

2023

## Efficient Resource Allocation of Latency Aware Slices for 5G Networks

Mai Ibrahim

Follow this and additional works at: <https://digitalcommons.aaru.edu.jo/erjeng>

---

### Recommended Citation

Ibrahim, Mai (2023) "Efficient Resource Allocation of Latency Aware Slices for 5G Networks," *Journal of Engineering Research*: Vol. 7: Iss. 1, Article 11.

Available at: <https://digitalcommons.aaru.edu.jo/erjeng/vol7/iss1/11>

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Journal of Engineering Research by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact [rakan@aar.edu.jo](mailto:rakan@aar.edu.jo), [marah@aar.edu.jo](mailto:marah@aar.edu.jo), [u.murad@aar.edu.jo](mailto:u.murad@aar.edu.jo).

# Efficient Resource Allocation of Latency-Aware Slices for 5G Networks

Mohamed T. Faheem Saidahmed, Mai Z. Ibrahim, Nada M. ElShennawy

Dept. of Computer and Control Engineering, Faculty of Engineering, Tanta University, Egypt

Emails: mohamed\_ahmed1@f-eng.tanta.edu.eg, mai\_zaky@f-eng.tanta.edu.eg, nada\_elshennawy@f-eng.tanta.edu.eg

**Abstract:** It's noticed that 5G mobile networks are considered to be an emerging technology that serves multiple users with varying types of applications having different quality of service (QoS) needs. Network slicing enables us to accommodate diverse services on the same infrastructure by using multiple virtual networks on the same physical infrastructure of the network. In this paper, a resource allocation framework to fairly share the network resources among different slices subject to delay sensitive requirements based on the priority factor of each slice to use the available radio resources efficiently is proposed. This priority factor depends on the weight of each slice and considering the quality of service of each one. The proposed framework ensures that each slice can share the high limits of allowable resources to achieve the least allowable latency as it is the most significant feature in 5G cellular networks. Packet loss and packet scheduling delay are used as performance metrics when compared with other existing resource allocation algorithms. The simulation results validated that our framework could serve the delay-sensitive slices with the least allowable delay and guaranteed throughput.

**Keywords:** 5G mobile networks, Network slicing, Resource allocation.

## I. INTRODUCTION

5G networks have a pivotal role in mobile networking systems. The growth of mobile communications and its incorporation into our daily life has had an apparent impact on social development and the economy in the last few years. This leverage has turned the 5G layout to be one of the most important poles of the future 2020 society [1-3]. Most devices and Internet of Things (IoT) services demand an Internet connection, that requires providing some network requirements (e.g., latency) and minimum processing in order to best realize the requested service [4]. In consequence, 5G network is anticipated to serve various users that often have conflicting demands with growing service complexity. Simultaneously, the requirement of accessing those services has been often omnipresent, whereas the demand for integration is intensifying.

The fifth generation of mobile networks is expected to deal with the growth of data traffic over cellular networks with diverse quality of service (QoS) requirements. It is very important for the mobile network operators to improve the throughput and utilization to meet the quality of service (QoS) requirements of 5G [5]. 5G supports two service classes to meet QoS requirements: Ultra-reliable and low-latency communications (URLLC) and enhanced mobile broadband (eMBB).

URLLC concentrates on the reliability of low-latency packets to support mission-critical applications such as industrial automation, remote surgery, and vehicular

communications. But eMBB focuses on higher data rates and massive machine- type communications (mMTC) [5,6].

The important challenge of 5G is the accommodation of the aforementioned services with the same network infrastructure. So, a promising solution is introduced by using a slicing mechanism. Slicing refers to the abstraction of the physical network infrastructure into end-to-end logical networks to operate as autonomous networks [7, 8].

5G slicing is enabled using both SDN (software- defined networking) which decouples the data plane from the control plane and NFV (Network Function Virtualization) architectures [9].

5G network architecture is divided into RAN (Radio Access Network) and CN (Core Network). These two architectures supported the slicing mechanisms to serve the distinct types of services [6].

This context is concerned with RAN slicing. It has many challenges such as handling the heterogeneous traffic demands with a variety of tenants using its limited radio resources. By using slicing, many isolated networks or slices with a particular radio resources are constructed. Each slice consists of virtual base station (BS) to cover it and manage the resource blocks (RBs) allocation between slice users.

To satisfy a spectrum efficiently and elasticity for heterogeneous users demands and services, efficient resource allocation mechanism between constructed slices is the major challenge in deploying 5G [10,11]. Deploying a resource allocation mechanism to balance the utilization, latency, and user satisfaction through all types of traffic is a critical task.

An efficient resource allocation mechanism to utilize and prohibit the wastage resources is required.

In this context, the research work address the RAN slicing problem and focus on the efficient allocation of resource blocks (RBs) to each slice according to its demands with QoS guarantee and increase the resource utilization. Our target will be on the latency aware slices and trying to allocate this type of slices first without making a considerable effect on the other slices on the network using the priority factor of each slice.

### A. Contribution

The contribution of this paper can be summarized:

- An efficient resource allocation mechanism for 5G slicing is proposed by sharing the network resources through various slices based on the use of priority factor.
- A performance evaluation for the proposed mechanism is introduced by comparing it with another two presented approaches.

## B. Related Work

Owing to the growth of data traffic services and types, 5G networks and other mobile networks have several requirements as provisions of radio resources for many mobile devices and latency. In literature, many solutions are introduced in concerned with 5G slicing to support strong allocation of radio resources and QoS guarantee.

In [12], a resource management approach for Mobile Edge Computing (MEC) called RELIABLE is implemented. It has two steps: resource aggregation and resource allocation in such a way to increase the availability of resources requested in MEC. It uses mobility prediction to minimize the unnecessary reallocations that are done by user mobility, and the service time for appropriate decision making. It also considers a multi-criteria mathematical method in order to deal with resource allocation in MEC and decide when and where to allocate MEC services. The results prove that: RELIABLE will be stable even with diverse loads of resource requests, so, it can serve a higher number of services and deny a lower percentage of services.

In [5], GPS- based scheduling approach with two operation modes is proposed. Its modes are static sharing resource (SSR) scheme and dynamic sharing resources (DSR) scheme. In SSR, allocation of the smallest quantity of transmission rate to a slice based on the QoS level is used. In DSR method, dynamically shares of the unutilized part of the C- RAN transmission rate are considered. DSR outperforms SSR in both delay and throughput.

The NVS approach mentioned in [13] introduces slicing on RAN resources that support two types of requirements: resource-based provisioning which made resource allocation in terms of the resources of the base station and bandwidth-based provisioning allocation in terms of the aggregate throughput obtained by the slice's flows. They emphasized on three basic requirements: isolation, customization, and efficient resource utilization. Utility functions are defined to permit coexistence of bandwidth- based provisioning and resource-based slice provisioning, then modeling the problem in a utility-optimization framework. Finally, an optimal resource allocation algorithm is introduced and show that NVS demonstrate can simultaneously meet the slices' requirements with bandwidth-based and resource-based reservations. None of the above types consider the effect of QoS as the latency on the RAN resources.

Authors in [14] provide a radio resource slicing system on three types of slices i.e. (a) Fixed slices that requests dedicated physical resources, (b) dynamic slices which asking for aggregate throughput, and (c) on-demand slices which have strict requirements regarding of latency with a certain quality of service (QoS) requirements. They also derive a utility-based slice scheduling method for 5G systems and derive the weight for each slice. Slice with the highest weight will be served first. Finally, they apply a resource allocation scheme to the three types of slices. Based on the fixed slices Service Level Agreement (SLA), they allocate the dedicated requested resources. Then, they allocate necessary resources following the QoS requirements of on-demand slices, and after that, they schedule resources for the dynamic slices. Authors proved that they maintain customization, isolation, and QoS guarantees, and

also outperforms other existing approaches like NVS. However, taking QoS into consideration, when the channel quality decreased, the delay will be high.

A resource allocation scheme that has dynamic priorities for Machine type communication (MTC) is proposed in [15]. The authors addressed the problem of scheduling MTC devices. This scheme has two main phases: medium access, and resource allocation. At medium access phase, it assigns priorities based on MTC devices' wait time and the broadcast nature of wireless signals/. At resource allocation phase, it assigns resources to MTC devices in the cellular band based on total induced transmission delay, Signal to Noise ratio (SNR), and transmission-awaiting MTC devices. This scheme is designed with common security aspects: authentication, access control, nonrepudiation, data confidentiality, data integrity. Success probability and outage probability analysis are employed to assess the proposed schedule and aggregation schemes. The results showed that the outage probability significantly reduced using the proposed scheme and also the probability of successful MTC transmissions increased.

A deep learning framework is developed in [16] to approximate the optimal resource allocation policy which can minimize the power consumption of a base station through allocation of power transmission and bandwidth optimization. A cascaded structure of neural networks (NNs) is used in which the first NN will approximate the optimal allocation of bandwidth, and the second NN will output the power transmission required to achieve the QoS requirement as well as bandwidth allocation. This work considers that the distribution of wireless channels and its services are non-stationary, and to update NNs in non-stationary wireless networks, a deep transfer learning is applied. They prove that the cascaded NNs made a better performance than the fully connected NN concerned with QoS. They also deduced that the number of training samples used to train the NNs is reduced because of using deep transfer learning.

A dynamic virtual resource allocation scheme is proposed in [17] based on RAN slicing in the uplink communications to satisfy QoS. It uses the optimal control policy based on the equivalent Bellman equation on the basis of subchannel allocation Q-factor with the value iteration algorithm in order to reduce the computational complexity. Also, it optimizes the value functions and Lagrangian multipliers (LMs) using distributed online stochastic learning algorithm. The authors prove that the proposed optimization scheme can improve the user performance compared with Random Control Scheme and queue-state information (QSI) -Based Scheme.

Authors in [18] propose a latency-aware dynamic resource allocation problem as a maximum utility optimization problem. It addresses the several service requirements through using three main slices: enhanced mobile broadband, massive machine-type communications and, ultra-reliable low-latency communications. It also focuses on both the received data rate QoS and latency requirement in the three slice use cases. It adopts the hierarchical decomposition technique in order to reduce the complexities while solving the optimization problem. In addition, a genetic algorithm (GA) intelligent latency-aware resource allocation scheme (GI-LARE) is proposed in a multi-

tenant, multi-tier heterogeneous network. The GI-LARE outperformed the static slicing resource allocation (SS) approach.

A survey that provides a comprehensive study of RL technique and DRL technique is introduced in [25] in order to perform resource allocation in 5G network slicing. This survey reviewed some research works and discussed the feasibility of the suggested solutions in admission control, resource scheduling and allocation challenges. RL and DRL network slicing solutions of resource management have achieved encouraging results in simulations. But their practicability need to be estimated, specifically in real time scenario, with ultrahigh volume of data and applications in various scale networks.

A distributed edge computing approach in [26] is proposed to achieve QoS-based resource allocation. This optimization problem maximizes the average throughput of the system considering the delay of the services and the constraints data rate through dividing this problem into global and local constraints. Authors calculate the local estimates by using the Lagrange method, then these values are sent to a consensus algorithm in order to find the global maxima.

### C. Organization of the paper

The paper is organized as follows: Section II introduces a review of the 5G system model and the types of slices. Section III presents the proposed resource allocation mechanism for slices. The simulation results that validate the efficiency of our proposed framework and the performance evaluation is provided in section IV. Finally, the conclusions of the paper are summarized in section V.

## II. 5G SYSTEM MODEL OVERVIEW

A network slice is considered to be a heterogeneous set of resources that are concatenated and optimized with each other in order to serve a nominated service. In 5G network model, the physical resources (i.e., core networks (CNs), radio access networks (RANs)) are divided into various virtual parts, thereby forming diverse network slices [19, 20]. In RAN slicing scenario of multi-cell cellular network, there are several BSs, and each one will be sliced to serve different tenants or users. First, the network determines the type of service of each user equipment (UE) and then it is assigned to the corresponding slice based on its demands.

Focusing on the next generation (NG) RAN, it can incorporate all the radio access functionalities. It consists of a next generation node (gNB) and/or an evolved node of Long-Term Evolution (eLTE eNB). Lately, 3GPP has proposed the term of (gNB) for the 5G NR base station. The eLTE eNB is used to make it possible to coexistence of the LTE and the 5G. gNB needs to be configured to serve different RAN behaviors through using the correct allocation of radio resources by realizing of RAN slices [21]. 5G supports two slices based on the services traffic types: Time-Sensitive Slice, Dynamic slice.

Time-Sensitive Slice is the one that concerned with QOS. Its mission is to achieve specific throughput within a rigid time requirement. So, it has a high priority. It is designated with more resources to accomplish its tasks with least allowable latency. It

used in serving Ultra-reliable and low-latency communications (URLLC) service delivered in the 5G.

Dynamic slicing achieves throughput by aggregating the bit rates of the available resources in the network. The delay has less importance in this slice. (eMBB) scenario uses this type of slice.

Another 5G slicing standardization is introduced by the Next Generation Mobile Networks (NGMN) [22]. It works based on 3-layer perspective as shown in Figure 1. The first layer is the service instance layer, which supports services includes the tenants. The second layer is network slice instance layer which provides the characteristics of network required by diverse service instances and sub network instance comprises the required set of network functions such as: processing functions, switching functions and IP routing functions and also it contains strips which mean slices that are created by the orchestrator for various tenants. Finally, resource layer which represents the logical or physical resources allocated to network slices.

Service Level Agreement (SLA) is used as a contract between the slice descriptor and orchestrator to indicate the slice requirements or the required resource parameters such as the guaranteed bit rate (GBR), maximum delay threshold, isolated CP processing and customized admission control rules [23,24]. In order to establish multiple virtual BS (for each slice) into one physical BS, there are some circumstances that should achieve: isolation of requested resources, utilizing resources efficiently and guaranteeing the performance level subjects to the service needs.

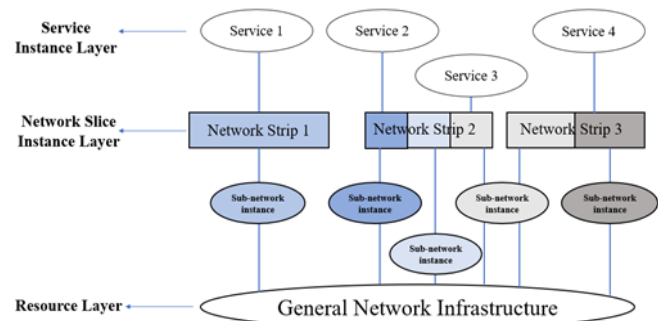


Figure 1. NGMN 3-layer of network slicing concept.

## III. PROPOSED FRAMEWORK

In this paper, an efficient resource allocation mechanism between slices based on latency for 5G networks is proposed, to enhance the overall performance of the network even if the channel quality degraded. The formulation of On-demand slices in [14] was used as the basis to our designed framework. First of all, we need to assign a weight for each slice which reflects the significance of each one according to the requirements mentioned in the SLA.

To determine the weight for each slice, a utility function and its helper equations was used as in [14]. Based on the slice requirements which sent in SLA, the number of bits that will be scheduled  $M_i$  will be signaled for slice  $i$  where  $i = 1, 2, \dots, n$ . Also, the slice operator can define the achieved resource share

as a portion of total resource blocks ( $R$ ) and the average bit rate per resource block ( $r_i$ ).

The ratio of the required resources for each slice ( $ReqR_i$ ) will be presented as a function in  $M_i$ ,  $R$  and  $r_i$ .

$$ReqR_i = \frac{M_i}{R * r_i} \quad (1)$$

The utility (revenue) function of the service for each slice  $U_i(AR_i)$  is calculated based on the ratio of the slice required resources and the achieved resource share ( $AR_i$ ) in each slice.

$$U_i(AR_i) = ReqR_i \log(AR_i) \quad (2)$$

According to the slice utility function, a weight value ( $T_i$ ) is deduced for each slice as:

$$\begin{aligned} T_i &= \frac{d}{dAR_i} U_i(AR_i) * N \\ &= \frac{ReqR_i}{AR_i} * N \\ &= \frac{M_i}{R * r_i * AR_i} * N \end{aligned} \quad (3)$$

Where  $N$  is the number of users per slice.

And also, for each slice, the average transported bits during each scheduling interval ( $b_i$ ) is computed using:

$$b_i = R * r_i * AR_i \quad (4)$$

Then this average rate  $b_i$  will be approximated by using exponential moving average rate  $b_{i,j}^{\text{exp}}$  for slice  $i$  at scheduling interval  $j$  through using the instantaneous achieved rate  $b_{i,j}^{\text{inst}}$ .

$$b_{i,j}^{\text{exp}} = (1-\beta) b_{i,j-1}^{\text{exp}} + \beta b_{i,j}^{\text{inst}} \quad (5)$$

where  $\beta$  is a small positive constant that ranges from “0 to 1”.

In order to determine the high limits of resources that will be allocated to each Time-Sensitive slice, the history of the allocated resources assigned to each slice will be taken into account. So, the average resource usage ratio  $AvR_{i,j}$  is first determined for slice number  $i$  in the  $j$ -th scheduling interval over a moving time scheduling window ( $\mu$ ) is calculated.  $AvR_{i,j}$  depends on the average resource usage ratio in the previous scheduling interval  $AvR_{i,j-1}$  and the historic resource share at the  $j$ -th scheduling interval  $HistR_{i,j}$ .

$$AvR_{i,j} = \left(1 - \frac{1}{\mu}\right) AvR_{i,j-1} + \frac{1}{\mu} HistR_{i,j} \quad (6)$$

Then, the formulation of the priority factor for each slice to present the proposed high limit of allowed resource share for the Time-Sensitive slices is proposed. The slice that has higher priority will take more resources to accomplish its task within the rigid time requirements. According to the slice weight  $T_i$  and its QoS requirements, a priority factor ( $P_i$ ) for each slice is driven to reveal the importance of each slice as illustrated below:

$$P_i = \frac{\varphi}{\alpha} + T_i \quad (7)$$

where  $\varphi$  is the packet loss probability, and  $\alpha$  is the delay limit in milliseconds (ms).

Finally, we derive the high limits of resources ( $MaxR_{i,j}$ ) for each slice  $i$  in the scheduling interval  $j$  using:

$$MaxR_{i,j} = P_i * MaxR_i + AvR_{i,j} \quad (8)$$

$MaxR_{i,j}$  is calculated based on the slice priority factor  $P_i$ , average resource usage ratio  $AvR_{i,j}$ , and the high limits of assigned resources for  $i$ -slice in the previous scheduling intervals  $MaxR_i$ . The final allocation resources decision is taken based on priority factor  $P_i$  and the high limit of allocated resources  $MaxR_{i,j}$ . If  $P_i$  is greater than  $MaxR_{i,j}$ , then the assigned resources equal  $MaxR_{i,j} * R$ . Else, equals  $P_i * R$ .

The following pseudocode summarizes the algorithm.

**Algorithm:** Resource allocation for **Time-Sensitive** slices

1: **Inputs:**

- The number of bits to be scheduled:  $M_i$
- Packet Loss Probability:  $\varphi$
- Delay limit:  $\alpha$
- High limits of resource share according to the ratio of RBs:  $MaxR_i$
- Average resource usage ratio at the  $j$ -th scheduling interval:  $AvR_{i,j}$

2. **while**  $i \in$  Time-Sensitive slices **do**

3. Calculate the utility function  $U_i$  as in (2)

4. Calculate the weight  $T_i$  as in (3)

5. Derive the average resource usage ratio  $AvR_{i,j}$  as in (6)

6. Calculate the Priority factor  $P_i$  as in (7)

7. Determine the high limit of resource share  $MaxR_{i,j}$  as in (8).

8. **end while**

9. **Assign resources for each slice using:**

10. **if** ( $P_i > MaxR_{i,j}$ ) **then**

11. No. of assigned resources =  $MaxR_{i,j} * \text{number of resources}$

12. **else**

13. No. of assigned resources =  $P_i * \text{number of resources}$

14. **end if**

#### IV. PERFORMANCE EVALUATION

The proposed algorithm is implemented and evaluated using MATLAB environment with the framework introduced in the concurrent slice scheduling approach [14]. Parameters of the system are described in 3GPP TS 36.212. The bandwidth used in simulations are 5MHz (25 RBs) in FDD (Frequency division duplex) mode single antenna.

The methodology defined in the concurrent slice scheduling approach, the basis to our framework, was adopted to evaluate the performance of the proposed algorithm. This approach was compared with the NVS approach [13], thus, our results is validated through making comparison with both the concurrent slice scheduling approach and the NVS approach.

The used performance metrics: cumulative distribution function (CDF) for the transmission delay of packets, the percentage of usable resources per slice, and packet loss.

Furthermore, the effect of varying the channel quality and also the number of frames on the delay was introduced. A set of slices with each one having a varying number of users is presented. Also, the downlink (DL) direction is considered with

varying channel qualities by allowing the change in dynamic modulation and coding scheme (MCS) per user and also per subframe.

#### A. Responses of packet scheduling

In our simulation, three dynamic slices are used. Each has two users as in [14], 4.0 Mbps throughput and 28 MCS. Also, one Time-Sensitive slice is used. It has four users and 20 MCS. A 16% maximum resource shares, scheduling window 25ms, 3000 frames and the traffic is CBR (Constant Bit Rate) that is preferred to be used in timing sensitive traffic, and VBR (Variable Bit Rate) which at most be used in burst data traffic applications. CBR has 10ms arrival rate with difference 2 ms. VBR claimed by a Poisson Process with mean inter-arrival time equal 10ms (100 packets for user per second).

Figure 2.(a) and Figure 2.(b) show the delay response of the proposed framework compared with the Concurrent and NVS approaches. As clearly seen from Figure 2.(a), in VBR, our framework schedules 99.999% of packets before 3.0 ms but Concurrent approach achieves the same percentile before 6.0ms and NVS before 10.0 ms.

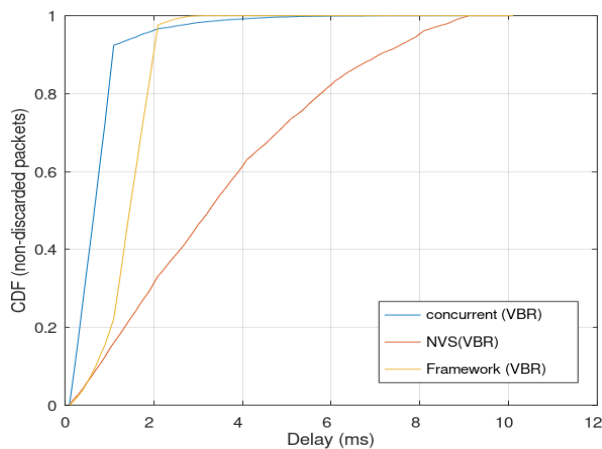


Figure 2. (a) CDF for VBR traffic using 3000 frames and 20 MCS between our framework, Concurrent and NVS

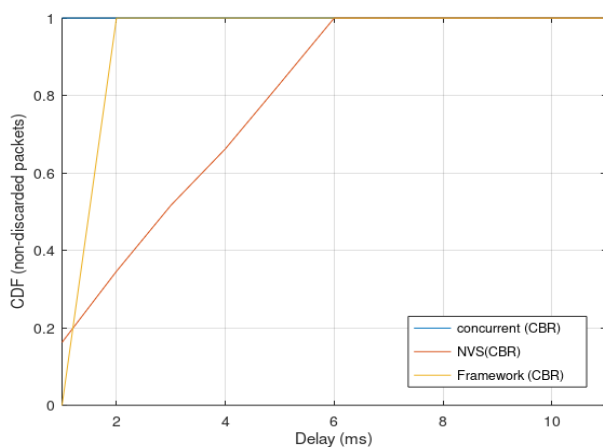


Figure 2. (b) CDF for CBR traffic using 3000 frames and 20 MCS between our framework, Concurrent and NVS.

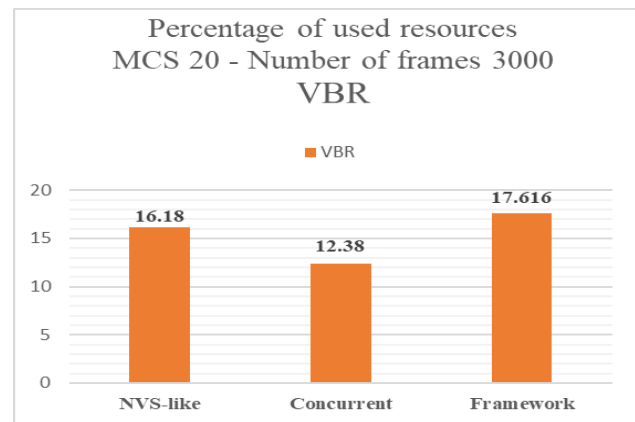


Figure 3. (a) Resource Usage through VBR traffic using MCS 20 and 3000 frames.

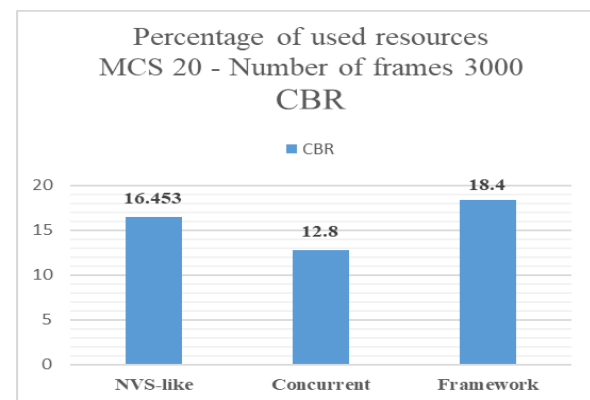


Figure 3. (b) Resource Usage through CBR traffic using MCS 20 and 3000 frames.

CBR is shown in Figure 2.(b), which proves that our proposed framework has 2.0 ms delay and NVS has 6.0 ms delay. Because of more resources were assigned for Time-Sensitive slice, the more packets were scheduled faster than the other approaches.

But Figure 3.(a) and (b) show the percentage of used resources with VBR and CBR, respectively. In VBR, NVS, Concurrent and the proposed framework have 16.18%, 12.38%, and 17.6%, respectively. Which indicate that our algorithm has higher percentage usage of resources. Also, in CBR, it has 18.4% to 12.8% and 16.45% increase than NVS and Concurrent, respectively.

Another delay performance study is introduced through decreasing the channel quality and increasing the number of frames by using 16 MCS and 5000 frames. Our framework proved its efficiency in both CBR and VBR even if the channel quality degraded. In Figure 4.(a), the proposed algorithm transmits 99.999% from packets at 4.0 ms for VBR traffic, whereas the performance of Concurrent approach is close to the NVS and both of them will achieve the same percentile at 9.0 ms. In Figure 4. (b) with CBR traffic, our framework effectively reaches its goal at 2.0 ms to transmit its traffic, but Concurrent approach takes 5.0 ms to do this task and NVS takes 6.0 ms.

Figure 5.(a) and Figure 5.(b) introduce the percentage of used resources in VBR and CBR, respectively. In VBR traffic, our framework has 20.3% and 13.58% and 16.07% for

Concurrent approach and NVS approach, respectively. For CBR traffic, 21.59% for the proposed framework, 16% for Concurrent approach and 16.38 for NVS approach.

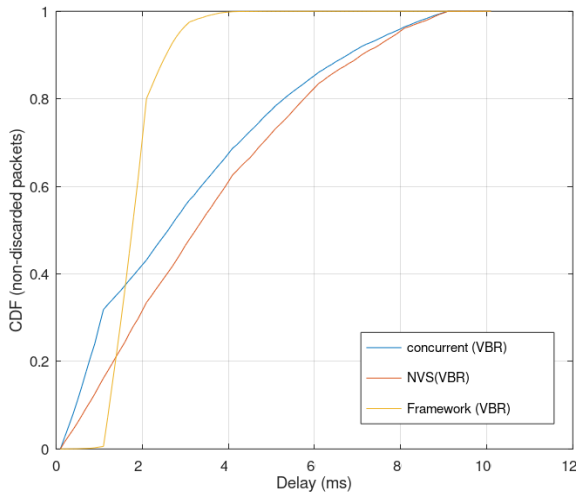


Figure 4. (a) CDF for VBR traffic using 5000 frames and 16 MCS between our framework, Concurrent and NVS.

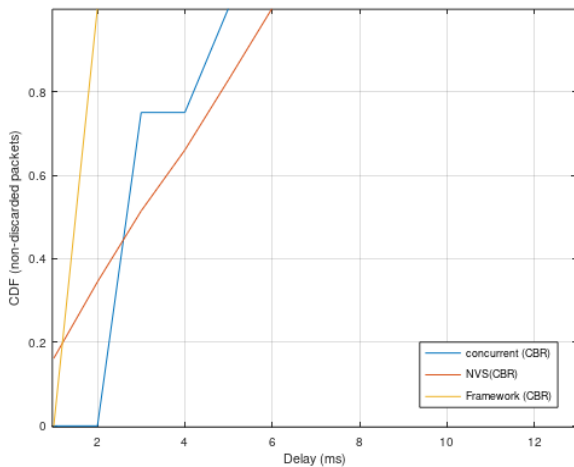


Figure 4. (b) CDF for CBR traffic using 5000 frames and 16 MCS between our framework, Concurrent and NVS.

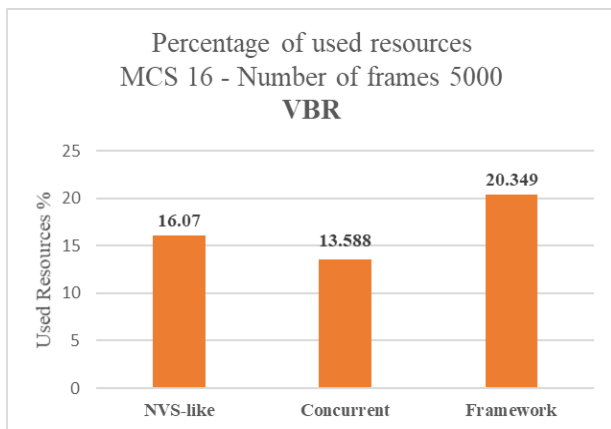


Figure 5. (a) Resource Usage through VBR traffic using MCS 16 and 5000 frames.

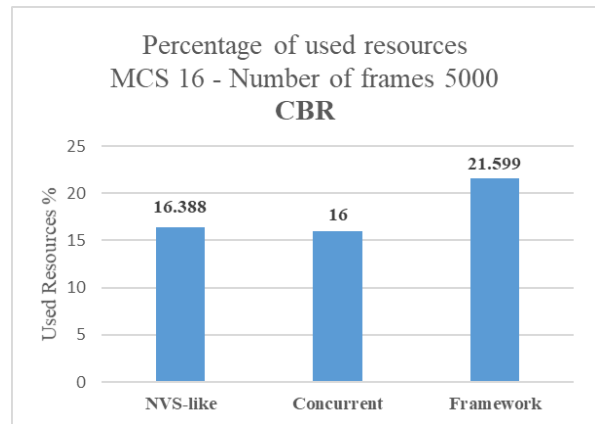


Figure 5. (b) Resource Usage through CBR traffic using MCS 16 and 5000 frames.

*B. The effect of the proposed resource allocation framework on the other slice types*

Our simulation model here consists of two dynamic slices and one Time-sensitive slice. The traffic has 3000 frames with 20 MCS. The proposal comparison with the Concurrent approach. This study is introduced to verify the fairness of our proposed framework in the used resources assigned to each slice. Also, in our framework, more resources are assigned to the third slice than those in Concurrent approach as it has more importance. As clearly seen from figure 6, all types of slices in the framework have an acceptable percentage of used resources.

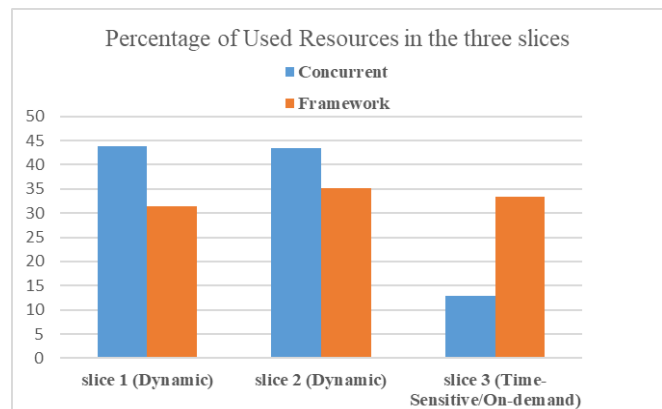


Figure 6. Resource usage in all slices of the network.

*C. Packet Loss*

For any resource allocation approach, there are two important things that should be taken into account: delay and packet loss. In many channel allocation qualities (MCS = 20 and MCS = 16), the performance of our proposed approach with respect to packet loss is introduced. Figure 7. (a) and Figure 7. (b) Show the impact of packet loss with VBR in 20 MCS and 16 MCS, respectively.

The results reveal that NVS has larger packet loss than Concurrent and the proposed framework. It has 1373 packets and both of Concurrent and our framework have zero loss of packets. But in 16 MCS, our proposed approach has a little loss

100 packets, which means 90% decrease in packet loss compared with the Concurrent approach. Concurrent has loss of 1019 packets and NVS has 2328 packets.

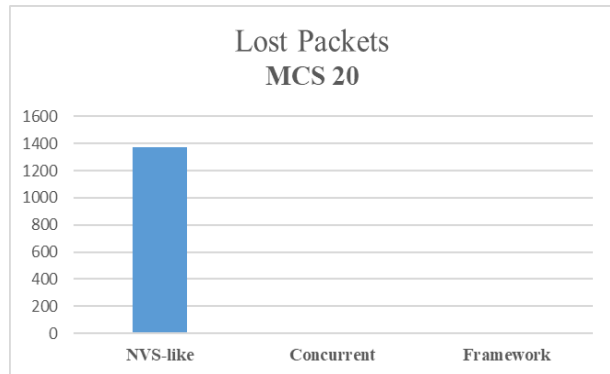


Figure 7. (a) Number of lost packets when using 20 MCS.

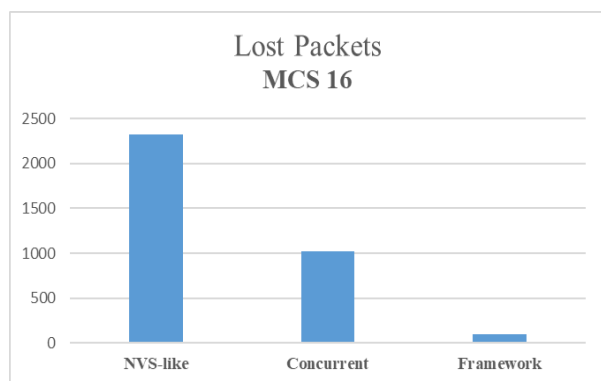


Figure 7. (b) Number of lost packets when using 16 MCS.

#### D. Delay effect with channel quality changes

In order to investigate the influence of the channel quality over the simulation on the Time-sensitive slices, two dynamic slices request throughput 5 Mbps and each serves two users with downlink bit rate 2.5 Mbps and a packet size 1250 B with number of frames 3500 frames are used. The third slice is a Time-Sensitive that serve four users with allowed delay 9 ms.

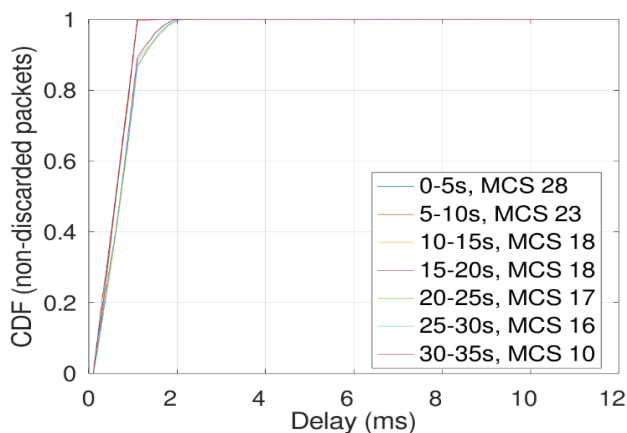


Figure 8. Scheduling Delay when changing MCS using one Time-Sensitive slice and 3500 frames.

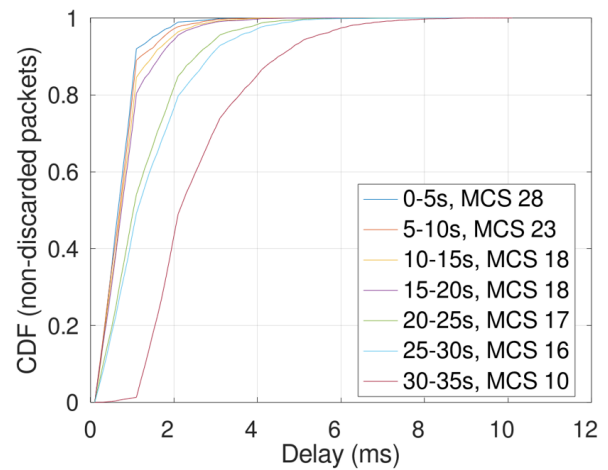


Figure 9. Scheduling Delay when changing MCS using two Time-Sensitive slices and 3500 frames.

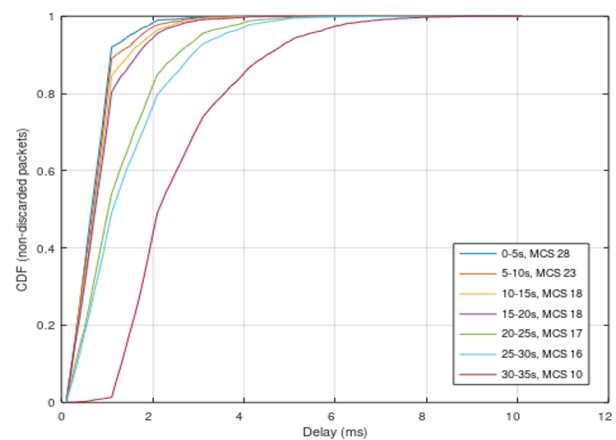


Figure 10. Scheduling Delay when changing MCS using two Time-Sensitive slices and 10000 frames.

The change of MCS applies: starting with 28 MCS, at time= 5s the MCS drops to 23, at time= 10 s the MCS drops to be 18, at time= 20s the MCS drops to be 17, at 25s the MCS will be 16 and finally MCS will drop to be 10 at time 30s. As it can be observed from Figure 8, the delay of the transmissions didn't exceed 2.0 ms even at MCS 10. Time-Sensitive slice seeks to serve users with high reliability requirements as it has sufficient resources to transmit its traffic.

The performance of our proposed approach with the channel quality changes is also introduced with increasing the network load by using two Time-Sensitive slices instead of one and study with 3500 and 10000 frames. As shown in Figure 9 and Figure 10.

## V. CONCLUSIONS

In this research, a resource allocation framework for the type of slices that is concerned with the delay requirements in 5G networks is proposed. Resources are assigned to the Time-Sensitive slices based on their priority factor which reveals the importance of effectively using the available resources. This factor depends on the SLA and the QAS requirements. Time-Sensitive slices can serve URLLC services in 5G cellular



networks. Through our simulations, the results validated that our framework can serve its users with the least allowable latency in traffic transmission and reliability as the resources allocated to the slice were increased so there will not be losses in packets and also outperforms existing approaches, as the Concurrent approach and NVS approach, in this respect.

**Funding:** This research has not been conducted under any fund.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## REFERENCES

- [1] B. Blanco, J. O. Fajardo, et al., "Technology pillars in the architecture of future 5G mobile networks: NFV, MEC and SDN", *Computer Standards and Interfaces archive*, Vol. 54, Issue P4, pp. 216- 228, 2017.
- [2] Y. Xu, G. Gui, et al., "A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges.", *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 668-695, doi: 10.1109/COMST.2021.3059896, Secondquarter 2021.
- [3] C. R. Storck and F. Duarte-Figueiredo, "A Survey of 5G Technology Evolution, Standards, and Infrastructure Associated with Vehicle-to-Everything Communications by Internet of Vehicles.", *IEEE Access*, vol. 8, pp. 117593-117614, 2020.
- [4] S. Li, L. Xu, et al., "5G Internet of Things: A survey", *J. Ind. Inf. Integr.*, vol. 10, pp. 1-9, 2018.
- [5] S. A. AlQahtani, "An efficient resource allocation to improve QoS of 5G slicing networks using general processor sharing-based scheduling algorithm.", *International Journal of Communication Systems*, vol. 33, no. 4, 2019. A. M. Mosa, A. H. Taher, and L. A. Al-jaberi, "Improvement of poor subgrade soils using cement kiln dust," *Case Studies in Construction Materials*, vol. 7, no. April, pp. 138-143, 2017.
- [6] P. Korrai, E. Lagunas, et al., "A RAN resource slicing mechanism for multiplexing of eMBB and URLLC services in OFDMA based 5G wireless networks.", *IEEE Access*, vol. 8, pp. 45674-45688, 2020. J. E. Edeh, A. O. Eberemu, and A. B. Abah, "Reclaimed asphalt pavements-lime stabilization of clay as highway pavement materials," *Journal of Sustainable Development and Environmental Protection*, vol. 2, no. 3, pp. 62-75, 2012.
- [7] S. O. Oladejo and O. E. Falowo, "Latency-aware dynamic resource allocation scheme for multi-tier 5G network: A network slicing-multitenancy scenario.", *IEEE Access*, vol. 8, pp. 74834-74852, 2020. S. S. Rajput and R. K. Yadav, "Effect of fly-ash on geotechnical characteristics of black cotton soil," *International Journal for Innovative Research in Science & Technology*, vol. 2, no. 03, pp. 9-13, 2015.
- [8] M. H. Abidi, H. Alkhalefah, et al., "Optimal 5G network slicing using machine learning and deep learning concepts," *Computer Standards & Interfaces*, vol. 76, 2021. P. Milburn, and L. Parsons, "Performance of soil stabilization agents," *Topeka, Kansas University of Kansas Lawrence, Kansas*, 2004.
- [9] A. A. Barakabitze, A. Ahmad, et al., "5G network slicing using SDN and NFV: A survey of taxonomy architectures and future challenges", *Comput. Netw.*, vol. 167, pp. 1-40, Feb. 2020. L. Bakaiyang, J. Madjadoumbaye, Y. Boussafir, F. Szymkiewicz, M. Duc, "Re-use in road construction of a Karal-type clay-rich soil from North Cameroon after a lime/cement mixed treatment using two different limes," *Case Studies in Construction Materials*, vol. 15, 2021.
- [10] H. Cao, Y. Hu, et al., "Towards intelligent virtual resource allocation in UAVs-assisted 5G networks", (*Elsevier*) *Computer Networks*, vol. 185, pp. 107660, Feb. 2021. I. V. Manikanta and P. S. Chandra, "Sustainable soil stabilization using plastic waste," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 4, no. 2, pp. 2062-2066, 2018.
- [11] H. Beshley, M. Beshley, et al. "QoS-Aware Optimal Radio Resource Allocation Method for Machine-Type Communications in 5G LTE and beyond Cellular Networks.", *Wireless Communications and Mobile Computing* 2021 (2021).
- [12] R. S. Pereira, D. D. Lieira, et al., "RELIABLE: Resource Allocation Mechanism for 5G Network using Mobile Edge Computing.", *Sensors*, vol. 20, no. 19, pp. 5449, 2020.
- [13] R. Kokku, R. Mahindra, et al., "NVS: A substrate for virtualizing wireless resources in cellular networks.", *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1333-1346, Oct. 2012.
- [14] R. Schmidt, C. -Y. Chang, et al., "Slice scheduling with QoS-guarantee towards 5G.", *IEEE Global Communications Conference (Globecom) 2019*, December 2019.
- [15] W. U. Rehman, T. Salam, et al., "Improved resource allocation in 5G MTC networks.", *IEEE Access*, vol. 8, pp. 49187-49197, 2020.
- [16] R. Dong, C. She, et al., "Deep learning for radio resource allocation with diverse quality-of-service requirements in 5g.", *IEEE Trans. Wireless Commun.*, Dec. 2020.
- [17] F. Song, J. Li, et al., "Dynamic virtual resource allocation for 5G and beyond network slicing.", *IEEE Open Journal of Vehicular Technology*, pp. 1-12, 2020.
- [18] S. O. Oladejo, and O. E. Falowo, "Latency-aware dynamic resource allocation scheme for multi-tier 5G network: A network slicing-multitenancy scenario.", *IEEE Access*, vol. 8, pp. 74834-74852, 2020.
- [19] H. Yu, F. Musumeci, et al., "Isolation-aware 5G RAN slice mapping over WDM metro-aggregation networks", *Journal of Lightwave Technology*, vol. 38, 2020.
- [20] T. Ma, Y. Zhang, et al., "Slicing Resource Allocation for eMBB and URLLC in 5G RAN", *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 6290375, 11 pages, 2020.
- [21] T. Costa Cañones, "Evaluation of radio access network Slicing mechanisms in 5G.", BS thesis. *Universitat Politècnica de Catalunya*, 2018.
- [22] N. Alliance, "Description of network slicing concept.", *NGMN 5G P1 Requirements & Architecture, Work Stream End-to-End Architecture, Version 1.0.*, 2016.
- [23] A. Papageorgiou, A. Fernández-Fernández, et al., "SLA Management Procedures in 5G Slicing-based Systems", *EuCNC 2020. IEEE*, pp. 7-11, 2020.
- [24] I. Afolabi, T. Taleb, et al., "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions.", *IEEE Commun. Surveys Tut.*, vol. 20, no. 3, pp. 2429-2453, 2018.
- [25] H. Sánchez, J. Andrea, et al., "Deep Reinforcement Learning for Resource Management on Network Slicing: A Survey." *Sensors* 22.8 (2022): 3031, 2022.
- [26] Nadeem, Lubna, et al. "Quality of Service based resource allocation in D2D enabled 5G-CN with network slicing." *Physical Communication* 52 (2022): 101703, 2022.