Check for updates

Received: 5 May 2022 | Revised: 11 December 2022 | Accepted: 13 December 2022

DOI: 10.1002/ett.4722

RESEARCH ARTICLE

WILEY

(Network value)-based adaptive dynamic bandwidth allocation algorithm for 5G network slicing

Lijuan Wu | **Chaoqin Gan** | **Zhongsen Xu** | **Jianqiang Hui**

Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, China

Correspondence
Chaoqin Gan, Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, 200444, China.
Email: cqgan@shu.edu.cn

Funding information
Higher Education Discipline Innovation Project; National Key Research and Development Program of China, Grant/Award Number: 2021YFB2900800; Science and Technology Commission of Shanghai Municipality, Grant/Award Numbers: 20511102400, 20ZR1420900; 111 project, Grant/Award Number: D20031

Abstract

In this article, a (network value)-based inter-slice and intra-slice adaptive dynamic bandwidth allocation algorithm is proposed for the coexistence scenario of URLLC slices, eMBB slices, and mMTC slices in 5G fronthaul networks. The network value model applicable to network slicing is constructed to provide a means for the value evaluation of virtual networks. According to the bandwidth requirements of the slices, the adaptive inter-slice bandwidth allocation is realized to ensure that the fronthaul network can obtain the maximum economic value under different load states. For the service characteristics of three slices, three intra-slice bandwidth allocation algorithms are proposed, respectively, to realize adaptive ONU bandwidth allocation. It not only meets the differentiated service demands of different slices, but also enables each slice to obtain the maximum QoS value. By simulation and analysis, the effectiveness of the proposed algorithm is demonstrated.

1 | INTRODUCTION

With the booming development of internet technology and smart terminal devices, network demand is gradually evolving towards diversification. Emerging applications such as smart home, smart transportation, virtual reality, and industrial industrial Internet of Things (IoT) require networks to provide differentiated services. In order to meet the needs of 5G vertical industry development, the traditional one-size-fits-all network architecture is no longer applicable, and network slicing technology comes into being.[1] Network slicing divides a physical network into multiple isolated logical networks to satisfy the demands of different services, and has become one of the key technologies of 5G network.[2-5] Therefore, it is important to adopt network slicing technology to solve the multiservice problem of 5G fronthaul network, especially to meet the differentiated demand of services in URLLC (ultra-reliable and low-latency communications), eMBB (enhanced mobile broadband) and mMTC (massive machine-type communications) scenarios.

Passive optical network (PON), which has the advantages of high capacity, low deployment cost, strong scalability, and flexible resource allocation, is regarded as the most cost-effective solutions for 5G fronthaul networks.[6] In recent years, several studies focus on the combination of PON and network slicing. The Reference 7 proposed a unified control

 © 2022 John Wiley & Sons, Ltd.

network slicing architecture for 5G fronthaul networks based on multi-tenant multi-standard PONs. The Reference 8 integrated MFH (mobile fronthaul) slice, IoT slice and control slice into single TDM-PON (time division multiplexing PON). The Reference 9 assigned wavelengths to optical network units (ONUs) based on delay sensitivity, and ONUs of the same wavelength in PON were constructed on the same slice. Moreover, References 10,11 constructed slices with different delay requirements in TDM-PON and performed different delay-guaranteed scheduling schemes on these slices. Virtual passive optical network (VPON) virtualizes the resources in PON, shares wavelength among multiple users, and can be networked on demand. In a PON-based fronthaul network, a VPON represents a network slice.[12] Different slices share the resources of the network, which enables dynamic allocation of resources.

Bandwidth allocation includes static and dynamic allocation. Static bandwidth allocation (SBA) allocates fixed bandwidth to users, which improves the efficiency of bandwidth allocation to a certain extent, but is not conducive to the sharing of bandwidth resources, and will cause great waste of resources. Dynamic bandwidth allocation (DBA) dynamically calculates the bandwidth allocated to users according to the service volume, which can effectively improve the bandwidth utilization of the system. The dynamic bandwidth allocation algorithm for network slicing in 5G is a current research hotspot. The Reference 8 proposes a DBA algorithm to realize the convergence of fronthaul network slicing and IoT slicing. The algorithm uses different methods to perform uplink bandwidth allocation on different slices, which can meet the needs of both slices. The Reference 13 performs mixed allocation of cloud resources and bandwidth resources for URLLC slices, eMBB slices and mMTC slices according to the comprehensive evaluation factor, and adopts a preemptive bandwidth allocation strategy to meet the delay requirements of URLLC slices. The Reference 14 proposes a unified radio access network (RAN) slice service provisioning framework, which guarantees the quality of service (QoS) of users by allocating the minimum requested bandwidth. These studies only consider the impact of bandwidth allocation between slices or bandwidth allocation within slices on network service quality alone. They do not perform differentiated slice bandwidth allocation from an end-to-end perspective, nor do they consider network costs in bandwidth allocation. In this article, we propose an inter-slice and intra-slice adaptive DBA algorithm from the perspective of maximizing network profit and maximizing slice QoS performance for the coexistence scenario of URLLC, eMBB, and mMTC slices.

The rest of this article is organized as follows: In Section 2, the system architecture is introduced. In Section 3, a network value model is constructed. In Section 4, an adaptive dynamic bandwidth allocation algorithm is proposed. The simulation results are displayed to confirm the rationality and effectiveness of the proposed algorithm in Section 5. Finally, we conclude the article in Section 6.

2 | SYSTEM ARCHITECTURE

The 5G fronthaul network architecture based on network slicing is shown in Figure 1. In this architecture, the fronthaul physical network consists of a PON. The ONU in the PON is directly connected to the radio unit (RU) and is responsible for receiving users' services. The optical line terminal (OLT) is connected to the distributed unit (DU) and is responsible for controlling the access and resource allocation of the ONU. The OLT and the ONU are connected through the optical distribution unit. This optical fronthaul physical network is logically divided into mutually isolated slices, namely, URLLC slice, eMBB slice, and mMTC slice. URLLC slices are mainly responsible for services with high latency requirements, such as autonomous driving and telemedicine. eMBB slices are responsible for large bandwidth services, such as high-definition video. mMTC slices are responsible for massive connectivity services, such as sensor networks, industrial IoT. Each slice composed of a VPON. In logical networks, the OLT's functions are abstracted into standard computing and storage resources through network function virtualization (NFV) technology.[15] Here, resource abstraction is the representation of a resource in terms of attributes that match predefined selection criteria while hiding or ignoring aspects that are irrelevant to such criteria, in an attempt to simplify the use and management of that resource in some useful way.[16] After that, the abstracted resources are unified and integrated into servers, storage devices and switches. Finally, specific functions are provided externally in the form of software, that is, functions are provided externally through multiple virtual OLTs (vOLTs). These vOLTs can share computing and storage resources in the OLT pool. Each vOLT is responsible for handling one type of ONU service. The vOLT, the ONUs corresponding to the vOLT, and the virtual link between the vOLT and the ONU constitute a VPON. Different VPONs transmit different ONU services. This is similar to T-CONT, but VPON is more flexible in classifying services than T-CONT, and VPON can serve both GPON and EPON. As shown in Figure 1, VPON1, VPON2, and VPON3 transmit URLLC, eMBB, and mMTC services, respectively. In the VPON, ONUs can exclusively use wavelength resources or share wavelength resources. The vOLT collects request information from

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

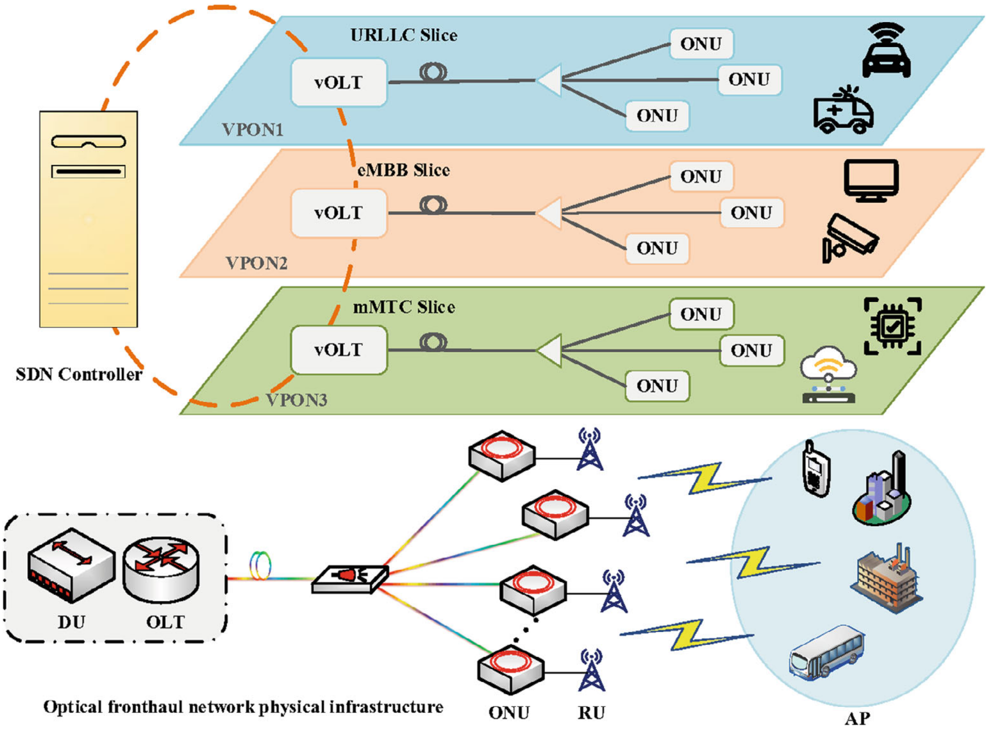

FIGURE 1 Slice-based 5G fronthaul network architecture

ONUs by establishing dynamic logical connections with ONUs, and performs intra-slice bandwidth allocation algorithm to complete the data transmission of the ONU. The resource allocation among the three slices is done by the software definition network (SDN) controller. The SDN controller receives the slice request information uploaded by the vOLT and performs global resource calculation to complete dynamic bandwidth allocation among different slices. When the SDN controller and the vOLT perform bandwidth allocation, the network value will be the evaluation indicator. Through this two-level resource management architecture of SDN-vOLT-ONU, the resource allocation of network slices will be more flexible and efficient.

3 | NETWORK VALUE MODEL

Based on our previous work, we define a new concept for network slicing to evaluate network performance, namely, network value.[17] Network value consists of two parts: economic value and QoS value. The economic value measures the profit of the network, and the QoS value measures the network service quality from the perspective of user perception. Economic value is determined by network revenue and network cost. Network revenue is the product of total user bandwidth and the unit price of bandwidth. Network costs include capital expenditures (CAPEX) and operating expenses (OPEX). Since network slices are virtual networks, the costs of all physical devices in the network are uniformly translated into the cost of building network slices. The CAPEX of the network is averaged into the cost of building slices per unit bandwidth size, that is, the CAPEX is calculated by multiplying the slice size with the slice unit cost. OPEX is the cost of transmitting bandwidth, which is calculated by multiplying the total bandwidth transmitted with the cost per unit of bandwidth transmitted. The economic value is calculated as follows:

$$V_E = \sum_{i=1}^{3} B_i * \alpha_i - \left(\sum_{i=1}^{3} C_i * \rho_0 + \sum_{i=1}^{3} B_i * \rho_1 \right). \tag{1}$$

In the above equation, the first term represents the network revenue. B_i represents the bandwidth allocated to users in slice i (i equal to 1 means URLLC slice, equal to 2 means eMBB slice, equal to 3 means mMTC slice), and α_i represents the bandwidth unit price of slice i. The second term represents the network cost. C_i denotes the capacity of slice i, ρ_0

2161915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

denotes the cost of constructing a unit capacity slice, and ρ_1 denotes the cost of transmitting a unit bandwidth. ρ_1 is related to the link load condition within a slice. When the load is too large, the links tend to cause congestion and the cost of transmission bandwidth becomes higher. The value of ρ_1 is calculated by the following equation:

$$\rho_1 = \begin{cases} \rho_{\min}, & \frac{B_i}{C_i} \le \theta_1, \\ \frac{\rho_{\max}-\rho_{\min}}{\theta_2-\theta_1}\left(\frac{B_i}{C_i}-\theta_1\right)+\rho_{\min}, & \theta_1 \le \frac{B_i}{C_i} \le \theta_2, \\ \rho_{\max}, & \frac{B_i}{C_i} \ge \theta_2, \end{cases} \tag{2}$$

where $\rho_{\min}$ and $\rho_{\max}$ represent the minimum and maximum cost of transmission unit bandwidth, respectively. θ_1 and θ_2 are slice resource utilization thresholds. ρ_1 is a piecewise function, and the inter-slice bandwidth allocation will be developed around this function.

QoS value is measured by user perception and contains four metrics: delay satisfaction, bandwidth satisfaction, jitter satisfaction, and packet loss rate satisfaction. Referring to References 18,19, the satisfaction function is represented by a sigmoid function. The sigmoid function can flexibly control the shape of the function by adjusting the parameters, so that it conforms to the influence trend of the QoS metrics on the network performance. The formulas for delay satisfaction, bandwidth satisfaction, jitter satisfaction and packet loss rate satisfaction are, respectively, shown in Equations (3)-(6):

$$Q_d = \frac{1+e^{-\delta d_m}}{1+e^{-\delta(d_m-d)}}, \tag{3}$$

$$Q_{bs} = \frac{2}{1+e^{\delta(b_m-b)}}, \tag{4}$$

$$Q_j = \frac{1+e^{-\delta d_{jm}}}{1+e^{-\delta(d_{jm}-d_j)}}, \tag{5}$$

$$Q_{pl} = \frac{2}{1+e^{\delta l}}. \tag{6}$$

In the above formula, d_m, b_m, and d_{jm} represent the user's expected delay threshold, requested bandwidth, and expected jitter threshold, respectively. d, b, d_j, and l denote the network delay, bandwidth allocated to users, delay jitter and packet loss rate, respectively. δ is a parameter that controls the shape of the curve. The larger δ is, the more the sigmoid function tends to be a step function. The four satisfaction functions can be intuitively shown in Figure 2. It can be seen from the figure that the delay satisfaction and jitter satisfaction are typical sigmoid functions. When the network jitter or delay is less than the threshold, both delay satisfaction and jitter satisfaction are relatively large. However, once the jitter and delay are larger than the threshold, both delay satisfaction and jitter satisfaction drop rapidly. The packet loss rate satisfaction is a monotonically decreasing function. When the packet loss rate is 0, the satisfaction is 1. Satisfaction decreases rapidly when the packet loss rate starts to increase. In this process, satisfaction decreases in a large and then small way until it drops to 0. In contrast to packet loss satisfaction, bandwidth satisfaction is a monotonically increasing function. When the allocated bandwidth is much smaller than the requested bandwidth, the bandwidth satisfaction tends to 0. As the allocated bandwidth increases, the bandwidth satisfaction increases more and more. Bandwidth satisfaction is 1 only if the allocated bandwidth is equal to the requested bandwidth.

The QoS value is the weighted sum of the above four satisfaction functions:

$$V_{\text{QoS}} = \sum_{i \in \{d,bs,j,pl\}} q_i Q_i, \tag{7}$$

where q_i represents the weight of the indicator. Different slices have different weights of indicators. Here, the four weight values that contribute to the QoS value from high to low calculated by the analytic hierarchy process (AHP) are 0.4750, 0.2684, 0.1546, 0.1020. According to the requirements of each slice, the QoS value of URLLC slice, eMBB slice, and mMTC slice can be defined as follows:

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

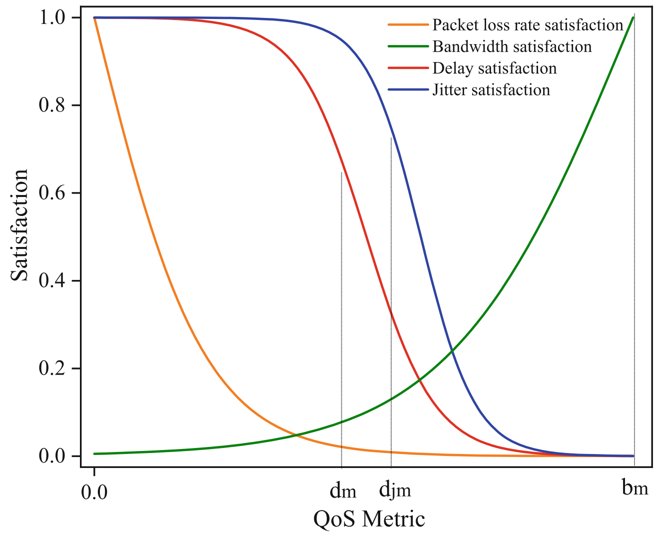

FIGURE 2 Satisfaction functions

$$V_{\text{QoS}}^{\text{URLLC}} = 0.4570Q_d + 0.2684Q_j + 0.1546Q_{pl} + 0.1020Q_{bs}, \tag{8}$$

$$V_{\text{QoS}}^{\text{eMBB}} = 0.4570Q_{bs} + 0.2684Q_{pl} + 0.1546Q_d + 0.1020Q_j, \tag{9}$$

$$V_{\text{QoS}}^{\text{mMTC}} = 0.4570Q_{pl} + 0.2684Q_{bs} + 0.1546Q_d + 0.1020Q_j, \tag{10}$$

The network value of the slice is the weighted sum of economic value and QoS value, that is:

$$V = v_1 V_E + v_2 V_{\text{QoS}}, \tag{11}$$

where v_1 and v_2 are the weights of economic value and QoS value, respectively. And V_E is the normalized value.

4 | ADAPTIVE DYNAMIC BANDWIDTH ALLOCATION ALGORITHM

The adaptive dynamic bandwidth allocation algorithm for 5G fronthaul slices proposed in this article consists of two parts: one is inter-slice bandwidth allocation algorithm, and the other is intra-slice bandwidth allocation algorithm. Among them, the intra-slice bandwidth allocation algorithm includes three parts, namely, URLLC slice bandwidth allocation, eMBB slice bandwidth allocation, and mMTC slice bandwidth allocation. The flowchart of the proposed adaptive dynamic bandwidth allocation algorithm is shown in Figure 3. Next, the inter-slice bandwidth allocation algorithm and the intra-slice bandwidth allocation (ISBA) algorithm are introduced in detail.

4.1 | Inter-slice bandwidth allocation algorithm

The main goal of inter-slice bandwidth allocation is to maximize the economic value of the network. Therefore, the bandwidth allocation algorithm in this section mainly revolves around the economic value expression.

When the total network requested bandwidth is less than the network capacity, the bandwidth allocated to users is equal to the requested bandwidth, that is, $B_i = R_i$. In this case, the demand of all slices can be satisfied. Since the slice capacity size and link resource utilization will affect the network cost, the optimal slice bandwidth allocation scheme needs to be further calculated. First, calculate the bandwidth allocation values of the three slices that maximize the economic value of the network in different resource utilization intervals. Then, compare the network economic value

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

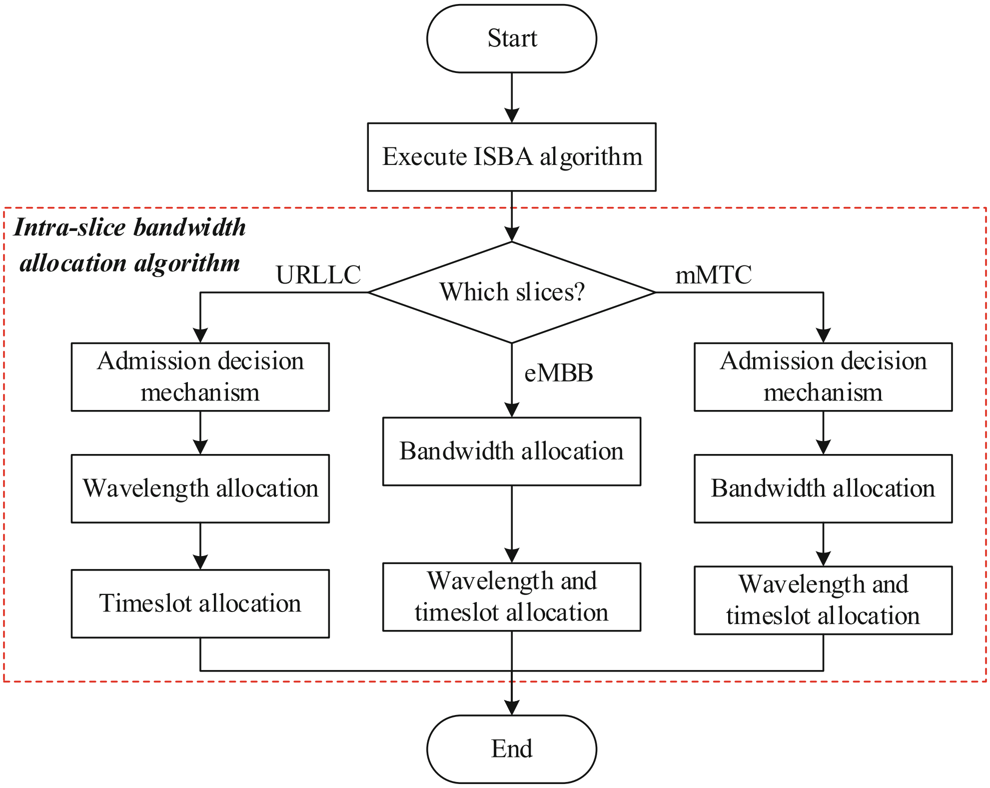

FIGURE 3 The flowchart of adaptive dynamic bandwidth allocation algorithm

calculated by the bandwidth allocation schemes in different intervals. The scheme with the greatest network economic value is the optimal slice bandwidth allocation scheme. The detailed solution process is as follows:

1. When $\frac{B_i}{C_i} \leq \theta_1$, ρ_1 is equal to $\rho_{\min}$. The economic value of the network is:

$$V_E = \sum_{i=1}^{3} B_i * \alpha_i - \left(\sum_{i=1}^{3} C_i * \rho_0 + \sum_{i=1}^{3} B_i * \rho_{\min} \right), \tag{12}$$

$$\text{s.t. } C_i \geq B_i, \quad \forall i. \tag{13}$$

It can be seen from formula (12) that V_E is inversely proportional to C_i. Therefore, in this interval, the network can obtain the maximum economic value when $C_i = \frac{B_i}{\theta_1}$. The maximum value is:

$$V_{E\max 1} = \sum_{i=1}^{3} B_i * \left(\alpha_i - \frac{\rho_0}{\theta_1} - \rho_{\min} \right). \tag{14}$$

2. When $\theta_1 \leq \frac{B_i}{C_i} \leq \theta_2$, ρ_1 is equal to $\frac{\rho_{\max}-\rho_{\min}}{\theta_2-\theta_1}\left(\frac{B_i}{C_i} - \theta_1\right) + \rho_{\min}$. The economic value of the network is:

$$V_E = \sum_{i=1}^{3} B_i * \alpha_i - \left[\sum_{i=1}^{3} C_i * \rho_0 + \sum_{i=1}^{3} B_i * \left(\frac{\rho_{\max} - \rho_{\min}}{\theta_2 - \theta_1} \left(\frac{B_i}{C_i} - \theta_1 \right) + \rho_{min} \right) \right]. \tag{15}$$

Derive the above equation to get:

$$\frac{\mathrm{d}V_E}{\mathrm{d}C_i} = -\rho_0 + \frac{\rho_{\max} - \rho_{\min}}{\theta_2 - \theta_1} \cdot \frac{B_i^2}{C_i^2}. \tag{16}$$

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Let $\frac{dV_E}{dC_i} = 0$, we get $C_i = B_i \cdot \sqrt{\frac{\rho_{\max}-\rho_{\min}}{(\theta_2-\theta_1)\rho_0}}$. Then it is necessary to judge whether C_i satisfies the condition of $\theta_1 \le \frac{B_i}{C_i} \le \theta_2$. When it is satisfied, the maximum economic value can be obtained by substituting C_i into Equation (15), which is expressed as $V_{E\max 2}$.

3. When $\frac{B_i}{C_i} \ge \theta_2$, ρ_1 is equal to $\rho_{\max}$. The economic value of the network is:

$$V_E = \sum_{i=1}^{3} B_i * \alpha_i - \left(\sum_{i=1}^{3} C_i * \rho_0 + \sum_{i=1}^{3} B_i * \rho_{\max} \right). \tag{17}$$

The above equation also needs to satisfy the condition of Equation (13), so $B_i \le C_i \le \frac{B_i}{\theta_2}$. In this interval, the network can obtain the maximum economic value when $C_i = B_i$. The maximum value is:

$$V_{E\max 3} = \sum_{i=1}^{3} B_i * (\alpha_i - \rho_0 - \rho_{\max}). \tag{18}$$

Finally, the values of $V_{E\max 1}$, $V_{E\max 2}$, and $V_{E\max 3}$ are compared, and the maximum value is the final network economic value. Meanwhile, the corresponding C_i is the bandwidth allocated for slice i.

When the total requested bandwidth of the network is greater than the network capacity, the total bandwidth allocated to users within each slice is the capacity of the slice, that is, $B_i = C_i$. And the resource utilization rate of all slices is greater than θ_2, so $\rho_1 = \rho_{\max}$. In this case, the economic value of the network is:

$$\begin{aligned} V_E &= \textstyle\sum_{i=1}^{3} C_i * \alpha_i - \left(\sum_{i=1}^{3} C_i * \rho_0 + \sum_{i=1}^{3} C_i * \rho_{\max} \right) \\ &= \textstyle\sum_{i=1}^{3} C_i * (\alpha_i - \rho_0 - \rho_{\max}). \end{aligned} \tag{19}$$

Since ρ_0 and $\rho_{\max}$ are constant, the value of $\alpha_i - \rho_0 - \rho_{\max}$ depends on α_i. Arrange in descending order of α_i, and perform bandwidth allocation on slice i in turn. Let the remaining network bandwidth be C_{remain}, and the requested bandwidth of slice i ($i \in \{1, 2, 3\}$, 1 means URLLC slice, 2 means eMBB slice, 3 means mMTC slice) be R_i.

When $C_{\text{remain}} \ge R_i$, C_i is eaual to R_i. Update C_{remain} to:

$$C_{\text{remain}} = C_{\text{remain}} - R_i. \tag{20}$$

When $C_{\text{remain}} < R_i$, C_i is eaual to C_{remain}. If the slices have the same α_i, to ensure the fairness among slices, slice bandwidth is allocated in proportion to the requested bandwidth, that is:

$$C_i = \frac{R_i}{\sum R_i} * C_{\text{total}}, \tag{21}$$

where C_{total} is network capacity.

After completing the inter-slice bandwidth allocation, the number of wavelengths for slices needs to be determined based on the bandwidth of each slice. Here, exclusive wavelengths and shared wavelengths are established for flexible inter-slice bandwidth adjustment. The exclusive wavelength of each slice is used only by that slice, while the shared wavelength can be shared by multiple slices. Let the capacity of one wavelength be C_λ. The number of exclusive wavelengths for slice i ($i \in \{1, 2, 3\}$, 1 means URLLC slice, 2 means eMBB slice, 3 means mMTC slice) is:

$$\text{wave}_i = \left\lfloor \frac{C_i}{C_\lambda} \right\rfloor. \tag{22}$$

The total number of shared wavelengths is:

$$\text{wave}_{\text{share}} = \text{wave} - \sum \text{wave}_i, \tag{23}$$

where wave represents the total number of wavelengths in the network. In Equation (22), the round-down function is used to determine the number of exclusive wavelengths for each slice, and these wavelengths cannot be shared. In

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Equation (23), the remaining bandwidth of all slices is allocated to the shared wavelength. This allocation scheme can make full use of wavelengths and improve bandwidth utilization.

4.2 | URLLC slice bandwidth allocation algorithm

The main goal of intra-slice bandwidth allocation is to maximize the QoS value of the slice. Different slices have different requirements for QoS metrics, and URLLC slices require high reliability and low latency performance. Therefore, the objective of bandwidth allocation for URLLC slices is to maximize packet loss satisfaction and delay satisfaction. For packet loss satisfaction, an ONU admission decision mechanism is proposed to reduce the packet loss rate. For delay satisfaction, a slot allocation algorithm considering ONU transmission distance is proposed to reduce the average packet delay.

4.2.1 | ONU admission decision mechanism

According to the delay requirements of ONU services, each ONU in the URLLC slice is defined with a priority p_i. The larger the p_i, the lower the delay required by the service. First, the ONUs are sorted in descending order by p_i and placed in queue. Then, based on the current queue order, the ONUs are sorted by request bandwidth in descending order. It means that when ONUs have the same priority, the one with the largest request bandwidth is placed at the front of the queue. This is to ensure that the total requested bandwidth of all admitted ONUs is the largest. Finally, traverse from the head of the queue. If the requested bandwidth of the current ONU is less than the remaining bandwidth of the slice, the ONU is admitted and the remaining bandwidth value is updated. Otherwise, the ONU is not admitted. Continue traversing the queue backward until all ONUs are accessed, or the network remaining bandwidth is zero. For all admitted ONUs, the bandwidth allocated to them is their requested bandwidth. The pseudo-code of the ONU admission decision and bandwidth allocation algorithm is shown in Table 1.

TABLE 1 Pseudo code for ONU admission decision and bandwidth allocation

Algorithm 1: ONU admission decision and bandwidth allocation
Input: $onu[]$, $R[]$, C
Output: $access[]$, $B[]$
function Access ($onu[]$, $R[]$, C)
$remain = C$
sort $onu[]$ according to priority and request bandwidth from large to small
for $i = 0 \rightarrow N-1$ **do**
if $R[i] \leq remian$ **do**
$access[i] = 1$
$B[i] = R[i]$
$remain = remain - R[i]$
else if $remain \leq 0$ **do**
break
else
$i += 1$
return $access[]$, $B[]$
end function

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.2.2 | Wavelength and time slot allocation

After completing the admission decision for ONUs, wavelength and time slot allocation is performed for the admitted ONUs. For the wavelength allocation, the bandwidth of ONUs is first arranged in descending order. Then ONUs are sequentially taken out of the queue and assigned to the earliest available wavelength for transmission. It should be noted that the bandwidth transmitted on the shared wavelength is limited. The maximum transmittable bandwidth can be calculated by:

$$\text{maxBandwidth}_{\text{share}} = \min\left(\sum R_i, C_1\right) - \text{wave}_i * C_\lambda. \tag{24}$$

Once the wavelength assignment is completed, the time slot assignment of the ONU can be performed. The goal of this phase is to maximize the delay satisfaction, as shown in Equation (25). Here, the end-to-end delay of the ONU only considers the queuing delay d_q and the propagation delay d_p. The queuing delay is determined by the start transmission time of the ONU, which is related to the queue size within the current buffer and the channel bandwidth.[20] It can be calculated by Equation (27).

$$\max: \sum \frac{1 + e^{-\delta d_m}}{1 + e^{-\delta(d_m - d_i)}}, \tag{25}$$

$$d_i = d_{qi} + d_{pi}, \tag{26}$$

$$d_q = \frac{\sum L_{\text{pkt}} * 8}{C_\lambda}, \tag{27}$$

where L_{pkt} is the size of a packet in Byte, $\sum L_{\text{pkt}}$ represents the total packet size of the buffer, multiplied by 8 to convert it into bits, and C_λ is the bandwidth of the channel. The average queuing delay of ONU_i can be regarded as the time waiting for the first $i-1$ ONUs in the queue to complete the transmission. The propagation delay is related to the transmission distance, which can be calculated by the following formula:

$$d_p = \frac{\text{dis}}{r} \tag{28}$$

Where dis is the distance between ONU and OLT, r represents the propagation speed of fiber, generally taken as 2×10^8 m/s.

The time slots assigned to ONUs affect the delay of the ONU. To maximize the delay satisfaction of slices, the delay of each ONU should be smaller than the delay threshold as much as possible. Therefore, in the process of allocating time slots to the ONU, the delay threshold, the propagation delay and the queuing delay should be comprehensively considered. Here, a dynamic programming algorithm is used to complete the time slot allocation of the ONU. The optimal slice satisfaction of each slot is transformed into the optimal slice satisfaction of the previous slot plus the maximum ONU delay satisfaction of the current slot. The detailed solution process is as follows.

First construct a two-dimensional matrix Q:

$$Q = \begin{bmatrix} q(1,1) & \cdots & q(1,K) \\ \vdots & \ddots & \vdots \\ q(N,1) & \cdots & q(N,K) \end{bmatrix}, \tag{29}$$

where $q(i,j)$ represents the total delay satisfaction of the slice when the jth time slot is allocated to the ith ONU, K represents the total number of time slots, and N represents the total number of ONUs. $q(i,j)$ is derived by the following formula:

$$q(i,j) = \begin{cases} \max\{q(:, j-1)\} + qds(i), & j \neq 1, \\ \frac{1+e^{-\delta d_m}}{1+e^{-\delta\left(d_m - d_{pi}\right)}}, & j = 1, \end{cases} \tag{30}$$

2161915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Pseudo code of ONU time slot allocation

Algorithm 2: ONU slot allocation algorithm

Input: $dm[]$, $dp[]$, $pkt[]$, c_λ

Output: slot[], maxvalue

function SlotAllocation ($onu[]$, $dm[]$, $dp[]$, $pkt[]$)
$qmax = []$
for $i = 1 \rightarrow N$ **do**
    $q(i, 1) = (1 + \exp(-\delta dm[i]))/(1 + \exp(-\delta(dm[i] - dp[i])))$
  $qmax[1] = \max\{q(:, 1)\}$
  pkt_size = 0
for $j = 2 \rightarrow K$ **do**
    pkt_size+ = pkt[$j - 1$]
    dq = pkt_size/c_λ
    for $i = 1 \rightarrow N$ **do**
        $qds(i) = (1 + \exp(-\delta dm[i]))/(1 + \exp(-\delta(dm[i]$
          $-dp[i] - dq)))$
        $q(i, j) = qmax[j - 1] + qds(i)$
    $qmax[j] = \max\{q(:, j)\}$
for $j = 1 \rightarrow K$ **do**
    for $i = 1 \rightarrow N$ **do**
      if $q(i, j) = qmax[j]$ **do**
          slot[j] = i
   maxvalue = $qmax[K]$
return slot[], maxvalue
end function

where $qds(i)$ represents the delay satisfaction of ONU_i. The average queuing delay for ONUs transmitting in the first time slot is zero.

Next, the ONU with the largest $q(i, j)$ in each time slot is allocated to the time slot for transmission. Finally, the slice delay satisfaction obtained by the optimal time slot allocation is $\max\{q(:, K)\}$. The pseudo code of ONU time slot allocation is shown in Table 2.

4.3 | eMBB slice bandwidth allocation algorithm

4.3.1 | Bandwidth allocation

The eMBB slices have low delay requirements but high bandwidth requirements. Therefore, the goal of bandwidth allocation for eMBB slice is to maximize the bandwidth satisfaction of the slices, as shown in Equation (31).

$$\max : \sum \frac{2}{1 + e^{\delta(b_m - b_i)}}, \tag{31}$$

$$\text{s.t.} \begin{cases} \sum b_i \le C_2, \\ 0 \le b_i \le b_m, \forall i, \end{cases} \tag{32}$$

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where C_2 is the capacity of eMBB slice. Equation (32) is the constraint of bandwidth allocation. The solution for this maximization problem can be solved by the Lagrange multiplier method.

In order to remove the constraints, the following Lagrange equation can be established:

$$L\left(b_i, x, y_i, z_i\right) = -\sum \frac{2}{1 + e^{\delta\left(b_m - b_i\right)}} + x\left(\sum b_i - C_2\right) + y_i\left(b_i - b_m\right) - z_i \cdot b_i, \tag{33}$$

where the minus sign before the first term is used to transform this maximization problem into a minimization problem. Partial derivation of $L\left(b_i, x, y_i, z_i\right)$ to b_i, x, y_i, z_i, respectively, can derive the results we want. The partial derivative equation is given as follows:

$$\begin{cases} \frac{\partial L\left(b_i,\ x, y_i, z_i\right)}{\partial b_i} = 0, \\ \frac{\partial L\left(b_i,\ x, y_i,\ z_i\right)}{\partial x} = 0, \\ \frac{\partial L\left(b_i,\ x, y_i, z_i\right)}{\partial y_i} = 0, \\ \frac{\partial L\left(b_i,\ x, y_i, z_i\right)}{\partial z_i} = 0. \end{cases} \tag{34}$$

The optimal bandwidth allocation value b_i is obtained by solving Equation (34). If b_i equals to 0, it means that the ONU is not admitted in the current cycle.

4.3.2 | Wavelength and time slot allocation

In order to realize the load balance of each wavelength in the eMBB slice, the wavelength utilization status should be considered when allocating the wavelength and time slot of the ONU. First, the ONUs in the slice are sorted in descending order of requested bandwidth. Next, the utilization of each wavelength is counted, and the wavelengths are arranged from low to high utilization. Then, the ONUs are taken out in sequence and allocated to the current time slot of the wavelength with the lowest utilization. Once an ONU has been assigned, the current utilization of each wavelength is updated. Repeat the above process until all ONUs complete the assignment of wavelengths and time slots.

4.4 | mMTC slice bandwidth allocation algorithm

The mMTC slice transmits massive machine-type communication services, which have low transmission rates and low requirements for delay and bandwidth. However, when a large number of machines request access to the network, intolerable delay, and packet loss will occur. Therefore, access control is needed for users of mMTC slices.

4.4.1 | ONU admission decision mechanism

In order to meet the service requirements of users, each ONU will be defined with a priority weight. Different from the priority defined by the URLLC slice, the request bandwidth and delay requirements of the ONU need to be integrated here. The definition formula is as follows:

$$w_i = \left(1 - \frac{R_i}{\sum R_i}\right) \cdot \frac{\tau_i}{\Gamma}, \tag{35}$$

where R_i represents the requested bandwidth of ONU$_i$, $\sum R_i$ represents the total requested bandwidth of mMTC slice. The value of τ_i is an integer ranging from 1 to 5, and each value represents a delay interval. The larger the τ_i, the lower the delay of the interval. For example, the τ_i for ONUs with delay requirements between 50 and 100 μs are all 5, while the τ_i for delay requirements between 1 and 10 ms are all 1. Γ is the sum of the priorities of all ONUs in the slice.

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

From Equation (35), it can be seen that the lower the requested bandwidth and the higher the delay requirement, the higher the priority of the ONU.

Let the admission threshold of ONU be w_{th}, and calculate the priority weight of all ONUs. If the priority weight of ONU is greater than the threshold, the ONU is admitted. Otherwise, it is not allowed to be admitted.

4.4.2 | Bandwidth allocation

For admitted ONUs, in order to ensure its fairness, the maximum-minimum fairness strategy is adopted for bandwidth allocation. First, the number of admitted ONUs is counted and denoted by N. Then, the initial bandwidth is assigned to each user:

$$B_{\text{initial}} = \frac{C_3}{N}, \tag{36}$$

where C_3 represents the capacity of mMTC slice.

Since the initial bandwidth may be greater than or less than the requested bandwidth, the ONUs with excess bandwidth allocated and the ONUs with insufficient bandwidth allocated should be counted, and represented by the $\text{Set}_{\text{excess}}$ and the $\text{Set}_{\text{insufficient}}$, respectively.

For the ONU in $\text{Set}_{\text{excess}}$, since the initial bandwidth is greater than its requested bandwidth, the final allocated bandwidth is the requested bandwidth, namely:

$$B_i = R_i. \tag{37}$$

The excess bandwidth of each ONU in this set is recycled into a resource pool, and the capacity of the resource pool is represented by pool. pool can be calculated by:

$$\text{pool} = \sum_{i\varepsilon \text{Set}_{\text{excess}}} B_{\text{initial}} - B_i. \tag{38}$$

For the ONUs in $\text{Set}_{\text{insufficient}}$, since the initial bandwidth is less than the requested bandwidth, bandwidth redistribution is required.

First, the priority weight ratio of ONUs in $\text{Set}_{\text{insufficient}}$ is calculated:

$$\omega_i = \frac{w_i}{\sum_{i\varepsilon \text{Set}_{\text{insufficient}}} w_i}. \tag{39}$$

Then, the bandwidth is reallocated according to the priority weighting ratio of ONUs:

$$B_i = B_{\text{initial}} + \omega_i \cdot \text{pool}. \tag{40}$$

At the same time, update the *pool*:

$$\text{pool} = \text{pool} - \omega_i \cdot \text{pool}. \tag{41}$$

If the requested bandwidth of ONU has been satisfied, the ONU is removed from $\text{Set}_{\text{insufficient}}$. Repeat the above bandwidth reallocation process for the ONUs in $\text{Set}_{\text{insufficient}}$ until the requested bandwidth of all ONUs is satisfied, or the pool is less than or equal to 0. Finally, end the bandwidth reallocation process.

4.4.3 | Wavelength and time slot allocation

mMTC slices allocate wavelengths and slots to ONUs according to the principle of the earliest idle wavelength first. At first, all admitted ONUs are sorted in descending order according to w_i. Then, ONUs are taken out from the head of the

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

queue in turn and assigned to the earliest free wavelength. Finally, when all ONUs are removed, the wavelength and time slot allocation of ONUs is completed.

5 | SIMULATION AND PERFORMANCE ANALYSIS

The proposed algorithm is evaluated with MATLAB and Python. The simulation scenario is a PON-based 5G optical fronthaul network. It consists of three slices: URLLC slice, eMBB slice, and mMTC slice. Each slice consists of a VPON. Both URLLC slice and eMBB slice contain one vOLT and 16 ONUs, while the mMTC slices contain one vOLT and 256 ONUs. The transmission distance of OLT and ONU is evenly distributed from 1 to 25 km. The ratio of requested bandwidth to total requested bandwidth for URLLC slice, eMBB slice, and mMTC slice are 0.25, 0.4, and 0.35, respectively. eMBB slice has the largest percentage of requested bandwidth. This is because the services transmitted in this slice are large bandwidth services. Even if the number of ONUs is small, the total bandwidth required by all ONUs will still be large. The second largest percentage of requested bandwidth is the mMTC slice. Although the amount of data that needs to be transmitted by each ONU in the mMTC slice is small, the total amount of data transmitted by 256 ONUs will also be relatively large. So, the percentage of requested bandwidth is relatively large. The least requested bandwidth is the URLLC slice. This is because the number of ONUs in a URLLC slice and the amount of bandwidth required by each ONU are both small. The bandwidth unit price of these three slices is set to 4, 3, and 2 RMB/Mbps, respectively. The priorities of ONUs in the URLLC slice are randomly selected from 1, 2, and 3, and the d_m corresponding to these three priorities are 50, 90, and 130 μs, respectively. The whole simulation network uses 15 wavelengths and the transmission rate of each wavelength is 10 Gb/s. ONUs' data streams are generated by Poisson distribution, and the packet size distribution was between 64 and 1518 bytes. The buffer size of each ONU is 20 Mbytes. Packet loss occurs when the amount of data arriving at the ONU exceeds the buffer size. ρ_0 is 0.3, $\rho_{\max}$ is 0.8, $\rho_{\min}$ is 0.2, θ_1 is 0.3, and θ_2 is 0.7. Both v_1 and v_2 in Equation (11) are set to 0.5. The admission threshold for ONUs in mMTC slices is set to 0.003. The grant cycle of SDN controller is set to 50 μs. The parameters of the simulation experiment are summarized in Table 3. During the simulation, the ONUs in each slice first report the received service request information to the vOLT. Then, the vOLT reports the total requested bandwidth of the slice to the SDN controller. Next, the SDN controller executes an inter-slice bandwidth allocation algorithm to assign wavelengths to each slice based on the received information. Finally, in each slice, the vOLT executes the intra-slice bandwidth allocation algorithm, allocating bandwidth to ONUs and scheduling ONUs to different time slots on the wavelength for data transmission according to ONU's request information and priority. For the convenience of description, in the simulation, the URLLC slice bandwidth allocation algorithm is represented by

TABLE 3 The parameters of simulation experiments

Parameters	value
The number of ONUs in URLLC, eMBB, mMTC slices	16, 16, 256
The ratio of request bandwidth for URLLC, eMBB, mMTC slices	0.25, 0.4, 0.35
The buffer size of each ONU	20 Mbytes
The grant cycle of SDN controller	50 μs
$\alpha_1, \alpha_2, \alpha_3$	4, 3, 2 (RMB/Mbps)
p_i	1, 2, 3
d_m	50, 90, 130 (μs)
wave	15
C_λ	10 Gb/s
ρ_0	0.3
$\rho_{\max}, \rho_{\min}$	0.8, 0.2
θ_1, θ_2	0.3, 0.7
v_1, v_2	0.5, 0.5

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

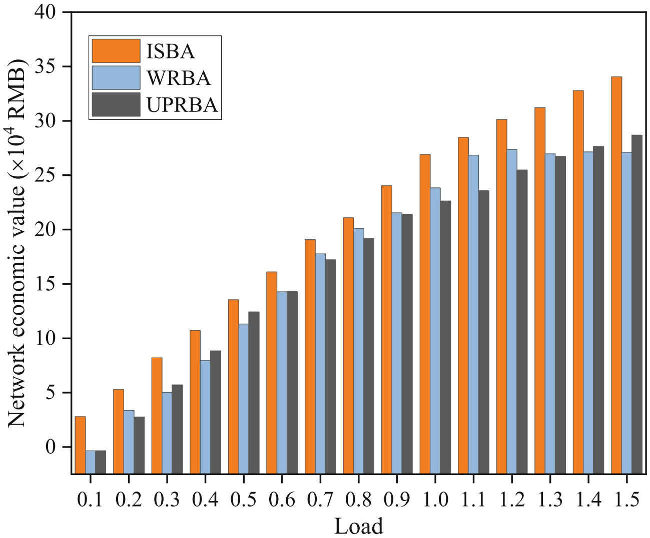

FIGURE 4 Comparison of network economic value of three algorithms

USBA, the eMBB slice bandwidth allocation algorithm is represented by ESBA, and the mMTC slice bandwidth allocation algorithm is represented by MSBA.

5.1 | Network economic value analysis

In this subsection, ISBA will be compared with the classic weighting ratio bandwidth allocation (WRBA) algorithm, and the unit price ratio bandwidth allocation (UPRBA) algorithm. For WRBA, the more bandwidth requested, the more bandwidth is allocated. For UPRBA, the higher the unit price, the more bandwidth resources are allocated. The comparison results of the network economic value of the three algorithms are shown in Figure 4. In the figure, the load is defined as the ratio of the total amount of data that needs to be transmitted in the network to the network capacity. When the load is between 0.1 and 1.5, the network economic value of ISBA is larger than that of the other two algorithms. And as the load increases, the difference between ISBA and the other two algorithms shows a trend of decreasing and then increasing. The reason is that the bandwidth allocated to each slice by WRBA and UPRBA remains largely unchanged regardless of the load. When the load is small, the network revenue is small while the cost of constructing slices is large, resulting in a lower economic value of the network. ISBA trades off the cost of constructing slices and the cost of transmission bandwidth. By calculating the optimal allocated bandwidth in different slice resource utilization intervals, the smallest network cost can be obtained. Therefore, with the same revenue, ISBA can obtain the maximum network economic value.

When the load increases but is less than 1, the difference of network cost calculated by the three algorithms becomes smaller, but the network benefit remains the same. Therefore, the difference of the network economic value decreases. When the load is greater than 1, the requirements of the three slices cannot be satisfied at the same time. At this time, ISBA will tilt network resources to slices with high bandwidth unit price, and give priority to meet the demand of slices with high unit price, so as to obtain higher network revenue. ISBA sacrifices a part of fairness, but can get the maximum network economic value. Due to insufficient bandwidth, the benefits and costs of WRBA will remain stable, and the obtained network economic value will remain basically unchanged. UPRBA will also tilt its resources to the slices with high unit price at this time, but the tilt is not large. Although the network economic value will increased, it is still lower than ISBA.

5.2 | QoS value analysis of URLLC slices

In this subsection, USBA is compared with the USBA without admission control (UWAC) and the longest-first first-available (LFFA) algorithm[21] with admission control. The strategy of the LFFA algorithm is that the ONU with the

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

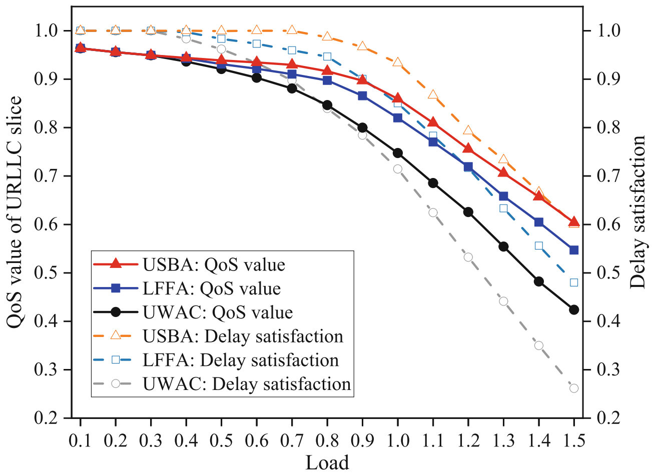

FIGURE 5 Comparison of QoS value and delay satisfaction of three algorithms

longest requested bandwidth will give priority to time slot allocation and be allocated to the earliest available wavelength. The comparison results of the QoS value and delay satisfaction of the three algorithms are shown in Figure 5. When the load is less than 0.3, the delay satisfaction and QoS value of the three algorithms are basically the same. This is because the needs of all ONUs can be met when the slice bandwidth is sufficient. And with low network link utilization, arriving packets can be transmitted without waiting, so there is basically no packet loss. When the load is less than 0.9, both the QoS value and the delay satisfaction value decrease, but the magnitude is small. When the load is greater than 0.9, the QoS value and delay satisfaction value of the three algorithms drop sharply, but the performance of USBA and LFFA algorithms is always better than that of UWAC. This is because USBA algorithm and LFFA algorithm have admission control. When the load is high, it will control the admission request of some ONUs, which greatly reduces the delay and the packet loss rate. The performance of the USBA algorithm is better than that of the LFFA algorithm because USBA algorithm takes into account the delay requirements and the propagation delay of ONUs when allocating time slots, and guarantees the maximum delay satisfaction for the ONU transmitted on each time slot.

5.3 | QoS value analysis of eMBB slices

The comparison results of QoS value and bandwidth satisfaction of ESBA and WRBA are shown in Figure 6. When the load is less than 0.9, the bandwidth satisfaction of both algorithms is 1 because the bandwidth resources of eMBB slices

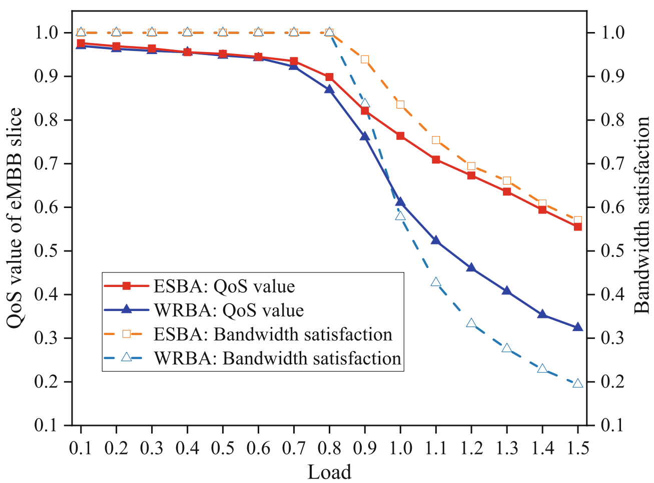

FIGURE 6 Comparison of QoS value and bandwidth satisfaction of two algorithms

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

are sufficient and the requests of all ONUs are satisfied. When the load is greater than 0.9, the bandwidth satisfaction of the network decreases sharply due to insufficient bandwidth resources. The ESBA algorithm maximizes the total bandwidth satisfaction of the network by finding the optimal solution. The WRBA algorithm allocates the bandwidth to each user proportionally. In the case of insufficient resources, the bandwidth of each user cannot be satisfied, and the whole network cannot obtain a larger bandwidth satisfaction value. In the QoS value model of eMBB slices, bandwidth satisfaction accounts for the largest proportion, so the curve of QoS value of slices has roughly the same trend as the curve of bandwidth satisfaction. Since QoS value is also affected by packet loss rate satisfaction, delay satisfaction, and jitter satisfaction, the overall QoS value will be lower than bandwidth satisfaction. Both in terms of bandwidth satisfaction and overall QoS value of slices, ESBA performs significantly better than WRBA.

5.4 | QoS value analysis of mMTC slices

The comparison results of the QoS value and packet loss rate satisfaction of MSBA and WRBA algorithms are shown in Figure 7. In each polling cycle, the MSBA algorithm performs ONU admission decision first, so it can still maintain a low packet loss rate even when the load is high. However, the WRBA algorithm allocates the bandwidth according to the requested bandwidth ratio of the ONU regardless of the load. When the load is large, a large number of data packets accumulate in the ONU's buffer, resulting in a sharp increase in the packet loss rate. Therefore, when the network is under heavy load, the packet loss rate satisfaction of WRBA algorithm is much smaller than that of MSBA algorithm. And as the load increases, this gap becomes even larger. Similarly, the QoS value of the WRBA algorithm is much smaller than that of the MSBA algorithm as the load increases. This is because the MSBA algorithm takes the user's delay requirement and bandwidth requirement as determinants when making the admission decision, so as to ensure that as many users as possible obtain the maximum delay satisfaction and bandwidth satisfaction. However, the WRBA algorithm allocates the bandwidth proportionally, so that all QoS performance requirements of users are not met, and the greater QoS value cannot be obtained.

5.5 | Slicing network value analysis

The comparison results of the network value of URLLC slice, eMBB slice, and mMTC slice are shown in Figure 8. In the figure, the solid line indicates the algorithm proposed in this article, and the dashed line indicates the comparison algorithm for each slice. In the legend, the left side of the plus sign is the inter-slice bandwidth allocation algorithm, and the right side of the plus sign is the intra-slice bandwidth allocation algorithm. As can be seen from the figure, when the load is less than 0.7, the network value of the three slices gradually increases as the load increases, but the difference between the three slices is not significant. This is because when the load is small, the QoS index requirements of slices

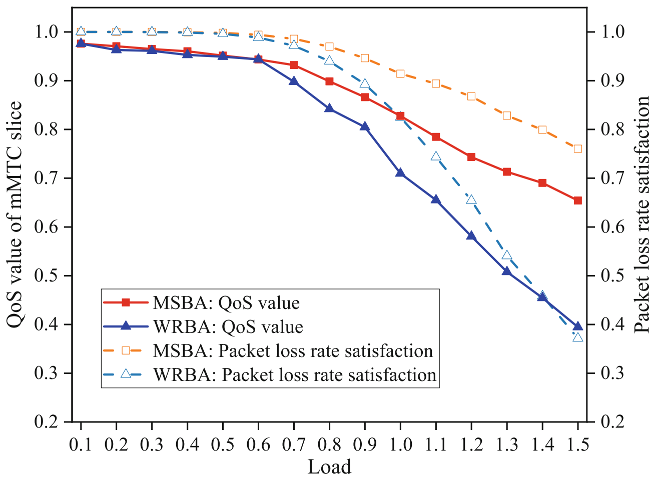

FIGURE 7 Comparison of QoS value and packet loss satisfaction of two algorithms

2161915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

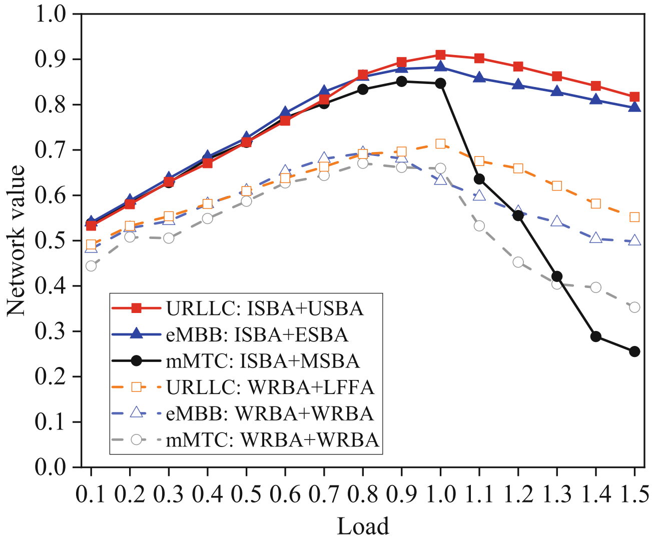

FIGURE 8 Comparison of network value of three slices

are all met and the QoS value is always at a large value. At this time, the network value of slices is mainly determined by the economic value of slices. As the load increases, the revenue of the three slices gradually increases, so the network value also increases. When the load is between 0.7 and 1, the economic value of the slice still keeps growing, but the QoS value begins to decrease, which makes the increase of the network value of each slice smaller. When the load is greater than 1, the QoS value of slices decreases faster. Due to the high unit price of resources, network resources are inclined to URLLC slice and eMBB slice, and the economic value of these two slices is still increasing. So, the network value of these two slices only decreases slightly. In contrast, the network value of mMTC slice decreases sharply due to the severe under-allocation of network resources. On the whole, when the load is less than 1, the performance of the algorithms proposed in this article is better than the comparison algorithms. When the load is greater than 1, the algorithm proposed in this article performs better in URLLC slice and eMBB slice, and their network value is greater than that of the comparison algorithm. The performance of the mMTC slice is inferior to that of the comparison algorithm due to the sacrifice of its fairness.

The average network value refers to the average of the network values of the three slices. The comparison results of the average network value between the proposed algorithm and the comparison algorithm are shown in Figure 9. From the figure, it can be seen that the average network value of the proposed algorithm in this article is larger than the comparison algorithm regardless of the load. This verifies that the proposed algorithm performs well in improving network profit and QoS performance. Combining Figures 8 and 9, it can be seen that when the load is between 0.7 and 1, the average network value is greater than 0.8, and the network value is at the higher value for all three slices. When the load exceeds 1, the total available bandwidth of the network is less than the total requested bandwidth of the users. Due to insufficient network bandwidth resources, users' demands cannot be satisfied, and the QoS value of slices decreases rapidly at this time, resulting in a lower average network value. Therefore, in order to ensure the best performance of the entire network and each slice, it is best to keep the network load between 0.7 and 1.

5.6 | Algorithm execution time analysis

The comparison results of the execution time of the proposed algorithm and the WRBA algorithm are shown in Figure 10. As can be seen from the figure, the execution time of the algorithm tends to increase as the load increases. This is because at higher loads, the system needs to transmit a larger number of packets, which imposes a greater computational effort on the system. Compared with WRBA, the algorithm proposed in this article has a shorter execution time. Since the ONU admission decision mechanism is set, when the load is large, the mechanism will reject the admission of some ONUs, which greatly reduces the computational workload of the system, and therefore, the execution time is also greatly reduced.

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

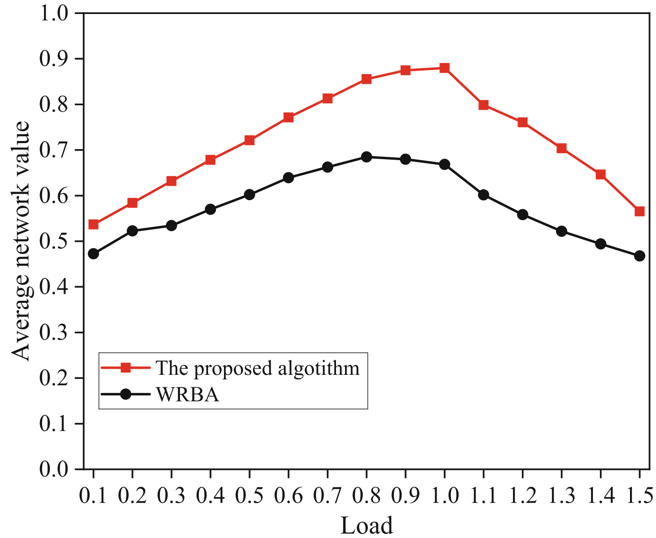

FIGURE 9 Comparison of average network value of two algorithms

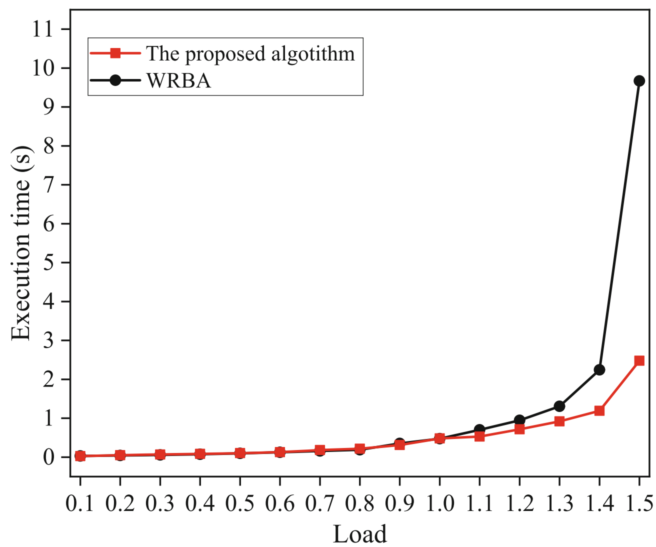

FIGURE 10 Comparison of execution time of two algorithms

5.7 | Algorithm complexity analysis

This subsection analyzes and compares the time complexity of the proposed algorithm and the WRBA algorithm. For the WRBA algorithm, it is necessary to traverse the requested bandwidth of all ONUs, so the time complexity is O(n) (n is the number of ONUs). For the algorithm proposed in this article, it consists of two parts: ISBA and intra-slice bandwidth allocation algorithm. The time complexity of ISBA is O(1). The intra-slice bandwidth allocation algorithm consists of three types: USBA, ESBA, and MSBA. USBA includes two steps: the first step is to conduct ONU admission decision and bandwidth allocation, and its time complexity is O(n). The second step is to allocate wavelengths and time slots for accessed ONUs. Since the number of admitted ONUs is unknown, the optimal time complexity of this step is O(1). In the worst case (all ONUs are accessed), the time complexity is O($k*n$) (k is the number of time slots on a wavelength) because a dynamic programming algorithm is used. Therefore, the average time complexity of USBA is O($n$log$n$). ESBA is also divided into two steps. The first step is bandwidth allocation. The Lagrangian algorithm is used for bandwidth allocation of eMBB slices, and its time complexity is related to the number of ONUs, that is, O(n). The second step is the allocation of wavelengths and time slots. In the allocation process, a merge sort algorithm is used to arrange the ONUs,

2161391 5, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and its time complexity is O($n$log$n$). Therefore, the average time complexity of ESBA is O($n$log$n$). MSBA includes three steps. The first step is ONU admission judgment, and its time complexity is O(n). The second step uses the maximum and minimum fairness strategy for bandwidth allocation. Since the number of admitted ONUs and the number of bandwidths requested by ONUs are uncertain, the optimal time complexity is O(1), the worst case time complexity is O(n^2), and the average time complexity is O($n$log$n$). The third step is wavelength and time slot allocation, and the time complexity is O($n$log$n$). Therefore, the average time complexity of MSBA is O($n$log$n$). Based on the above analysis, the average time complexity of the algorithm proposed in this article is O($n$log$n$). Although the time complexity of the proposed algorithm is slightly higher than the WRBA, the proposed algorithm can bring higher network value. Meanwhile, the slight increase in complexity is not so unacceptable. Thus, the proposed algorithm is reasonable and effective.

6 | CONCLUSION

We propose an inter-slice and intra-slice adaptive dynamic bandwidth allocation algorithm based on network value model for 5G fronthaul network. First, a network value model is proposed by comprehensively considering network profit and network service quality. In this model, network profit is related to link utilization status, and network service quality is measured by user perception. Then, the allocated bandwidths of slices with the greatest network economic value under different link utilization conditions are calculated to obtain the optimal inter-slice bandwidth allocation scheme. Furthermore, considering slice characteristics, three intra-slice bandwidth allocation algorithms are proposed so that URLLC slice, eMBB slice, and mMTC slice can obtain the maximum QoS value. Finally, the simulation results demonstrate the effectiveness of the proposed model and algorithm.

ACKNOWLEDGEMENTS

This work was supported in part by National Key Research and Development Program of China (2021YFB2900800), the Science and Technology Commission of Shanghai Municipality (Projects No. 20511102400, 20ZR1420900) and 111 project (D20031).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Chaoqin Gan https://orcid.org/0000-0002-9737-9381

REFERENCES

1. Zhang S. An overview of network slicing for 5G. *IEEE Wirel Commun*. 2019;26(3):111-117.
2. Alliance N. Description of network slicing concept. NGMN 5G P, 2016, 1(1): 1–11.
3. Huang S, Guo B, Liu Y. 5G-oriented optical underlay network slicing technology and challenges. *IEEE Commun Mag*. 2020;58(2):13-19.
4. Wang G, Feng G, Qin S, Wen R, Sun S. Optimizing network slice dimensioning via resource pricing. *IEEE Access*. 2019;7:30331-30343.
5. Han B, Tayade S, Schotten HD. Modeling profit of sliced 5G networks for advanced network resource management and slice implementation. Paper presented at: IEEE symposium on computers and communications (ISCC); 2017:576-581; IEEE.
6. Zhang J, Xiao Y, Song D, Bai L, Ji Y. Joint wavelength, antenna, and radio resource block allocation for massive MIMO enabled beamforming in a TWDM-PON based fronthaul. *J Light Technol*. 2019;37(4):1396-1407.
7. Liang D, Gu R, Guo Q, et al. Demonstration of multi-vendor multi-standard PON networks for network slicing in 5G-oriented mobile network. Paper presented at: Asia Communications and Photonics Conference Optical Society of America; 2017:1-3; IEEE.
8. Uzawa H, Honda K, Nakamura H, et al. Dynamic bandwidth allocation scheme for network-slicing-based TDM-PON toward the beyond-5G era. *J Opt Commun Netw*. 2020;12(2):A135-A143.
9. Zhang L, Zhang M, Song C, et al. Service-aware network slicing supporting delay-sensitive services for 5G fronthaul. Paper presented at: 2018 23rd Opto-electronics and communications conference (OECC); 2018:1-2; IEEE.
10. Wang F, Tian Q, Zhang Q, et al. A dynamic bandwidth allocation scheme for internet of thing in network-slicing passive optical networks. Paper presented at: 2020 IEEE computing, communications and IoT applications (ComComAp); 2020:1-5; IEEE.
11. Liang X, Tian Q, Wang F, Yu W, Xin X. A dynamic resource allocation based on network traffic prediction for sliced passive optical network. Paper presented at: 2021 19th International Conference on Optical Communications and Networks (ICOCN); 2021:1-3.
12. Soumplis P, Kokkinos P, Varvarigos E. Network slicing in fiber wireless fronthaul for 5G networks. ONDM; 2021.
13. Song C, Zhang M, Zhan Y, et al. Hierarchical edge cloud enabling network slicing for 5G optical fronthaul. *J Opt Commun Netw*. 2019;11(4):B60-B70.

14. Sun Y, Qin S, Feng G, Zhang L, Imran MA. Service provisioning framework for RAN slicing: user admissibility, slice association and bandwidth allocation. *IEEE Trans Mob Comput*. 2021;20(12):3409-3422.
15. Mijumbi R, Serrat J, Gorricho JL, Bouten N, de Turck F, Boutaba R. Network function virtualization: state-of-the-art and research challenges. *IEEE Commun Surv Tutor*. 2015;18(1):236-262.
16. Ordonez-Lucena J, Ameigeiras P, Lopez D, Ramos-Munoz JJ, Lorca J, Folgueira J. Network slicing for 5G with SDN/NFV: concepts, architectures, and challenges. *IEEE Commun Mag*. 2017;55(5):80-87.
17. Wu L, Gan C, Xu Z, Hui J. Multi-dimensional quantitative model of communication network value. *Sci Rep*. 2022;12(1):1-14.
18. Zhan N, Gan C, Hui J, Guo Y. Fair resource allocation based on user satisfaction in multi-Olt virtual passive optical network. *IEEE Access*. 2020;8:134707-134715.
19. Antonioli RP, Rodrigues EB, Maciel TF, Sousa DA, Cavalcanti FRP. Adaptive resource allocation framework for user satisfaction maximization in multi-service wireless networks. *Telecommun Syst*. 2018;68(2):259-275.
20. Haiyan MA, Jinyao YAN, Georgopoulos P, et al. Towards SDN based queuing delay estimation. *China Commun*. 2016;13(3):27-36.
21. Zhang Y, Gan C, Gou K, Hua J. (box-filling-model)-based ONU schedule algorithm and bandwidth-requirement-based ONU transfer mechanism for multi-subsystem-based VPONs' management in metro-access optical network. *Opt Fiber Technol*. 2017;36:10-18.

How to cite this article: Wu L, Gan C, Xu Z, Hui J. (Network value)-based adaptive dynamic bandwidth allocation algorithm for 5G network slicing. *Trans Emerging Tel Tech*. 2023;34(3):e4722. doi: 10.1002/ett.4722

21613915, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.4722 by South African Medical Research, Wiley Online Library on [21/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License