Research Article

Bo Ren*, Zhicheng Zhu, Fan Yang, Tao Wu, and Hui Yuan

High-altitude satellites range scheduling for urgent request utilizing reinforcement learning

https://doi.org/10.1515/astro-2022-0033
received June 10, 2022; accepted July 10, 2022

Abstract: High-altitude satellites are visible to more ground station antennas for longer periods of time, its requests often specify an antenna set and optional service windows, consequently leaving huge scheduling search space. The exploitation of reinforcement learning techniques provides a novel approach to the problem of high-altitude orbit satellite range scheduling. Upper sliding bound of request pass was calculated, combining customized scheduling strategy with overall antenna effectiveness, a frame of satellite range scheduling for urgent request using reinforcement learning was proposed. Simulations based on practical circumstances demonstrate the validity of the proposed method.

Keywords: satellite range scheduling, reinforcement learning, slide

1 Introduction

To ensure the required and reliable satellite-ground communication capability, a set of ground stations are placed over the country and abroad to provide TT&C services, satellite payload data reception services, and launcher tracking services for supported missions. A suite of applications that can cope with satellite range scheduling were developed. These applications are automated and rules based, but the scheduling strategy is evolved in line with the space access requirement of current and future missions. Satellite range scheduling means scheduling communications between satellites in space and antennas on the ground. Satellite communicate with ground station antenna only during visible window. With the arrival of new request for satellite-antenna pass, the scheduling algorithm is in charge of solving conflict among passes. All the requests are expected to be satisfied with certain constraints. Urgent request from satellite reaches the scheduling application due to many kinds of uncertainties, such as random satellite observation, scheduled antenna breakdowns, mission cancellations, and other urgent events. In general, regardless of normal request or urgent request, the ground antennas are oversubscribed, and all the requests cannot be satisfied.

For high-altitude orbit satellites, there are two special situation of satellite-antenna pass in the scheduling problem. First, the visible time is much longer than the low-altitude satellites, but only part of the visible time can be arranged to request satellite. Second, to overcome the congestion issue between the scheduled pass and request in future (before the start time of the scheduled pass), allocated passes may be adjusted in the visible window on different antennas. Since the search space is usually large and with specific constraints, solving the problem of high-altitude satellite range scheduling is very computationally demanding.

In the past decades, various techniques are studied to generate conflict free and high-efficient solutions for the problem of satellite range scheduling (Petelin *et al.* 2021, Luo *et al.* 2017, Li *et al.* 2015, Li *et al.* 2014, Zufferey *et al.* 2008, Zufferey and Michel 2015, Ling *et al.* 2013, Badaloni *et al.* 2007, Bagchi 2009, Xhafa *et al.* 2013, Zhang *et al.* 2014). Heuristics methods provide a solution quickly, but fails to scale to larger and more complex problems. Metaheuristics such as genetic algorithm, simulated annealing and tabu search take more time but output results closer to optimal solution. Barbulescu *et al.* (2004) proved that satellite range scheduling is NP complete, Marinelli *et al.* (2011) and Brown *et al.* (2018) developed a new heuristic method based on Lagrangian relaxation to overcome the difficulty that large-scale variables yield, and the former applied the method to GALILEO constellation. As far as we know, there exist few studies about satellite range scheduling using reinforcement learning.

Reinforcement learning is about learning how to act to achieve a goal in an uncertain environment to maximize the reward in a particular situation, and it is a specialized application of machine learning and deep learning

* **Corresponding author: Bo Ren,** State Key Laboratory of Astronautic Dynamics, Xi'an Satellite Control Center, Xi'an 710043, China, e-mail: renbo1313@126.com
Zhicheng Zhu, Fan Yang, Tao Wu, Hui Yuan: State Key Laboratory of Astronautic Dynamics, Xi'an Satellite Control Center, Xi'an 710043, China

techniques to solve various types of decision making problems (Sutton and Barto 2019, Wiering and Otterlo 2018). The agent is trained to seek a new way to maximize the reword, and by interacting with the environment, a neural network storing experiences improves the performance. The reinforcement learning algorithms are usually classified into three types: value-based approach, policy-based approach and model-based approach. In a value-based reinforcement learning algorithm, the agent will try to maximize a value function and is expecting a long-term return of current state. In a policy-based algorithm, the agent will try to come up with such a policy that the action performed in every state improves the return in the future. In a model-based algorithm, a model was created in every environment and the agent learns to perform in them. Recently, reinforcement learning has also gained more and more attention in scheduling. Cunha *et al.* (2021) created a novel architecture that incorporates reinforcement learning into scheduling systems. Shyalika *et al.* (2020) addressed the reinforcement learning techniques used for dynamic task scheduling. Luo (2020) developed a deep Q-network to solve the dynamic job shop scheduling problem. Huang *et al.* (2021) used deep deterministic policy gradient algorithm to solve the satellite task scheduling problem. Inspired by the aforementioned applications of reinforcement learning in scheduling, solving the problem of satellite range scheduling by using reinforcement learning became feasible.

The main contribution of this article is the proposal of an innovative high-altitude satellite range scheduling method for urgent request by adopting reinforcement learning. We utilized Q learning as the learning model, and Q learning is a value-based algorithm of providing information to inform which action an agent should take (Watkins 1989, Watkins & Dayan 1992). Our article is organized as follows. First, we introduce the background of the problem and several basic definitions. Section 2 is the analysis of high-altitude orbit satellites urgent request scenario. Calculation of sliding bound of request pass is described in Section 3. A novel frame of satellite range scheduling using reinforcement learning is proposed in Section 4. The simulation and discussion are presented in Section 5. Finally, we conclude this study and put forward some future ideas in Section 6.

2 Analysis of high-altitude orbit satellite request for ground antenna

To introduce the problem of satellite range scheduling clearly, several major definitions are listed as follows:

1) **High-altitude orbit satellite request**: When a high-altitude orbit satellite need a supported service from an antenna, it will send a resource requests message to the scheduling system. The message mainly includes the least last time (LLT, minimum duration) and the request time (RT), LLT is the shortest service time that the satellite need, RT is a time scope in which the service time should be arranged. There are two types of resource requests, normal request and urgent request. The normal requests are periodic corresponding to the regular service demand, and the urgent request is temporary corresponding to the service demand in an emergency. The present article focuses on the resources scheduling for urgent request. Whenever an urgent request arrives, there was already an antenna work plan generated by the normal request scheduling process.
2) **Resources scheduling system**: It is the core system in charge of antenna scheduling for normal and urgent requests, a plan of scheduled passes for all antennas would be generated after scheduling process.
3) **Antenna service switch time**: It is the antenna repositioning time between adjacent service time for different satellites, and it is a constant value for all ground station antennas. The constraints of the switch time, minimum elevation, and visibility are taken into account, but they are neglected in the notations for simplicity.
4) **Scheduled pass and request pass**: The scheduled passes are the components of the service plan of an antenna, it is the output of resources scheduling system. A request pass is the pass with a length, which is equal to the request LLT in the RT. All the passes must be in the satellite-antenna visible time.
5) **Service conflict**: The conflict between two service time is defined by pass conflict that one pass is overlapped with the adjacent one, while the antenna service switch time must be considered.
6) **Sliding operation**: In a scheduling environment, a request pass should be arranged on an antenna and then all the scheduled passes should be conflict free. We define the arrange operation as insert. When a pass insert operation causes a conflict, all the passes in the RT should try to move backward (to the earlier time or to the left side) or forward (to the later time or to the right side) to overcome the confliction. The move of pass on an antenna is called slide, the request pass after slide operation should be in the RT, and other pass after slide operation should be in the visible time.

The design of urgent request data is to create an interface for interoperability between the satellite and the scheduling system, and it specifies the service detail that may be provided by an antenna. An urgent request contains the following elements:

- a request satellite;
- a set of ground station antennas;
- request time (RT);
- least last time (LLT);
- the type of request service.

In the RT on an antenna belonging to the request antenna set, there may be scheduled high-altitude satellite normal/urgent passes, or scheduled low-altitude satellite passes. To achieve the state of conflict free of all passes in the RT, the scheduled high-altitude satellite passes and request pass need to slide in a request-relevant time scope, and the scheduled low-altitude satellite passes cannot slide due to the relative short visible time. Let us assume that *Req* is the urgent request of satellite s and the request antenna set is $D = \{d_1, d_2, \ldots, d_N\}$, where N is the number of request antennas, the request time window is W_r, the LLT is W_s. Let $W[i]$ denote the visible time of s on the antenna d_i. Figure 1 shows the urgent request scheduling scenario, and the approach to find a conflict free position for request pass is as follows:

1) Calculate the intersection of visible time and request time, that is, $W[i] \cap W_r$. Subtract the switch time from the start time of intersection and the end time of intersection plus the switch time, and then we obtain the maximum request time (MRT). The difference of RT and MRT is the antenna switch time, and we use MRT in the following sections.
2) Let W(LEO) denotes the scheduled low-altitude satellite passes in MRT, and we obtain the scheduled high-altitude satellite passes in MRT as $W_x = W[i] \cap W_r - W$ (LEO).
3) Constraints of the process of urgent request scheduling. W_x and *Req* may both need to slide, and the sliding scope of the terms in W_x are determined by the corresponding original request.
4) The satellite range scheduling. This will be further illustrated in Section 4.

3 Calculation of the sliding upper bound

In this section, the sliding bound of the high-altitude satellite request pass will be provided. We call upper bound of sliding operation a maximum time scope, such that in which one high-altitude satellite pass can move. The initial condition for a conflict free scheduling solution is $p_x \geq llt_x$, $r_x \leq let_x$, where p_x is the visible time of request satellite on antenna d, llt_x is the LLT of the urgent request t_x, r_x is the start time of p_x and let_x is latest ending time which equal to the end time of the request. Each high-altitude satellite scheduled pass had an corresponding original request, and the parameters of the original request will be parsed for calculation of the upper sliding bound.

For the case of sliding backward in Figure 2, s_v and f_v are the start time and the end time of a scheduled pass t_v, respectively. p_{x-1} is the visible time of the satellite of scheduled pass t_{x-1} at the left of t_x on antenna d, t_{x-2} and t_{x-3} are scheduled high-altitude satellite passes and t_i and t_{i-1} are scheduled low-altitude orbit satellite passes. For low-altitude satellite, the whole visible time will be arranged by the scheduling system if conflict does not exist; consequently, the scheduled pass cannot slide in visible time. For scheduled passes of high-altitude satellites, the time length of sliding backward is determined by two items: one is the start time of visible window and the other is the end time of the left hand scheduled pass. If t_{x-1} is a scheduled high-altitude orbit satellite pass, and there are u_1 scheduled high-altitude orbit satellite passes between t_{x-1} and the low-altitude orbit satellite scheduled passes, then the sliding upper bound of t_{x-1} to the left side is expressed as follows:

$$\begin{aligned}\Delta_l = {} & f_{x-1} - \max(f_{x-2}, r_{x-1}) - llt_{x-1} \\ & + \sum_{j=1}^{u_1-1}(f_{x-1-j} - \max(f_{x-2-j}, r_{x-1-j}) - llt_{x-1-j}) \\ & + f_{x-1-u_1} - \max(f_i, r_{x-1-u_1}\big) - llt_{x-1-u_1}.\end{aligned} \quad (1)$$

For the case of sliding forward in Figure 3, t_{x+1} and t_{x+2} are scheduled high-altitude orbit satellite passes, and

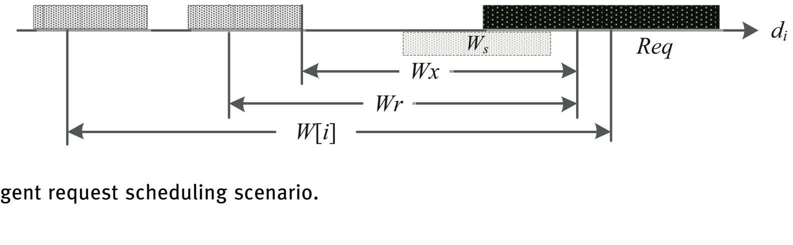
Low altitude orbit satellite pass
Low altitude orbit satellite pass
High altitude orbit satellite pass
d_i
W_s
Req
Wx
Wr
$W[i]$

Figure 1: Example of an urgent request scheduling scenario.

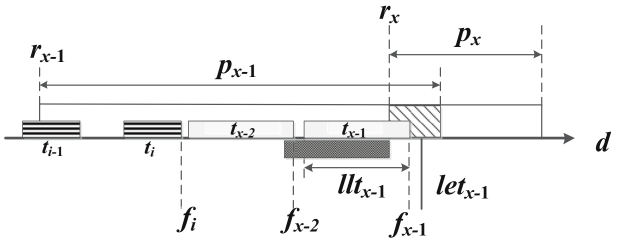

Figure 2: Scheduled high-altitude orbit satellite passes slide backward.

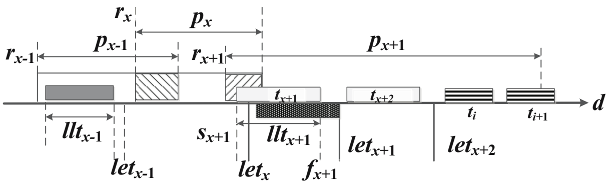

Figure 3: Scheduled high-altitude orbit satellite passes slide forward.

t_i and t_{i+1} are scheduled low-altitude orbit satellite passes. llt_{x+1} is the LLT of the request of t_{x+1} and let_{x+1} is the latest ending time of request of t_{x+1}. For scheduled passes of high-altitude satellites, the time length of sliding forward is determined by three items: one is the end time of visible window, the second is the latest ending time, and the third is the start time of the right hand scheduled pass. The high-altitude satellite passes from the right side slide one by one, and start time of the sliding pass will be updated. If there are u_2 scheduled high-altitude orbit satellite passes between t_{x+1} and the low-altitude orbit satellite scheduled passes, then the sliding upper bound of t_{x+1} to the right side is expressed as follows:

$$\begin{aligned}\Delta_r = {} & \min(r_{x+1} + p_{x+1}, let_{x+1}, s_{x+2}) - f_{x+1} \\ & + \sum_{j=1}^{u_2-1}(\min(r_{x+1+j} + p_{x+1+j}, let_{x+1+j}, s_{x+1+j}) - f_{x+1+j}) \\ & + \min\left(r_{x+1+u_2} + p_{x+1+u_2}, let_{x+1+u_2}, s_i\right) - f_{x+1+u_2}.\end{aligned} \tag{2}$$

Updated values of s_{x+1+j} after every slide were used in calculation of Δ_r in Eq. (2).

We learn from Eqs. (1) and (2) that the length of the total time window for request pass after sliding on d is

$$\Delta T = \Delta_l + \Delta_r + s_{x+1} - f_{x-1}. \tag{3}$$

The time length ΔT is the maximum idle time window on antenna d that an urgent request can be arranged. If ΔT is not longer than LLT, the corresponding request will be rejected on antenna d immediately. Consequently, scheduling process ended before reinforcement learning module.

4 A frame of satellite range scheduling using reinforcement learning

Scheduling strategy is the reference policy for the scheduling method and is the principle of sliding operation (Pinedo 2016, Conway *et al.* 2003). The strategies describe that how the request pass slide to avoid overlap. We design two scheduling strategies for high-altitude satellite urgent request, mono-layer, and multi-layer scheduling.

- **Mono-layer scheduling strategy**. The request pass is only permitted to be arranged on the current antenna, the scheduled passes in the MRT of the request cannot be transferred to the other antenna belonging to the corresponding request antenna set.
- **Multi-layer scheduling strategy**. The request pass can be arranged on an antenna belonging to the request antenna set, and the scheduled high-altitude satellite passes can be transferred to the any other antenna belonging to the corresponding original request antenna set.

Compared with the mono-layer strategy, the multi-layer strategy may change the work plan of other antenna, this will bring instability to the entire plan of the network and further increase the searching space. However, the advantage of multi-layer strategy lies in providing more scheduling solutions. It can utilize various optimization techniques, such as load balancing between different antennas, user-oriented antenna preference, and overall antenna service efficiency. If the scheduling method utilizing mono-layer strategy is failed to generate a conflict free scheme, then the multi-layer strategy could be selected as the second stage of the high-altitude satellite request scheduling.

After the discussion of scheduling strategy, we propose an overall antenna effectiveness (OAE) function to represent the benefit of different sliding operations:

$$O = \frac{1}{T_1 - T_0}\sum_{i=0}^{N}(\min(T_i^e, T_1) - \max(T_i^s, T_0)), \tag{4}$$

where T_0 is the start time of MRT and T_1 is end time of MRT, T_i^s is the start time of request pass or scheduled pass

t_i, and T_i^e is the end time of request pass or scheduled pass t_i. N is the total number of passes in MRT. The OAE function will be calculated on the basis of conflict free scheduling result, that is, if the scheduling is failed, then the value of OAE will be 0. If the passes in MRT trying to occupy the time in the MRT, not out of MRT, then the OAE function may obtain a bigger value. A bigger value means a higher percentage of service time, while the antenna is online. If part of one pass is out of MRT, then the upper bound of OAE is expressed as follows:

$$O_m = \frac{1}{T_1 - T_0}\left(\Delta t_1 + \Delta t_2 + \sum_{i=1}^{N-1}(T_i^e - T_i^s)\right), \tag{5}$$

where $\Delta t_1 = \max(T_0^e - T_0)$, $\Delta t_2 = \max(T_1 - T_N^s)$, If all the passes are perfectly in the MRT, then the value of OAE will be a constant:

$$O_c = \frac{1}{T_1 - T_0}\left(\sum_{i=0}^{N}(T_i^e - T_i^s)\right). \tag{6}$$

From Eqs. (5) and (6), we find that the value of OAE ranges from O_c to O_m for a conflict free scheme.

The new satellite range scheduling frame for urgent request of high-altitude orbit satellites begins from some main components of reinforcement learning. The details are as follows:

- **Agent**: There are two types of time windows, LLT and MRT in high-altitude satellite request. These two terms are regarded as entities, which slide forward or backward on an antenna in the scheduling environment.
- **Action**: The action of the agents is sliding from one state to another, which means that the agent can stay still, move forward, or backward on the antenna, and the time scope of sliding operation is limited.
- **Environment**: In a scheduling scenario, there would be one ground station antenna or several antennas, one high-altitude satellite request, and scheduled passes in the request-relevant MRT. To achieve the objective that all the scheduled passes are conflict free after sliding actions, the agents need to slide according to the value of reward. The state of agents will be updated after each sliding. The direction and magnitude of the sliding are the key factors of each action.
- **State**: The state of the scheduling environment consists of the start time and the end time of each time window on the corresponding antenna, and the result of the value function at the end of the sliding action.
- **Value function**: We use overall antenna effectiveness as value function, it is a measure of how well a scheduling operation is utilized in the MRT, compared to full potential of an antenna during the visible time window. It specifies the value of a state, and the agents should be expected maximizing the overall antenna effectiveness in the MRT on an antenna for a high-altitude satellite request. If clashes between two passes are detected, then the value of OAE will be 0.

In Figure 4, we design an architecture to illustrate the scheduling mechanism that employs reinforcement learning techniques. When a new urgent request arrives, visible time, antennas, RT, and LLT will be parsed from request. Mono-layer scheduling strategy is selected if single antenna was requested. Then sliding upper bound is adopted as a basic filter for the following process, which was described at the end of Section 3. User-oriented design of value function makes this frame flexible for various situations, and current design comes from the perspective of service provider. This scheduling frame works as a loop according to the requested antennas, and the process continues with next antenna if the previous antenna failed to output a conflict free window.

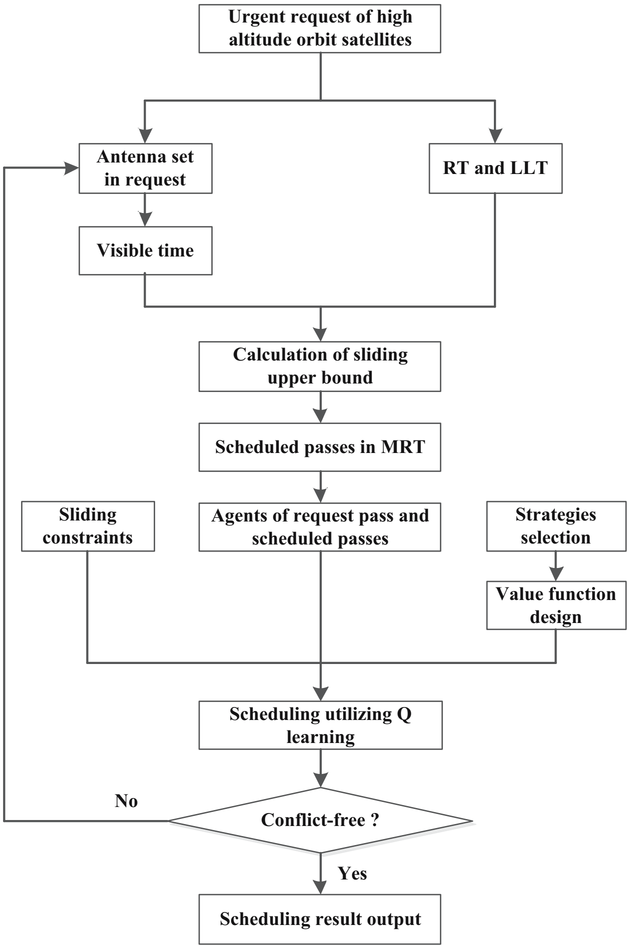

Figure 4: Scheduling procedure utilizing reinforcement learning.

The scheduling frame breaks out of the loop if a valid result is generated. Two major factors are considered in the real circumstance of satellite range scheduling: the quality of the scheduling result and the time consumed to obtain it. Both the scheduling strategy and the value function of reinforcement learning will react on the objective. The proposed frame of scheduling is not to generate the best solution, but a result good enough that is obtained by an intelligent scheduling system using reinforcement learning.

5 Simulation

The simulation is implemented in a real operational context, 30 ground station antennas, 250 satellites including 30 high-altitude orbit satellites, and a plan of scheduled passes for normal requests with a period of 7 days was involved. The data environment for mono-layer strategy scheduling is presented in Table 1, where *s*_3 represents a low-altitude orbit satellite, other satellites are high-altitude orbit satellites. For high-altitude orbit satellite's pass in MRT of *s*_1, it has an original request that is listed in its second line. The antenna service switch time is considered in the start time and end time in Table 1, where "Sat" represents a satellite or a satellite pass, "ST" represents the start time of the pass, "ET" represents the end time of the pass, "Type" means that the pass of line is a high-altitude orbit satellite scheduled/request pass or a low-altitude orbit satellite pass. The MRT of urgent request of *s*_1 is from 13:49:00UTC to 16:39:20UTC.

We apply Q learning to the scheduling for urgent request of *s*_1. The pass of *s*_3 cannot slide, the reinforcement learning parameters are as follows: initial pass position of *s*_1 is 13:55:12UTC-14:10:12UTC, learning rate is 0.001, and slide step is 3 s. Figure 5 shows the variation of OAE for different values of training times.

Table 1: Data environment for mono-layer strategy simulation

Sat	ST/UTC	ET/UTC	LLT/second	Type
*s*_1	13:55:00	16:33:20	900	Urgent request
*s*_2	13:55:12	14:10:12	900	Scheduled pass
*s*_2	13:55:00	16:33:20	900	Original request
*s*_3	14:31:19	14:41:28	—	Low altitude
*s*_4	14:58:36	15:54:06	3,330	Scheduled pass
*s*_4	14:58:30	16:21:46	3,330	Original request
*s*_5	16:35:33	20:22:13	13,600	Scheduled pass
*s*_5	16:11:40	20:50:35	13,600	Original request

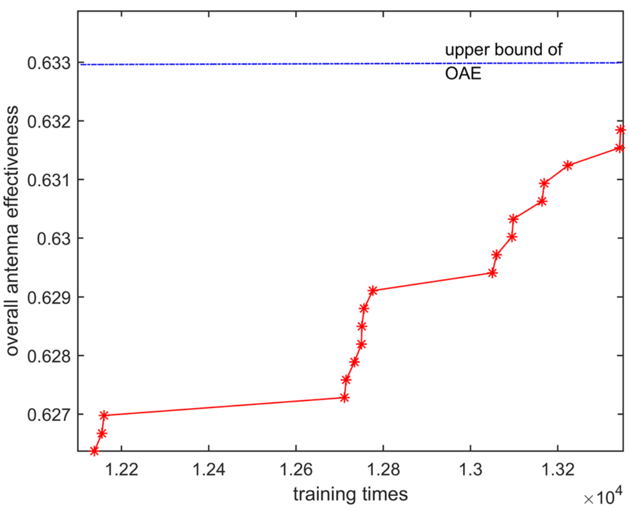

Figure 5: Variation of OAE for different values of training times with mono-layer strategy.

We find from Figure 5 that the maximum value of OAE is 0.632 (the corresponding scheduled result for *s*_5 is 16:31:09UTC-20:17:49UTC) and lies in the range of O_c = 0.582 and O_m = 0.63296 (the corresponding theoretical position for *s*_5 is 16:30:58UTC-20:18:58UTC).

The data environment for multi-layer strategy scheduling is presented in Tables 2 and 3. The request pass and scheduled passes in the MRT of *s*_1 on the first preferred antenna *d*_1 are listed in Table 2, The scheduled passes in the MRT of *s*_2 on the first preferred antenna *d*_2 except *d*_1 are listed in Table 3. Five agents are request of *s*_1, scheduled passes *s*_2, *s*_7, *s*_8, and *s*_9, learning rate is 0.001, and slide step is 3 s. The training result is shown in Figure 6, due to the failure in the first stage of mono-layer strategy scheduling, *s*_2 was transferred to *d*_2 and the multi-layer strategy is implemented.

Table 2: Data environment for multi-layer strategy simulation on *d*_1

Sat	ST/UTC	ET/UTC	LLT/second	Type
*s*_1	13:55:00	15:50:00	900	Urgent request
*s*_2	13:45:00	14:00:00	900	Scheduled pass
*s*_2	12:00:00	14:00:00	900	Original request
*s*_3	14:08:57	14:28:57	1,200	Scheduled pass
*s*_3	08:59:57	14:57:30	1,200	Original request
*s*_4	14:50:58	15:10:58	1,200	Scheduled pass
*s*_4	14:43:58	17:54:49	1,200	Original request
*s*_5	15:17:00	15:32:00	900	Scheduled pass
*s*_5	14:36:00	16:31:00	900	Original request
*s*_6	15:38:12	15:53:12	900	Scheduled pass
*s*_6	14:55:00	16:50:00	900	Original request

Table 3: Data environment for multi-layer strategy simulation on *d*_2

Sat	ST/UTC	ET/UTC	LLT/second	Type
*s*_7	12:02:06	12:11:46	—	Low altitude
*s*_8	12:30:23	12:45:23	900	Scheduled pass
*s*_8	12:00:00	14:00:00	900	Original request
*s*_9	13:05:57	13:25:57	1200	Scheduled pass
*s*_9	13:00:00	15:00:00	1200	Original request
*s*_10	14:18:58	14:30:14	—	Low altitude

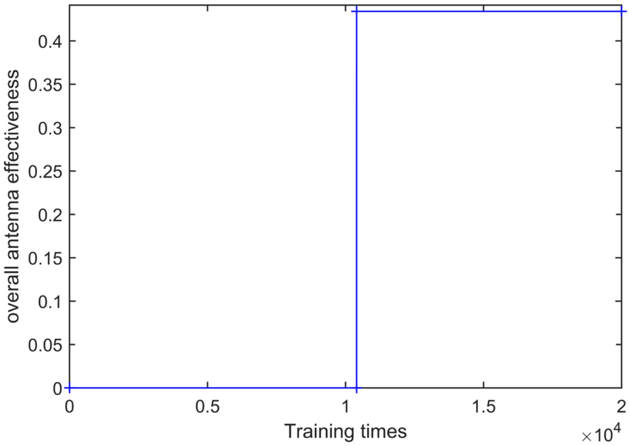

Figure 6: Variation of OAE for different values of training times with multi-layer strategy.

Then the value of OAE equal to $O_c = 0.434$ after the successful slide of *s*_2, *s*_8, and *s*_9 on *d*_2.

In the experiment with a more larger scale, for example, there are more than 15 high-altitude orbit satellite passes in one MRT, it is difficult to have a good control on the search. An additional high-altitude scheduled pass in MRT can lead to a slightly drawback in run time. On the contrary, a low-altitude scheduled pass in MRT instead of high-altitude pass can result in a similar OAE. Improving the overall performance of the method on large-scale scenario will be an important research topic in the future.

6 Conclusion

This article demonstrates how reinforcement learning can be used in high-altitude satellites range scheduling for urgent request. In the frame of our implementation, calculation of pass sliding bound and overall antenna effectiveness was introduced, Q learning was utilized as the learning model. Two kinds of scheduling strategies make our novel approach feasible to be extended to any other user-oriented scheduling method. The trained model can directly generate a scheduling result without retraining a new instance. Although not considered in our article, the proposed method can easily incorporate urgent request scheduling for request from low-altitude satellites. Numerical experiments are conducted on a large-scale of practical context, and results of simulations prove that the proposed method is feasible.

Funding information: This work was supported by the National Natural Science Foundation of China (Grant No. U21B2050).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

References

Badaloni S, Falda M, Giacomin M. 2007. Solving temporal over-constrained problems using fuzzy techniques. J Intell Fuzzy Sys. 18:255–265.

Bagchi TP. 2009. Near optimal ground support in multi-spacecraft missions: a GA model and its results. IEEE Trans Aero Elec Sys. 45:950–964.

Barbulescu L, Watson JP, Whitley LD, Howe AE. 2004. Scheduling space-ground communications for the air force satellite control network. J Scheduling. 7:7–34.

Brown N, Arguello B, Nozick L, Xu N. A heuristic approach to satellite range scheduling with bounds using Lagrangian relaxation. 2018. IEEE Sys J. 12:3828–3836.

Conway RW, Maxwell WL, Miller LW. 2003. Theory of scheduling. Mineola (NY), USA: Dover Publications.

Cunha B, Madureira A, Fonseca B, Matos J. 2021. Intelligent scheduling with reinforcement learning. Appl Sci. 11:3710.

Huang YX, Mu ZC, Wu SF, Cui BJ. 2021. Revising the observation satellite scheduling problem based on deep reinforcement learning. Remote Sens. 13:2377. doi: https://doi.org/10.3390/rs13122377.

Li YQ, Wang RX, Liu Y, Xu MQ. 2015. Satellite range scheduling with the priority constraint: an improved genetic algorithm using a station ID encoding method. Chinese J Aeronaut. 28:789–803.

Li YQ, Wang RX, Xu MQ. 2014. An evolution algorithm for satellite range scheduling problem with priority constraint. Appl Mech Mater. 568–570:775–780.

Ling XD, Zhu WK, Wu JM, Wu XY. 2013. Research of multi-satellite T&C scheduling problem. Appl Mech Mater. 263–266:476–484.

Luo KP, Wang HH, Li YJ, Li Q. 2017. High-performance technique for satellite range scheduling. Comput Oper Res. 85:12–21.

Luo S. 2020. Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning. Appl Soft Comput. 91:106208. doi: https://doi.org/10.1016/j.asoc.2020.106208.

Marinelli F, Nocella S, Rossi F, Smriglio S. 2011. A Lagrangian heuristic for satellite range scheduling with resource constraints. Comput Oper Res. 38:1572–1583.

Petelin G, Antoniou M, Papa G. 2021. Multi-objective approaches to ground station scheduling for optimization of communication with satellites. Optim Eng. 2021:1–38. doi: https://doi.org/10.1007/s11081-021-09617-z.

Pinedo ML. 2016. Scheduling: theory, algorithms and systems. 5th ed. Berlin Heidelberg, Germany: Springer.

Shyalika C, Silva T, Karunananda A. 2020. Reinforcement learning in dynamic task scheduling: a review. SN Comput Sci. 1:306. doi: https://doi.org/10.1007/s42979-020-00326-5.

Sutton RS, Barto AG. 2019. Reinforcement learning: an introduction. 2nd ed. Beijing, China: Publishing House of Electronics Industry.

Watkins CJCH. 1989. Learning from delayed rewards, PhD thesis, King's College, London, UK.

Watkins CJCH, Dayan P. 1992. Q-learning. Mech Learn. 8:279–292.

Wiering M, Otterlo M. 2018. Reinforcement learning: state-of-the-art. Beijing, China: China Machine Press. doi: https://doi.org/10.3390/app11083710.

Xhafa F, Herrero X, Barolli A, Barolli L, Takizawa M. 2013. Evaluation of struggle strategy in genetic algorithms for ground stations scheduling problem. J Comput Syst Sci. 79:1086–1100.

Zhang ZJ, Zhang N, Feng ZR. 2014. Multi-satellite control resource scheduling based on ant colony optimization. Expert Syst Appl. 41:2816–2823.

Zufferey N, Amstutz P, Giaccari P. 2008. Graph colouring approaches for a satellite range scheduling problem. J Scheduling. 11:263–277.

Zufferey N, Michel V. 2015. A generalized consistent neighborhood search for satellite range scheduling problems. Oper Res. 49:99–121.