Proceedings Paper:
Zhong, J, Zhang, L orcid.org/0000-0002-4535-3200, Gautam, P et al. (2 more authors) (2023) An Improved Q-Learning Based Handover Scheme in Cellular-Connected UAV Network. In: Proceedings of the 2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC). WPMC 2022 - The 25th International Symposium on Wireless Personal Multimedia Communications, 30 Oct - 02 Nov 2022, Herning, Denmark. IEEE , pp. 520-525. ISBN 978-1-6654-7319-4

https://doi.org/10.1109/WPMC55625.2022.10014798

An Improved Q-Learning Based Handover Scheme in Cellular-Connected UAV Network

Jihai Zhong, Li Zhang, Prabhat Raj Gautam
School of Electrical and Electronic Engineering
University of Leeds
Leeds, UK
E-mail:{ml18j34z, L.X.Zhang, elprg}@leeds.ac.uk

Jonathan Serugunda, Sheila Mugala
Electrical and Computer Engineering Department
Makerere University
Kampala, Uganda
E-mail:{jonathan.serugunda, sheila.mugala}@mak.ac.ug

***Abstract*—Cellular-Connected Unmanned Aerial Vehicles (UAVs) have been used for a variety of applications, from remote sensing, monitoring to the search and rescue operations. Many applications of UAVs rely on the seamless and reliable communication link to control the UAVs remotely and transmit the data. Thus, the challenges due to the mobility in three dimensional space, such as the air-to-ground channels, interference and handover (HO) problems must be addressed. The HO failure and unnecessary HO seriously affect the quality of service. Q-learning is an effective method to address the challenges in HO and has attracted a lot of attention. In this work, we introduce some changes in an existing HO algorithm and proposed an improved Q-Learning based HO algorithm. The formation of the Action Space considers both the Signal-to-Interference-plus-noise ratio (SINR) and the distance between Base station (BS) and UAV with different weights. The results show the proposed algorithm can further reduce the HO rate by increasing the weight of distance with slightly compromising the throughput and the outage rate. An optimal distance weight is suggested based on the analysis of the three performance indicators.**

***Index Terms*—Unmanned Aerial Vehicle, handover, Q-learning**

I. Introduction

Unmanned Aerial Vehicles (UAVs), widely known as drones, are getting popular in various industries because of some unique characteristics. For example, they can move fast in three-dimensional space because of the high agility and adjustable working altitude. Comparing with other devices, UAVs are easy to deploy with low cost in a short time [1]. Therefore, they have great potential to significantly improve the transmission performance over the direct ground-to-UAV communications when they are connected with cellular networks as cellular-connected UAVs. Cellular-connected UAV is a promising technology to integrate UAVs for various applications into the existing cellular networks as aerial user equipments (UEs) [2]. When UAVs move in the air, Handover (HO) takes place to transfer the connection from the current base station to the next one and ensure a smooth transmission. Interruption and package loss can happen in case of handover failure. Therefore, suitable HO technology is essential to ensure seamless transmission. However, HO for cellular connected UAVs is a challenging problem because UAVs can move in three-dimensional space. Unlike the ground UE, the BS can only serve UAVs with side lobes which has lower transmit power because the main lobes are tilted downwards for ground UEs. In addition, when the UAVs are served by the ground BS, the Air-to-Ground (A2G) channel could be line-of-sight (LOS) and Non-line-of-sight (NLOS). In fact, most of the A2G channel is more likely to be line-of-sight (LOS) since there are less obstacles in the air. As a result, cellular-connected UAVs may experience more interference in both uplink and downlink as UAVs can still detect strong signals from other user equipment (UE) and BSs. Meanwhile, the fast moving UAVs may cause frequent and unnecessary HOs, which increase the risk of interruption and package loss. Therefore, reducing the number of HO with appropriate link quality is critical for cellular-connected UAVs [3].

In order to reduce the unnecessary HO, it is important to make an appropriate decision so that the UAVs can always connect to the suitable BS. This problem has attracted some attention recently. Fuzzy inference is useful to take into account multiple input and output parameters which have unclear and non-linear relationship between each other [5]. In [6], multiple variables such as the received signal strength (RSS), altitude, and speed are considered in the HO decision making. They are divided into two groups to measure network and terminal information. Then information is converted into membership function values which are 'Speed Limit' and 'Coverage'. The final output estimation level related to HO decision is determined by the values and fuzzy rules. Reinforcement Learning (RL) method is used in [7], where the authors proposed a novel learning-based strategy to dynamically adjust the speed of UAVs. As a result, the strategy achieves a balance between time of mission accomplishment, energy consumption, connectivity time and HO rate. Another RL based HO optimization mechanism for a cellular-connected UAVs system is proposed in [8]. In the algorithm of [8], the state is the combination of current position and current serving BS, the action consists of the list of potential target BSs, and the reward equation is related to the reference signal received power (RSRP) and HO cost indicator $I\left(HO\right)$. A complete Q table for a certain trajectory is generated after enough iteration, which can make HO decisions leading to reduced HO number. The simulation results have shown that the proposed approach can significantly reduce the number of HOs at the cost of decreased quality of service of the communication link.

In this paper, we use the similar method with [8]. But unlike [8], this paper adopts signal-to-interference-plus-noise-ratio (SINR) as the criteria of HO to determine the potential target BSs. To reduce interference, the classical reuse-3 frequency reuse scheme is considered in this work. Apart from SINR, the distance between BSs and UE is also taken into account in order to improve the Action Space of the algorithm. The results show that the proposed algorithm can further decrease the HO rate and achieve the best trade-off between HO rate, outage rate and throughput if a suitable proportion of distance is identified for a certain trajectory.

The rest of this paper is organized as follows. System model is described in Section II. The framework of the proposed RL method is introduced in Section III. Section IV presents the simulation results. Conclusion of this paper follows in section V.

II. System Model

It is common that the antennas are installed above the rooftop levels in Urban environment, and Urban-macro with aerial vehicles (UMa-AV) scenario is used to represent it [9]. Therefore, an UMa-AV scenario is considered in the paper, where K BSs are deployed with a fixed distance d between each other. There are three carrier frequencies CF_1, CF_2, CF_3, and the BSs deployed adjacent to each other have different frequencies. A UAV is assumed to move in the air along a fixed trajectory. The trajectory is generated randomly in three-dimensional space in the specific area. When a UAV is moving, it will pass through the coverage areas of deployed ground BSs, and be served by different BSs. To maintain a reliable network link, HO must be performed so that the UAV can always connect with a suitable BS. Along the path, there are many way-points separated by interval t. At every way-point, the UAV should make the decision of whether the HO is triggered, and which is the target BS.

A. Trajectory

A fixed trajectory is applied in the proposed learning scheme, which is generated randomly. The key parameters that decide the trajectory are the start point l_1 and n way-points along the trajectory. l_1 is generated randomly in the area, and the direction and speed of UAV are randomly picked within a certain range. The direction is decided by the horizontal angle θ_1 and the vertical angle θ_2, where $\theta_1 \in [0°, 360°)$ and $\theta_2 \in [-90°, 90°]$. The maximum speed is V_{max}, so the speed V is generated between 0 and V_{max}. After a specific interval time t, a new way-point is generated, and all the related parameters are generated again until the way-point l_n.

B. Propagation Model

In this paper we only consider A2G channels, therefore the propagation is different with the conventional two-dimensional channel. As mentioned above, A2G channel could be LOS and NLOS, which have different propagation models. The probability of a LOS channel in UMa-AV is defined in [4]:

$$P_{LOS} = \begin{cases} 1 & d_{2D} \le d_1 \\ \frac{d_1}{d_{2D}} + e^{\frac{-d_{2d}}{P_1}} \left(1 - \frac{d_1}{d_{2D}}\right) & d_{2D} > d_1 \end{cases}, \quad (1)$$

where

$$P_1 = 233.98 \times log_{10}(h_{UT}) - 0.95, \quad (2)$$

$$d_1 = max(460 \times log_{10}(h_{UT}) - 700, 18), \quad (3)$$

where d_{2D} is the two-dimensional distance between the UE and the BS, and $h_{UT} \in (22.5, 100]$ is the height of the current position of the UE. According to [4], if the channel is LOS, the path loss between the UE and the BS is defined as:

$$PL_{LOS} = 28.0 + 22 \times log_{10}(d_{3d}) + 20 \times log_{10}(f_c), \quad (4)$$

where d_{3d} is the 3D distance between the UE and the BS, and f_c is the carrier frequency. On the contrary, when the channel is NLOS, its path loss to BS is calculated as

$$PL_{NLOS} = -17.5 + (46 - 7 \times log_{10}(h_{UT})) \times log_{10}(d_{3d}) + 20 \times log_{10}\left(\frac{40\pi f_c}{3}\right). \quad (5)$$

SINR is the metric used in HO decision making. To calculate the SINR, downlink interference from other BSs and RSS between UE and serving BSs should be calculated. The RSS in dBm is calculated as follows

$$RSS = P_{tr} + G_1 + G_2 - PL - SF, \quad (6)$$

where P_{tr} is the transmit power of the BS, G_1 and G_2 are the antenna gains of the BS and UE respectively which are both 0 in dB in this paper, PL is the Path loss, and SF is the shadowing between a BS and a UE. SF is Gaussian distribution with mean 0 and deviation σ_{SF}:

$$SF = \mathcal{N}(0, \sigma_{SF}). \quad (7)$$

For the LOS channel, σ_{SF} is calculated as

$$\sigma_{SF} = max\left(4.64 \times e^{-0.0066 h_{UT}}, 2\right), \quad (8)$$

and for NLOS channel, $\sigma_{SF} = 6$. When the UE is served by the BS i, the interference received by the UE from other BSs is expressed as

$$I = \sum_{j=1, j \neq i}^{k-1} RSS_j, \quad (9)$$

where $k-1$ is the total number of interfering BSs, j represents a certain BS, and unit of RSS_j is mW. The SINR measured when UE is served by BS i can be obtained as

$$SINR = \frac{RSS_i}{I + {\sigma_n}^2}, \quad (10)$$

where σ_n is the noise power, the calculation of σ_n in dBm is

$$\sigma_n = -174 + 10 \times log_{10}(B), \quad (11)$$

where B is the bandwidth. Combining (10) and (11), the SINR calculation is represented as

$$SINR = \frac{RSS_i}{\sum_{j=1,j\neq i}^{j=k-1} RSS_j + {\sigma_n}^2}. \quad (12)$$

Throughput is a crucial indicator of the quality of link and can be expressed as

$$T = B \times log_2 (1 + SINR). \quad (13)$$

III. IMPROVED RL-BASED HO SCHEME

In this section, an improved RL-based HO scheme is introduced. RL is a learning algorithm where the agent takes action to change its state based on the reward and interacts with the environment [11]. It is a model-free learning method, and it works based on Markov process. The objective of the scheme is to make HO decision at each way-point along the trajectory. Unlike [8], the proposed HO scheme involves distance as a parameter in action space forming.

A. Q-learning Algorithm in [8]

1) Definition:

- State: The state in the RL framework consists of the positions of the UAV and the current serving BS, which is represented as $S = [P_s, c_s]$. P_s is the coordination of way-points when UE is at the state s and c_s is the current BS, where $c_s \in C_s$, and C_s is the set of candidate BSs at state s.
- Action: The action A_s ($A_s \in A$) at every state s is the decision made to choose a serving BS for the next state $s^{'}$. The decision determines the c_s when the UE reaches the next way-point. Therefore, different action can lead to different state. In [8], the action space A consists of BSs that the UAV can select at a way-point, which can be considered as candidate list $C_{s'}$.
- Reward: The reward is a metric of different actions at different states. In the reward equation, the importance of HO criteria, which is RSRP in [8], should be reflected. Apart from that, to reduce the number of HOs, if c_s of s is still qualified in the candidate list of $s^{'}$, then it is more likely that UE stays with c_s. The equation of the reward is

$$R = -\omega_{HO} \times I(HO) + \omega_{RSRP} \times RSRP_{s'}, \quad (14)$$

where ω_{HO} and ω_{RSRP} are the weights for $I(HO)$ and $RSRP_{s'}$. $RSRP_{s'}$ is the normalized RSRP of the target BS, and $I(HO)$ is the HO indicator which will be 1 if HO happens, otherwise it will be 0.

2) Algorithm: The algorithm in [8] is as follow. An Q-table is firstly created and represented as Q. After a trajectory generated with L way-points, Q with state set S and action set A is generated to store the content of Q which is the reward R for each state and action. For every state s, the action space A consists of the BSs that can be selected, which can be represented as candidate BS list. Action of state s denoted as A_s consists of k BSs ($k < K$), which are potential to be the target BS. The options of action for an agent are known as action space. So, the size of Q is $L \times k \times k$, and the rewards of all the decisions are calculated and stored in Q at every state. After that, it is a phase of Q-table training with Q value iterations in each episode. The policy is Epsilon-Greedy (ϵ - greedy) Policy, which is followed in every iteration episode. At every state s, the option with higher Q value is chosen with the probability of ϵ, and there is $1 - \epsilon$ chance to select an option randomly. Then, the Q value of the action will be updated by Bellman Equation

$$Q(s,a) \leftarrow Q(s,a) + \alpha \times [R + \gamma \times max_{a'} Q(s',a') - Q(s,a)], \quad (15)$$

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. After iterations, an updated Q is created, which instructs the sequence of HO decisions for every way-point of the trajectory.

B. Improved Algorithm

In the improved algorithm, the channel quality is SINR instead of RSRP because it quantifies the relationship between channel condition and throughput better. Therefore the equation of reward for Q table in the improved scheme is

$$R = -\omega_{HO} \times I(HO) + \omega_{SINR} \times SINR_{s'}, \quad (16)$$

where ω_{SINR} is the weights for $SINR_{s'}$. $SINR_{s'}$ is the normalized SINR of the target BS. The formation of Action space A in the proposed algorithm is improved. A is the BS options that the agent can choose. In [8], A consists of the k BSs with the highest channel quality, which is RSRP. In other word, A in [8] only considers the channel quality with BS. However, in this work, for the further number of HO reduction, A is not only related to the SINR but also the distance between BS and the way-point l since the involvement of distance can improve the similarity of action spaces at adjacent states, so UAV can possibly stay with the current BS to avoid HO. To balance the two parameters, there is a reward parameter $\mathbb{R}$ proposed to decide which BS should be included in A. The k BSs with the highest $\mathbb{R}$ can be included in A, and $\mathbb{R}$ is determined by both SINR and distance:

$$\mathbb{R}_{l,i} = \omega_{sinr} \times sinr_{l,i_n} - \omega_d \times d_{l,i_n}, \quad (17)$$

where $\mathbb{R}_{l,i}$ represents the value of the reward of BS i when UAV is at the way-point l, $sinr_{l,i_n}$ is the normalized SINR of BS i, d_{l,i_n} is the normalized distance in three-dimensional area between way-point l and BS i, ω_{sinr} and ω_d are the weight of SINR and distance. With different weight combination of ω_{sinr} and ω_d, A is different, subsequently affecting each HO decision. The algorithm of the formation of A is shown in Algorithm 1. After A in every way-point is formed, the Q value is iterated following Algorithm 2.

IV. SIMULATION RESULTS

In this section, comparison of proposed HO scheme with the work in [8] is presented in terms of HO rate, HO outage

Algorithm 1 Formation of Action Space in improved HO scheme

Initialize input parameters:
  Waypoints L, number of BS K
  Action Space A, $A \leftarrow 0_{L\times k}$
  Set ω_d, ω_{sinr}
for l ($l \in L$) **do**
  for BS $i \leftarrow 1$ to K **do**
    get SINR $sinr_{l,i}$,
    get Distance $d_{l,i}$
  end for
  $sinr_{l,i}$ = Normalized SINR
  $d_{l,i}$ = Normalized Distance
  for BS $i \leftarrow 1$ to K **do**
    Calculate $\mathbb{R}_{l,i}$ with Equation (17)
    $\mathbb{R}_l(i) = \mathbb{R}_{l,i}$
  end for
  $A[l,:]$ = k BSs with best $\mathbb{R}_{l,i}$
end for
return A

Algorithm 2 Q value iteration

Initialize input parameters:
  Waypoints L, number of BS K
  Q table Q, $Q \leftarrow 0_{L\times k\times k}$
  $HO(k,k) \leftarrow 0_{k\times k}$
  Action Space A in Algorithm 1
  Set $\omega_{HO}, \omega_{SINR}, \omega_d, \omega_{sinr}, \lambda, \alpha, \gamma$
for l in $L-1$ **do**
  for $x = 1 : k$ **do**
    for $y = 1 : k$ **do**
      get normalized $SINR_s$ of $A(l,:)$
      $Q(l,x,y)$=ω_{SINR}×SINR$_s(y)$
      if $A(l,x) \neq A(l+1,y)$ **then**
        $HO(x,y)$=0
      else
        $HO(x,y)$=1
      end if
    end for
  end for
  $Q(l,x,y)$ is calculated with equation (16)
  Reward matrix $R = Q$
  while training $< n$ **do**
    $j = 0$;
    for l in L **do**
      if random value i($i \in$[0,1])$< \epsilon$ **then**
        picked BS $j_{new} \leftarrow argmax$Q(l,j,u)(u$\in k$)
      else
        $j_{new} \leftarrow$randomly picked from $A(l, 1:k)$
      end if
      Update $Q(l,j,j_{new})$ with equation (15)
    end for
    $j = j_{new}$;
  end whilereturn Q

rate and throughput. The performance is evaluated by taking the average of 100 times assuming the UAV moves along the same trajectory. To reflect the effect of different combination of ω_{sinr} and ω_d, the ratio of ω_{HO} and ω_{SINR} in the Q table reward function is fixed at 3/7. To illustrate the trend clearly with different proportion of ω_d, the abscissas of all the performance results are ω_d from 0 to 10, and ω_{sinr} is $10-\omega_d$. Fig.1 shows the simulation scenario, in which the asterisks are the locations of ground BSs, and the line is the trajectory in the area which is generated randomly. The simulation parameters are listed in Table I.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
Bandwidth (B)	10 MHz
Transmit Power (P_{tr})	44 dBm
BSs distance (D)	500m
Number of BSs (K)	19
Way-points (L)	1000
Action Space Size (k)	5
Side length of area	2000 m
Height of area	20 m∼300 m
Height of BSs	10 m
Carrier Frequency (CF)	CF_1 = 1990 MHz CF_2 = 2000 MHz CF_3 = 2010 MHz
Threshold (P_{th})	3 dB
Max speed of UAV (V_{max})	20 m/s

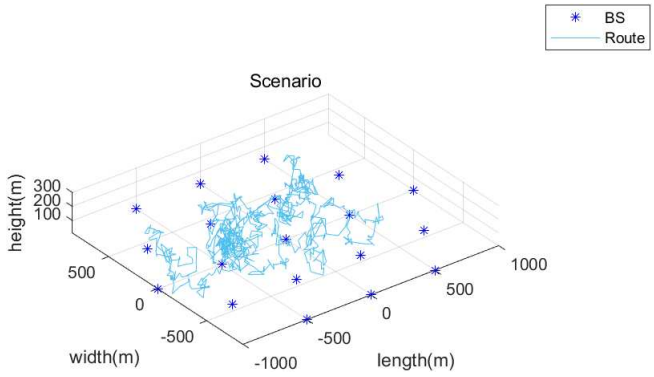

Fig. 1. Simulation scenario.

A. Performance

1) HO Rate: HO rate R_{HO} reflects the HO frequency in a trajectory. It is the proportion of HO happened in the 1000 way-points along the trajectory. If the serving BS is different with the previous serving BS after the HO decision at a way-point, HO happens at that way-point.

2) Outage Rate: Outage rate R_{out} is defined as the proportion of way-points where SINR cannot meet the requirement after the HO decision in all way-points. Outage happens when the SINR of serving BS is less than the threshold P_{th} after the HO decision.

3) Throughput and Throughput Loss Rate: Throughput T is the average value of throughput at all way-points along the trajectory after the HO decision. Throughput loss rate R_{TL} is defined as the proportion of the margin between the proposed average throughput and the baseline:

$$R_{TL} = \frac{T_{baseline} - T}{T_{baseline}}, \tag{18}$$

where the $T_{baseline}$ is the average throughput when UAV is only connected with the BS with the best SINR at every way-point in the same trajectory. Both of them reflect the throughput results. However, the higher throughput or the less throughput loss rate means the better communication quality.

4) Performance Indicator: Performance Indicator $I(P)$ consists of the three indicators mentioned above. $I(P)$ is a synthetic performance, which is related to HO rate, outage rate and throughput loss rate. Eventually, the calculation of $I(P)$ is represented as

$$I(P) = R_{HO} + R_{out} + R_{TL}, \tag{19}$$

where R_{HO} is HO rate, R_{out} is the outage rate, and R_{TL} is the throughput loss rate. The smaller value of $I(P)$ indicates the better performance.

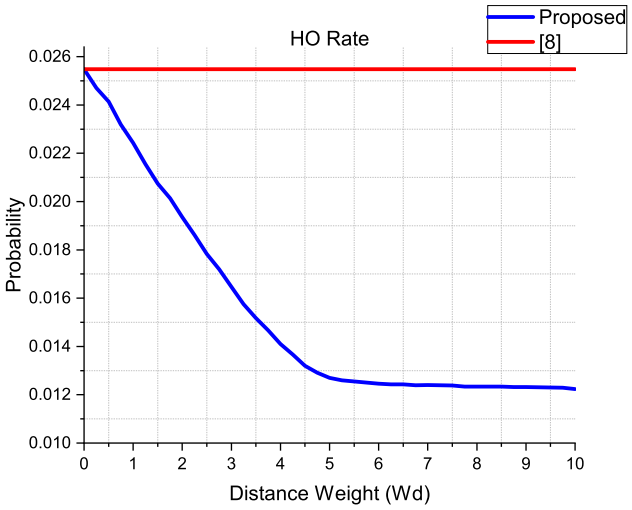

Fig. 2. HO rate with the increase of ω_d.

B. Performance Analysis

In Fig.2, we can see the simulation results of HO rate with different ratios between ω_d and ω_{sinr}. It shows that when ω_d equals to 0, which means distance is not involved in forming the Action Space, the performance is the same with the result of [8]. With the increase of ω_d , the HO rate decreases significantly. Therefore, HO rate can be reduced with the consideration of distance.

With the increase of ω_d, there are negative effect on throughput and outage rate. Compared with [8], the outage rate is increased and the throughput is reduced. However, the

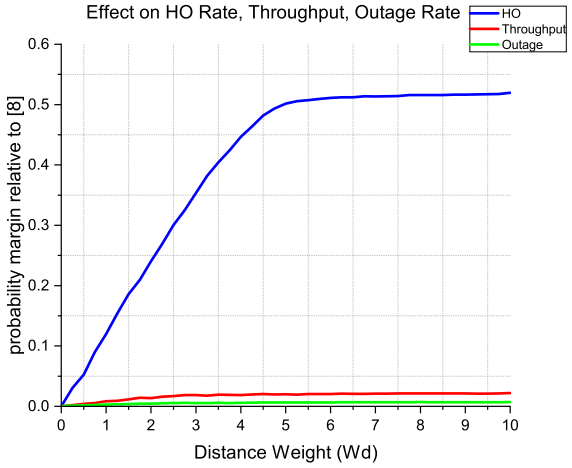

Fig. 3. Effect on HO rate, throughput, outage rate of different value of ω_d.

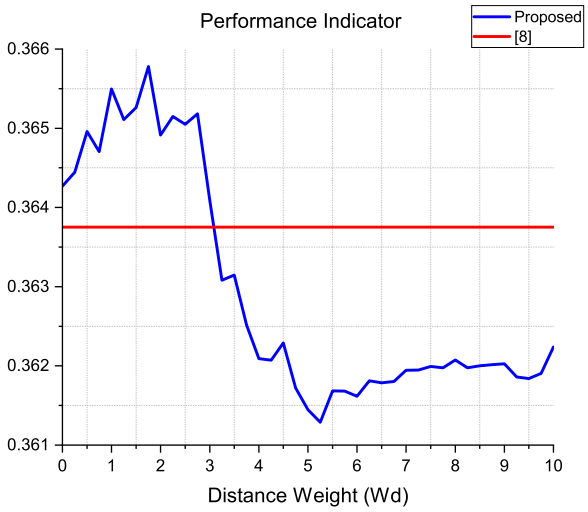

Fig. 4. Performance indicator of different value of ω_d.

effects on the three indicators are different. The effects are reflected by the margin between the results of the proposed scheme and that of [8]. Fig.3 shows the the margin relative to the corresponding results in [8] with the increase of ω_d. From the graph, the effect on HO rate is significantly larger than outage rate and throughput. Therefore, the proposed scheme can remarkably reduce HO rate without greatly compromising outage rate and throughput.

To find the best trade-off between HO rate, and the HO outage rate as well as the throughput, a performance indicator $I(P)$ which is the combined performance indicator as defined in (18) is illustrated in Fig.4. The value of $I(P)$ fluctuates with the increase of ω_d. From Fig.4, it is obvious that $I(P)$ of the proposed algorithm is less than [8] when the weight increases to a certain extend. Additionally, when ω_d is around 5.25 and ω_{sinr} equals to 4.75, $I(P)$ achieves its minimum,

corresponding to the best performance trade-off of HO rate, outage rate and throughput.

V. Conclusion

In this work, we consider a scenario where HOs are required to ensure a seamless connection between UAV and ground BSs as it is served by different ground BSs while moving along a trajectory. However, frequent handovers may take place which is undesirable. In this paper, we present an RL-based scheme for HO decision by proposing improvements in an existing algorithm. In the proposed scheme, the forming action space A considers not only RSS or SINR, but also the distance with selected weights. Different combination of ω_{sinr} and ω_d can lead to different A and different performance. Simulation results show that, the proposed method can significantly reduce the HO rate compared to the existing algorithm at the cost of slightly increased outage rate and reduced throughput. Based on the analysis of simulation results, the best ratio of ω_{sinr} and ω_d are $\omega_d = 4.75$ and $\omega_{sinr} = 5.25$. In this way, the improved algorithm can achieve greatly reduced HO rate without compromising much the HO outage and the throughput.

References

[1] M. Mozaffari, W. Saad, M. Bennis, Y. -H. Nam and M. Debbah, "A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems," in IEEE Communications Surveys and Tutorials, vol. 21, no. 3, pp. 2334-2360, thirdquarter 2019, doi: 10.1109/COMST.2019.2902862.

[2] M. Hassanalian and A. Abdelkefi, "Classifications, applications, and design challenges of drones: a review," Prog. Aerosp. Sci., vol. 91, pp. 99-131, May 2017.

[3] J. Angjo, I. Shayea, M. Ergen, H. Mohamad, A. Alhammadi and Y. I. Daradkeh, "Handover Management of Drones in Future Mobile Networks: 6G Technologies," in IEEE Access, vol. 9, pp. 12803-12823, 2021, doi: 10.1109/ACCESS.2021.3051097.

[4] 3GPP TR 36.777, "Enhanced LTE support for aerial vehicles," Online: ftp://www.3gpp.org/specs/archive/36 series/36.777.

[5] L. A. Zadeh, "On fuzzy algorithms," in Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers, L. A. Zadeh, Ed. Singapore: World Scientific, 1996, pp. 127–147.

[6] E. Lee, C. Choi and P. Kim, "Intelligent Handover Scheme for Drone Using Fuzzy Inference Systems," in IEEE Access, vol. 5, pp. 13712-13719, 2017, doi: 10.1109/ACCESS.2017.2724067.

[7] M. M. Azari, A. H. Arani and F. Rosas, "Mobile Cellular-Connected UAVs: Reinforcement Learning for Sky Limits," 2020 IEEE Globecom Workshops (GC Wkshps, 2020, pp. 1-6, doi: 10.1109/GCWkshps50303.2020.9367580.

[8] Y. Chen, X. Lin, T. Khan and M. Mozaffari, "Efficient Drone Mobility Support Using Reinforcement Learning," 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea (South), 2020, pp. 1-6, doi: 10.1109/WCNC45663.2020.9120595.

[9] S. D. Muruganathan et al., "An Overview of 3GPP Release-15 Study on Enhanced LTE Support for Connected Drones," in IEEE Communications Standards Magazine, vol. 5, no. 4, pp. 140-146, December 2021, doi: 10.1109/MCOMSTD.0001.1900021.

[10] M. Toril, S. Pedraza, R. Ferrer and V. Wille, "Optimization of handover margins in GSM/GPRS networks," The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring., 2003, pp. 150-154 vol.1, doi: 10.1109/VETECS.2003.1207520.

[11] R. S. Sutton, A. G. Barto et al., Introduction to reinforcement learning. MIT press Cambridge, 1998, vol. 2, no. 4.