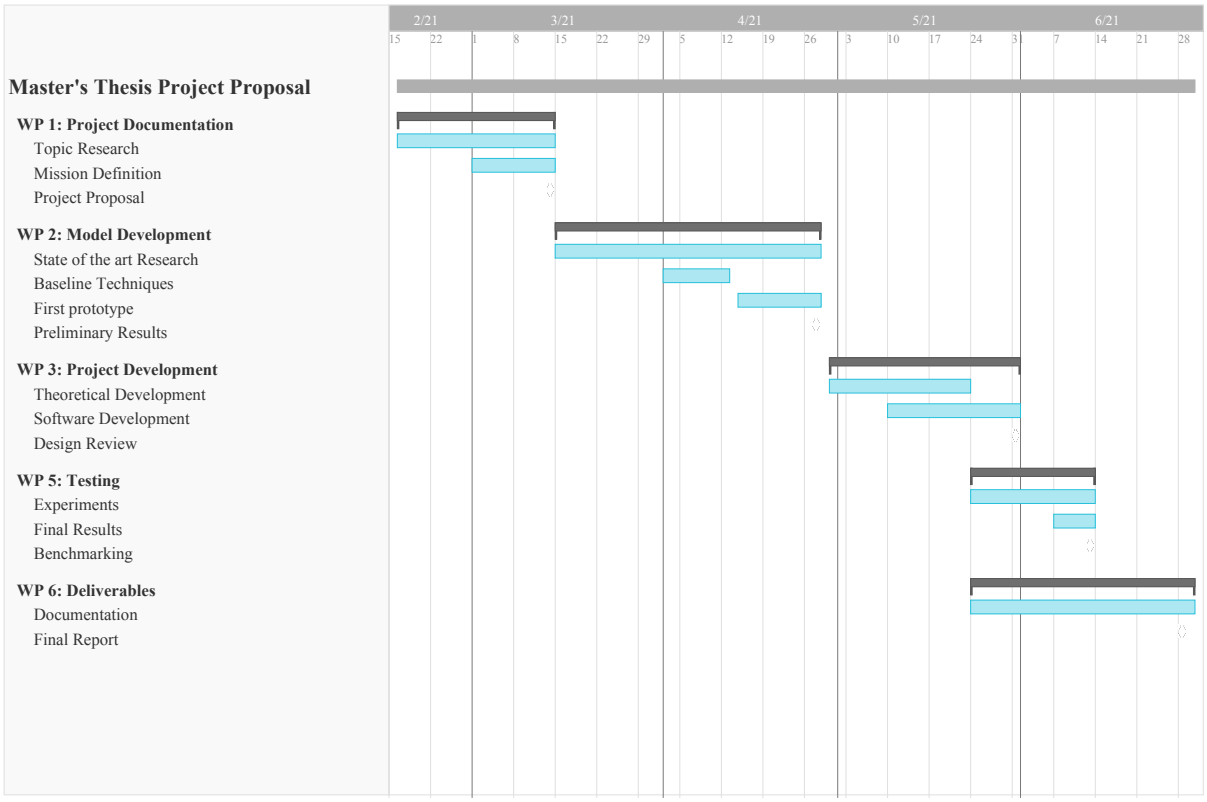


Figure 1: Gantt diagram of the project.

2 State of the Art

This section highlights the cutting edge research that leads to developing energy-efficient algorithms for joint computation and communication in satellite networks. First, the concept of energy-efficiency in wireless networks is introduced and a brief description of satellite networks for Earth Observation is conducted. Then, we examine the Edge Computing paradigm as a means to reduce the distance between users and services and how this can be transferred to satellite networks. Finally, we propose a prospective graph-based optimization approach for radio and computation resource management.

2.1 Energy-efficiency in wireless networks

Energy consumption is a major concern due to the increase in its demand and the challenges it entails. Via smart grids, industry automation, intelligent logistics, among other services, Information and Communication Technologies (ICT) are a facilitator for reducing global energy consumption. Even though, the volume of network traffic has been increasing sharply in the last decade [2], which evinces the existence of an intrinsic trade-off. The authors in [3] suggest the advent of energy-efficiency in wireless communications is traced to the following fundamental drivers:

- Development of more complex and diverse capabilities;
- Proliferation of ad-hoc networks and support for mobility;
- Reduction of OPEX by means of energy costs minimization;
- The green communications concept due to the environmental footprint concern;
- Smaller form-factor devices.

Energy represents a fundamental restriction in communication networks and, particularly, in wireless settings as the energy consumption affects the network lifetime. One metric that allows to assess it is the energy-per-bit E_b, defined as the energy needed to transmit one bit. From the Shannon's capacity formula

$$R = B \log_2 \left(1 + \frac{P}{N}\right), \tag{1}$$

where B is the bandwidth allocated for transmission, P is the transmission power and N is the noise power. The energy-per-bit can be expressed as

$$E_b = P\, t_b = P\, \frac{1}{R} = \frac{P}{B \log_2 \left(1 + \frac{P}{N}\right)}, \tag{2}$$

where t_b is the time-per-bit. The dependence of the energy spent in communication in terms of the transmission power reveals the relevance of power control as a state of the art procedure for managing energy consumption, besides influencing other communication challenges (e.g., transmission delay, interference, error probability, etc.).

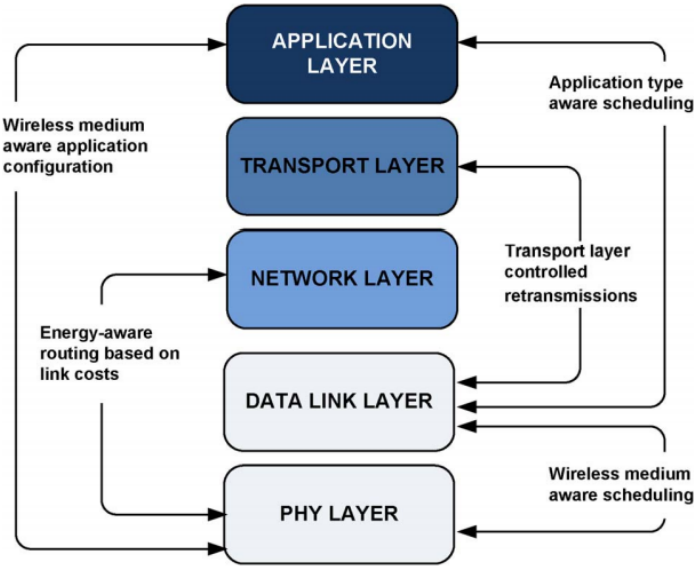


Figure 2: (Retrieved from [3]) Cross-layer mechanisms for energy efficiency. The signaling and information exchange between nonadjacent layers facilitate cross-layer optimization of energy consumption.

In a communication system, the energy consumption arises from five different components [4]: the direct current power supply module, the base band module, the radio frequency module, the power amplifier module and an active cooler and battery backup module. From those, the power amplifier is a critical source of power consumption and, paradoxically, not considered in many energy consumption models. This module may represent more than 50% of power consumption in base stations [5] and it is expected to be larger in satellite communications.

The vast majority of studies on transmission power control have focused on the traditional TCP/IP layered protocol stack. While this allows for robust and scalable solutions, the limited interaction between layers restricts the optimality of the algorithms. Recently, the topic of cross-layer design for energy-efficiency is gaining attention since it favours the expansion of functionalities of the network. Particularly, we are interested in the connection between the physical and the network layer (see Figure 2). A significant analysis and discussion on the subject is presented in [3].

2.2 Satellite Networks for Earth Observation

In many locations environmental monitoring or agritech are out of range of terrestrial IoT networks. Satellite-delivered IoT services could help provide connectivity for these growing markets. The European Space Agency (ESA) has estimated that there are at least 20 satellite companies looking to deliver space-based IoT services [6]. For instance, the satellite mission PAZ [7] focuses on the acquisition of SAR data and it is equipped with a transponder able to detect Automatic Identification System (AIS) data, which allows vessels to intercommunicate their positions among other parameters in order to

avoid collisions [8]. This mission is a relevant solution for maritime monitoring systems since data from SAR images and AIS data can be merged in order to detect malicious vessels. By deploying a large number of down-scaled low-cost PAZ satellites, the dual sensing functionalities may be expanded to collaborative data fusion or collaborative remote classification, inference and learning, aiming to provide a solution for this delay-constrained service.

This notion of a satellite constellation for Earth Observation (EO) was promoted more than 10 years ago [9]. This was the case of Disaster Monitoring Constellation (DMC) [10] and RapidEye [11] which disappeared during its deployment. DMC is composed of 5 micro-satellites capable of providing multi-spectral images in less than one day with a relevant resolution. Recently, the company Planet acquired RapidEye and it is leading the market of EO with small satellites missions, aiming to have 12 satellites available [12]. Yet another relevant actor of non-geostationary (NGEO) Earth Observation is the Finish company IceEye [13].

The design of a satellite network depends ultimately on the service and can be assessed through the following factors [14]:

- **Service requirements**: for the given use case we seek for real-time applications. In order to have global coverage dense constellations are needed, which allow to obtain data with high-time resolution.
- **Orbit**: attending to the height of deployment with respect to the Earth's surface we find: Geostationary Orbit (GEO) satellites, located at 35.786 km and whose position is fixed with respect to the Earth; Low Earth Orbit (LEO) satellites, located between 200 and 3000 km and Medium Earth Orbit (MEO) satellites, located anywhere between LEO and GEO. The majority of deployed satellites are GEO as they provide wide coverage at expenses of large delays and high power consumption, both because of large distances. Conversely, LEO satellites are adequate for delay-constrained services because of their low altitude with respect to the Earth surface. This also complements energy-efficiency, as it allows to use smaller transmission powers.
- **Coverage**: unlike GEO, which can provide global coverage with only three satellites, the coverage of LEO satellites is smaller, so dense constellations have to be deployed. Moreover, the cost of deployment for LEO satellites is smaller and they provide a better spatial resolution because of its proximity to the surface.
- **Frequency band**: the major source of power consumption is due to the free-space path-loss (FSPL), which can be modelled as

$$FSPL = \left(\frac{4\pi df}{c}\right)^2, \tag{3}$$

where d is the distance between the antennas, f is the transmission frequency and c is the speed of light. Because of large distances between satellites, the large attenuation is compensated with antennas providing large gains. Nonetheless, the frequency affects the size of the antenna, the magnitude of the Doppler shift and, for frequencies above 10 GHz, the attenuation produced by rain and other atmospheric conditions is

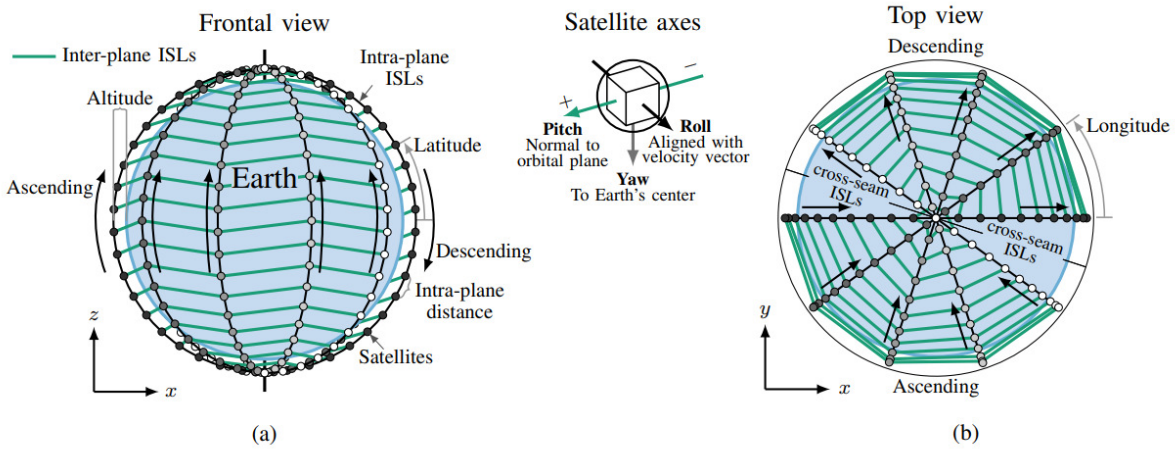


Figure 3: (Retrieved from [17]) (a) Frontal view and satellite axes and (b) top view of a Walker star constellation with 200 satellites and 5 orbital planes.

noticeable. Current development is moving towards employing Free-Space Optical (FSO) links, as they provide higher rates. For the downlink (DL) to the ground station this technology cannot be used because the atmospheric conditions affect light communications.

As a consequence, dense LEO constellations are suitable for EO missions and will be integrated in 5G and beyond networks [15, 16]. Their operational complexity comes from high orbital speeds, resulting in a very dynamic topology. Moreover, on-board processors usually have low capabilities, which promotes the collaboration among satellites to enhance the sensed data and reduce the consumed data rate. Besides that, GEO or MEO can incorporate more information or just transmit it directly to the Earth data center as a data relaying system. This favours the inception of multilayered networks via inter-orbital links (IOL) that allow to trade-off and take advantage of the desired properties of every type of orbit. Nevertheless, in this thesis we focus on dense constellations of LEO satellites only.

LEO constellations are typically organized in groups of satellites called orbital planes that travel in the same trajectory. Communication between satellites occurs through inter-satellite links (ISL). We discern between intra-plane or inter-plane ISLs, whether the communication occurs within the same plane or not, respectively. Figure 3 illustrates a classic Walker star constellation. The white and black satellites are orbiting in opposite directions, known as cross-seam ISLs, and resulting in large relative velocities between them. Since it is very challenging compensating the Doppler effect, it is a common practice leaving these ISLs unimplemented [17]. To avoid collisions where the orbital planes intersect (e.g., the poles), they are deployed at slightly different altitudes. Conversely, this produces orbital period deviations so that the inter-plane ISLs are not fixed, which emphasizes the dynamic structure of the network. Otherwise, satellites need to implement active monitoring to avoid collisions, translating in more energy consumption.

Current implementations do not route the data throughout the satellite segment. In fact, the LEO satellite caches the data until it can establish a communication to a ground station. This represents a drawback for delay-sensitive services, because the orbital period of LEO satellites is around 90 minutes. Moreover, even if the data is not constrained in time, the data storage of the satellite may be limited. This promotes the development of routing algorithms for NGEO constellations. [17] proposes a new methodology for establishing dynamic ISL links that maximize the throughput of the LEO network.

With the advancement of aerospace technology, commercial satellite companies such as OneWeb [18], O3b [19], and SpaceX Starlink [20] have proposed new LEO and MEO satellite constellation plans, increasing investment and research on satellite-related technologies, making satellites more powerful and smarter. Nevertheless, most existing applications need to be cloud-assisted. This motivates the formulation of the satellite segment in the context of Mobile Edge Computing (MEC).

2.3 Edge Computing

Edge computing reduces the distance between users and services. It is designed to move computing capabilities from centralized cloud servers to edge nodes near the user end, so that tasks can be performed at the edge of the network. The hybrid design of the satellite-terrestrial network very easily adapts to the edge computing paradigm, which can reduce the amount of data transmitted from satellites and communication delays, improve the bandwidth utilization of ISLs and reduce the pressure on data processing of satellite ground stations [21].

The emergence of dense LEO constellations has shifted the research of MEC towards including the satellite segment. In fact, throughput and latency suffer greatly in the traditional access network, essentially cancelling out much of the gain from technologies such as optical line transfer and fiber-to-the-home, and 5G networks. LEO can alleviate part of the edge terrestrial access. In [22], a framework for MEC to improve QoS in satellite-terrestrial networks is proposed. Like terrestrial MEC, the satellite MEC extends the computation capacity and can store content to reduce redundant transmissions from the remote cloud. The authors consider a dynamic Network Function Virtualization (NFV) system with a dynamic resource allocation to check whether a Virtual Network Function (VNF) is connected or gone. Among the computation offloading strategies that are considered, the cooperative computation offloading, in which MEC servers cooperate to complete computation tasks of the user tasks, is shown to provide the best performance.

The authors in [23] studied a cloud-edge computing architecture (Figure 4) where the edge IoT layer is the LEO constellation, and the cloud side is divided into two layers, the GEO satellite cloud and the ground cloud. Notice that the GEO satellite cloud represents a longer routing path, but the links are always available and it allows a faster data transfer. This layer could be removed in the future once the FSO technology is mature and can compete with the GEO relay system. Using this architecture, satellite IoT edge nodes and satellite IoT cloud nodes can cooperate with each other.

The other set of applications use the satellite network as a service for edge computing

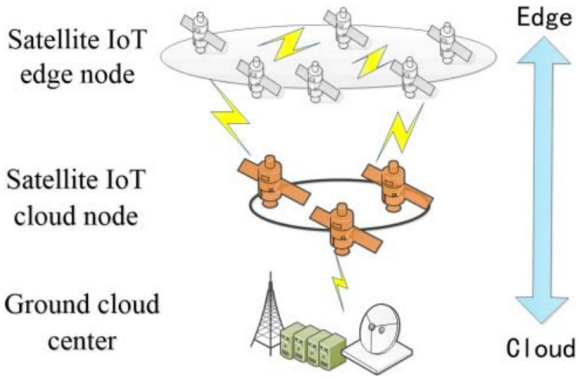


Figure 4: (Retrieved from [23]) Cloud-edge computing architecture on satellite IoT.

required from terrestrial services (e.g., IoT devices). The data is generated on the Earth and the satellite network accepts computing tasks from ground devices. This kind of scenario is considered in [24], where computation tasks from IoT device users in remote areas are offloaded to satellite edge servers. Specifically, satellite edge computing is combined with NFV, such that the available resources of LEO satellite can be abstracted into a resource pool and provide agile service provisioning to the end-users. In [25] a hybrid satellite-terrestrial architecture is proposed for task allocation to optimize the energy consumption and delay. By exploiting the satellite segment a reduction in 50% is achieved in both metrics.

However, in the two preceding studies no Radio Resource Management (RRM) is considered. This would significantly reduce the energy consumption and would provide more tailor-made algorithms. In both approaches, the dynamic topology of the satellite segment is substituted by a simplified architecture where each satellite is assumed to have four satellite neighbors. This makes the analysis more tractable though oversimplifies the satellite constellation, which can be further exploited. To date, several studies have investigated the fundamental trade-offs of Edge Computing. Since the optimization problems are, in general, NP, different authors focus on different approaches to find suboptimal solutions based on, for instance, Markov decision processes [26], game theory [27] or Lyapunov optimization [28].

The energy-efficiency driver is a prevalent metric for designing offloading algorithms under the MEC architecture. In fact, radio resource optimization is a common practice to compensate for the channel conditions. In [29] an OFDMA scheme is chosen to jointly optimize power and subcarrier allocation in order to reduce the energy consumption. [30] develops a joint Computing, Caching, Communication, and Control algorithm that exploits the collaboration between MEC servers for Big Data applications to reduce bandwidth and meet delay constraints. For non latency-critical applications a long-term metric is more suitable. For instance, [31] adopts an stochastic analysis via communication and computation resource management to reduce the average sum power consumption.

2.4 Graph-based resource optimization

Graphs are powerful data structures that capture dependencies through edge connections and provide high flexibility for representing non-Euclidean spaces. A diverse variety of algorithms have been developed for routing, such as for shortest paths [32] and clustering with constraints [33].

The studies presented thus far provide evidence that edge computing in LEO satellite networks is an active field of research. Nevertheless, most of the existing literature focuses on either routing or RRM, but they do not address both simultaneously. There would therefore seem to be a definite need for a joint optimization.

Little is known whether a graph is useful for resource optimization. In [17], the authors present a greedy algorithm to establish dynamic ISL that maximize the sum rate of the network. In the same vein, [34] models each communication link as a node in a graph and allocates resources for a vehicular network. Nevertheless, the graph is used as a support for information and not fully exploited to optimize the resources.

In a recent study, [35] investigates whether the RRM problem can be formulated via Graph Neural Networks (GNN). The authors found this method outperforms classic optimization-based approaches, besides being highly scalable and with low computational complexity. The use GNNs is out of the scope of this thesis, even though further research should be undertaken to determine the usefulness of graphs for RRM.

3 System Model

In this research we focus on the use case of remote surveillance via a LEO constellation (e.g., the PAZ mission). The processing task is related to the detection of events in images captured by the satellites and it must be executed within a time window. This scenario is depicted in Figure 5 for 5 satellites. The collaborative task between satellites within an edge computing environment is left for future research.

Let us consider a set of LEO satellites $\mathcal{N} = \{1, \ldots, N\}$ gathering EO data. This EO data has a size of D_n bits for $n \in \mathcal{N}$. The aim of the satellite system is to offer the processed data, F_n, to a final user. We consider an arbitrary processing function, in which the ratio

$$\rho_n = \frac{D_n}{F_n} \geq 1, \tag{4}$$

is coined as the processing ratio and it provides a notion of the processing performed. In case the processing is a compression, ρ_n becomes the compression ratio. This processing is assumed that it can be done at the source LEO satellite ($n \in \mathcal{N}$), or at the Ground Station (GS). The connection between the LEO satellites and the GS is done by direct connection via a feeder link. Even though in practice there are several GS, they all will be envisioned under a unique node, which eases notation.

The data collected by the n-th LEO satellite shall be delivered to the GS with a maximum delay of τ_n. To fulfill this requirement, the satellite system shall decide i) a route between the LEOs and the GS through; ii) where the processing is going to be performed (locally or at the GS) and iii) the RRM parameters to be employed in the communication.

3.1 Satellite-Terrestrial Network

Let us consider a Walker star LEO constellation of M orbital planes and N_a satellites per orbital plane, where $a = 1, \ldots, M$. All satellites are deployed at the same altitude H. The satellites maintain four ISLs whenever possible: two intra-plane ISLs with the closest intra-plane neighbors, and two inter-plane ISLs with the closest inter-plane neighbors (see Figure 3).

We model the LEO network as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with vertex set $\mathcal{V} = \mathcal{U}_{GS} \bigcup_{a=1}^{M} \mathcal{V}_a$, where $\mathcal{U}_{GS}$ is the destination GS ($|\mathcal{U}_{GS}| = 1$) and $\mathcal{V}_a$ is the set of LEO satellites deployed in orbital plane a. The edge set $\mathcal{E}$ represents the wireless links established for communication between them. The intra-plane ISLs within an orbital plane a constitute the set of edges $\mathcal{E}_a \subset \left\{ij : i, j \in \mathcal{V}_a^{(2)}\right\}$ and $|\mathcal{E}_a| = N_a$. The inter-plane ISLs constitute the set of edges $\mathcal{E}_{\text{inter}} \subset \{ij : i \in \mathcal{V}_a, j \in \mathcal{V}_b, a \neq b\}$. LEO satellites and the GS are connected with the DL edges $\mathcal{E}_{DL} \subset \{ij : i \in \mathcal{U}_{GS}, j \in \mathcal{V}_a, a = 1, \ldots, M\}$. Finally, we define the edge set as $\mathcal{E} = \mathcal{E}_{DL} \cup \mathcal{E}_{\text{inter}} \bigcup_{a \in \mathcal{M}} \mathcal{E}_a$.

The weights $\omega(e)$ for all $e \in \mathcal{E}$ are defined by the communication cost. Without loss of generality we consider the ISL and DL distances. Assuming a spherical model of the

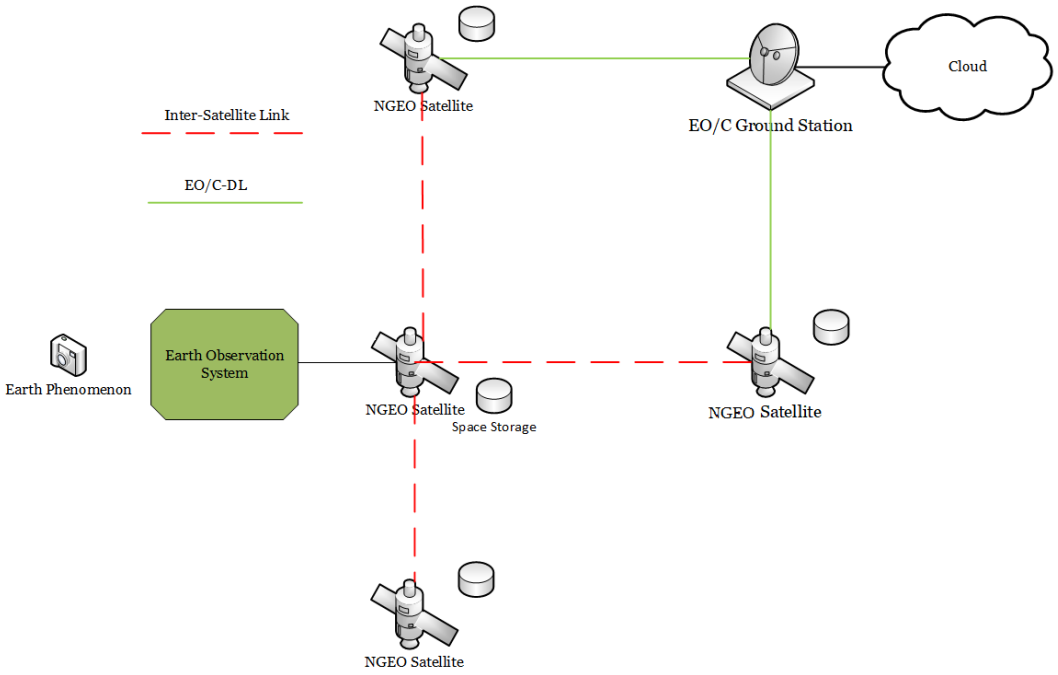


Figure 5: Scenario for remote surveillance.

Earth, the maximum distance that allows Line of Sight (LoS) between satellites u and v is

$$d_{\max,\mathrm{ISL}}(u,v) = \sqrt{h_u^2 + R_E h_u} + \sqrt{h_v^2 + R_E h_v}, \tag{5}$$

where h_i is the height of satellite i and R_E is the radius of the Earth [17]. Thus, this metric bounds which potential satellites can establish a communication.

3.2 Routing

As the GS may not always be visible for all LEO satellites, there might be cases where a LEO opts to forward the data information towards another LEO satellite. While the constellation is dynamic, the time scale for communication and computation tasks is much shorter. Thus, there is no time dependence and we observe the entire network at particular time instants. In this context, we consider a route

$$\mathcal{S}_n = \{n, \ldots, m_n\} \cup \{\mathrm{GS}\}, \tag{6}$$

defined as a set of ordered indexes starting from n and finalizing at the GS. Note that this last hop between m_n and GS is supported via direct DL connection. Connection through a relay consisting of NGEO-GEO-GS link is out of the scope of this project.

Given a route $\mathcal{S}_n$, the resulting time delay due to this transmission can be described as

$$T_n^{DL} = \sum_{i \in \mathcal{S}_n}^{m_n} (T_i^{comm} + T_i^{prop}) = \sum_{i \in \mathcal{S}_n}^{m_n} \left(\frac{L_n}{R_i} + T_i^{prop} \right), \tag{7}$$

where R_i is the data-rate of the link between the i-th and $(i+1)$-th nodes and T_i^{prop} is the corresponding propagation time. We do not need to distinguish between the ISL and DL links as they can be modelled individually with particular constraints in their data rates and distances. Observe that T_i^{prop} is a function of the distance between the nodes, this is, the edge weight $w(e)$; conversely, T_i^{comm} is a function of $w(e)$ along with information of the nodes (e.g., transmission power).

3.3 Data Offloading

Data processing can be performed either locally (at the source LEO satellite) or at the GS. In case it is decided to perform the data processing locally, the amount of transmitted data is substantially reduced. Assuming that the decision variable l_n takes 1 value when the processing is performed locally and 0 when it is offloaded, the data size transmitted through the network becomes

$$L_n = l_n F_n + (1 - l_n) D_n = D_n \left(l_n \frac{1}{\rho_n} + (1 - l_n) \right) \tag{8}$$

The energy employed for performing this processing is assumed to be governed by a function $C_p(D_n)$ [Joules] that only depends on the data size. This function is assumed to be the same either the processing is performed at the space segment or at the GS. As a result, the energy consumption of n-th LEO satellite due to the processing is defined as

$$E_n^{proc} \leq l_n C_p(D_n), \tag{9}$$

we can particularize for certain energy consumption model such as

$$C_p(D_n) = D_n z f_{CPU}^2 \nu, \tag{10}$$

where z is the number of CPU cycles to process 1 bit of data, f_{CPU} is the number of CPU cycles per second and ν the parameter that relates the required processing with the consumed energy in Joules as en equivalent capacitance. In general, it is challenging describing the energy consumption of a processor. A more precise model depends on the number of instructions executed per task, the energy and the time spent per instruction. Nevertheless this parametrization is unmanageable because these specifications are complex to estimate: the number of instructions heavily depends on the algorithm implementation, and the energy and time per instruction depend on the instruction itself (e.g., type of memory access, CPU architecture, supply voltage, etc.). The consumption model we propose captures the influence of the most relevant CPU parameters, has been used in previous research [29, 30, 25, 36, 37] and is consistent with experimental results [38], in such a way that it is used for mobile devices.

Furthermore, we define the delay associated to processing the D_n bits as

$$T_n^{proc} = \frac{D_n z}{f_{CPU}} \tag{11}$$

As stated in (10), the processing parameters are considered constant and identical for all processors, even though these parameters heavily affect the energy consumption. For

instance, the larger the clock frequency the more energy will be consumed and less time the task will invest. This is the reason why several research lines focus on optimizing these parameters in order to lead to convenient trade-offs (see [36] and references therein).

3.4 Radio Resource Management

We now describe the attainable rates of each mentioned link. We assume all links to be half-duplex and used for unicast communication. Besides, ISL and DL links do not interfere among each other, this is, the beam directivity allows Space Division Multiple Access (SDMA) without interference. The bandwidth that is allocated to each of them is B. For the $(i, i+1)$ link the data rate is upper bounded by

$$R_i = B \log_2 \left(1 + p_i h_i^2\right), \tag{12}$$

where p_i is the transmission power between the i-th and the $(i+1)$-th nodes and h_i is the respective channel coefficient, normalized by the receive noise power.

In this context, the energy consumption due to the communication subsystem can be modelled by means of $C_t(L_n, R_i, p_i)$ [Joules], a function that models the energy consumption of transmitting L_n bits at certain data rate R_i and power p_i. A simple yet representative energy consumption model can be written as

$$C_t(L_n, R_i, p_i) = p_i \frac{L_n}{R_i}, \tag{13}$$

3.4.1 Power Amplifier subsystem

Radio-frequency (RF) power amplifiers are a key component in satellite communications and its energy consumption cannot be neglected. For satellite RF communications, Traveling Wave Tube Amplifiers (TWTAs) are the main amplifier choice because they offer higher power at higher frequencies [39]. From [40], the power consumption at the amplifier can be linearly modelled as

$$P_{c,i} = P_{fix} + \frac{c_0}{\eta} p_i, \tag{14}$$

where P_{fix} is the power consumption independent of the output power of the amplifier; c_0 is a scaling coefficient for the power loading dependency, which depends on other subsystems such as the base band module, though these will not be considered; η is the drain efficiency of the amplifier, defined as the ratio between the output and consumed power by the amplifier; p_i is the output power of the amplifier, this is, the transmitted power.

Power amplifiers operate close to the saturation point, where the efficiency is maximum. The Output Back-Off (OBO) prevents the device from trespassing the linear region and determines the efficiency of the amplifier [40]. The OBO characterizes the drain efficiency of the amplifier. There is also a maximum output power P_{out}^{max} that limits the transmitted power of the amplifier.

The overall energy consumption model due to the RF subsystem including the power amplifier can be represented by means of $C_t(L_n, R_i, p_i + P_{c,i})$ as

$$E_n^{RF} = \sum_{i \in \mathcal{S}_n}^{m_n} (P_{c,i} + p_i) \frac{L_n}{R_i} = \sum_{i \in \mathcal{S}_n}^{m_n} (P_{fix} + \mu p_i) \frac{L_n}{R_i}, \tag{15}$$

where $\mu = 1 + c_0/\eta$.

4 Joint Satellite Computation and Communication Optimization: Sat2C

In this section we formulate the energy minimization with transmission power and time delay constraints as an optimization problem. Next we discuss the convexity of the problem and divide it into two subproblems: the first one considers routing through the LEO constellation and the latter manages the resource allocation and task offloading optimally with respect to the previous routing algorithm.

4.1 Problem Statement

The objective to be minimized is the total energy, this is, the sum over all paths. Considering expression (4), the data size D_n is a common factor to all terms, in such a way that the energy can be expressed per bit:

$$E_b^{total} = \sum_{n=1}^{N} E_{b,n}^{total} = \sum_{n=1}^{N} \frac{1}{D_n} E_n^{total} = \sum_{n=1}^{N} \frac{1}{D_n} \left(E_n^{RF} + E_n^{proc}\right) = \sum_{n=1}^{N} E_{b,n}^{RF} + E_{b,n}^{proc}, \quad (16)$$

where E_b^{RF} is the energy per bit definition used in communications. The following optimization problem merges routing, data offloading and radio resource management:

$$\underset{\{\mathcal{S}_n\}_{n=1}^N, \{p_{n,i}\}_{n=1}^N, \{l_n\}_{n=1}^N}{\text{minimize}} \quad E_b^{total} \quad \textbf{(P1)}$$

$$\begin{aligned}
\text{subject to} \quad & \mathcal{S}_n \subset \mathcal{N} \cup \{\text{GS}\}, n = \mathcal{S}_n(1), GS = \mathcal{S}_n(m_n + 1) && n = 1, \ldots, N && \text{(C1-C3)} \\
& g_i \frac{E_n^{proc}}{T_n^{proc}} + p_i \leq P_i, \ i \in \mathcal{S}_n && n = 1, \ldots, N && \text{(C4)} \\
& p_i \leq P_{out}^{max}, \ i \in \mathcal{S}_n && n = 1, \ldots, N && \text{(C5)} \\
& T_n \leq \tau_n && n = 1, \ldots, N && \text{(C6)} \\
& l_n \in \{0, 1\} && n = 1, \ldots, N && \text{(C7)}
\end{aligned}$$

Constraints (C1-C3) restrict the paths to start at each LEO gathering data and end at the GS. In constraint (C4), g_i is a predefined parameter (i.e., not to be optimized) that takes the value 1 when $i = \mathcal{S}_n(1)$ and 0 otherwise. This constraints the power to be below the available at the satellite payload P_i. Likewise (C5) restricts the transmitted power to be below the maximum available at the amplifier. Constraint (C6) refers to the total time, corresponding to computation and communication:

$$T_n = \sum_{i \in \mathcal{S}_n}^{m_n} T_n^{DL} + l_n T_n^{proc} = \sum_{i \in \mathcal{S}_n}^{m_n} \left(T_i^{comm} + T_i^{prop}\right) + l_n T_n^{proc} \quad (17)$$

4.2 Communication time analysis

Problem (P1) is a mixed-integer program (MIP) because of the integrality constraints (C7). Additionally, the simultaneous optimization of routing and RRM is NP and highly non-convex. Thus, finding the optimal solution to (P1) is not possible.

In light of [41] we will analyse the problem in terms of the communication time. For sake of clarity we drop the superscript in T_i^{comm} to simply T_i, since the other delays (processing and propagation) are fixed. Considering (1), we can define the rate according to the capacity and, also, as the ratio between data size and communication, which combined lead to the following expression of the power:

$$p_i = \frac{1}{h_i^2}\left(2^{\frac{L_n}{BT_i}} - 1\right), \tag{18}$$

and the energy cost function can be expressed as

$$E_b^{total} = \sum_{n=1}^{N} \frac{1}{D_n}\left(\sum_{i\in\mathcal{S}_n}^{m_n}\left(P_{fix} + \frac{\mu}{h_i^2}\left(2^{\frac{L_n}{BT_i}} - 1\right)\right)T_i + l_n C_p(D_n)\right) \tag{19}$$

4.2.1 Excluding the power amplifier subsystem

Most of the literature does not consider the power amplifier model, even though the conditions of the problem change dramatically depending on its inclusion. Without loss of generality, the exclusion of this model corresponds to $P_{fix} = 0$ and $\mu = 1$, however only the former changes the geometry of the problem.

Proposition 1. *For $P_{fix} = 0$, the objective (19) is convex and decreasing in T_i.*

Proof. The derivative of the objective with respect to T_i is

$$\begin{aligned}
\frac{\partial E_b^{total}|_{P_{fix}=0,\ \mu=1}}{\partial T_i} &= \frac{1}{D_n h_i^2}\frac{\partial}{\partial T_i} T_i\left(2^{\frac{L_n}{BT_i}} - 1\right) \\
&= \frac{1}{D_n h_i^2}\left(2^{\frac{L_n}{BT_i}}\left(1 - ln(2)\frac{L_n}{BT_i}\right) - 1\right) \\
&= \frac{1}{D_n h_i^2}\Big((1+S_i)\left(1 - ln\left(1+S_i\right)\right) - 1\Big) \\
&\leq 0
\end{aligned} \tag{20}$$

The last equality follows from the rate definition $R_i = \frac{L_n}{T_i} = Blog_2(1+S_i)$, S_i being the SNR between nodes $(i, i+1)$. Let $f(S)$ be expression (20) with derivative $f'(S) = -ln(1+S)$. Since $ln(S)$ is an increasing function in S and $f'(0) = 0$, we have that $f'(S) < 0$ for $S > 0$. Along with that $f(0) = 0$, thus $f(S) < 0$ for $S > 0$. This proves the last inequality, and the objective being a decreasing function in T_i.

Regarding convexity, the second derivative of the objective with respect to T_i is

$$\begin{aligned}\frac{\partial^2 E_b^{total}|_{P_{fix}=0,\ \mu=1}}{\partial T_i^2} &\propto \frac{\partial}{\partial T_i}\left(2^{\frac{L_n}{BT_i}}\left(1-ln(2)\frac{L_n}{BT_i}\right)-1\right)\\ &= ln(2)\left(\frac{L_n}{B}\right)^2\frac{1}{T_i^3}2^{\frac{L_n}{BT_i}}\\ &\geq 0\end{aligned} \tag{21}$$

The last inequality follows from all multiplying terms being positive for $T_i > 0$. Equation (21) is the diagonal entry of the Hessian. Since there are no cross-term elements in the objective, the Hessian is a diagonal matrix with positive entries. Thus it is positive definite and the objective convex. □

This same conclusion is found in [41]. Intuitively, this means that using less power corresponds to a lower SNR and an overall smaller rate, which increases the communication time. Nevertheless, due to (C6) there is a time constraint to be fulfilled such that the energy cannot be decreased infinitely, which leads to Corollary 1.

Corollary 1. *For $P_{fix} = 0$, the minimum energy is obtained with equality in the (C6) constraint.*

$$\sum_{i\in\mathcal{S}_n}^{m_n}(T_i + T_i^{prop}) + l_n T_n^{proc} = \tau_n \tag{22}$$

On the contrary, the propagation time plays no role in the objective, so it can be minimized to allow for a larger communication time, as seen in Corollary 2. Unipath routing via minimization of the propagation time is equivalent to using path-loss: the lower the time, the smaller the distance and the smaller the FSPL, as seen in (3). Besides, according to [42], unipath multirate routing in LEO constellations via path-loss provides a smaller overall latency and favours paths with large data rates in contrast to other metrics.

Corollary 2. *For $P_{fix} = 0$, the minimum energy is obtained with minimum propagation time, since*

$$\sum_{i\in\mathcal{S}_n} T_i = \tau_n - \sum_{i\in\mathcal{S}_n} T_i^{prop} - l_n T_n^{proc} \leq \tau_n - \min_{\mathcal{S}_n}\sum_{i\in\mathcal{S}_n} T_i^{prop} - l_n T_n^{proc} \tag{23}$$

Remark 1. *Notice that the communication time is maximized for $l_n = 0$, but this does not ensure energy minimization. The fact that the propagation time is decoupled from energy allows to treat it independently.*

4.2.2 Including the power amplifier subsystem

Besides increasing the power consumption towards a more realistic model, the power amplifier plays an essential role because it shifts the geometry of the problem.

Proposition 2. *For $P_{fix} \neq 0$, the objective (19) is convex and non-monotonic in T_i.*

Proof. The derivative of the objective with respect to T_i is

$$\begin{aligned}\frac{\partial E_b^{total}}{\partial T_i} &= \frac{\mu}{D_n}\frac{\partial}{\partial T_i}\left(\frac{1}{h_i^2}T_i\left(2^{\frac{L_n}{BT_i}}-1\right)+\frac{P_{fix}}{\mu}T_i\right)\\ &= \frac{\mu}{D_n}\left(\frac{1}{h_i^2}2^{\frac{L_n}{BT_i}}\left(1-ln(2)\frac{L_n}{BT_i}\right)+\left(\frac{P_{fix}}{\mu}-\frac{1}{h_i^2}\right)\right)\end{aligned} \tag{24}$$

A simple way to prove that the function is not monotone is computing the limits of the function in its domain, this is, $T_i \in (0, \infty)$:

$$\lim_{T_i \to 0} \frac{\partial E_b^{total}}{\partial T_i} = -\infty < 0 \tag{25}$$

$$\lim_{T_i \to \infty} \frac{\partial E_b^{total}}{\partial T_i} = \frac{\mu}{D_n}\left(\frac{P_{fix}}{\mu}-\frac{1}{h_i^2}\right) > 0 \tag{26}$$

In (26) we assume domain knowledge to justify the sign of the expression. Indeed, the normalized channel is larger than 1 once it is compensated through antenna gains. This proves that the function is not decreasing in T_i.

For convexity, we need to compute the Hessian. By inspection, expression (24) differs from (20) by a constant term so both Hessian matrices match and this function is convex as well. □

From Proposition 2, the minimum energy is not obtained with maximum propagation time. Considering a unique link, this is, $N = 1$ and a fixed set $\mathcal{S}_n = \{n, GS\}$, we can particularize (P1) and rewrite it in terms of T_i. Without loss of generality we can assume $l_n = 0$:

$$\begin{aligned}\underset{T_i}{\text{minimize}} \quad & \frac{\mu}{D}\left(\frac{1}{h_i^2}\left(2^{\frac{D}{BT_i}}-1\right)+\frac{P_{fix}}{\mu}\right)T_i && \textbf{(P1a)}\\ \text{subject to} \quad & T_i \leq \tau_n \triangleq T_{max} && \text{(C1)}\\ & T_i \geq \frac{L_n}{Blog_2(1+h_i^2 P_{max})} \triangleq T_{min} && \text{(C2)}\end{aligned}$$

where $P_{max} = min\{P_i - E_n^{proc}/T_n^{proc}, P_{out}^{max}\}$. Figure 6 illustrates the objective of (P1a) along with both time constraints (refer to Section 5.1 for the parameters' setup). Depending on the scenario the problem may be posed in different zones: in region I (T_{max} below $E_{b,n}^{min}$) the function is decreasing in T_i and so equivalent to the case with $P_{fix} = 0$; in region II (as in Figure 6) there is an optimal time delay (and thus transmission power) such that the energy is minimized; and in region III (T_{min} above $E_{b,n}^{min}$) the energy is minimized at minimum delay, this is, transmitting at maximum power. Unlike Corollary 2, in regions II and III the minimum energy is not achieved with maximum communication time, which leads to the following corollary.

Corollary 3. *For $P_{fix} \neq 0$, the minimum energy is not obtained in general with equality in the (C6) constraint.*

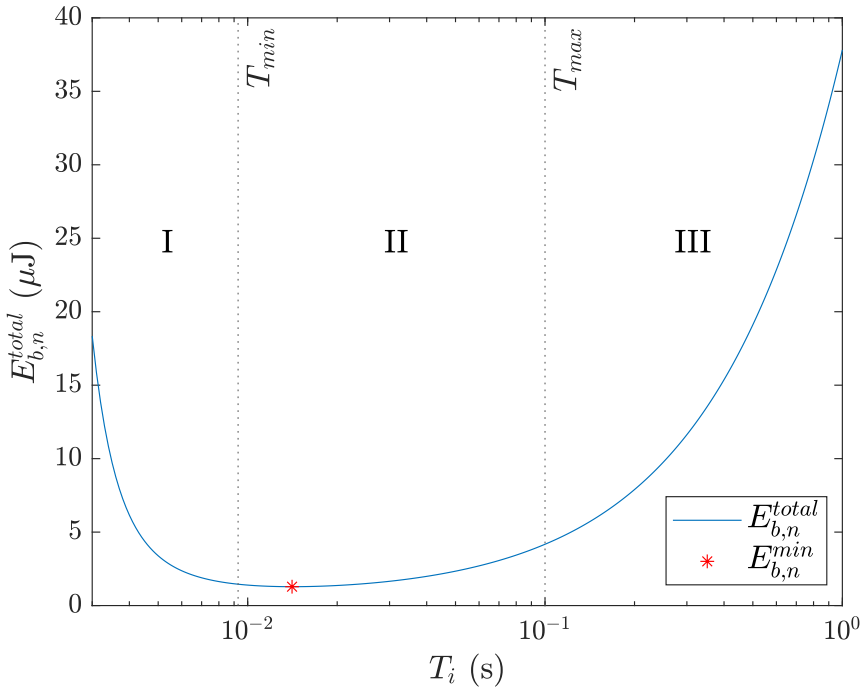


Figure 6: Energy dependence on T_i.

Nevertheless, minimizing the propagation time as Corollary 2 states is advantageous for different reasons: first, minimizing the propagation time reduces the total time delay, which is suitable. When the time constraint is inactive (i.e., regions II and III) minimizing the propagation time pushes T_{max} to the right with no effect on the energy minimization, but on reducing the total delivery time. Second, when the time constraint is active (i.e. region I), minimizing the propagation time leads to minimum energy. Heuristically, minimizing the propagation time always reduces the total delivery time and it may reduce the energy consumption when the problem is posed in region I or may not affect the energy in the other two cases. As a conclusion, minimizing the propagation time (or path-loss) will be the metric used to assess the routing procedure.

4.3 Routing algorithm

From the previous result, we will determine the sets $\mathcal{S}_n$ as the paths minimizing the propagation time. Given the set $\mathcal{N}$ with satellites having data to be transmitted and processed, we define the optimization problem (P2).

$$\begin{aligned}
&\underset{\{\mathcal{S}_n\}_{n=1}^{N}}{\text{minimize}} && \sum_{n=1}^{N}\sum_{i\in\mathcal{S}_n} T_i^{prop} && && \textbf{(P2)}\\
&\text{subject to} && \mathcal{S}_n \cap \mathcal{S}_m = \{GS\},\ \forall m \neq n && n = 1,\dots,N && \text{(C1)}\\
& && \mathcal{S}_n(1) = \mathcal{A}(n) && n = 1,\dots,N && \text{(C2)}\\
& && \mathcal{S}_n(m_n+1) = GS && n = 1,\dots,N && \text{(C3)}
\end{aligned}$$

The constraint (C1) ensures path deconfliction, this is, that any two paths share no intermediate nodes, whereas (C2) and (C3) ensure feasibility, that each path starts at the origin satellite and terminates at the GS, respectively. We assume that τ_n is much larger (e.g., some order of magnitude) than any propagation time so that all solutions remain feasible.

Finding the optimal solution to the shortest path problem with constraints is NP-hard. There is up-to-date literature dealing with path search for multiple agents with node deconfliction [43]-[47]. This has been tackled through relaxation methods to apply convex optimization [48] and via tailored algorithms [49]. A more robust approach is via submodularity, a functional tool with similar properties to convexity in the discrete domain. As opposed to the previous methods, submodularity can provide solutions that are scalable and with provable near-optimality guarantees [50].

4.3.1 Greedy maximization of submodular set functions

Definition 1. A set function $f : 2^{\mathcal{N}} \to \mathbb{R}$ defined over a set $\mathcal{N}$ is submodular if for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$ and $e \in \mathcal{N} \backslash \mathcal{B}$ it holds that

$$f(e \mid \mathcal{A}) \geq f(e \mid \mathcal{B}), \tag{27}$$

where $f(e \mid \mathcal{S}) = f(\mathcal{S} \cup e) - f(\mathcal{S})$ is termed as the marginal gain of function f at $\mathcal{S}$ with respect to e.

Definition 2. A set function $f : 2^{\mathcal{N}} \to \mathbb{R}$ is monotone if for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$

$$f(\mathcal{A}) \leq f(\mathcal{B}) \tag{28}$$

Definition 3. A set function $f : 2^{\mathcal{N}} \to \mathbb{R}$ is normalized if

$$f(\varnothing) = 0, \tag{29}$$

where $\varnothing$ is the empty set.

The objective of problems dealing with multiway partitions is to divide a set $\mathcal{S}$ into k disjoint sets $\mathcal{S}_1, \ldots, \mathcal{S}_k$, $\mathcal{S}_i \cap \mathcal{S}_j = \varnothing$, $\cup_{i=1}^{k} \mathcal{S}_i = \mathcal{S}$ such that $\sum_{i=1}^{k} f_i(\mathcal{S}_i)$ is maximized. A simple approach to solve normalized monotone submodular function maximization under a multiway partition constraint is via a greedy algorithm that allocates each element e to the best subset at that point $\mathcal{S}_j^* \leftarrow \mathcal{S}_j^* \cup \{e\}$, where $j^* = \text{argmax}_j \; f_j(\mathcal{S}_j \cup \{e\})$ and $e = \text{argmax}_{j,\; e \in \mathcal{N} \backslash \mathcal{S}_j} \; f_j(e \mid \mathcal{S}_j)$.

This greedy algorithm provides a one-half-approximation guarantee [50]. The performance of this low-complexity algorithm can be explained through the marginal gain: by definition (27), the elements added at the beginning are more relevant because they provide a larger marginal gain. Similarly, the greedy algorithm only optimizes over the next element, which maximizes the marginal gain at each iteration.

4.3.2 SHIELD algorithm

Consider a set $\mathcal{S}$ with all possible paths from all LEO satellites in $\mathcal{N}$ to the GS. The goal is using submodularity to obtain a multiway partition of N sets, such that each set contains only one path. This could be designed as a submodular minimization with cardinality constraints, though we will follow an unconstrained approach. The function

$$f(\mathcal{S}_i) = \sum_{p \in \mathcal{S}_i} T_p^{prop} + K|\mathcal{S}_i| \tag{30}$$

computes the propagation time for all paths in $\mathcal{S}_i$, where T_p^{prop} is the total propagation time for path p. Then, it adds an extra term weighted by the cardinality of the set $|\mathcal{S}_i|$, i.e., the number of paths in $\mathcal{S}_i$.

Lemma 1. *$f(\mathcal{S}_i)$ is modular.*

Proof. Consider $\mathcal{S}_j \subseteq \mathcal{S}_i \subseteq \mathcal{S}$ and $e \in \mathcal{S} \backslash \mathcal{S}_i$, it holds that

$$\begin{aligned} f(e \mid \mathcal{S}_j) &= f(\mathcal{S}_j \cup \{e\}) - f(\mathcal{S}_j) \\ &= \sum_{p \in \mathcal{S}_j} T_p^{prop} + T_e^{prop} + K|\mathcal{S}_j \cup \{e\}| \; - \sum_{p \in \mathcal{S}_j} T_p^{prop} - K|\mathcal{S}_j| \\ &= T_e^{prop} + K \end{aligned} \tag{31}$$

The result in (31) does not depend on $\mathcal{S}_j$ because f is a sum of linear set functions and, hence, linear. Thus $f(e \mid \mathcal{S}_i)$ provides the same result and the relationship in (27) is fulfilled with equality. Similarly, $-f$ also accomplishes the relationship with equality. Hence, f is submodular and supermodular and, thus, f is modular. □

It is easy to see that f is also monotone and normalized, so it is appropriate for the previously defined greedy algorithm. Even though it was defined as a maximization problem, because of the linearity, the problem can be minimized as well and the theoretical convergence is guaranteed.

The role of the last term in (30) is that it allows to generate subsets of cardinality one. The intuition behind it follows from that at a given iteration of the greedy algorithm, the minimum path may be found for a node that already has allocated a path. When only considering the propagation time, this second path would be added to the existing ones, which is undesirable. This cardinality term adds a penalty to the sets that already have a path to ensure that they are not eligible.

Theorem 1. *The multiway partition greedy algorithm computed over $f(\mathcal{S}_i)$ produces subsets of cardinality one.*

Proof. At a given iteration, consider two sets $\mathcal{S}_1 = \varnothing$ and $\mathcal{S}_2 = \{p\}$. At the next iteration there are two paths $e, e' \in \mathcal{N} \backslash \mathcal{S}_2$ that minimize the marginal gain (i.e., propagation time) of every set, respectively. Then, the function evaluation for every set is

$$f(\mathcal{S}_1 \cup e) = T_e^{prop} + K \tag{32}$$

$$f(\mathcal{S}_2 \cup e') = T_p^{prop} + T_{e'}^{prop} + 2K \tag{33}$$

Comparing both results and rearranging the terms, we can find an expression for K:

$$f(\mathcal{S}_2 \cup e') > f(\mathcal{S}_1 \cup e) \implies K > T_e^{prop} - T_p^{prop} - T_{e'}^{prop} \tag{34}$$

With a sufficiently large K, if the propagation time of p and e is smaller than e', the algorithm will never select a non-empty set. □

We introduce the original SHIELD algorithm (**S**ubmodular **HIE**rarchica**L** **D**ijkstra's algorithm) in Algorithm 1, used for routing. In practice it is not necessary to compute $\mathcal{S}$, this is, all feasible paths from $\mathcal{N}$ to the GS. For a given node, only the shortest path until the GS minimizes the marginal gain, meaning that we only need to compute the shortest path for each feasible node at every iteration. We use the Dijkstra's algorithm, which is a well-known algorithm used to efficiently compute the shortest path between two nodes in a graph [32]. Besides that, K was only used theoretically to demonstrate the correctness of the algorithm. In practice, only the empty subsets are considered at each iteration.

Since the SHIELD algorithm runs the Dijkstra's algorithm N^2 times, its time complexity can be bounded by $\mathcal{O}(N^2(\mathcal{V} + \mathcal{E}log\mathcal{V}))$.

Algorithm 1 SHIELD

Input: $\mathcal{G}$ and $\mathcal{N}$
Output: $\mathcal{S}_i$ for $i = 1, \ldots, N$

Initialize $\mathcal{S}_i = \varnothing \ \forall i$
for $i = 1, \ldots, N$ **do**
  for $j = i, \ldots, N$ **do**
    $e_j = \text{Dijkstra}(\mathcal{G}, \mathcal{N}(j), GS)$
    $j^* = \text{argmin}_j \ f_j(\mathcal{S}_j \cup \{e\})$
  end
  $\mathcal{S}_i \leftarrow \{e_{j^*}\}$
  $\mathcal{G} \leftarrow \{\mathcal{G} \setminus \mathcal{S}_i\} \cup \{GS\}$
end

Theorem 2. *SHIELD provides a Pareto efficient solution.*

Proof. An allocation $\mathcal{S}_A = \{\mathcal{S}_1, \ldots, \mathcal{S}_N\}$ is Pareto optimal if there is no $\mathcal{S}_B = \{\mathcal{S}'_1, \ldots, \mathcal{S}'_N\}$ such that $f(\mathcal{S}'_i) \leq f(\mathcal{S}_i) \, \forall i$, with $f(\mathcal{S}'_j) < f(\mathcal{S}_j)$ for some j. Without loss of generality, we assume that there are no two different paths providing the same distance.

By contradiction, suppose that there exists a set $\mathcal{S}'_j$ fulfilling $f(\mathcal{S}'_j) < f(\mathcal{S}_j)$. At iteration j the allocation S'_j cannot contain non-assigned nodes in $\mathcal{S}_A$, this is, $\mathcal{S}'_j \cap \{\mathcal{G} \setminus \bigcup_{i=1}^{j-1} \mathcal{S}_i\} = \varnothing$, because the optimality of the Dijkstra's algorithm ensures finding the path with minimum length [51]. Thus, $\mathcal{S}'_j \cap \mathcal{S}_i \neq \varnothing$ for some $i = 1, \ldots, j-1$, meaning that $\mathcal{S}'_j \neq \mathcal{S}_j$. However, the algorithm assigns nodes greedily such that $e_j = \text{argmin}_{j, \ e \in \mathcal{N} \setminus \mathcal{S}_j} \ f(e \mid \mathcal{S}_j)$, and so $f(\mathcal{S}'_j) \geq f(\mathcal{S}_j)$, which contradicts the assumption. □

4.4 Radio Resource Allocation and Offloading strategy

As a result of the routing procedure, the RRM and offloading problem is decoupled for every path $\mathcal{S}_n$. From Corollary 3, we can express the time constraint as

$$\sum_{i\in\mathcal{S}_n}^{m_n}\frac{1}{log_2\left(1+p_ih_i^2\right)} \leq \frac{B}{L_n}\underbrace{\left(\tau_n - \sum_{i\in\mathcal{S}_n}^{m_n} T_i^{prop} - l_n T_n^{proc}\right)}_{\gamma_n(l_n)} \tag{35}$$

Thus for each path $\mathcal{S}_n$ we need to determine the optimal power allocations fulfilling constraint (35). Following the initial approach, we propose problem (P4), which is a MIP. However, due to the low dimensionality of l_n, the optimal offloading corresponds to allocating optimal powers for $l_n = \{0,1\}$ and choose the l_n minimizing the energy. Thus in the following we focus on the RRM problem.

$$\underset{\{p_i\}_{i=1}^{m_n},\ l_n}{\text{minimize}} \quad E_{b,n}^{total} = \mu\frac{L_n}{D_nB}\sum_{i\in\mathcal{S}_n}^{m_n}\frac{p_i+\frac{P_{fix}}{\mu}}{log_2\left(1+p_ih_i^2\right)} + l_n\frac{1}{D_n}C_p(D_n) \tag{P4}$$

$$\text{subject to} \quad \sum_{i\in\mathcal{S}_n}^{m_n} T_i \leq \gamma_n(l_n) \tag{C1}$$

$$g_i\frac{E_n^{proc}}{T_n^{proc}} + p_i \leq P_i \qquad i=1,\dots,m_n \tag{C2}$$

$$p_i \leq P_{out}^{max} \qquad i=1,\dots,m_n \tag{C3}$$

$$l_n \in \{0,1\} \tag{C4}$$

Proposition 3. *The objective of Problem (P4) is not convex in p_i.*

Proof. The second derivative with respect to p_i is

$$\frac{\partial^2 E_{b,n}^{total}}{\partial p_i^2} = ln(2)\mu\frac{L_n}{D_nB}h_i^2\frac{2h_i^2\left(p_i+\frac{P_{fix}}{\mu}\right)+\left(h_i^2\left(\frac{P_{fix}}{\mu}-p_i\right)-2\right)ln\left(1+p_ih_i^2\right)}{\left(1+p_ih_i^2\right)^2 ln^3\left(1+p_ih_i^2\right)} \tag{36}$$

To know the sign of the numerator in (36) we can compute its limits once again in the domain of the function:

$$\lim_{p_i\to 0}\frac{\partial E_b^{total}}{\partial p_i} = \lim_{p_i\to 0} 2h_i^2\left(p_i+\frac{P_{fix}}{\mu}\right) = 2h_i^2\frac{P_{fix}}{\mu} > 0 \tag{37}$$

$$\begin{aligned}\lim_{p_i\to\infty}\frac{\partial E_b^{total}}{\partial p_i} &= 2h_i^2\frac{P_{fix}}{\mu} + \lim_{p_i\to\infty} p_ih_i^2\left(2-ln(1+p_ih_i^2)\right)+\left(h_i^2\frac{P_{fix}}{\mu}-2\right)ln(1+p_ih_i^2)\\ &= \lim_{p_i\to\infty} p_ih_i^2\left(2-ln(1+p_ih_i^2)\right) = -\infty < 0\end{aligned} \tag{38}$$

Thus, we cannot ensure the Hessian is positive semidefinite and the function is not convex nor concave. □

Even though the energy function is not mathematically convex, we see that it may be geometrically convex within the feasible region. In Figure 7(a) we represent the total energy per bit for one dimension and $h(p_i)$ is the numerator of the second derivative, this is, expression (36). Up to around 35 W the function switches convexity. Since we allow for a maximum power of 10 W (see Table 2), an interior-point algorithm can reach the optimal solution if we consider this upper bound.

4.4.1 Energy minimization via log-log convex program (LLCP)

We seek for a more robust approach, this is, that we can define an equivalent optimization problem that is always convex. We observe that both functions (energy cost and delay constraint) in (P4) are log-log convex in p_i, meaning that we can perform a bijective transformation in the domain ($p_i > 0$) such that the problem is convex. Problem (P5) is the log-log transformation of problem (P4).

$$\underset{\{p_i\}_{i=1}^{m_n},\ l_n}{\text{minimize}} \quad \tilde{E}_{b,n}^{total} = log\left(\mu \frac{L_n}{D_n B} \sum_{i \in \mathcal{S}_n}^{m_n} \frac{e^{u_i} + \frac{P_{fix}}{\mu}}{log_2\left(1 + e^{u_i} h_i^2\right)} + l_n \frac{1}{D_n} C_p(D_n)\right) \tag{P5}$$

$$\text{subject to} \quad log\left(\sum_{i \in \mathcal{S}_n}^{m_n} \frac{1}{log_2\left(1 + e^{u_i} h_i^2\right)}\right) \leq log\left(\gamma_n(l_n)\right) \tag{C1}$$

$$log\left(g_i \frac{E_n^{proc}}{T_n^{proc}} + e^{u_i}\right) \leq log\left(P_i\right) \qquad i = 1, \ldots, m_n \tag{C2}$$

$$log\left(p_i\right) \leq log\left(P_{out}^{max}\right) \qquad i = 1, \ldots, m_n \tag{C3}$$

$$l_n \in \{0, 1\} \tag{C4}$$

Log-log convex programs (LLCP) are the generalization of Generalized Geometric Programs (GGP) and the log-log transformation is equivalent to expressing the objective and constraints in log-log scale [52]. In the following we will prove that (P5) corresponds to the log-log transformation of (P4) and that it is a convex optimization problem. To do so, we will need to prove of several lemmas.

Lemma 2. *The sum of log-log convex functions is convex.*

Proof. We need to show that the log-log transformation is convex. The sum function is

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \tag{39}$$

and its log-log transformation is

$$F(u) = log f(e^{u_1}, \ldots, e^{u_n}) = log \sum_{i=1}^{n} e^{u_i}, \tag{40}$$

which is the *log-sum-exp*, a well-known convex function [53]. □

Lemma 3. *Function* $f(x) = \frac{x+a}{ln(1+x)}$ *for* $a \geq 0$ *is log-log convex.*

Proof. A necessary and sufficient condition [52] for a twice-differentiable function to be log-log convex is

$$f''(x) + \frac{f'(x)}{x} - \frac{f'(x)^2}{f(x)} \geq 0 \tag{41}$$

Analytically we can compute the derivatives to obtain that

$$\begin{aligned} &f''(x) + \frac{f'(x)}{x} - \frac{f'(x)^2}{f(x)} = \\ &\frac{2(x+a) - (x+2-a)ln(1+x)}{(1+x)^2 ln^3(1+x)} + \frac{ln(1+x) - \frac{x+a}{1+x}}{x ln^2(1+x)} - \frac{\left(ln(1+x) - \frac{x+a}{1+x}\right)^2}{x ln^3(1+x)} = \\ &\frac{x(x+a)^2 - (3x+a-2)(x+a)ln(1+x) + a(1+x)^2 ln^2(1+x)}{x(x+a)(1+x)^2 ln^3(1+x)} \geq 0 \end{aligned} \tag{42}$$

In (42) the denominator is always positive in the domain ($x > 0$), and the numerator is trivially shown to be positive. Another way to show the log-log convexity of $f(x)$ is to show that its log-log transformation is convex, which corresponds to expression (43). Figure 7(b) shows the plot of $F(u)$, which is convex.

$$F(u) = lnf(e^u) = ln\left(\frac{e^u}{ln(1+e^u)}\right) = u - ln\left(ln(1+e^u)\right) \tag{43}$$

□

Lemma 4. *Function $g(x) = \frac{1}{ln(1+x)}$ is log-log convex.*

Proof. The easiest way to show it is by means of the convexity of its log-log transformation:

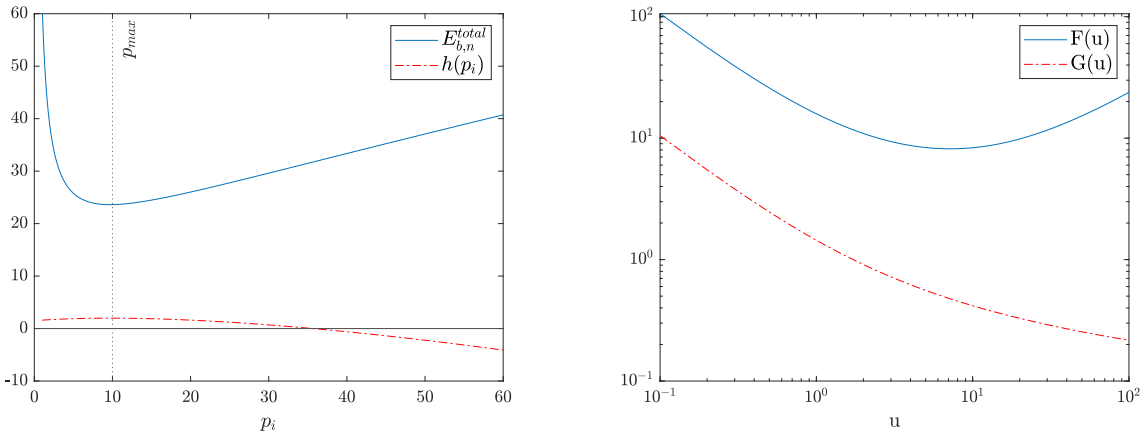
$$G(u) = lng(e^u) = ln\left(\frac{1}{ln(1+e^u)}\right) = -ln(ln(1+e^u)), \tag{44}$$

which is convex since it differs from (43) by a linear term. Similarly, the plot is depicted in Figure 7(b) to show its convexity. □

Theorem 3. *Problem (P5) is the log-log transformation of (P4) and it is a convex optimization problem in p_i.*

Proof. $\tilde{E}_{b,n}^{total}$ is the sum of several $f(x)$ of Lemma 7 and the delay function in (C1) is the sum of several $g(x)$ functions of Lemma 8. They only differ by a constant (the logarithm base) and a linear transformation of the optimization variable ($u_i \leftarrow p_i h_i^2$), which do not affect convexity. In light of Lemma 5 both functions are log-log convex. Constraints (C2) and (C3) are clearly convex, which can be also derived from Lemma 5. Thus we obtain (P5), a convex optimization problem in p_i. □

Remark 2. *At first sight the RRM optimization problem might seem to be posed at the Pareto frontier, this is, there are several power allocations that can reach the same energy consumption. However, since there are not cross-terms and the problem is convex in every variable the optimal power allocations are unique.*



(a) (b)

Figure 7: (a) Convexity of $E_{b,n}^{total}$ and (b) log-log transformation of $f(x)$ and $g(x)$.

4.5 Sat2C Algorithm

Algorithm 2 describes the overall optimization for routing, resource allocation and task offloading. The independence between the routing and the rest of tasks provides a basis for future updates without interfering in the optimality of the RRM. This is particularly useful because Sat2C is dominated by the routing algorithm: the RRM increases with N, whereas the former with N^2.

We envision its implementation as a centralized algorithm. Sat2C needs to be deployed in some agent of the network that receives the information of the satellites and sends the updates back to them with the output of the algorithm. We can assume it is deployed in one or several GEO satellites because of the following reasons: the on-board processor may have more capabilities than the one from the LEO satellites and the whole coverage of the constellation is guaranteed with only three GEO satellites. Besides, the position of the satellites can be estimated without having feedback of all satellites.

Algorithm 2 Sat2C

Input: $\mathcal{G}$ and $\mathcal{N}$
Output: $\mathcal{S}_n$, l_n, $\{p_i\}_{i=n}^{m_n}$ for $n = 1, \ldots, N$

$\mathcal{S}_i \leftarrow$ SHIELD$(\mathcal{G}, \mathcal{N})$
for $n = 1, \ldots, N$ **do**
  $\{p_i\}_{i \in \mathcal{S}_n} \leftarrow$ solve (P5)
end

5 Results

In this chapter we test the performance pf the Sat2C algorithm with respect to alternative approaches. We simulate an scenario and analyse the effectiveness of the SHIELD algorithm for routing, along with the capabilities of the whole Sat2C algorithm. These experiments have been implemented in MATLAB and the optimizations via the built-in interior-point procedures [58].

5.1 Setup

We consider a Kepler constellation of $M = 7$ orbital planes with $N_a = 20$ LEO satellites per plane. The orbital planes correspond to polar orbits, this is, perpendicular to the equatorial plane, and all deployed at the same altitude of $H = 600$ km. There are 26 GSs, whose location can be found in Table 1. We append an artificial node linked to all GSs and with null weight that allows to be conceived as the destination GS. There are $N = 14$ randomly selected satellites with EO data.

The communication parameters are adjusted according to the spectral efficiency for the DVB-S2X system [59]: the modulation and coding schemes cover the range of SNRs $[-2.85, 19.57]$ dB, such that the spectral efficiency is 0 if SNR < -2.85 dB and 5.90 if SNR ≥ 19.57 dB. Hence, the antenna design and the selection of transmission power is focused on achieving SNRs within this interval. The SNR between satellites u and v is computed in linear as

$$\mathrm{SNR}(u, v) = \frac{p_i G_{tx} G_{rx}}{\sigma^2 L(u, v) L_p^2}, \tag{45}$$

where σ^2 is the noise power and computed as $\sigma^2 = \kappa_B T_N B$, where κ_B is the Boltzmann constant, T_N is the thermal noise in Kelvin, B is the channel bandwidth in Hertz, and L_p is the pointing loss. Similarly, the antenna sizes were adjusted after setting $P_i = 10$ dB via

$$G = \eta_a \left(\frac{\pi D}{\lambda} \right)^2, \tag{46}$$

where η_a is the efficiency of the parabolic antenna, D is the diameter of the dish in meters and λ is the wavelength in meters.

The minimum distance to ground is, naturally, H. Therefore, the transmission power and antenna sizes were selected to provide maximum SNR in the ground to satellite link of $\mathrm{SNR}(h) \approx 19.57$ dB. Conversely, for the inter-plane ISL, the maximum LoS distance corresponds to expression (5).

Equivalently, the transmission power and antenna sizes were selected to provide a minimum SNR in the inter-plane ISL of $\mathrm{SNR}(d_{\max,\mathrm{ISL}}(u, v)) \approx -2.85$ dB. With this, the rate is greater than zero for any satellite that has at least one neighbor within LoS. For a communication system located in the Ka-band, Table 2 lists a configuration of parameters satisfying these requirements.

Regarding the power amplifier subsystem, we consider using a single-carrier per transponder (SCPT), which corresponds to an OBO of 0 dB and $\eta = 0.65$ [54]. According to [4],

Table 1: Locations of the Ground Stations

#	Location	Latitude	Longitude
1	Troll, Antarctic	−72.01	2.53
2	Cordoba, Argentina	78.33	15.99
3	Tolhuin, Argentina	−32.09	−63.79
4	Inuvik, Canada	−54.51	−67.19
5	Punta Arenas, Chile	68.36	−133.72
6	Nemea, Greece	−53.16	−70.90
7	Nuuk, Greenland	37.82	22.66
8	Bangalore, India	64.17	−51.73
9	Tokyo, Japan	12.97	77.59
10	Mauritius	35.68	139.75
11	Awarua, New Zealand	−20.27	57.57
12	Svalbard, Norway	−46.50	168.36
13	Tromsø, Norway	69.64	18.95
14	Vardø, Norway	70.33	30.96
15	Panama	8.55	−81.13
16	Azores, Portugal	37.80	−25.47
17	Singapore	1.34	103.83
18	Hartebeesthoek, South Africa	−25.63	28.08
19	Puertollano, Spain	38.68	−4.11
20	Dubai, United Arab Emirates	25.07	55.18
21	Fairbanks, Alaska, US	64.83	−147.71
22	Los Angeles, California, US	34.05	−118.24
23	Hawaii, US	19.58	−155.42
24	Maspalomas, Canary Islands, Spain	27.76	−15.58
25	Jeju, South Korea	33.38	126.53
26	Mingenew, Australia	−29.11	115.44

we set the power loading factor to $c_0 = \frac{\pi}{4}\eta$, $P_{fix} = 5$ W and $P_{out}^{max} = 10$ W.

With respect to the processing we use a standard configuration of $f_{CPU} = 250$ MHz, $z = 737.5$ CPU cycles/bit and $\nu = 10^{-27}$ J/Hz^3 [30, 25].

High resolution images, such as Google's Aerial Orthoimagery [55] or Planet [56] are typically used. For instance, the latter equips three cameras per satellite to capture a region 24 km × 16 km within a scene of 2560 pixels × 1080 pixels. For 16 bits of depth, this results in images around 128 Mb. However, the processing time needed for this images exceeds 60 seconds, which is inconsistent with the assumption of working with a snapshot of the network. We set the data size per image to $D_n = 1.2$ Mb, which results in a processing time below 4 seconds, and $\tau_n = 10$ s, $\forall n$. Finally, the compression ratio for images for on-board LEO satellites is set at $\rho_n = 4, \forall n$ [57].

Table 2: Communication parameters for the Kepler constellation

Parameter	Symbol	LEO to GS	Inter-plane ISL
Carrier frequency (GHz)	f	20	26
Bandwidth (MHz)	B	500	500
Transmission power (W)	p	10	10
Antenna diameter (Tx – Rx) (m)	$(D_{tx} - D_{rx})$	(0.26 – 0.33)	(0.26 – 0.26)
Antenna gain (Tx – Rx) (dB)	$(G_{tx} - G_{rx})$	(32.13 – 34.20)	(34.41 – 34.41)
Pointing loss (Tx – Rx) (dB)	L_p	(0.3 – 0.3)	(0.3 – 0.3)
Antenna efficiency (Tx – Rx) (–)	η_a	(0.55 – 0.55)	(0.55 – 0.55)
Noise temperature (K)	T_N	50	290
Noise figure (dB)	N_f	1.5	2
Noise power (dB)	σ^2	-119.32	-114.99

5.2 SHIELD Performance

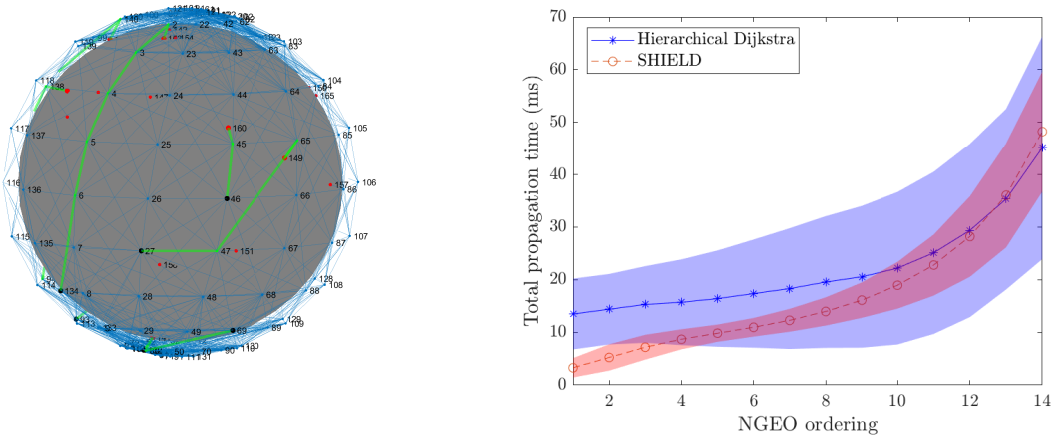
In the following we evaluate the SHIELD routing algorithm. We compare it to a simpler algorithm called Hierarchical Dijkstra, which orders $\mathcal{N}$ according to τ_n, such that $\tau_i \leq \tau_{i+1}$. Then, it finds the shortest path for every node hierarchically using the Dijkstra's shortest path algorithm. This algorithm is less expensive as it increases with N and not N^2. Figure 8(a) displays an instance of the routes determined by the SHIELD algorithm, where red nodes are GSs, black nodes are satellites with EO data and highlighted green edges are the respective paths.

Using a Monte Carlo method, both algorithms are run 1000 times and averaged. Figure 8(b) shows the mean propagation time per path according to the satellite's route assignment order and the shaded areas represent the standard deviation. The first conclusion is that it is effective to route in satellite networks, as both algorithms outperform the existing solutions (i.e., the satellite sends the data directly to the GS when it has LoS). Both algorithms produce shorter paths for early assigned routes, since in late iterations some nodes have been already assigned and the graph is more constrained. Likewise, both route lengths increase exponentially with the number of satellites. The SHIELD algorithm, however, provides shorter routes in mean and with smaller variance, which means it is a more robust and reliable algorithm. On the other hand, with the Hierarchical Dijkstra one can control which satellite receives the shortest route, even though there is high variance in the solution.

5.3 Parametric analysis

The performance of Sat2C is compared to three alternative approaches:

- **Never-Offload Policy**: this algorithm optimizes the radio resources and always computes the task locally. It is equivalent to Sat2C with $l_n = 1, \forall n$.
- **Always-Offload Policy**: this algorithm optimizes the radio resources and always



(a) (b)

Figure 8: (a) An instance of the routes determined by SHIELD and (b) mean propagation time per route depending on the satellite's route assignment order.

offloads the task to the GS. It is equivalent to Sat2C with $l_n = 0, \forall n$.

- **Maximum Power**: this algorithm transmits at maximum power (i.e., does not perform RRM) and computes the optimal task offloading decision for each route.

We use a Monte Carlo setup of 1000 samples and average the results for all routes and all instances. In other words, we will not appreciate the effect of the routing algorithm, since the energy consumption is averaged for all N paths. Figure 10(a) represents the mean total energy per bit, and clearly Sat2C outperforms the other algorithms. Notice that the Maximum Power policy consumes less than the Never-Offload and Always-Offload policies. This is because the consumption due to processing is much significant than the used for communication. Thus, under this specific scenario, optimizing the task allocation is much more relevant than optimizing the transmission power.

These results highly depend on the CPU specifications, which motivates a parametric analysis of the energy consumption. In the following, we analyse the energy consumption varying one parameter at a time and observing the optimality of Sat2C. With these experiments, we can find the conditions under which Sat2C can be implemented.

We do not show the influence of the data size D_n nor the time delay τ_n because the energy optimization is not affected by these variables. Conversely, they do affect the time delay constraint. Nonetheless, this constraint remains inactive most of the time.

5.3.1 Clock frequency

As f_{CPU} increases, the energy spent in processing does as well. In Figure 10(b), the Never-Offload policy is heavily affected by the speed of the CPU since it always computes the task locally. On the contrary, the Always-Offload algorithm remains constant because it

always offloads the task to the GS (i.e., it is not affected by the computation model). The optimal solution approaches the Never-Offload strategy when the frequency is low (computation energy is low) and the Always-Offload policy when it is very expensive to compute it locally. Around 250 MHz it heavily depends on the channel conditions, thus some tasks are computed locally and the rest are offloaded, evincing the optimality of this criterion. This finding is unexpected and suggests that the faster the processor does not correlate with energy minimization.

5.3.2 CPU cycles per bit

Both the energy and the computation time increase with z, as seen in Figure 10(c). When this parameter is low, the optimal policy degrades to the Never-Offload policy, as the task can be computed locally while satisfying the time constraint. Conversely, when the cost of computing locally is expensive, the optimal policy tends to the Always-Offload strategy.

5.3.3 Equivalent capacitance

ν relates the capabilities of the processor with the energy consumption and, so, it is a measure of the current technological development. As seen in Figure 10(d), these results are consistent with the previous ones: when this parameter is small, the energy consumed at the processor is smaller and Sat2C always computes the task locally. Otherwise, the task is offloaded to the GS.

5.3.4 Compression ratio

As Figure 10(e) shows, the Never-Offload policy decreases with ρ because the more compressed is the data, the less bits are transmitted. Besides, this parameter does not affect the processing energy. Thus, the Never-Offload policy shows how the communication energy is reduced as the processing compresses the information. Sat2C degrades to the Always-Offload policy when the data is not compressed, as the satellite would spend energy processing and then transmit the same amount of information. However, once the data is somewhat compressed, Sat2C offloads some of the tasks to the GS and some are computed on-board, providing minimum energy consumption.

5.4 Current technological development

As stated previously, these results are highly affected by the capabilities of the on-board processor. Even though the data size does not affect the energy consumption per bit, it does affect the optimal offloading decision: for instance, for $D = 10$ Mb, the time delay cannot be ensured with local processing and the task is always offloaded to the GS. Thus, in the following we analyse the current technological development of the CPU consumption.

For these experiments we set $T^{proc} = 4$ s and $z = 737.5$ cycles/bit, $\forall n$. With expression (11), we find the required CPU frequency. We consider data sizes ranging $D = [10^{-3}, 10]$

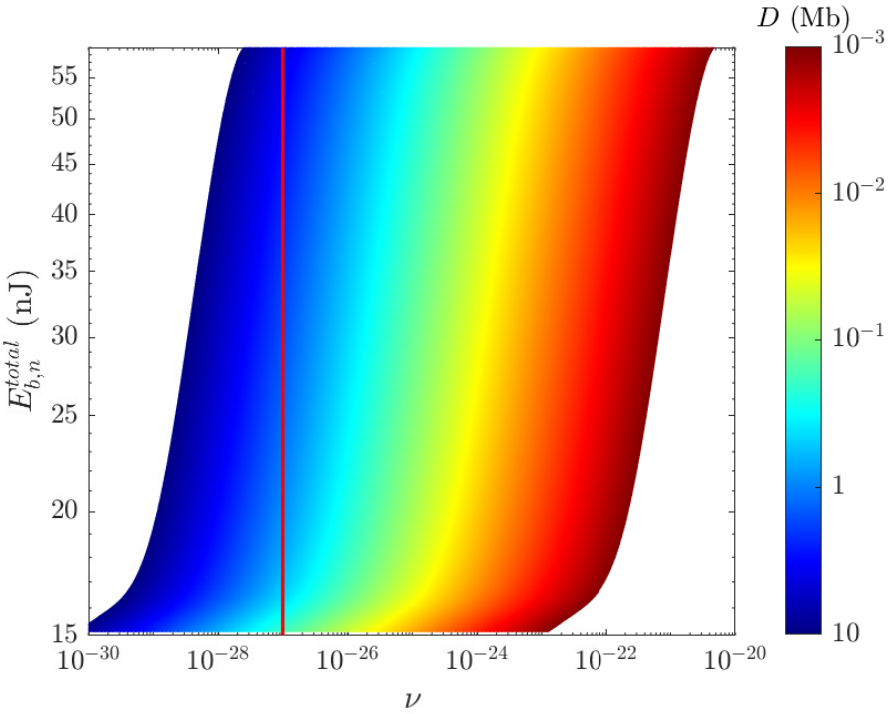


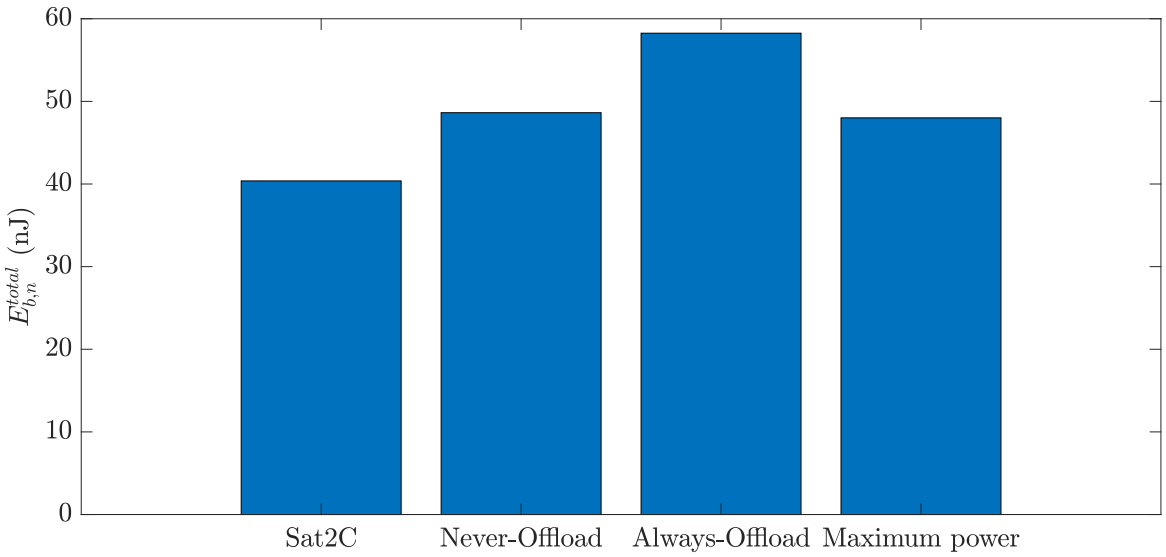
Figure 9: Data sizes that support satellite computation for current CPU technologies.

Mb. Larger data sizes require $f_{CPU} > 2$ GHz, which is unattainable for current off-the-shelf processors.

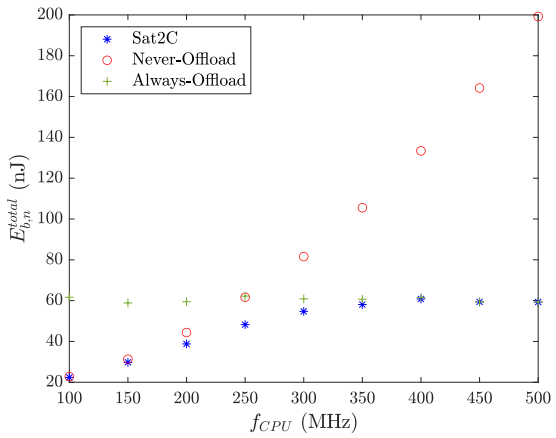
In Figure 9 we represent the energy consumption of Sat2C for different ν and D, for the cases where the algorithm does not degrades to the Always-Offload policy. In this way we can analyse in which scenarios the satellites process data locally, this is, we can dimension what kind of data fits with Sat2C. The red vertical line denotes the current development in CPUs ($\nu = 10^{-27}$): the right side contains attainable energy consumptions, whereas the left side has not been achieved yet for current processors.

For light data (e.g., $D = 10$ kb), there is a wide range of ν where Sat2C decides to compute some tasks locally. Conversely, for $D = 10$ Mb the plot is located at the left side, meaning that Sat2C will always offload the task to the GS. In conclusion, for $D \geq 10$ Mb, the current processors cannot process data locally in order to meet the time delay constraints for dynamic LEO constellations.

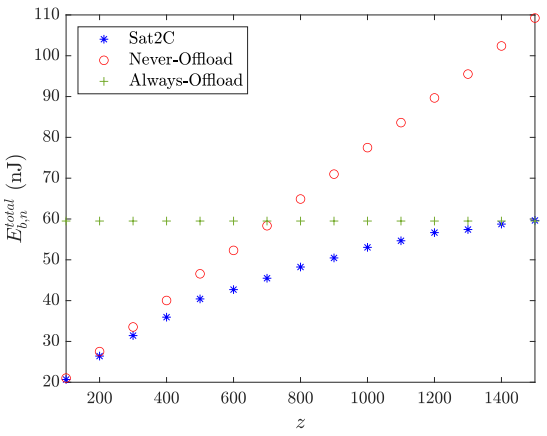
Increasingly, many services require large batches of data and the current processors are not enough to satisfy the needs of this scenario. A different approach for large amounts of data is to distribute the processing among satellites, which is equivalent to moving towards the right in Figure 9. Along with that, an active research issue is to develop new acceleration engines capable of supporting more demanding applications in the latest mobile systems on chips [60].



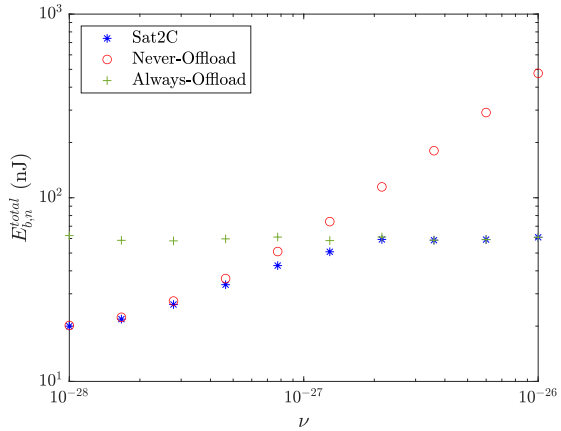
(a)



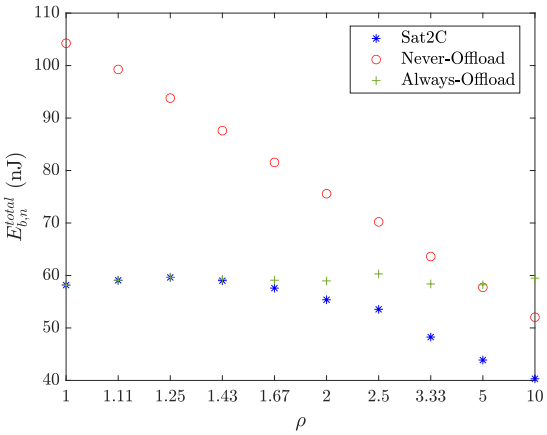
(b)



(c)



(d)



(e)

Figure 10: (a) $E_{b,n}^{total}$ for four different algorithms and (b-e) $E_{b,n}^{total}$ dependence on different parameters.

6 Conclusions and Future Development

In this thesis we have examined the ongoing state of the art of non-terrestrial LEO networks for Earth Observation, which will be a fundamental driver for the next generation networks in the path to 6G. The increasing performance of on-board processors allows to conceive satellites as edge computing devices, beyond a relay system to the data centers. Besides, the current environmental concern emphasizes the development of green technologies. All these considerations lead to the formulation of Sat2C, an energy-efficient algorithm for joint routing, radio resource management and task offloading optimization for delay-sensitive services.

Particularly, we have developed SHIELD, an original routing algorithm for dense LEO constellations based on submodularity, and the RRM and data offloading tasks are solved via log-log convex programs. In the former, we formulate a greedy algorithm, that provides one-half-approximation guarantee, with low complexity and low variance in the mean path distance; the latter provides the optimal transmission powers and data offloading for each satellite to minimize the energy consumption. Apart from these innovative mathematical tools, the novelty of this research resides in a precise model of the energy consumption for satellite networks, including the power amplifier module, which lacks in most of the literature. Overall, the optimization problem we have formulated and analysed in this research was not studied before in the literature, leading to a submission of a publication in the 2021 IEEE Vehicular Technology Conference.

In the era of Big Data, we have concluded that satellites cannot process large batches of data on-board and meet the delay requirements at the same time. This motivates the distribution of the computation among satellites, coined as partial offloading. This complicates the optimization because the data has to be scheduled among the nodes. A further study could assess the trade-off in energy consumption for partial offloading, since a more sophisticated algorithm is more consuming.

Moreover, with the adoption of FSO, communication between satellites is faster and produces less interference. Nevertheless, both technologies will coexist because many constellations that are already deployed use RF links. Under these heterogeneous conditions, the scenario changes dramatically: FSO links provide rates with several orders of magnitude larger than RF, at the expense of a repointing that it is not instantaneous. This is critical for LEO constellations, because of the dynamism of the network.

We forecast that these studies should be supported by the modelling of the network via graphs. This will allow to take advantage of the architecture and ease the joint optimization of routes and resources, apart from including new capabilities (e.g., caching data). For example, by means of graph signal processing, a clustering algorithm could be designed to determine which satellites process data and which forward only. More broadly, research is also needed to determine the usefulness of GNNs for resource optimization.

Since the network may be used for other services beyond EO, the simultaneous optimization of several objectives would be a fruitful area for further work. Apart from energy minimization, time minimization is suitable for delay-constrained applications. Conversely, some companies may aim to minimize the DL rate for several considerations:

the GS may be running other tasks, so we seek to minimize its workload; by reducing DL rate (i.e., data in the DL), more satellites can communicate with the GS; if the terrestrial infrastructure does not belong to the satellite company, reducing the workload of the GS corresponds to a reduction in OPEX.

References

[1] X. Lin, S. Cioni, G. Charbit, N. Chuberre, S. Hellsten, and J. Boutillon, "On the Path to 6G: Low Orbit is the New High," *CoRR*, vol. abs/2104.10533, 2021.

[2] M. Kakemizu and A. Chugo, "Approaches to Green Networks," *Fujitsu scientific technical journal*, vol. 45, 10 2009.

[3] F. Alagoz and G. Gur, "Energy Efficiency and Satellite Networking: A Holistic Overview," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1954–1979, 2011.

[4] J. Joung, C. K. Ho, and S. Sun, "Spectral Efficiency and Energy Efficiency of OFDM Systems: Impact of Power Amplifiers and Countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 208–220, 2014.

[5] O. Blume, D. Zeller, and U. Barth, "Approaches to energy efficient wireless access networks," in *2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 1–5, 2010.

[6] Ofcom, "Technology Futures Spotlight on the technologies shaping communications for the future," tech. rep., Ofcom, United Kingdom, January 2021. Accesed: 08-06-2021.

[7] ESA, "PAZ SAR satellite mission of Spain." `earth.esa.int/web/eoportal/satellite-missions/p/paz`. Accesed: 08-06-2021.

[8] A. Harati-Mokhtari, A. Wall, P. Brooks, and J. Wang, "Automatic Identification System (AIS): Data Reliability and Human Error Implications," *Journal of Navigation*, vol. 60, no. 3, p. 373–389, 2007.

[9] M. N. Sweeting, "Modern Small Satellites-Changing the Economics of Space," *Proceedings of the IEEE*, vol. 106, no. 3, pp. 343–361, 2018.

[10] A. da Silva Curiel, L. Boland, J. Cooksley, M. Bekhti, P. Stephens, W. Sun, and M. Sweeting, "First results from the disaster monitoring constellation (DMC)," *Acta Astronautica*, vol. 56, no. 1, pp. 261–271, 2005. 4th IAA International Symposium on Small Satellites for Earth Observation.

[11] D. G. Tyc, K. Ruthman, and J. Steyn, "The RapidEye Low Cost Mission Design," in *55th International Astronautical Congress of the International Astronautical Federation*, pp. 1–11, 2004.

[12] Planet, "Our Constellation." `storage.googleapis.com/planet-ditl/day-in-the-life/index.html`. Accesed: 08-06-2021.

[13] ICEYE. `www.iceye.com`. Accesed: 08-06-2021.

[14] Z. Sun, *Satellite Networking: Principles and Protocols*. Wiley Publishing, 2nd ed., 2014.

[15] J. Sedin, L. Feltrin, and X. Lin, "Throughput and Capacity Evaluation of 5G New Radio Non-Terrestrial Networks with LEO Satellites," 2020. arxiv.org/pdf/2012.02136.pdf. Accesed: 09-06-2021.

[16] G. P. T. Board, "Edge Computing for 5G Networks," tech. rep., January 2021.

[17] I. Leyva-Mayorga, B. Soret, and P. Popovski, "Inter-Plane Inter-Satellite Connectivity in Dense LEO Constellations," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2021.

[18] OneWeb, "How OneWeb's system works." www.youtube.com/watch?v=8_kytEDxC0A&t=4s, December 2020. Accesed: 09-06-2021.

[19] SES, "Our Coverage." www.ses.com/our-coverage#/explore. Accesed: 09-06-2021.

[20] M. Wall, "SpaceX's 60-Satellite Launch Is Just the Beginning for Starlink Megaconstellation Project," *Space*, May 2019. Accesed: 09-06-2021.

[21] J. Wei, J. Han, and S. Cao, "Satellite IoT Edge Intelligent Computing: A Research on Architecture," *Electronics*, vol. 8, no. 11, 2019.

[22] Z. Zhang, W. Zhang, and F.-H. Tseng, "Satellite Mobile Edge Computing: Improving QoS of High-Speed Satellite-Terrestrial Networks Using Edge Computing Techniques," *IEEE Network*, vol. 33, no. 1, pp. 70–76, 2019.

[23] J. Wei and S. Cao, "Application of Edge Intelligent Computing in Satellite Internet of Things," in *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, pp. 85–91, 2019.

[24] X. Gao, R. Liu, and A. Kaushik, "Virtual Network Function Placement in Satellite Edge Computing with a Potential Game Approach," 2021. arxiv.org/pdf/2012.00941.pdf. Accesed: 09-06-2021.

[25] Y. Wang, J. Zhang, X. Zhang, P. Wang, and L. Liu, "A Computation Offloading Strategy in Satellite Terrestrial Networks with Double Edge Computing," in *2018 IEEE International Conference on Communication Systems (ICCS)*, pp. 450–455, 2018.

[26] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Energy-Efficient Offloading for Mobile Edge Computing in 5G Heterogeneous Networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.

[27] Y. Wang, J. Yang, X. Guo, and Z. Qu, "A Game-Theoretic Approach to Computation Offloading in Satellite Edge Computing," *IEEE Access*, vol. 8, pp. 12510–12520, 2020.

[28] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic Computation Offloading for Mobile-Edge Computing With Energy Harvesting Devices," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3590–3605, 2016.

[29] K. Cheng, Y. Teng, W. Sun, A. Liu, and X. Wang, "Energy-Efficient Joint Offloading and Wireless Resource Allocation Strategy in Multi-MEC Server Systems," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2018.

[30] A. Ndikumana, N. H. Tran, T. M. Ho, Z. Han, W. Saad, D. Niyato, and C. S. Hong, "Joint Communication, Computation, Caching, and Control in Big Data Multi-Access Edge Computing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 6, pp. 1359–1374, 2020.

[31] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic Joint Radio and Computational Resource Management for Multi-User Mobile-Edge Computing Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5994–6009, 2017.

[32] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[33] X. Wang, B. Qian, and I. Davidson, "On constrained spectral clustering and its applications," *Data Mining and Knowledge Discovery*, vol. 28, pp. 1–30, Jan 2014.

[34] R. Zhang, X. Cheng, Q. Yao, C.-X. Wang, Y. Yang, and B. Jiao, "Interference Graph-Based Resource-Sharing Schemes for Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 4028–4039, 2013.

[35] Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "Graph Neural Networks for Scalable Radio Resource Management: Architecture Design and Theoretical Analysis," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 101–115, 2021.

[36] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[37] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, 2013.

[38] A. P. Miettinen and J. K. Nurminen, "Energy Efficiency of Mobile Clients in Cloud Computing," HotCloud'10, p. 4, USENIX Association, 2010.

[39] W. Q. Lohmeyer, R. J. Aniceto, and K. L. Cahoy, "Communication satellite power amplifiers: current and future SSPA and TWTA technologies, journal = International Journal of Satellite Communications and Networking," vol. 34, no. 2, pp. 95–113, 2016.

[40] J. Joung, C. K. Ho, and S. Sun, "Spectral Efficiency and Energy Efficiency of OFDM Systems: Impact of Power Amplifiers and Countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 208–220, 2014.

[41] Q. Zeng, Y. Du, K. Huang, and K. K. Leung, "Energy-Efficient Radio Resource Allocation for Federated Edge Learning," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020.

[42] J. W. Rabjerg, I. Leyva-Mayorga, and B. Soret, "Exploiting topology awareness for routing in LEO constellations," 2021. `arxiv.org/pdf/2006.12242.pdf`. Accesed: 11-06-2021.

[43] M. S. Hughes, B. J. Lunday, J. D. Weir, and K. M. Hopkinson, "The multiple shortest path problem with path deconfliction," *European Journal of Operational Research*, vol. 292, no. 3, pp. 818–829, 2021.

[44] F. Grenouilleau, W.-J. van Hoeve, and J. N. Hooker, "A Multi-Label A* Algorithm for Multi-Agent Pathfinding," in *ICAPS*, vol. 29, pp. 181–185, May 2021.

[45] G. Sharon, R. Stern, A. Felner, and N. R. Sturtevant, "Conflict-based search for optimal multi-agent pathfinding," *Artificial Intelligence*, vol. 219, pp. 40–66, 2015.

[46] S. Almagor and M. Lahijanian, "Explainable Multi Agent Path Finding," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, (Richland, SC), p. 34–42, International Foundation for Autonomous Agents and Multiagent Systems, 2020.

[47] H. Ma, G. Wagner, A. Felner, J. Li, T. K. S. Kumar, and S. Koenig, "Multi-Agent Path Finding with Deadlines," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pp. 417–423, International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[48] F. He and K. Dai, "A Shortest Path Algorithm with Constraints in Networks," in *Applied Informatics and Communication* (J. Zhang, ed.), pp. 598–604, Springer Berlin Heidelberg, 2011.

[49] R. Ananthalakshmi Ammal, P. C. Sajimon, and V. C. S S, "Canine Algorithm for Node Disjoint Paths", booktitle="Advances in Swarm Intelligence," pp. 142–148, Springer International Publishing, 2020.

[50] E. Tohidi, R. Amiri, M. Coutino, D. Gesbert, G. Leus, and A. Karbasi, "Submodularity in Action: From Machine Learning to Signal Processing Applications," *IEEE Signal Processing Magazine*, vol. 37, no. 5, pp. 120–133, 2020.

[51] K. Subramani, "Correctness of Dijkstra's algorithm." `community.wvu.edu/~krsubramani/courses/fa05/gaoa/qen/dijk.pdf`. Accesed: 23-06-2021.

[52] A. Agrawal, S. Diamond, and S. Boyd, "Disciplined geometric programming," *Optimization Letters*, no. 13, pp. 961–976, 2019.

[53] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[54] K. N. R. Rao, *Satellite Communications: Concepts and Applications*. PHI Learning, 2013.

[55] Google, "Google Maps Imagery," tech. rep. Accesed: 19-06-2021.

[56] Planet, "Planet Imagery Product Specifications," tech. rep., February 2021. Accesed: 19-06-2021.

[57] G. Yu, T. Vladimirova, and M. N. Sweeting, "Image compression systems on board satellites," *Acta Astronautica*, vol. 64, no. 9, pp. 988–1005, 2009.

[58] MATLAB, *version R2020a*. The MathWorks Inc., 2021.

[59] ETSI, "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)," standard, France, October 2014.

[60] W. Pan, Z. Li, Y. Zhang, and C. Weng, "The New Hardware Development Trend and the Challenges in Data Management and Analysis," *Data Science and Engineering*, vol. 3, pp. 263–276, Sep 2018.