



Reinforcement Learning Based Signal Quality Aware Handover Scheme for LEO Satellite Communication Networks

Menting Chen¹(✉), Yong Zhang¹, YingLei Teng¹, Baoling Liu¹, and Lili Zhang²

¹ School of Electronic Engineering,
Beijing University of Posts and Telecommunications,
Beijing, People's Republic of China
{mengting.chen,yongzhang,lilytengt,bliliu}@bupt.edu.cn

² Shuangchuang Center (Laboratory),
State Grid Shengwang Location Based Service Co., Ltd.,
Beijing, People's Republic of China
lily3286000@126.com

Abstract. With the increasing of Low Earth Orbit (LEO) satellites emission, utilizing existing LEO satellite network systems have lower CAPEX/OPEX than deploying fixed terrestrial network systems in the remote area. Due to the high mobility of LEO satellites, mobility management mechanisms such as handover schemes are the key issue should be settled on leveraging the merits of LEO satellites telecommunications (e.g. lower propagation delay than GEO satellites). In traditional handover schemes, choosing which one is the next-hop satellite for one user is only determined by evaluating some specific criteria in the current state, not guaranteeing long-term and global optimization. To solve this problem, we use the cumulative signal quality that involves the remaining service time and signal quality and propose a Q-Learning based handover scheme. The simulation results show that the proposed scheme improves the overall signal quality and reduce the average handover number of users compared with other handover schemes.

Keywords: LEO satellite network · Handover · Q-Learning · Signal quality

1 Introduction

A satellite network system is a viable option to cover remote areas where there are not adequate available terrestrial communication networks [13], an integral part of the communication system to achieve truly ubiquitous coverage. The LEO satellite as a typical small satellite is becoming fashionable and changing the

This work is supported by National Key R&D Program of China No. 2018YFB1201500, National Natural Science Foundation of China under Grant No. 61771072, and Beijing Natural Science Foundation under granted No. L171011.

economics of space [18]. LEO satellites introduce lower end-to-end delay, require lower power and exert more efficient spectrum allocation than the geostationary orbit satellites, making them suitable for future personal telecommunication [16]. Even for fixed users, handover in non-geostationary satellite communication system will be continuous.

To improve the QoS of users, diverse handover schemes have been extensively developed in link-layer and network-layer. Satellite handover occurs in the link layer when the existing connection of one end user with the satellite transfers to the other satellite [6, 19]. Due to the high mobility of the LEO satellite system, the handover happens frequently in many circumstances, causing call interruption, thus influencing directly the quality of experience of the users.

For satellite handover schemes, most of the works only considered a specific handover criterion (e.g. remaining service time [8], number of free channels [7], elevation angle [10]) or highly relied on geometric information, lacking an overall solution [20]. It true that there are some schemes combined with two criteria. For instance, Zhao et al. take the user position and the signal strength into consideration to reduce the call termination probability [24]. Besides, Wu et al. firstly presented a graph-based handover framework for LEO satellite network [20], integrating all kinds of satellite networks into one topology graph, where the process for the end-user to switch between different serving satellite during its call period can be considered as to find a way among these consecutive covering satellites. However, the handover strategy of the proposed scheme is the same as that of the traditional largest service time scheme. Among these schemes, single handover criteria leave the user shortsighted and make the network is unable to meet the complex QoS requirements of users.

Reinforcement Learning (RL) method has been used in many problems of network communication [22]. All of the problems (e.g. base station on-off switching strategies [11], routing [1], adaptive tracking control [12] and power allocation [23]) can be formulated as Markov Decision Processes (MDPs). Specifically, Wang et al. first formulated the traffic variations as an MDP, then used actor-critic algorithm (a typical RL algorithm) to improve energy efficiency in radio access networks [11]. In [1], Al-Rawi et al. introduce that RL can address routing challenges by gathering information, learning, and making routing decision efficiently. Besides, Liu et al. propose an online RL method to achieve adaptive tracking control for nonlinear discrete-time MIMO systems [12], and Zhang et al. maximize the overall capacity of a multi-cell network by deep RL method [23].

In horizontal handover, where users switch in the same network layer such as satellite handover in one LEO satellite network system, the signal quality is a key criterion. Moreover, the remaining service time of satellites contributes a lot to satellite handover. The optimal trade-off between the two criteria needs to be achieved. For example, if a user always chooses the satellite of best signal quality but he is about to leave the satellite's coverage area, the user has to switch to another satellite soon. In this way, the ping pong switch occurs, which is intolerable in satellite networks due to a long propagation delay. Therefore,

how to lowest the ping pong switching rate by balance the signal quality and remaining service time is our most concerned.

To best of our knowledge, the satellite handover problem for optimal cumulative signal quality is untouched. Our contributions are summarized as follows: (1) we model the signal quality by Ornstein-Uhlenbeck process, then first propose a criterion that combines signal quality and remaining serving time; (2) we propose a handover scheme based on reinforcement learning method to maximize the overall signal quality and minimum the ping pong switching rate in the long term.

2 System Model

2.1 LEO Satellite Network Model

Consider an LEO satellite constellation with a specific topology structure, which operated in slotted time $t \in \{0, 1, 2, \dots, T\}$, containing N satellites and M users. The speed of the mobile terminals, which is much lower than the speed of the LEO satellite (about 25000 km/h relative to the earth rotation), can be ignored.

2.2 Satellite Handover Process

A satellite handover is divided into three steps: handover information collection, handover decision-making, and handover execution.

Handover Information Collection. Assume the user connected with LEO satellite can easily obtain its exact position by using the existing Global Position System (GPS) infrastructure (or using the other ways), and also the covering satellites in the future period by predicting the motion of them. The prediction can be made by the predicting method [2,3], or by a centralized controller in Software-Defined Network (SDN) and Network Function Visualization (NFV) architecture. Under such conditions, users can obtain the Received Signal Strength (RSS), remaining service time and user elevation of the connected satellites.

When the RSS of a user's connection with a satellite that may be switched is less than a certain hysteresis threshold RSS_{min} for a period of time, the satellite is placed into the user's candidate satellite sequence and the switching decision is triggered. This threshold needs to be set properly: if the hysteresis threshold is too large, it will lead to frequent switching, causing large switching delay; if it is too small, the satellite will be blocked due to stranded users' service.

Handover Decision-Making. In view of uneven user distribution restricted by the imbalance of regional development and the vast ocean area, certain restrictions should be made on the fairness of handover, i.e. the channel usage $q(t)$ at any time should be less than a threshold:

$$q_n(t) \leq \xi q_{max} \quad (1)$$

q_{max} is the channel number of the satellite, and ξ is the maximum satellite channel utilization rate.

The satellite for next hop from the candidate satellite sequence should be determined by the collected network information. Due to the limited satellite coverage area, the remaining service time directly affects the number of users' handover during a service period. For example, if the satellite time remaining for each hop is very short during a service period, users may be in the process of continuous handover and ping-pong handover is more likely to occur. Therefore, we should consider the compromise between channel quality and remaining time.

In the actual scenario, channel changes are more dynamically. We consider a mean-reverting Ornstein-Uhlenbeck process [4] as the instantaneous dynamics of non-stationary channel model, where a time-varying additive Gaussian channel model is given by $|h_{i,t}|^2$. Thus, the dynamics of the channel are given by:

$$d|h_{i,t}| = \theta(\mu_h - h_{i,t})dt + \sigma_h d\mathcal{B}_{i,t} \quad (2)$$

where $\mu_h > 0$, $\sigma_h > 0$, and $\mathcal{B}_{i,t}$ is a standard Brownian motion. μ_h refers to the geocentric angle α between users and satellites, which related to the elevation angle of users. Since the larger the elevation angle is, the smaller the geocentric angle is, μ_h can be defined as:

$$\mu_h = h_0 \cos \alpha - \cos \alpha_0 \quad (3)$$

where h_0 represents the channel gain, α_0 is the maximum geocentric angle in the coverage area of one satellite. Equation 3 means that the smaller the geocentric angle is, the better the channel quality is.

Handover Execution. Once the next satellite to be handed over is determined, the user will start the soft handover process, and many signaling exchanges will take place during the handover process. Each signaling exchange may fail. The handover failure rate of each user should be less than a threshold to avoid frequent handover:

$$1 - (1 - p)^{n_m} \leq \mu \quad (4)$$

where p is the failure probability of one handover, and n_m is the total number of user m 's switches in his communication period.

This possibility of failure is also related to the channel quality, which can be defined by SNR as shown in Eq. 5.

$$\gamma_{i,t} = \frac{p_0 |h_{i,t}|^2}{n_0} \quad (5)$$

Then, integrating SIR representing cumulative signal quality can be obtained:

$$\int \gamma_{i,t} dt = \frac{p_0}{n_0} \int |h_{i,t}|^2 dt \quad (6)$$

Therefore, the utility function of one handover (from satellite i to satellite j) can be defined as follows:

$$u_m(t) = \int_t^{t_j} \gamma_{j,t} dt \quad (7)$$

where t_j are the last coverage times of satellites j .

2.3 Problem Formulation

To sum up, considering the whole process of handover, the overall goal is determined as follows:

$$\begin{aligned} \min_{\Omega} \quad & \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{m=1}^M u_m(t) \right] \\ \text{s.t.} \quad & q_n(t) \leq \xi q_{max}, \forall n, n = 1, \dots, N \\ & 1 - (1 - p)^{n_m} \leq \mu, \forall m, m = 1, \dots, M \end{aligned} \quad (8)$$

This problem is a stochastic programming problem that is control dependent state evolution where the control action (handover) will influence the state process. Also, in the dynamic scenario of LEO satellite handover, the transition probability is unknown and the state space is infinite. In this way, the problem is NP-Hard, difficult to find the best solution, but we can try to make a policy to minimize the effective transmission rate of all users. Also, the handovers of users are independent as a Markov decision process. Thus, a model-free learning technique such as QL is needed to find a policy Ω to better adapt to the dynamic scenario and gain long-term control.

3 Reinforcement Learning Handover Scheme

3.1 Decision-Making Process in Q-Learning

QL is a technique of reinforcement learning where the agent learns to take actions to maximum the cumulative reward (Eq. 9) by trial-and-error interactions with its environment:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (9)$$

where $0 \leq \gamma \leq 1$ indicates the weight of the experience value and r is the numerical reward obtained at each optimization epoch which comprised by one iteration. Basically, QL has two steps: policy evaluation and policy improvement. It evaluates the policy by calculating value function (the expected value of the return at state \mathcal{S} accumulated by action \mathcal{S} in a limited period of time):

$$Q_{\pi} = \mathbb{E}_{\pi} [R_t | s_t = s, s_t = a], \quad (10)$$

it then improve the policy by updating the value in the action's corresponding index in the Q matrix. In this model-free approach, the premise of fully evaluating the policy value function is that each state can be accessed, so an

exploration strategy called $\epsilon - greedy$ is adopted to update Q value in step of policy improvement. For an iteration for the user m , we have:

$$Q_{t+1}(s, a) = (1 - \alpha)Q_t(s, a) + \alpha Q_t(s, a') \quad (11)$$

where the higher the learning rate $\alpha(0 \leq \alpha \leq 1)$ is, the less the previous training results will be retained.

3.2 Maximum Cumulative Signal Quality Handover Scheme (MCSQ)

Using the QL framework discussed above, the agent, state \mathcal{S} , action $\mathcal{A}(s)$ and reward $r : \mathcal{S} \times \mathcal{A}(s) \times \mathcal{S} \rightarrow \mathbb{R}$ can be designed as follows:

- **Agent:** the handover controller.
- **States:** $s = (s^{(1)}, s^{(2)}, \dots, s^{(4)}) \in \mathcal{S} := \hat{\mathcal{U}} \times \mathcal{I} \times \mathcal{J} \times \mathcal{H}$. $\hat{\mathcal{U}}$ denotes the discretized space of the user position in slot t , also following the initial consideration of QL, where the space of states is a set of discretized value. $\hat{\mathcal{U}}$ is defined as the area discretized by a grid g_1, g_2 on longitude and latitude correspondingly, which is represented by

$$\hat{\mathcal{U}} := \{([\![u_1/g_1]\!], [\![u_2/g_2]\!]) | (u_1, u_2) \in \mathcal{U}\} \quad (12)$$

where $[\cdot] : \mathbb{R} \rightarrow \mathbb{N}$ is the floor function. $\mathcal{I} = \{i_0, \dots, i_m\}$ represents the connecting satellites set, where i_m represents the number of user m 's connecting satellite. $\mathcal{J} = \{J_0, \dots, J_m\}$ indicates the set of adjacent satellites that can exert handover in next hop, where J_m represents the set of user m 's adjacent satellites. $\mathcal{H} := \{0, 1\}$ indicates whether in each decision epoch the handover is being processed or not.

- **Actions:** $a \in \mathcal{A}(s)$ is the number of satellite selected in state \mathcal{S} . The set of actions is defined as

$$\mathcal{A}(s) = \begin{cases} \mathcal{I}, & \text{if } \mathcal{H} = 0 \\ \mathcal{J}', & \text{Otherwise} \end{cases} \quad (13)$$

which means that the user only can connect with one of the satellites when the handover is not processed, and otherwise the user can choose an adjacent satellite from the set $\{s^{(3)}\}$ in the time slot of handover. \mathcal{J}' represents the selected satellites.

- **Reward:** $r(s', a, s)$ is defined as the cumulative signal quality of one link defined before, which is influenced by the RSS, the remaining service time:

$$r(s', a, s) = \begin{cases} u_m(t), & \text{if } \mathcal{H} = 1 \\ 0, & \text{Otherwise} \end{cases} \quad (14)$$

where $r : \mathbb{R} \rightarrow \mathbb{R}$ is for controller to choose the optimal handover satellites of maximum cumulative quality of users received signal.

After many iterations, the Q matrix tends to converge and be stable, since almost every state is accessed at least once.

4 Simulation Results

4.1 Simulation Setup

The simulation referring to the Globalstar system runs using python to conduct the comparison. Satellites and users are randomly sprinkled in the Region \mathcal{U} by a Poisson Point Process (PPP). The maximum geocentric degree α_0 of satellite coverage is determined by the height of satellite orbit and the user-visible elevation angle α_e :

$$\alpha_0 = \sin(\arccos(\frac{R_e}{R_e + h})) \cos \alpha_e - \alpha_e \quad (15)$$

Consider users moving following a straight certain route, and assume the relative velocity between users and satellites is 60 km/s for better simulation effect (about 6 km/s in reality).

Table 1 summarizes the simulation parameters of the network model.

Table 1. Simulation parameters.

Parameter	Value
Height of the satellite orbit (h)	1414 km
User visible elevation angle (α_e)	10°
Radius of satellite coverage (R_0)	2670 km
Radius of the Earth (R_e)	6378 km
Relative velocity	60 km/s
Duration of user service period	120 s
Region \mathcal{U}	6000 km × 6000 km
Number of satellites	25
Satellite load threshold (ξp_{max})	5
Failure probability of one handover (p)	99%
RSS threshold (RSS_{min})	0.5×10^{-8} w
Hysteresis threshold	3 s
Transmit power of satellite (p_0)	50 w
Channel gain (h_0)	-1.85 dB
θ in the dynamics of channel	1
σ_h in the dynamics of channel	2.5×10^{-5}
Learning rate (α)	0.4
Greedy policy (ϵ)	0.6
Experience rate (γ)	0.9
Granularity of the location (λ)	15
Number of iterations	120

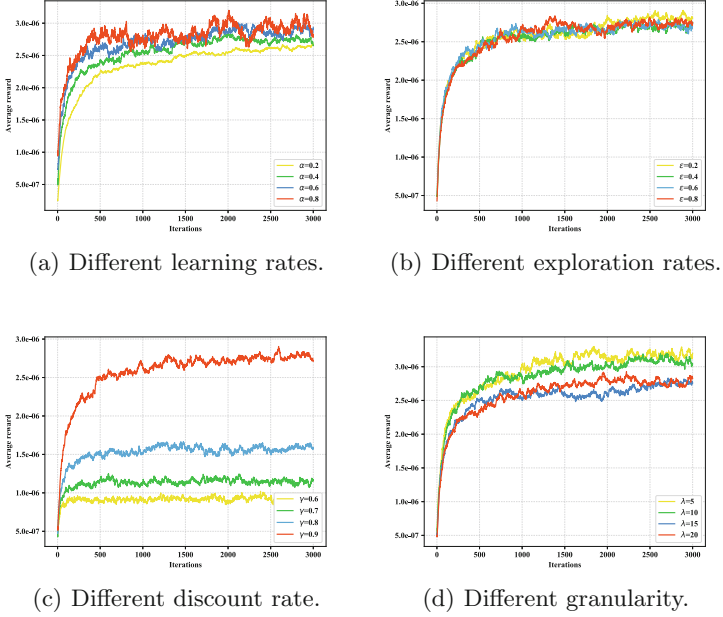


Fig. 1. Performance comparison of parameters in the MCSQ scheme.

4.2 Simulation Results

Figure 1 illustrates the convergence property and the parameter evaluation of the MCSQ algorithm. According to the impact of the learning rate can be shown in Fig. 1(a), we choose $\alpha = 0.4$ as our learning rate of all QL methods because the average reward performs more stable. Similarly, we choose $\epsilon = 0.6$ as the exploration rate because of the performance comparison in Fig. 1(b) and $\gamma = 0.9$ as the discount rate according to Fig. 1(c). The granularity of the location λ in the sate \mathcal{U} is determined by grid g_1, g_2 . Suppose $g_1 = g_2 = 2000/\lambda$, so the impact of λ is shown in Fig. 1(d). It can be seen that when $\lambda = 5$, average reward remains at a higher level because of more user in one grid region and cumulated Q in the same position of Q table. In consideration of the area of the grid region, we choose $\lambda = 15$ as the granularity of location.

It is worth noting that, in all simulations, the signal quality will be generated randomly according to Eq. 2 in every iteration, and the initial location of satellites and users and relative movement relationship will remain the same in one epoch.

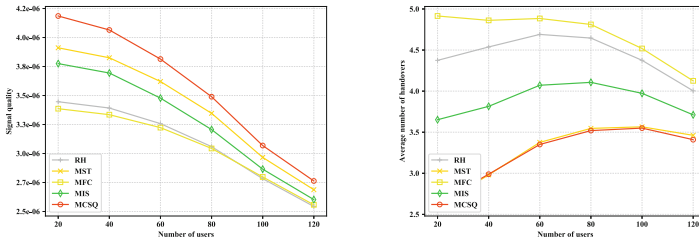
Although the Q table will be stable after about over thousands of iterations, the complexity of the proposed algorithm is low because this paper is dedicated to giving an optimal handover strategy for an initial status of satellites and users. Once we obtain the optimal policy Ω , it acts as a look-up table. Besides, the dynamic of Earth orbit is periodic, several time periods of satellite running can construct a cycle, wherein the location of satellites will be the same as that in

the same slot of the previous cycle. Therefore, the table can be reused and be calculated very few times.

Here, cumulative signal quality in the service time period of all users is represented by the average cumulative RSS. It only computes the RSS from the serving satellites, which means that if a user m has switched from satellite i to satellite j , his total cumulative RSS is cumulative RSS in a serving period of satellite i plus that in the serving period of satellite j . Therefore, the total cumulative of all users is as follows.

$$c = \mathbb{E}_M \left[\sum_{m=0}^M \sum_{n=0}^N \int_{t_{nms}}^{t_{nme}} \gamma_{n,t} dt \right] \quad (16)$$

where the t_{nms} is the start time of a serving time period of satellite n for user m , and t_{nme} is the end time respectively.



(a) Signal quality comparison. (b) Average handover number comparison.

Fig. 2. Performance comparison of handover schemes.

Figure 2 shows the performance comparison of the proposed algorithm and all baseline algorithms. In this comparison, 100 independent tests are conducted, where the topology of satellites and users is renewed in every test. The baseline algorithms are several traditional schemes:

- **Maximum number of free channels scheme (MFC)** [8]: In this scheme, users will choose the satellite that reserves most free channels. In this way, it tends to achieve the highest fairness index between satellites and uniform distribution of the telecommunication traffic in the LEO network [14]. Therefore, it will avoid overload satellites, but for users, they should find the satellite with more free channels regardless of location or signal quality, leading to more number of handovers.
- **Maximum service time scheme (MST)** [7]: According to this criterion, the user will be served by the satellite that offers the maximum remaining service time [14]. It will get the lowest number of handovers and ping-pong

switching rate, but the user will be reluctant to change serving satellite even if another satellite may offer more free channels. In an LEO satellite network with a certain constellation, the long coverage time means that in this serving period, the user will get a bigger maximum communication elevation angle [17], so the overall signal quality will be higher than other baseline algorithms.

- **Maximum instantaneous signal strength scheme (MIS)** [10]: This criterion is almost equivalent to the maximum elevation angle, because, for satellite communication, small elevation angle between the mobile terminal and the satellite leads to frequent shadowing and blockage events for the signal due to trees, buildings, hills, etc. [5]. Thus, it will avoid link failures and get relative high signal quality, but due to the dynamic characteristics of the channel, it is only able to guarantee the instantaneous signal quality. Once the channel changes rapidly, the handover decision will be unstable.
- **Random handover scheme (RH)**: According to this criterion, the user will randomly choose an available satellite in the slot of handover decision-making.

It can be seen that MCSQ, after iterations, gets the greatest signal quality and obtains almost the same low average handover number as MST. Since the cumulative signal quality combines the effect of elevation angle and remaining serving time, according to the criterion, the MCSQ combines the advantages of MIS scheme and MST scheme in data representation. Besides, the reinforcement learning framework makes the proposed scheme more telescopic in handover decision, so the MCSQ achieves the overall stable improvement in the whole satellite cycle.

5 Conclusion

In this paper, we have investigated a handover problem for LEO satellite networks, and we further formulated it as a stochastic optimization problem to get optimal signal quality of users. To solve this problem, we model the channel by the O-U process and introduce a criterion that combines signal quality and remaining serving time. Then, we propose a reinforcement learning handover strategy, which has shown that it has a good effect on improving the overall signal quality between satellites and users in a period, also greatly reduces the average handover number of users. In reality, the distribution of signal received cannot be predicted by a simple model, but the Q table after training could be a good reference when users are choosing the next hop satellites. Therefore, it can be generalized into a different network model easily, ensuring the flexibility of the algorithm. In further work, some novel reinforcement learning methods can be applied in this network model and handover procedure to improve the efficiency of this scheme.

References

1. Al-Rawi, H.A., Ng, M.A., Yau, K.L.A.: Application of reinforcement learning to routing in distributed wireless networks: a review. *Artif. Intell. Rev.* **43**(3), 381–416 (2015)
2. Ali, I., Al-Dhahir, N., Hershey, J.E.: Predicting the visibility of LEO satellites. *IEEE Trans. Aerosp. Electron. Syst.* **35**(4), 1183–1190 (1999)
3. Bertiger, B.R., Leopold, R.J., Peterson, K.M.: Method of predicting cell-to-cell hand-offs for a satellite cellular communications system, uS Patent 5,161,248, 3 Nov 1992
4. Charalambous, C.D., Menemenlis, N.: Stochastic models for short-term multipath fading channels: chi-square and Ornstein-Uhlenbeck processes. In: *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No. 99CH36304)*, vol. 5, pp. 4959–4964. IEEE (1999)
5. Chini, P., Giambene, G., Kota, S.: A survey on mobile satellite systems. *Int. J. Satell. Commun. Netw.* **28**(1), 29–57 (2010)
6. Chowdhury, P.K., Atiquzzaman, M., Ivancic, W.: Handover schemes in satellite networks: state-of-the-art and future research directions. *IEEE Commun. Surv. Tutorials* **8**(4), 2–14 (2006)
7. Del Re, E., Fantacci, R., Giambene, G.: Efficient dynamic channel allocation techniques with handover queuing for mobile satellite networks. *IEEE J. Sel. Areas Commun.* **13**(2), 397–405 (1995)
8. Del Re, E., Fantacci, R., Giambene, G.: Handover queuing strategies with dynamic and fixed channel allocation techniques in low earth orbit mobile satellite systems. *IEEE Trans. Commun.* **47**(1), 89–102 (1999)
9. Duan, C., Feng, J., Chang, H., Song, B., Xu, Z.: A novel handover control strategy combined with multi-hop routing in LEO satellite networks. In: *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 845–851. IEEE (2018)
10. Gkizeli, M., Tafazolli, R., Evans, B.: Modeling handover in mobile satellite diversity based systems. In: *Proceedings of the IEEE 54th Vehicular Technology Conference on VTC Fall 2001, (Cat. No. 01CH37211)*, vol. 1, pp. 131–135. IEEE (2001)
11. Li, R., Zhao, Z., Chen, X., Palicot, J., Zhang, H.: TACT: a transfer actor-critic learning framework for energy saving in cellular radio access networks. *IEEE Trans. Wireless Commun.* **13**(4), 2000–2011 (2014)
12. Liu, Y.J., Tang, L., Tong, S., Chen, C.P., Li, D.J.: Reinforcement learning design-based adaptive tracking control with less learning parameters for nonlinear discrete-time mimo systems. *IEEE Trans. Neural Netw. Learn. Syst.* **26**(1), 165–176 (2015)
13. Moura, J., Edwards, C.: Future trends and challenges for mobile and convergent networks. arXiv preprint [arXiv:1601.06202](https://arxiv.org/abs/1601.06202) (2016)
14. Papapetrou, E., Karapantazis, S., Dimitriadis, G., Pavlidou, F.N.: Satellite handover techniques for LEO networks. *Int. J. Satell. Commun. Netw.* **22**(2), 231–245 (2004)
15. Papapetrou, E., Pavlidou, F.N.: QoS handover management in LEO/MEO satellite systems. *Wireless Pers. Commun.* **24**(2), 189–204 (2003)
16. Sadek, M., Aissa, S.: Personal satellite communication: technologies and challenges. *IEEE Wirel. Commun.* **19**(6), 28–35 (2012)
17. Seyedi, Y., Safavi, S.M.: On the analysis of random coverage time in mobile LEO satellite communications. *IEEE Commun. Lett.* **16**(5), 612–615 (2012)

18. Sweeting, M.N.: Modern small satellites-changing the economics of space. *Proc. IEEE* **106**(3), 343–361 (2018)
19. Taleb, T., Hadjadj-Aoul, Y., Ahmed, T.: Challenges, opportunities, and solutions for converged satellite and terrestrial networks. *IEEE Wirel. Commun.* **18**(1), 46–52 (2011)
20. Wu, Z., Jin, F., Luo, J., Fu, Y., Shan, J., Hu, G.: A graph-based satellite handover framework for LEO satellite communication networks. *IEEE Commun. Lett.* **20**(8), 1547–1550 (2016)
21. Yue, P.C., Qu, H., Zhao, J.H., Wang, M., Wang, K., Liu, X.: An inter satellite link handover management scheme based on link remaining time. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC), pp. 1799–1803. IEEE (2016)
22. Zhang, C., Patras, P., Haddadi, H.: Deep learning in mobile and wireless networking: a survey. arXiv preprint [arXiv:1803.04311](https://arxiv.org/abs/1803.04311) (2018)
23. Zhang, Y., Kang, C., Ma, T., Teng, Y., Guo, D.: Power allocation in multi-cell networks using deep reinforcement learning. In: 2018 Technology Conference (VTC-Fall). IEEE (2018)
24. Zhao, W., Tafazolli, R., Evans, B.: Combined handover algorithm for dynamic satellite constellations. *Electron. Lett.* **32**(7), 622–624 (1996)