Multi-Objective Optimization for Integrated Ground-Air-Space Networks

Jingjing Cui, *Member, IEEE,* Soon Xin Ng, *Senior Member, IEEE,* Dong Liu, *Member, IEEE,* Jiankang Zhang, *Senior Member, IEEE,* Arumugam Nallanathan *Fellow, IEEE* and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—With space and aerial platforms deployed at different altitudes, integrated ground-air-space (IGAS) networks will have multiple vertical layers, hence forming a three-dimensional (3D) structure. These 3D IGAS networks integrating both aerial and space platforms into terrestrial communications constitute a promising architecture for building fully connected global next generation networks (NGNs). This article presents a systematic treatment of 3D networks from the perspective of multi-objective optimization. Given the inherent features of these 3D links, the resultant 3D networks are more complex than conventional terrestrial networks. To design 3D networks accommodating the diverse performance requirements of NGNs, this article provides a multi-objective optimization framework for 3D networks in terms of their diverse performance metrics. We conclude by identifying a range of future research challenges in designing 3D networks and by highlighting a suite of potential solutions.**

I. Introduction

Next generation networks (NGNs) are expected to strike an improved trade-off amongst the key quality of service (QoS) metrics such as the data rate, latency, energy efficiency and integrity [1], [2]. As reported in [3], commercial air traffic has grown by a factor of 2.3 since 2000 and it is expected to further double in the next 20 years. The seamless integration of space and aeronautical applications into terrestrial networks has attracted substantial interests both in industry and in academia in the context of future networks.

In contrast to terrestrial networks, integrated ground-air-space (IGAS) networks exhibit a multi-layer architecture in the vertical domain, forming a three-dimensional (3D) network such as the five-layer architecture described in [1], [4]. Moreover, in [5], the characteristics of heterogeneous air-ground network integrating both low- and high-altitude unmanned aerial vehicles (UAVs) were discussed. In terms of modelling the space to ground channel, the path-loss model of the satellite to ground channel was presented in urban environments in [6]. In addition to supporting terrestrial communications, 3D IGAS networks, abbreviated 3D/IGAS networks, rely on 'heterogeneous' architectures and are expected to support flawless global on-demand connectivity. With the proliferation of Internet services and applications, the end devices require high-speed broadband access on board of ships, airliners, trains, etc even in remote rural areas. In this context, the IGAS concept becomes a promising solution for supporting 3D connectivity relying on emerging technologies.

There are several compelling advantages in integrating the ground, air and space networks. Firstly, the available frequency bands may be integrated with the aid of wide-ranging spectrum sharing; Secondly, the flexible deployment of aerial platforms may be combined with those of satellites for providing ubiquitous high-quality connectivity for NGNs. Recently, the SaT5G Project has announced the demonstration of 5G signal transmissions over satellites [7]. Moreover, supporting connectivity both by space and aerial systems has also been considered in 3GPP standardization [2]. Against this background, the goal of this article is to provide an overview of how interconnected 3D networks may deliver ubiquitous, high-quality connectivity by invoking multi-objective optimization for striking a compelling trade-off amongst the diverse criteria to be jointly satisfied at the same time.

In Section II, we summarize the distinguishing features of interconnected 3D networks, namely their potentially high bandwidth, global coverage, heterogeneity and flexibility. In Section III,, we introduce both the architecture and the requirements of 3D networks relying on multi-objective optimizations. In Section IV, we present a multi-objective throughput and delay optimization problem. Then, in Section V we introduce the properties of Pareto-optimal solutions and illustrate the impact of Pareto-optimal solutions on the system's performance. Finally, in Section VI, we discuss the associated research challenges and some promising future directions followed by our conclusions in Section VII.

II. Typical Features of 3D Networks

3D IGAS networks seamlessly integrating terrestrial and non-terrestrial layers can be characterized by the vertical multi-layer architecture of Fig. 1[1]. The terrestrial networks on the ground include both the networks evolving from 2G to 5G as part of the operational heterogeneous networks (HetNets) and the various types of entities operating on the ground such as macro-BSs, smart-phones and vehicles [2]. Low-altitude platforms (LAPs) such as various aircraft, airships

This work would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/Noo4558/1, EP/PO34284/1, COALESCE, of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

J. Cui, S. Ng, D. Liu and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, Southampton (UK) (e-mail: jingj.cui,d.liu,jz1n08@soton.ac.uk and sxn,lh@ecs.soton.ac.uk)).

J. Zhang is with Department of Computing & Informatics, Bournemouth University, BH12 5BB, U.K. (e-mail: jzhang3@bournemouth.ac.uk).

Arumugam Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K

[1]Note that intra-layer connections are not shown by this projection, since the main challenge of 3D networks is the efficient cooperation among the entities of different layers.

Table I: Comparison of different communication properties: typical characteristics [8].

Places	Entities	Altitude (h)	Examples	Propagation delay	Transmission speed	Motion
Space	GEO (L_6)	≥ 35786km	Satellites	≥ 119.286ms	1-140Gbps	Circular orbits; Earth fixed position
	MEO (L_5)	> 2000km		> 6.666ms	0.5-5.6Gbps	Circular orbits or elliptical orbits around the earth
	LEO (L_4)	> 160km		> 0.533ms	0.01-2Gbps	
Air	HAP (L_3)	$17 - 30$km	Airships, balloons, airplanes, UAVs	56-100us	Up to 1.25Gbps	Typically to be programmed for specific missions
	LAP ($L2$)	$5 - 10$km		< 34us	High	
Ground	Mobile networks ($L1$)	–	BS and end devices	Ultra-low	High	Slow or application-driven motions

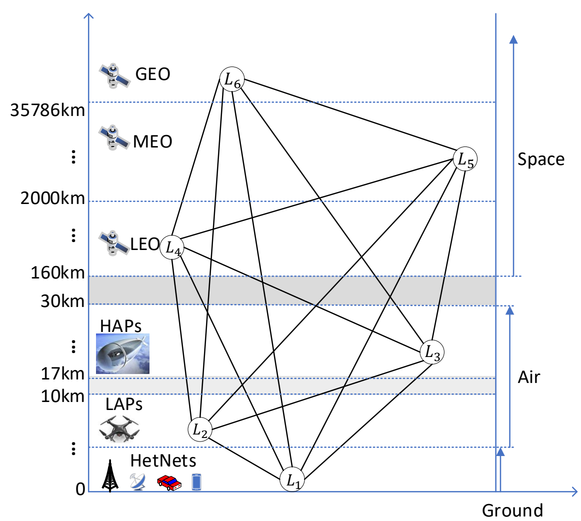


Figure 1: Illustrations of a topological framework of the interconnected 3D IGAS network projected onto the vertical plane, which comprises six layers of different altitudes of the infrastructure entities.

and small unmanned aerial vehicles (UAVs) fly at various altitudes in the troposphere between a few dozen meters to a few thousand meters. High-altitude platforms (HAPs) [9] operate in the stratosphere at altitudes up to 30km, including balloons and unmanned solar-powered planes. Both HAPs and LAPs form part of the airborne segment. Space segments encompass satellites operating in three different orbits: low earth orbit (LEO), medium earth orbit (MEO) and geostationary or geosynchronous orbit (GEO). Table I summarises the salient features of ground, air and space communications and contrasts them. Note that each layer may consist of different elements having distinct characteristics such as size, payload capabilities, communication capacity, endurance, etc. For instance, there are many types of UAVs designed for diverse applications, which can be classified by size, range and endurance. Furthermore, the UAVs can also be classified by the specific type of the aerial platform used, such as single-rotor helicopter, multi-rotor UAVs, fixed-wing UAVs and fixed-wing hybrid UAVs etc. As discussed in [5], the different types of UAVs usually have different performance including payload capabilities, communication capacity, etc. However, these considerations are beyond the scope of this work. Note that regarding to the communications links in the IGAS network, there are a number of solutions for providing the link options in the IGAS network, such as radio frequency (RF), free space optical (FSO), etc, which should be selected carefully according to the specific system requirements. For instance, due to the low absorption and scattering loss in space, FSO links can be used for connecting satellites in space. However, the optical signal in the atmosphere is gravely attenuated by absorption and scattering owing to rain, snow, fog and and clouds as studied in [4]. RF links become a beneficial option for interconnecting aerial and ground platforms. Moreover, with the development of the mm-wave technologies, mm-Wave links based on advanced beamforming techniques can potentially support the air-to-air and air-to-ground connections as in [4], [5]. By exploiting the diverse potential of different segments, the IGAS system becomes capable of efficiently amalgamating ground, air and space networks by evolving them to the qualitatively new IGAS network concept, which posses the characteristics of potentially high bandwidth, global coverage, heterogeneity and flexibility.

Wide bands at high-frequency carrier: Spectrum-sharing may be exploited in 3D networks relying on various terrestrial frequency bands as well as air-to-air (A2A), air-to-ground (A2G) and a variety of satellite frequency bands. The higher carrier frequency naturally facilitates access to wider bandwidths and allows the use of more compact antennas than their low-frequency counterparts, which dramatically reduces the space and weight requirements, whilst facilitating sophisticated interference and resource management.

Wide-area service coverage: The variety of satellites covers a wide area of the earth and the aerial platforms including HAPs as well as LAPs having a relatively low cost are capable of providing a high-quality coverage. Hence there is potentially increased capacity for supporting hitherto unserved areas, such as isolated/remote rural areas and on board aircraft through the seamless integration of space and aerial platforms into ground networks. Additionally, communications in poorly covered areas such as remote rural areas having a low tele-traffic density are also expected to rely on these integrated 3D networks for upgrading them in a cost-efficient manner. Therefore, 3D networks have the potential of creating a fully-connected world.

Super-heterogeneity of multi-layer networks: The super-heterogeneity of 3D networks includes the heterogeneity of architectures, service requirements, base stations (BSs) and handsets. It will also impose additional hardware and software constraints on the infrastructure, users, vehicles and many other paraphernalia of 3D networks. These constraints will have a substantial impact on the network's organization and resource allocation.

Mobility and flexibility in 3D networks: The convenience of HAPs and LAPs provides substantial grade of flexibility for communications, given their flexible and prompt deployments. However, a particular challenge imposed by the rather different velocity of different platforms is the handoff issue. For instance, when an on-board GEO satellite user moves from one beam to another or between different non-GEO constellations, handovers would potentially impose unacceptable disruptions.

III. Multi-objective Optimization in 3D Networks

In this section, the potential architectures and requirements of 3D NGNs are presented. Furthermore, inspired by the demanding features of emerging applications in terms of their conflicting objectives, the concept of multi-objective optimization is proposed for 3D networks in the face of their design constraints.

A. *Architectures and Requirements for 3D IGAS Networks*

The terminals of the near future are likely to have a blend of connectivity options, seamlessly transitioning between different types of access points when they move from place to place. Hence, the space and aerial platforms will have to achieve new levels of flexibility and collaboration capabilities with ground networks. 3GPP distinguishes a pair of satellite and aerial access networks based on their operating frequency bands [2]: 1) Broadband access networks operating above 6GHz, which serve terminals having very small aperture that can either be fixed or mounted on a moving platform such as a bus, train, vessel and aircraft etc; 2) Narrow- or wide-band access networks operating below/above 6GHz, which serve terminals equipped with omni- or semi-directional antennas, such as handheld terminals and IoT devices.

As it can be observed from Fig. 1, based on the specific types of space and aerial platforms, a large variety of 3D architectures can be attained by investigating the relationship between the edges and vertices in the graph. More specifically, the degree of a node denotes the number of all possible links with the other layers and a circle in the graph corresponds to a relay. Given the inherent features of the entities in each

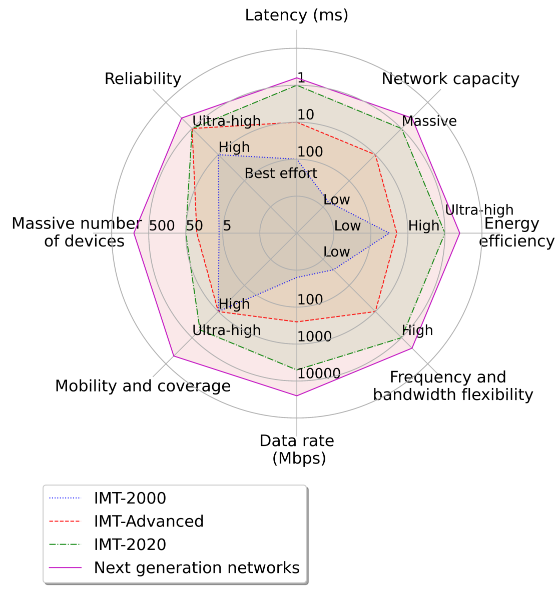


Figure 2: Requirements of KPIs [10] for IMT-2000, IMT-Advance, IMT-2020 and the future networks.

layer such as their channel characteristics or link budgets, the weight of each edge can be described by these characteristics, bearing in mind the specific applications. Moreover, with the evolution of social networking, the ubiquitous wireless connectivity, coupled with the availability of low-cost and powerful computing devices as well as with the increasing demands for improved rate and reliability, they jointly shape the evolution of next generation networks. Specifically, the ever-increasing demands on wireless networks can be distinguished from the evolution of the requirements identified by International Mobile Telecommunications (IMT) documents. In Fig. 2, we provide an example of the conflicting key performance requirements [10], also termed as key performance indicators (KPIs), as they have evolved in two decades from IMT-2000 to IMT-2020 and to next generation networks in future.

As it can be observed from Fig. 2, the key capabilities of future networks are expected to further enhance the performance metrics. For example, the number of devices and autonomous vehicles deployed in mission-critical and application-centric settings is expected to grow substantially, along with the velocity of users and coverage areas. In these scenarios, challenging network designs that meet ultra-high reliability ($\geq$ 99.99%) and low end-to-end latency ($\leq$ 1 ms) are expected, even in some high-mobility scenarios on board of both planes and trains. Hence, several factors of the KPIs shown in Fig. 2 have to be jointly considered in the face of diverse application requirements. In this case, defining a multi-component OF in terms of different KPIs that forms a multi-objective optimization and determining the Pareto front of all optimal solutions would allow us to activate that particular optimal mode of operation, which best fits the specific application to be supported.

B. Necessity of Multi-objective Optimization

To elaborate a little further, given our limited wireless resources such as energy and frequency bands, the objectives of different applications may become conflicting, as exemplified by minimizing the latency, maximizing the reliability, and minimizing the energy, etc. At the time of writing most of the existing investigations tend to concentrate on transforming multi-objective optimization problems (MOOPs) to single-objective ones, even if ideally we would like to jointly maximize the weighted sum rate and to minimize the total energy consumption [4], [11]. This is because it is quite challenging to strike a trade-off by incorporating multiple conflicting metrics into a single OF, especially when the resultant solution space is non-convex. Moreover, 3D networks, having an ultra-large coverage associated with a complex architecture result in decentralized and distributed optimization scenarios. Implementing their centralized control becomes even more challenging when increasing the size of the network. In this context, multi-objective optimization becomes a promising formulation for simultaneously satisfying multiple key performance requirements with respect to different optimization goals.

Having said that, when multiple metrics are incorporated into a MOOP, the search space is expanded compared to directly treat them as constraints, hence potentially resulting in complex and particularly challenging optimization problems, especially when the number of objectives is high. Furthermore, in contrast to single-objective optimization problems (SOOPs) having a well-defined search space, the search space may become less well ordered, when we try to optimize several objectives at the same time [12]. The solution methods of MOOPs can be broadly categorized into two types: one of them combines the individual OFs into a single composite function or relegates all but one objective to the constraint set; the second type of approaches determines the entire Pareto-optimal solution set or a representative subset of them by evolutionary algorithms.

C. Design Constraints of 3D networks

Based on the typical features of the 3D networks discussed in Section II and on the challenging requirements of flawless telepresence services, both the constraint formulations of these features and the OF formulations play a vitally important role in deriving Pareto optimal solutions for MOOPs. To this end, we introduce some specific design constraints of 3D networks that are different from those of conventional terrestrial communications [2]:

- The propagation channel models between space and aerial platforms and ground-terminals have different multi-path and Doppler-spectrum models from those of the terrestrial networks. Furthermore, given the mobility of the infrastructure's transmission equipment and terminals, the Doppler effect will continuously modify the carrier frequency and phase, resulting in destructive inter-carrier-interference (ICI), which depends both on the relative satellite/HAP velocity with respect to the terminal, and on the specific carrier frequency. To compensate the Doppler shift and Doppler fluctuation rate, the satellite/HAP motion trajectories and the ground terminal location are expected to be known.
- Satellite systems exhibit much higher propagation delay than ground systems, while the one way delay of the aerial platforms is comparable to that of cellular networks, as illustrated in Table I. Note that the high satellite-delay will affect not only the data links but also all the signalling loops both during access and data transport.
- Space and air systems tend to support much larger coverage areas (cells) than the BSs of ground networks. This has its pros and cons, because the hand-over rate is beneficially reduced, but given their large 'footprint' on the ground, they can only accommodate a limited number of ground users across a large area in sparsely populated regions. Hence their area spectral efficiency is limited. As a further challenge, their coverage footprint may be moving (without a fixed earth reference point) in case of non-geostationary satellite orbits (NGSOs). On the other hand, the ratio between the propagation delay at the cell-centre and cell-edge is likely to be higher in the context of aerial systems having lower operational elevation angles than the geostationary satellite systems.
- Agile radio resource management accurately adapted to the challenging tele-traffic demands has to be incorporated into our 3D network design. These constraints encompass the transmit power, carrier frequency and channel bandwidth, the service-continuity during handover procedures as well as mobility control etc. In this challenging context, efficient 3D resource management techniques are required for reducing the network cost and mitigating the inter/intra-layer interference, whilst improving the coverage.

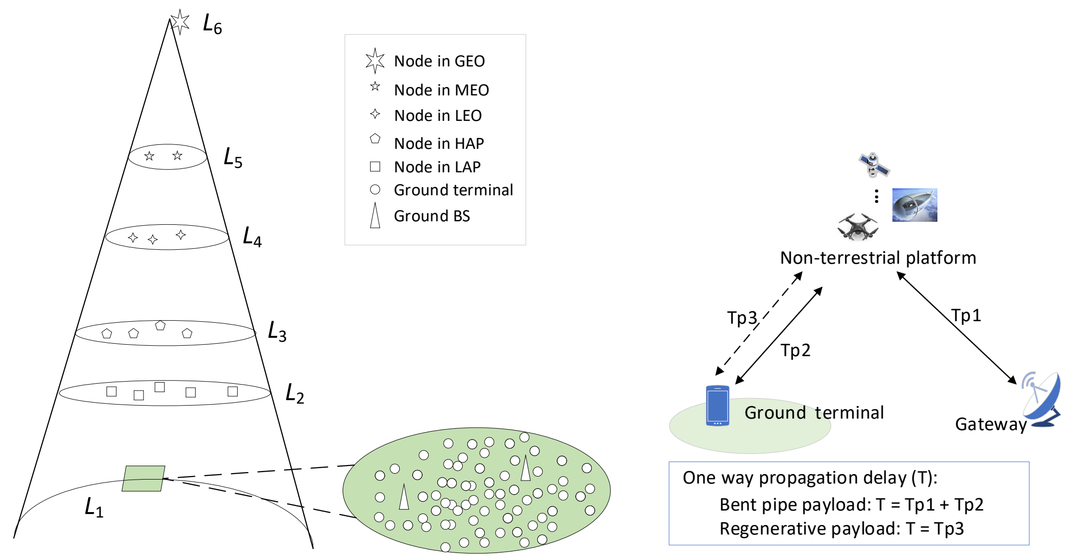
As a tangible design example, a general multi-objective optimization model of maximizing the throughput and minimizing the delay of the system can be formulated as follows:

$$\begin{aligned} \max \quad & f_1: \text{Total throughput} \\ \min \quad & f_2: \text{Total delay} \\ \text{s.t} \quad & g_1: \text{Guarantee the ground devices' service constraints,} \\ & g_2: \text{Guarantee each platform's constraints,} \end{aligned}$$

where the variables can be any arbitrary design parameters related to our 3D network, such as task scheduling or resource allocation. Note that this expression can be directly extended to a general form for MOOPs having more than two objectives and constraints by attaching additional objectives and constraints. For instance, we can also consider the third objective f_3 for minimizing the energy consumption and the third constraint g_3 for satisfying a certain maximum tolerable handover rate constraint.

IV. Multi-objective Scheduling Problems in 3D networks

Scheduling schemes delegate the service requests of the nodes (e.g., handheld or IoT devices) to different access points (space and aerial platforms). Due to the diversity of use cases,



(a) A simple 3D network model with $N_i \in \{1, 4, 4, 3, 2, 1\}$ from the first layer to the sixth layer.

(b) Propagation delay model

Figure 3: Illustrations of 3D network model and propagation delay model. (a) illustrates a hierarchical network consists of six layers spanning from the ground to the space level; (b) illustrates the propagation delay model for non-terrestrial platforms with a bent-pipe and a regenerative payload configuration, respectively.

satellite networks, air networks and ground networks co-exist and collaborate with each other in support of emerging use cases. For ease of clarification, we consider that there are N_i entities at layer L_i, denoted as $\mathcal{L}_i = \{1, 2, \cdots, N_i\}$ with $i \in \mathcal{I} = \{1, 2, \cdots, 6\}$ and M ground users, denoted as $\mathcal{M} = \{1, 2, \cdots, M\}$, which are randomly distributed in the region considered. Explicitly, Fig. 3(a) illustrates an example of the 3D network structure having different number of entities in each layer.

With the heterogeneity of 3D networks, the different types of entities generally behave quite differently, in terms of their spectral efficiency, power efficiency, service delay, the users' connectivity, etc. As a result, there is a need to strike a trade-off between a high throughput and a low delay by beneficially exploiting the specific features of the nodes in the different layers of Fig. 3(a), which results in a scheduling problem subject to a pair of conflicting objectives.

Throughput: The high-throughput GEO as well as new MEO and LEO constellations cooperating with advanced HAPs and LAPs are capable of dramatically expanding the capacity of ground networks. Having said that, optimizing the resource allocation of 3D networks constitutes a significant challenge. The maximum throughput per spot beam of a satellite depends on several parameters, such as the number and construction of antennas, the transmission technology or the available bandwidth, etc. However, the percentage of the maximum throughput that can be reserved by a single node directly depends on the terrestrial user-density within each spot beam. Therefore, the achievable throughput per scheduled node can be modeled by the maximum possible throughput based on the specific choice of platforms accessed. To elaborate further, the throughput of 3D networks also depends on the transmission mode, which can be broadly categorized into three types – unicast, multicast and broadcast [2], [8]. For instance, public safety alerts and automatic upgrades of application software and operating system may be transmitted in unicast/multicast mode; By contrast, social media and entertainment requests (e.g., live broadcasts and TV), automotive and IoT use cases, broadcast or mixed modes are expected to satisfy demanding QoS requirements.

Delay: Space and air communications tend to have rather different propagation parameters, such as path loss, propagation delay, fading properties, etc compared to terrestrial communications. These are critical factors in terms of supporting a user's service level agreement and QoS, especially in delay-sensitive applications of NGNs. Two specific types of payloads may be considered: the bent pipe payloads and the regenerative payloads characterized in [2]. As illustrated in Fig. 3(b), in bent pipe payloads, the one-way propagation delay T is the sum of the feeder link's propagation delay Tp_1 and the user link's propagation delay Tp_2, as exemplified by the total propagation delay T between the gateway and the ground terminal via satellites; By contrast, in regenerative payloads, the one-way propagation delay T is the delay Tp_3 between the satellite and the ground terminal. Therefore, delay-sensitive traffic would have to be scheduled through low-latency links, while delay-tolerant traffic can be scheduled over high-latency satellite links, as it becomes explicit in Fig. 3(a). In 3D networks, the actual propagation delay depends on the space/airborne platform altitude and the relative positions of the gateway and terminals on the ground. Furthermore, the delay of ground devices also depends on the tele-traffic volume as well as on many other factors. However, the propagation delay has a key impact on 3D networks owing to the rather different propagation features of their diverse links.

More particularly, let us consider an example for demonstrating the specific framework of MOOPs in our 3D networks using a pair of OFs, where all the entities in the same layer are assumed to have the same spectral efficiency and propagation delay. Furthermore, all ground devices would access orthogonal resources for avoiding the interference. Let C_i and T_i denote the data rate and the propagation delay of the entity in layer L_i to the ground devices, respectively. Here $x_{i,k}$ is a binary variable to be optimized, where $x_{i,k} = 1$ represents that the ground device k is served by the platform in layer i; otherwise we have $x_{i,k} = 0$. The resultant optimization problem can be formulated as follows:

$$\begin{aligned} \max_{x} \quad & f_1 = \sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{M}} C_i x_{i,k} \\ \min \quad & f_2 = \sum_{i\in\mathcal{I}} T_i x_{i,k} \\ \text{s.t.} \quad & \sum_{i\in\mathcal{I}} x_{i,k} \geq 1, \ \forall k \in \mathcal{M} \\ & x_{i,k} \in \{0,1\}, \ \forall k \in \mathcal{M}, \forall i \in \mathcal{I}, \end{aligned} \tag{1}$$

where f_1 denotes the sum of the spectral efficiency of all layers and f_2 is the total propagation delay suffered by the ground devices, respectively. The first constraint represents that each ground device has to be served by at least one platform. Note that the goal of Problem (1) is to find the best combinations of f_1 and f_2 simultaneously, which results in a set of optimal solutions representing the best trade-offs between the spectral efficiency and the delay.

V. Pareto-optimal Solutions for 3D networks

We consider a scheduling problem having two OFs, which are equally important. In contrast to SOOPs, there is no single globally optimal solution in MOOPs, we rather have a set of optimal operating points. These optimal points jointly form the Pareto front and none of the associated metrics can be improved without degrading at least one of the others. Generating the full set of Pareto-optimal solutions can be computationally expensive and it often becomes infeasible, owing to its excessive complexity. Evolutionary algorithms are well-known techniques that are eminently suitable for solving MOOPs, without visiting the entire solution space, yet finding all optimal solutions of the entire Pareto front with a high probability. Therefore, we may argue that the goal of solving Problem (1) is to find its Pareto-optimal front. In the absence of any further information, none of these Pareto-optimal solutions can be said to be better than the others for a specific application. Hence, ideally we have to find all Pareto-optimal solutions. As discussed in [12], a graphical depiction of a MOOP having two objectives is illustrated in Fig. 4, where a general optimization framework based on evolutionary computation is also illustrated.

Multi-objective evolutionary algorithms (MOEAs) using bio-inspired search paradigms are eminently suitable for solving MOOPs by relying on a population of search agents that collectively approximate the Pareto front. Moreover, evolutionary algorithms are flexible in terms of solving MOOPs, since additional objectives can be easily added, removed or modified [12]. These advantages make MOEAs particularly suitable for optimizing 3D IGAS systems in terms of a variety of objectives. More specifically, NSGA-II, as a typical evolutionary algorithm, first generates offsprings from a set of candidate solutions (parents) of the problem considered using a specific type of crossover as well as mutation operations statiscally proceeding gradually improving the front obtained and then selects the next generation of candidate solutions based on the so-called non-dominated sorting and crowding distance comparisons detailed in [13].

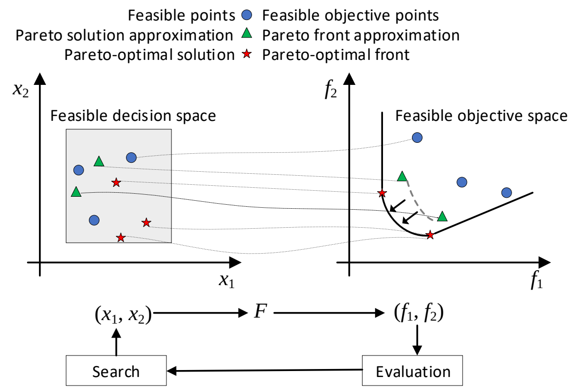


Figure 4: Illustration of solutions for a MOOP having a twin-component OF.

For solving Problem (1), the first OF-component f_1 is equivalently transformed to be the minimization of its negative counterpart $-f_1$, which results in a standard constrained MOOP that can be directly solved by NSGA-II. Let us investigate the Pareto front of Problem (1) in the specific scenarios of Fig. 5, where a GEO satellite in L_6 and a HAP in L_3 collaboratively serve K ground devices[2]. As discussed in Section IV, the GEO station supports a data rate C_6 and has a delay of $T_6 = 120$ms, while the HAP station has a rate of C_3 and a delay of 100μs, respectively. We first consider the scenario of $C_3 = C_6 = 10$ Mbit/s supporting $K = 2$ ground devices. In Fig. 5(a), we plot the Pareto front of $-f_1$ and f_2 as well as all individuals. We can see that there are three Pareto-optimal points on the Pareto front, which indicates the trade-offs between f_1 and f_2. Note that none of the objective values on the Pareto front can be improved without sacrificing the others. Moreover, we can also see from Fig. 5(a) that multiple solutions can map to a single point on the Pareto front. By contrast, in Fig. 5(b), we consider the scenario of having a three-component OF associated with $K = 6$, $C_3 = 10$ Mbit/s and $C_6 = 20$ Mbit/s. Specifically, we introduce the third objective of minimizing the total transmit power, formulated as $f_3 = P_3 \sum_{k=1}^{K} x_{3,k} + P_6 \sum_{k=1}^{K} x_{6,k}$, where $P_3 = 25$ dBm,

[2] More sophisticated system models can be developed for further investigating the impact of system parameters, such as the channel model, on the Pareto front, but given our strict page-limit, this is beyond the scope of this paper.

$P_6 = 40$ dBm. As a result, a three-dimensional Pareto front emerges for the three different OFs in Fig. 5(b), where each point on the Pareto front corresponds to an optimal trade-off between the throughput, the delay and the transmit power of the system. Furthermore, we can see that there are more points on the Pareto front of Fig. 5(b) than in Fig. 5(a), since more ground devices are involved in Fig. 5(b), which provides more flexible alternatives for satisfying the specific QoS preference of the decision makers.

VI. Future Research Challenges

When aiming for ubiquitous connectivity, while satisfying multiple optimization criteria in NGNs, Pareto-optimal optimization provide a useful paradigm for designing efficient and flexible 3D IGAS networks. Based on the above rudimentary investigations of our 3D network in terms of its throughput and delay, our results provide a stepping stone towards their multi-objective optimization. However, there are numerous open research challenges.

A. Resource Management

Naturally, resource management in 3D networks substantially differs from typical cellular networks in terms of the associated altitudes, mobility as well as coverage. Here we have only considered two or three conflicting objectives, namely, the throughput, delay and power. It would be beneficial to determine the Pareto front for even more conflicting performance metrics, as illustrated in Fig. 2. As the objective functions relating to the resource management relies on the wireless resources available, which affects the resultant Pareto front. Therefore, it would be more practical to formulate and solve MOOPs for 3D networks by incorporating realistic resources such as the carrier frequency, bandwidth, power, delay, coverage and Doppler frequency etc, which have to be carefully considered.

B. Network Architecture Designs

Given the recent advances in enabling technologies, protocols and network architectures, 3D IGAS networks in NGNs are heterogeneous and are thus expected to encompass a huge number of diverse near-instantaneously reconfigurable operating entities as well as operating modes in support of flawless Pareto-optimal services at the same time. These entities may have different types of design objectives in specific scenarios. In fact, even a single entity may have multiple design objectives, hence the emerging 3D networks become vastly more complex than their terrestrial counterparts. For instance, satellites are capable of providing global coverage, but they tend to suffer from high propagation delay, while LAPs only impose low latency. Therefore, the Pareto-optimal design of 3D networks satisfying multiple objectives is indeed a promising research direction. Moreover, the extremely agile mobility of planes and LEO satellites makes the design of 3D networks challenging, especially in the face of multiple objectives. The design of 3D networks to support constant service qualities in dynamic environments still requires further investigations.

C. Intelligent Algorithms

In the face of dynamically fluctuating fading channels, it is a challenge to formulate accurate system models and/or perceptually meaningful OFs. Hence MOEAs become computationally demanding. This challenge may be circumvented with the aid of artificial intelligence (AI). When using AI and machine learning in designing 3D networks, it is important to find the most suitable learning agents, especially when aiming for real-time learning. In this context, the MOOPs of 3D networks may be transformed into multi-agent and/or multi-task learning problems. Although centralized learning algorithms are conceptually simple, their computational complexity is extremely high, especially in large-scale networks. Therefore, decentralized and distributed learning algorithms constitute more promising solutions for designing 3D networks, especially when the environments are not completely known. However, to benefit from the collaboration of decentralized and distributed learning processes, bespoke learning mechanisms have to be conceived. Moreover, given that 3D networks face a potentially uncertain complex environment, reinforcement learning has to commence its action from totally random trials and gradually develop sophisticated strategies by performing many rewards guided action-trials, which is another promising research direction.

D. Quantum Algorithms

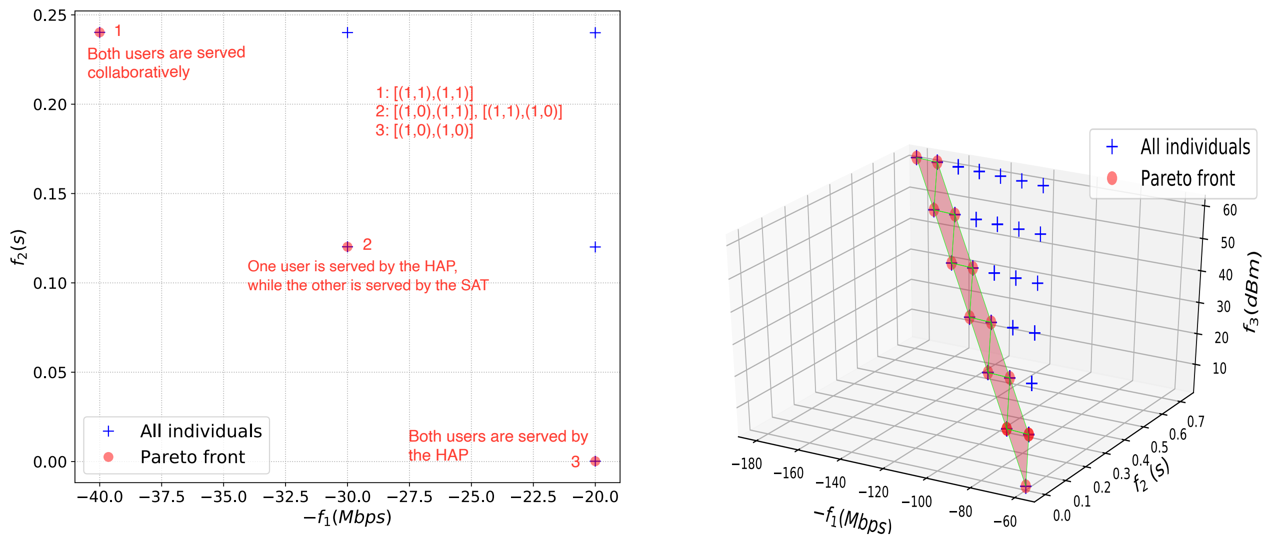
Again, 3D IGAS networks will encompass a huge number of components in the quest for ubiquitous connectivity, which often results in exponentially escalating computational complexity requirements. Fortunately, at the time of writing quantum computing is developing at a fast pace, fuelled by huge investments across the globe. This has also expedited the development and employment of quantum search algorithms, which are eminently suitable for solving large-scale Pareto-optimization problems. In [14], the authors have provided an indication of their computational benefits. Explicitly, given a search-space size of N entries, their quantum algorithm is capable of finding the optimal solutions with a near unity probability by evaluating as few as $22.5\sqrt{N}$ OF values.

VII. Conclusions

The seamless integration of 3D IGAS networks has the potential of providing global coverage, high capacity and always-on connectivity. However, due to the heterogeneity of 3D networks as well as the large variety of performance metrics to be satisfied, numerous new challenges have to be tackled. Single-objective optimization fails to reach the full potential of 3D networks. Hence, we have proposed a framework of multi-objective optimization for 3D networks. Finally, we concluded with a range of future research challenges in optimizing 3D networks.

References

[1] H. Yanikomeroglu, "Integrated terrestrial/non-terrestrial 6G networks for ubiquitous 3D super-connectivity," in *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM '18. New York, NY, USA: ACM, 2018, pp. 3–4.



(a) Two-component OF

(b) Three-component OF

Figure 5: Pareto-optimal solutions for our 3D IGAS network having two layers, where two scenarios with different number of OFs are considered.

[2] T. 38.811. (2019) Study on new radio (NR) to support non-terrestrial networks (Release 15). [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3234

[3] E. Schulz, “Global networks, global citizens, global market forecast 2018-2037,” *Airbus (Ed.). Blagnac Cedex, France*, 2018.

[4] X. Huang, J. A. Zhang, R. P. Liu, Y. J. Guo, and L. Hanzo, “Airplane-aided integrated networking for 6G wireless: Will it work?” *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 84–91, Sep. 2019.

[5] J. Qiu, D. Grace, G. Ding, M. D. Zakaria, and Q. Wu, “Air-ground heterogeneous networks for 5G and beyond via integrating high and low altitude platforms,” *IEEE Wireless Commun.*, vol. 26, no. 6, pp. 140–148, 2019.

[6] A. Al-Hourani and I. Guvenc, “On modeling satellite-to-ground path-loss in urban environments,” *IEEE Communications Letters*, pp. 1–1, 2020.

[7] SaT5G, “SaT5G project announces successful demonstrations of 5G over satellite use cases at EuCNC 2019 event in valencia,” https://www.sat5g-project.eu/eucnc-2019-press-release, accessed: 2019-07-12.

[8] J. N. Pelton, S. Madry, and S. Camacho-Lara, *Handbook of Satellite Applications*, 2nd ed. Springer Publishing Company, Incorporated, 2017.

[9] T. C. Tozer and D. Grace, “High-altitude platforms for wireless communications,” *Electronics Communication Engineering Journal*, vol. 13, no. 3, pp. 127–137, Jun. 2001.

[10] D. Kennedy. (2019) To-Euro-5G – Supporting the European 5G PPP. [Online]. Available: https://cordis.europa.eu/project/rcn/211057/results/en

[11] J. Cui, Y. Liu, and A. Nallanathan, “Multi-agent reinforcement learning based resource allocation for UAV networks,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, 2020.

[12] E. Zitzler, M. Laumanns, and S. Bleuler, “A tutorial on evolutionary multiobjective optimization,” in *Metaheuristics for Multiobjective Optimisation*, X. Gandibleux, M. Sevaux, K. Sörensen, and V. T’kindt, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 3–37.

[13] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Trans. Evolut. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[14] P. Botsinis, D. Alanis, Z. Babar, H. V. Nguyen, D. Chandra, S. X. Ng, and L. Hanzo, “Quantum search algorithms for wireless communications,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1209–1242, Secondquarter 2019.

Jingjing Cui (jingj.cui@soton.ac.uk) is a research fellow with the School of Electronics and Computer Science, University of Southampton, United Kingdom. Her research interests include optimization theory for wireless communications and quantum communications. She is a member of the IEEE.

Soon Xin Ng (Michael) [S’99-M’03-SM’08] received the B.Eng. degree (First class) in electronic engineering and the Ph.D. degree in telecommunications from the University of Southampton, Southampton, U.K., in 1999 and 2002, respectively. He is currently a Professor of Next Generation Communications at the University of Southampton. His research interests include radio communications, quantum communications, artificial intelligence and machine learning. He is a Senior Member of the IEEE, a Fellow of the Higher Education Academy in the UK, a Chartered Engineer and a Fellow of the IET.

Dong Liu (S’13-M’19) received his B. Eng. and Ph.D. degree from Beihang University (BUAA), Beijing, China in 2013 and 2019, respectively. He is currently a research fellow at the University of Southampton, UK. His recent research interests include mobile/wireless AI and space-air-ground integrated networks.

Jiankang Zhang (S’08-M’12-SM’18) is a Senior Lecturer at Bournemouth University. Prior to joining in Bournemouth University, he was a senior research fellow at University of Southampton, UK. Dr Zhang was a lecturer from 2012 to 2013 and then an associate professor from 2013 to 2014 at Zhengzhou University. His research interests are in the areas of aeronautical communications, aeronautical networks, evolutionary algorithms and edge computing. He serves as an Associate Editor for IEEE ACCESS.

Arumugam Nallanathan (a.nallanathan@ qmul.ac.uk) is a professor of wireless communications and the head of the Communication Systems Research Group, Queen Mary University of London, U.K. His research interests include Beyond 5G wireless networks and Internet of Things. He is an IEEE Fellow.

Lajos Hanzo (lh@ecs.soton.ac.uk) holds the chair in telecommunications at the School of Electronics and Computer Science, University of Southampton, United Kingdom. He is currently working on a range of research projects in the field of wireless multimedia communications. He is a Fellow of the IEEE.