A Reinforcement Learning Approach to Call Admission Control in HAPS Communication System

Shu Yan Ni[1], Pan Feng He[1,2,a] and Nai Ping Cheng[1]

[1]*Equipment Academy, 101416 Beijing, China*
[2]*NanjingTelecommunicationTechnologyInstitute, 210007 Nanjing, China*

Abstract. The large changing of link capacity and number of users caused by the movement of both platform and users in communication system based on high altitude platform station (HAPS) will resulting in high dropping rate of handover and reduce resource utilization. In order to solve these problems, this paper proposes an adaptive call admission control strategy based on reinforcement learning approach. The goal of this strategy is to maximize long-term gains of system, with the introduction of cross-layer interaction and the service downgraded. In order to access different traffics adaptively, the access utility of handover traffics and new call traffics is designed in different state of communication system. Numerical simulation result shows that the proposed call admission control strategy can enhance bandwidth resource utilization and the performances of handover traffics.

1 Introduction

Wireless communication system based on HAPS is a new type of wireless communication system which is being studied in the world at present. Due to the influence of the stratospheric winds and the limitation of the position and attitude keeping technique, HAPS is usually in a quasi-stationary state, also known as perturbation. The perturbation of HAPS will cause the dynamic change of cells, resulting to the cell edge users to handover frequently, which brings more challenges to the call admission control strategy. The purpose of studying call admission control strategy is to ensure the lower new call blocking probability and handover dropping probability without degrading the performance of the system. Since the cellular coverage area can improve the system performance and increase the flexibility of the system[1], Li use location information of HAPS and users to make the overlapping area to ensure the handover strategy, which helps determine and block new calls that may cause a handover failure, resulting in a near-zero handover dropping rate[2]. These methods can adapt to dynamic change of platform position, but cannot deal with the situation of attitude change.

An call admission control for adaptive resource allocation is introduce in [3], which reduces the blocking rate of new call and the handover dropping rate by service degradation. However, since the fixed admission parameters are set for new and handover calls, it has poor adaptability for changes of system state. In [4], the system resource allocation is optimized to maximize the number of users to improve the system performance. Du to this method only deal with the current call request and system state, the adaptability of dynamic state is poor.

[a] Corresponding author : hepanfeng01@126.com

The problem of call admission control can be modelled as a Markov decision processes (MDPs) or Semi-Markov decision processes (SMDPs), which may adapt to the dynamic change of system state[5]. The methods such as policy iteration and dynamic programming can solve this kind of problem [6], but the current research assumes that the link capacity is constant. In fact, due to the presence of signal fading, interference, user mobility and using adaptive modulation and coding(AMC), the link capacity is time-varying in actual communication networks (such as World Interoperability for Microwave Access(WiMAM), Long Term Evolution(LTE) etc.)[7].The quasi-stationary state of HAPS will also cause the change of link capacity in HAPS communication system.

Although the above research can ensure the performance of handover in a certain extent, the dynamic changes of users in cell and link capacity caused by perturbation and mobility of users are considered inadequate, as well as different business requirements. In addition, it is difficult to obtain the exact model parameters of the MDPs. To solve the above problems, we introduced the reinforcement learning which does not need the exact model parameters to maximize the long-term benefits of system, aiming at deal with different traffics. By cross layer interaction and business downgrade ideas, the adaptive call admission control strategy based on reinforcement learning was put forward.

2 System model

Thornton pointed out that the change of carrier interference ratio from cell centre to edge is large, because of the power rolling down quickly at the edge zone[8]. So the adaptive modulation coding can get very good spectrum efficiency. Meanwhile call admission control strategy should not only consider the change of link capacity, but also guarantee the QoS of different traffics.

The multi-beam antennas from HAPS produce an elliptical beam that is projected onto the ground as a circular. For each deviation θ of the beam direction from HAPS, the beam directivity gain $D(\theta)$ can be approximated as

$$D(\theta) = G_{\max} \cdot \max[\cos^{n}\theta, \mathrm{S}_{f}] \tag{1}$$

Where $G_{\max}$ is directivity gain at the centre of the cell and n is the main lobe decay rate. S_{f} is sidelobes parameter. Assuming that the center coordinates of cell i is (x_i, y_i), the coordinates of user q in cell i is expressed by (x, y), as shown in Figure 1. The deviation $\theta_{i,q}$ from user q to the beam boresight of cell i can expressed as

$$\theta_{i,q} = \mathrm{a}\cos\left(\left(xx_i + yy_i + h^2\right) / \sqrt{x_i^2 + y_i^2 + h^2} / \sqrt{x^2 + y^2 + h^2}\right) \tag{2}$$

Where h is the height of HAPS. Due to the quasi-stationary state of the HAPS, the directivity gain of the user changes dynamically in the cell even under the condition of user's still, which will cause the change of signal to interference plus noise ratio(*SINR*) . In addition, the wireless channel fading characteristics also cause changes of *SINR*.

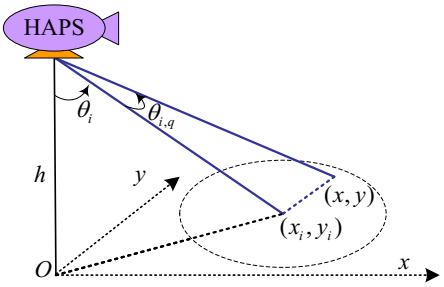


Figure 1. the Relationship between Users Position and the Beam Boresight.

Assuming that the radio channel of the multi-user OFDMA communication system from HAPS is flat fading and the channel conditions are stable and independent in each call admission control period. The subcarriers used in the cell are allocated equal power. The instantaneous *SINR* for the subcarriers *n* in cell *i* can be expressed as

$$SINR_{i,q}^{n}=\frac{P_i^n D(\theta_{i,q})\,|\,H_{i,q}^{n}\,|^2}{N_q^n+\sum_{j=1,j\neq i}^{n_{cc}} P_j^n D(\theta_{j,q})\,|\,H_{j,q}^{n}\,|^2} \tag{3}$$

Where P_i^n is the allocated transmit power of the subcarrier *n* in the cell *i*, N_q^n and n_{cc} are the Gaussian white noise and the number of all cells in same frequency. $H_{i,q}^{n}$ is channel response of the user *q* on the subcarrier *n* of the cell *i*, which generally include path loss, small scale fading and large scale fading. $D(\theta_{i,q})$ is the beam gain of the user q in the cell i, and $D(\theta_{j,q})$ is interference gain of the user q from the interference cell j. $\theta_{j,q}$ and $\theta_{i,q}$ can be calculated by (2).

For the AMC mechanism based on the Multiple Quadrature Amplitude Modulation(MQAM), the instantaneous bit rate of the carrier *n* in the cell *i* can be expressed as

$$R_{i,q}^{n}=B/N\cdot\log_2(1+SINR_{i,q}^{n}/\Gamma) \tag{4}$$

Where $\Gamma=-2/3\cdot\ln(5\,\mathrm{BER})$ is the signal to noise ratio difference between the spectral efficiency of the AMC mechanism based on MQAM and the Shannon capacity[9]. *B* and *N* are the channel bandwidth and the number of subcarrier of the cell.

According to the changing state of the channel environment, the instantaneous bit rate of OFDM subcarriers can be dynamically changed. So the link capacity can be updated based on the capacity of the subcarrier instantaneous state. It is assumed that a single subcarrier can carry the 1bit information at its minimum, its basic transmission capability is R_B bps. If the subcarrier can carry M bit information at a certain time, then the transmission capacity is MR_B bps. For the subcarrier which is not allocated, it has the basic capacity. Therefore, the total reachable link capacity ξ of the cell *i* can be expressed by

$$\xi=\sum_{N_{occ}+1}^{N} R_{\mathrm{B}}+\sum_{n=1+1}^{N_{occ}} R_{\mathrm{B}}\sum_{q=1}^{Q}\left\lfloor R_{i,q}^{n}/R_{\mathrm{B}}\right\rfloor\cdot R_{\mathrm{B}} \tag{5}$$

where N_{occ} is the number of subcarriers occupied. Assuming that the link reachable capacity has a total of L states and the link capacity is represented by ξ_l, then $\xi_l<\xi_{l+1}, l=1,...,L-1$.

3 Call Admission Control scheme based on reinforcement learning

Reinforcement learning (RL) interacts with the environment to obtain optimal strategy by learning agents. Q learning a kind of model free reinforcement learning algorithm , mainly through MDPs modelling and iterative method to approach the optimal solution. The call admission control(CAC) problem is modelled as a discrete time MDPs, the following are the state space and action set, reward function design, realization process of Q learning and adaptive CAC.

3.1 system state space and action sets

The vector $\boldsymbol{x}=(x_1,x_2,...,x_k,...,x_K)$ represents the number of different traffic types in a cell. The number of occupied subcarriers in the cell at the time *t* is $n_c(t)$, and the link capacity of the cell at the time t is represented by $\xi(t)$. The state space $\boldsymbol{S}(t)$ of the cell at the time *t* can be expressed as $(\boldsymbol{x}(t),n_c(t),\xi(t))$, and can be divided into three kinds according to the link capacity and the number of occupied subcarriers.

$$\boldsymbol{s}=\{(\boldsymbol{x}(t),n_c(t),\xi(t)\}=\begin{cases}\boldsymbol{S}_1, n_c(t)<N, \sum_{k=1}^{K} x_k rb_k \le \xi(t)\\ \boldsymbol{S}_2, n_c(t)=N, \sum_{k=1}^{K} x_k rb_k \le \xi(t)\\ \boldsymbol{S}_3, n_c(t)=N, \sum_{k=1}^{K} x_k rb_k \ge \xi(t)\end{cases} \tag{6}$$

Where rb_k and x_k are the average rate and number of the traffic type k. There are available subcarriers can allocate for the new call request or handover request in state $\boldsymbol{S}_1$. Although there is no available subcarrier in state $\boldsymbol{S}_2$, the sum of link rate of all traffics is less than link capacity of the system, which means that it is possible to allow call request by redistributing subcarriers to release some subcarriers. Since the sum of link rate exceeds the link capacity in state $\boldsymbol{S}_3$, some traffic must be degraded or interrupted. No matter what state the cell is in, when call request of the traffic k arrives, the system must choose to reject or accept the request. Then the action sets can be expressed by

$$\boldsymbol{A}(\boldsymbol{s})=\{\boldsymbol{a}(\boldsymbol{s})=(a(\mathbf{s},1),...,a(\boldsymbol{s},K)) \mid a(\boldsymbol{s},k)\in\{0,1\}\}, \forall \boldsymbol{s}\in \boldsymbol{S}_1\cup\boldsymbol{S}_2\cup\boldsymbol{S}_3 \tag{7}$$

3.2 Reward Function

The goal of maximizing the reward of the system can be expressed as the sum of the utility functions of traffics in service, we consider four types of traffics, voice, fixed data, multi-media and best effort traffic. The priority of these four traffics are $\beta_k(k=1,...,4)$, and the priority of handover traffic is β_0.

The voice and fixed rate data traffic are belong to constant bit rate(CBR) with strict guaranteed bit rate(GBR) limits, and the quivalent bandwidth rc_k must reach a certain value. While the requirement of multimedia traffic for bit error rate and delay is not high, which belongs to the GBR service with minimum rate guarantee. So the equivalent bandwidth is in a certain range[$rc_{\min}, rc_{\max}$]. The best effort traffic generally does not have constant bandwidth requirements, and the equivalent bandwidth is represented by rc_e .Since QoS requirements for different are measured by the equivalent bandwidth requirements of the traffics, the utility function of new call request with different traffics can be

$$U_r=\begin{cases}\beta_k\cdot rc_k\varepsilon(r-rc_k) & ,k=1,2\\ \beta_k\cdot\sin\left(\frac{\pi}{2}\cdot\frac{r-rc_{\min}}{rc_{\max}-rc_{\min}}\right)[\varepsilon(r-rc_{\min})-\varepsilon(r-rc_{\max})]+rc_e\cdot\delta(r-rc_{\max}) & ,k=3\\ \beta_k\cdot r/rc_e[\varepsilon(r)-\varepsilon(r-rc_e)]+rc_e\cdot\delta(r-rc_e) & ,k=4\end{cases} \tag{8}$$

Where $\varepsilon(\cdot)$ is the utility function of handover request with different traffics can be $U_{rh}=\beta_0\cdot U_r$. When the state is $\boldsymbol{S}_1$, there are idle subcarriers. So allowing access to the traffic can increase the overall utility, and the reward $r(\boldsymbol{s},\boldsymbol{a})$ equals to $a(\boldsymbol{s},k)\cdot U_r(k)$ for the traffic k request. If the average blocking ratio $\bar{f}_d$ is zero or blocking ratio for the traffic k request $f_d(k)$ is zero, the reward is also $a(\boldsymbol{s},k)\cdot U_r(k)$ .If the value of $\bar{f}_d\cdot f_d(k)$ is greater than 0 in state $\boldsymbol{S}_2$,the reward value is $a(\boldsymbol{s},k)\cdot\gamma_k f_d(k)/\bar{f}_d\cdot U_r(k)$ to access more traffics k than others (γ_k is the blocking probability weighting factor of traffic k). While the state is $\boldsymbol{S}_3$, the reward value includes both the access revenue and the inherent loss, which can be expressed by $a(\boldsymbol{s},k)\cdot(\gamma_k f_d(k)/\bar{f}_d-1)\cdot U_r(k)-\bar{f}_d\bar{U}_r$. Therefore the designed reward function $r(\boldsymbol{s},\boldsymbol{a})$ is

$$r(\boldsymbol{s},\boldsymbol{a})=\begin{cases} a(\boldsymbol{s},k)\cdot U_r(k)\ , \quad \forall \boldsymbol{s}\in \boldsymbol{S}_1 or\left(\overline{f}_d\cdot f_d(k)=0,\forall \boldsymbol{s}\in \boldsymbol{S}_2\right),k=1,...,K \\ a(\boldsymbol{s},k)\cdot \gamma_k f_d(k)/\overline{f}_d\cdot U_r(k),\ \ \overline{f}_d\cdot f_d(k)>0,\forall \boldsymbol{s}\in \boldsymbol{S}_2,k=1,...,K \\ a(\boldsymbol{s},k)\cdot(\gamma_k f_d(k)/\overline{f}_d-1)\cdot U_r(k)-\overline{f}_d\overline{U}_r,\ \ \forall \boldsymbol{s}\in \boldsymbol{S}_3,k=1,...,K \end{cases} \tag{9}$$

3.3 implementation process of reinforcement learning

We use the Q value iteration method to achieve reinforcement learning in this paper. The iterative formula is[10]

$$Q_{t+1}(\boldsymbol{s},\boldsymbol{a})=(1-\alpha)Q_t(\boldsymbol{s},\boldsymbol{a})+\alpha\left[r_t(\boldsymbol{s},\boldsymbol{a})+\gamma\max_{\boldsymbol{a}'\in A}Q_t(\boldsymbol{s}',\boldsymbol{a}')\right] \tag{10}$$

where $\alpha\in[0,1)$ is the learning rate and $\gamma\in[0,1)$ is the discount factor. α can be expressed by $\alpha=1/(1+T(\boldsymbol{s},\boldsymbol{a}))$ ($T(\boldsymbol{s},\boldsymbol{a})$ represents the number of visits on state-action). Obviously, when α tends to 0, $Q_t(\boldsymbol{s},\boldsymbol{a})$ converges to $Q_t^*(\boldsymbol{s},\boldsymbol{a})$. The optimal action set can be obtained by repeated learning and decision making. In order to balance the exploration and exploitation, the standard algorithm ε-greey is used to select the action of the controller.

(1) Parameter initialization: make $Q_0(\boldsymbol{s},\boldsymbol{a})$ equal to $r(\boldsymbol{s},\boldsymbol{a})$ and initialize α , γ and the exploration rate ε_0 ;

(2) Obtain the current state: When a new call arrives or a call leaves the cell, the controller collects the number of traffic in the current cell, the number of subcarriers occupied, and the current cell link capacity;

(3) Select action: the random decision is made to access or reject the arrival request for the exploration stage. While the action of the current state maximum rewards value is selected to decide whether to access or reject the arrival request for exploitation stage.

(4) Update the value $Q(\boldsymbol{s},\boldsymbol{a})$: Caculate $Q(\boldsymbol{s},\boldsymbol{a})$ according to the reward function $r(\boldsymbol{s},\boldsymbol{a})$ of the current state-action-pair ;

(5) Update parameters: the parameter α and ε are gradually reduced by anti-scaling function after each iteration, and returned (2).

3.4 adaptive subcarrier allocation strategy

To ensure the long-term gains of the system, the effective subcarrier resource allocation strategy can ensure the QoS requirements and efficient resource utilization. According to the most stringent QoS requirements for CBR traffics, the service of these traffics should not downgrade and the AMC scheme is always successful implementation. While the AMC scheme of VBR and BE traffic is controlled by the adaptive subcarrier allocation strategy to accept more users and reduce the call blocking probability.

Suppose that the equivalent bandwidth of the user q for the traffic k corresponds to the number of subcarriers required is $n_{k,q}^r$, the number of allocated subcarriers is $n_{k,q}^a$. We use N_{occ} to express the total number of occupied subcarrier.

(1) $N_{\text{occ}}<\eta N$: Since the cell is in a state of light load, all users does not distinguish the traffic types. The allocated subcarriers equals to the need of this traffic, namely $n_{k,q}^a=n_{k,q}^r$.

(2) $N_{\text{occ}}\geq\eta N$: The cell should distinguish traffic type in the heavy load state. For the CBR traffics, the number of allocated subcarriers equals to the requirements ($n_{k,q}^a=n_{k,q}^r,k\in\{CBR\}$). While the number of the allocated subcarriers for VBR or BE traffic is discounted on the number of requests

($n_{k,q}^{a}=\gamma_{k,q}n_{k,q}^{r}, k\in\{VBR,BE\}$). The discount factor $\gamma_{k,q}$ is associated with the current load state, the link capacity, the new call blocking rate and the handover dropping rate in the cell. Assuming the minimum number of subcarriers allocated is $n_{k,q}^{\min}$ for VBR and BE traffic, the number of subcarriers that can be released by the service degradation is:

$$N_{\text{release}}=\sum_{k=3}^{K}\sum_{q=1}^{x_k}(n_{k,q}^{a}-n_{k,q}^{\min}) \tag{11}$$

According to the relationship between $N-N_{\text{occ}}-N_{rel}$ and n_k^r, there are two cases:

1) $n_k^r \le N-N_{\text{occ}}-N_{rel}$: the traffic k can be accessed through service degradation in the cell. In order to In order to access serve more traffics, the cell will allow access to the traffic k.

2) $n_k^r > N-N_{\text{occ}}-N_{rel}$: Although service degradation can release some subcarriers, it is not enough to accept traffic k. For the purpose of reducing handover dropping rate of the GBR (including CBR and VBR), the system can force call termination of the BE traffic, which has the largest number of occupied subcarriers currently to release subcarriers. If the arrival traffic is belong to the GBR and the required subcarriers n_k^r is satisfied with the formula

$$n_k^r \ge \max_q\left\{n_{4,q}^{a}\right\},\ 1\le q\le x_4 \tag{12}$$

the system allows this handover access to the cell. Otherwise the access is refused.

4 simulation analysis

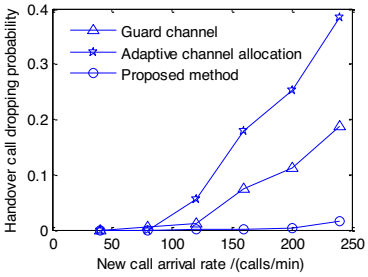
The simulation parameters of HAPS communication system are consistent with [11]. The mapping relationship between AMC and SINR used in this system can refer the air interface of IEEE802.16, which can be found in [12]. The simulation traffic parameters are shown in Table 1. It is assumed that the relationship between handover call arrival rate λ_{h} and new call arrival rate λ_{n} is $\lambda_{\text{h}}=\rho\lambda_{\text{n}}$, in which ρ is a random variable with mean and variance of 0.5.

Table 1. Traffic Parameters

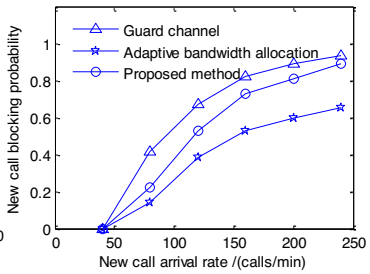
Traffic types	Voice	Fixed bit rate data	Multimedia	Best effort
Equivalent bandwidth/Kbps	64	56	1128-512	0-120
Call duration/s	100	120	400	250
Traffic priority weight	9	87	2	1
Percentage of all new calls	0.25	0.15	0.4	0.2

In order to verify the performance of the proposed algorithm, the strategy based on guard channel (guard channel with 10%), adaptive bandwidth allocation with handover priority strategy in [3] and the proposed algorithm are compared under the same conditions. We use five metrics to test the performance of the algorithms, which are handover call dropping probability, new call blocking probability, bandwidth utilization, forced call termination probability and Average system revenue. Forced call termination probability include the current dropped call resulting from Link capacity reduction or arrival GBR handover request.

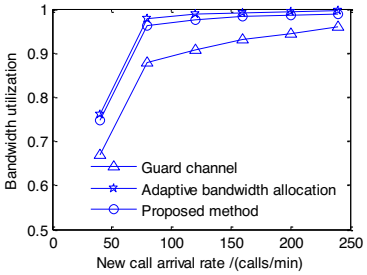
The performances of this three strategies are shown in figure 2. Compared with the guard channel strategy, the proposed method have better performance in handover dropping probability, new call blocking probability and bandwidth channel utilization, the price is slightly higher rate of call termination probability. Although adaptive bandwidth allocation strategy has lowest new call blocking probability and highest bandwidth utilization, the handover dropping probability rate and forced call termination probability are worst. In comparison, the performance of the proposed method is superior, especially handover dropping probability and bandwidth utilization.



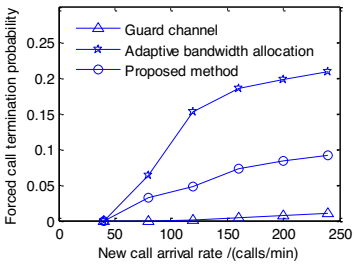
(a) Handover dropping probability



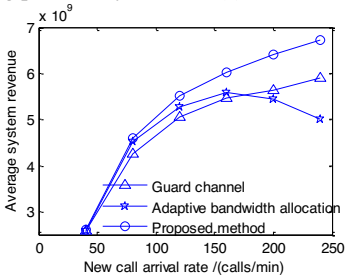
(b) New call blocking probability



(c) Bandwidth utilization



(d) Forced call termination probability



(e) Average system revenue

Figure 2. The performances of the three strategies

Figure 2 (e) shows that the proposed method always has the most average system revenue, The reason is that this method takes full account of the changing system state , the performance of the user and the system, in order to ensure long-term revenue.

5 conclusions

It has always been a hot issue in the research of wireless communication system that how to effectively access various types of traffics, and ensure the QoS requirements of various traffics services and improve the utilization of system resources. By introducing cross layer interaction and service downgrade, the proposed strategy considers the utility and the blocking rate of different traffics and maximizes long-term gains of system to balance the handover performance and system resource utilization. In the future work, we also need to consider the dynamic interference between cells to further improve the system performance.

References

1. D. Grace, C. Spillard, T.C. Tozer, *Wireless Personal Multimedia Communications Conference* (IEEE, New York, 2003)
2. S. FLi ,J. Wei, D. T. Ma, L. Wang, Journal on Communications. **32**, 131 (2011)
3. M. Z. Chowdhury, Y. M Jang, Z. J.Haas, Journal of Communications and Networks, **15**, 15 (2013)
4. A. Ibrahim, A. S. Alfa, IEEE Trans.Wireless Comm., **14,** 5823 (2015)
5. M. Panfili, A. Pietrabissa, International Journal of Control, **89**, 1428 (2016)
6. M. K. Luka ,A. A. Atayero, O. I. Oshin,*Sai Computing Conference* (IEEE, New York, 2016)
7. A. Pietrabissa. . European Journal of Control, **17** 89(2011)
8. J. Thornton, D. Grace, M. H. Capstick ,IEEE Trans.Wireless Comm., **2,** 484 (2003)
9. A. J. Goldsmith , S. G. Chua,IEEE Trans. Comm., **45**, 1218(1997)
10. D. Xiong, Y. Li, Journal on Communications, **36**, 252 (2015)
11. P. He, N. Cheng, S. Ni, *Wireless Communications & Signal Processing* (IEEE, New York, 2016)
12. P802.16Rev3/D5, Air Interface for Broadband Wireless Access System (IEEE, New York, 2012)