IOP Conf. Series: Materials Science and Engineering **719** (2020) 012044 doi:10.1088/1757-899X/719/1/012044

A QoE-Oriented Network Selection Mechanism in Heterogeneous Wireless Networks

Chao He[1,2], Zhidong Xie[1,2] and Chang Tian[1]

[1] College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China
[2] National Innovation Institute of Defense Technology, Academy of Military Sciences of PLA, Beijing 100071, China
Email: xzd313@163.com

Abstract. In order to guarantee the UAV video uplink in heterogeneous networks, a novel access network selection mechanism was proposed in this paper. Due to the unknown Channel State Information(CSI), the issue was first modeled as the Multi-Armed Bandit (MAB) problem. Then, the reward function was built and the Quality of Experience (QoE) of video was considered. By means of three typical solving algorithms, decisions could be made between utilization and exploration. The simulation results show that the mechanism can effectively solve the network selection problem of video transmission in heterogeneous networks.

1. Introduction

The rapid development of Internet of Things and aviation technologies has made aerial photography through Unmanned Aerial Vehicles (UAVs) more and more popular. Video streaming captured by the UAVs could be sent to the ground by all kinds of wireless networks in real time. The overlaps of heterogeneous wireless networks indeed provide more convenience for UAV communication. But, at the same time, they also trigger the access network selection problem. Because of the particularity of video transmission service, the selection strategy needs to start from the application layer and take into account the characteristics of different wireless channels.

Due to the time-varying property of wireless channel and the mobility of UAV, the bandwidth of access network will not always keep the same. This bandwidth fluctuation adds uncertainties to the selection of access network. At the same time, due to channel fading effect and other reasons, packet loss is inevitable in video transmission, which is always random and increases the complexity of network selection. In the case of unknown CSI, the UAV needs to constantly perceive the transmission environment, learn from the network, so as to provide reference for their next decision. Therefore, the solution of this problem will be a sequential decision-making process, which can be solved by the Multi-Armed Bandit (MAB)[1] model. As a classical reinforcement learning algorithm, the model has been widely used in many aspects, such as spectrum access and networks handover[2,3]. It generally includes a decision maker and multiple optional arms. The decision maker needs to make selection among the optional arms and keep a balance between utilization and exploration.

Traditional access network selection methods often focus on Quality of Service(QoS) of the network, ignoring the needs of users. When considering the special scenario of UAV video transmission, it is necessary to start from the perspective of improving QoE for users. The main contributions of this paper are as follows:

- The network selection mechanism of UAV video uplink is modeled as a Multi-Armed Bandit problem, and the optimization objective is determined.
- Based on a QoE function according to video content classification, the reward function of the UAV video transmission is established.
- The access network selection model is solved by three different algorithms.
- The model is verified by some simulations and the differences of the three solving algorithms are analyzed.

2. Network Selection Model Based on MAB

We consider a specific service area W in the communication. It is in the overlap area of M heterogeneous access networks, as shown in the red circular area in figure 1. UAV D performs video shooting and uplink missions during flight. Assuming that UAV D is always flying in area W and T is the total transmission time slots of video, the optional access network set can be expressed as $\mathbf{S} = \{s(1), s(2), \ldots, s(T)\} \subseteq \mathcal{M}$, and the reward function set is $\mathbf{U} = \{u(1), u(2), \ldots, u(T)\}$.

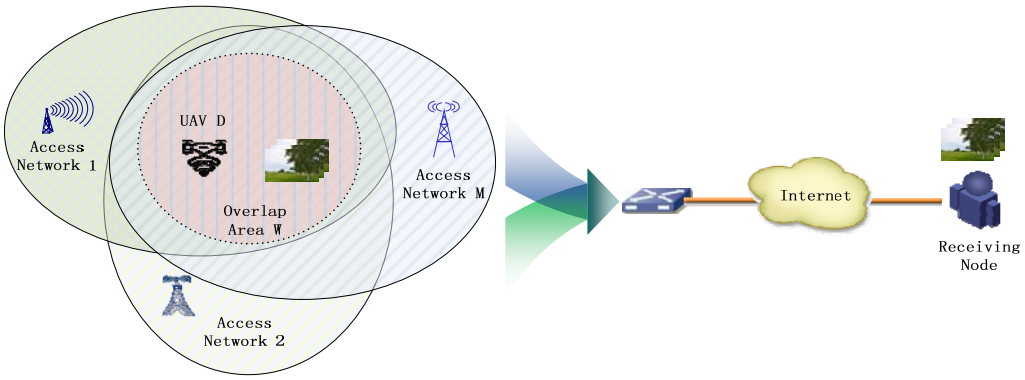


Figure 1. System Model

Because the CSI of each access network is time-varying, we will make a statistical description of its uncertainty. It is assumed that the CSI remains unchanged in each slot, and the changes between different slots follow an independent and identical distribution. First, in the slot $i \in [1, 2, ..., T]$, the transmission rate that the network v can carry, i.e. the available channel bandwidth of v, is a fixed value, $r_v(i)$. Then, in this slot, the bandwidth information vectors of M heterogeneous networks can be expressed as $\Re(i) = \{r_1(i), r_2(i), \ldots r_M(i)\}$. The total mean bandwidth vector of the M networks is $\overline{\Re} = \{\overline{R_1}, \overline{R_2}, \ldots, \overline{R_M}\}$. Similarly, in the slot i, the packet loss rate of the network v is set to a fixed value $l_v(i)$. Then the packet loss information vector of M channels can be expressed as $\Gamma(i) = \{l_1(i), l_2(i), \ldots, l_M(i)\}$, and the mean vector of M channels is $\overline{\Gamma} = \{\overline{L_1}, \overline{L_2}, \ldots, \overline{L_M}\}$.

Next, we model the network selection problem as an online learning problem based on the MAB. The heterogeneous network set $\mathcal{M} = \{1, 2, \ldots, M\}$ can be regarded as a set of optional arms, each of which corresponds to a certain reward $\mathcal{Q} = \{q(1), q(2), \ldots, q(M)\}$. Because of the randomness of CSI, the value of this reward will change in different time slots, and its expected value in all the time remains unchanged, which is recorded as $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_M\}$. As a decision maker, UAV D needs to continuously select an arm from the set of optional arms $\mathcal{M}$, so as to make a selection in an access network. Generally speaking, the selection set made in different time slots can be recorded as $\mathbf{S_{MAB}} = \{s_{MAB}(1), s_{MAB}(2), \ldots, s_{MAB}(T)\} \subseteq \mathcal{M}$, and the corresponding reward function set is $\mathbf{U_{MAB}} = \{u_{MAB}(1), u_{MAB}(2), \ldots, u_{MAB}(T)\}$. We record the historical information of this reward, and feedback this information to the system continuously for decision-making. Therefore, the reward

function is not only an evaluation method of current slot selection strategy, but also a reference basis for future decision-making. The selection and feedback process is iterated until the selection results do not change, and the algorithm converges, when the UAV will ultimately choose the arm (access network) that maximizes the cumulative reward function. The optimal selection strategy $S_{opt} = \arg\max_{s} \boldsymbol{\mu}$ is called Deterministic Stationary Policy (DSP).

Because of the unknown CSI, UAV D is not always able to make optimal judgments. Therefore, there will be a difference between the actual choice and the optimal choice, also known as regret value, which can be used to measure and judge the performance of decision-making. When the actual selection vector is $\mathbf{S_{MAB}}$ and the reward function vector is $\mathbf{U_{MAB}}$, the regret value can be expressed as $R_{\mathbf{S_{MAB}}}(i) = i \cdot \mu^* - E[\sum_{k=1}^{i} u_{MAB}(k)]$, where $\mu^* = \max \boldsymbol{\mu}$ is the optimal selection reward expectation. When UAV D transmits video in M heterogeneous access networks, we expect greater cumulative reward through continuous decision-making. This problem can be described as

$$\begin{aligned} \mathbf{S_{MAB}} &= \arg\max_{\mathbf{S}} \sum_{k=1}^{i} u_{MAB}(k) \\ s.t.\ \ & \mathbf{S} \subseteq \mathcal{M}, \mathbf{S_{MAB}} \subseteq \mathcal{M} \\ & i \in \{1, 2, \ldots, T\} \end{aligned} \tag{1}$$

$\mathbf{S_{MAB}}$ is the vector of selection results in different time slots in the iteration process.

3. QoE-oriented Reward Function

In order to effectively evaluate whether the access network selection algorithm can meet the needs of video users, we consider using QoE as a part of the reward function. We hope to establish a good matching relationship between CSI observations and users' real viewing experience by reasonably setting QoE-oriented reward functions in each slot. Usually, video QoE is closely related to the bit rate of video coding, packet loss rate in the network and frame rate in video playback. In this paper, we assume that the frame rate is a fixed value. This part of the reward function can be modeled as a function of bandwidth information vector, packet loss information vector and access network selection strategy of heterogeneous networks, $\mathbf{U_{QoE}} = K\{\mathbf{S}[\Re(i), \Gamma(i)], \Re(i), \Gamma(i)\}$. $\mathbf{S}[\Re(i), \Gamma(i)]$ represents the network selection vectors made by UAV D in M networks with CSI of $\Re(i)$ and $\Gamma(i)$ in each slot.

In addition to the above three factors of bit rate, packet loss rate and frame rate, the impact of video content characteristics on QoE can not be ignored. Literature [4] proposes that video can be divided into three categories: Slight Movement (SM), Gentle Walking (GW) and Rapid Movement (RM). In practical applications, the video uplinked by UAV may totally cover the three categories mentioned above. Assuming that the selected network in the slot i is $s(i) \in \mathcal{M}$, the channel rate is $r_s(i)$ and packet loss rate is $l_s(i)$, then the reward function QoE can be expressed as[5]

$$u_{QoE}(i) = K\{\phi, s(i), r_s(i), l_s(i)\} = \frac{c_1(\phi) + \ln[\frac{1 + r_s(i)}{c_2(\phi)}] + c_3(\phi) \cdot f}{c_4(\phi) + e^{c_5(\phi) \cdot l_s(i)}} \tag{2}$$

$\phi \in \{1, 2, 3\}$ represents that the video segment in current slot can be classified into one of the three categories mentioned above. $c_1(\phi), c_2(\phi), c_3(\phi), c_4(\phi), c_5(\phi)$ represents the relevant constants when video content belongs to different categories and their typical value can be found in [5]. f is frame Rate for Video.

In addition to the QoE, the reward function also needs to take into account the related cost. This paper considers the difference between QoE function and cost as the reward function, $\mathbf{U} = \mathbf{U}_{\mathbf{QoE}} - \mathbf{U}_{\mathbf{COST}}$. Usually, the cost needs to consider two factors: channel bandwidth leasing and energy consumption, which are linearly related to video transmission rate. If the correlation coefficient of the total loss of each network is set as $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_M\}$, then, the total reward function of the slot can be expressed as $u(i) = u_{QoE}(i) - u_{COST}(i) = u_{QoE}(i) - \lambda_s \cdot r_s(i)$.

In all, in order to solve the problem of UAV D's access selection in heterogeneous networks, we need to find a strategy to maximize the cumulative reward function by using the MAB model

$$\begin{aligned} \mathbf{S}_{\mathbf{MAB}} &= \arg\max_{s(i)} \sum_{k=1}^{i} \{K\{\phi, \mathrm{s}(k), r_s(k), l_s(k)\} - \lambda_s \cdot r_s(k)\} \\ s.t.\ \ & i \in \{1,2,\ldots,T\}, s(i) \in \mathcal{M}, \phi \in \{1,2,3\} \end{aligned} \tag{3}$$

When using regret value to measure the performance, the expected reward of each network can be expressed as $\boldsymbol{\mu} = K\{\overline{\Re}, \overline{\Gamma}\} - \lambda \cdot \overline{\Re}$, and the regret value is $R_{\mathbf{S}_{\mathbf{MAB}}}(i) = i \cdot \max \boldsymbol{\mu} - E[\sum_{k=1}^{i} u_{MAB}(k)]$.

4. Solution Algorithms

In Section 2, we have modeled the network selection as the MAB problem. Then, how to balance the utilization and exploration of the MAB becomes very crucial. We adopt three algorithms to solve the problem.

4.1. ε-Greedy Algorithm[6]

ε-Greedy Algorithm first sets a smaller probability value $\varepsilon \in (0,1)$. As the decision maker, UAV D will choose the arm with the largest expected reward with probability $1-\varepsilon$, and the other $M-1$ arms with small probability ε. Suppose that in slot i, the choice made by UAV D is network v. Then in slot $i+1$, the probability that D still takes the access network v as the selection arm is $p_v(i+1) = \begin{cases} 1-\varepsilon+\varepsilon/M & if \quad v = \arg\max \mathcal{Q} \\ \varepsilon/M & otherwise \end{cases}$. Suppose the average sampling value of the reward is $\overline{\mathcal{Q}} = \{\overline{q(1)}, \overline{q(2)}, \ldots, \overline{q(M)}\}$, then

$$s_{MAB}(i) = \arg\max \overline{\mathcal{Q}} \tag{4}$$

4.2. UCB Algorithm[7]

The basic idea of UCB Algorithm is to maintain confidence intervals of various expected rewards by using appropriate centralized inequalities. With the increasing number of samples, the confidence interval of sample expectations becomes more and more tight, so that the real average reward can be estimated more accurately. Let $\mathcal{A} = \{A(1), A(2), \ldots, A(M)\}$ represent a cumulative value of the number of times each arm has been selected, then

$$s_{MAB}(i) = \arg\max_{s \in \mathcal{M}} [\overline{\mathcal{Q}} + \sqrt{\frac{\alpha \cdot \ln i}{2\mathcal{A}}}] \tag{5}$$

4.3. SoftMax Algorithm[8]

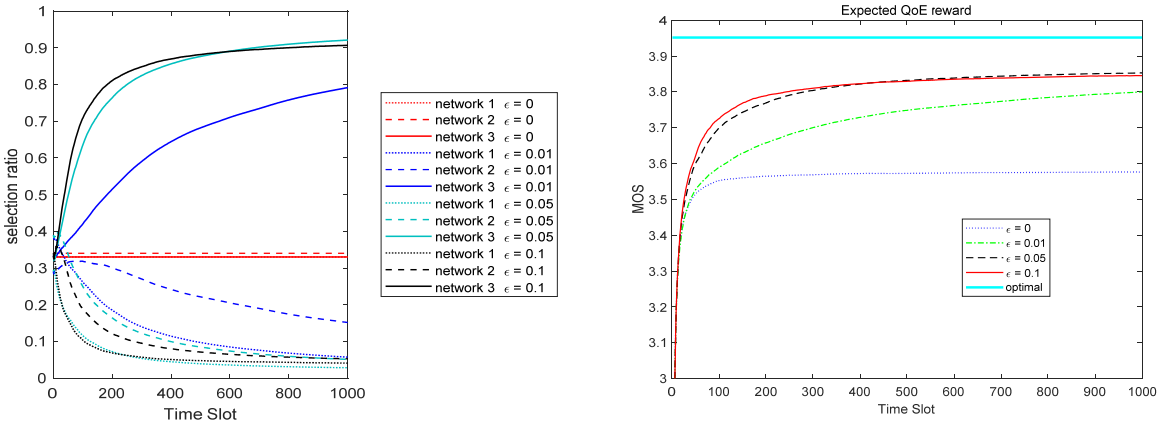
In SoftMax Algorithm, the probability of an arm being selected is related to both its own expected reward and all the expected reward of other arms. This probability is the ratio of the two. The improved SoftMax Algorithm based on Boltzmann exploration can be expressed as

$$s_{MAB}(i) = \arg\max_{v \in \mathcal{M}} \left[\frac{e^{\overline{q(v)}/\tau}}{\sum_{k=1}^{M} e^{\overline{q(k)}/\tau}} \right] \tag{6}$$

5. Simulations and Analyses

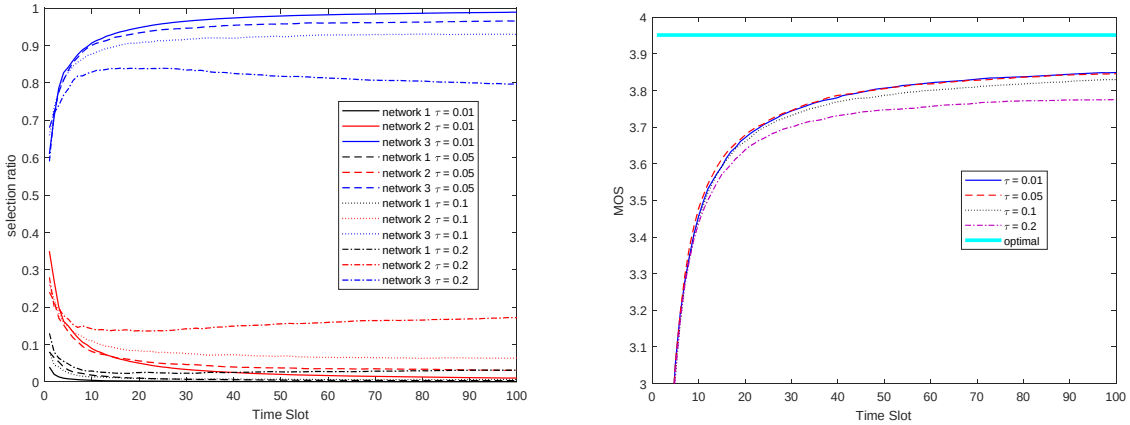
We conduct some simulations to evaluate the mechanism. The number of arms is set 3, which means there are three networks for selection. Firstly, we assume that the video transmitted by UAV D has fast moving speed, and it belongs to RM class. We compare the three solving algorithms under different parameters in figure 2. As the number of iterations increases, all curves tend to be smooth. This means all of the three algorithms will make the model converge to a stable state. 'Selection ratio' means the probability a network would be selected. From figure 2(a), 2(c) and 2(e) we can find SoftMax Algorithm has the fastest convergence rates. All of its choices are closer to zero or one, which means we could make a clear judgement of selection. The reward values of MOS in figure 2(b), 2(d) and 2(f) have similar trend.

The Performance of ε-Greedy Algorithm, under Different ε



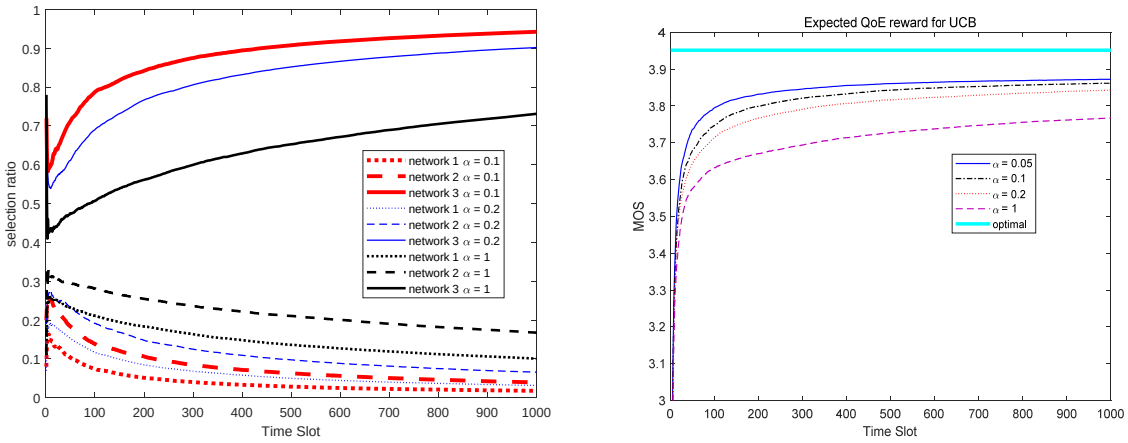
(a) (b)

The Performance of SoftMax Algorithm, under Different τ



(c) (d)

The Performance of UCB Algorithm, under Different α



(e) (f)

Figure 2. The Results of Three Algorithms under Different Parameters

Then, we set $\varepsilon = \tau = \alpha = 0.05$ and analyze the performance of different algorithms when three different types of video, SM, GW and RM, appear randomly. Five algorithms are compared, including the ideal one which brings expection MOS and the no learning one which is completely random selection. The MOS results are shown in figure 3. We can find that the results of SoftMax Algorithm is most close to the ideal one, while the worst one is no learning algorithm.

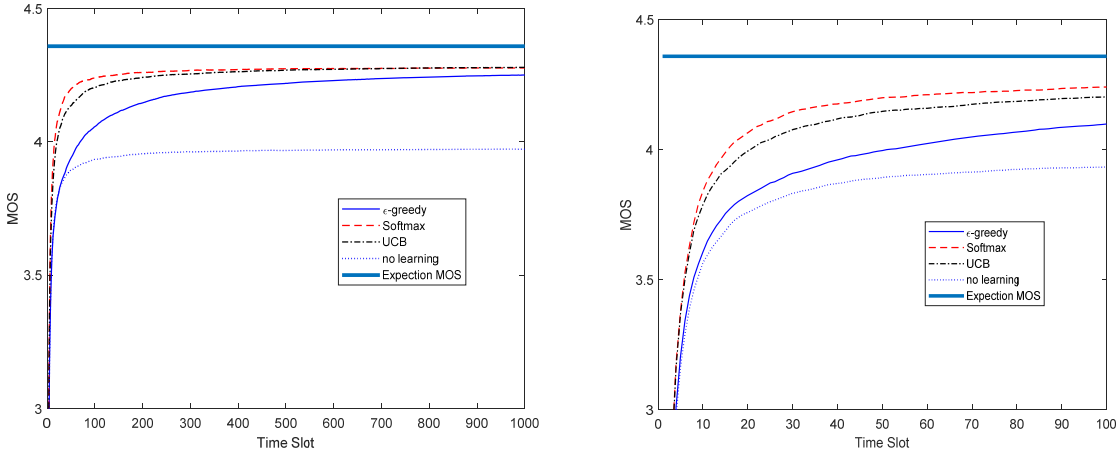


Figure 3. Comparison of Different Algorithms when Video Types are Random

6. Conclusion

By taking the video content classification and unknown CSI into consideration, this paper proposes an access network selection mechanism oriented to video QoE. The sequential decision-making problem is described as the MAB model. And three different algorithms are adopted to effectively solve it. The simulation results show that the SoftMax Algorithm outperform other solving method, and the proposed mechanism could effectively solve the problem of access network selection of UAV video transmission in heterogeneous networks.

7. Acknowledgments

This work was funded by the Project of Natural Science Foundations of China (No. 91738201 and 61401507) and China Postdoctoral Science Foundation (No.2017M613403).

8. References

[1] Robbins H. Some Aspects of the Sequential Design of Experiments[M]//Herbert Robbins Selected Papers: Springer New York, 1985:169–177.
[2] M. Bande and V. V. Veeravalli, "Multi-User Multi-Armed Bandits for Uncoordinated Spectrum Access," 2019 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, 2019, pp. 653-657.
[3] S. C. Pakhrin and D. R. Pant, "Multi-Armed Bandit Learning Approach with Entropy Measures for Effective Heterogeneous Networks Handover Scheme," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018, pp. 451-455.
[4] Khan A, Sun L, Ifeachor E. Content Clustering Based Video Quality Prediction Model for MPEG4 Video Streaming over Wireless Networks[C]//2009 IEEE International Conference on Communications, 2009:1–5.
[5] Z. Deng, Y. Liu, J. Liu, X. Zhou and S. Ci, "QoE-Oriented Rate Allocation for Multipath High-Definition Video Streaming Over Heterogeneous Wireless Access Networks," in *IEEE Systems Journal*, vol. 11, no. 4, pp. 2524-2535, Dec. 2017.
[6] Vermorel J, Mohri M. Multi-armed Bandit Algorithms and Empirical Evaluation[C]//European Conference on Machine Learning.[S.l.]: Springer-Verlag, 2005.
[7] K. Saito, A. Notsu, S. Ubukata and K. Honda, "Performance Investigation of UCB Policy in Q-learning," *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, 2015, pp. 777-780.
[8] Kuleshov V, Precup D. Algorithms for the multi-armed bandit problem[J]. Machine Learning Research, 2010, 4(10):1–48.