Flying Mobile Edge Computing towards 5G and beyond: An Overview on current use cases and challenges

Pedro Cumino, Susana Sargento

Abstract—The increasing computational capacity of multiple devices, the advent of complex applications, and data generation create new challenges of scalability, ubiquity, and seamless services to meet the most diverse network demands and requirements, such as reliability, latency, battery lifetime. For this reason, the 5th Generation (5G) network comes to mitigate the most diverse challenges inherent to the current dynamic mobile networks and their increasing data rates. Unmanned Aerial Vehicles (UAVs) have also been considered as communication relays or mobile base stations to assist mobile users with limited or no available wireless infrastructure. They can provide connections for mobile users in hard-to-reach areas, replacing damaged or overloaded ground infrastructure and working as mobile clouds, providing low but increasing computational power. However, the feasibility of a Flying Edge Computing requires special attention in terms of resource allocation techniques, cooperation with existing ground units and among multiple UAVs, coordination with user mobility, computation efficiency, collision avoidance, and recharging approaches. Thus, the cooperation among UAVs and the current terrestrial Mobile Edge Computing can be relevant in some cases once the computation power of a single UAV might be insufficient. It is important to understand the feasibility of current proposals and establish new approaches that consider the usage of multiple UAVs and recharging approaches. In this paper we discuss the challenges of a 5G extended network through the help of UAVs. The proposed multi-tier architecture employs UAVs with different mobility models, providing support to ground nodes. Moreover, the support of the UAVs as edge nodes will also be evaluated.

Index Terms—UAV, Mobile Edge Computing, Service Migration, Task offloading, Task Allocation

I. INTRODUCTION

The increasing computational capacity of multiple devices, the advent of complex applications, and data generation create new challenges of scalability, ubiquity, and seamless services to meet the most diverse network demands and requirements, such as reliability, latency, battery lifetime, and QoE. In the context of a Smart City, wireless networks play a crucial role, and its evolution needs to be continuous. For this reason, the network of the 5th Generation (5G) comes to mitigate the most diverse challenges inherent to the current dynamic mobile networks and their increasing data rates.

The employment of Small Cells (SCs), building an Ultra-Dense Network (UDN), comes as a standard approach in order to fulfill the current network requirements. However, there are still technical challenges to address, such as the high data rate provisioning in a short time range, more efficient resource allocation, and the need to coexist within complex environments of multiple heterogeneous technologies [1]. Despite being a promising approach to meet the growing end-user demand for communication capacity and providing a better coverage area, the densification of SCs base stations in a 5G architecture and the usage of such terrestrial hubs still suffer from an inefficient non-Line of Sight link limitations and unavailability of adequate infrastructure in an urban and remote area. Therefore, the usage of UAVs as network platforms can potentially provide more flexible coverage. Such devices can adapt dynamically according to the mobility of its users, a mission, and communication and operational costs.

With the emergence of UAVs, also followed by their rapid evolution, their use has been progressively viable in 5G scenarios. For instance, to attend not only the most restricted and hard-to-reach scenarios, but also areas with high data rate requirement, being considered as communication relays or mobile base stations (BS) to assist mobile users with limited or no available resources or wireless infrastructure. Research has also opened the perspective for UAV-based edge computing, where the devices make their resources available for different network requirements, such as data caching and processing. Besides, the development of resource allocation by UAVs has also been essential to making smart city applications feasible [2].

The continuous and rapid increase in the computational capability of these devices has enhanced the use of UAVs in the composition of Unmanned Aerial Networks (UANs). Improved hardware for wireless communication has been equally essential to build the UANs, making the UAVs the key to enabling more ubiquitous Mobile Edge Computing in hard-to-reach scenarios with high computational demand.

The rest of this position paper highlights the main applications, scenarios and insights of network topology and architecture designs in Section II, and presents the most relevant and current solutions of related works in Section III. The main current issues, future contributions as well as a conclusion completes the paper as described in Section IV.

II. APPLICATIONS AND SCENARIOS

In the 5G and high-bandwidth scenarios as, shown in Figure 1, flash crowd events are the ones that require an increased data rate in a short time range. UAVs have been considered as an option to either substitute or cooperate with pre-existing methods. UAVs are usually cheaper, faster, effective and can be used for short-term events, due to their ability to facilitate on-demand deployment without highly constrained and expensive infrastructure, high flexibility in network reconfiguration by dynamically changing their positions to provide on-demand communications to ground users in emergency or crowded events’ situations, high chance of having Line of Sight (LoS) communication links. In addition,
multiple UAVs might collaborate with each other to form a UAV network in order to establish a backbone integrated with the ground stations. This network can add more capacity in non-permanent scenarios, provide high-bandwidth communication to the ground and widespread coverage through the extension of, e.g., 5G coverage in multi-hop high-bandwidth wireless links.

Fig. 1. 5G and UAVs enabled scenarios.

Having the chance to feel immersed in a specific environment or venue (sport fields, music halls, etc.) is one of the most promising benefits of 5G networks and beyond; they may allow people to feel more connected to the main actors of an event, and even provide them with a whole new level of freedom to improve their experience. For instance, in a soccer stadium, users are limited to the perspective from where they stand throughout the entire game, instead of having the possibility to freely select a viewpoint they find interesting (e.g. behind the goalie, above the player carrying the ball, always following the referee, etc.). In a concert, the need for high bandwidth communication to share videos, video lives, and all envisioned services, requires a temporary strong network to handle this. UAV networks will be able to extend and increase the capacity of the backhaul network for these crowd and sudden events.

UAVs working as aerial BS provide a promising solution to enable fast, flexible, and reliable wireless communications in public safety scenarios, disaster-affected areas, or as a network of aerial sensors to collect data in large areas [3], [4]. UAVs are autonomous and low-cost aerial devices with the possibility of being equipped with sensing and communication devices, and can be placed in specific locations to help BSs that are providing a bad Quality of Service (QoS) or replacing the damaged ones [5]. Therefore, mobile users can choose to connect to a nearby BS or UAV in order to get better connectivity [6], [7].

Real-time surveillance and monitoring scenarios, post-disaster recovery, access to rural broadband, and military operations are always generating critical, information-rich sensing data. The collected data usually needs to be quickly sent and processed in order for the information generated to be as accurate as possible. In most cases, cloud computing is not reachable via a direct link. Thus, taking into account the Mobile Edge Computing paradigm, the use of multiple UAVs as forwarding nodes, as well as processing units can be a promising approach to provide reachable computational resources in hard-to-reach scenarios [8].

UAVs have become more and more capable of processing different tasks. Some proposals have been employing UAVs as relay nodes for video dissemination and aerial BS for user data offloading [9]. Others have been employing it as flying processing units to support applications such as detection and tracking systems, image analysis, and object recognition [10]. Thus, Flying Ad hoc Networks (FANETs) can flexibly provide computational resources and improve the Quality of Service in many other scenarios. For instance, the installation of computing devices, such as Road Side Units (RSUs), at the edge of the network and closer to the users, has been a solution to support delay-sensitive applications in VANET scenarios. The main goal of such approach is to reduce the delay in communication and then to improve the quality of service. However, these devices become overloaded due to a large number of computing requests when considering congested areas, resulting in degradation of overall system performance. Due to its peculiar characteristics, Flying Edge Computing (FEC) can potentially be a solution for this issue once it can provide and allocate resources more flexibly and dynamically [8].

A. Topology and Architectures

In order to provide smart city services in a high-density scenario, high bandwidth links between the UAVs, the UAVs and the users, and the UAVs and other ground infrastructure need to be in place. Moreover, beyond providing a high communication link and efficient network for communication, UAVs are able to be controlled to improve the link and connection reliability and quality. This is a major difference from other networks: the possibility to adjust the position and orientation of the network nodes dynamically and in real-time to improve the network communication. Positioning the UAVs between themselves and between themselves and the users will be critical for the high-bandwidth links communications.

A multi-tier perspective can bring advantages such as scalability, reliability, information delivery, network formation management [11]. In 5G, scalability and reliability issues must be efficiently addressed. Software-Defined Network (SDN) has been one of the most efficient approaches to address those issues. It introduces complete network programmability by separating the control plane and data plane. SDN in Flying Ad hoc Networks (FANETS) also considers a centralized SDN controller responsible for all control functions such as UAVs placement, collision avoidance, network routing, and other services [12]. However, an SDN architecture usually employs ground BS as the controller unit. This fact can bring scalability and coverage limitations because it is directly related to the communication range between the controller and the rest of the network. Moving the controller entity to the air might lead to a more flexible network in terms...
of topology formation. Besides, once defining a set of UAVs as part of the controller entity, the position of the controller can follow more dynamic scenarios.

It is also important to highlight that the higher the altitude of a UAV, the greater the probability of having an air-to-ground communication link, since there is a higher probability of LoS. However, the increased distance between a UAV and a ground node also increases the path loss. The trade-off between the coverage of a UAV and its distance to the ground mobile users is an issue that a multi-tier UAV architecture can mitigate [13]. A multi-tier architecture employs UAVs with different mobility models, where the stationary ones fly in a higher altitude, providing coverage for UAVs of lower altitude, which in turn are responsible for giving support to most of the main services that runs on ground nodes. This approach can also be suitable to manage the path planning challenges of aerial vehicles and its peculiar mobility in 3 dimensions [14].

On the one hand, a set of higher altitude UAVs along with abundant resource nodes, such as ground BSs, RSUs, VANETs, comprises the control plane in our architecture. In the control plane, all nodes reach each other through Air to Ground (A2G) and Air to Air (A2A) connections, regarding the connection range limitations of each node. Higher altitude UAVs are responsible for maintaining the connection between nodes in the control plane, with the primary objective of ensuring that control messages are transmitted as efficiently as possible. On the other hand, the set of UAVs with lower altitude comprises the service or application plane. The architecture relies on the connection between the lower altitude UAVs and the controller entity. Its goal is to provide services to nodes with a lack of resources.

III. RELATED WORKS

The emerging mobile applications, such as augmented reality, face recognition, and mobile online gaming, are latency-sensitive and need a high computation capability. It becomes more critical once the requirements are needed in a short time range, such as during traffic accidents, emergencies, post-disaster relief, and police tracking services. A UAV network can potentially be employed in order to provide the needed resources for User Equipment (UE) and IoT devices in such scenarios [15]. The UAVs can also opportunistically use resources from ground BSs, RSUs, or vehicular nodes [16]. These scenarios require reliable path planning management, which must consider the available resources and the trajectory of every node. This information must be jointly addressed with the computation offloading scheduling in order to meet the service requirements. For instance, during flash-crowd and unpredictable events, the mobility of nodes can perform differently from the usual pattern. Thus, it is crucial to perform the task offloading service based on the mobility of each node following a prediction approach. Besides, the UAV network must be able to follow the nodes based on their computation capabilities and energy constraints. Thus, a task offloading approach must prioritize the critical nodes, where there is a lack of resources, and be aware of the nodes with higher capacity, which will provide the required resources.

In the literature, a UAV has been broadly considered a limited device in terms of energy. A vast amount of proposals show that an efficient management system must consider energy as the main restriction leading to computational power limitations. For instance, the authors in [9] proposed a task offloading formulation based on game theory, where the UAVs offload their tasks to available resources such as base station or another nearby wireless network. The offloading of intensive computational tasks approach, as the authors claim, improves energy overhead, and decreases execution delay.

Artificial Intelligence (AI) related applications are getting more and more out of UAVs. The flexibility of such devices allows their use in real-time applications such as AI processing, object recognition and navigation. However, even considering multiples UAVs to run the same application, they still face processing power and communication limitations, as highlighted in [17]. Thus, the work in [17] presented an edge computing model to extend the computational resources — processing power and communication limitations — of UAV swarms. The UAV, cloud, and edge computing models compose the proposed solution, considering all their single aspects and advantages. Firstly, this model regards not only the cooperation among UAVs for performing tasks, processing data, and communicate collaboratively, but also the cooperation from edge servers to reduce the communication latency. The edge servers compute the tasks close to the UAV swarm, which can potentially enhance collision and obstacle avoidance procedures. Secondly, the cloud computing model provides abundant computational resources to help process and store large-scale gathered data for the UAV swarms, and provide a solution for UAV swarms to sporadic wireless network connectivity, or in scenarios where UAV swarms are unable to connect a cloud access point directly. Moreover, the authors also claim the need for a joint task placement and routing planning optimization. The algorithm for optimization follows the Markov Approximation, where the task allocation procedure considers a timer for each pair of tasks, and the timer value will decide whether a route to a computing entity (UAV, edge, or cloud) will be updated or not. This work shows a better system performance when considering the joint of routing and task allocation optimization.

In post-disaster relief scenarios, applications such as face recognition in search and rescue operations require a short response time since they are life-and-death matters. The usage of UAVs’ computational resources has become a convenient solution to address the response time issue in these applications, due to their mobility characteristic and low deployment cost. To mitigate those issues, the authors in [18] presented an hybrid architecture that integrates FANET and Vehicle Fog Computing (VFC), considering UAVs, ground vehicles and RSUs as processing unit candidates. In this architecture, the tasks offloaded to UAVs can then be allocated to vehicles or RSUs with idle computational resources. The proposed architecture is divided into three layers: the sensing, the management, and the processing layers. The sensing layer is responsible for gathering the information using the sensing capabilities of the UAVs. The processing layer pre-processes the gathered information in order to remove redundancies before the offloading decisions. Besides, it gathers contextual
information such as weather conditions and the status information of each UAV — the battery level, speed, location, topological information, and so. The paper evaluates the performance by comparing the average response delay, which increases faster in a UAV to cloud approach than in a UAV to fog approach.

With technological advances, new perspectives have been considered. In this case, the computational capacity of UAVs has become an increasingly relevant factor, since this device becomes a potential candidate to make its resources available to offload tasks from ground nodes. For instance, the authors in [8] presented an image analysis application that is processed in a network that considers three in-line UAVs and a ground control station. In this work, the UAVs’ processing power is employed as an alternative for task computation, making the offloading process to UAVs possible. The authors evaluate the performance in terms of computation response time and flight time by allocating the tasks in different sets of nodes. The image analysis application tasks were processed at the UAVs or the Ground Control Station (GCS). There is a trade-off between processing the task entirely at the UAVs or the GCS. The performance of the processing tasks relies on the number of iterations and the available cores. Besides, flight time is also affected depending on the number of cores used.

Applications such as face/fingerprint recognition augmented/virtual reality and some lightweight machine-learning tasks are in surging demands for computation capacity of resource-constrained IoT mobile devices (IMDs). In parallel, UAVs with cloud-like computing capabilities has also been considered as a potential candidate to support the ever-increasing requirement for computational and communication resources. The authors in [19] present a joint optimization of trajectory and bit allocation leading to the energy consumption optimization in a task offloading scenario. Similarly, the authors in [15] proposed an optimized trajectory to provide computation offloading from ground users to a single UAV. It relies on a convex problem formulation, which is the minimization of the total energy consumption regarding the number of the offloading computation bits of the users, the trajectory of the UAV, and the CPU frequency of both. The simulation scenario considered four users and a single UAV. The presented results show energy and trajectory optimization.

In congested environments, the number of computational resource requests significantly increases in a short time range. Such situations may happen during traffic jams or any other people’s short-time agglomerations. Thus, the authors in [16] designed a computing framework that takes into account the computational limitations of congested RSUs, by employing additional compute resources in the form of a set of UAVs. The set of UAVs acts as a flying fog-based resource provisioning, giving access to computational resources from nearby available RSUs during a leasing time, which varies according to the number of tasks and the time needed to compute the tasks queue in the vehicles and RSUs. Firstly, the implementation considers the vehicle task generation, processing, and offloading procedures, and each vehicle decides whether the task is computed locally or at an RSU. Secondly, an RSU enqueues all the offloaded tasks, calculates the time to process its queue, and eventually sends flying fog resources requests to a nearby base station. Finally, the UAV allocation is performed by the base station, and based on the workload of the requesting RSUs, which considers the current task queue length and average computation time of the RSU, and the entire system workload.

On the one hand, unanimously, energy consumption by UAVs is a cost that must be mitigated in all circumstances: both in task unloading operations from terrestrial nodes to be processed in UAVs, and in the allocation of tasks in high computational capacity nodes (BSs, vehicles, and RSUs), which further strengthens the success of using UAVs for both contexts. On the other hand, some works consider the computational capacity of a UAV as an advantageous factor, while other works consider it a disadvantage. This divergence occurs since, in the beginning, UAVs were introduced for sensing, thus not having sufficient computational capacity for more complex tasks. However, nowadays, its processing is becoming more and more effective and viable. This becomes most evident when its computational power is combined in a network format.

UAVs have been more and more integrated to network solutions and applications throughout the development of UAV based solutions to manage and improve many services, such as task offloading and migration. Bringing computational capabilities closer to users through such devices has been proven to be an essential approach to improve QoS in terms of reduced latency and increased throughput. UAVs can be potentially employed as computational units closer to the respective clients. However, there are still challenges in terms of topology, migration, and path planning management in order to provide more reliable services. Therefore, this work proposes an architecture that employs UAVs as part of a flying fog network and able to manage different tasks through their offloading and migration among UAVs and other nodes with different computational and energy resources.

IV. OPEN CHALLENGES AND FUTURE WORKS

As highlighted in the previous section, many approaches have been proposed to enable the usage of UAVs in different ways. The works show how versatile a UAV can be in a smart city scenario once it flexibly provides connectivity and processing power. However, the feasibility of a Flying Edge Computing requires special attention in terms of resource allocation techniques, cooperation with existing ground units and among multiple UAVs, coordination with user mobility, computation efficiency, collision avoidance, and recharging approaches.

The usage of multiple UAVs remains a challenge. Grouping multiple users and UAVs is typically an integer non-convex optimization problem, making it difficult to find an optimal solution. In order to investigate this issue, clustering algorithms shall be studied as well as other Machine Learning and Artificial Intelligence techniques. Besides, ground device mobility must be taken into account due to the dynamics of the scenario, such as in post-disaster, traffic congestion, object tracking, and search and rescue operations scenarios.
Resource allocation and task offloading are deeply related. These services can assist devices with low available resources by providing computation power. However, the decision-making process can be challenging once it relies on contradictory requirements, such as low latency, increased throughput, low overhead, support of a massive number of devices, and dynamic conditions. Thus, the cooperation among UAVs and the current terrestrial Mobile Edge Computing can be relevant in some cases once the computation power of a single UAV might be insufficient. In other cases, where increased signaling overhead and transmission latency is a challenge, there is a need to improve current algorithms and approaches.

The energy will always be a challenge when considering UAVs. Despite being narrowly investigated, recharging approaches put a light in the shadow. However, the coordination of recharging schedule and service provisioning shall be investigated, especially when the number of UAVs is limited.

The feasibility of a Flying Edge Computing network architecture has been notably studied. However, there are still crucial challenges that shall be addressed in order to improve the state-of-arts. Future works aim to understand the feasibility of current proposals, and develop a new architecture that consider the usage of multiple and cooperative UAVs to support ground network in the most diverse scenarios. Besides, new UAV position optimization, mobility and trajectory planning, and task offloading and allocation approaches will be studied.

ACKNOWLEDGMENT

This work has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie ITN TeamUp5G, grant agreement No. 813391.

REFERENCES

[1] S. A. W. Shah, T. Khattab, M. Z. Shakir, and M. O. Hasna, “A distributed approach for networked flying platform association with small cells in 5G+ networks,” in GLOBECOM 2017-2017 IEEE Global Communications Conference. IEEE, 2017, pp. 1–7.

[2] F. Zhou, R. Q. Hu, Z. Li, and Y. Wang, “Mobile edge computing in unmanned aerial vehicle networks,” IEEE Wireless Communications, vol. 27, no. 1, pp. 140–146, 2020.

[3] S. Hayat, E. Yamaz, and R. Muizzar, “Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint,” IEEE Communications Surveys Tutorials, vol. 18, no. 4, pp. 2624–2661, 2016.

[4] A. Merwaday and I. Guvenc, “Uav assisted heterogeneous networks for public safety communications,” in 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), March 2015, pp. 329–334.

[5] Y. Zeng, R. Zhang, and T. J. Lim, “Wireless Communications with Unmanned Aerial Vehicles: Opportunities and Challenges,” IEEE Communications Magazine, vol. 54, no. 5, pp. 36–42, 2016.

[6] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, “On the number and 3d placement of drone base stations in wireless cellular networks,” in 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Sep. 2016, pp. 1–6.

[7] C. Luo, W. Miao, H. Ullah, S. McLean, G. Parr, and G. Min, “Unmanned aerial vehicles for disaster management,” in Geological Disaster Monitoring Based on Sensor Networks. Springer, 2019, pp. 83–107.

[8] G. Secinti, A. Trotta, S. Mohanti, M. Di Felice, and K. R. Chowdhury, “Focus: Fog computing in uas software-defined mesh networks,” IEEE Transactions on Intelligent Transportation Systems, 2019.

[9] M.-A. Messous, H. Seddajmaci, N. Houari, and S.-M. Senouci, “Computation offloading game for an uav network in mobile edge computing,” in 2017 IEEE International Conference on Communications (ICC). IEEE, 2017, pp. 1–6.

[10] X. Zhao, F. Pu, Z. Wang, H. Chen, and Z. Xu, “Detection, tracking, and geolocation of moving vehicle from uav using monocular camera,” IEEE Access, vol. 7, pp. 101 160–101 170, 2019.

[11] J. Wu, “A multi-tiered network with aerial and ground coverage,” Computer Communications, vol. 131, pp. 39–42, 2018.

[12] Z. Zhao, P. Cummo, A. Souza, D. Rosário, T. Braun, E. Cerqueira, and M. Gerla, “Software-defined unmanned aerial vehicles networking for video dissemination services,” Ad Hoc Networks, vol. 83, pp. 68–77, 2019.

[13] S. Sekander, H. Tabassum, and E. Hossain, “Multi-tier drone architecture for 5g/b5g cellular networks: Challenges, trends, and prospects,” IEEE Communications Magazine, vol. 56, no. 3, pp. 96–103, 2018.

[14] I. K. Singgih, J. Lee, and B.-I. Kim, “Node and edge surveillance using drones considering required camera altitude and battery charging,” in IFIP International Conference on Advances in Production Management Systems. Springer, 2018, pp. 172–180.

[15] F. Zhou, Y. Wu, H. Sun, and Z. Chu, “Uav-enabled mobile edge computing: Offloading optimization and trajectory design,” in 2018 IEEE International Conference on Communications (ICC). IEEE, 2018, pp. 1–6.

[16] N. Madan, A. W. Malik, A. U. Rahman, and S. D. Ravana, “On-demand resource provisioning for vehicular networks using flying fog,” Vehicular Communications, p. 100252, 2020.

[17] W. Chen, B. Liu, H. Huang, S. Guo, and Z. Zheng, “When uav swarm meets edge-cloud computing: The qos perspective,” IEEE Network, vol. 33, no. 2, pp. 36–43, 2019.

[18] C. Tang, C. Zhu, X. Wei, H. Peng, and Y. Wang, “Integration of uav and fog-enabled vehicle: Application in post-disaster relief,” in 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS). IEEE, 2019, pp. 548–555.

[19] J. Xiong, H. Guo, and J. Liu, “Task offloading in uav-aided edge computing: Bit allocation and trajectory optimization,” IEEE Communications Letters, vol. 23, no. 3, pp. 538–541, 2019.