Intelligent RAN Slicing Orchestration Framework For Healthcare Application in 5G

Srikanth Sailada, Vineeth Aitipamula, Suresh V , Anil Kumar Gupta

CDAC, Pune, India

ABSTRACT

With the increase in the number of internet-connected devices, there is a need to improve reliability, lower latency, higher capacity, more security, and high-speed connectivity. Every application has its performance metrics in terms of QoS parameters. Network slicing enables slicing an extensive broadband network into multiple virtual networks to serve applications more cost-efficiently. With the advancements in Artificial Intelligence (AI), the performance of network decision-making accelerates. In this paper, a dynamic RAN slicing framework is proposed for healthcare applications and a static Radio Access Network slice simulation model is developed by implementing KNN to predict the class. The deep slice data set from the public domain was leveraged to train the model and predict appropriate slice service types for healthcare applications.

Keywords: K-Nearest Neighbors (KNN), Network slicing (NS), Orthogonal Frequency Division Multiplexing (OFDM), Quality of Service (QoS)

INTRODUCTION

Mobile and internet connectivity has become a part of human lives. There has been an exponential increase in the number of mobile and other devices which use the internet. To escalate the capacity of the mobile networks and support very high data rates, 5G will extend the range of frequencies used for mobile communication. With the advancements in technology, there is also a need for high-speed connectivity and low latency for proper functioning.

As the requirement of every application differs, mobile networks have been divided into three main classes of services known as enhanced Mobile Broadband (eMBB), massive Machine Type Communication (mMTC), and ultra-Reliable Low Latency Communication (URLLC). Each of these classes has its requirements regarding throughput, mobility, reliability, latency, jitter, energy efficiency, etc. Packets of these different classes are marked using QFI (QoS Flow Identifier).

The Third Generation Partnership Project (3GPP) considers network slicing a critical enabling technology of 5G. Network slicing allows operators to cut the physical network into multiple instances (slices) by demands of applications to satisfy their Quality of Service (QOS) requirements. Software-Defined Network (SDN) and Network Function Virtualisation (NFV) together allow us to manipulate these slices as and when needed without touching multiple different physical pieces of equipment in the network. Each network slice can be isolated, have individual control and policy management systems. The inclusion of Machine Learning here will allow us to predict the future unknowns and take necessary actions.

Various simulation tools provide access for the simulation of communication networks. Some of the tools are NS3, NetSIM, Riverbed, MATLAB, etc. Each of these tools has its advantages. Some tools provide GUI-based simulation like NetSIM. Some tools are based entirely on code, such as NS3. MATLAB provides GUI and code-based simulation. MATLAB provides various modules for Machine Learning algorithms, and there are multiple modules for graphical representation.

This project aims to simulate a Machine Learning model in MATLAB to predict a class for the given parameters and analyze the output for the given parameters.

RELATED WORK

A lot of experimentation has been taking place to enhance the capabilities of the 5G network. As a result, many simulation platforms are providing access to simulate 5G networks and their technologies.

Authors in (Anurag Thantharate, 2019) present a deep slice model which uses Random Forest (RF) algorithm to create an ML model and uses a convolution neural network (CNN) classifier to help decide which network slice to use for given input information. 65,000 unique input combinations of the dataset are used for both RL and CNN models. Slices are divided into four classes that are URLLC, eMBB, mMTC, and master slice. Whenever there is a heavy load on any slice or a disturbance in one of the three slices, the master slice is chosen. The master slice works as a backup for the other three classes.

Selecting a simulation tool to perform 5G simulations has become a significant task.

Many simulation tools perform network analysis, but very few have implemented 5G analysis, as discussed in (Christos Bouras, 2015). NetSIM is a platform with GUI for 5G end-to-end network simulation, but it is only available for commercial purposes. NS3 is an open-source tool that supports 5G simulations through the "mmWave Cellular Network Simulator module." NS3 supports Python and C++ to make code. Other network simulations tools such as OMNET++ and Riverbed Modeler are also discussed in (Christos Bouras, 2015).

Network slicing can be applied in two dimensions: Vertical network slicing and horizontal network slicing, as discussed in (Qian Li, 2016). Vertical slicing and horizontal slicing form independent slices in the network. The end-to-end traffic flow in a vertical network slice is between the core network and the terminal devices. On the other hand, the end-to-end traffic flow in a horizontal slice usually transits between edge devices. Slice pairing functions between radio, RAN, and Control Network (CN) is also discussed.

In modern communication systems, Orthogonal Frequency Division Multiplexing (OFDM) is an efficient modulation format. Research work focusing on OFDM techniques is discussed in (Umesha G B, 2018). OFDM improves the quality of long-distance communication by eliminating inter-symbol Interference (ISI) and improving the Signal-to-Noise ratio (SNR). In OFDM, ISI is countered by using long equalizers with high-speed signal processing. Paper (Umesha G B, 2018) provides a detailed explanation of the OFDM transmitter and receiver and the effect of OFDM on attaining high data rates.

Dynamic Slice Allocation Framework (DSAF) has been proposed in (Danish Sattar, 2019). DSAF can perform slice allocation and deallocation as well as provide on-demand intra-slice isolation. This framework is evaluated on a real testbed. DSAF provides user-interaction to request slices and any other required services that need to run on slices. It accepts single or multiple slice allocation requests, and it dynamically allocates them. Furthermore, a novel framework for real-time RAN network slicing resources sharing has been proposed in (M. Maule, 2020) (Lea Skorin-Kapov, 2010). The core idea behind this framework is the partitioning of each slice in different modes, according to the traffic load and custom Service Provider specifications.

ROLE OF 5G IN MEDICAL HEALTH

With the advancements in the 5G networks, the potential has increased to deliver various healthcare services anywhere, anytime. There are a wide variety of healthcare services such as Electronic Health Record (EHR), telemedicine services, health portals, Internet of Medical Things (IoMT), Remote/ Robotic Surgery, connected ambulance, HD virtual consultations, video-enabled prescription management, and many more. QoS requirements for few healthcare services are given below.

CONNECTED AMBULANCE:

A Connected ambulance equipped with telemedical devices allows the available crew to diagnose a patient at the scene effectively. A connected ambulance can vastly improve Healthcare services. The network requirements for a connected ambulance can be implemented using 5G capabilities. It needs low latency to send data and video in real-time to the hospital, as in emergency situations, split-second decisions can have a significant impact. High bandwidth is required to live stream video from emergency respondent body cams without losing quality or buffering. A network slice can be used for these types of emergency services. With respect to QoS parameters required, a URLLC class meets the requirement for a connected ambulance.

MULTIMEDIA CONFERENCING:

Multimedia conferencing can be used for various healthcare communication scenarios, including patient-doctor, doctor-doctor, and patient-patient interactions. A two-way HD video can be used between a patient and a doctor to conduct initial screening assessments, routine check-ups, therapy sessions, and many more. By conducting these appointments over the air reduces the burden on the patient and decreases the cost of each appointment. These interactions are considered to be “less critical” when compared to emergencies. These audio or video applications are highly jitter sensitive. An HD video needs higher bandwidth and reliability. An eMBB slice can be used for these types of applications.

REMOTE/ROBOTIC SURGERY:

In 5G, “telesurgery” is a context where a specialist can perform an operation from a remote location. A 5G enabled AR/VR headset is used to allow a specialist to watch in on a surgery taking place in real-time, guiding the in-person surgeon. It requires the transmission of both still and streaming images. In the case of robotic telesurgery, the key requirement is minimal delay time from when a surgeon’s hand movement is initiated, the remote manipulator moves, and images are shown on the surgeon’s monitor. This case of healthcare service is highly delay and jitter sensitive. It also requires high bandwidth to send extremely high-definition video. A URLLC slice meets the QoS requirements for a telesurgery application.

There are other healthcare applications as given in (Lea Skorin-Kapov, 2010) and (International Journal of Telemedicine and Applications, 2017). Implementation of 4G network slicing in e-health services based on priority is presented in (Lea Skorin-Kapov, 2010), whereas the role of 5G network slicing in e-health services is presented in (International Journal of Telemedicine and Applications, 2017). In order to meet the requirements of healthcare service traffic delivered over networks along with other traffic (such as audio and video calls, streaming, internet traffic, etc.), a QoS mechanism such as class-based traffic prioritization is necessary. Different e-health services require different Quality of Service (QoS) requirements. One requirement is delay tolerance that ranges from strict real-time and delay intolerant data transmission to delay-tolerant services. Another requirement is application data sensitivity to packet loss. Depending on the QoS requirements, there are three classes in 5G networks: eMBB, mMTC, and URLLC.

DYNAMIC RAN SLICING

In this paper, a Dynamic RAN slicing Framework is proposed for healthcare services. Figure 1 shows the proposed Dynamic RAN Slicing Framework. The input data is available from different e-health services. This framework uses Machine Learning (ML) algorithms such as KNN, SVM, Logistic Regression, etc., to train a model with the previous data available. Then, the ML model is used to predict the class for new data. After predicting the class, considering the availability of infrastructure resources and the current situation of the network (i.e., active devices and connected devices), the Resource Allocation Controller (RAC) will decide whether resource demands are achievable or not. The RAC mechanism ensures that the provision of the resources, their management, and automation are feasible for concurrent use in the corresponding 5G slices. Once RAC ensures the feasibility for the infrastructure to support the defined requirements, the slice allocation takes place.

RAC will allocate that slice for a new request if a particular slice is already enabled and is free to use. If there are no available resources for a slice for a particular application, the application gets denied. If the particular slice is not used, RAC will deallocate that slice.

SIMULATION MODEL

Network Slicing (NS) plays a significant role in improving the data rate and in decreasing latency. NS provides multiple paths for the data depending on the QoS requirements. Machine learning is used widely in the present generation to work on the classification of massive data.

In this model, a K-Nearest Neighbor (KNN) model is developed to predict the class of input data. KNN is a straightforward machine learning model for classification to implement when compared to other models. The data set for training is also small, so that KNN would be a good choice. In this project, K = 5 is chosen to have less error and high accuracy for prediction.

MATLAB is used as a simulation tool since it provides support for various machine learning algorithms. It is used to predict the class and also to simulate the transmitter and receiver blocks of the model. In the transmitter block, STBC encoding, OFDM modulation, and equalizer implementation at the receiver side are simulated in MATLAB.

For static slice model, three classes are eMBB, URLLC, mMTC are considered. A sample of dataset used in this prediction model is given in Table I. Here, reliability and latency (in ms) are considered as the parameters for the simulation. The dataset required for the simulation is taken from (DeepSlice & Secure - 5G - 5G & LTE Wireless Dataset, 2020), which is suggested in (Anurag Thantharate, 2019). It is modified to match the requirements of the simulation.

Table I: Sample Data Set for Slice Prediction

S.No.	Reliability	Latency	Class
1	0.1	50	eMBB
2	0.1	100	eMBB
3	0.0001	10	URLLC
4	1	50	mMTC
5	0.0001	300	eMBB

Figure 2 shows us the flow diagram of MATLAB simulation model. First, a KNN model is created by using a training data set. Then, since there are three classes, three OFDM channels are created which represent those slices.

Figure 3 shows us the transmitter model simulated in Matlab. After KNN classification according to the class, the number of subcarriers for OFDM modulation is decided. As shown in the transmitter model, data is serial to parallel converted, and then STBC encoding is done using the encoding matrix for two transmitting antennas. Next, IFFT is performed for OFDM modulation, and again this parallel stream is converted as a serial stream. Finally, it is passed through a low pass filter for digital to analog conversion. Finally, this transmitted signal is passed through the Rayleigh fading channel and received at the receiver end.

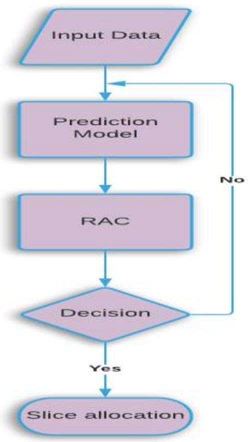

Figure 1. Dynamic RAN Slicing Framework

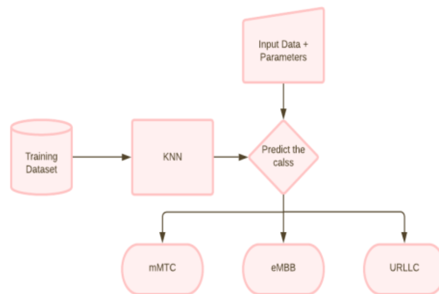

Figure 2. Flow Diagram of the Static RAN Slicing Model

As shown in Figure 4, the received signal is the first analog to digital converted, and demodulation is performed by implementing FFT. After parallel to serial conversion of this stream, equalization is performed by different signal detection algorithms according to the class requirements.

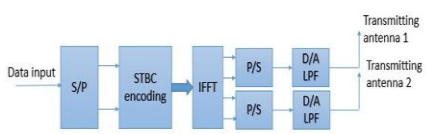

Figure 3. Transmitter Model

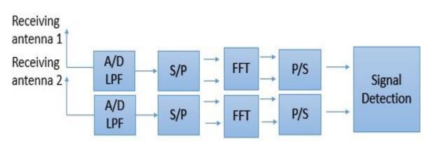

Figure 4. Receiver Model

Differentiation of classes is done by maintaining variable data rates and by using different equalizers.

DATA RATE:

It is one of the QoS parameters which can be considered for slicing. Each slice requires a different data rate. In this project, variation in data rate is compared with variation in the number of subcarriers. By varying the number of subcarriers, the amount of data transmitting in a given time can be varied, as discussed in (Umesha G B, 2018). As presented in (URLLC, difference between 5G eMBB mMTC, 2012), eMBB needs a high data rate compared to URLLC and mMTC, so 64 subcarriers are provided for the eMBB slice. On the other hand, since mMTC needs less data rate, so 16 subcarriers are offered for mMTC. The URLLC slice is simulated with 32 subcarriers.

EQUALIZER:

Equalization is a process of compensating for the effect caused by a channel on the transmitted data. Different equalizers have different orders of compensating the channel effect. Signal Detection has been implemented by using ZF, MMSE, and ML equalizers in this project. These equalizers increase the reliability, reducing the bit error rate (BER) by a different amount. Usually, URLLC needs high reliability. Therefore, a Maximum Likelihood (ML) equalizer is used, which has a high rate of reducing the probability of error. For the eMBB class, Minimum Mean Square Error (MMSE) equalizer is used, which minimizes the mean square error (MSE). Finally, Zero Forcing (ZF) equalizer, which removes all ISI, is used in mMTC.

For MMTC class, BPSK modulation is used with 16 subcarrier OFDM passed through a simulated Rayleigh fading channel, and detection at the receiver end is done using a ZF equalizer.

In equation (1), the pseudo matrix is calculated by using pseudoinverse $(H^H.H)^{-1}$ and Hermitian H^H , where H is the channel matrix. The weight matrix for ZF, G_{ZF} is calculated by multiplying a pseudo matrix (H^+) with a fraction to normalize the weight matrix as shown in equation (2), where $\mathrm{E_s}$ is the energy of the transmitted symbol.

$$H^+ = (H^H.H)^{-1}.H^H \tag{1}$$

$$G_{ZF} = \sqrt{\frac{2}{E_s}}.H^+ \tag{2}$$

For eMBB class, BPSK modulation is used with 64 subcarrier OFDM passed through a simulated Rayleigh Fading channel, and detection at the receiver end is done using an MMSE equalizer.

Using equation (3), the weight matrix for MMSE is calculated. I_{n_t} is identity matrix of order number of transmitting antennas, whereas the number of transmitting antennas is two. Here σ^2 is variance of the Channel coefficients, and ρ is scaling factor. Using this weight matrix (G_{MMSE}), the MMSE algorithm is implemented.

$$G_{MMSE} = H^H(HH^H + \frac{\sigma^2}{\rho} I_{n_t})^{-1} \quad (3)$$

For URLLC class, BPSK modulation is used with 32 subcarrier OFDM passed through a simulated Rayleigh fading channel, and detection at the receiver end is done using an ML equalizer.

Code vector using ML equalizer is detected using the below equation. Equation (4) shows the algorithm for Maximum Likelihood detection. Here error is calculated for every codeword with the received codeword, and then code word with minimum error is considered.

$$\hat{s}_{ML} = \underset{\hat{s}\epsilon\ codeword}{\text{argmin}} ||y - H\hat{s}|| \quad (4)$$

Figure 5 shows the variation of Bit Error Rate (BER) against Signal to Noise Ratio (SNR) to show the variations of BER in different classes. As shown in the plot, BER for URLLC is low when compared to eMBB and mMTC. So, it shows that the URLLC slice provides more reliability when compared to eMBB and mMTC.

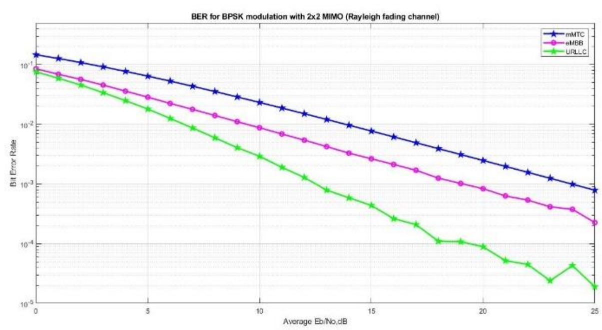

Figure 5. BER vs SNR comparison for Network Slices

A KNN prediction model is developed using Python, similar to the MATLAB model, to train and predict the class label. Figure 9 shows us the prediction of class for given input parameters.

```
C:\Users\Srikanth\Desktop>python Slice_predictor.py
Enter Reliability[%]:0.0001
Enter Latency[ms]:1.5
URLLC
```

Figure 6. Prediction Model Output Using Python

CONCLUSIONS

In this paper, a Dynamic Slicing Framework is proposed for healthcare applications. A prediction model for static network slicing according to the QoS parameter requirement is developed. Python and MATLAB are used to develop prediction model. Transmitter and receiver blocks of OFDM channel for each slice are simulated in MATLAB. Differentiation of these slices is done by varying data rates using variable OFDM subcarriers and different equalizer techniques. BER performance and data rate achieved shows the fulfillment of application requirement. This model implementation performs different reliability and latency static slices according to the application.

REFERENCES

Anurag Thantharate, R. P. (2019). DeepSlice: A Deep Learning Approach towards an Efficient and Reliable Network Slicing in 5G Networks. IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference.

Christos Bouras, G. D. (2015). A Comparative Study of 4G and 5G Network Simulators.

Danish Sattar, A. M. (2019). Dynamic Slice Allocation Framework for 5G Core Network.

DeepSlice & Secure - 5G - 5G & LTE Wireless Dataset. (2020). Kaggle Website: https://www.kaggle.com/anuragthantharate/deepslice

International Journal of Telemedicine and Applications. (2017). Website: https://stlpartners.com/digital_health/10-5g-healthcare-use-cases/

Lea Skorin-Kapov, M. (2010). Analysis of QoS Requirements for e-Health Services and Mapping to Evolved Packet System QoS Classes. International Journal of Telemedicine and Applications.

M. Maule, P. -V. (2020). Dynamic partitioning of radio resources based on 5G RAN Slicing. GLOBECOM 2020 - 2020 IEEE Global Communications Conference.

Qian Li, G. W. (2016). An end-to-end network slicing framework for 5G wireless communication systems.

Umesha G B, S. S. (2018). OFDM System for High Data Rate and High Mobility. International Journal of Engineering Research & Technology.

URLLC, difference between 5G eMBB mMTC. (2012). RF wireless world Website: https://www.rfwireless-world.com/Terminology/5G-eMBB-vs-mMTC-vs-URLLC.html