A Unified Optimisation Framework for QoS Management and Congestion Control in VHTS Systems

Tomaso de Cola, *Member, IEEE*, Mario Marchese, *Senior Member, IEEE*, Maurizio Mongelli, *Member, IEEE*, and Fabio Patrone, *Member, IEEE*

***Abstract*—The design of Very High Throughput Satellite (VHTS) systems will be an important milestone for future telecommunications systems in the light of the ongoing integration between satellite and 5G technologies as well as with the upcoming 6 G evolution. To this end, satellite systems will be operating at EHF frequency bands, hence requiring a special attention to the overall network design in terms of ground and space segment optimisation. In particular, the mitigation of feeder link outage events through advanced prediction techniques and the exploitation of smart gateway diversity concepts will be combined with an overarching network resources' orchestration, jointly conceived to meet the QoS requirements of the end users. In this respect, this paper proposes a novel optimisation framework able to achieve very high throughput system figures, building on an adaptive QoS management concept. The proposed framework has been compared with other existing solutions from the literature and validated by means of extensive simulation campaigns, whose collected results show significant performance improvements.**

***Index Terms*—Congestion control, feeder link outage, QoS management, satellite communications.**

I. Introduction

THE conception of the 5G communication paradigm represents an important shift for the telecommunication industry [1] in that it will provide existing and enable new services with unprecedented performance figures such as pervasive content distribution at very high data rates, ubiquitous data delivery, connectivity for massive deployments of sensors, ultra-low latency services, just to cite a few. Moreover, the scientific community is also discussing the expectations for the forthcoming generation (i.e., 6 G), which will aim at more extreme service requirements. This technology evolution is the necessary response to the ever-increasing data traffic volume and the widespread use of mobile devices as confirmed by the traffic forecast reports from CISCO [2], claiming the advent of the Zettabyte era. Many countries are targeting data rates of at least 100 Mbit/s offered to all users by 2025, independently of the specific geographic areas, i.e. in rural as well as in metropolitan areas. However, in spite of the large national investments and the important technology advances testified in the last years, providing very large data rates to all users according to an *anytime-anywhere* paradigm will be hardly achievable because of the digital divide still experienced in some regions and because of the inherent capacity limitations of the wireless terrestrial technology.

In order to circumvent the performance constraints imposed by the present terrestrial technology, satellite systems have been receiving consensus in the light of their capability to offer very high data rate to a vast population independently of the specific users' location [3]. This concept is also supported by the ongoing 3GPP Rel. 17 standardisation where non-terrestrial networks (NTN) are identified as fundamental pillars to achieve most of the planned 5G-performance [4]. In particular, the role of satellite communications is considered particularly relevant with respect to eMBB services (enhanced Mobile Broadband). To this regard, the deployment of Very High Throughput Satellite (VHTS) systems will be instrumental to achieve such performance requirements [5], by means of tight convergence with the (wireless) terrestrial infrastructure. In more detail, providing the 5G (and beyond) ecosystem with a platform able to provide Tbps capacity [6] is the ultimate goal, whose achievement is however conditioned to the solution of specific technology challenges. First of all, providing high data rates requires re-engineering the overall satellite system design in both ground and space segments. In particular, the exploitation of EHF (Extremely High Frequency, 30–300 GHz) frequency band is taken as the reference approach to meet such a requirement, by specifically moving user links to Ka-band (20 – 40 GHz) and operating the feeder links at higher frequency bands, which are offering larger portions of spectrum currently inutilized [7]. This frequency allocation plan, however, introduces important design challenges in the links operating at EHF, which may be affected by important outage events as a result of harsh atmospheric conditions (e.g., heavy rain). Secondly, the integration of terrestrial and satellite networks [8] requires efficient resource allocation strategies, able to adapt to the dynamically fluctuating conditions of traffic and network conditions.

These technology challenges have pushed the scientific community to work out sophisticated design options, which orbit around the concept of smart-ground diversity (SGD) [9],

Manuscript received April 16, 2020; revised June 25, 2020; accepted July 19, 2020. Date of publication August 10, 2020; date of current version October 22, 2020. The review of this article was coordinated by Prof. Pascal Lorenz. *(Corresponding author: Tomaso de Cola.)*

Tomaso de Cola is with the Deutsches Zentrum für Luft- und Raumfahrt e.V., Wessling 82234, Germany (e-mail: tomaso.decola@dlr.de).

Mario Marchese and Fabio Patrone are with the University of Genoa, 16145 Genoa, Italy (e-mail: mario.marchese@unige.it; f.patrone@edu.unige.it).

Maurizio Mongelli is with the National Research Council, 16149 Genoa, Italy (e-mail: maurizio.mongelli@ieiit.cnr.it).

Digital Object Identifier 10.1109/TVT.2020.3015252

allowing to deploy a number of ground stations sufficient to target very high levels of service availability (i.e., $>99.5\%$) at limited bandwidth redundancy. Moreover, advanced orchestration [10]–[12] concepts combined with novel re-routing concepts have enabled new QoS management frameworks, able to guarantee the demanded performance figures. As a matter of fact, the satellite scientific community and industry have worked hard in the last decade to explore these concepts and to conduct important experimentation campaigns, by dedicating particular attention to the design of the ground segment from a physical layer point of view, whereas less importance has been given so far to the higher protocol layers. Consequently, a comprehensive system analysis is missing. In particular, very few scientific works have addressed, at the best of the authors' knowledge, the problem of providing an efficient QoS management framework to VHTS systems. The limited existing literature elaborates mostly on the case of a few interconnected gateways and investigates the overall QoS system performance only partially.

To bridge the existing scientific gaps, this paper provides a holistic framework to address QoS management in VHTS systems when sophisticated link outage prediction algorithms have to be combined with efficient re-routing algorithms aimed at minimising network congestion events. To this end, the framework drafted in [13] has been enriched and extensively validated through simulation campaigns, which have proved the superiority of the proposed concept with respect to other existing schemes. It is worth noting that the proposed framework has been validated by taking into consideration fixed users, although its rationale applies to the case of mobile users too without loss of generality.

The remainder of this paper is organised as follows. Section II surveys the current state of the art and outlines the main contribution of this paper. Section III introduces the reference scenario, while the proposed strategies with respect to link outage prediction and smart gateway diversity concepts are elaborated in Section IV. The performance analysis is in turn presented in Section V, whereas the final remarks and the main lessons learnt are drawn in Section VI.

II. Overview

A. State of the Art

As introduced in Section I, a key requirement for the future generation of satellite systems is the capability to offer Tbps capacity [14] and hence to make use of EHF frequency band. To this end, the ground segment has to implement the consolidated concepts of smart-gateway diversity (SGD) [9] that have been largely investigated in the last decade [5], [15], [16]. In this respect, different concepts have been elaborated to determine the optimal number of gateways. On the one hand, 'N+P' gateway concept has been worked out to allow for the instantiation of a number of redundancy gateways (P) in addition to the nominal ones (N) with a slight increase in the network complexity because of the interconnection between gateways through a terrestrial backbone. On the other hand, an alternative approach to reduce CAPEX/OPEX costs on the ground segment is to avoid using backup gateways and to provide instead each user with additional bandwidth, of which a portion will be used only upon feeder link outage, i.e. to support link handover procedures. As such, this option is typically referred to as 'N+0' SGD configuration. A drawback is that it requires more sophisticated signalling processing functions in both in ground and space segments. The implementation of an SGD-based architecture has engineering implications in the design both of the ground segment and of the space segment, especially for what concerns carrier multiplexing and corresponding handover procedures at beam level, as analysed in [15]. Moreover, an overall system view is required to get a proper resource allocation strategy, which can follow different theoretical concepts such as the matching theory as discussed in [17].

Besides the ground segment design, a special note has to be dedicated to the integration of the satellite system [18] with the rest of the infrastructure, with a special emphasis on networking aspects. To this regard, some attention has been given by the scientific community to the end-to-end architecture supporting the integration of terrestrial and aerial platforms in a broad sense (i.e., including satellites too) [3]. In particular, the introduction of SDN/NFV concepts is becoming more and more compelling to boost the convergence of satellite and terrestrial infrastructures [10]. Different architecture solutions have been worked out in many research and project initiatives. As such, the programmability of all network nodes involved in the telecommunication chain has been addressed to getting a more effective service and system performance in terms of data traffic key performance indicators such as throughput and delay [19]. Last but the least, some attention must be dedicated also the use of multi-access edge computing (MEC) and in particular of edge caching concepts that have been extensively considered in mobile networks [20] and are receiving increasing interest for the application in satellite systems. These networking aspects automatically point to the definition of a suitable QoS management framework able to bring together the requirements coming from both satellite and terrestrial networks and translate them into a suitable optimisation approach. Given the complexity and high dimensionality of this problem, only a few studies have attempted to work out a proper networking approach towards QoS management both with respect to SDN-enabled gateway positioning as in [21], [22] and in terms of handover procedures as in [23], [24]. Additionally, [25] addressed this problem by introducing network coding to improve the robustness of data communication against link outage and proposed an analytical framework to assess the delay performance of traffic flows, though assuming only two gateways. [12] provided other insights into the problem mostly focusing on the migration of virtual functions into the *cloud* and on the assessment of the overall performance. [11] provided a general framework.

B. Main Paper Contributions

According to the available state of the art, important research gaps deal with the definition of a comprehensive framework able to capture the main architecture aspects and the corresponding networking implications. In the attempt to address some of the underlying technology challenges, the main contributions of this

TABLE I
TABLE OF ACRONYMS

Abbreviation	Definition
3GPP	Third Generation Partnership Project
CBR	Constant-Bit-Rate
DVB	Digital Video Broadcasting
DWAS	Decreased Weight Association Scheme
eMBB	enhanced Mobile BroadBand
EHF	Extremely High Frequency
GEO	Geostationary Earth Orbit
IMT	Interference Mitigation Techniques
ModCod	Modulation and Coding
NCC	Network Control Center
NFV	Network Function Virtualization
NT	Normalized Throughput
PLR	Packet Loss Rate
NTN	Non-Terrestrial Network
OF	Open Flow
QoS	Quality of Service
RLNC	Random Linear Network Coding
SDN	Software-Defined Networking
SGD	Smart Gateway Diversity
SINR	Signal-to-Interference-plus-Noise Ratio
SOT	SINR Outage Threshold
UOS	Unified Optimised Strategy
UPC	Uplink Power Control
VHTS	Very High Throughput Satellite
VIM	Virtual Infrastructure Manager
WRR	Weighted Round Robin

paper are to extend the preliminary findings reported in [13] and, in particular, to highlight the following aspects:

- To propose a satellite system architecture built on an SGD configuration and enabling end-to-end orchestration functionalities by means of SDN concepts;
- To propose an optimisation framework to select alternate satellite gateways on the basis of an effective link outage prediction strategy. The developed framework is aimed at optimising the overall system throughput, by also mitigating the occurrence of congestion events on the involved gateways;
- To show the effectiveness of the proposed approach with respect to existing solutions by means of realistic simulation campaigns.

C. Main Acronyms

To ease the comprehension of the rest of the paper and the recall of the considered concepts, we include a table (Table I) containing the main mentioned acronyms and the related meanings.

III. Reference Scenario

A. Satellite System Architecture

A multi-beam star satellite network built on a Geostationary Earth Orbit (GEO) satellite platform is the reference scenario considered in this paper.

The satellite feeder link is operated in EHF frequency band (30–300 GHz), whereas the user link is allocated in Ka-band (i.e. $20 - 30$ GHz). The forward link implements the DVB-S2 standard[1] whereas the return one applies DVB-RCS2. The main scientific challenge is to cope with the important propagation impairments introduced by EHF frequency band. The use of classical interference mitigation techniques (IMT) such as uplink power control (UPC) and data rate adaptation (through ModCod) is not sufficient. On the contrary, smart-gateway diversity turns out to be a more efficient approach as it allows exploiting additional resources provided by the ground segment, without necessarily introducing excessive link margin that would result in large bandwidth waste. To this end, the overall ground segment is partitioned in clusters of gateways, wherein the corresponding gateways are interconnected through a fiber optic backbone. Each cluster is defined in such a way that its gateways are separated by at least 100 Km (to minimise the correlation of rain fading). Each gateway is associated to one local Network Control Center responsible for the control of the available feeder link resources utilisation on the basis of the periodically collected measures of the feeder link Signal-to-Interference-plus-Noise Ratio (SINR) (e.g., by means of the received satellite beacon). A *Central Node* connected to the local NCCs is playing the role of *GW manager* and is in charge of predicting future outage events by exploiting the information received from the NCCs and of consequently selecting the alternate gateways on the feeder link handover procedure. Without loss of generality, the rest of the paper is focused on the analysis of a single cluster of gateways, as shown in Fig. 1, which also contains details specified in the remainder of the paper.

B. Integration of Satellite and Terrestrial Networks

A special note has to be dedicated to the integration of the satellite with the terrestrial networks, so that the incoming traffic flows can be efficiently treated and the corresponding data packets forwarded to the most appropriate of the N available gateways ($G_1, \ldots, G_N$), each of them associated with the corresponding local NCC ($NCC_1, \ldots, NCC_N$), in order to guarantee performance requirements. In this respect, dynamic and automatised feeder link selection is of paramount importance. This action can be affected by current and future network status. This requirement translates into deploying a programmable satellite ground segment infrastructure, built on Software Defined Networking (SDN) [26] and Networking Function Virtualisation (NFV) [27], allowing to dynamically set up different routing paths for different traffic flows. To this end, the Central Node also includes a SDN controller responsible for canalizing the traffic flows to any of the available N GWs through N SDN-enabled switches ($S_1, \ldots, S_N$), see Fig. 1, encountered along the path. These nodes store the received forwarding table in a dedicated table (*flow table*): as soon as a new traffic flow enters the network, the involved SDN switch applies the forwarding rule available in its flow table. The specific rule computation is carried out by the SDN-controller on the basis of the specific application requirements (e.g., maximum acceptable delay, minimum required throughput, maximum tolerated loss,

[1] DVB-S2x standard can be applied too, with no impact on the networking concepts elaborated in the next section.

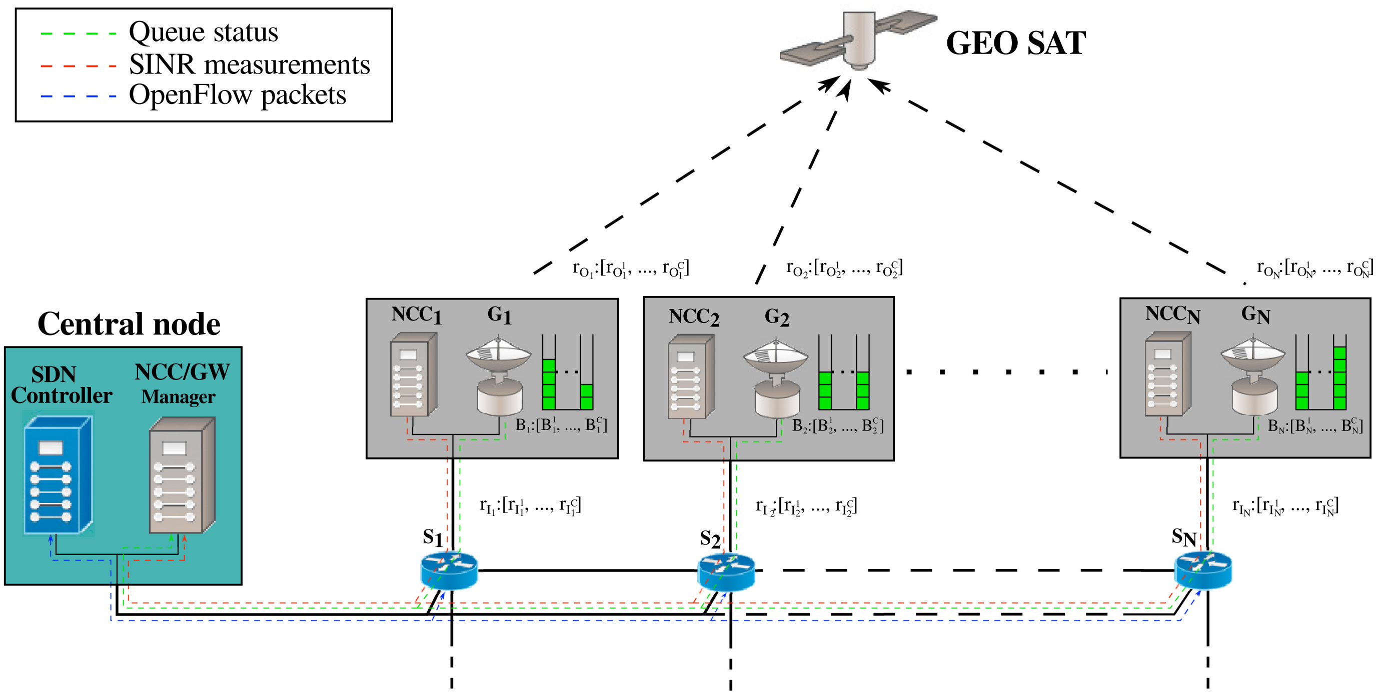


Fig. 1. Smart Gateway Diversity scenario, single cluster of gateways.

etc.) provided by the overarching network management architecture through the so called *north-bound* interface. In more detail, prediction of link outage, estimation of congestion occurrences in GWs, and packet re-routing strategy are implemented as virtual functions on top of the SDN Controller, which exposes its functionalities (flow/route map and routing manager) to the north-bound applications through the Virtual Infrastructure Manager (VIM). The re-routing engine is responsible for the identification of the traffic flows whose routing path needs to be modified and then notifies the Routing manager/Flows Rule Generator, creating the rules for the SDN switches. Network status information is instead received from the SDN switches through the so-called south-bound interface typically making use of the OpenFlow (OF) [28] protocol. In this respect, it is worth noting that the OpenFlow messages already existing have been considered without any modification, hence making the system more appealing from an implementation view point. Yet, measured SINR satellite links values will be used by the central node to perform feeder link outage prediction and consequently select the alternate gateways to forward the incoming traffic after the link outage event.

The diagram architecture of The Central Node is depicted in Fig. 2.

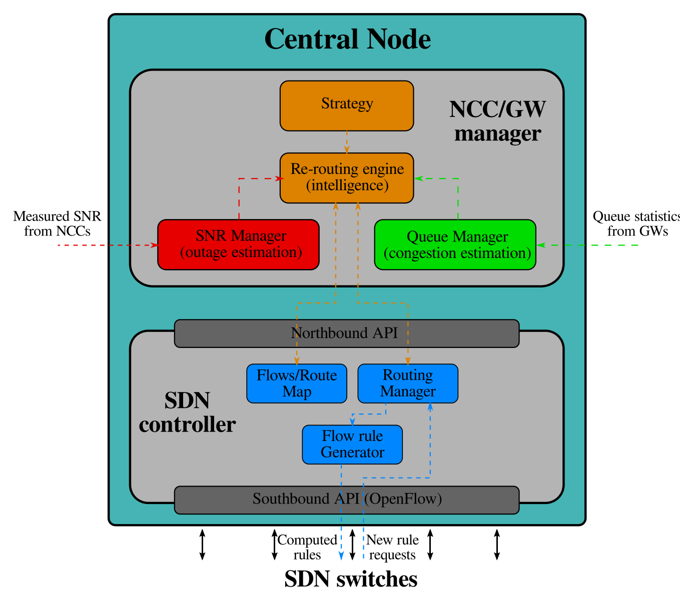


Fig. 2. Central Node architecture.

C. Data Traffic Characterisation

We assume the satellite network as deputed to transport eMBB-services, which are categorised into C traffic priority classes depending on the peculiarities of the specific traffic flows (i.e., web-browsing, data transaction, video streaming, etc.). As mentioned in the introduction, the focus of this paper is on content distribution to fixed users, although the optimisation framework here proposed can be applied to the case of mobile users too. In this latter scenario, however, additional QoS requirements coming from the peculiarities of vehicular communications should be considered too. Moreover, a more sophisticated satellite design concept should be also considered in order to take possible performance degradation into account such as those introduced by beam handover, doppler effects, and other propagation impairments just to cite a few. All these aspects are not addressed in this paper and are subject to future investigations.

The flows belonging to a given traffic class c are characterised by an average bit rate r_f^c and a minimum guaranteed throughput r_g^c. As such, the cumulative average bit rate of all flows insisting to a given gateway G_n can be formulated as $r_{In} = \sum_{c=1}^{C} r_{In}^c$, with $r_{In}^c = \sum_{i=1}^{F_n^c} r_{fi}^c$, r_{fi}^c and F_n^c being the average bit rate of any traffic flow i belonging to service class c and the total number of traffic flows, respectively. The output data rate r_On of each gateway can be computed according to the offered symbol rate and ModCod set according to the measured SINR.

Another key parameter to control the overall system performance is the occupancy of the buffer implemented in any G_n. It is assumed that each gateway implements a pool of C buffers, each of them assigned to a given class of service c. The overall buffer occupancy can be defined as $B_n = \sum_{c=1}^{C} B_n^c$, B_n^c being the buffer allocated to the traffic flows belonging to service class c. Each buffer will be served through an appropriate scheduling policy, such as Weighted Round Robin (WRR), in order to assure a proper portion of the overall available bandwidth to each class c, such that $r_On = \sum_{c=1}^{C} r_On^c$.

IV. Theoretical Framework

A. Aim and Scope

This section introduces the main building blocks of the overall system design, aimed at optimising the performance in terms of overall offered throughput. In particular, specific details about the tools necessary to support efficient handover procedures and, accordingly, packet re-routing procedure will be provided in the next sections in terms of the adopted outage prediction algorithm and routing process.

B. Feeder Link Outage Prediction

The prediction of an upcoming outage event consists in tracking the evolution over time of the SINRs samples received by the GW and, accordingly, assessing whether, in a future time instant, SINR will go below a given threshold.[2] To this end, the prediction algorithm based on Linear Regression (LR) already discussed in [29] is taken as a reference here for its simplicity and effectiveness. For the sake of completeness, the main features of the algorithm are shortly summarised in the following.

Let $s(t)$ denote the SINR sample received at time instant t. The computation of the SINR prediction in a future time instant is based on the information vector $I(t) = [s(t-T+1), \ldots, s(t-j), \ldots, s(t)]$, T being the size of the observation window over the past and $j = 0, \ldots, T-1$. Assuming that the prediction algorithm is used to assess the link quality (i.e., outage or not) at time instant $t + t_d$ and that γ_{th} corresponds to the link outage threshold, called SINR outage threshold (SOT), then the generic prediction function can be expressed as $f(I(\cdot), \cdot)$. An outage event occurring at $t + t_d$ is foreseen at time t if $f(I(t), t + t_d) < \gamma_{th}$. The linear regression strategy applies on the information vector $f(I(\cdot), \cdot)$ at each time instant t. In particular, the coefficients for the slope $m(t)$ and the intercept $q(t)$ are computed according to a classical least-squares approach as follows:

$$m(t) = \frac{T \cdot \sum_{j=0}^{T-1} j \cdot s_j(t) - \sum_{j=0}^{T-1} j \cdot \sum_{j=0}^{T-1} s_j(t)}{T \cdot \sum_{j=0}^{T-1} j^2 - \left(\sum_{j=0}^{T-1} j\right)^2} \tag{1}$$

$$q(t) = \frac{\sum_{j=0}^{T-1} s_j(t) - m(t) \cdot \sum_{j=0}^{T-1} j}{T} \tag{2}$$

Once the parameters have been updated at time t, the prediction s^* at any time t^* in the future ($t^* > t$) is $s^*(t^*) = m(t) \cdot t^* + q(t)$. The corresponding prediction is derived by verifying whether $s^*(t + t_d) \leq \gamma_{th}$ (outage) or not (no outage). It is worth noting that the prediction accuracy is sensitive to the specific time lag t_d in the future, i.e. larger t_d, higher the prediction error.

C. Gateway Handover Strategy

We take the $N + 0$ SGD approach as a reference, so that no additional backup gateways are necessary. All the deployed gateways are operational and make available their allocated bandwidth also to the flows initially handled by other gateways temporarily subject to feeder link outage. Moreover, it is assumed that all gateways nominally exploit all available bandwidth, differently from what elaborated in other research contributions where spare bandwidth or additional inactive carriers were considered (e.g. in [25], [6]). The reason for this assumption is to carry out a performance analysis of the system under stress, i.e to check the capability of the system to effectively react to outage events and accordingly take the proper decision in terms of alternate gateway selection.

The objective of the proposed gateway handover strategy, called Unified Optimised Strategy (UOS), is twofold. On the one hand, it is aimed at distributing the affected traffic flows to alternate gateways in a such a way to meet the performance requirements of each traffic class c in terms of the minimum guaranteed throughput r_g^c. On the other hand, it pursues the exploitation of the available feeder link bandwidth as long as available (i.e. until an outage event occurs), provided that this does not impact the data traffic requirements, i.e. in terms of packet loss or delay. Practically speaking, the gateway manager estimates the available link data rate $r_O^* n^c(p)$ at a time instant $t = p$ offered to a given class of service c from a gateway n on the basis of the measured $SINR_n$ and of the ModCod to be adopted according to the DVB-S2 standard specification. The gateway manager also computes the average input rate of the data traffic belonging to any class c and destined to gateway n, denoted as $r_I n^c$. Hence, the objective of the handover strategy is to select the gateway $G_{\bar{n}}$ able to maximise the availability of buffer space over time denoted as Λ_n^c, computed as the difference between the weighted average of the output data rate and the corresponding input rate, subject to the constraint of guaranteeing minimum throughput per class of service c. Mathematically speaking, it yields:

$$\bar{n} = \arg\max_{n \in \mathcal{GW}} [\Lambda_n^c]; \Lambda_n^c = \sum_{p=0}^{P} \alpha_p r_O^* n^c(p) - r_I n^c$$

$$\Lambda_n^c \geq r_g^c \tag{3}$$

[2]This threshold is representing the minimum link quality to allow successful symbol synchronisation and signal demodulation/decoding.

where the average output data rate is computed over $P+1$ samples and each sample is weighted by factor α_p, which has to be conveniently selected so that $\alpha_p \in [0,1]$, $p = 0,\ldots,P$, $\sum_{p=0}^{P} \alpha_p = 1$. $\mathcal{GW}$ denotes the set of gateways that can be associated to the traffic flows being processed by the gateway manager. Such a set contains all the gateways belonging to the reference cluster, with the exclusion of the ones currently or predicted for the near future under feeder link outage.

Once the SINR s_k measured on a given feeder link k gets smaller, the offered data rate r_{Ok} is reduced accordingly through the application of ModCod at lower spectral efficiency, possibly implying $r_{Ik} \geq r_{Ok}$ and eventually resulting in a congestion event (i.e., B_k getting saturated). Under these conditions, a re-allocation of one or more traffic flows to one or more alternate GWs would relieve the congestion situation and possibly avoid packet losses due to imminent buffer overflow. Actually, a decision about the traffic flows to be re-allocated is taken on the basis of the $r_O^* n^c$ knowledge. As such, the estimation of future SINR values according to the feeder link outage prediction algorithm (described in Section IV-B) allows also predicting future congestion events and reacting in advance. In practice, the GW manager periodically updates all Λ_n^c values and initializes the *traffic flow re-allocation procedure* in case of predicted congestion by re-allocating the minimum number of traffic flows. First, it selects the traffic class of the GW that will give rise to a congestion event, i.e. the c_k-th class of the k-th GW such that $\Lambda_k^{c_k} < 0$. Then, it re-selects an alternate GW among the ones not congested according to (3). The most recently flows injected into the network will be the first re-allocated traffic flows in order to minimize the frequency of the re-allocation process and, consequently, the total number of re-allocated flows. Consequently, the flow tables implemented in the SDN switches will be updated so that the new incoming packets belonging to the re-allocated flows will be forwarded to the new selected GWs, hence contributing to lower r_{Ik}.

When at time instant t the prediction algorithm detects an upcoming outage event on the feeder link n at $t + t_d$, the SDN controller has to properly react, on one hand, to allow the affected G_n to empty its buffer before the outage begins and, on the other hand, to avoid wasting the available satellite bandwidth so to go on sending packets as long as possible. Then, the GW manager triggers the *handover procedure* as follows: 1) it removes G_n from the set of available gateways $\mathcal{GW}$, then 2) it estimates the amount of data λ_n which G_n can send out before the predicted outage event takes place:

$$\lambda_n = \sum_{c=1}^{C} \lambda_n^c = \sum_{c=1}^{C} \int_t^{t+t_d} r_O n^c(\tau)\, d\tau \tag{4}$$

If $\lambda_n \geq B_n$, the SDN switches keep forwarding packets to G_n, since the data volume currently stored in the gateway buffer is less or equal than the one the outgoing link capacity is able to accommodate. On the contrary, if $\lambda_n < B_n$, the re-allocation process starts re-allocating all traffic flows from G to one or more alternate $G_i \in \mathcal{GW}$. It is important to note that if $\lambda_n > B_n$ when a link outage is already predicted, G_n will not be able to send all the packets stored in its buffer in time before the outage begins.

TABLE II
SIMULATED SCENARIO DESIGN PARAMETERS

Number of GWs N	5
GW buffer size	10 Gb
Number of priority classes C	3
Flow guaranteed throughput r_g^c	[8 4 2] $Mbps$
Packet inter-generation time	[25 50 100] ms
Packet size	200 kb
Flow duration	1 200 s
Number of predicted samples P	10
Size of the observation window T	10
Predicted sample weights α_p	$\frac{1}{P+1} = \frac{1}{11}$
SINR outage threshold (SOT)	−2.35 dB
Simulation duration	100 $days$

To avoid a long waiting time until the link outage finishes or any packet drop, a *packet re-routing process* is invoked so to move packets from G_n to one or more alternate $G_i \in \mathcal{GW}$. In practice, the amount of data belonging to the c-th priority class needing re-routing is defined as $\sigma_n^c = \lambda_n^c - B_n^c$, with $\sigma_n^c > 0$. The selection of the alternate GW eligible to receive the re-routed packets of class c is computed according to (3). Further to this, an additional control is performed on the selected GWs' buffer occupancy to avoid packet losses due to buffer overflow.

As soon as the $SINR_n$ measured on the feeder link n temporarily on outage is again above the threshold, the outage event is considered concluded after performing a hysteresis loop to avoid any 'false positive'. As such, G_n is considered available again and re-inserted in the $\mathcal{GW}$ set.

V. PERFORMANCE EVALUATION

A. Simulation Setup

To simulate the scenario depicted in Fig. 1, we used a discrete event simulator written in Python based on the simulation library Simpy.[3]

The design parameters of the considered scenario are summarized in Table II.

The simulated traffic flows are Constant-Bit-Rate (CBR) flows[4] whose parameters are reported in Table II. Traffic flows' start times are randomly generated with a uniform distribution for the entire simulation duration. Each flow's priority class is also randomly generated with a uniform distribution. Since we are interested in assessing UOS which acts on the feeder link, simulated traffic is generated from the ground segment and forwarded up to the GEO satellite. User link is not included in the simulation environment.

The α_p values are set equal for all the considered samples, i.e. current and estimated values contribute to the traffic flow

[3]https://simpy.readthedocs.io/en/latest/

[4]Additional tests with VBR-traffic flows have been carried out, essentially confirming the conclusions derived from the results obtained from the CBR traffic flows. These have been omitted from the present paper because of the lack of space.

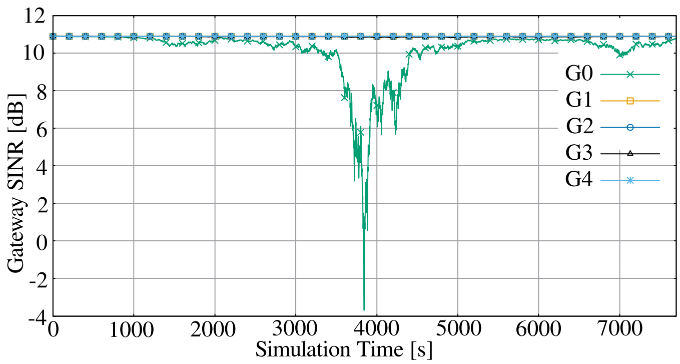


Fig. 3. GWs' SINR traces.

association and re-allocation with the same weight. Samples are picked up with $1s$ granularity. We selected $P = 10$ samples and consequently $t_d = 10$ seconds.

The link outage threshold is set accordingly to the minimum SINR tolerable with the strongest ModCod available from DVB-S2 (QPSK 1/4).

Rain attenuation has been modeled as a first order Gauss Markov process of the Ornestein-Uhlenbeck type according to [16]. N different and independent attenuation traces (expressed in dB) have been generated for links operating at central frequency of 50 GHz. An SINR trace has been consequently derived from each attenuation trace on the basis of typical satellite system parameters to close the link budget. Finally, the simulator computes the feeder link offered data rate on the basis of the corresponding SINR sample and a dedicated bandwidth of 400 MHz by taking the spectral efficiency of DVB-S2 ModCods as a reference.

In order to better highlight the network behaviour in case of outage events, the first part of our investigation addressed the specific dynamics of the system during the various phases of satellite system operations, i.e. in preparation, during, and after an outage event. Fig. 3 shows the GWs SINR trend in the considered time window. In this timeframe, one of the GWs (G_0) is affected by an increasing rain attenuation, which eventually leads to feeder link outage, while the others are still operative.

Looking at G_0's SINR behaviour, it starts fluctuating in the proximity of the outage event and then drops until it reaches the outage threshold. To mitigate the impact of fluctuations, we employed a hysteresis mechanism that prevents the frequent transitions between *outage* and *non-outage* states and vice versa by "merging" together the shortest outage intervals into a single one.

B. Other Handover Strategies Used for Comparison

The performance obtained by using UOS, has been compared with other three strategies available in the literature and considered here as benchmarks:

- *Static strategy:* when an outage event is predicted for the GW G_n, the Central Node always re-allocates the traffic flows from G_n to another GW, selected a priori without any optimisation methodology. In more detail, such a fixed selection can follow different criteria, such as the geographical distance between GWs where no information about the current or predicted status of the network is needed. Likewise, the GW selection to perform packet re-routing after outage is carried out in the same way. This strategy offers a simple outage event management even if can have drawbacks such as traffic flow performance degradation and congestion on the selected GWs.
- *Network Coding-based strategy:* it essentially consists in the application of *random linear network coding (RLNC)* to a block of data packets, with the aim of protecting it from packet loss possibly occurring in the case of link outage. Let Q_n^c denote the number of class c's packets stored in the G_n buffer: once the GW manager predicts an outage event in G_n, m^c network coding packets are generated as a linear combination of the k^c original data packets, $k^c = Q_{\mathcal{N}}^c$, $c = 0, \ldots, C$, and the configuration of m^c is dynamically computed. In more detail, the GW manager periodically estimates the amount of data λ_n (and, λ_n^c consequently) that G_n can transmit before the predicted outage begins according to (4). If $\lambda_n > B_n$, the SDN switches keep forwarding packets to G_n; when $\lambda_n \leq B_n$, the outage procedure is triggered. The number of generated network coding packets is computed as $m^c = \mathcal{Q}(B_n^c - \lambda_n^c)$, where the function $\mathcal{Q}(x)$ calculates the smallest number of stored packets exceeding x. In this way, assuming a reliable outage prediction estimation, the first $\mathcal{Q}(\lambda_n^c)$ packets stored in G_n will be directly sent out by G_n, whereas the m^c network-coded packets will be forwarded through another GW selected according to (3). Finally, once the outage starts, the packets still stored in the G_n buffer are dropped. At the receiver side, decoding a network-coded packet block is considered successful if at least k^c out of the transmitted $k^c + m^c$ packets for each class c are correctly received, otherwise, the block decoding fails and the corresponding packets are considered lost. It is immediate to see that this strategy allows high robustness against packet loss due to outage at the cost of a slight increase in the traffic volume. Hence, particular attention has to be paid to the actual value of m^c to keep an efficient trade-off between increased traffic volume and packet loss avoidance.
- *Drop strategy:* when an outage event is predicted for G_n, the GW manager re-allocates the traffic flows from G_n to another GW selected according to (3). If G_n buffer is not empty when the outage event starts, all the stored packets are dropped, i.e. no packet re-routing among GWs is possible. This strategy aims to reduce the traffic volume which "horizontally" traverses the ground segment, though at risk of increasing the overall packet loss and hence degrading users' QoS.

C. Performance Comparison

The obtained performance has been assessed by considering two typical performance metrics: *Normalized Throughput (NT)* and *Packet Loss Rate (PLR)*. Gateway SINR values and considered window time will be the ones reported in Fig. 3. Fig. 4 shows

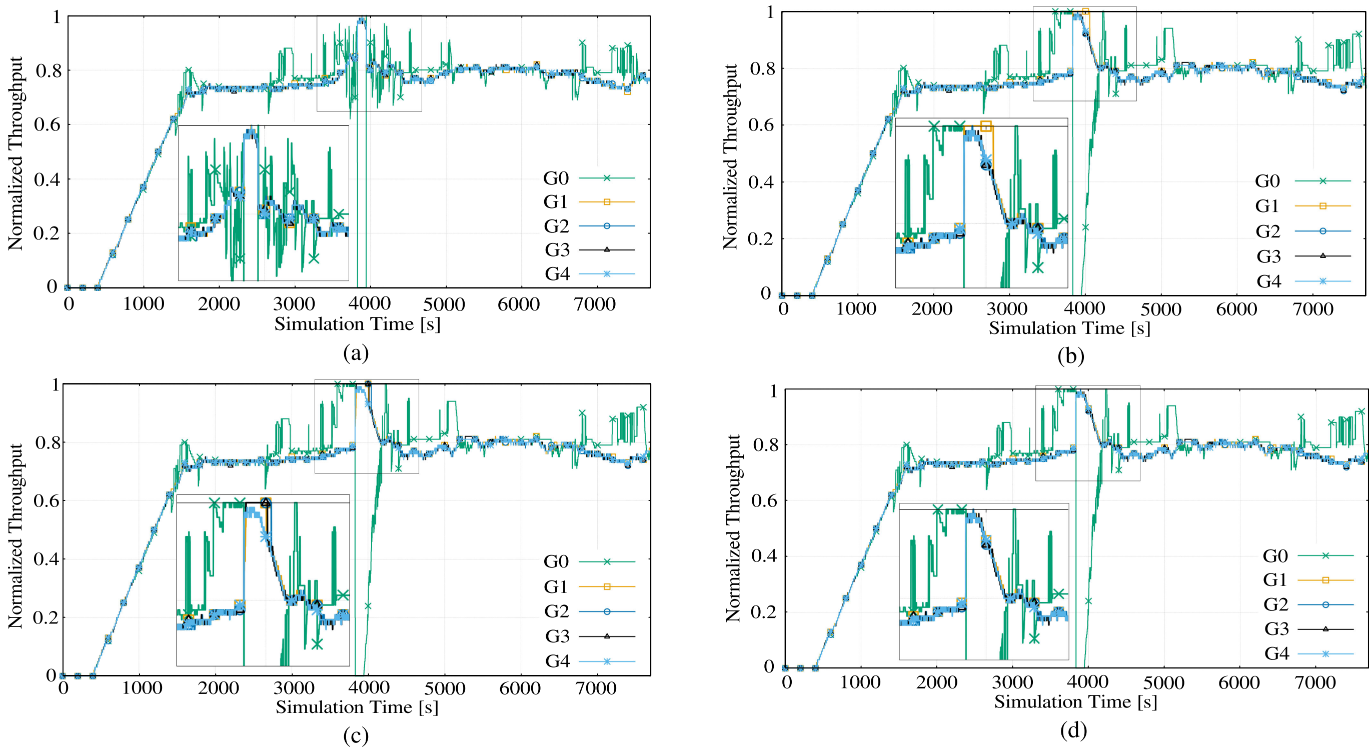


Fig. 4. NT for the GW uplink channels: Unified Optimised Strategy (UOS) (a), static (b), network-coding based (c), and drop (d) solutions.

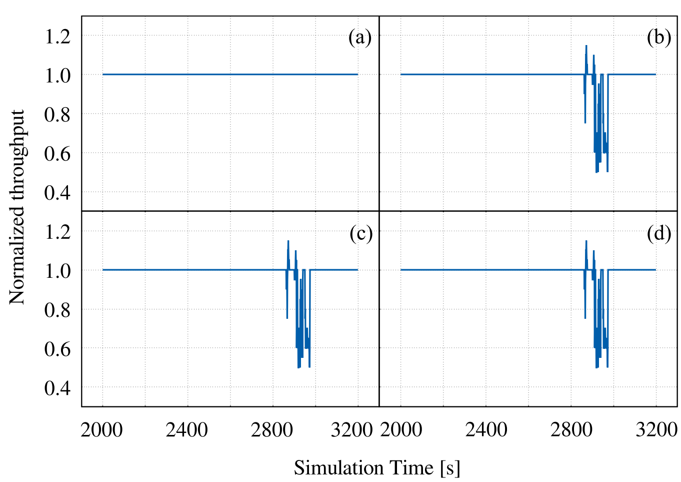


Fig. 5. NT of one of the traffic flows associated to G_0 when it is affected by congestion: OUS (a), static (b), network-coding based (c), and drop (d) solutions.

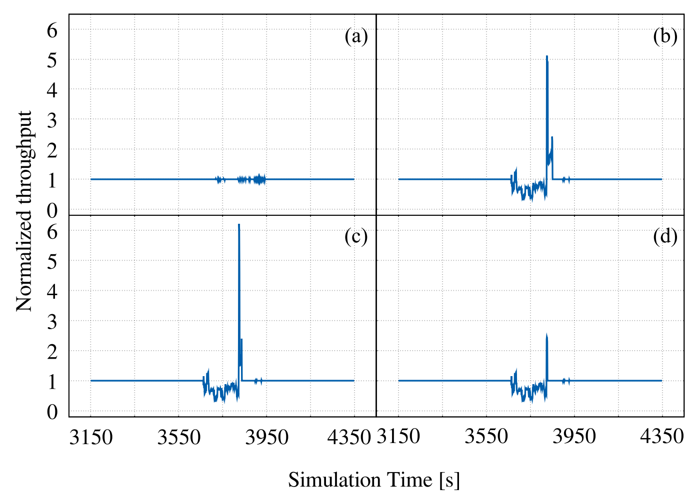


Fig. 6. NT of one of the traffic flows associated to G_0 when its uplink is affected by outage: OUS (a), static (b), network-coding based (c), and drop (d) solutions.

the NT measured for GW uplink channels. Fig. 5 shows the NT of one of the traffic flows associated to G_0 when it gets congested. Fig. 6 shows the NT of one of the traffic flows associated to G_0 when there is a link outage.

As it can be observed, when all the GWs have the same SINR values, the GW manager homogeneously distributes the new incoming traffic flows among all the GWs, as confirmed by the GW NT values in Fig. 4. When G_0 channel quality decreases, G_0 NT starts fluctuating due to its unbalanced load situation which can lead to possible congestions in G_0, such as at about 1500 s and between 2500 s and 3000 s. Our proposed solution is the only one able to detect this rate imbalance and keep guaranteeing each traffic flow required throughput r_g^c, whose values are reported in Table II. Moreover, when G_0 feeder link is about to experience the outage, the channel quality rapidly decreases leading to congestion due to the significant unbalance between overall input and output rates. Looking at NT values taken by G_0 right before the outage starts, they never reach the maximum value if OUS is used (Fig. 4(a)), hence avoiding an increase of the GW buffer occupancy and, consequently, possible losses due to buffer overflow. As soon as the link outage starts, G_0

is no longer able to send packets and its NT drops to zero, while the value of NT observed for the other GWs has different trends, depending on the considered strategy (Fig. 4). This also leads to performance variations from a traffic flow viewpoint, as highlighted in Fig. 6 and Fig. 5 and commented below.

Concerning OUS, the GW manager starts re-allocating traffic flows as soon as it detects an imminent congestion event, even when an outage event has not been predicted. Besides, when an outage event is predicted, all data packets possibly still stored in G_0 buffer are homogeneously distributed to all the other GWs upon the start of the link outage. It has to be reminded again that also under this configuration, the required throughput of the analysed traffic flow is guaranteed both in case of congestion (Fig. 5(a)) and in case of outage (Fig. 6(a)). No packet loss has been observed in this case.

Concerning the static strategy, it does not consider traffic flow re-allocation due to congestion if an outage has not been predicted, with consequent performance decrease (Fig. 5(b)). The traffic flow re-allocation takes place only during the time interval between the outage prediction and the outage beginning. All packets temporarily stored in G_0 buffer which cannot be directly transmitted to the satellite will be forwarded to another GW (G_1), independently of its buffer occupancy and channel quality. This leads to a rapid increase of G_1 NT, as clear in Fig. 4(b). The analysis of NT values in Fig. 6(b) confirms this behaviour, where the experienced low values are due to the congestion situation occurring in G_0 before the link outage and conversely the peak is due to the re-routing of the block of packets still stored in G_0 buffer when the link outage starts. An overall network PLR of 0.06% has been measured.

As far as the network coding-based strategy is employed (NT for each GW in Fig. 4(a)), the traffic flow re-allocation takes place only during the time interval between the outage prediction and the link outage start, with the consequent performance decrease due to congestion not related to outage previously described commenting the static strategy (Fig. 5(c)). All packets stored in G_0 buffer are not re-routed, and the generated network coding packets per traffic class are sent to other GWs. In this case, the three packet blocks, one for each defined priority class, are forwarded to three different GWs (G_1, G_2, and G_3). Distributing the packets to more GWs than to a single one as done by the static strategy reduces the impact of the congestion event, hence offering better performance as can be seen looking at the higher peak in Fig. 6(c). A lower overall network PLR of 0.03% has been measured.

Also when the drop strategy is used (NT for each GW in Fig. 5(b)), the traffic flow re-allocation takes place only during the time interval between the outage prediction and the link outage start, with the performance decrease in case of congestion not related to outage already highlighted describing the previous two strategies (Fig. 5(d)). All packets stored in G_0 buffer are not re-routed and the ones which cannot be sent before the link outage are dropped. This obviously affects the value of NT which increases to compensate the link outage experienced at G_0 and eventually leads to a severe performance degradation of the traffic flow influenced by the link outage, as can be seen from the values of NT depicted in Fig. 6(d). An overall network PLR of 0.09% has also been measured.

When the link outage terminates and G_0 link becomes active again, the value of NT at G_0 with OUS almost instantly reaches the one experienced by the other GWs, hence avoiding any waste of G_0 link bandwidth. On the contrary, when the other strategies are applied, the recorded value of NT at G_0 gradually increases and then decreases until it reaches the balanced pre-outage situation.

D. Assessment of the Proposed OUS Strategy

Some design parameters can influence the behaviour of OUS, such as the accuracy of the outage prediction and the load burden due to traffic flow re-allocation and data packets re-routing, consequently affecting the obtained performance. In this subsection, we investigate the behaviour of our strategy in dependence of different values of the design parameters in order to find out the configuration ensuring the most satisfactory performance.

The first parameter we consider is the size of the observation windows T. Fig. 7 shows the obtained NT of the GW uplink channels with different T values $[T = 10,\ 20,\ 30,\ 40\ s]$.

Increasing T a higher number of past samples are taken into account in the estimation of SINR and achievable output rate. This can provide a more precise estimation of these variables, but it could also lead to a lower estimation accuracy depending on the variables' statistics.

Looking at the behaviour of NT for all GWs depicted in Fig. 4, no significant performance differences are recorded by varying this design parameter. For the sake of completeness, we also analysed additional information related to mean overall NT, NT of the GW affected by outage, number of reallocated traffic flows, lost and re-routed packets, summarised in Table III for the same T values reported in Table II.

These data indicate very similar performance in terms of mean overall NT and mean NT for G_0 (as indicated in the 2^{nd} and 3^{rd} rows of the table, respectively). Similar considerations also hold for the load in terms of the number of re-allocated traffic flows due to predicted congestion and outage events (as indicated in the $4th$ and $5th$ rows of the table, respectively). The prediction accuracy and the proper reaction to both congestion and outage events offered by OUS are further confirmed by the obtained packet losses due to buffer overflow ($6th$ row of the table) and by the number of re-routed packets stored in G_0 buffer at the time the link outage starts ($7th$ row of the table). However, the presence of re-routed packets in the $T = 40$ case let us infer that further increasing T value could actually lead to lower estimation accuracy, possibly due to large-scale non-linear behaviour of the estimated variable (i.e., SINR), and to consequent erroneous timing in the traffic flow re-allocation decisions. In order to explore further this expected trend, we run additional tests with increased values of T, ranging from 100 to 400, whose performance results are reported in Table IV.

It can be noticed that the number of re-routed packets is higher due to the fact the designed strategy was able to predict a substantial degradation of G_0 link SINR but was not able to

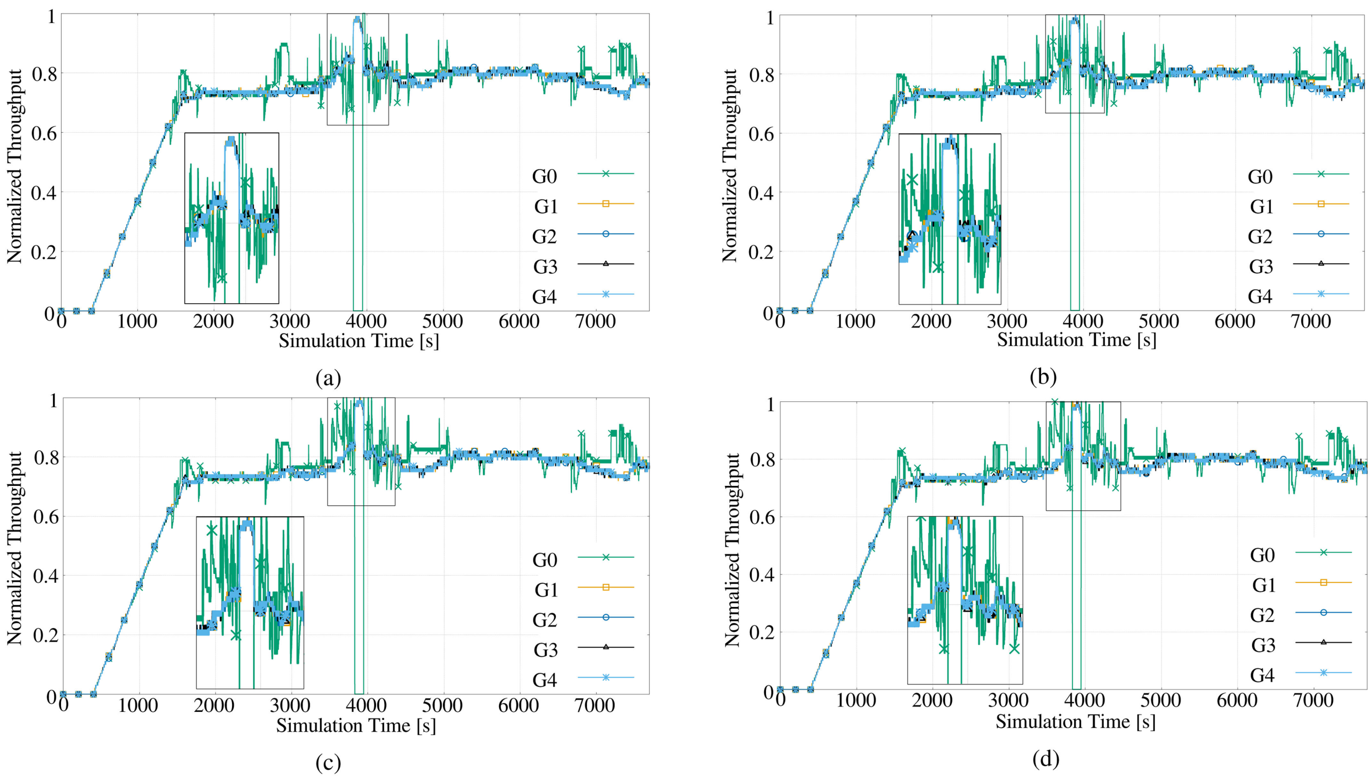


Fig. 7. GW uplink channels' NT: $T = 10\ s$ (a), $T = 20\ s$ (b), $T = 30\ s$ (c), and $T = 40\ s$ (d), OUS.

TABLE III
ADDITIONAL INFORMATION ABOUT OBTAINED PERFORMANCE AND OVERHEAD WITH $T = [10, 20, 30, 40]\ s$, OUS

***T* [*s*]**	10		20		30		40	
Mean network NT	0.672		0.672		0.672		0.672	
Mean *G0* NT	0.668		0.670		0.674		0.675	
Number of re-allocated traffic flows (predicted congestion)	364	0.004%	250	0.003%	184	0.002%	187	0.002%
Number of re-allocated traffic flows (predicted outage)	192	0.002%	210	0.002%	222	0.002%	215	0.002%
Number of lost packets (buffer overflow)	0	0%	0	0%	0	0%	0	0%
Number of re-routed packets (outage)	0	0%	0	0%	0	0%	448	$4 \cdot 10^{-6}\%$

TABLE IV
ADDITIONAL INFORMATION ABOUT OBTAINED PERFORMANCE AND OVERHEAD WITH $T = [100, 200, 300, 400]\ s$, OUS

***T* [*s*]**	100		200		300		400	
Mean network NT	0.672		0.672		0.672		0.672	
Mean *G0* NT	0.676		0.678		0.678		0.676	
Number of re-allocated traffic flows (predicted congestion)	121	0.001%	91	0.001%	83	0.001%	86	0.001%
Number of re-allocated traffic flows (predicted outage)	255	0.003%	227	0.003%	212	0.002%	204	0.002%
Number of lost packets (buffer overflow)	0	0%	0	0%	0	0%	0	0%
Number of re-routed packets (outage)	3192	$3 \cdot 10^{-5}\%$	2774	$3 \cdot 10^{-5}\%$	2759	$3 \cdot 10^{-5}\%$	13125	$10^{-4}\%$

react in time, with a consequent need of packet re-routing to empty the G_0 buffer after the link outage starts.

Another key parameter influencing the overall performance is the number of prediction samples P (and the prediction of t_d), for which a dedicated sensitivity analysis has been carried out through additional simulation campaigns, whose results in terms of NT are shown in Fig. 8 for t_d set to $10,\ 20,\ 30,\ 40\ s$ and $T = 10\ s$.

A performance decrease is evident by analysing the behaviour of NT recorded at G_0. Additional information about the obtained performance is also provided in Table V to shed more light on the observed performance degradation.

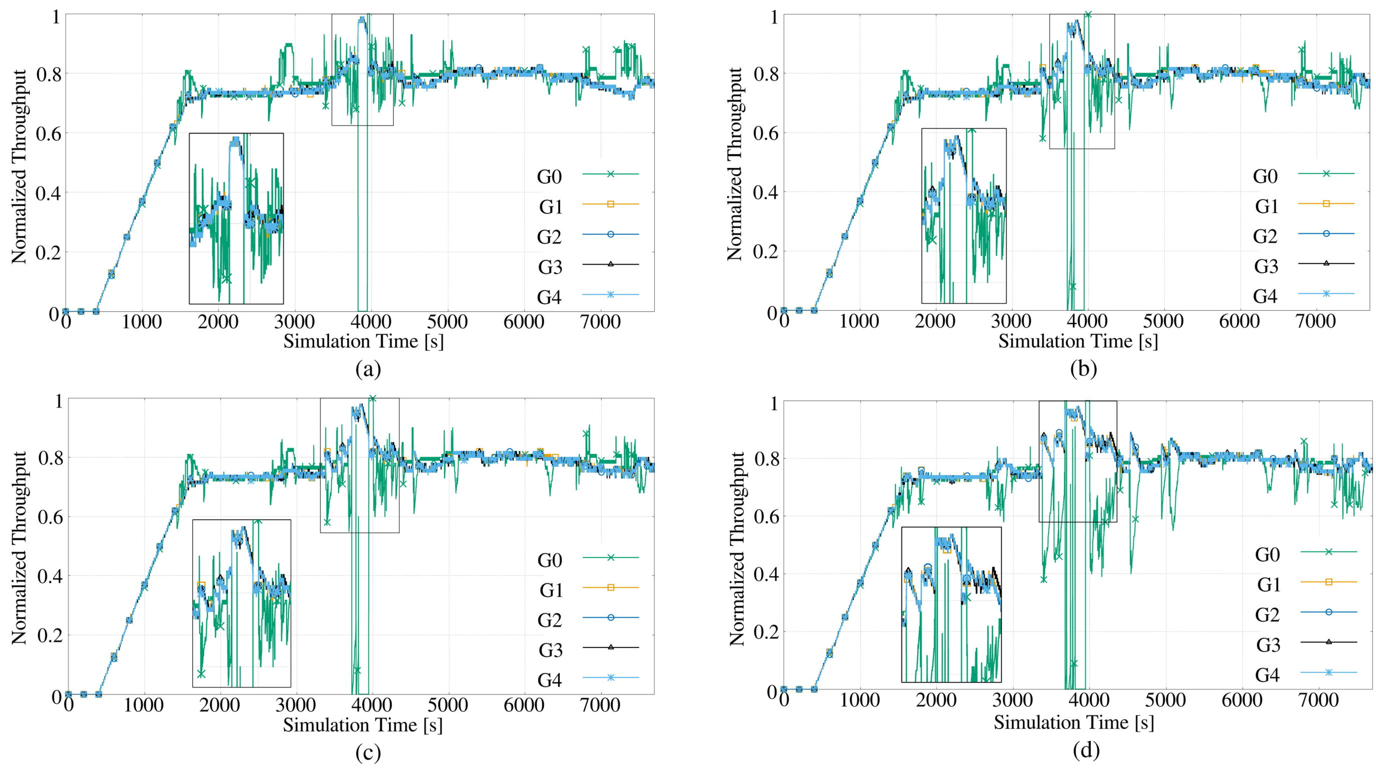


Fig. 8. GW uplink channels' NT: $t_d = 10\ s$ (a), $t_d = 20\ s$ (b), $t_d = 30\ s$ (c), and $t_d = 40\ s$ (d), OUS.

TABLE V
ADDITIONAL INFORMATION ABOUT OBTAINED PERFORMANCE AND OVERHEAD WITH $t_d = [10, 20, 30, 40]\ s$, OUS

t_d **[s]**	10		20		30		40	
Mean network NT	0.672		0.672		0.672		0.672	
Mean *G0* NT	0.668		0.648		0.631		0.612	
Number of re-allocated traffic flows (predicted congestion)	364	0.004%	742	0.009%	1273	0.015%	1747	0.020%
Number of re-allocated traffic flows (predicted outage)	192	0.002%	229	0.003%	296	0.003%	364	0.003%
Number of lost packets (buffer overflow)	0	0%	27456	$3 \cdot 10^{-4}$%	27463	$3 \cdot 10^{-4}$%	27679	$3 \cdot 10^{-4}$%
Number of re-routed packets (outage)	0	0%	0	0%	0	0%	0	0%

Particular attention has to be paid to the case where the t_d value is set higher than $T = 10\ s$, since the link outage prediction algorithm performs the SINR estimation on the basis of a reduced set of past samples, implying two main side effects:

- *False positive predictions:* the outage event is incorrectly predicted (either several seconds before or after the actual outage start), as shown by the upside-down spike of NT at G_0, leading to unnecessary traffic flow re-allocation ($4th$ and $5th$ rows of the table).
- *False negative predictions:* during the outage event, the outage termination is predicted to take place too early, leading to a premature traffic flow re-allocation to G_0 and consequent packet losses due to buffer overflow ($6th$ row of the table).

As already mentioned in Section IV-B, more the prediction algorithm tries to estimate in a future far from the present, lower its accuracy is. Even if we can properly set T and t_d values to estimate the future in a fixed time window dependent on the amount of available knowledge of the past, this precision decrease will always affect the estimation. For this reason, we investigated the behaviour of OUR by assigning a decreasing set of weights to the estimated achievable output rate values over time, i.e. setting the α_p values in Eq. (3) in order to give higher weights to the estimated values closer to the current time (lower p values). The employed scheme, called *decreased weight association scheme (DWAS)*, can be described as:

$$\alpha_p = \frac{P - p}{\sum_{k=1}^{P} k} = \frac{(P - p) \cdot 2}{P \cdot (P + 1)} \tag{5}$$

The NT performance of the GW uplink channels do not show any evident change compared to the one obtained by using the *uniform association scheme* applied in OUS and are not reported here. However, a small performance improvement can be noticed if additional performance statistics are analysed, as reported in Table VI.

TABLE VI
ADDITIONAL INFORMATION ABOUT OBTAINED PERFORMANCE AND OVERHEAD WITH $T = [10, 20, 30, 40]$ s, DWA α_p ASSOCIATION SCHEME, OUS

***T* [*s*]**	10		20		30		40	
Mean network NT	0.672		0.672		0.672		0.672	
Mean *G*0 NT	0.670		0.672		0.673		0.674	
Number of re-allocated traffic flows (predicted congestion)	264	0.003%	210	0.002%	170	0.002%	165	0.002%
Number of re-allocated traffic flows (predicted outage)	199	0.002%	213	0.002%	224	0.003%	216	0.003%
Number of lost packets (buffer overflow)	0	0%	0	0%	0	0%	0	0%
Number of re-routed packets (outage)	0	0%	0	0%	0	0%	0	0%

Comparing the data shown in Tables III and VI, a lower number of re-allocated traffic flows has been obtained by using the DWA scheme, i.e. relying more on the near future estimated values for both the initial traffic flow association and the traffic flow re-allocation in the case of congestion prediction.

VI. CONCLUSION

The paper has focused on the design of future VHTS satellite systems, by paying particular attention to the ground segment optimisation for providing end users with very high throughput and at the same time guarantee content distribution continuity in spite of possible satellite feeder link outage. To this end, a novel optimisation framework combining advanced prediction algorithms as well as congestion-aware routing concepts has been proposed in order to overcome some of the limits of the currently available network satellite solutions typically addressing feeder link outage and resource allocation problems separately. On the contrary, the proposed approach helped shed some light on the advantage of developing a joint optimisation framework able to reliably predict link outage and accordingly identify the most convenient alternate gateways to use for forwarding the corresponding traffic flows. Moreover, the developed proposal shows also important features in what concerns the mitigation of congestion problems when excessive traffic load insists on the same satellite link as a result of handover procedures.

Given the large number of degrees of freedom introduced by the investigated problem as well as the rich set of possible key performance indicators to assess the effectiveness of the proposed solutions, additional studies in the direction of resource allocation and network orchestration based on MANO-based architectures are needed. In particular, the availability of flexible satellite payloads and the use of SDN concepts in space seems particularly appealing for the satellite systems to adapt to the diverse operating conditions and the various traffic flavours injected by different classes of users. These aspects along with the corresponding network architecture implications though very relevant could not be addressed in this paper because of lack of space and will be subject of future investigations.

REFERENCES

[1] G. Fettweis and S. Alamouti, "5G: Personal mobile internet beyond what cellular did to telephony," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 140–145, Feb. 2014.

[2] "Cisco annual internet report (2018–2023) white paper," Tech. Rep., 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[3] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tut.*, vol. 20, no. 4, pp. 2714–2741, 2018.

[4] A. Guidotti *et al.*, "Architectures and key technical challenges for 5G systems incorporating satellites," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2624–2639, 2019.

[5] Z. Katona, F. Clazzer, K. Shortt, S. Watts, H. P. Lexow, and R. Winduratna, "Performance, cost analysis, and ground segment design of ultra high throughput multi-spot beam satellite networks applying different capacity enhancing techniques," *Int. J. Satell. Commun. and Netw.*, vol. 34, no. 4, pp. 547–573, 2016.

[6] N. Jeannin, L. Castanet, J. Radzik, M. Bousquet, B. Evans, and P. Thompson, "Smart gateways for terabit/s satellite," *Int. J. Satell. Commun. Netw.*, vol. 32, no. 2, pp. 93–106, 2014.

[7] T. Rossi, M. De Sanctis, F. Maggio, M. Ruggieri, C. Hibberd, and C. Togni, "Smart gateway diversity optimization for EHF satellite networks," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. 56, no. 1, pp. 130–141, Feb. 2020.

[8] Y. Shi, J. Liu, Z. M. Fadlullah, and N. Kato, "Cross-layer data delivery in satellite-aerial-terrestrial communication," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 138–143, Jun. 2018.

[9] P. Angeletti, R. De Gaudenzi, E. Re, N. Jeannin, and A. Vernucci, "Multi-beam satellite communication system and method, and satellite payload for carrying out such a method," 2016, U.S. Patent 9 356 685.

[10] Y. Shi, Y. Cao, J. Liu, and N. Kato, "A cross-domain sdn architecture for multi-layered space-terrestrial integrated networks," *IEEE Netw.*, vol. 33, no. 1, pp. 29–35, 2019.

[11] T. Ahmed, E. Dubois, J.-B. Dup, R. Ferrs, P. Glard, and N. Kuhn, "Software-defined satellite cloud ran," *Int. J. Satell. Commun. Netw.*, vol. 36, no. 1, pp. 108–133, 2018.

[12] L. Bertaux *et al.*, "Software defined networking and virtualization for broadband satellite networks," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 54–60, 2015.

[13] F. Patrone, M. Marchese, A. Moheddine, T. de Cola, and M. Mongelli, "QoS-aware handover strategies for Q/V feeder links in VHTS systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–7.

[14] R. De Gaudenzi, P. Angeletti, D. Petrolati, and E. Re, "Future technologies for very high throughput satellite systems," *Int. J. Satell. Commun. Netw.*, vol. 38, no. 2, pp. 141–161, 2020.

[15] A. Kyrgiazos, B. Evans, P. Thompson, P. Mathiopoulos, and S. Papaharalabos, "A terabit/second satellite system for european broadband access: A feasibility study," *Int. J. Satell. Commun. Netw.*, vol. 32, no. 2, pp. 63–92, 2014.

[16] A. Gharanjik, B. Shankar M. R., P. Arapoglou, and B. Ottersten, "Multiple Gateway Transmit Diversity in Q/V Band Feeder Links," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 916–926, Mar. 2015.

[17] A. J. Roumeliotis, C. I. Kourogiorgas, and A. D. Panagopoulos, "Dynamic capacity allocation in smart gateway high throughput satellite systems using matching theory," *IEEE Syst. J.*, vol. 13, no. 2, pp. 2001–2009, Jun. 2019.

[18] Y. Vasavada, R. Gopal, C. Ravishankar, G. Zakaria, and N. BenAmmar, "Architectures for next generation high throughput satellite systems," *Int. J. Satell. Commun. Netw.*, vol. 34, no. 4, pp. 523–546, 2016.

[19] T. Ahmed, A. Alleg, R. Ferrus, and R. Riggio, "On-demand network slicing using sdn/nfv-enabled satellite ground segment systems," in *Proc. 4th IEEE Conf. Netw. Softwarization Workshops (NetSoft)*, 2018, pp. 242–246.

[20] M. Dong, H. Li, K. Ota, and J. Xiao, "Rule caching in sdn-enabled mobile access networks," *IEEE Netw.*, vol. 29, no. 4, pp. 40–45, Aug. 2015.

[21] Y. Cao, Y. Shi, J. Liu, and N. Kato, "Optimal satellite gateway placement in space-ground integrated network for latency minimization with reliability guarantee," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 174–177, Apr. 2018.

[22] J. Liu, Y. Shi, L. Zhao, Y. Cao, W. Sun, and N. Kato, "Joint placement of controllers and gateways in sdn-enabled 5G-satellite integrated network," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 221–232, Feb. 2018.

[23] N. Jeannin, L. Castanet, I. Dahman, V. Pourret, and B. Pouponneau, "Smart gateways switching control algorithms based on tropospheric propagation forecasts," *Int. J. Satell. Commun. Netw.*, vol. 37, no. 1, pp. 43–55, 2019.

[24] R. Nebuloni *et al.*, "Optimization of Q/V-band Smart Gateway Switching in the framework of Q/V-Lift Project," in *Proc. IEEE Aerosp. Conf.*, 2019, pp. 1–8.

[25] M. Muhammad, G. Giambene, and T. de Cola, "QoS support in SGD-based high throughput satellite networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8477–8491, 2016.

[26] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, 2015.

[27] H. Li, K. Ota, and M. Dong, "Virtual network recognition and optimization in sdn-enabled cloud environment," *IEEE Trans. Cloud Comput.*, pp. 1–1, 2018.

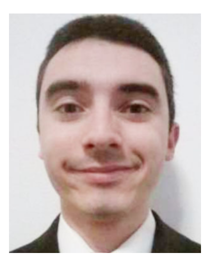
[28] N. McKeown *et al.*, "Openflow: Enabling innovation in campus networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, 2008.

[29] M. Mongelli, T. De Cola, M. Cello, M. Marchese, and F. Davoli, "Feeder-link outage prediction algorithms for sdn-based high-throughput satellite systems," in *Proc. IEEE Int. Conf. Commun.*, 2016, pp. 1–6.

Tomaso de Cola (Member, IEEE) was born in Manosque, France, on April 28, 1977. He received the *Laurea* degree (with honors) in telecommunication engineering, in 2001, the qualification degree as Professional Engineer, in 2002 and the Ph.D. degree in electronic and computer engineering, robotics and telecommunications from the University of Genoa, Italy, in 2010. From 2002 until 2007, he has worked with the Italian Consortium of Telecommunications (CNIT), University of Genoa Research Unit, as Scientist Researcher. Since 2008, he has been with the German Aerospace Center (DLR), where he is involved in different European Projects focusing on different aspects of DVB standards, CCSDS protocols and testbed design. He has been serving since 2015 as Deputy Area Director of the Space Internetworking Services (SIS) area within Consultative Committee for Space Data Systems (CCSDS) and taking part to the CCSDS Engineering Steering Group (CESG). He is member of the IEEE Communications Society, where he was Chair of the Satellite and Space Communications (SSC) technical committee from 2017 to 2020. He has been Editor for many IEEE journals such as IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, SYSTEMS JOURNAL, COMMUNICATION LETTERS, WIRELESS COMMUNICATION LETTERS. He was also Guest Editor for IEEE JSAC, *IEEE Wireless Communications Magazine*, and IEEE NETWORK. He is author of more than 100 scientific papers.

Mario Marchese (Senior Member, IEEE) was born in Genoa, Italy, in 1967. He received the *Laurea* degree cum laude at the University of Genoa, Italy in 1992, and the Ph.D. in T*elecommunications* from the University of Genoa, in 1997. From 1999 to January 2005, he worked with the Italian Consortium of Telecommunications (CNIT), by the University of Genoa Research Unit, where he was Head of Research. From February 2005 to January 2016 he was Associate Professor at the University of Genoa. Since February 2016 he has been Full Professor at the University of Genoa. He was the Chair of the IEEE Satellite and Space Communications Technical Committee from 2006 to 2008. He is Winner of the IEEE ComSoc Award "2008 Satellite Communications Distinguished Service Award" in "recognition of significant professional standing and contributions in the field of satellite communications technology." He is the author of the book *Quality of Service over Heterogeneous Networks* (John Wiley & Sons, Chichester, 2007), and author/co-author of more than 300 scientific works, including international magazines, international conferences and book chapters. His main research activity concerns: networking, Quality of Service over heterogeneous networks, software defined networking, satellite DTN and nano-satellite networks, networking security.

Maurizio Mongelli (Member, IEEE) received the Ph.D. degree in electronics and computer engineering from the University of Genoa (UNIGE), in 2004. The doctorate was funded by Selex Communications S.p.A. (Selex). He worked for both Selex and the Italian Telecommunications Consortium (CNIT) from 2001 until 2010. During his doctorate and in the following years, he worked on the quality of service for military networks with Selex. From 2007 to 2008, he coordinated a joint laboratory between UniGe and Selex, dedicated to the study and prototype implementation of Ethernet resilience mechanisms. He was the CNIT Technical Coordinator of a research project concerning satellite emulation systems, funded by the European Space Agency; spent three months working on the project at the German Aerospace Center in Munich. Since 2012 he is a Researcher at the Institute of Electronics, Computer and Telecommunication Engineering (IEIIT) of the National Research Council (CNR), where he deals with machine learning, bioinformatics and cyber security, having the responsibility and coordination, for the CNR part, of several funded projects in those sectors. He is co-author of over 100 international scientific papers and 2 patents.



Fabio Patrone (Member, IEEE) was born in Genoa, Italy in 1988. He received the bachelor's degree in telecommunication engineering, in 2010 and the master's degree in telecommunication engineering, in 2013 from the University of Genoa. He got his Ph.D. at the Satellite Communications and Networking Laboratory (SCNL) with a thesis on routing and scheduling algorithms in Satellite Delay and Disruption Tolerant Networks (DTNs). He is currently a Post-Doc research fellow at the SCNL. His main research activity involves satellite, vehicular, and sensor networks. In particular, it concerns the design of Routing, Scheduling, and Congestion Control algorithms and the employment of networking technologies, such as Network Function Virtualization (NFV) and Software-Defined Networking (SDN), for the integration of these networks with the terrestrial infrastructure within the 5G ecosystem.