IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 November 2023, accepted 4 February 2024, date of publication 22 February 2024, date of current version 4 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3368916

RESEARCH ARTICLE

A Hybrid Scheme Using TOPSIS and Q-Learning for Handover Decision Making in UAV Assisted Heterogeneous Network

JIHAI ZHONG[1], LI ZHANG[1], (Senior Member, IEEE), MOHANAD ALHABO[2], JONATHAN SERUGUNDA[3], AND SHEILA N. MUGALA[3]

[1]School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT Leeds, U.K.
[2]Informatics and Telecommunication Public Company (ITPC), Iraqi Ministry of Communications, Mosul 10013, Iraq
[3]Electrical and Computer Engineering Department, Makerere University, Kampala, Uganda

Corresponding author: Li Zhang (l.x.zhang@leeds.ac.uk)

ABSTRACT An increasing number of users are expected to be served by wireless network with heterogeneous requirements. Unmanned aerial vehicles (UAV) can be deployed to augment the heterogeneous network from aerial area by taking advantages of the characteristics of UAV such as mobility, manoeuvrability, low cost, and Line-of-Sight (LoS) communication. However, the deployment of UAV can also cause problems. For example, in UAV-assisted HetNet, the number of handover (HO) will increase because of the dense distribution of small base stations (SBS) and UAV base stations (UBS). Also, because of the small coverage of SBS and the LoS communication links from neighbour UBSs, the number of unnecessary HO will also rise. Frequent HO and unnecessary HO can result in interruption, increased overhead and energy consumption, which is not desirable for battery powered UAVs. In this paper, to solve the problem, a HO decision making algorithm adopting TOPSIS and Q-learning (QL) is proposed with the aim to reduce HO number and improve energy efficiency. Q-learning can be applied to address decision-making challenges in communication systems widely. However, a large volume of training data can pose challenges and complexities, therefore, the TOPSIS is utilised to reduce the size of the action space in Q-learning. The proposed hybrid TOPSIS-Q-learning method enhances both the handover performance and the scalability. In the method, signal to interference and noise ratio (SINR), time of stay (ToS) and average energy efficiency (EE) are taken into account. The simulation results show that the number of HO and unnecessary HO is remarkably reduced and the average EE is notably improved in comparison with other existing methods.

INDEX TERMS UAV, heterogeneous network, handover, TOPSIS, Q-learning.

I. INTRODUCTION

Recently, heterogeneous network (HetNet) has been deployed widely to provide access service for user equipment (UE) to combine different access technologies and transmission solutions [1]. However, the increasing demand for bandwidth, capacity, and coverage, along with challenges such as interference and spectrum management complexities, necessitate incorporating unmanned aerial vehicles (UAVs) in HetNet due to their flexibility, cost-efficiency, and ability to facilitate quick deployment [2], [3], [4]. What is more, in HetNet, UAVs can serve as aerial base stations (BSs) or relay nodes, leveraging the air-to-ground (A2G) channel for excellent communication with ground UE due to high line-of-sight (LoS) probability [5]. With low propagation delay, high-quality communication links, and efficient maintenance, UAV-assisted HetNet enhances wireless network coverage, capacity, reliability, and energy efficiency to meet future communication demands [1].

Although UAV-assisted Hetnets provide such significant benefits, there are still open challenges due to the need for handovers (HOs) between different types of BSs [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Xujie Li.

TABLE 1. Description of the abbreviations used in the paper.

Symbol	Description
λ_s, λ_u	Distribution density of SBS and UBS
L	The set of waypoints
v_{UE}, v_{UBS}	The velocity of UBS and UE
θ^l_{UE}, θ^l_{UBS}	The moving angle of the UE and UBS at waypoint l
P_{LoS}, P_{NLoS}	The probability of LoS and NLoS
a, b	The environment dependent constants
PL_{LoS},PL_{NLoS} and PL_{G2G}	The path loss for LoS, NLoS and G2G links
PL_{ave}	The average received path loss od A2G link
f_c	The carrier frequency
h_U	The height of the UAV
d_{3d}	The distance between the UE and a BS in 3-dimension environment
d_{2d}	The distance between the UE and a BS in 2-dimension environment
P_t	The transmit power from the serving BS
P_c	The receiver's circuit power
G_t, G_r	The antenna gain of transmitter and receiver
σ_{SF}	The standard deviation of shadowing deviation
I_s, I_u	The interference in mW from SBSs and UBSs
$RSRP_i$	The reference signal received power from BS i
σ_n	The noise power
B	The bandwidth
T_{th}	Time threshold
R_{BS}	The radius of a BS
ω	The angle between the path of UE and the connection between UE and a BS
η_{SE}	The spectral efficiency of a UE
$a^{norm}_{i,p}$	The normalized ith values of pth attributes
σ_p	The standard deviation of pth attribute
μ_p	The mean value of pth attribute
w^{sd}_p	The weight of attribute p from SD weighting technology
R	The reward from QL
$\mathrm{SINR}^{\mathrm{norm}}$, $\mathrm{ToS}^{\mathrm{norm}}$ and $\mathrm{EE}^{\mathrm{norm}}$	The normalized value of SINR, ToS and EE
$I\,(HO)$	The HO indicator
α	The learning rate
γ	The discount rate
$Q(s, a)$	The Q value of the certain state s and a action a

HO is the process of transferring an ongoing call/data session from a serving BS to a target BS without degrading the quality of service for the connection. However, frequent HOs will lead to various consequences, such as increased signaling overhead, increased call drops, packet loss rates as well as increased latency. In ground HetNet involving only macro BSs (MBSs) and small BSs (SBSs), frequent HO and unnecessary HO are usually caused by dense distribution of SBSs [7]. Additionally, in UAV assisted HetNet, HO is also caused by the movement of UAV BSs (UBSs). UEs easily connect to UAVs with poor positions, resulting in increased HO occurrences [8].

Several existing studies have addressed HO challenges in UAV wireless networks. In [9], the author proposed a UAV-aided network model that considers the advantages and disadvantages of HO procedures. The study concluded that UAVs should employ directive antenna systems, fly at low altitudes, and maintain slower speeds. In [10], a path-loss-plus-fading model incorporating HO parameters was considered for HO in UAV networks. This study specifically addressed frequent HO problems caused by the transition between LoS and Non-Line-of-Sight (NLoS) in UAV networks. A discrete-time Markov chain was employed, and HO failure and ping-pong probabilities were derived by analyzing HO state probabilities during time-to-trigger (TTT). The results revealed a trade-off between HO failure and ping-pong probabilities through TTT adjustments. Recently, novel reinforcement learning methods have been applied in wireless communications research, such as in [11], [12], and [13]. They have also been employed to address handover decision making problems. Authors of [14] have proposed an RL-based HO mechanism to achieve UAV connectivity in a cellular-connected UAV network. HO decisions are dynamically optimized be applying the proposed algorithm. In [15], a reinforcement learning-based HO algorithm and mobility control optimization method were proposed for UAV networks. Authors in [16] consider a network consisting of only UAV base stations, and proposed another intelligent HO method. It introduced a deep learning-based trajectory prediction model and made HO decisions based on predicted channel characteristics. The results showed a higher HO success rate compared to traditional HO methods. In [17], a Q-Learning-based HO management solution was proposed between UBSs and static base stations, improving UE capacity. Additionally, TOPSIS is used in many researches. Reference [18] proposed a TOPSIS method to the problem of network selection with a proposed iterative approach to improve the result. In [19], the authors presented a to rank BSs based on the weighted attribute for HO optimization. In this work, authors compared two techniques which are entropy weighting technique and standard deviation weighting technique to scale the importance of the attributes. While the need for UAV-assisted network systems is growing, there is limited research focusing on optimizing HO performance in UAV-assisted HetNets.

In this paper, a hybrid HO decision making algorithm based on Technique for Order Preference by Similarity to Idea Solution (TOPSIS) and Q-learning (QL) is proposed. Q-learning can be applied in various decision-making challenges in communication systems. However, in ultra-dense network with a significant number of BSs, handling a large volume of training data can pose challenges and complexities. The presence of noise or errors in the training data can negatively impact the performance. Therefore, we employ the TOPSIS method to reduce the dimensions of the action space of Q-learning. This hybrid design has the potential to enhance both the performance and the scalability, which is crucial in efficiently handling an increasing number of BSs. In the proposed method, three decision criteria are used to select the best BS at a specific position, which are signal to interference and noise ratio (SINR), time of stay (ToS) and average energy efficiency (EE). Furthermore, in the ToS measurement, we take into account the movements of both the UEs and the UAVs. Finally, the results show that the proposed scheme can reduce the number of HO and unnecessary HO with higher EE comparing with the existing methods.

The rest of paper is organized as follows. Section II presents the related work, in Section III, the system model is introduced. The proposed HO scheme is illustrated in Section IV. Section V shows the simulation results and the analysis. Finally, the conclusion is presented in Section V.

II. SYSTEM MODEL

A. UAV ASSISTED HETNET MODEL

A three-tier UAV assisted HetNet model is considered in this work, as shown in Fig.1. In this model, a round coverage area of a MBS is considered. The MBS is located at the centre of the area, and some SBSs and UBSs are distributed in the area following Poisson Point Process (PPP) with a specific density λ_s and λ_u. In this system, SBSs operate at 5G millimetre wave frequency and UBSs work at a frequency of LTE. Also, when UAVs are operating, they hover and fly in the air following their own path which is generated randomly. At the same time, a UE moves in the area following a randomly generated trajectory. During the movement of UBSs and the UE, HOs are performed to keep the UE connect to a suitable BS. In this scenario, MBS will not be the target BS in any UE HO decisions, and UE will only connect to SBSs and UBSs to offload the traffic from the congested MBS.

In the model, UAVs and UE are moving at the same time following the trajectories with L waypoints. Every waypoint is generated after UAVs or UE move for T time from the last waypoint. Each UAV is moving with an unchanged velocity v_{UBS}, but the direction will be changed at every waypoint by

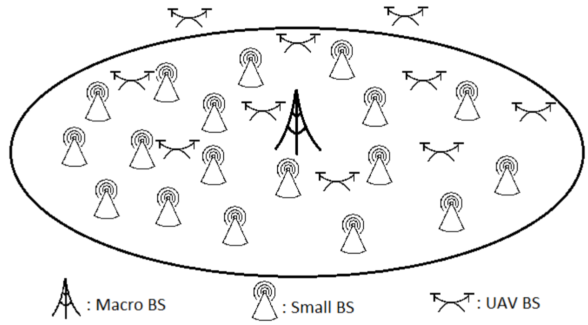


FIGURE 1. UAV assisted three-tier HetNet.

randomly generating a moving angle $\theta^l_{UBS_i}$, where i indicates the ith UAV and l ($l \in L$) is the lth waypoint.

The moving model for ground UE is slightly different. The initial point of UE is randomly generated in the area, and UE is moving with a fixed v_{UE} all the time. At every waypoint l, a new value of direction θ^l_{UE} is randomly selected in $\theta^l_{UE} \in [\theta^{l-1}_{UE} - \frac{\pi}{6}, \theta^{l-1}_{UE} + \frac{\pi}{6}]$, where θ^{l-1}_{UE} is the direction of the UE at the at last waypoint.

B. PROPAGATION MODEL

In this work, there are two different types of channels which are ground-to-ground (G2G) and air-to-ground (A2G), corresponding to ground and air BSs respectively.

The A2G channel consists of both LoS and NLoS links. The A2G channels are easily affected by the obstacles such as buildings between the UBSs and UEs. In this study, whether the UE has an unobstructed LoS channel to a given BS is determined using a ray tracing model [20], where the buildings are generated randomly following Poisson distribution with density β, and the random height of each building follows Rayleigh distribution with scale parameter κ. Therefore, when the ground UE connects a UBS, if there is not a building blocking the path line between the UBS and the UE, the channel is a LoS channel. Otherwise, it is an NLoS channel.

The path loss for LoS and NLoS links is calculated as follows [21]:

$$PL_{LoS} = 30.9 + 20\log_{10}(f_c) + (22.25 - 0.5\log_{10}(h_U))\log_{10}(d_{3d}), \quad (1)$$

$$PL_{NLoS} = 32.4 + 20\log_{10}(f_c) + (43.2 - 7.6\log_{10}(h_U))\log_{10}(d_{3d}), \quad (2)$$

where f_c is the carrier frequency, h_U is the height of the UAV and d_{3d} is the distance between the ground UE and the UAV in 3D environment.

In contrast, the LoS probability for G2G channels is low, therefore, it is assumed that G2G channels are NLoS. From [22], the path loss model for G2G channels can be given as:

$$PL_{G2G} = 32.4 + 20\log_{10}(f_c) + 30\log_{10}(d_{3d}), \quad (3)$$

In order to improve the quality of communication link, SINR, ToS and EE are measured for HO decision making, as explained in the next section.

C. HO DECISION CRITERIA

1) DOWNLINK SINR FOR GROUND UE

SINR is related to the reference signal received power (RSRP), interference and noise.The RSRP of both SBSs and UBSs in dBm can be expressed as

$$RSRP = P_t + G_t + G_r - PL + SF, \tag{4}$$

where the P_t is the transmit power from the serving BS, G_t and G_r are the antenna gain of transmitter and receiver respectively, PL represents the path loss from the serving BS to ground UE, and SF is the shadowing fading which follows Gaussian distribution with mean 0 and standard deviation σ_{SF}

$$SF \sim \mathcal{N}(0, \sigma_{SF}). \tag{5}$$

The values of σ_{SF} for LoS A2G, NLoS A2G and G2G channels are given below [24]:

$$\sigma_{SF}^{LoS} = \max(5 \times e^{(-0.01 \times h_U)}, 2), \tag{6}$$

$$\sigma_{SF}^{NLoS} = 8, \tag{7}$$

$$\sigma_{SF}^{G2G} = 7.8. \tag{8}$$

Downlink interference is caused by other BSs. It is assumed that there are K_s SBSs and K_u UBSs deployed in the area. Then, when the UE is served by BS i, which could be SBSs or UBSs, the interference in mW can be expressed as:

$$I_s = \sum_{j=1, j\neq i}^{K_s} RSRP_j, \tag{9}$$

$$I_u = \sum_{j=1, j\neq i}^{K_u} RSRP_j, \tag{10}$$

where j represents a BS. Interference is the sum of $RSRP$ from all BSs with same frequency except the serving BS i. Therefore, SINR is given by:

$$SINR = \frac{RSRP_i}{I + {\sigma_n}^2}, \tag{11}$$

where I represents the sum of interference I_s and I_u, and σ_n is noise power, which is calculated in Watt as:

$$\sigma_n = 10^{-3} \times 10^{(-174 + 10 \times \log_{10} B)/10}, \tag{12}$$

where B is the bandwidth.

2) TIME OF STAY

ToS is the sojourn time of the UE in a BS. Unnecessary HO is related to ToS. If the ToS of the target BS is lower than a time threshold T_{th} after a HO completes, then the HO is an unnecessary HO. As Fig.2 shows, the position could be at A, B and C when a UE is passing a BS. The corresponding ToS is expressed respectively as

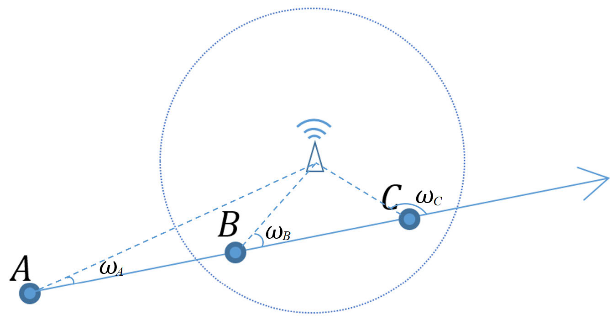


FIGURE 2. UE is moving through a BS coverage.

$$ToS_A = \frac{2 \times R_{BS} \times \sqrt{1 - (\frac{d_{2d} \times \sin(\omega_A)}{R_{BS}})}}{v_{UE}}, \tag{13}$$

$$ToS_B = \frac{\sqrt{{R_{BS}}^2 - (d_{2d} \times \sin(\omega_B)^2)} + d_{2d} \times \cos(\omega_B))}{v_{UE}}, \tag{14}$$

$$ToS_C = \frac{\sqrt{{R_{BS}}^2 - (d_{2d} \times \sin(180 - \omega_C)^2)}}{v_{UE}} - \frac{d_{2d} \times \cos(180 - \omega_C))}{v_{UE}}, \tag{15}$$

where R_{BS} is the radius of the BS, d_{2d} is the distance between the UE and the BS in 2-dimension environment, and the ω_A, ω_B, ω_C are the angle between the path of UE and the connection between UE and the BS in degree.

Additionally, when the target BS is a UBS, the calculation of ToS must also consider the movement of the UBS, including both the velocity and the direction. This is achieved by measuring ToS using relative velocity, which is calculated using vector representation as illustrated in Fig.3, where A and B are the objective movement of UE and UAV. Assuming UAV is static, in relative to the UAV, the UE's path is modified from A to C, and ToS will be measured using the relative velocity.

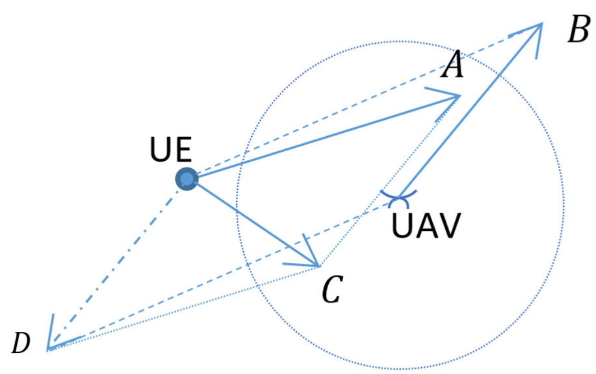


FIGURE 3. Relative velocity of UE.

3) ENERGY EFFICIENCY

Considering the limted life time of the UAV battery, it is critical to improve the EE of UAVs at every HO decision making. According to [23], EE is defined as the ratio of the spectral efficiency to power consumption.

$$\eta_{EE} = \frac{\eta_{SE}}{P_t + P_c}, \tag{16}$$

where P_t and P_c is the transmit power and receiver's circuit power respectively. Also η_{SE} is the spectral efficiency of the UE, which can be calculated as

$$\eta_{SE} = \log_2(1 + SINR), \tag{17}$$

where the unit of spectral efficiency is bits/s/Hz.

III. PROPOSED TOPSIS-QL HANDOVER SCHEME

The proposed TOPSIS-QL handover scheme combines a multiple criteria decision making (MCDM) method TOPSIS and a reinforcement learning method in the HO decision making for UAV assisted HetNet. The HO decision aims to select a proper target BS at every waypoint to optimise the performance when UE is moving in the network. In TOPSIS, we adopt three criteria to rank the BSs and construct a candidate list. The actions of QL are the BSs in the candidate list, and from which the QL selects the target BS. Algorithm 1 depicts the process of the TOPSIS-QL HO scheme.

A. TOPSIS

A MCDM technique is used to rank the alternatives based on multiple criteria. Generally, the best and worst values of criterion from different alternatives are extracted to construct a set of positive ideal solution and a set of negative ideal solution. The main idea of TOPSIS is, by calculating the Euclidean distance of each alternative to the solutions, finding the alternative that is closest to the positive ideal solution, and furthest from the negative ideal solution with weighted criterion [24].

In order to obtain the appropriate weight for each criterion, a method called standard deviation weighting technique is applied. The method can provide each criterion a weight according to its standard deviation [25]. The criterion that is similar for all alternatives gets a smaller weight, on the other hand, the weight is higher for the criterion with larger variance. The weights represent the relative importance of each criterion in the decision-making process. Assume there are P attributes, and there are m values in every attributes p $(p \in P)$, the standard deviation of any criterion can be calculate as:

$$\mu_p = \frac{1}{m}\sum_{i=1}^{m} a_{i,p}^{norm}, \tag{18}$$

$$\sigma_p = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (a_{i,p}^{norm} - \mu_p)^2}, \tag{19}$$

where $a_{i,p}^{norm}$ represents the normalized values of different attributes, i represents a certain value in an attribute.

Algorithm 1 TOPSIS-QL Handover Scheme

Initialize input parameters:
  Waypoints L, number of both SBSs and UBSs K
  Q table($L \times 10 \times 10$)
for $l (l \in L)$ **do**
  for BS$i = i : K$ **do**
    get $SINR_{l,i}$, $ToS_{l,i}$, $EE_{l,i}$
  end for
  Cell List $CL(l, :) \leftarrow$ Top 10 ranked by TOPSIS
end for
QL initialize:
for $l (l \in L)$ **do**
  for $x \in CL(l-1, :)$ **do**
    for $y \in CL(l, :)$ **do**
      get $Q(l, x, y)$ according to equation (24)
    end for
  end for
end for
while training ≤ 2000 **do**
  $j = 0$
  for l in L **do**
    if random value $i (i \in [0, 1]) \leq \epsilon$ **then**
      picked BS $j_{new} \leftarrow$ argmax $Q(l, j, u) (u \in 10)$
    else
      $j_{new} \leftarrow$ randomly picked from $A(l, 1 : k)$
    end if
    Update $Q(l, j, j_{new})$ with equation (25)
  end for
  $j = j_{new}$;
end while
Return Q table

In addition, σ_p and μ_p are the standard deviation and the mean value of pth attribute. Therefore, the weight of attribute p is calculated by

$$w_p^{sd} = \frac{\sigma_p}{\sum_{j=1}^{P} \sigma_j}, \tag{20}$$

where j represents different attributes.

Therefore, the higher standard deviation of the criterion, the more important the criterion is, and its weight is higher.

With the weighted criterion, TOPSIS can rank all the BS candidates for the waypoint in terms of the Euclidean distance to the positive and negative ideal solutions. However, as TOPSIS can only make decision for the current waypoint, the effect of the next waypoint cannot be taken into account. To further improve the HO performance, we proposed a hybrid scheme using TOPSIS to select the top ten BSs to form a reduced candidate list for the QL method, which is applied to make HO decisions at each waypoint.

B. Q-LEARNING

QL is a model-free reinforcement learning method. Usually, Markov decision process (MDP) is employed. It can be described as $\{S, A, P, R\}$, where S is the state, A is the possible actions, P is the transition probability for states, and R is the reward from the environment. The process of QL is the agent takes action to change the state based on the reward and interaction with the environment [14]. At the training stage, agent optimises every action decision in iterations achieved through exploitation and exploration. For exploitation, agent takes the actions with the highest reward at the present, while for exploration, agent takes actions randomly to discover the actions that have low reward now but benefit for the long-term reward. Therefore, QL method helps to get the best returns for a sequence of decisions.

1) STATE (S)

In this work, the ground UE is the agent, and the states include the coordinates of UE and the current serving BS, since the position of UE can affect the channel condition and the current connected BS can affect the selection of next BS.

2) ACTION (A)

The actions are the potential BSs. At each waypoint, UE decides which BS to connect to achieve good performance. Different actions can lead to different state of agent. It is meaningless and complicated that all the BSs in the area are selected as actions, therefore the candidate list generated by the TOPSIS is used as the action space for QL so that the agent only chooses a BS from the candidate list. According to [14] and the density of BSs, the size of action space is fixed at 10.

3) REWARD (R)

In the algorithm, a reward equation is given below involving SINR, ToS and EE:

$$R = \mathrm{SINR}^{\mathrm{norm}} \cdot w_{sinr} + \mathrm{ToS}^{\mathrm{norm}} \cdot w_{tos} + \mathrm{EE}^{\mathrm{norm}} \cdot w_{ee} - I(HO), \tag{21}$$

where $\mathrm{SINR}^{\mathrm{norm}}$, $\mathrm{ToS}^{\mathrm{norm}}$ and $\mathrm{EE}^{\mathrm{norm}}$ are the normalized value of SINR, ToS and EE, and $I\,(HO)$ is the HO indicator which will be 1 if HO happens, and 0 otherwise. The weights w_{sinr}, w_{tos} and w_{ee} are also generated using the above SD weighting technique.

In the beginning of QL, a Q-table is created containing the states and the corresponding actions of every state. All the BSs are ranked by TOPSIS, and the action space will consist of the top 10 BSs. As the size of action space of Q-learning is preliminarily reduced by TOPSIS, the proposed hybrid method can achieve more efficient learning and enhanced scalability, which is crucial in efficiently handling an increasing number of BSs. Since there are L waypoints and the length of candidate list is 10, the size of Q-table is $L \times 10 \times 10$. At the training stage, the agent makes decisions for exploitation and exploration according to ϵ-greedy policy at every waypoint iteratively. Following the policy, the agent decides to exploit with the probability of ϵ, and there is $1-\epsilon$ chance to explore [26]. After every decision, a new reward of the action at a specific waypoint will be calculated to replace the previous value by equation (21). Then, Bellman equation is applied to renew the previous Q value after getting the new value [26]:

$$Q\left(s,a\right) \leftarrow Q\left(s,a\right) + \alpha \times [R + \gamma \times \max_{a'} Q\left(s',a'\right) - Q\left(s,a\right)], \tag{22}$$

where α is the learning rate, γ is the discount rate, and $Q(s, a)$ is the Q value of the certain state and action stored in the current Q-table. The above procedure is repeated until a trained Q-table is generated. Following the Q-table, a sequence of HO decisions can be made by choosing the best Q value.

IV. PERFORMANCE AND RESULTS ANALYSIS

The performance of the hybrid HO scheme is evaluated in terms of the number of HO, number of unnecessary HO and energy efficiency. The performance of the proposed method is compared with that of the conventional, TOPSIS-only and QL-only methods. For the conventional method, HO will take place when the RSRP of the target BS is higher than the serving BS. For TOPSIS-only method, we used the same criteria as explained before, and the top ranked BS is the target BS. For the QL-only method, the action space is also formed by all BSs. In simulation, the learning rate α and discount rate γ in training stage are fixed at 0.5 and 0.3 respectively according to [14]. The other simulation parameters are given in the table 1.

TABLE 2. Simulation parameters.

Parameters	Values
Bandwidth (B)	$B_{SBS} = 500\ MHz$ $B_{UBS} = 20\ MHz$
Transmit Power (P_t)	$P_t^{SBS} = 35\ dBm$ $P_t^{UBS} = 20\ dBm$
Carrier Frequency (CF)	$CF_{SBS} = 28\ GHz$ $CF_{UBS} = 2\ GHz$
Density of SBSs (λ_s)	$60\ (10^{-6}/m^2)$
Density of Buildings (β)	$100\ (10^{-6}/m^2)$
Radius of BSs (R)	$R_{MBS} = 1000\ m$ $R_{SBS} = 50\ m$ $R_{UBS} = 150\ m$
Number of Way-points (L)	1000
Height of BSs(h)	$h_{SBS} = 4\ m$ $h_{UBS} = 100\ m$
Time Threshold(T_{th})	$0.72\ s$
Receiver Circuit Power (P_c)	$0.001\ W$
Velocity of UAVs (v_{UAV})	$10\ m/s$
Scale Parameter (κ)	20
Learning Rate (α)	0.5
Discount Rate (γ)	0.3

A. NUMBER OF HO

The number of HOs for the four methods versus the UE velocity is shown in Fig.4. The density of SBSs and density of UAV are fixed at $60 \times 10^{-6}/m^2$ and $40 \times 10^{-6}/m^2$

respectively, and the velocity of UE increases from $5m/s$ to $55m/s$. It is obvious that the numbers of HOs increases with the increase in the density of UAV for all methods. This is because with the increase of velocity, it is more likely for the UE to cross cell boundaries at the next waypoint, resulting in a higher probability of HO. Thus, the number of HO increases with increasing UE velocity. Fig.4 clearly shows that the proposed TOPSIS-QL method, labelled as the TOPSIS-QL, requires the lowest number of HOs. The conventional method has the highest number of HO as the UE only connect to the BSs with the best RSRP. However, in the TOPSIS-only method, the number of HOs increases slightly slower than that of the conventional method as it avoids some unnecessary HOs. Although QL-only method considers multiple criteria and future decisions, it performs worse than the proposed method because of the action space in QL-only method is formed by all BSs, while in the proposed method we consider multiple criteria to form the action space, especially ToS to avoid unnecessary HOs.

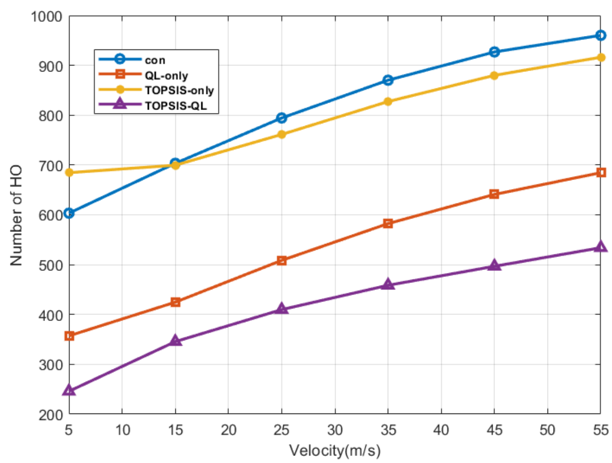


FIGURE 4. Number of HO against UE velocity.

B. NUMBER OF UNNECESSARY HO

Fig.5 shows the performance of the number of unnecessary HO against distribution density of UAVs from $10 \times 10^{-6}/m^2$ to $100 \times 10^{-6}/m^2$. In addition, the velocity of UE is fixed at 25 m/s. With the increase of the density of UAVs, the number of unnecessary HOs for all methods decrease. This is because the more UAVs are distributed around UE, it is more likely to choose a BS with higher ToS. It is observed from the figure that the proposed method has the lowest number of unnecessary HOs and the conventional method has the highest number of unnecessary HO for various density of UAVs. Although both the TOPSIS-only and QL-only methods also use ToS as a criterion, the proposed algorithm performs the best because of the use of the ToS in both the candidate list generation and the final selection of BS.

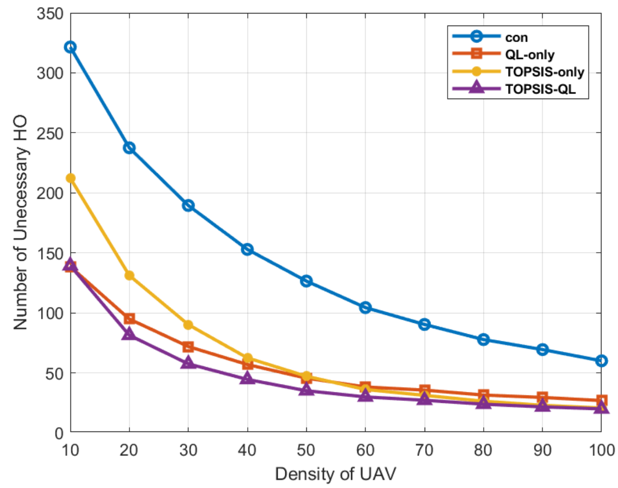


FIGURE 5. Number of unnecessary HO against density of UAV.

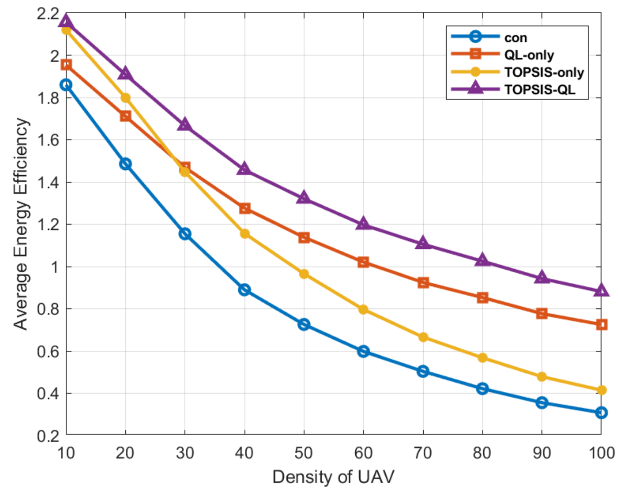


FIGURE 6. Average EE against UAV density.

C. AVERAGE ENERGY EFFICIENCY

The average EE is evaluated by averaging the EE cross the UE trajectory. The performance of the average EE against density of UAV is shown in Fig.6. In this simulation, the velocity of ground UE is fixed at 25 m/s. As Fig.6 shows, the average EE of all methods are decreasing with the increase of the UAV density. This is because the increase of UAV density leads to higher number of HOs, which interrupt the transmission more, leading to lower EE. Compared with the other three methods, the proposed method can achieve the best average EE.

V. CONCLUSION

A hybrid HO decision making scheme employing both TOPSIS and QL for UAV assisted three-tier HetNet is proposed in this paper. Specifically, TOPSIS with SD weighting technique is adopted to reduce the dimensions of the action space of QL by selecting the candidate BSs that provide better performance, and the QL can select a BS from the candidate BSs to achieve the best overall performance in terms of number of HOs, number of unnecessary HOs and the average EE. In the proposed scheme, SINR, ToS and EE are adopted as criteria. Simulation results demonstrate the effectiveness of the proposed TOPSIS-QL method in comparison with the conventional method, TOPSIS-only method, and the QL-only method. The hybrid scheme enhances both the performance and the scalability, allowing the applications in large scale networks. However, it is important to note that the proposed algorithm is designed for scenarios with fixed trajectories, such as deliver drivers, users in public transportation etc. In practice, users often move without prior knowledge of their paths. In our future work, we will consider HO decision making without prior path information; we will also tune the hyperparameters α and γ to make the Q-Learning more efficient and accurate.

REFERENCES

[1] A. Gupta, S. K. Gupta, M. Rashid, A. Khan, and M. Manjul, "Unmanned aerial vehicles integratedHetNetfor smart dense urban area," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 10, p. e4123, Oct. 2022, doi: 10.1002/ett.4123.

[2] A. Almohamad, F. Yilmaz, M. Hasna, and K. Qaraqe, "UAV-assisted HetNets: Association probability with instantaneous received power," in *Proc. Int. Balkan Conf. Commun. Netw. (BalkanCom)*, Aug. 2022, pp. 162–167, doi: 10.1109/BalkanCom55633.2022.9900672.

[3] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, Mar. 2021, doi: 10.1109/MNET.011.2000493.

[4] A. Sharma, P. Vanjani, N. Paliwal, C. M. W. Basnayaka, D. N. K. Jayakody, H.-C. Wang, and P. Muthuchidambaranathan, "Communication and networking technologies for UAVs: A survey," *J. Netw. Comput. Appl.*, vol. 168, Oct. 2020, Art. no. 102739, doi: 10.1016/j.jnca.2020.102739.

[5] H. Sedjelmaci, M. A. Messous, S. M. Senouci, and I. H. Brahmi, "Toward a lightweight and efficient UAV-aided VANET," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 8, p. e3520, Aug. 2019, doi: 10.1002/ett.3520.

[6] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016, doi: 10.1109/MCOM.2016.7470933.

[7] M. Alhabo, L. Zhang, and N. Nawaz, "GRA-based handover for dense small cells heterogeneous networks," *IET Commun.*, vol. 13, no. 13, pp. 1928–1935, Aug. 2019, doi: 10.1049/iet-com.2018.5938.

[8] T. Ayass, T. Coqueiro, T. Carvalho, J. Jailton, J. Araújo, and R. Francês, "Unmanned aerial vehicle with handover management fuzzy system for 5G networks: Challenges and perspectives," *Intell. Robot.*, vol. 2, no. 1, p. 2036, Feb. 2022, doi: 10.20517/ir.2021.07.

[9] S. Mignardi, P. M. Pulcinella, and R. Verdone, "The role of ground-to-air handovers in B5G UAV-aided mobile networks," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–2, doi: 10.1109/CCNC49032.2021.9369612.

[10] Y. He, W. Huang, H. Wei, and H. Zhang, "Effect of channel fading and time-to-trigger duration on handover performance in UAV networks," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 308–312, Jan. 2021, doi: 10.1109/LCOMM.2020.3024686.

[11] Z. Yao, X. Liang, G.-P. Jiang, and J. Yao, "Model-based reinforcement learning control of electrohydraulic position servo systems," *IEEE/ASME Trans. Mechatronics*, vol. 28, no. 3, pp. 1446–1455, Jun. 2023, doi: 10.1109/TMECH.2022.3219115.

[12] A. Prado, F. Stöckeler, F. Mehmeti, P. Krämer, and W. Kellerer, "Enabling proportionally-fair mobility management with reinforcement learning in 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1845–1858, Jun. 2023, doi: 10.1109/JSAC.2023.3273705.

[13] Y. Ju, Y. Chen, Z. Cao, L. Liu, Q. Pei, M. Xiao, K. Ota, M. Dong, and V. C. M. Leung, "Joint secure offloading and resource allocation for vehicular edge computing network: A multi-agent deep reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 5555–5569, May 2023, doi: 10.1109/TITS.2023.3242997.

[14] Y. Chen, X. Lin, T. Khan, and M. Mozaffari, "Efficient drone mobility support using reinforcement learning," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6, doi: 10.1109/WCNC45663.2020.9120595.

[15] Q. Li, M. Ding, C. Ma, C. Liu, Z. Lin, and Y.-C. Liang, "A reinforcement learning based user association algorithm for UAV networks," in *Proc. 28th Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, Nov. 2018, pp. 1–6, doi: 10.1109/ATNAC.2018.8615400.

[16] B. Hu, H. Yang, L. Wang, and S. Chen, "A trajectory prediction based intelligent handover control method in UAV cellular networks," *China Commun.*, vol. 16, no. 1, pp. 1–14, Jan. 2019.

[17] A. Madelkhanova, Z. Becvar, and T. Spyropoulos, "Optimization of cell individual offset for handover of flying base stations and users," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3180–3193, May 2023, doi: 10.1109/TWC.2022.3216342.

[18] F. Bari and V. Leung, "Multi-attribute network selection by iterative TOPSIS for heterogeneous wireless access," in *Proc. 4th IEEE Consum. Commun. Netw. Conf.*, Jan. 2007, pp. 808–812, doi: 10.1109/CCNC.2007.164.

[19] M. Alhabo and L. Zhang, "Multi-criteria handover using modified weighted TOPSIS methods for heterogeneous networks," *IEEE Access*, vol. 6, pp. 40547–40558, 2018.

[20] B. Galkin, E. Fonseca, R. Amer, L. A. DaSilva, and I. Dusparic, "REQIBA: Regression and deep Q-learning for intelligent UAV cellular user to base station association," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 5–20, Jan. 2022, doi: 10.1109/TVT.2021.3126536.

[21] *Enhanced LTE Support for Aerial Vehicles*, document TR 36.777, 3GPP, 2017. [Online]. Available: http://www.3gpp.org/specs/archive/36 series/36.777

[22] *5G Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document TR 38.901, 3GPP, 2017. [Online]. Available: https://www.3gpp.org/specs/archive/36 series/38.901

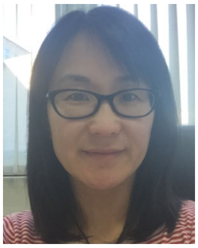
[23] P. Dastranj, V. Solouk, and H. Kalbkhani, "Energy-efficient deep-predictive airborne base station selection and power allocation for UAV-assisted wireless networks," *Comput. Commun.*, vol. 191, pp. 274–284, Jul. 2022, doi: 10.1016/j.comcom.2022.05.001.

[24] E. Obayiuwana and O. E. Falowo, "Network selection in heterogeneous wireless networks using multi-criteria decision-making algorithms: A review," *Wireless Netw.*, vol. 23, no. 8, pp. 2617–2649, Nov. 2017, doi: 10.1007/s11276-016-1301-4.

[25] Y.-M. Wang and Y. Luo, "Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making," *Math. Comput. Model.*, vol. 51, nos. 1–2, pp. 1–12, Jan. 2010, doi: 10.1016/j.mcm.2009.07.016.

[26] V. Yajnanarayana, H. Rydén, and L. Hévizi, "5G handover using reinforcement learning," in *Proc. IEEE 3rd 5G World Forum (5GWF)*, Sep. 2020, pp. 349–354, doi: 10.1109/5GWF49715.2020.9221072.

JIHAI ZHONG received the B.S. degree in electronic information engineering from Guangdong University of Petrochemical and Technology, Maoming, China, in 2019, and the M.S. degree in communication and signal processing from the University of Leeds, Leeds, U.K., in 2020. He is currently pursuing the Ph.D. degree with the School of Electronic and Electrical Engineering, University of Leeds. His current research interests include the handover management in UAV-assisted heterogeneous networks and resource allocation in UAV-assisted mobile edge computing systems.



LI ZHANG (Senior Member, IEEE) received the Ph.D. degree in communications from the University of York, in 2003. Currently, she is an Associate Professor and leads the Wireless Communication Group, School of Electronic and Electrical Engineering, University of Leeds, U.K. Her research interest is focused on wireless communications and signal processing techniques, such as massive MIMO, mmWave communications, heterogeneous network, device to device communications, and 5G systems. She has been the Ph.D. examiner for numerous universities. In 2006, she became a fellow of the Higher Education Academy. She has selected as a member of the prestigious U.K. EPSRC Peer Review College, in 2006, and regularly helps reviewing grant applications for research councils and book proposals. In 2005, she received the Nuffield Award for a Newly Appointed Lecturer. In 2011, she was awarded as an IEEE exemplary reviewer. She has been serving on the technical program committees of most major IEEE conferences in communications, since 2006. She is an associate editor of IEEE journals.

MOHANAD ALHABO received the B.Sc. degree in computer and information engineering from the Faculty of Electronics Engineering, University of Mosul, Mosul, Iraq, in 2007, the M.Sc. degree in computer network administration and management engineering from the University of Portsmouth, Portsmouth, U.K., in 2009, and the Ph.D. degree from the School of Electronic and Electrical Engineering, University of Leeds, Leeds, U.K., in 2018. He was a Network Engineer for over four years. He is currently with Informatics and Telecommunication Public Company, Iraqi Ministry of Communications. His research interests include mobility management and handover and interference management for heterogeneous networks.

JONATHAN SERUGUNDA received the B.Sc. degree in electrical engineering from Makerere University, in 2005, the M.Sc. degree in communication engineering from The University of Manchester, U.K., in 2008, and the Ph.D. degree in electrical and electronic engineering from the University of Bristol, U.K., in 2015. He is a Lecturer with the Department of Electrical and Computer Engineering, Makerere University. He is a member of netLabs!UG, which is a wireless research center based at Makerere University. His research interests include radio wave propagation and antenna design, design and analysis of wireless networks, physical layer security, and unmanned aerial vehicles (UAVs) assisted communication systems.

SHEILA N. MUGALA received the B.S. and M.S. degrees in electrical engineering from Makerere University, Kampala, Uganda, in 2006 and 2013, respectively.

She has been teaching with Makerere University, since 2006. Her research interests include wireless communications, with a focus on unmanned aerial vehicles, green communications, and electronics, with a focus on low cost humanitarian products, instrumentation, and control systems.