

Research Article

Scheduling and Optimization of Wireless Virtual Network Resources Based on Reinforcement Learning

Junxiang Wang 

School of Information and Intelligent Transportation, Fujian Chuanzheng Communications College, Fuzhou 350007, China

Correspondence should be addressed to Junxiang Wang; 2001002@fjccpc.edu.cn

Received 24 February 2022; Accepted 15 March 2022; Published 31 March 2022

Academic Editor: Zhiguo Qu

Copyright © 2022 Junxiang Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the booming and proliferation of 5G wireless network services in the future, a large number of wireless virtual network resources will emerge, and the densification and heterogeneity of the wireless communication networks that provide services for them will become the trend of development. To this end, a wireless virtual network resource scheduling model based on user satisfaction is constructed, and a reinforcement learning-based wireless virtual network resource scheduling mechanism, IRSUP, is proposed. IRSUP is designed with an intelligent optimization module for user service preferences to address the personalized needs of user service customization, and a reinforcement learning-based intelligent scheduling module is designed to address the challenge of joint optimization of multistar resources. Simulation results show that IRSUP can effectively improve resource scheduling rationality and link resource utilization and user satisfaction, among which service capacity is improved by 30% to 60% and user satisfaction is more than doubled.

1. Introduction

Network virtualization is currently considered one of the most promising ways to achieve network innovation; using network virtualization technology, a variety of network regulation and control functions can be stripped from the hardware, breaking the original In P (InfrastructureProviders) and SP (ServiceProvider) one-to-one correspondence model [1]. The purpose is to unify the management and control of the network and to improve the efficiency of coordination and management, and the SP can rent the In P to provide the required services for users, crossing the barrier of physical resources and making the network more flexible, and excel in solving the rigid problem of backward compatibility of the Internet architecture [2, 3].

With the rapid growth of wireless communications and services, the natural combination of network virtualization technology and wireless communications technology has led to the creation of wireless network virtualization [4]. By applying wireless network virtualization technology, the more expensive wireless physical underlying resources are leased and shared by multiple virtual network operators

(VNOs), improving resource utilization and reducing spend. Wireless virtualization and the capabilities of VNOs are described in detail in the next section; at the same time, wireless network virtualization has good potential to solve the problem of coexistence and interoperability of wireless network heterogeneity; in addition, wireless network virtualization technology can be more easily ported to new products or technologies [5]. Although wireless network virtualization has many advantages, there are still many technical challenges in the research and implementation process, such as the definition of interfaces between different layers of the wireless virtual network, signalling control, resource discovery and allocation, mobility management, network management and operation, and network security [6].

In particular, in solving the problem of resource sharing, wireless virtualization is different from wired virtualization, which can accomplish the abstraction and independence of resources at the hardware level; unlike wired networks that can establish tunnels on physical devices to achieve virtual nodes and virtual links, the inherent link broadcast characteristics of wireless communication and the volatility of the

channel cannot directly accomplish the abstraction of wireless resources; and it is difficult to achieve independence of resources [7]. In addition, the spectrum resources of wireless networks are very valuable resources, and due to the limitations of current wireless network technology, there is the problem of unreasonable allocation of spectrum resources, resulting in the shortage of resources [8]. The introduction of wireless network virtualization technology centralizes spectrum resources and provides a means for effective management of spectrum resources; similarly, wireless network virtualization enables multiple wireless virtual networks to share expensive wireless physical devices and also provides a platform for resource management for the effective use of wireless physical device resources. In short, wireless network virtualization solves the problem of shared coexistence and efficient reuse of resources, so the management and allocation of resources are the most important technical issues, which determine how to embed the wireless virtual network into the physical underlying network, as well as the good operation of resources' update [9].

The research of wireless network virtualization has just started, the research results are still immature, and the technology implemented is also very different from that of wired network virtualization. In recent years, the technology implementation of wireless network virtualization has become a research hotspot for scholars and has significant research significance.

2. Related Work

In recent years, research on wireless network virtualization has been carried out worldwide in order to promote in-depth research on wireless network virtualization [10], proposing a holistic implementation architecture on wireless sensing networks and the ability to provide advanced services. This solution enables the decoupling of applications from wireless sensor network deployments, allowing the dynamic collaboration of sensor nodes to accomplish new services or add value to applications beyond the original deployment [11]. A wireless LTEv_RAN virtualization solution based on a common platform is presented in [12], which is based on a common virtualized processing platform that enables resource sharing and is cost effective, highly reliable, and easy to deploy. The idea of related base station virtualization is introduced in [13], which studies the virtualization of base stations for LTE networks in 3GPP. The study is based on the idea of base station virtualization mentioned in [14], but the difference is that instead of introducing a hypervisor, a resource management system is used to virtualize the radio access network. In [15], a set of management strategies is proposed, from the extraction of wireless resource information to the formulation of reasonable decisions to the final implementation, with a complete system and well-defined functions; only by collecting information from all aspects of the network, such as users, underlying devices and network conditions, can the information be analyzed to provide a good solution strategy; and the formulation of each strategy contains a rigorous scientific logic and is supported by a strong mathematical theory. The imple-

mentation process is also multifaceted and involves close collaboration between all members of the network [16]. The paper [17] proposes a new heterogeneous network convergence platform based on a centralized access network architecture, which allows the sharing and unified management of processing and wireless resources in heterogeneous networks on the basis of a centralized resource pool; it also investigates techniques such as virtualization of base station resources, on-demand allocation of processing resources, and dynamic spectrum sharing. The paper [18] proposes a new business model to serve the convergence of heterogeneous networks and explains that a resource management model can meet customer needs and reasonably utilize the underlying resources which is the key to effective implementation of resource allocation. The literature [19] addresses the problems of poor scalability and autonomy of network virtualization resource management architecture and designs a hierarchical virtual network resource management architecture using a hierarchical and domain-based management mechanism.

In summary, wireless network virtualization has attracted widespread attention in recent years and has made great progress in just a few years. With in-depth research, wireless network virtualization still faces many challenges and requires the hard work and struggle of researchers. The research of wireless network virtualization technology has achieved more results, but the architecture for wireless network virtualization implementation needs to be improved, especially the research based on different levels of slicing approach or hybrid approach, the new division of labor for different roles in the business model, and the establishment of a perfect virtualization system. At the same time, on the basis of the new model, the wireless virtual resource management system should be designed or improved. Wireless network virtualization technology still has numerous problems that need to be solved and has a large research space. It is necessary to improve the management technology of wireless virtual resources in the existing wireless network virtualization architecture.

3. Wireless Network Virtualization

We assume that the architecture is as shown in Figure 1, and that the system is logically a 3-layer model, with the In P layer, the virtual layer, and the SP layer from the bottom up. The virtual layer is responsible for segmenting and reorganizing the physical resources from the In P to form virtual network resources, i.e., virtual spectrum resources and virtual wireless self-backhaul links. The abstracted virtual resources are then leased out to SPs belonging to different virtual networks, and the SPs provide end-to-end customized services to their registered users through the leased virtual resources. The SPs provide end-to-end customized services to their registered subscribers through the leased virtual resources. In order to maximize the total revenue of the SPs and to meet the rate requirements of each SP, the virtual network manager in the virtual layer needs to execute the virtual resource allocation algorithm reasonably and efficiently [20].

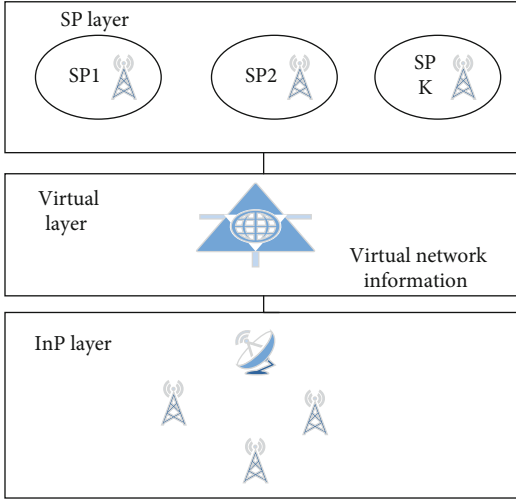


FIGURE 1: Wireless network virtualization framework.

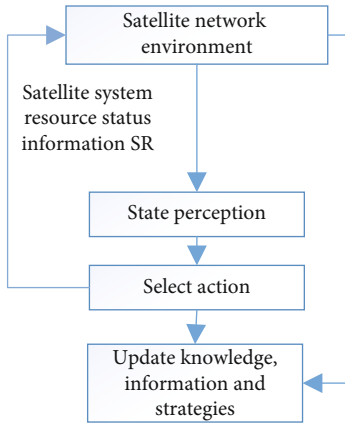


FIGURE 2: Wireless network virtualization resource scheduling decision process.

4. System Modelling and Problem Modelling

4.1. Physical Network Model. The instantaneous spectral efficiency of user k_u in base station n at scheduling period t can be expressed as

$$r_{k_u}^n(t) = \log_2 \left(1 + \frac{P_{k_u} h_{k_u}^n(t)}{\sum_{l=0, l \neq n}^N \sum_{k_u \in S_l} P_{k_u} h_{k_u}^l(t) + \sigma} \right), \quad (1)$$

where P_{k_u} and P_{k_u}' are the transmit power of user k_u and user k_u' , respectively; $h_{k_u}^n(t)$ and $h_{k_u}'^n(t)$ are the instantaneous channel gains from user k_u and user k_u' to base station n , respectively; S_l represents the set of users served by base station l ; and σ is the Gaussian white noise [21].

At each scheduling cycle, the virtual network manager dynamically adjusts the allocation of spectrum resources based on the channel quality and queue state of each SP user. Thus, the instantaneous data rate of user k_u on time slot t

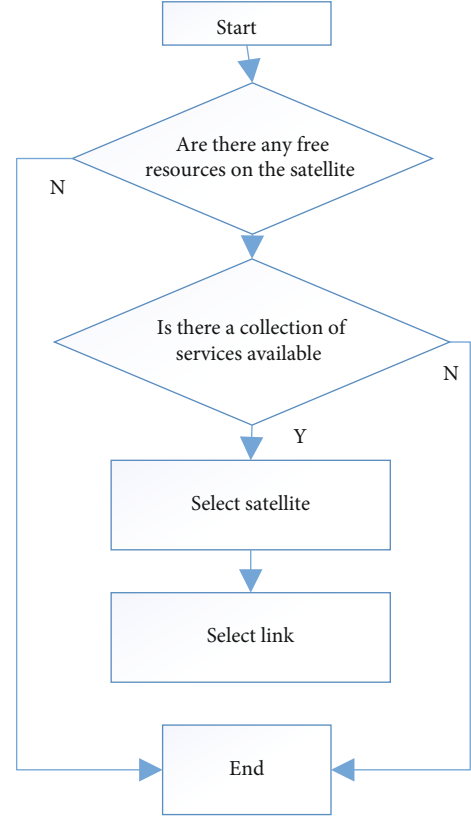


FIGURE 3: Flow chart of the machine learning scheduling module.

can be expressed as

$$R_{k_u}(t) = \sum_{n=0}^N y_{k_u}^n(t) \alpha B r_{k_u}^n(t), \quad (2)$$

where $y_{k_u}^n(t)$ represents both the connection identifier of the user k_u and the proportion of spectrum resources that the user receives in base station n , and $y_{k_u}^n(t) \neq 0$ represents the user k_u connected to base station n . In addition, under a particular virtual resource allocation strategy, the long time average expectation and rate of SPk can be defined as

$$\bar{R}_k = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left\{ \sum_{k_u \in \mathcal{Z}_k} R_{k_u}(t) \right\}. \quad (3)$$

In the uplink, a static spectrum allocation is used to avoid interference problems for small base stations on the backhaul link. Assume that small base station n is allocated to bandwidth ratio g_n for data return, so that the instantaneous return rate of small base station n is

$$R_n(t) = g_n (1 - \alpha) B \log_2 \left(1 + \frac{p_n h_n^0(t)}{\sigma} \right), \quad (4)$$

where p_n represents the transmit power of small base station

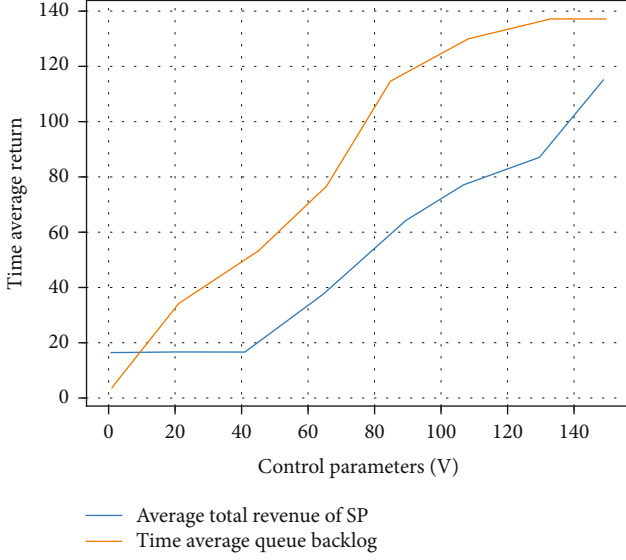


FIGURE 4: Average revenue and average queue backlog versus control parameter V .

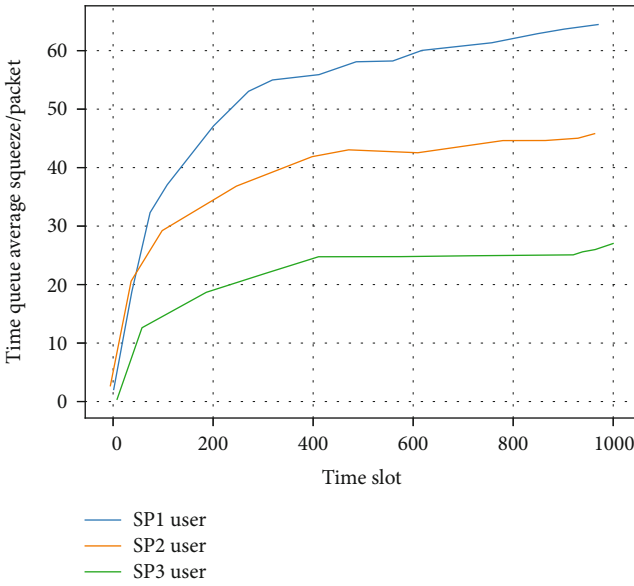


FIGURE 5: Queue variation for different SP users on consecutive time slots.

n in the backhaul link and $h_n^0(t)$ represents the channel gain from small base station n to the macrobase station.

4.2. Problem Modelling. In the business model of wireless network virtualization, SPs lease the appropriate amount of virtual resources to provide end-to-end services to registered users. In a virtualized network based on bandwidth slicing, each SP can schedule users for service and allocate the necessary bandwidth resources to them based on quality of service requirements [22]. Therefore, the net revenue of each SP can be defined as the difference between the total revenue earned by the subscriber for the service provided and the total cost of resource depletion for leased access link resources and back-

haul link resources, which can be expressed as

$$G_k(t) = U\left(\sum_{k_u \in \mathcal{U}_k} \omega_k R_{k_u}(t)\right) - \gamma_k G_k^a(t) - G_k^b(t). \quad (5)$$

The first part of the right-hand side of the above equation represents the revenue received by the SP for providing services to users in the time slot, where $U(\bullet) = \log(\bullet)$ is a utility equation, usually defined as an increasing concave function, and ω_k represents the unit cost charged by the SP to the users of the service. The second component represents the cost to the SP of leasing spectrum resources in the access link, γ_k , representing the agreed price between the In P and the SP for the use of the radio access link. At the time slot, the access side spectrum resources leased by the SP are

$$G_k^a(t) = \sum_{k_u \in \mathcal{U}_k} \sum_{n=0}^N y_{k_u}^n \bullet \alpha B. \quad (6)$$

The third part of the right-hand side of Equation (6) represents the cost incurred by the SP for leasing a wireless backhaul link for data backhaul. Since multiple SPs share the same backhaul link instead of having exclusive access to it, the backhaul cost per SP depends on the amount of data backhauled and the amount of backhaul spectrum resources available. Similar to the results in the literature [20], the SP's overhead on the backhaul link at time slot t can be expressed as

$$G_k^b(t) = \rho_k \sum_{n=1}^N g_n(1 - \alpha) B \{k_u \in \mathcal{U}_k, k_k \in \mathcal{S}_n\} Z_{k_u}(t), \quad (7)$$

where ρ_k represents the unit price of the leased backhaul link resource and \mathcal{S}_n represents the set of users served by small base station n . Since the in-band self-backhaul mechanism utilizes in-band spectrum resources and avoids the need to build additional infrastructure or develop new band resources for small base stations, this technique is clearly more economically efficient than the traditional backhaul method. The resource scheduling problem can be modelled mathematically as follows:

$$\begin{aligned} P1 : \max_{\mathbf{y}, \alpha} \bar{G} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left\{ \sum_{k \in \mathcal{K}} G_k(t) \right\}, \\ \text{s.t. } C1 : \bar{R}_k &\geq R_k, \forall k, \\ C2 : \sum_{k_u \in \mathcal{S}_n} y_{k_u}^n(t) \alpha B r_{k_u}^n(t) &\leq R_n(t), \forall n, n \neq 0, \\ C3 : 0 &\leq y_{k_u}^n(t) \leq 1, \forall n, k_u, \\ C4 : \sum_{k \in \mathcal{K}} \sum_{k_u \in \mathcal{U}_k} y_{k_u}^n(t) &\leq 1, \forall n, \\ C5 : 0 &\leq \alpha \leq 1, \\ C6 : \overline{Q}_{k_u} &< \infty, \forall k_u. \end{aligned} \quad (8)$$

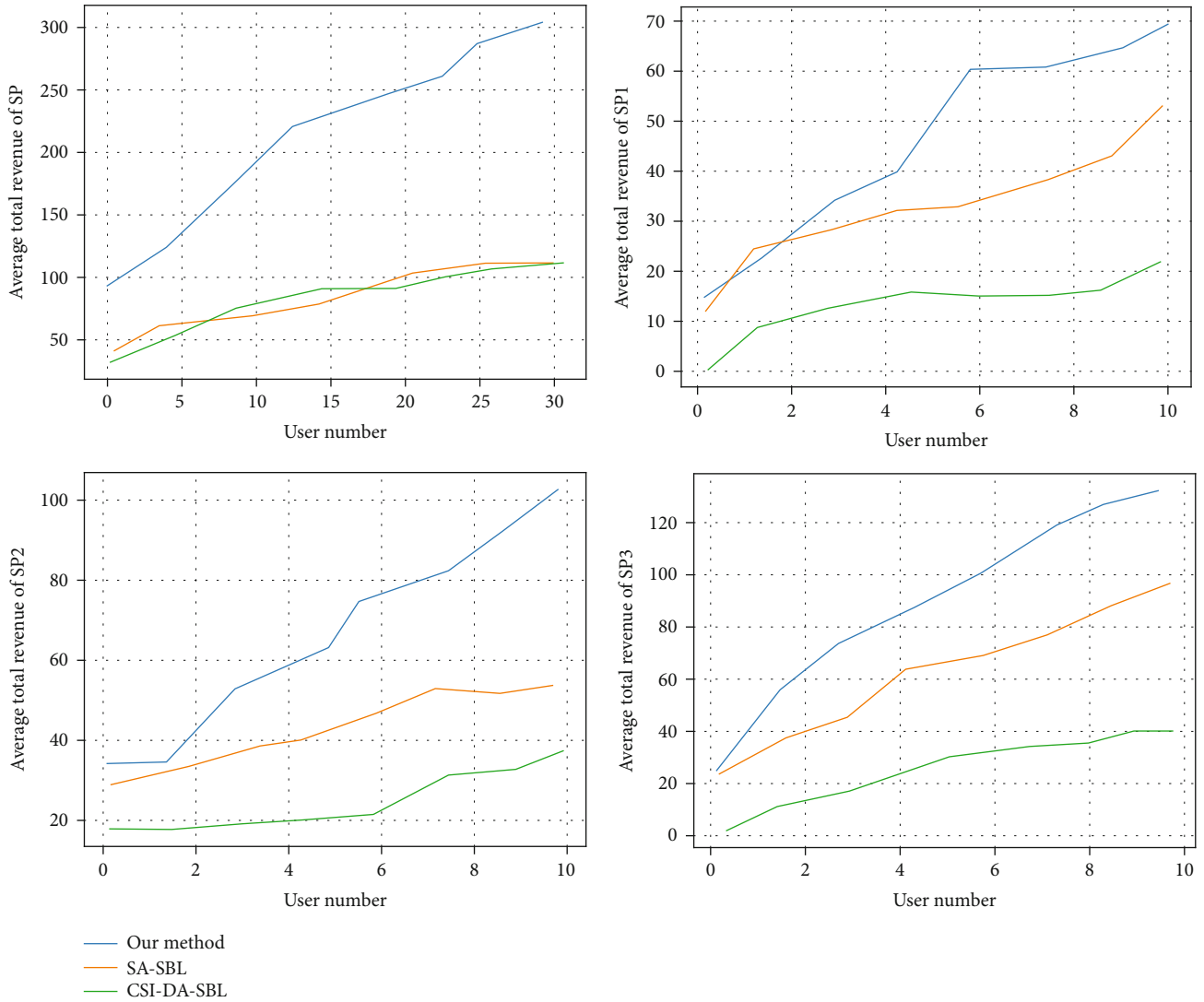


FIGURE 6: Comparison of average SP returns for different virtualization scenarios.

In the above equation, $C1$ represents the minimum average rate guarantee that needs to be provided to each SP during resource scheduling, R_k , being the minimum rate requirement for SPk. $C2$ represents the rate of return that must not be less than the rate at the access end for any small base station $n \in \{1, \dots, N\}$ to avoid an infinite backlog of user data in the small base station that could cause data loss or unnecessary processing delays. $C3$ and $C4$ represent that the sum of the band resources allocated to the users connected to any base station must not exceed the bandwidth limit of that base station and the restriction of $C6$ ensures queue stability [23, 24].

5. Intensive Learning Scheduling Module

An agent for reinforcement learning generally consists of three parts: a state perceiver, a learner, and an action selector. The state perceiver maps the environmental state to internal perception; the learner updates the agent's knowledge strategy according to the reward value of the environmental state and the internal perception; the action selector

selects the action a to act on the environment E according to the current strategy, which will cause the environmental state of E to change under the action a . The basic principle is that if an action of the agent is positively rewarded by the environment (i.e., reinforcement signal), the tendency of the agent to produce this action in the future will be strengthened, while on the contrary, the tendency of the agent to produce this action will be weakened.

The wireless network virtualization resource scheduling model needs to make efficient and reliable choices between each wireless network virtualization and each interplanetary link in order to optimize resource utilization. Therefore, the reinforcement learning scheduling module can be used to maximize the resource utilization and resource scheduling rationality of the wireless network virtualization links [25, 26].

The wireless network virtualization resource scheduling problem is a dynamic decision making problem, in which the decision maker (scheduling system) has to be able to decide on-the-fly whether to process or wait for the current

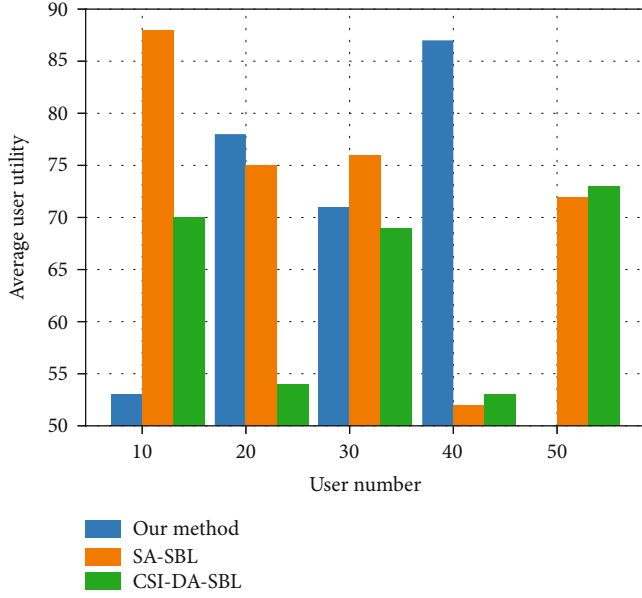


FIGURE 7: Comparison of average user utility for different virtualization scenarios.

arriving user services based on the current resource status. The decision maker has to consider not only the immediate impact of the decision in the current state, but also the future impact of the decision. Specifically in the wireless network virtualization resource scheduling problem, if the decision maker processes the current subscriber service, it consumes the wireless network virtualization resources and is likely to make future tasks with large benefits unable to be executed and must wait longer; if the decision maker allows its subscriber service to enter the waiting queue, it will make the current subscribers less satisfied. Therefore, the decision maker needs to consider the attributes of the current user service and the state of the system resources virtualized by the wireless network to make a decision.

As shown in Figure 2, each wireless network virtualization resource scheduling decision process can be divided into three steps: obtaining current system resource status information, business task decision, and system resource status information update.

The agent obtains the current wireless virtual network resource system resource state information SR through the state perceptron, and then determines the starting wireless network virtualization set S_o [27], the ending wireless network virtualization set S_f , and the link set L_1 that are likely to accept the task after matching and analyzing the user service request state information WSj with SR, which should satisfy the following conditions:

$$\begin{aligned}
 & \exists S_o, S_f, L_1, S_o, S_f \subseteq S, L_1 \subseteq L_r \\
 & B_{ij_c} > 0 \\
 & VB_{v_j_s} > 0 \\
 & WB_j \leq aB_{ik} \cap WB_j \leq aB_{kx} \cap WB_j \leq aB_{qy},
 \end{aligned} \tag{9}$$

where $S = \{i, k, \dots, y\}$ (i etc. is the wireless network virtualization number, i.e., wireless network virtualization i etc.)

$$L_r = \{(ik, kx, \dots, gy), \dots, (fp, pt, \dots, xq)\}. \tag{10}$$

At this point, the action selector then selects the corresponding S_o , S_f , and L_1 to execute action a based on the state perceiver's mapping of the environmental resource state into an internally perceived policy.

The agent learner updates the knowledge information by the current decision maker's decision and the current resource state, and the update formula is shown in

$$Q(s_k, a_k) = Q(s_k, a_k) + \lambda \{R + \max_{a_{k+1}} \{\gamma Q(s_{k+1}, a_{k+1})\} - Q(s_k, a_k)\}, \tag{11}$$

where s_k is the current state, a_k is the current action taken, S_{k+1} is the next state, a_{k+1} is the action taken in the next state, R is the reward obtained after the execution of action a_k , λ is the learning rate, and γ is the decay factor.

The machine learning scheduling module flow chart is shown in Figure 3.

Therefore, the machine learning scheduling module can update the policy and choose the next action in real time according to the system's current resources and the current allocation action, so as to maximize the system's resource scheduling benefits.

According to the principles of machine learning, after an agent's implementation of the rules (S, a_k, r_k) , its learner needs to check the validity of action k (i.e., a_k) against what is running in the actual environment. Its effectiveness can be tested based on resource utilization and QoS. If a_k is the resource utilization and QoS remains at high values after running on E , then, the action a_k is valid, otherwise a_k is not. In turn, the effectiveness of a_k can be used as a reward function for reinforcement learning R .

Let u_{1i}, u_{2i}, u_{3i} be the actual utilization of processor, memory, and link bandwidth, respectively, and counted after a_k running on the wireless network virtualization i ; let b_1, b_2, b_3 be the actual service quality of service in terms of completion time, transmission rate, and jitter of the service after execution, respectively; Let $4\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ be the weighting factor, and $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 = 1$. Then, the reward function is

$$R = R_u + R_b, \tag{12}$$

where R_u, R_b is the reward function for actual resource utilization and service QoS, respectively, which is calculated as shown in

$$\begin{aligned}
 R_u &= u_{1i} * \alpha_1 + u_{2i} * \alpha_3 + u_{3i} * \alpha_3, \\
 R_b &= b_1 * \alpha_4 + b_2 * \alpha_5 + b_3 * \alpha_6.
 \end{aligned} \tag{13}$$

Let R_0 ($0 < R_0 \leq 1$) be the threshold of the reward function and R and C be the corrected amount of the rule reward degree k (i.e., r_k) made by the rule r . Then

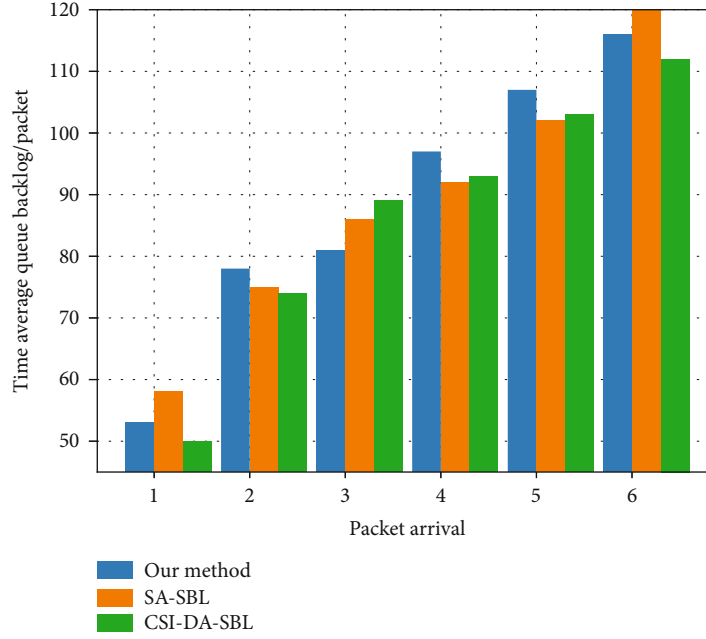


FIGURE 8: Comparison of the average pairwise backlog for different virtualization scenarios.

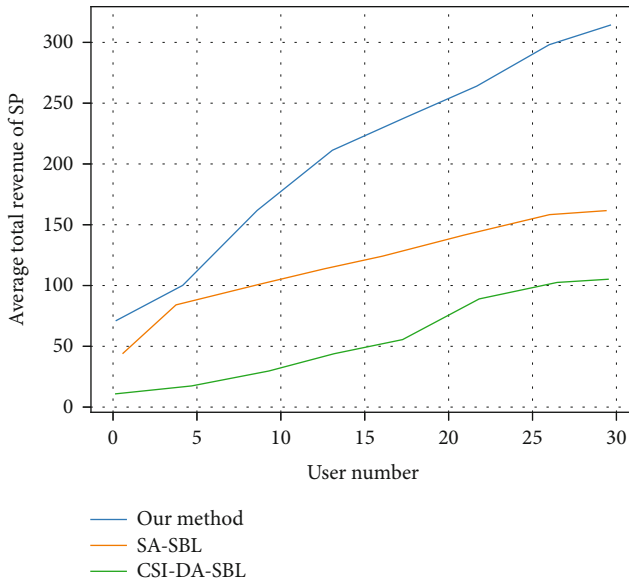


FIGURE 9: Comparison of average SP returns for different backhaul scenarios.

- (1) If $R \geq R_0$, then, the rule reward degree $r_k = r_k + C$; rule r making α_k under this E shows an increasing trend
- (2) If $R < R_0$, then, the rule reward degree $r_k = r_k - C$; rule r making α_k under this E shows a weakening trend

The agent will determine the next rule selection action based on the magnitude of the rule reward degree value.

Therefore, the autonomous learning update strategy model of reinforcement learning can achieve a closed-loop

resource scheduling mechanism, which integrates the influence of multiple factors to better update knowledge decisions and achieve reliable and satisfactory resource allocation.

6. Simulation Results and Analysis

In this subsection, the proposed algorithm is simulated and validated. we considers a physical network coverage area for three SPs to provide resource sharing services, where the physical network consists of one macrobase station and three small base stations with self-backhaul capabilities, and the macrobase station is fixed at the center of the area, and the locations of the small base stations and the users are randomly deployed. The available bandwidth in the system and the transmit power of the user and the small base station are 20 dBm and 33 dBm, respectively. We evaluate the performance of the algorithm over $T = 1000$ cycles, with each cycle time slot length set too.

First, the performance of our proposed Lyapunov optimization-based algorithm is evaluated. Figure 4 plots the average total SP gain and time-averaged queue backlog versus the control parameters. As shown in the figure, the average total SP gain gradually increases and plateaus as it increases. In Lyapunov optimization, the larger the control parameter, the more the system tends to optimize the penalty function, which in this paper is optimizing the SP average total return. However, as it increases, the time-averaged queuing backlog increases almost linearly, due to the interconnection between the system revenue and the time delay. Therefore, in order to make the system work in the ideal state, a reasonable selection of control parameters is required.

In addition, to visualize the average queue backlog of users in consecutive times and to demonstrate the

convergence of queue stability, we arbitrarily selected three users belonging to different SPs. As shown in Figure 5, the horizontal coordinate is the time index and the vertical coordinate represents the time-averaged queue backlog, from which it can be seen that the average cache queue of SP users increases first and then gradually tends to smooth out, which also implies that the Lyapunov optimization-based resource allocation algorithm proposed in this paper can effectively guarantee the stability of the system queue; in other words, it also confirms the effectiveness of our proposed algorithm.

Secondly, we validate the performance of the proposed algorithm on wireless network virtualization and compare it with two algorithms, namely, the static resource allocation (SA-SBL) strategy and the user CSI-based dynamic resource allocation (CSI-DA-SBL) strategy. For SA-SBL, the virtual layer allocates a fixed number of virtual resources to each SP and does not vary over time; for CSI-DA-SBL, the virtual layer dynamically schedules resources based on the channel quality of the SP users with the goal of maximizing system capacity.

Figure 6 shows the variation of the average SP revenue for different numbers of access users. x -axis represents the number of users accessed by the system, and y -axis represents the time-averaged total revenue. As the number of subscribers increases, the average revenue of the SP also tends to increase, but at a slower rate. This is because even though access to more users generates more revenue for the SP, the SP has to pay relatively high resource rental costs to the In P. In addition, the virtual network manager will allocate the optimal amount of resources to each SP rather than the full amount available, so the SP's net revenue will be somewhat limited. The figure also shows that our proposed algorithm significantly outperforms SA-SBL and CSI-DA-SBL because our proposed algorithm jointly considers the channel quality of the users and the utility of the SPs and aims to maximize the total system revenue by allocating the right amount of virtual band resources to different SPs. In contrast, SA-SBL allocates a fixed amount of resources to each SP regardless of whether the SP needs it or not, resulting in a waste of resources and therefore no significant improvement in the SP's revenue in this scheme.

Figure 7 shows the relationship between the average utility of users and the number of access users in different scenarios, where the user utility is the amount paid by the user to the SP. It can be seen from the figure that the average user utility of the proposed algorithm tends to decrease as the number of access users increases. This is because the more dense the system is, the proportion of users with poor channel conditions will increase, and in order to ensure the delay performance of these users, the proposed algorithm will allocate relatively more resources to them to avoid an infinite queue backlog; thus, the average rate obtained by all users in the system will be slightly reduced.

Figure 8 depicts the average queue backlog versus packet arrival rate in different scenarios. For an arbitrary average packet arrival rate, our proposed virtual resource allocation scheme has a lower queue backlog. This is because our proposed algorithm dynamically allocates resources at each time slot in conjunction with the user's cache state, thus ensuring

that all queues in the system are smoothed out, and thus, our proposed algorithm has better latency performance.

Finally, we tested the performance of the backhaul mechanism and compared two backhaul schemes, namely, the wireless self-backhaul (FBB) mechanism with a fixed band cut ratio and the wired backhaul (WB) mechanism, where in the FBB scheme, the band ratio between the access and backhaul links is fixed for each time slot, while the traditional WB mechanism enables small base stations to transmit data from associated users over fiber or digital subscriber lines to macrobase station or core network. Next, the performance of the different schemes on backhaul will be evaluated in terms of both SP average total utility and In P average utility.

As can be seen from Figure 9, our proposed dynamic in-band self-backhaul mechanism can bring more benefits to both SPs and In P. Since the FBB mechanism uses a statically configured band cut on the radio backhaul and access links, the self-backhauling small base stations are unable to match the backhaul and access capacity, wasting a certain amount of band resources and reducing the benefits for In P and SP. Although WB can provide larger backhaul capacity, the cost of backhaul will also increase, and In P's backhaul revenue will be used to build backhaul equipment, so the SP and In P revenue in the WB mechanism is significantly lower than the other two options.

7. Conclusions

In the face of increasingly complex network environments and the scarcity of wireless resources, there is an urgent need for new scalable wireless network technologies, and network virtualization technology has great potential. We propose a machine learning-based resource scheduling mechanism for wireless virtual networks and design an intelligent optimization module for user service preferences in response to the personalized needs of user service customization. The results show that the scheme in this paper can effectively improve indicators such as resource scheduling rationality, link resource utilization, and user satisfaction.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- [1] H. Wang, Y. Wu, G. Min, J. Xu, and P. Tang, "Data-driven dynamic resource scheduling for network slicing: a deep reinforcement learning approach," *Information Sciences*, vol. 498, pp. 106–116, 2019.

- [2] Y. Wang, H. Liu, W. Zheng et al., "Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning," *IEEE Access*, vol. 7, pp. 39974–39982, 2019.
- [3] S. Vimal, M. Khari, N. Dey, R. G. Crespo, and Y. H. Robinson, "Enhanced resource allocation in mobile edge computing using reinforcement learning based MOACO algorithm for IIOT," *Computer Communications*, vol. 151, pp. 355–364, 2020.
- [4] R. Atallah, C. Assi, and M. Khabbaz, "Deep reinforcement learning-based scheduling for roadside communication networks," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 1–8, Paris, France, 2017.
- [5] L. Gu, D. Zeng, W. Li, S. Guo, A. Y. Zomaya, and H. Jin, "Intelligent VNF orchestration and flow scheduling via model-assisted deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 2, pp. 279–291, 2020.
- [6] Y. Wei, F. R. Yu, M. Song, and Z. Han, "User scheduling and resource allocation in Het nets with hybrid energy supply: an actor-critic reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 680–692, 2018.
- [7] Z. H. Zhengwan, Z. H. Chunjong, L. I. Hongbing, and X. I. Tao, "Multipath transmission selection algorithm based on immune connectivity model," *Journal of Computer Applications*, vol. 40, no. 12, p. 3571, 2020.
- [8] Y. Qian, J. Wu, R. Wang, F. Zhu, and W. Zhang, "Survey on reinforcement learning applications in communication networks," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 30–39, 2019.
- [9] Q. Zhiguo, H. Sun, and M. Zheng, "An efficient quantum image steganography protocol based on improved EMD algorithm," *Quantum Information Processing*, vol. 20, no. 53, pp. 1–29, 2021.
- [10] Q. Zhiguo, S. Chen, and X. Wang, "A secure controlled quantum image steganography algorithm," *Quantum Information Processing*, vol. 19, no. 10, pp. 1–25, 2020.
- [11] J. Yang, X. You, G. Wu, M. M. Hassan, A. Almogren, and J. Guna, "Application of reinforcement learning in UAV cluster task scheduling," *Future Generation Computer Systems*, vol. 95, pp. 140–148, 2019.
- [12] A. I. Orhean, F. Pop, and I. Raicu, "New scheduling approach using reinforcement learning for heterogeneous distributed systems," *Journal of Parallel and Distributed Computing*, vol. 117, pp. 292–302, 2018.
- [13] I. S. Comşa, S. Zhang, M. E. Aydin et al., "Towards 5G: a reinforcement learning-based scheduling solution for data traffic management," *IEEE Transactions on Network and Service Management*, vol. 15, no. 4, pp. 1661–1675, 2018.
- [14] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 9, pp. 3391–3402, 2021.
- [15] X. Ning, K. Gong, W. Li, and L. Zhang, "WSAA: joint weak saliency and attention aware for person re-identification," *Neurocomputing*, vol. 453, pp. 801–811, 2021.
- [16] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang, and L. C. Wang, "Deep reinforcement learning for mobile 5G and beyond: fundamentals, applications, and challenges," *IEEE Vehicular Technology Magazine*, vol. 14, no. 2, pp. 44–52, 2019.
- [17] Z. Ning, K. Zhang, X. Wang et al., "Joint computing and caching in 5G-envisioned internet of vehicles: a deep reinforcement learning-based traffic control system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5201–5212, 2021.
- [18] A. Elgabli, H. Khan, M. Krouka, and M. Bennis, "Reinforcement learning based scheduling algorithm for optimizing age of information in ultra reliable low latency networks," in *2019 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–6, Barcelona, Spain, 2019.
- [19] Z. Ning, P. Dong, X. Wang, J. J. Rodrigues, and F. Xia, "Deep reinforcement learning for vehicular edge computing: an intelligent offloading system," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 6, pp. 1–24, 2019.
- [20] J. Prados-Garzon, T. Taleb, and M. Bagaa, "LEARNET: reinforcement learning based flow scheduling for asynchronous deterministic networks," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, Dublin, Ireland, 2020.
- [21] J. Wang, J. Hu, G. Min, W. Zhan, Q. Ni, and N. Georgalas, "Computation offloading in multi-access edge computing using a deep sequential model based on reinforcement learning," *IEEE Communications Magazine*, vol. 57, no. 5, pp. 64–69, 2019.
- [22] H. Che and J. Wang, "Bicriteria sparse nonnegative matrix factorization via two-timescale duplex neurodynamic optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 3, 2021.
- [23] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to mixed-integer optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 36–48, 2021.
- [24] X. Chen, C. Wu, T. Chen et al., "Age of information aware radio resource management in vehicular networks: a proactive deep reinforcement learning perspective," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2268–2281, 2020.
- [25] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4005–4018, 2019.
- [26] R. F. Atallah, C. M. Assi, and J. Y. Yu, "A reinforcement learning technique for optimizing downlink scheduling in an energy-limited vehicular network," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 4592–4601, 2017.
- [27] Q. Luo, C. Li, T. H. Luan, and W. Shi, "Collaborative data scheduling for vehicular edge computing via deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9637–9650, 2020.