Computers, Materials & Continua
DOI:10.32604/cmc.2021.016087
Article

Tech Science Press

Reinforcement Learning-Based Optimization for Drone Mobility in 5G and Beyond Ultra-Dense Networks

Jawad Tanveer[1], Amir Haider[2], Rashid Ali[2] and Ajung Kim[1,*]

[1]School of Optical Engineering, Sejong University, Seoul, 05006, Korea
[2]School of Intelligent Mechatronics Engineering, Sejong University, Seoul, 05006, Korea
[*]Corresponding Author: Ajung Kim. Email: akim@sejong.ac.kr

Received: 22 December 2020; Accepted: 23 February 2021

Abstract: Drone applications in 5th generation (5G) networks mainly focus on services and use cases such as providing connectivity during crowded events, human-instigated disasters, unmanned aerial vehicle traffic management, internet of things in the sky, and situation awareness. 4G and 5G cellular networks face various challenges to ensure dynamic control and safe mobility of the drone when it is tasked with delivering these services. The drone can fly in three-dimensional space. The drone connectivity can suffer from increased handover cost due to several reasons, including variations in the received signal strength indicator, co-channel interference offered to the drone by neighboring cells, and abrupt drop in lobe edge signals due to antenna nulls. The baseline greedy handover algorithm only ensures the strongest connection between the drone and small cells so that the drone may experience several handovers. Intended for fast environment learning, machine learning techniques such as Q-learning help the drone fly with minimum handover cost along with robust connectivity. In this study, we propose a Q-learning-based approach evaluated in three different scenarios. The handover decision is optimized gradually using Q-learning to provide efficient mobility support with high data rate in time-sensitive applications, tactile internet, and haptics communication. Simulation results demonstrate that the proposed algorithm can effectively minimize the handover cost in a learning environment. This work presents a notable contribution to determine the optimal route of drones for researchers who are exploring UAV use cases in cellular networks where a large testing site comprised of several cells with multiple UAVs is under consideration.

Keywords: 5G dense network; small cells; mobility management; reinforcement learning; performance evaluation; handover management

1 Introduction

In 5th generation (5G) wireless networks, drone technology has a significant impact due to its wide range of applications. Large companies and entrepreneurs worldwide carry out numerous tasks using drones, and thus the popularity of these flying mini robots is increasing rapidly [1]. Drones play roles in education, defense, healthcare, disaster relief, surveillance, telecommunications, space, journalism, food services, and emergency response applications [2,3]. Since the deployment of 5G, drone applications have gradually increased with the reshaping of use cases and technology. Over the last decade, the growth of unmanned aerial vehicles (UAVs) has been magnificent, and low altitude commercial drone endeavors have led to a sufficiently high air traffic. With this increased air traffic, the safe flight operation while maintaining reliable connectivity to the network has been the most critical issue in drone mobility faced by the cellular operators [4,5]. For safe and secure flight, some critical use case applications require milliseconds end–end latency, e.g., in medical operations and emergency response teams. Moreover, cyber confrontations, confidentiality and privacy concerns, and public protection are also leading challenges. Time-sensitive drone applications required seamless connectivity via cellular infrastructure with ultra-reliable low-latency communications [6].

5G empowers a new era of the internet of everything. A user will facilitate high data rate internet speed with ultra-reliable low latency communications (URLLC) services. The enhanced mobile broadband (eMBB) services will enable high-speed internet connectivity for several use cases, such as public transportation, large-scale events, and smart office. In contrast, low-power, wide area technologies include narrow-band internet of everything for massive machine-type communications (mMTC) [7,8]. Cellular technologies such as the 5G new radio spectrum provide abundant higher data throughput rates, and ultra-dense networks offer additional capacity using offloading. The inclusion of higher-order modulation and coding schemes, such as millimeter-wave in 5G new radio, can enable data rates beyond 10 Gbps while using less bandwidth. Moreover, massive multiple-input multiple-output with beam-steering offers energy efficiency and user tracking services. The ultra-lean design of 5G new radio architecture promises to reduce energy consumption and interference by combining multiple subchannels within a single channel. Furthermore, the transmit power of the base station (BS) can be focused in a particular direction to increase the coverage range of the cell [9]. 5G also enables traffic management of unmanned aircrafts at a commercial level. New drone applications will be entertained beyond visual line of sight flights, where low-altitude operations such as those below 400 ft/120 m are allowed worldwide. Sensor data transmission will be used for live broadcasting data transmission [10]. Although 5G NR ensures seamless and ubiquitous connectivity for low-mobility users, there remain some key challenges to be addressed in case of the high-mobility users, particularly for the UAV-based communication. A UAV may carry different apparatuses up to hundreds of kilograms depending on weight, route interval, and battery capacity.

Moreover, in contrast to terrestrial vehicles, UAVs suffer from several key limitations such as inadequate communication links, limited energy sources, variation in the network topology, and Doppler effect due to high air mobility. Machine learning (ML) techniques are anticipated to deliver improved network performance solutions, channel modeling, resource management, positioning, interference from the terrestrial node, and path-loss in drone handover, all with minimum computation. ML algorithms have been proposed as key enablers for decisions making in UAV-based communications, e.g., in the UAV swarm scenario, numerous drones share network resources in an optimal manner [11].

These technologies support many UAV services, and as a result, the drone will obtain more than 500 km/h high-speed mobility with a maximum latency of 5 ms. There are also some mobility challenges in modern technologies with respect to drones in a 5G cellular network [12]. First, rapid changes in reference signals such as the received signal strength indicator (RSSI) fluctuate due to flying within a 3D space at high speed. Therefore, the RSSI will rise and fall suddenly, and the drone will face situations where the handover decision is challenging. Second, high constructive and destructive interference of uplink (UL) and downlink (DL) channels from neighbor cells due to drone line of sight propagation conditions also results in handover. Third, the main lobe of BS antennas covers a large portion of the cell with height limitations due to their tilt settings, thus focusing only on ground users or the users inside the buildings. This height limitation results in availability of only the side lobes of BS antennas for UAV connectivity. Since the drone frequently flies along the side lobes of BS antennas, the signals at the antenna lobe edges may drop abruptly, which causes handover [13]. The drone might also fly on the strongest signals from far away BS antennas rather than the closest one with a significantly weaker signal. Since the side lobes of the antenna have limited gain, they can only ensure connectivity over a small area when compared to the main lobe of the BS. This results in unnecessary handover from the main lobe of the serving BS to side lobe of a neighboring BS, and hence the drone's flight may face dis-connectivity [14]. Thus, frequent handover occurs due to the split coverage area provided by BS side lobes and to maintain the best reference signal received power (RSRP) value [15,16]. As mentioned above, unnecessary ping-pong handover will turn into high signaling costs, dis-connectivity, and radio link failure. Hence, ultra-reliable low latency communication between the drone and BS requires an efficient handover mechanism for drone mobility management in the ultra-dense network.

The remainder of this article is organized as follows. Section 2 presents the related work. The problem statement, motivation, and the proposed solution are presented in Section 3. In Section 4, the handover scenario under consideration for UAV mobility along with the reinforcement learning-based solutions to optimize the mobility are discussed. In Section 5, we show a Q-learning-based optimized handover scheme. In Section 6, experimental results and a discussion are presented, and we conclude the paper in Section 7.

2 Related Work

The 3rd generation partnership project (3GPP) specifications of Release-15 and the 5G Infrastructure Public Private Partnership (5G PPP) reports (D1.2-5, D2.1) for drone-based vertical applications in the ultra-dense 5G network were studied in Refs. [17,18]. These studies found that drone mobility is one of the main concerns in existing 5G networks to provide reliable connectivity in ultra-dense scenarios. The simulation results in [19] were acquired under different UAV environments and with UAVs flying at different heights and velocities. The results showed that a UAV may have depreciated communications performance compared to ground UEs since they fly on lobe edges; furthermore, DL interference may cause low throughput and low SINR. In [20], the authors discussed the solutions to mitigate the interference in both the uplink and downlink to maintain an optimal performance. Additionally, mobility scenarios were considered to ensure that the UAVs remained connected to the serving BS despite the increase in altitude and situations where the neighbor BS transmitted signals at full power. In [21], the authors discussed the impact of change in cell association on the SINR of the UAV flying at different altitudes. They also compared performance gains for the 3D beamforming technique and fixed array pattern in existing LTE through simulations. The trajectory design, millimeter-wave cellular-connected UAVs, and cellular-connected UAV swarm are still open issues to accomplish high data rate and ultra-reliable low latency for robust connectivity in 5G drone use cases.

In [22], the authors discussed challenges such as low power, high reliability, and low latency connectivity for mission-critical machine-type communication (mc-MTC). To meet these mc-MTC applications' requirements, the authors considered drone-assisted and device-to-device (D2D) links as alternative forms of connectivity and achieved 40 percent link availability and reliability. In D2D and drone-assisted links, the handover ratio is maximized; however, reliability is still an open issue when dealing with $>$ 500 km/h mobility. Simulation results in [23] showed successful handover in a 4D short and flyable trajectory using multiple-input-multiple-output ultrawideband antenna that considers kinematic and dynamic constraints. However, the authors did not address the issue of accuracy in following the 4D planned trajectory with the drone at low altitude, resulting in several unnecessary handovers while completing a trajectory. In [24], the authors proposed an analytical model using stochastic geometry to illustrate the cell association possibilities between the UAV and BS by considering the handover rates in a 3D n-tier downlink network. However, the proposed model did not result in cost-efficient handover in drone flights with a constant altitude. In [25], the authors proposed interference mitigation techniques for uplinks and downlinks to ensure that the UAV remains in LTE coverage despite increased altitude or worst-case situations in which the neighbor BS transmits a signal at full power. With their proposed technique, a strong target cell for handover could be identified to maintain drone connectivity, but the unneeded handover increased the handover cost.

In [26], the authors proposed a UAV and network-based solutions for UAV performance. The authors considered coverage probability, achievable throughput, and area spectral efficiency as performance metrics. They concluded that as the UAV's altitude rises, the coverage and performance decline; accordingly, drone antenna tilting/configuration can increase the drone's coverage and throughput. For reliable connectivity, the proposed solution enhances the coverage area, channel capacity, and spectral efficiency. However, this solution is not cost-efficient as a fast-moving drone may face number of handovers wherever it meets the strongest BS signal. Meanwhile, in [27], the authors proposed a framework to support URLLC in UAV communication systems, and a modified distributed antenna system was introduced. The link range between the UAV and BS was increased by optimizing the altitude of the UAV and the antenna configuration; additionally, increasing the antenna's range also improved the reliability of drone connectivity. Making decisions using ML will surely reduce the handover frequency rate and achieve the required latency and reliability for URLLC applications in the 5G drone system. In [28], the authors detailed drone handover measurements and cell selection in a suburban environment. Experimental analysis showed that handover rate increases with an increase in drone altitude; however, these results only focused on drone altitude. Drones may face several cell selection points in a fast-moving trajectory where an intelligent algorithm is needed for cost-efficient decision-making. In [29], the authors proposed a fuzzy inference method in an IoT environment where the handover decision depends on the drone's characteristics, i.e., the RSS, altitude, and speed of the drone. Perceptive fuzzy inference rules consider the rational cell associations that rely on the handover decision parameters. Simultaneously, an algorithm that can learn an environment can make a better decision about whether a handover is needed. In [30], the authors proposed a handover algorithm for the drone in 3D space. Their technique is cost-efficient because it avoids recurrent handovers.

Furthermore, Ref. [30] evaluated the optimal coverage decision algorithm based on the probability of seamless handover success rate and the false handover initiation. Their algorithm focused on maximum reward gain by minimizing handover costs. Still, the trajectory was not the optimal route for drone flight. To address this drawback, ML tools can be applied to learn the environment and provide the optimal route. In [31], the authors evaluated the handover rate and sojourn time for a network of drone-based stations. Another factor to consider is that the drone's speed variations at different altitudes introduce Doppler shifts, cause intercarrier interference, and increase the handover rate. In [32], the authors proposed a scheme to avoid unnecessary handovers using handover trigger parameters that are dynamically adjusted. The proposed system enhanced the reliability of drone connectivity as well as minimized the handover cost. Regardless, the complete trajectory was not cost-efficient because the proposed technique did not employ any learning model such as reinforcement learning, and therefore the drone cannot decide which path is most cost-efficient in terms of handover.

3 Problem Statement

In an ultra-dense small-cell scenario, the coverage area among cells is small, and drones may observe frequent handover due to their fast movement. Furthermore, channel fading and shadowing cause of ping-pongs. According to 3GPP, user equipment and drones are focused on strengthening RSRP, as illustrated in Fig. 1. Δ and β values are used to overcome the unnecessary handover. An A3 event is triggered when the RSRP of the neighboring cell becomes higher than the RSRP of the serving cell, resulting in a handover. This is a continuous process, and whenever a drone finds a stronger signal, handover occurs. Thus, these unacceptable handovers are mainly caused by delay and loss of packets, and the link remains unreliable, particularly in the case of mission-critical drone use cases.

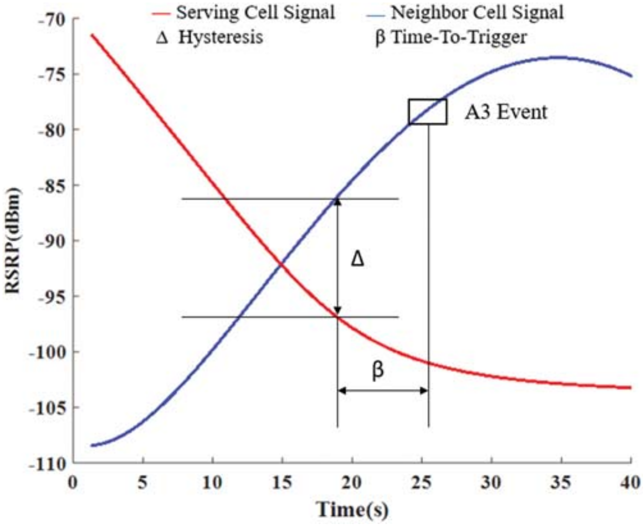

Figure 1: Illustration of the handover mechanism in a cellular network [33]

3.1 Motivation

In 5G dense small-cell deployment, drones face frequent handovers due to the short range of cells. This results in a high signaling cost, cell–drone link reliability issues, and bad user experience, particularly in time-sensitive drone use cases. In real-time scenarios, 5G can meet the requirements of several use cases; however, optimized handover remains an open issue. The baseline handover mechanism in 5G requires critical improvements to ensure seamless connectivity while maintaining a lower handover cost. Hence, a reinforcement learning-based solution will optimize the existing solution since we will compromise the signal strength at some points. By taking the signaling overhead to account, the optimized tradeoff between RSRP of serving cells and handover occurrence will efficiently lower the handover cost.

3.2 Proposed Solution

This study optimizes the handover procedure for a cellular-connected UAV drone that ensures robust wireless connectivity. The drone handover decisions for providing efficient mobility support are optimized with Q-learning algorithms, which are reinforcement learning techniques. The proposed framework considers handover rules from the received signal strength indicator (RSSI) and UAV trajectory information to improve mobility management. Handover signaling overhead is minimized using the Q-learning algorithm, within which the UAV needs to decide whether handover is required, and which handover is the most efficient path. The proposed algorithm depends on RSRP, which aids in the efficient handover decision and minimizes the handover cost. Moreover, the tradeoff between RSRP of serving cells and handover occurrence clearly shows that our Q-learning technique helps optimize this tradeoff and achieve minimum cost for the UAV route, all while considering the handover signaling overhead.

4 System Model

As shown in Fig. 2, the UAV is served by a cellular network within which several BS actively participate. We assume that UAVs fly at a fixed altitude with a two-dimensional (2D) trajectory path, and all information regarding UAV is known to the connected network. The UAV needs to connect to different BSs and perform more than one handover along its route to accomplish the trajectory path with reliable connectivity.

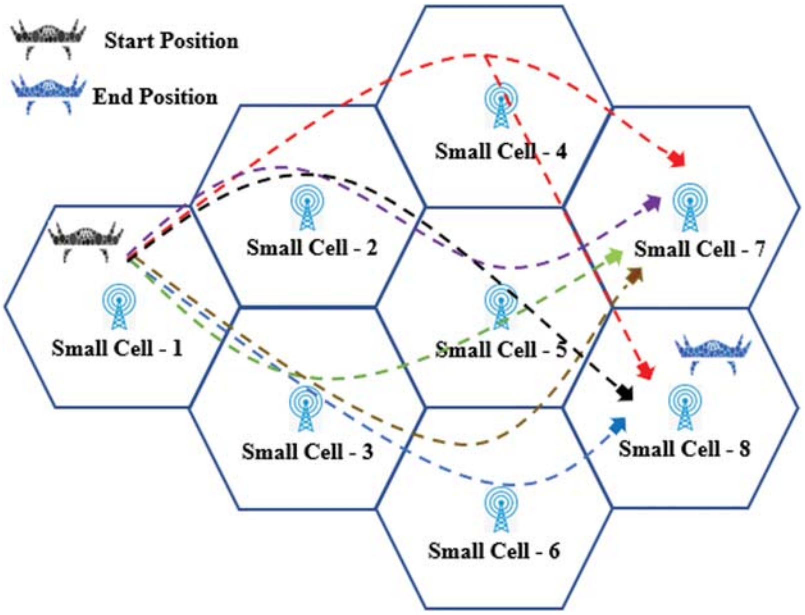

Figure 2: Illustration of the network model

Handover continuously changes the association between the UAV and BS until the UAV reaches its destination. Our study assumes predefined positions along the trajectory wherever the UAV needs to change its association to the next BS or for better connectivity. At every location, the UAV needs to decide whether to do a handover. The steps and signal measurement report involved in handover commands/procedure and admission control are shown in Fig. 3. The BS distribution, UAV speed and trajectory path, RSSI, RSRP (dBm) = RSSI − 10 × log(12 × N), and reference signal received quality RSRQ $=$ (N $\times$ RSRP/RSSI) govern the result of a complete handover process.

UAVs are always hunting for more reliable and robust BSs, such as those with a maximum RSRP value; however, this behavior may be disadvantageous for signaling overhead and reliable connectivity. For instance, every time upon receiving an RSRP value from a neighboring cell BS that is higher that the RSRP value of the serving cell BS, the UAV will trigger handover along its trajectory path, which is costly. This baseline approach introduces ping–pong handover with connectivity failures, such as hasty shifting in RSRP. This expensive solution leads us to construct an efficient UAV handover mechanism in a cellular network that ensures robust wireless connectivity with minimum cost.

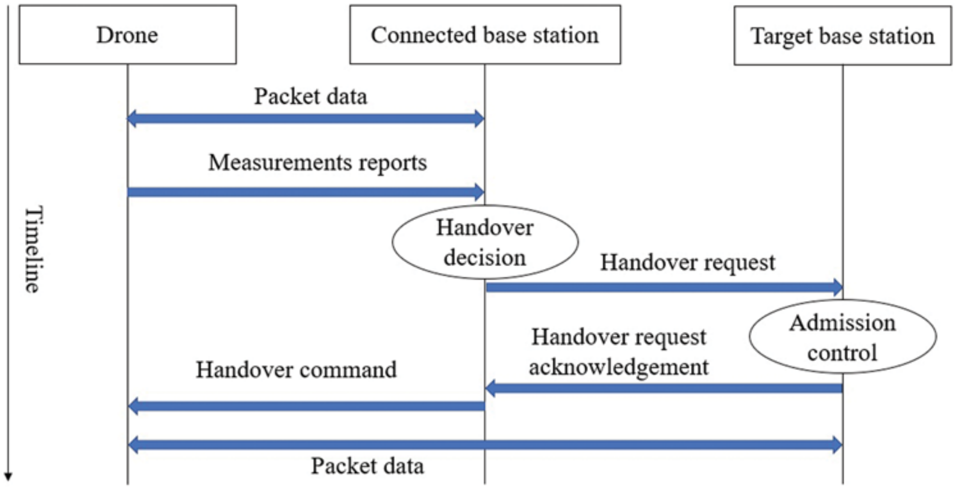

Figure 3: Illustration of the handover process

In this study, we propose a framework for the handover decisions based on the Q-learning technique, which ensures robust connectivity while considering the handover signaling cost. Our proposed Q-learning-based framework will view measurement reports (RSSI, RSRP, and RSRQ) and handover cost as key characteristics for handover decisions. The proposed framework also considers the tradeoff between RSRP values (required maximum) and several handovers (required minimum). Moreover, in the handover decisions, we consider Σ_{HO} and Σ_{RSRP} as weights to adjust the tradeoff between RSRP of the serving cell and the number of handovers. We consider RSRP a substitution for the robust connectivity and number of handovers as a signaling overhead along the whole trajectory. Inherently, our proposed Q-learning-based framework will maintain a good RSRP value with a minimum number of handovers throughout the trip.

4.1 Background of Q-Learning

Reinforcement learning is part of the broad area of ML. It is all about what to do and how to map situations to an action, where an agent takes appropriate action to maximize the reward in a particular state. As shown in Fig. 4, first, the RL agent observes state St and then takes an action $\mathcal{A}_t$ at time t. In response to the action, the agent receives feedback about that action (i.e., reward R_t), and to increase the anticipated action's reward accumulated over time, the agent must choose the appropriate actions. This continues until the algorithm obtains the maximum reward value. RL, an ethical framework described by the Markov decision process (MDP), depends on its problem statement. MDP can be represented as a tuple $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{A}}, \{P_{sa}\}, \lambda, R)$, where S denotes the number of states, A is the set of actions taken by the agent, and P_{sa} provides probability of state transitions for state belongs to the set of states and set of actions. The discount factor is denoted by $\lambda \in [0, 1]$, and R is the reward obtained by $R: S \times \boldsymbol{\mathcal{A}} \to \mathbb{R}$. MDP always aims to obtain the optimum policy, which depends upon the action taken at each state while looking forward to the maximized reward.

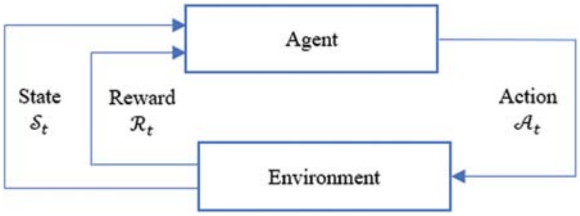

Figure 4: Illustration of Q-learning

Q-learning [34] is an off-policy, model-free, and values-based reinforcement learning algorithm. The objective is to maximize the reward and learn the optimum policy for the given Markov decision process. Let us assume the Q-value $Q^{\pi}(s, a)$ that anticipated the maximized reward for policy π when the Q-learning agent takes an action a in state s and then selects an action regarding policy π. After some learning episodes, the agent will ultimately learn optimal Q-values $Q^{*}(s, a)$, and the highest Q-value for each state establishes an optimal policy. In this study, we donate the Q-value as $Q_t(s, a)$ at time t throughout the process; after receiving the updated reward R_t for the current state s, the learning agent takes action A_o to get a transition to next state s' with reward R_{t+1}. Evaluation of the updated Q-value can be performed using

$$Q_{t+1}(s, a) \leftarrow (1 - \alpha) Q_t(s, a) + \alpha \left[R_{t+1} + \lambda \frac{max}{\alpha' \in \boldsymbol{\mathcal{A}}} Q_t(s', a') \right]. \tag{1}$$

In Eq. (1) Q_{t+1} is the next state value, λ is the discount factor, and α is the learning rate. After performing 250 computations, the Q-learning algorithm learns the optimal Q-values for all states using successive approximation.

4.2 Q-Learning-Based Drone Mobility Framework

This section will briefly describe the state, action, and reward to decide whether the handover is needed along a trajectory path. Moreover, we propose a Q-learning-based algorithm for making the optimal handover decision for the given trajectory path. The main parameters used for the proposed Q-learning-based algorithm for handover optimization are listed in Tab. 1.

4.2.1 Definitions

State: In Fig. 5a, we considered three parameters: the drone's position, represented by $P_{s_o}: (x_{s_o}, y_{s_o})$; the drone's movement direction θs_o, where θ could be $\{k\pi/4, k = 0, \ldots, 7\}$; and the currently connected cell, represented by $C_{s_o} \in C$ (set of all neighbor cells). In our proposed algorithm, we detail the trajectory's initial $(\mathcal{T}_o)$ and final $(\mathcal{T}_m)$ positions. The drone's selected path is the shortest trajectory from the initial to the final position, and the drone always connects to the next predefined BS along the optimized trajectory. In our proposed model, the complete trajectory is not necessarily a straight line because of the fixed number of possible movement directions. Reinforcement learning algorithms are commonly used in drone trajectory path planning. Compared to conventional techniques, the proposed methodology considers the optimized trajectory computed by Q-learning, rather than by adopting a fixed predefined trajectory.

Algorithm 1: $\mathcal{Q}$-value iteration for drone handover scheme using $\mathcal{Q}$-learning

```
Initialize input parameters:
Drone trajectory $\mathcal{T} = \{Pi \mid i = 0, 1, \ldots l - 1\}$;
Set $\mathcal{Q} \leftarrow \mathbf{0}_{l\times\mathcal{V}\times\mathcal{V}}$; $\boldsymbol{HO}_{s_o,s'} \leftarrow \mathbf{0}_{\mathcal{V}\times\mathcal{V}}$;
Set $\Sigma_{HO}, \Sigma_{RSRP}, \epsilon, \alpha, \lambda$;
$C_{\mathcal{V}_s} \leftarrow \mathcal{V}$ strongest cells at starting waypoint $\mathcal{T}_0$;
for $i$ in length $(\mathcal{R}) - 1$ do
    $C_{\mathcal{V}_{s'}} \leftarrow \kappa$ strongest cells at waypoint $\mathcal{T}_{i+1}$;
    $RSRP_{\mathcal{V}_{s'}} \leftarrow$ RSRP values for cells in $C_{\mathcal{V}_{s'}}$;
    $\forall y \in I_{\mathcal{V}}$, set $\mathcal{Q}[i, :, y] \leftarrow, \Sigma_{RSRP} \cdot \text{RSRP}_{\mathcal{V}_{s'}}$;
    Binary matrix $HO_{s_o,s'} \leftarrow C_{\mathcal{V}_s} \neq C_{\mathcal{V}_s'}$;
    $\mathcal{Q}[i, :, :] \leftarrow \mathcal{Q}[i, :, :] - \Sigma_{HO} \times \boldsymbol{HO}_{s_o,s'}$;
    $C_{\mathcal{V}_s} \leftarrow C_{\mathcal{V}_s'}$;
end for
Reward matrix $\boldsymbol{R} \leftarrow \mathcal{Q}$;
While training step < $n$ do
    $j = 0$;
    $\epsilon$-greedy algorithm:
    for $i$ in length $(\mathcal{R}) - 1$ do
        If $\epsilon >$ uniform random value on interval, $[0, 1]$ then
            $j_{new} \leftarrow \frac{\text{argmax}}{u \in I_{\mathcal{V}}} \mathcal{Q}[i, j, u]$;
        else
            $j_{new} \leftarrow$ pick a random number from $I_{\mathcal{V}}$;
        end if
        Update $\mathcal{Q}$–values:
            $\mathcal{Q} = \mathcal{Q}[i, j, j_{new}]$;
            $\mathcal{Q} = (1 - \alpha) \cdot \mathcal{Q} + \alpha \cdot \boldsymbol{R}\left[i, j, j_{new}\right] + \alpha\lambda \underset{u \in I_{\mathcal{V}}}{\mathit{argmax}} \mathcal{Q}[i + 1, j_{new}, v]$;
            $\mathcal{Q}\left[i, j, j_{new}\right] = \mathcal{Q}$;
        end for
    $j = j_{new}$;
end while
return $\mathcal{Q}$
```

Action: Selection of the next serving cell for the next state s' depends on the drone's action A_{s_o} at the current state s_o. For instance, as shown in Fig. 5b, the drone shifts to the serving cell $C_{s'} = 4$ if $A_{s_o} = 4$ at state s'.

Table 1: Definitions in our model related to RL

Label	Definition
$\mathbb{C}\,(HO)$	Handover cost
Σ_{HO}	Weight for handover cost
Σ_{RSRP}	Weight for serving cell RSRP
R	Reward defined as a weighted combination of handover cost and RSRP
$\boldsymbol{\mathcal{S}}_{\mathrm{o}}$	State defined as$[x_o, y_o, \theta_o, c_o]$
(x_o, y_o)	Position coordinate at state $\boldsymbol{\mathcal{S}} = s_o$
θ_{s_o}	Movement direction at state $\boldsymbol{\mathcal{S}} = s_o$
c_{s_o}	Serving cell at state $S = s_o$
$\boldsymbol{\mathcal{S}}'$	Next state of $\boldsymbol{\mathcal{S}}_o$
A_o	Action performed at state S_o
A'	Actioned performed at state S'
$\boldsymbol{\mathcal{Q}}\,(\boldsymbol{\mathcal{S}}_o, a_o)$	Q-value of taking action a_o at state $\boldsymbol{\mathcal{S}}_o$
α	Learning rate
γ	Discount factor
ε	Exploration rate
$\mathcal{n}$	Number of training episodes

Reward: In our optimized drone's handover mechanism, we aim to decrease the number of handovers by maintaining reliable connectivity, as shown in Fig. 5c. The drone should also connect to the lower RSRP cell in the trajectory path, and frequent handover can be avoided. Our proposed model considers handover cost weight Σ_{HO} and the serving cell RSRP weight Σ_{RSRP} at a future state in the reward function given by

$$R = -\Sigma_{HO} + \mathbb{C}\,(HO) + \Sigma_{RSRP} \times RSRP_{s'}. \tag{2}$$

These weights in Eq. (2) with the indicator function $\mathbb{C}\,(HO)$ balance our two contradictory goals. The handover cost $\mathbb{C}\,(HO)$ will be "1" if the serving cells at state s and s' are different; otherwise, the cost will be "0".

4.2.2 Q-learning-based Algorithm for Handover Optimization

In our proposed model, at every single state, action space A is constrained to the strongest $\boldsymbol{\mathcal{V}}$ candidate cells *denoted* by a set $I_{\boldsymbol{\mathcal{V}}} = \{0, 1, \ldots, \boldsymbol{\mathcal{V}} - 1\}$. Q-table $Q \in \boldsymbol{\mathcal{R}}^{l \times \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{V}}}$ is updated according to Eq. (1) for a drone trajectory path. *The complexity of* Algorithm 1 is given by $O\,(cn)$, *where "n" denotes* the number of training episodes, *and "c" is a constant value equivalent* to the *total route length "l"*. Steps 2–9 produce the preliminary Q-table for the given drone's path, and a binary square matrix (*size* $\boldsymbol{\mathcal{V}}$) is generated in step 6.

Furthermore, if the p-th strongest cell at state s is different from the $q - th$ strongest cell in state s' in $(p, q) - th$ entity of matrix, then the entity is "1"; else, it is "0". Steps 11–24 execute and update each training episode's Q-value; for example, the greedy exploration runs in steps 14–18, and the Q-value is updated in the table at step 20. Finally, the Q-table contains values that can be chosen for different actions, and the highest value indicates the optimal choice. Consequently, for an efficient mechanism, the handover decisions can be attained from the maximum Q-value at each state along the given trajectory. The block diagram for the proposed model is shown in Fig. 6.

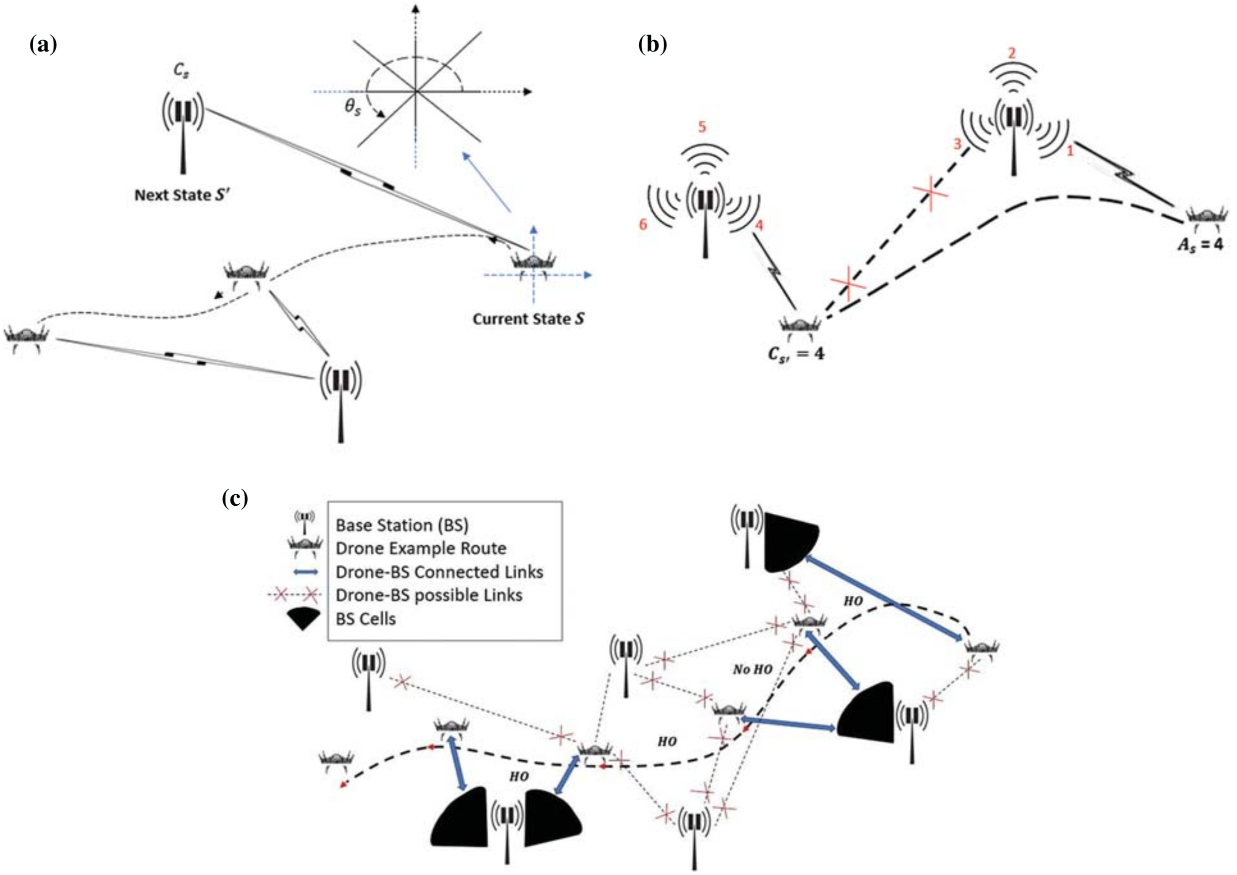

Figure 5: Illustration of the proposed Q-learning-based framework. (a) Illustration of current and next state; (b) illustration of action; (c) example of handover decisions during a trip

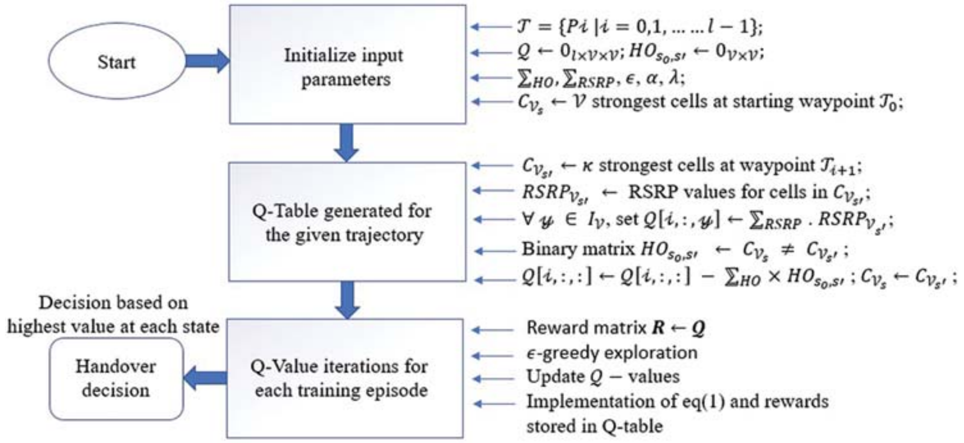
Start
Initialize input parameters
$\mathcal{T} = \{Pi \mid i = 0,1, \ldots \ldots l-1\}$;
$Q \leftarrow 0_{l\times\mathcal{V}\times\mathcal{V}}$; $HO_{s_0,s'} \leftarrow 0_{\mathcal{V}\times\mathcal{V}}$;
$\sum_{HO}, \sum_{RSRP}, \epsilon, \alpha, \lambda$;
$C_{\mathcal{V}_s} \leftarrow \mathcal{V}$ strongest cells at starting waypoint $\mathcal{T}_0$;
Q-Table generated for the given trajectory
$C_{\mathcal{V}_{s'}} \leftarrow \kappa$ strongest cells at waypoint $\mathcal{T}_{i+1}$;
$RSRP_{\mathcal{V}_{s'}} \leftarrow$ RSRP values for cells in $C_{\mathcal{V}_{s'}}$;
$\forall\, y \in I_{\mathcal{V}}$, set $Q[i,:,y] \leftarrow \sum_{RSRP} \cdot RSRP_{\mathcal{V}_{s'}}$;
Binary matrix $HO_{s_0,s'} \leftarrow C_{\mathcal{V}_s} \neq C_{\mathcal{V}_{s'}}$;
$Q[i,:,:] \leftarrow Q[i,:,:] - \sum_{HO} \times HO_{s_0,s'}$; $C_{\mathcal{V}_s} \leftarrow C_{\mathcal{V}_{s'}}$;
Decision based on highest value at each state
Handover decision
Q-Value iterations for each training episode
Reward matrix $\boldsymbol{R} \leftarrow \boldsymbol{Q}$
ϵ-greedy exploration
Update Q − values
Implementation of eq(1) and rewards stored in Q-table

Figure 6: Block diagram of handover decisions based on Q-learning

5 Simulation and Results

This section will evaluate the proposed Q-learning-based handover scheme with the 3GPP access-beam-based method (greedy handover algorithm), where drones are always connected to the strongest cell. We calculate the handover ratio as a performance metric for every drone trajectory. In the baseline scenario, the drone is always connected to the strongest cell, and the handover ratio will be constant at 1. The performance evaluation for diverse weight combinations of Σ_{HO} and Σ_{RSRP} in the reward function represents the tradeoff between upcoming RSRP values and the number of handovers. The handover ratio approaches zero and the number of handovers decreases when the ratio $\frac{\Sigma_{HO}}{\Sigma_{RSRP}}$ increases. Meanwhile, the proposed Q-learning-based handover and baseline algorithm will show similar results in a special scenario where there is no handover cost $\mathbb{C}(HO)$, that is, when no handover occurs.

Extensive simulations are conducted to evaluate the performance based on the number of episodes and accumulated reward gained in each episode. The proposed algorithm is evaluated by varying the parameters $(\alpha, \epsilon, \gamma)$ of Q-learning to find the optimal results. The results show that the proposed algorithm converges on the maximum reward for each randomly generated route. As shown in Tab. 2, parameters are set based on exploration and exploitation with a greedy algorithm: $\alpha = 0.1, 0.5, 0.9$, $\epsilon = 0.1, 0.5, 0.9$, and $\gamma = 0.1, 0.5, 0.9$. The variations in $\alpha, \epsilon, \text{and} \gamma$ show the optimized results for the drone's trajectory.

Table 2: Parameters in each scenario

Scenario 1			Scenario 2			Scenario 3		
γ	ε	α	α	ε	γ	γ	α	ε
0.9	0.5	0.1 0.5 0.9	0.3	0.5	0.1 0.5 0.9	0.9	0.3	0.1 0.5 0.9

In Fig. 7, we show the accumulated reward when the learning rate (α) varies in the range of $0.1 - 0.9$; meanwhile, the values for epsilon (ε) and discount factor (γ) are 0.5 and 0.9, respectively. After 7 initial episodes, the best-accumulated reward of the proposed algorithm is at learning rate $\alpha = 0.9$. For $\alpha = 0.5$, the accumulated reward stays higher than $\alpha = 0.1$ from episode 7 to 70 but degrades afterward up to episode 250. The best parameters are $\alpha = 0.9, \epsilon = 0.5, \text{and} \gamma = 0.9$. Since the proposed Q-learning-based algorithm reduces the ping-pong handovers, RSRP is also reduced.

In Fig. 8, we show the accumulated reward as the discount factor (γ) varies in the range of $0.1 - 0.9$; meanwhile, the values for learning rate (α) and epsilon (ε) are 0.3 and 0.5, respectively. After 70 initial episodes, the best-accumulated reward of the proposed algorithm is at $\gamma = 0.5$. For $\gamma = 0.9$, the accumulated reward stays higher than $\gamma = 0.1 \text{and} 0.5$, from episode 20 to 70 but degrades afterward up to episode 250. Meanwhile, $\gamma = 0.1$ yields the worst accumulated reward. As shown in the results, the best parameters are $\alpha = 0.3, \epsilon = 0.5, \text{and} \gamma = 0.5$. There will be no handover when the proposed scheme is equivalent to the baseline scheme.

In Fig. 9, we show the accumulated reward when epsilon (ε) varies in the range of $0.1 - 0.9$; meanwhile, the values for α *and* γ are 0.3 and 0.9, respectively. After 18 initial episodes, the best-accumulated reward of the proposed algorithm is at $\epsilon = 0.9$. For $\epsilon = 0.1$, the accumulated reward stays higher than $\epsilon = 0.1$, from episode 1 to 31 but degrades afterward up to episode 250.

As shown in the results, the best parameters are $\alpha = 0.3, \epsilon = 0.9, \text{and} \gamma = 0.9$. To avoid unnecessary handovers, we need to decrease the ratio $\frac{\Sigma_{HO}}{\Sigma_{RSRP}}$, and then the cost of handover also decreases.

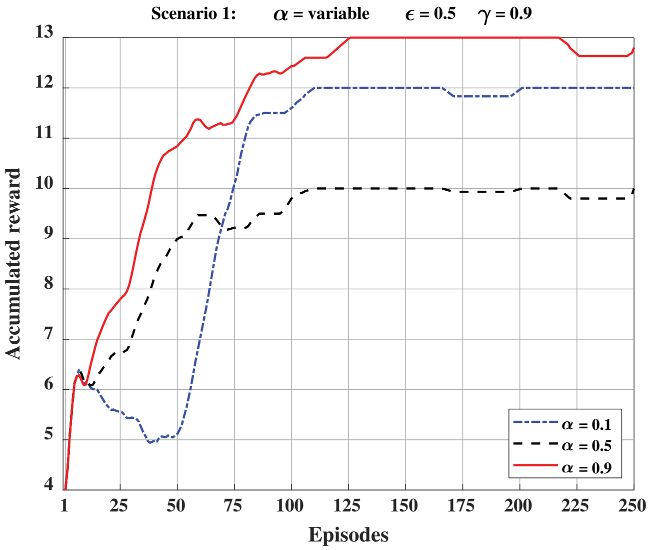

Figure 7: Accumulated reward at $\gamma = 0.9$, $\varepsilon = 0.5$, and $\alpha = 0.1$–0.9

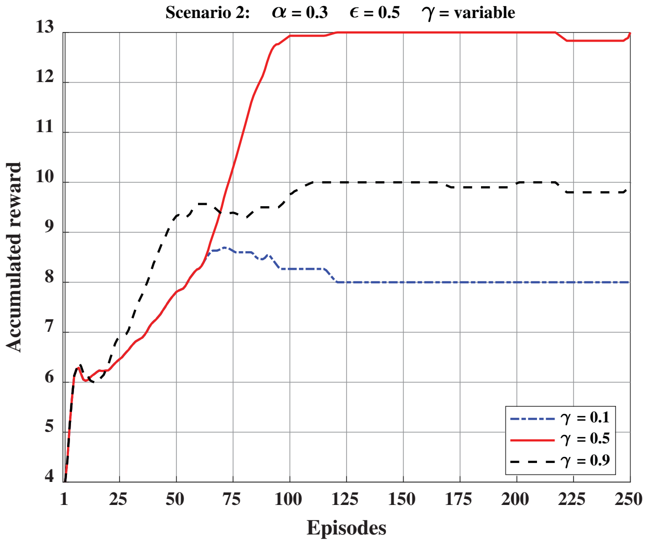

Figure 8: Accumulated reward at $\alpha = 0.3$, $\varepsilon = 0.5$, and $\gamma = 0.1$–0.9

In Fig. 10, we show the accumulated rewards of the proposed algorithm in each scenario ($\alpha = 0.9$, $\gamma = 0.5$, and $\epsilon = 0.9$). The simulation results show that the right parameters will affect drone performance throughout the learning phase and also enhance the learning curve. The proposed technique demonstrates that the learning process is the best way to optimize drone mobility in dense scenarios.

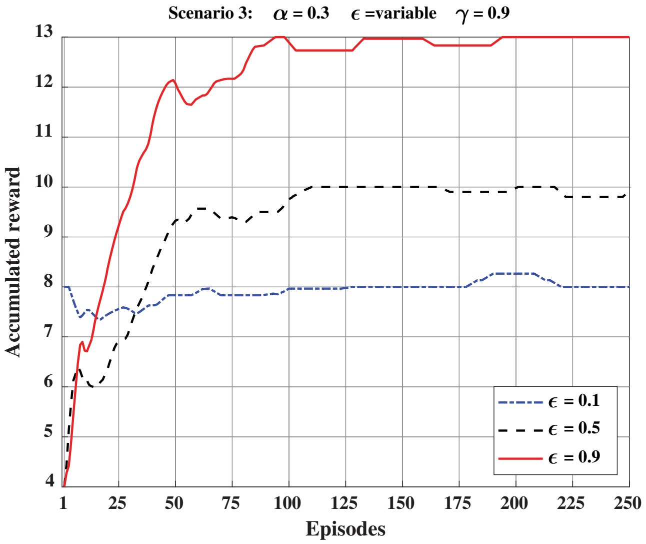

Figure 9: Accumulated reward at $\gamma = 0.9$, $\alpha = 0.3$, and $\epsilon = 0.1$–0.9

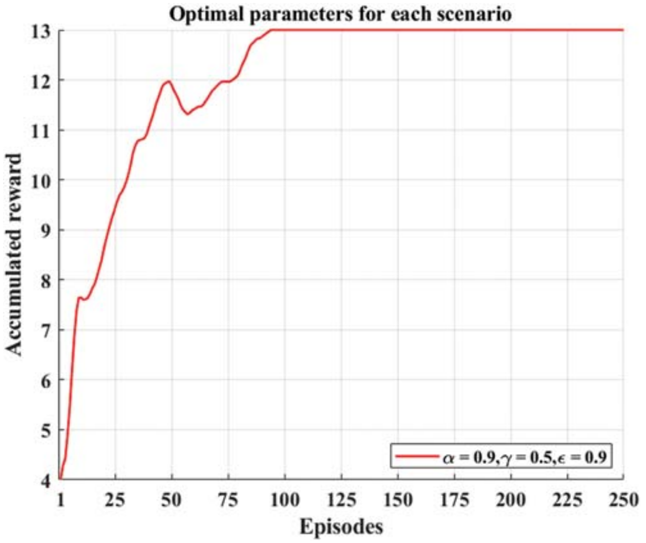

Figure 10: Accumulated reward at $\alpha = 0.9$, $\gamma = 0.5$, and $\epsilon = 0.9$

6 Conclusions and Future Work

In this work, we proposed a machine learning-based algorithm to accomplish strong drone connectivity with less handover cost such that the drone will not always connect to the strongest cell in a trajectory. We suggested a robust and flexible way to make handover decisions using a Q-learning framework. The proposed scheme reduces the total number of handovers, and we can observe a tradeoff between received signal strength and the number of handovers while always connecting the drone to the strongest cell. This tradeoff can be adjusted by changing the weights in the reward function. There are many potential directions for future works such as exploring which additional parameters may further enhance reliability during handover decision-making.

This work presents a notable contribution to determine the optimal route of drones for researchers who are exploring UAV use cases in cellular networks where a large testing site comprised of several cells with multiple UAVs is under consideration. Finally, the proposed framework studies 2D drone mobility; a 3D mobility model would introduce more parameters to aid efficient handover decision.

Acknowledgement: The authors would like to thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Funding Statement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2018R1D1A1B07049877) and the Strengthening R&D Capability Program of Sejong University.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Mozaffari, W. Saad, M. Bennis, Y. Nam and M. Debbah, "A tutorial on UAVs for wireless networks: applications, challenges, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[2] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano *et al.,* "Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3417–3442, 2019.

[3] A. Sharma, P. Vanjani, N. Paliwal, C. M. W. Basnayaka, D. N. K. Jayakody *et al.,* "Communication and networking technologies for UAVs: a survey," *Journal of Network and Computer Applications*, vol. 168, no. 2, pp. 1–18, 2020.

[4] G. Yang, X. Lin, Y. Li, H. Cui, M. Xu*et al.,* "A telecom perspective on the internet of drones: from LTE-advanced to 5G," *Networking and Internet Architecture*, 2018. https://arxiv.org/ abs/1803.110484.

[5] A. Kamran, H. X. Nguyen, Q. T. Vien, P. Shah and M. Raza, "Deployment of drone-based small cells for public safety communication system," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2882–2891, 2020.

[6] X. Lin, V. Yajnanarayana, S. D. Muruganathan, S. Gao, H. Asplund *et al.,* "The sky is not the limit: LTE for unmanned aerial vehicles," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 204–210, 2018.

[7] A. Anpalagan, M. Bennis and R. Vannithamby, "Dense networks of small cells," In: *Design and Deployment of Small Cell Networks*, vol. 5, pp. 1–30, 2015. [Online]. Available: https://www.cambridge.org/core/books/design-and-deployment-of-small-cell-networks/5227E03ADA00439B.

[8] I. Shayea, M. Ergen, M. H. Azmi, S. A. Çolak, R. Nordin *et al.,* "Key challenges, drivers and solutions for mobility management in 5G networks: A survey," *IEEE Access*, vol. 8, pp. 172534–172552, 2020.

[9] A. A. Zaidi, R. Baldemair, H. Tullberg, H. Bjorkegren, L. Sundstrom *et al.,* "Waveform and numerology to support 5G services and requirements," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 90–98, 2016.

[10] K. W. Li, C. Sun and N. Li, "Distance and visual angle of line-of-sight of a small drone," *Applied Sciences*, vol. 10, no. 16, pp. 1–12, 2020.

[11] P. S. Bithas, E. T. Michailidis, N. Nomikos, D. Vouyioukas and A. G. Kanatas, "A survey on machine-learning techniques for UAV-based communications," *Sensors*, vol. 19, no. 23, pp. 1–39, 2019.

[12] E. Gures, I. Shayea, A. Alhammadi, M. Ergen and H. Mohamad, "A comprehensive survey on mobility management in 5G heterogeneous networks: Architectures, challenges and solutions," *IEEE Access*, vol. 8, pp. 195883– 195913, 2020.

[13] J. A. Besada, L. Bergesio, I. Campaña, D. Vaquero-Melchor, J. López-Araquistain *et al.,* "Drone mission definition and implementation for automated infrastructure inspection using airborne sensors," *Sensors*, vol. 18, no. 4, pp. 1–29, 2018.

[14] E. Klavins and Z. Valery, "Unmanned aerial vehicle movement trajectory detection in open environment," in *Proc. ICTE*, Riga, Latvia, pp. 400–407, 2017.

[15] Z. Ullah, F. A. Turjman and L. Mostarda, "Cognition in UAV-aided 5G and beyond communications: A survey," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 872–891, 2020.

[16] X. Lin, R. Wiren, S. Euler, A. Sadam, H. L. Maattanen *et al.,* "Mobile network-connected drones: field trials, simulations, and design insights," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 115–125, 2019.

[17] 3GPP TR 36.777, "Enhanced LTE support for aerial vehicles (Release 15)," in *3GPP TSG RAN Meeting Proc.*, 2018, [Online]. Available: https://portal.3gpp.org/ngppapp/TdocList.aspx?meetingId=31948.

[18] 5G!Drones, "5G for Drone-based Vertical Applications," 2019. [Online]. Available: https://5gdrones.eu/wp-content/uploads/2020/05/.

[19] J. Stanczak, I. Z. Kovacs, D. Koziol, J. Wigard, R. Amorim *et al.,* "Mobility challenges for unmanned aerial vehicles connected to cellular LTE networks," in *Proc. VTC*, Spring, Porto, Portugal, pp. 1–5, 2018.

[20] S. D. Muruganathan, X. Lin, H. L. Maattanen, J. Sedin, Z. Zou *et al.,* "An overview of 3GPP release-15 study on enhanced LTE support for connected drones," *Networking and Internet Architecture*, 2018, https://arxiv.org/abs/1805.00826.

[21] Y. Zeng, J. Lyu and R. Zhang, "Cellular-connected UAV: Potential, challenges, and promising technologies," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 120–127, 2018.

[22] A. Orsino, A. Ometov, G. Fodor, D. Moltchanov, L. Militano *et al.,* "Effects of heterogeneous mobility on D2D-and drone-assisted mission-critical MTC in 5G," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 79–87, 2017.

[23] W. P. Nwadiugwu, W. Esther, L. Jae-Min and K. D. Seong, "Communication handover for multi-dimensional UAVs in ROI using MIMO-ultrawideband," in *Proc. ICAIIC*, Fukuoka, Japan, pp. 47–50, 2020.

[24] A. Rabe, E. Hesham, L. Lutz and M. J. Hossain, "Handover rate characterization in 3D ultra-dense heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10340–10345, 2019.

[25] V. Yajnanarayana, Y. P. E. Wang, S. Gao, S. Muruganathan and X. Lin, "Interference mitigation methods for unmanned aerial vehicles served by cellular networks," in *Proc. 5GWF*, Silicon Valley, CA, USA, pp. 118–122, 2018.

[26] M. M. Azari, F. Rosas and S. Pollin, "Cellular connectivity for UAVs: Network modeling, performance analysis, and design guidelines," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3366–3381, 2019.

[27] S. Changyang, L. Chenxi, Q. S. Q. Tony, Y. Chenyang and L. Yonghui, "Ultra-reliable and low-latency communications in unmanned aerial vehicle communication systems," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3768–3781, 2019.

[28] A. Fakhreddine, C. Bettstetter, S. Hayat, R. Muzaffar and D. Emini, "Handover challenges for cellular-connected drones," in *Proc. DroNet'19*, Seoul, Republic of Korea, pp. 9–14, 2019.

[29] E. Lee, C. Choi and P. Kim, "Intelligent handover scheme for drone using fuzzy inference systems," *IEEE Access*, vol. 5, pp. 13712–13719, 2017.

[30] K. N. Park, J. H. Kang, B. M. Cho, K. J. Park and H. Kim, "Handover management of net-drones for future internet platforms," *International Journal of Distributed Sensor Networks*, vol. 12, no. 3, pp. 1–9, 2016.

[31] W. Dong, M. Xinhong, H. Ronghui, L. Xixiang and H. Li, "An enhanced handover scheme for cellular-connected UAVs," in *Proc. ICCC*, China, pp. 418–423, 2020.

[32] M. Salehi and E. Hossain, "Handover rate and sojourn time analysis in mobile drone-assisted cellular networks," *Networking and Internet Architecture*, 2020. https://arxiv.org/abs/2006.05019.

[33] V. Yajnanarayana, H. Rydén and L. Hévizi, "5G handover using reinforcement learning," in *Proc. 5GWF*, 2020. https://arxiv.org/pdf/1904.02572,

[34] R. S. Sutton and A. G. Barto, "Finite markov decision processes," In *Introduction to reinforcement learning: An introduction*, Second ed., vol. 3, Massachusetts London, England: The MIT Press, pp. 53–79, 2018. [Online]. Available: https://mitpress.mit.edu/books/reinforcement-learning.