Lightweight Network Slicing Scheme for 5G Coexistence Network in Multitenant and Services

A Thesis Submitted to the Department of Computer Science and Communications Engineering, the Graduate School of Fundamental Science and Engineering of Waseda University in Partial Fulfillment of the Requirements for the Degree of Master of Engineering

Submission Date: July 18th, 2022

FANG Tianshun

(5120FG37-3)

Advisor: Prof. Shigeru SHIMAMOTO

Research guidance: Research on Wireless Access Scheme

Acknowledgement

I would first like to thank my supervisor, Professor Shigeru Shimamoto, whose immense knowledge, and plentiful research experience have encouraged me in my academic research and daily life. I am extremely grateful for his invaluable advice, continuous support, and patience during my two-year master's study. Prof. Shimamoto gave me a lot of room for research topics and gave me plenty of time for research progress. Because of his calmness, he cultivated my passion and drive for research, which was crucial to my entire master's study. His outstanding personal charm and rigorous academic attitude will influence me all my life, urging me to become a rigorous and independent thinking researcher.

I would like to thank my research group leader PAN. She gave me a lot of constructive suggestions and academic guidance. Without her urging, I might not have completed my research. I would like to thank Dr. Hao Xue. Thanks to him for sharing his knowledge and ideas, he guided me in the direction of my research.

I would like to thank Professor LIU Jiang. She gave me a lot of courage and motivation to make my master's career so smooth and positive in the past two years. I would like to thank my fellow students and other members of the lab. Thank you for your kindness and enthusiasm, and very fortunate to study with you. I would like to thank my family for supporting my choice to come to Japan. Thanks to their material and spiritual help, I can concentrate on my research and see the wider world.

Abstract

Resource allocation of 5G cellular network slicing has been gaining huge interest recently. However, it is not well-explored under the Multi-tenant and Multi-service Coexistence Network scenario. This paper proposes a dynamic network slicing scheme for the 5G Coexistence Network, which considers tenants, priority, baseband resources, slice granularity, service characteristics, quality of service (QoS), and fairness index. The framework of the network slicing scheme consists of a group-level, which performs clustering preprocessing, setting the priority of service, and filtering the desired characteristic service groups; and a slice-level, which performs radio resource allocation among services and follows the changes in the network environment closely. Simulation results show that the proposed scheme can significantly improve compared to the three baseline schemes. For the simulated goal service group, QoS is improved by 20%.

Keywords: Network slicing, Resource allocation, Clustering algorithm, Slice granularity, Fairness index

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background and Motivation

With the emergence of diversified new services in the 5G and cloud era, different industries, services, or users have put forward various service quality requirements for the network. For example, for services such as mobile communication, smart home, environmental monitoring, smart agriculture, and smart meter reading, the network needs to support massive device connections and the frequent transmission of many small packets. Services such as webcasting, video backhaul, and mobile medical care require a higher transmission rate. higher requirements. Services such as the Internet of Vehicles, smart grids, and industrial control require millisecond-level latency and nearly 100% reliability. Therefore, 5G networks should have the capabilities of massive access, deterministic latency, and extremely high reliability, and a flexible and dynamic network needs to be built to meet the diverse business needs of users and vertical industries[1]. This situation has led to a sharp increase in the number of tenants in the network, and the characteristics of services are complex. In the research of designing future 5G-beyond and sixth generation(6G) networks, network slicing and Artificial Intelligence (AI) assisted communication and networking technologies have attracted attention.

The network slicing technology divides the physical network into multiple virtual slices, and each slice achieves different QoS services. In 5G end-to-end network slicing, the main function of IP bearer network slicing is to provide customized network topology connections between network elements and services in the network slice of the radio access network and the core network, and to provide customized network topology connections for different network slices. The business provides differentiated service quality SLA guarantees. In addition, to realize the collaborative management of end-to-end network slices with radio access network and core network slices, and to provide network slices as a new service to tenants in vertical industries, the bearer network also needs to provide an open network slice management interface. Lifecycle management for network slices. The classic network slicing scheme focuses on service classification and corresponds to the three classic application scenarios of 5G (eMBB, uRLLC, and mMTC) [2]. Most research on network slicing also follows this setting [3]. The slicing granularity of these slicing schemes is coarse, which cannot meet the diversified QoS, and it is very cumbersome to dynamically adjust the slicing. With the development of network technology, such network slicing configuration will have unsuitable application scenarios and cannot solve some resource allocation problems.

1.2. Accessibility

Through network slicing, operators can build multiple dedicated, virtualized and isolated logical networks on top of a common physical network to meet the differentiated requirements of different customers for network capabilities. The value of network slicing is mainly reflected in four aspects: resource and security isolation, deterministic latency, flexible customized topology connections, and automated slice management.

Network slice instances can be abstracted as models for complex user classification. It can well solve the complicated service classification in network slicing. Cluster analysis, also known as group analysis, is a statistical analysis method for studying (sample or index) classification problems and is also an important algorithm for data mining. Cluster analysis consists of several patterns, usually a pattern is a vector of measures, or a point in a multidimensional space. Fuzzy c-mean (FCM) is the most popular fuzzy clustering algorithms [4]. For a multi-tenant and multi-service coexistence network, FCM is used to process each service, and finally, services with similar characteristics are allocated to the same slice. Although clustering algorithms are a mature research topic in computer science, there is still much to be done to apply them to network slicing. In view of the characteristics of FCM, we apply this algorithm to our proposed scheme.

1.3. Related Works

According to the description of 3GPP TR 28.801[5], a network slice instance includes a set of network functions and the supporting network resources, which are arranged and configured to form a complete logical network that meets specific network characteristics required by a service instance [5]. Related work can be categorized into two major research directions, network scenarios and resource allocation.

The research direction of the network scenario of network slicing is mainly to optimize the user experience. Through the research direction, operators can reduce the cost of constructing multiple private networks and can provide highly flexible on-demand network services according to business needs, thereby enhancing the network value. The challenges in designing the RAN slicing solutions for multi-tenant and multi-service scenarios can be elaborated as follows:[6]-[9]

Priority of different QoS services: When the network is overloaded, we need to give priority to ensure that the services of the target users are met, meanwhile we also need to filter out the main tasks of the current network.

The high complexity of services characteristics: The RAN slicing framework needs to estimate the network resource requirements of the slicing intelligently and accurately according to the QoS requirements of the service. And according to the characteristics of services, suitable slices with number, size, and other performance indicators.

Spatial-temporal dynamics of network: On one hand, user location distributions and deployment of mobile nodes in different regions may be different, which introduces spatial inhomogeneity of service traffic between different cells. On the other hand, the time dynamics of cellular networks include long-term fluctuations (that is, the dynamics of business traffic) and short-term fluctuations (that is, the dynamics caused by wireless channels).

Fairness indicators for the overall service: Since the plan will focus resources on goal service, this action will have an impact on the overall service, so macro indicators need to be introduced to evaluate the scheme.

The research direction of resource allocation for network slicing is mainly to solve the problem of resource shortage. Although 5G technology has further improved communication resources, in practical applications, there will still be overloading, that is, services that need to be processed exceed the processing capacity of the network processing unit (base station or base station group). In the Radio Access Network (RAN), network slicing also needs to face the challenges of the randomness of the wireless network, the multi-dimensional QoS requirements of the service, and the highly dynamic service flow.

Most network slicing solutions are implemented through four stages: planning, deployment, maintenance, and optimization. Slice planning is to plan the scope, bandwidth, and delay of slices according to service guarantee requirements. Slice deployment is when the controller completes the deployment of slice instances, including creating slice interfaces, configuring slice bandwidth, and configuring VPNs and tunnels. Slice maintenance is the controller uses iFIT and other technologies to monitor service delay and packet loss indicators. The network slice traffic, link status, and service quality information are reported through the Telemetry technology, and the network slice status is presented in real time. Slicing optimization is based on business service level requirements and seeks the best balance between slicing network performance and network cost. The two main implementations of slicing optimization are bandwidth optimization and slice capacity expansion. We summarize and abstract three schemes with different concerns.

[10]-[12]

1)The fair scheme is to not process each slice, and only distribute communication resources fairly based on the proportion of the number of user requests. This solution can theoretically achieve the fairest distribution, but it cannot improve the satisfaction of major network users in a targeted manner.

2)The smallest class priority scheme is to first satisfy the user request of the smallest slice. This kind of scheme can improve the satisfaction of network user groups, but it sacrifices the service quality of some user groups and loses fairness.

3)The largest class priority scheme is to give priority to satisfy the user request of the largest slice. In this scheme, judging the largest slice is the most important task of the entire network, but in many cases, this assumption does not hold.

1.4. Dissertation Outlines

Our scheme considers both application scenarios and resource allocation. To adapt to the Multitenant network scenario, we design an FCM algorithm with preprocessing to reduce the computation time. To simulate and deal with resource allocation flexibly, we use the concept of slice group to link the service and the slice result. The main contributions are summarized as follows:

First, starting with the granularity of fine network slicing, and introducing preprocessing to set the priority of different services.

Secondly, using fuzzy clustering algorithms to perform deep clustering of users, prioritizing the allocation of resources to goal users.

Finally, flexibly adjusting slicing resource regarding various user situation in a complex network environment, which can dynamically adapt to communication requirement changes.

The simulation results show that the proposed scheme can significantly improve the QoS of target user groups, which considers fairness, and QoS of the entire system.

Chapter 2

Network Slicing Scheme

This scheme considers a 5G New Radio (NR) based multi-tenant and multi-service network model, which consists of one macro cell and multiple small cells. And each tenant will produce different types of services in its core network. In addition, the number of tenants will increase or decrease over time, and the services required by the same tenant at different times will also change. Our scheme considers both application scenarios and resource allocation. To adapt to the Multitenant network scenario, we design an FCM algorithm with preprocessing to reduce the computation time. To simulate and deal with resource allocation flexibly, we use the concept of slice group to link the service and the slice result.

2.1. Scheme framework

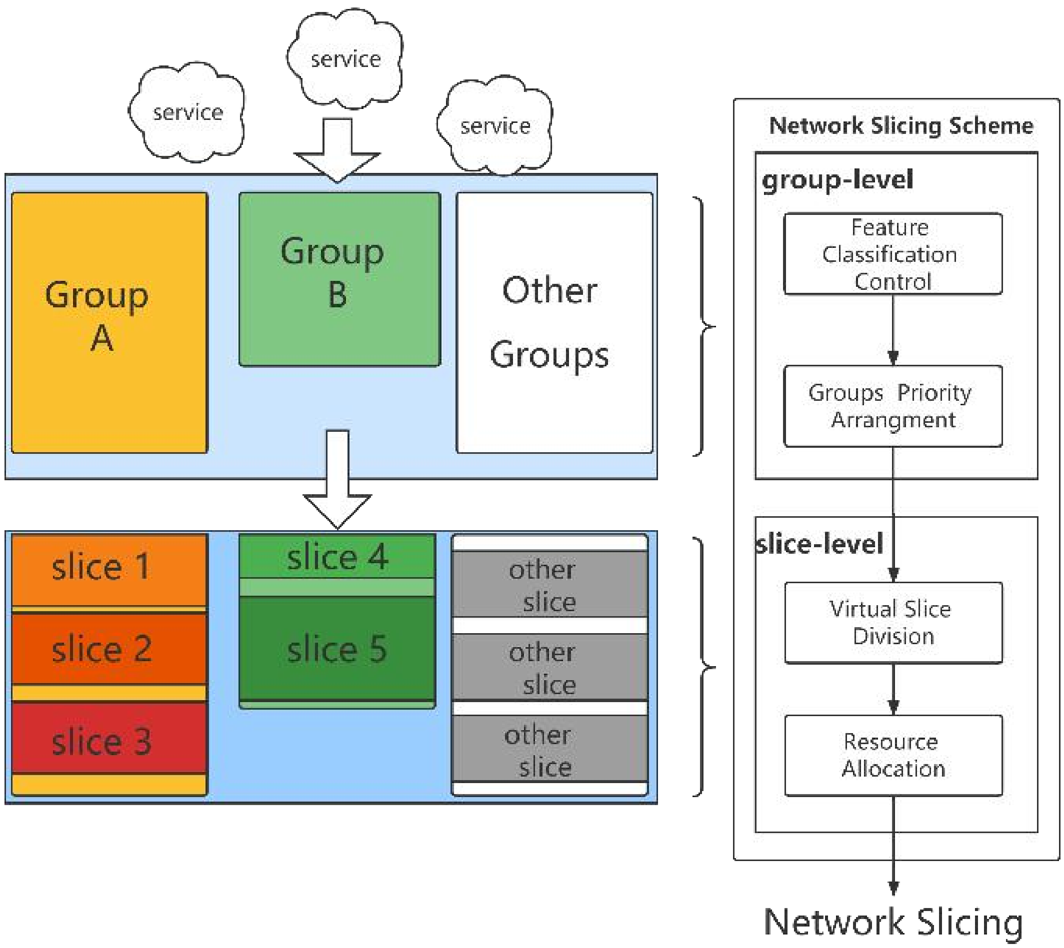

Fig. 1. Framework of the proposed network slicing scheme

This scheme is divided into two levels: Group-level and Slice-level. We follow the NR specification, where the channel bandwidth is divided into physical resource blocks (PRB), each physical resource block occupies sub-6GHz in the frequency domain and 0.25 milliseconds in the time domain. In NR, since resource allocation is performed every 0.5 milliseconds transmission time interval (TTI), a minimum of two PRBs (in the time domain) can be allocated to the UE. After that, we assume that 1) each PRB experiences a flat and slow fading.2) the network is perfectly synchronized. 3) using the shared spectrum mode, each base station can access the entire channel bandwidth.

It is also worth mentioning that our scheme focuses on slice granularity. The slice size ranges from the smallest L7 to the largest L1, and the existing slice sizes (eMBB, uRLLC, mMTC) are around L1-L2. However, the slicing granularity is not as small as well. If the service complexity is not high, selecting the smallest slicing granularity will result in a bloated and cumbersome slicing scheme. The key point of slicing granularity is to be suitable. It is necessary to select a suitable slicing granularity according to the complexity of the service to be processed. In order to achieve this purpose, this scheme adopts the concept of group.

In this slicing scheme, we propose the concept of group, which links the actual user service with the slicing in network slicing technology. Through the operation of group, our scheme can adapt to different application scenarios, and can guarantee the user's QoS.

2.2 Clustering algorithm

Fuzzy c-means (FCM) algorithm is a clustering algorithm that uses membership to determine the extent to which each identified element belongs to a particular category and find clustering centers for each class to minimize fuzzy objective function that is shown in (1).

$$J = \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^{m} \left\| x_i - v_j \right\| \quad (1)$$

where x_i is individual data, u_{ij} is the fuzzy membership degree of individual u_{ij} belonging to j class; m is the fuzzy weight index; v_j is the cluster center of the j class, and the calculation formulas of u_{ij} and v_j are as follows:

$$u_{ij} = \begin{cases} \left[\sum_{k=1}^{c} \frac{\|x_i - v_j\|^{\frac{2}{m-1}}}{\|x_i - v_k\|^{\frac{2}{m-1}}} \right]^{-1}, \|x_i - v_k\| \neq 0 \\ 1, \|x_i - v_k\| = 0 \cap k = j \\ 0, \|x_i - v_k\| = 0 \cap k \neq j \end{cases} \quad (2)$$

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \quad (3)$$

Clustering algorithms are well suited to the problem of classifying many services. For more mature clustering algorithms, there are the K-means algorithm and its improvement (hard clustering) and the FCM algorithm (soft clustering). To adapt to the network slicing scene, our scheme finally selects the FCM algorithm as the main algorithm and the K-means algorithm as the auxiliary. Such an approach can take advantage of FCM, that each service is classified according to its membership probability for each slice and facilitates a dynamic adjustment step. This advantage is compared to hard clustering algorithms, such as the ISODATA algorithm, which can flexibly adjust the cluster center (the number of corresponding slices) to achieve the effect of membership probability classification, but the extra complexity they introduce will increase the consumption of the slicing scheme.

In addition, we add a preprocessing step, using the most classic K-means algorithm as an auxiliary, this optimization makes our algorithm more efficient than simply using the FCM algorithm. The k-means algorithm is a heuristic algorithm that cannot guarantee convergence to the global optimal solution, and the clustering result will depend on the initial clustering. But the algorithm usually runs very fast, so it is common to run it multiple times with different starting states to get better results. For scenarios with known network service characteristics or low complexity, manual setting methods can also be selected in the preprocessing link.

2.3 Fairness index

Fairness measures or metrics are used in network engineering to determine whether users or applications are receiving a fair share of system resources. There are several mathematical and conceptual definitions of fairness.FI is a fairness measurement or measurement used in network engineering to determine whether users or applications have obtained a fair share of system resources. Referring to the multi-rate multi-node fair data packet length adjustment strategy in IEEE802.11, Jain's index is defined as follows [13].

$$FI = \frac{(\sum_{i=1}^{n} \frac{T_i}{O_i})^2}{n\sum_{i=1}^{n} (\frac{T_i}{O_i})^2} \qquad (4)$$

Among them, FI is the fairness index, T_i is the transmission capacity of the *i*-th link in the network; O_i is the actual throughput of the *i*-th link when all n links are active.

The value range of FI is $[n^{-1}, 1]$. When FI=1, the entire system achieves absolute fairness, and the actual throughput of each link is proportional to its transmission capacity, and each link uses wireless resources (channels) in proportion to its transmission capacity; when FI = n^{-1} at this time, the system will be completely unfair. In this case, a certain link completely monopolizes the wireless resource (channel), and the traffic of other links is 0.

Today's cellular networks tend to improve user satisfaction and fairness when allocating network resources, so these two indicators are also the focus of the evaluation of this program. In network slicing scenarios, focusing on fairness indicators can avoid some extreme situations. For example, in order to achieve the overall optimal situation, the network completely ignores the service requests of some users. By improving the fairness index, our scheme can treat different subgroups of users fairly. This issue is most widely studied and discussed in the field of machine learning application "classification", and in recent years several quantitative measures of fairness have been proposed, resulting in a series of algorithms designed to meet these requirements. These algorithms are mainly suitable for offline and small decision-making problems.

In addition, network slicing also involves the method of resource allocation, which can be expanded according to user needs to maximize the indicators of user satisfaction. In other words, the fairness index and user satisfaction are not two completely opposite indicators. They respectively represent two aspects of the performance of the network slicing scheme, and they are also the two most intuitive indicators.

2.4 Distributed Network and Blockchain

The distributed network topology structure is generally grid-like. Different from the centralized network structure, the communication between nodes is no longer a point-to-center communication method, but a point-to-point communication method. This change in the way of communication makes the client/server network model and the computing information processing model of the network easier to implement in a distributed manner. In the distributed network structure, the concept of data processing center has been diluted, because each network site is both a network service object and a network service provider.

A blockchain is a collection of many different types of nodes (or other nodes on the chain) that can interact with each other with clearly defined rights and control. The core of blockchain technology is distributed data storage to ensure security, high availability and high availability. The blockchain Internet is composed of blocks, and each block contains the following information: block data, transaction data, timestamp and block transaction structure.

Chapter 3

Methodology of Algorithms and Optimization

Our proposed slicing scheme is divided into two layers, group-level and slice-level. This section describes the whole scheme from three aspects, flow chart, silhouette value and slice parameter table. The flow chart shows the overall structure of the scheme and the functional modules of each part and describes the working steps of the scheme implementation. Silhoutte value shows the application of the clustering algorithm in this scheme from the algorithm results and shows the relationship between user services and slices and slice groups. The slice parameter table shows the characteristics of the user service considered in this scheme and is also a specific consideration for slice setting.

When the entire slicing scheme is deployed, adjustments will be made according to the network equipment conditions. If it is a highly distributed network, the Group-level can be arranged on distributed nodes to reduce the resource consumption of the control center.

3.1 Group-level

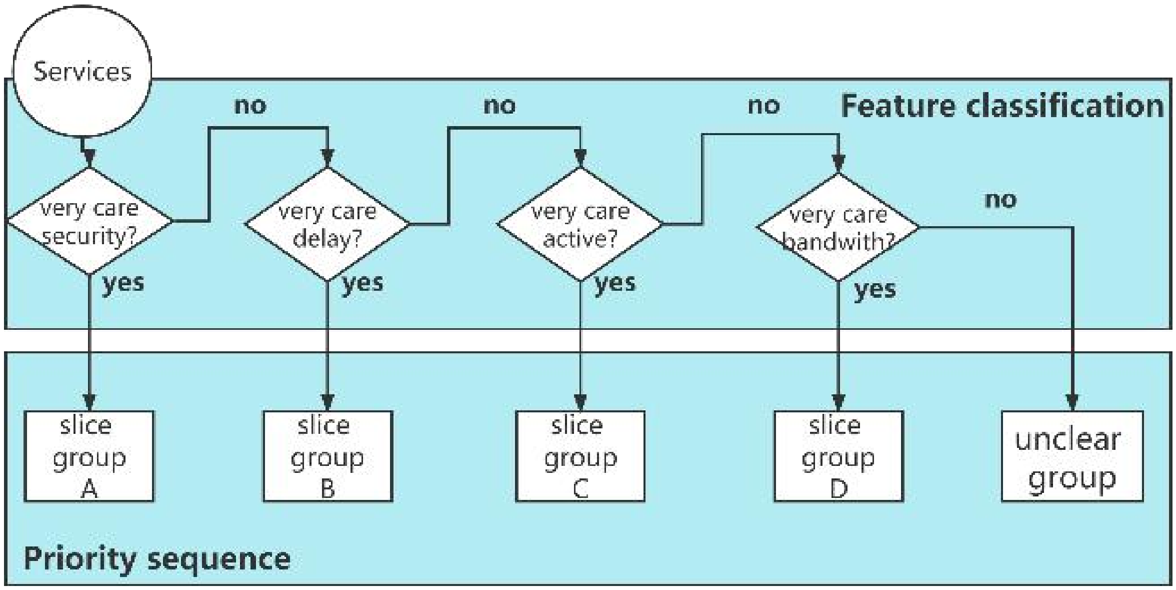

Fig. 2. Group-level flow chart

At the group-level, we apply groups. Group-level considers the complexity of the future communication network. To cope with the changes and specific requirements of the network, this solution can be solved by adjusting and changing the settings of the group-level. Such an approach enhances the compatibility and robustness of the scheme.

Fig. 2 shows the flow at the Group-level. At the group-level, we plan to do two parts, Feature classification and Priority sequence. The main task of this level is to classify services in the network for the first time and group them by the main characteristics of services. This level is the preprocessing work. After this level of classification, users with similar main feature services will be classified into the same slice group, which will speed up the processing speed of the slice-level and improve the accuracy of classification. At this level, we set the priority sequence of features, and the order of the priority sequence can be changed according to the network situation. This level introduces the K-means algorithm to assist processing.

Feature classification is a preprocessing. This part classifies all services in the scenario for the first time, which can reduce the complexity (number, difference) of the services that need to be processed at the slice-level. The judgment conditions of this part are pre-set or algorithm-assisted settings. Reducing the service complexity can reduce the classification time of the next-level clustering algorithm and improve the accuracy of the classification results.

Priority sequence is to sort the priority of all services in the network environment. In this way, the services that need to be prioritized in the network environment can be evaluated and filtered out. The finite sequence in the figure is in our envisaged network environment, where high-security and low-latency services are prioritized. This kind of service scenario is common in some application scenarios of Internet of Vehicles.

3.2 Slice-level

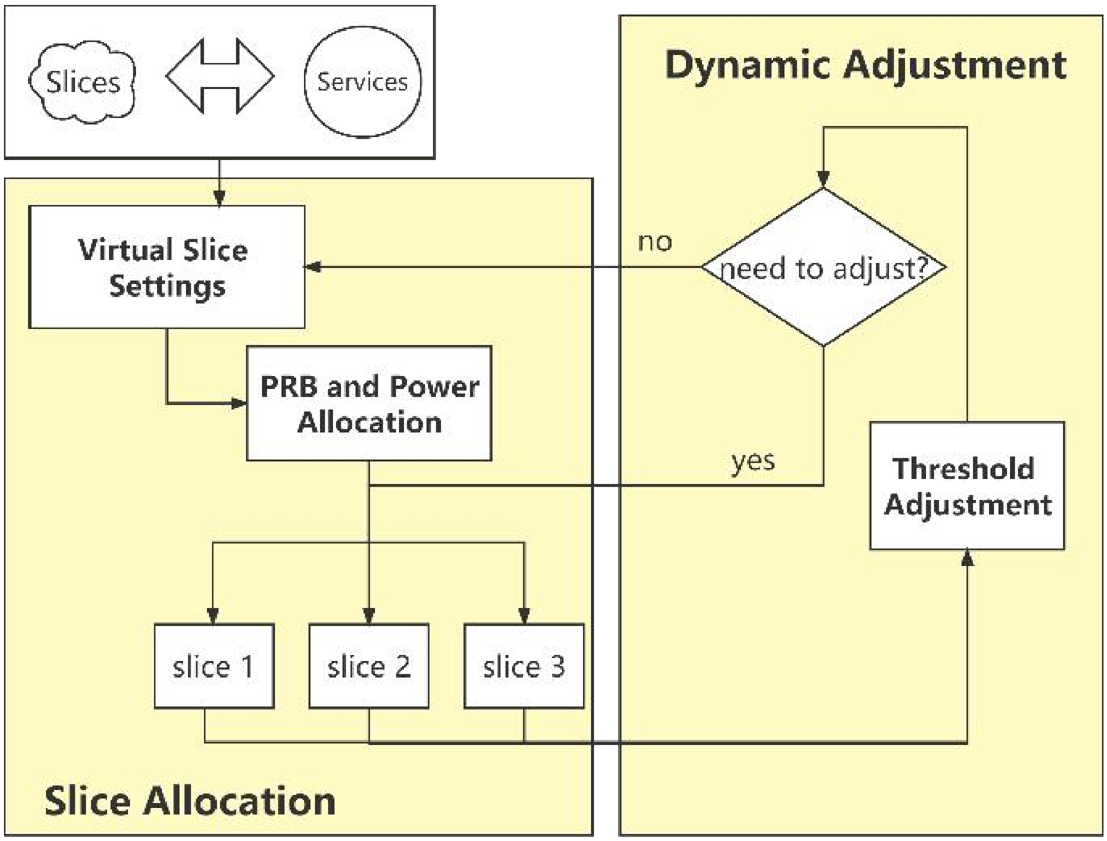

Fig. 3. Slice-level flow chart

At the slice-level, we focus on the granularity and stability of slices. Like the traditional slicing scheme, our scheme can dynamically adjust the parameters of slicing according to the situation of the network service. But our scheme slice granularity will be finer, reaching L5. To cope with the frequent adjustment of slices caused by increasing slice granularity, we set a threshold.

Fig. 3 shows the flow in Slice-level. This level will further subdivide services into classes of appropriate size, allocate slices for each class, and finally divide frequency band resources for each slice. This level also sets the steps of slice allocation and dynamic adjustment.

In the slice allocation step, it is to allocate slice groups according to the number of services. Dynamic Adjustment is now a necessary part of all network slicing. We have adopted the method of setting the threshold, which can adjust the adjustment frequency of slices, which increases the stability of the scheme. This threshold can adjust the slice size to ensure that the slice granularity is not too fine, leading to trivial phenomena. In addition, this threshold will also be used as a judgment condition for dynamic adjustment.

In the dynamic adjustment step, a determination is made in each time slot. If the service change does not exceed the threshold interval, the slice will not change. On the contrary, according to the change, the number of slices and distribution will be adjusted. The scheme can control the frequency of slice allocation by adjusting the threshold, so that the entire network is in a state of dynamic balance. If the slice changes too frequently, the burden on the entire network will increase and the communication quality will decrease. If the slice changes too slowly, the sensitivity of the entire network will be reduced, and it will not be able to respond quickly to emergencies.

3.3 Characteristic Parameter

TABLE I. SERVICE CHARACTERISTIC PARAMETER SET

Service Characteristic Parameter Set	
Definition	**Range**
Bandwidth	0.5Mbps~50Mbps
Min. required bit rate	0.5Mbps~50Mbps
Min. required latency	10ms~1s
Security and reliability	5 levels
Mobility	5 levels
Number of devices	10~100k(km^2)
Network energy efficiency	5 levels
User experienced data rate	0.5Mbps~50Mbps
Coverage	5 levels

Under the trend of high real-time requirements and high-reliability requirements of 5G services, the traditional TCP protocol is gradually abandoned, and the evolution trend of the UDP-based protocol is obvious. User Datagram Protocol is a transport layer protocol in the OSI reference model. It is a connectionless transport layer protocol that provides transaction-oriented simple and unreliable information transfer services. It is very fast in passing data and is relatively secure. The UDP protocol will bring the advantages of high throughput and low latency. According to this feature, and referring to 3GPP TR38.913, we selected 9 parameters that describe various services, such as Min. required bit rate, Min. required latency, Bandwidth, Number of devices, Security and reliability, etc.

This includes the main parameters of network slice classification, such as Min. required latency(uRLLC), Number of devices(mMTC), Mobility(eMBB), and also includes some characteristics of services in the future network, such as Security and reliability, User experienced data rate. Taking these parameters in the simulation can describe each service more vividly and help to refine the slice granularity. For some parameters that are not specified by clear protocols, in order to distinguish service characteristics, we adopt a hierarchical method, such as Network energy efficiency, Coverage.

It is worth mentioning that when performing the clustering algorithm, not all parameters are selected for classification. This is because of the limitations of the service itself and the consideration of actual effects. In this scheme, three parameters are generally selected as the classification basis during simulation. In most application scenarios, the parameters of user services can generally be used as a classification basis for about five features. Selecting different parameter combinations as the classification basis can achieve different classification results, and there is an optimal solution in these results.

3.4 Implementation

We propose corresponding solutions to the challenges that existing slices need to face.

- Solve the problem of priority of different QoS services by setting the priority sequence.
- Set the unclear group to simplify the problem of the high complexity of services characteristics.
- For the case of spatial-temporal dynamics of the network, a dynamic solution based on virtual slice capacity is given.
- Fairness indicators for the overall service are considered in the final Dynamic adjustment strategy.

For the judgment of the main users of the network, we choose to adjust the feature judgment conditions at the Group-level and prioritize the conditions that need attention. In addition, we set up an indeterminate group to allocate services that are not easy to make judgments, and services that are scattered or not worthy of attention.

In the Slice-level, we set the virtual slice capacity. The number of slices in the entire slice group is determined by the ratio of the number of user requests in the slice group to the capacity of the virtual slice. Then use the FCM algorithm to allocate each slice to the corresponding slice. For the dynamic adjustment strategy, we adopt the mechanism of checking every time slot. For each slice group, if the change services by the user are greater than the capacity of a single virtual slice size, the existing slice structure needs to be adjusted. This mechanism ensures that the slicing scheme can well cope with changes in the communication environment, and prevents frequent changes in the slicing configuration, that is, to ensure stability.

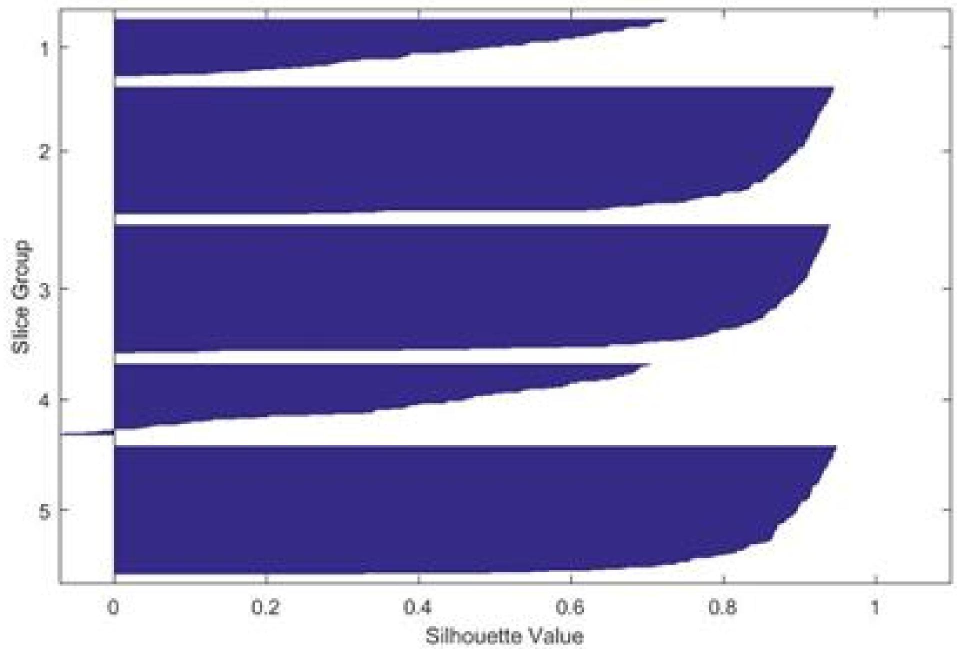

Fig. 4. Distribution of services in each group

Silhouette refers to a method of interpretation and validation of consistency within clusters of data. Positive values for these Silhouette values represent suitable user classifications, while negative values represent service categories with indistinct features or small numbers. This part of the service needs to be manually processed or grouped into the unclear group.

The Fig.4 shows the result of Feature classification at the group-level layer. Here we use algorithm-assisted preprocessing. The vertical axis represents the situation in each group, and the horizontal axis represents the similarity between each service and the standard features of this group. Among them, we can see that the fourth group has negative values, which shows that this part of the service is quite different from other services in this group. In this case we set the unclear-group. We assign all services with negative values to the unclear-group. This part of the service is usually sparse and very different from other services.

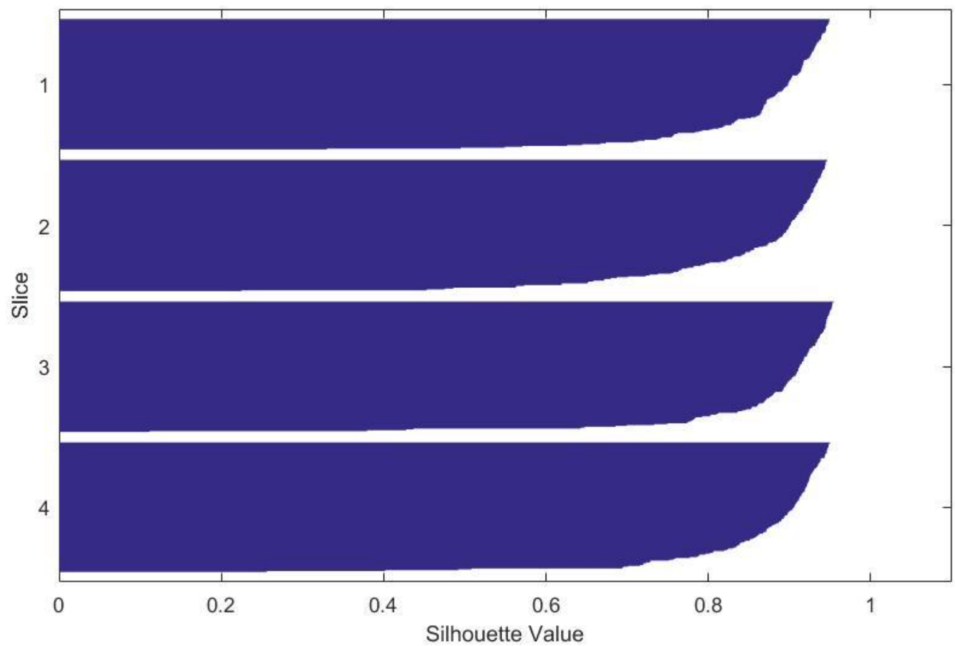

Fig. 5. Distribution of services in each slice

The Fig.5 shows the result of slice-level resource allocation. All slices of this graph originate from a slice group. The average distribution means that the characteristics of each service in the slice group are very close. There are no negative values here, indicating that the clustering results are good.

Under different simulation conditions, the number of services for each slice is different, which may be different from the set virtual slice capacity. In this case, the slice sizes in Silhouette picture will be different. In addition, setting different classification adjustment combinations will also result in different slice sizes.

Chapter 4

Experiment and Result

In this section, we consider two main research directions of network slicing, application scenarios and resource allocation.

We will simulate the final resource allocation results to show the contribution of this scheme. For the simulation of the scheme, we set the network processing capacity of 600 services, and the actual network services that randomly vary in the range of 600-800. Each service has all the features in the parameter table we set, and these features are generated according to the rules of the application scenario.

We have chosen three conventional schemes, fair scheme, smallest-class priority scheme, and largest-class priority scheme for comparison. By comparing related works, we summarize the core indicators they care about, and estimate the possible resource allocation results under this kind of goal.

4.1 Evaluation standard

In the case of simulation, we verify the differences in the three evaluation indicators of goal user satisfaction, fairness, and overall user satisfaction between this scheme and these three schemes.

TABLE II. EXPECTATIONS OF EVALUATION INDICATORS

Expectations of evaluation indicators				
Scheme	Fair scheme	Minority priority scheme	Majority priority scheme	Our proposed scheme
Goal user satisfaction	80%	72.72%	75%	93.75%
Fairness index	100%	97.6%	98.6%	99.6%
Overall user satisfaction	81%	85%	77.5%	80%

Here, we define QoS (Quality of service) to specify how the slicing scheme handles the service. Our proposed Expectations of evaluation indicators scheme optimizes the QoS of services belonging to the main task of the network, so we set the goal user satisfaction to focus on the optimization effect of our scheme. The US (user satisfaction) of these two indicators can be expressed uniformly as

$$US = \frac{\sum_{i=1}^{n} QoS_i}{n} \tag{5}$$

where n is the number of services in the slice group. In addition, QoS indicates whether the slice scheme satisfies the service, and is given as

$$QoS_i = \begin{cases} 1, Satisfy\ service \\ 0, Not\ satisfy\ service \end{cases} \tag{6}$$

The fairness index is another important indicator for studying network slicing schemes. We refer to other paper for the indicator used here. [14][15] The fairness index and overall user satisfaction draw on related papers on evaluating communication network optimization based on QoS indicators. The results can show the impact on the original communication network after the introduction of our scheme.

The most concerning indicator in this paper is a certain type of user satisfaction. This indicator represents the most concerned task in the network. With the help of the clustering algorithm, our scheme can filter out all goal services and flexibly process them in slices. This is what the current slicing scheme does not consider, and it is also the advantage brought by our digging out the slicing granularity of subdivisions.

4.2 Simulation results

Our scheme allocates slices according to service characteristics and can be applied to emerging 5G application scenarios. In the vehicle-to-infrastructure scenario, most services belong to uRLLC, such as infotainment, driver information service, and assisted driving. Our scheme can further subdivide them. Smart grids are built on high-speed two-way communication networks, including eMBB, uRLLC, and mMTC services. Better results can be obtained by managing suitable slices.

We intercepted a period of continuous time in the simulation and observed the changes of the three indicators of the four schemes. It is worth noting that the three comparison schemes perform well in certain indicators but produce significant deficiencies in the other two indicators. For the test data, we introduce relatively frequent and large changes and use this situation to simulate the future complex multi-tenant and multi-service coexistence network.

To prove the theoretical expectation, we use a line graph to track and analyze the resource allocation of the four scenarios. We show the changes of the four schemes for the three performance indicators we care about at the same time and under the same network environment. We found that with the change of the network environment, the simulation results will be different from the theoretical expectations at certain moments, but the overall simulation results tend to be infinitely close to the theoretical expectations.

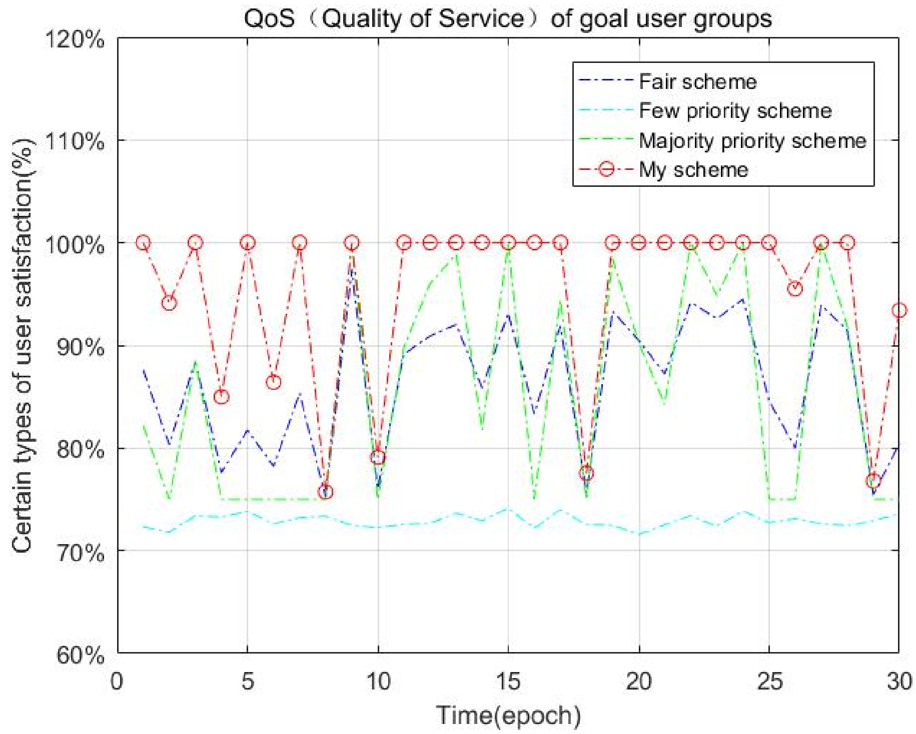

Fig. 6. QoS (Quality of service) of goal service group

From Fig. 6, we can see that in the case of limited resources, we give priority to meeting the identified goal service, and the effect is very significant. The simulation results are basically in line with our predictions. This is because when the classification of the FCM algorithm is effective, we can filter out the core services and give priority to meeting the resource requirements of these slices. The next best scheme is the blue Majority priority scheme. It prioritizes the needs of the most active part of the service, and in most cases, the most active service is equivalent to the core service. But in the simulation scenarios we designed, such equivalence does not always correct.

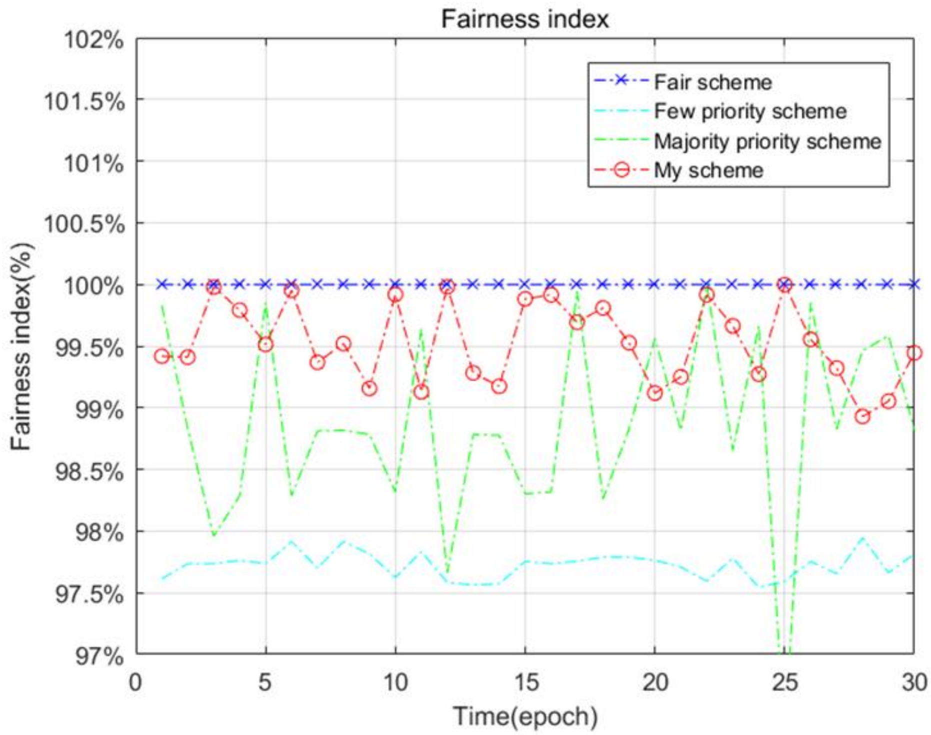

Fig. 7. Fairness index of the three schemes under the set conditions

For the other three schemes, the minority priority scheme performs the worst because it prioritizes the smallest group of users in pursuit of overall user satisfaction and does not conduct user characteristic analysis. The fair scheme performs evenly because the scheme is determined by the service composition of the current network and does not differentiate the priority of each service. The majority priority scheme is close to our scheme because it considers the priority of services and the number of services as the most important characteristic of users, but this scheme does not trade off overall user satisfaction and fairness.

All optimizations come at a cost. The Fig.7 shows the impact on other non-core users after we prioritize core service satisfaction. In the fairness index, our scheme is second only to the fair scheme and far superior to the other two schemes, which is the result of setting the priority sequence.

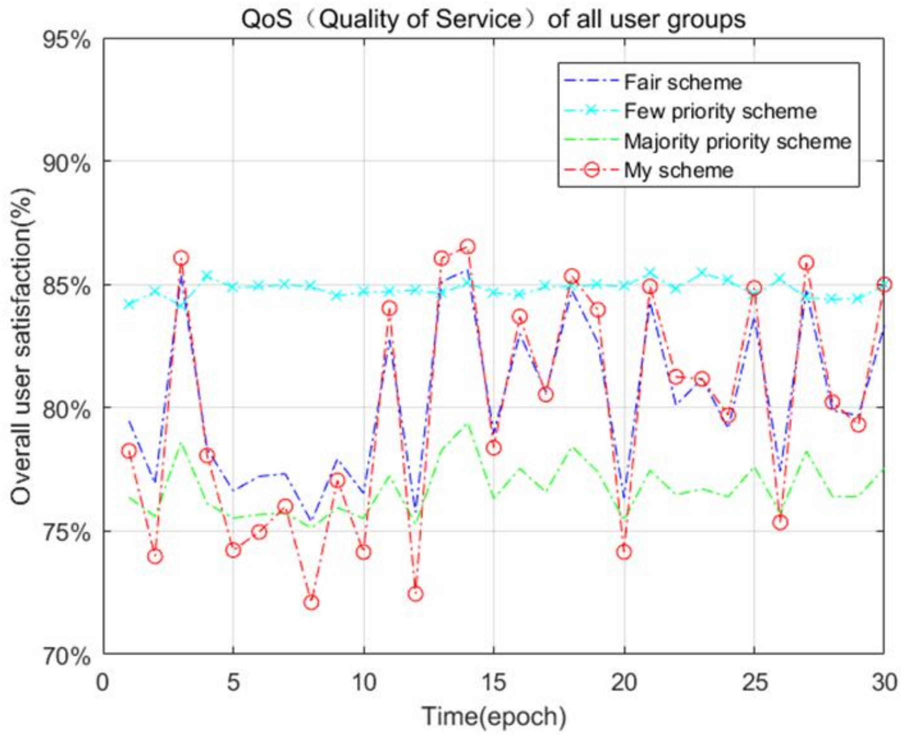

Fig. 8. QoS (Quality of service) of total service group

In the Fig.8, our scheme can be seen that the cost of achieving this effect is acceptable and even better than some schemes. In overall user satisfaction, our scheme is close to the fair scheme and outperforms the majority priority scheme, which is the result of setting the virtual slice. After our scheme is optimized, it has a certain impact on non-core services, but due to the use of groups and thresholds, it becomes a sub-optimal scheme.

Through the simulation results, we can see that after applying this slicing scheme, the satisfaction of the target user group has been greatly improved, and fairness and overall user satisfaction have been considered.

4.3 Application scenarios

Proposed scheme that the expected use cases for next-generation communications will require quality of service requirements. This diversity includes higher end-user data rates and cell-level throughput than today's networks to meet enhanced Mobile Broadband (eMBB) requirements, as well as Ultra Reliable Low Latency Communications (uRLLC) and massive Machine Type Communications (mMTC), these considerations focus on:

1) 5G networks need to support multiple QOS for various services, rather than building a network specifically for a limited set of QoS.

2) The network slice management and orchestration layer can manage the current network slice performance and coordinate the necessary resource allocation in the virtual environment to network functions (VNFs) in different domains (RAN, CN and transport) to meet the required QoS by evaluating the supporting QoS.

3) Effective E2E QoS negotiation requires application and service awareness on various network points.

4) Machine learning and artificial intelligence are the keys to enabling multi-point data sources and real-time streaming analysis in the future.

Our scheme allocates slices according to service characteristics and can be applied to emerging 5G application scenarios, such as smart grids and smart cities.

TABLE III. VARIOUS SERVICES IN SMART GRID SCENARIOS

Various Services in Smart Grid Scenarios				
Service component	Min. required latency	Bandwidth	Number of devices (system)	Application type
Service 1	≤15ms	≥2Mbps	6	uRLLC
Service 2	≤50ms	≥2Mbps	6	uRLLC
Service 3	≤200ms	20~50Mbps	4	eMBB/ uRLLC
Service 4	≤3s	1~2Mbps	100	mMTC
Service 5	≤50ms	1.13Mbps	6	uRLLC

Smart grid is a fully automated power transmission network capable of monitoring and controlling each user and grid node, ensuring bidirectional flow of information and electrical energy between all nodes in the entire transmission and distribution process from the power plant to the end user. Smart grid is built based on high-speed two-way communication network, using advanced sensing and measurement technology, control method and decision support system to realize the intelligence of power grid, which runs through the five links of power generation, transmission, transformation, distribution, and consumption. The low-voltage centralized copying scenario is a typical wide-connection business requirement. The main business fields of smart grid include eMBB, uRLLC and mMTC. From the perspective of network challenges, mMTC is a bigger challenge.

In the table, we omit the specific service names, select three service characteristics, and list the possible slice types allocated to each service in the traditional network slicing scheme. In Smart grid, there are many uRLLC services, which can be further distinguished in proposed scheme.

TABLE IV. VARIOUS SERVICES IN CONSUMER DEVICES SCENARIOS

Various Services in Consumer Devices Scenarios				
Service component	Min. required latency	Min. required bit rate	Number of devices (km^2)	Reliability and security
eMBB service supporting 4k+ video	<100ms	10 Mbps	≥10	High
eMBB – consumer portable devices	<100ms	10 Mbps	≥10	Very High
eMBB supporting 360-degree video	<100ms	50 Mbps	≥150k	Very High
AR/VR-based eMBB with low latency	<10ms	50 Mbps	≥150k	Suitable

The Internet of Things is undergoing a transformation. From the perspective of the telecom ecosystem, 5G IoT will bring about a new business model flip. There are many connectivity technologies and standards available today. These have been tested and many of them have been successfully used in various deployments around the world. For example, in a smart city, city halls can use smart parking sensors, smart trash can sensors, and smart street light sensors. Smart parking sensors are not only capable of directing drivers to vacant parking spaces (thus reducing driving time, pollution, etc.), but also correlating occupancy data with payment data. Smart bin sensors detect when bins need to be emptied exactly, improving pick-up schedules, and saving council money. Smart street light sensors can adjust light usage based on ambient light conditions and movement on the street.

In smart cities, there are various services for consumer portable devices such as smartphones, tablets, and laptops. They belong to eMBB services but have certain differences in specific characteristics. In the table, we select a part of eMBB user services, and select four service characteristics. Proposed scheme can further refine the slice granularity through the clustering algorithm.

TABLE V. VARIOUS SERVICES IN VEHICLE-TO-INFRASTRUCTURE SCENARIOS

Various Services in Vehicle-to-Infrastructure Scenarios				
Service component	Min. required latency	Min. required bit rate	Number of devices (km^2)	Reliability and security
Intelligent traffic signal control	>100 ms	0.1Mbps	≥100	Suitable
V2I – infotainment	<100ms	10 Mbps	≥20	Very High
V2I – driver information service	<100ms	0.5 Mbps	≥100	Very High
V2I – assisted driving	<100ms	0.5 Mbps	≥50	High

Vehicle-to-Infrastructure (V2I) is a communication model that allows vehicles to share information with components that support the National Highway System. Such components include overhead RFID readers and cameras, traffic lights, lane markings, street lights, signage and parking meters. V2I communication is usually wireless and bidirectional: data from infrastructure components can be delivered to the vehicle via an ad hoc network, and vice versa. In intelligent transportation systems, V2I sensors can capture infrastructure data and provide travelers with real-time advice on road conditions, traffic congestion, accidents, construction areas and parking space availability. Likewise, traffic management monitoring systems can use infrastructure and vehicle data to set variable speed limits and adjust traffic signal phase and timing (SPaT) to improve fuel economy and traffic flow.

In smart cities, vehicle-to-infrastructure V2I is an important business in the future. In times of high server load or emergencies, better results can be achieved by managing units of appropriate slices. In the table, we selected four common V2I services in smart cities, and selected four service characteristics.

Chapter 5

Conclusion and Future Plan

5.1 Conclusion

This paper proposes a slicing scheme for multi-tenant and multi-service networks. The motivation of the proposal is that with the development of communication networks, communication services will become more complex, and the existing slicing schemes are rough and cumbersome in dealing with this situation. This solution optimizes the granularity of slices by setting priorities and clustering algorithms so that slices can more appropriately adapt to complex services. And by setting the uncertain group at the Group-level and the virtual slice size at the Slice-level, the slice can be dynamically adjusted more flexibly. Finally, fairness and overall satisfaction are considered in resource allocation, which greatly improves the satisfaction of the goal service group.

Under the premise of considering the complexity of communication services, this paper proposes to use a clustering algorithm to optimize the slice granularity and points out that the combination of clustering algorithm and refinement of slice granularity brings new advantages in screening and identifying current network goal services. It further leverages the flexible characteristics of network slicing technology to process services in units of slices.

5.2 Challenges and Future Plan

In the future development of network slicing technology, the software-defined network direction will involve more applications of machine learning and artificial intelligence algorithms. The introduction of these optimizations will affect the original communication network and consume new network resources. In this regard, we need to introduce some parameters in the computer network to describe these changes and compare them. These are not considered in the traditional communication network field. Next, our scheme will give priority to some more prominent features, such as network resource allocation under the concept of computing power network, as well as the training time cost and possible network delay caused by the introduction of machine learning.

Research Achievements

Tianshun Fang, Hao Xue, Zhenni Pan and Shigeru Shimamoto, "A Lightweight Network Slicing Scheme for 5G Coexistence Network in Multitenant and Services,"2022International Symposium on Networks, Computers and Communications（ISNCC）

Reference

[1] C. Yang, W. M. Shen, and X. B. Wang, "The internet of things in manufacturing: Key issues and potential applications", IEEE Syst. Man Cybern. Mag., vol. 4, no.1, pp. 6–15, 2018.

[2] D. Feng, L. Lai, J. Luo, Y. Zhong, C. Zheng, and K. Ying, "Ultrareliable and low-latency communications: applications, opportunities and challenges", SCIENCE CHINA Information Sciences, vol. 64, no. 2, pp.120-301, 2021.

[3] R. H. Wen, G. Feng, J. H. Tang, T. Q. S. Quek, G. Wang, W. Tan, and S. Qin, "On robustness of network slicing for next-generation mobile networks", IEEE Trans. Commun., vol. 67, no. 1, pp. 430–444, 2019.

[4] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A comparison of neural network and fuzzy Clustering techniques in segmenting magnetic resonance images of the brain", IEEE Trans. Neural Netw., no. 5, vol.3, pp. 672–682, Sep. 1992.

[5] J. Mei, X. B. Wang, and K. Zheng, "Intelligent network slicing for V2X services toward 5G", IEEE Netw., vol. 33, no. 6, pp. 196–204, 2019.

[6] F. Li, J. Cao, X. Wang, Y. Sun, and Y. Sahni, "Enabling software defined networking with qos guarantee for cloud applications", in 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), pp. 130–137, June 2017.

[7] X. Xu, Q. Liu and E. Steinbach, "Toward QoE driven dynamic control scheme switching for time-delayed teleoperation systems: A dedicated case study." [Online], May 2017.

[8] Available:https://arxiv.org/abs/1705.05613

[9] Study on Management and Orchestration of Network Slicing for Next Generation Network, 3GPP, TR 28.801 V15.1.0,2018.

[10] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?", IEEE Trans. Commun., vol. 67, no. 10, pp. 7331-7376, 2019.

[11] Z. Hou, C. She, Y. Li, T. Q. S. Quek, and B. Vucetic, "Burstiness

[12] aware bandwidth reservation for uplink transmission in tactile Internet", in Proc. IEEE ICC Workshops, pp. 1–6, May 2018.

[13] M.J. Neely, M. Eytan and L. Chih-Ping, "Fairness and optimal stochastic control for heterogeneous networks", IEEE/ACM Trans Networking, vol. 16, no. 2, pp. 396-409, Apr. 2008.

[14] J. Jin, W. Wei-Hua and P. Marimuthu, "Utility max−min fair resource allocation for communication networks with multipath routing", Computer Communications, vol. 32, no. 17, pp. 1802-1809, 2009.

[15] M. Li, C. Zhenzhong and T. Yap-Peng, "A MAXMIN resource allocation approach for scalable video delivery over multiuser MIMOOFDM systems", Circuits and Systems (ISCAS), 2011 IEEE International Symposium, pp. 2645 – 2648, May 2011.

[16] H. Dahrouj, A. Douik, O. Dhifallah, T. Y. Al-Naffouri, and M.-S. Alouini, "Resource allocation in heterogeneous cloud radio access networks: Advances and challenges", IEEE Wireless Commun. Mag.,vol. 22, no. 3, pp. 66–73, Jun. 2015.

[17] P. Tang, et al. "QoE-based resource allocation algorithm for multiapplications in downlink LTE systems", 2014 International Conference on Computer, Communications and Information Technology (CCIT 2014). Atlantis Press, pp. 1011-1016, 2014.

Appendix I

Network slicing technology is an emerging research direction. When I was building a simulation system, I found that there was very little reference material. I'll list some of the code and ideas I've used below. Facilitate future research needs.

Clustering Algorithms and Preprocessing

```
[X,textdata] = xlsread('test.xlsx');
X=X';
X2=X(1:end,2:3);

%kmeans
idx = kmeans(X2,4);
[S,H]=silhouette(X2,idx);

%fcm

options =[2,200,1e-5,0];
[center, U, obj_fcn] = fcm(X2, 5, options) ;
id1 = find(U(1,:)==max(U));
id2 = find(U(2,:)==max(U));
id3 = find(U(3,:)==max(U));
```

```
id4 = find(U(4,:)==max(U));

id5 = find(U(5,:)==max(U));

%id6 = find(U(6,:)==max(U));

%id7 = find(U(7,:)==max(U));

%id8 = find(U(8,:)==max(U));


idx(id1)=1;idx(id2)=2;idx(id3)=3;idx(id4)=4;idx(id5)=5;

%idx(id6)=6;idx(id7)=7;idx(id8)=8;


silhouette(X2,idx);
```