Journal of Engg. Research Vol.10 No. (2A) pp. 60-71 DOI: 10.36909/jer.13331

A hybrid resource allocation approach for 5G IOT applications

Mohammad Alkandari*', Jassim Alfadhli* and Lamis Waleed***

** Computer Engineering Department, College of Engineering and Petroleum, Kuwait University.*
*** Corresponding Author: m.kandari@ku.edu.kw*

Submitted ***:*** 21/02/2021
Revised ***:*** 03/10/2021
Accepted ***:*** 12/10/2021

ABSTRACT

5G cellular networks are expected to sustain various Quality of Service (QoS) requirements and provide customers with multiple services based on their requirements. Implementing 5G networks in Internet of Things (IoT) infrastructure can help meet the requirements of IoT devices in a 100x faster and more efficient manner. This objective can be accomplished by applying the network slicing approach, where it partitions a single physical infrastructure into multiple virtual resources that can be distributed among different devices independently. This paper merges the benefits of both the static allocation and the network slicing approach to propose a mechanism that can allocate resources efficiently among multiple customers. The allocation mechanism based on a predefined policy between the slice provider and the customer is to specify the attributes that will be computed before any allocation process. Network slicing is the idiosyncratic latest 5G technology that produces diverse requirements to sustain the traditional network infrastructure's adequate granularity level. The main objective of this paper is to present a simulation suite for a network consists of base stations, including clients whose probable scenarios of 5G can attain high standards of network operation and perform a better and easier analysis of various concepts. Network slicing methodology is enhanced at blocking. Further, it was obvious that the block ratio correspondingly increased the usage of the bandwidth. Based on the results, network slicing methodology enhanced blocking, and the block ratio correspondingly increased the usage of the bandwidth.

Keywords: 5G Cellular network; Internet of things; Network slicing; Resource allocation; Blocking probability.

INTRODUCTION

The full deployment of the IoT, which is anything connected to the Internet and communicates data with other devices, makes the recent researches focus on how to implement an infrastructure that can support the requirements of the IoT traffic data efficiently. The IoT serving broad-based requirements lead to additional considerations such as the ability to provide services to IoT devices installed anywhere. It requires sufficient coverage that can extend anywhere to cover as many IoT devices as possible (Agiwal, Mamta et al., 2019). Besides, both technologies can play significant roles in delivering substantial value to many applications when implemented together. 5G networks ensure QoS requirements that are needed in many scenarios in an IoT environment such as the following: Massive Machine Type Communications (mMTC), which requires the support of having a large number of devices in a small area, send a massive amount of data between each other, such as an Internet of Things. The second QoS requirement is the Ultra-Reliable Low Latency Communications (URLLC) in which these applications built for critical communications do not tolerate any latency such as surgeries and autonomous vehicles. The third one is the Enhanced Mobile Broadband (eMBB) in which these applications require a high data rate within a wide coverage area (Popovski et al., 2018).

Network Slicing allows a single physical network that is sliced into many logical or virtual networks. Network slicing allows many virtual networks to function autonomously across a shared infrastructure (Samdanis et al., 2018). Since the slices are designed to be created dynamically based on the service requirements, the RAN infrastructure needs to be flexible in slicing the network to support various QoS requirements. The slicing approach can help improve the spectrum's efficiency since it can dynamically allocate resources to the slice based on the IoT diverse requirements. By considering the infrastructure proposed in Wang et al. (2018), it divides the 5G I-IoT architecture into three main components: Sensing Regions, Object Processor, and Processing Center. The principal issue in slicing a wireless network can be found in the Object Processor at the base stations and the links that connect the sensors with the base station. A network slice is a connected chain of network functions, building a devoted virtual network that assists the service business model for a particular functions. The variability of wireless link's capacity and the limited resources can cause a problem in providing an effective slicing mechanism that fulfills the IoT device's requirements, which is known as a resource allocation problem in wireless networks. This paper proposes a mechanism based on sharing the virtual resources that are based on a single physical infrastructure to the tenants or requesters, both statically and dynamically. This ensures satisfying their requirements and maximizing the utilization of these resource blocks. The main contribution of this paper is summarized as follows:

- This research presents a simulation suite for a network consisting of base stations plus clients whose potential situations of 5G can fit into analyzing various concepts more apparent.
- A hybrid resource allocation technique is based on merging a static and a dynamic allocation mechanism to benefit from them in the allocation process and utilize the resources as much as possible to enhance overall performance.
- Implementing a predefined allocation agreement between the slice provider and the slice tenant or requester can be based on several matrices such as the RB size and reservation time.
- The performance of the hybrid methodology is validated by comparing it with other existing techniques based only on a static mechanism, or a dynamic approach under different attributes such as blocking probability and latency.
- To conduct an experiment that shows 5G network slicing lets several traffic types to receive various amounts of resources added to the most effective use of the connection for the use case.
- To analyze and assesses the advantages of static allocation, including network slicing in the resource's allocation to different clients effectively.

LITERATURE REVIEW

Background

This literature presents several methodologies for network slicing, whether in 5G IoT architecture on in 5G networks only. In Kapassa et al. (2018), the authors focused on how to satisfy the demands of the IoT devices by dynamically provisioning of 5G slices. An IoT oriented architecture was proposed, which has three different components: Aggregation Mechanism, which is responsible for gathering the IoT application's requirements, checking if any requirement conflict occurs, resolving this conflict, and finally providing an optimal result for the next entity. Feasibility Analysis takes place to check if the incoming request can be achievable or not, based on the number of available resource blocks and the current situation of the network. Dynamic Slicing Manager is responsible for deciding the most appropriate way of allocating the resources to the current request.

In Wu et al. (2018), the authors proposed a resource allocation scheme that analyzes the available slices and increases the overall utilization. This approach uses a shared resource pool among different applications controlled by an Interslice Resource Controller. This controller analyzes multiple slice characteristics among users, such as Networks, Services, and social characteristics, to derive the utility function of resource allocation. Simulation results showed a comparison between the traditional virtual resource allocation schemes and the proposed scheme, which significantly impacted resource utilization.

The authors in Rost, et al. (2017) examined network slicing as an effective solution that discusses 5G mobile networks' various requirements, therefore, presenting certain elasticity and the capacity to be modified linked with expected implementations of the network. The authors explain the difficulties that appear when planning 5G networks, which rely on network slicing. They concentrate on the architectural characteristics linked with the existence of shared and dedicated slices in the network.

Research Methodology

To address the challenges mentioned earlier, this paper introduces a hybrid resource allocation approach that provides the needed services for IoT applications. This approach combines the advantages of both the static and dynamic/network slicing mechanisms. It is based on employing the static approach, which guarantees a fixed number of resource blocks to all tenants. Further, it applies the network slicing technique that dynamically allocates the resource blocks on-demand on the rest of the resource blocks. The dynamic approach concentrated on the 5G slicing mechanism shares the bandwidth resources among different IoT slices in order to maximize its utilization. The network slicing in 5G networks divides the physical network into many logical subnetworks that consist of shared resources such as radio spectrum or Virtual Network Functions (Ordonez-Lucena et al., 2012). This virtualization feature can help in building the needed resources for each IoT services with diverse requirements.

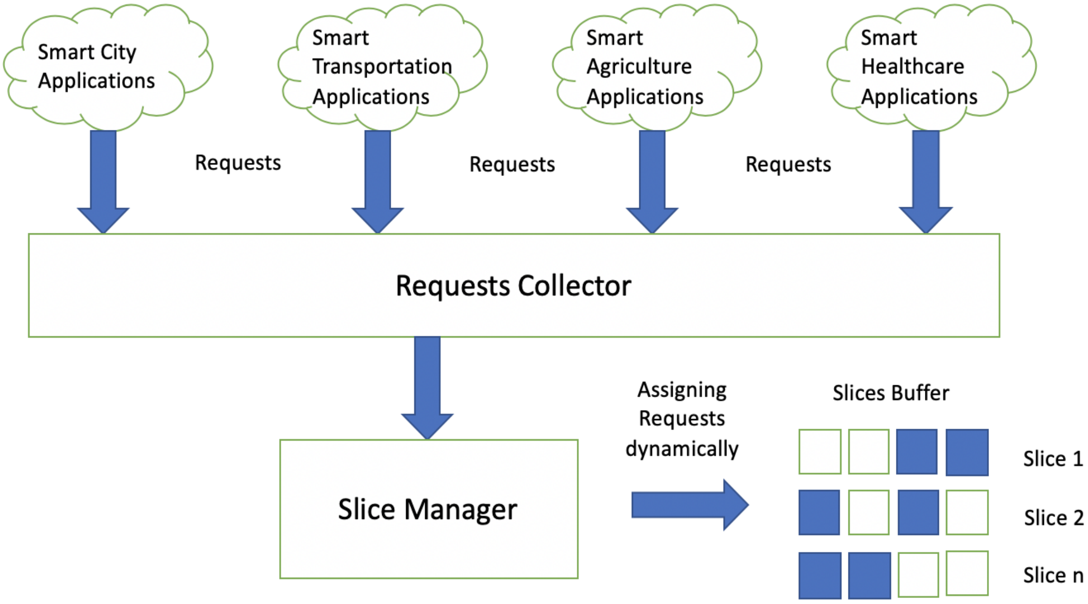

Figure 1. Overall View of the proposed methodology.

As shown in Figure 1, the proposed solution consists of the following:

a) Requests Collector: This entity is responsible for collecting all the requests generated by the different IoT applications that request different resources to fulfill their requirements. It also analyzes the requirements of each application and resolves any conflicts requirements. It then sorts them based on the highest priority request, which depends on many matrices:
b) Size: The requests with the lowest number of resource blocks (RBs).
c) Reservation Time: The time that the application needs to reserve the requested resources.
d) Throughput: The requests that require low bandwidth have a higher priority to guarantee more bandwidth for other requests to be executed.

1. Slice Manager: It analyzes the available resources for each slice provider along with the incoming requirements from the Requests Collector to decide the best approach and assign the needed resources to each request.
2. Resource Allocation process: The allocation process is based on two phases: Static and Dynamic phases. In the static phase, the slice manager assigns a fixed number of resource blocks for all the tenants; for example, 30% of the available resource blocks will be equally distributed among all the tenants. This approach satisfies the requirements of the user and ensures fairness among them as well. In the dynamic or the network slicing phase, the other 70% of the resource blocks will be distributed based on the incoming requests and their priorities. Some IoT applications might need more resource blocks at a specific time, while others might need fewer resource blocks. The requests that are related to the critical application or need a low latency communication to perform, such as uRLLC efficiently, will have a higher priority than other applications. Another approach to order the incoming requests will be based on their blocking probability, which is computed after every allocation process. The tenants with a higher blocking probability will have a higher chance of getting more resource blocks in the next allocation process.

Resource Allocation

The multiple virtual resources that operate on one physical infrastructure are stored in a resource blocks pool (RBP), where different slice tenants (or requesters) ask for these resource blocks from the entities which own these resource blocks (known as slice providers). The resource allocation strategy is based on having requests from multiple requesters, and then, the slice manager will be responsible of serving these requests by communicating with the available slice providers. The user traffic is modelled by a Markov process (Nerlich et al., 2012) where the next user traffic state depends only on the current traffic state and not on the previous state.

In the first resource allocation process, the static resource allocation will take place by allocating 30% of the free resource blocks to all the requesters. For instance, if there are 100 resource blocks, then 30 resource blocks will be distributed equally among all the requesters in order to ensure that all the requesters or tenant have been allocated with resource blocks to serve their users and avoid starvation.

After generating the user's requests and statically allocating RBs to requesters, the requests collector will collect all the requests and order them based on the resource allocation policy that is pre-defined between the requester and the provider. The resource allocation policy can be defined as follows:

1) The requests with the least number of requested resource blocks will be served first.

2) The requests with the higher application priority (e.g., mMTC, uRLLC) will have a higher priority than other applications.
3) The requester that have the highest blocking probability will have the higher priority.

Many other resource allocation policies can be defined to maximize the utilization of the resource blocks. This research defines the resource allocation policy by the requesters with the highest blocking probability as a higher priority, which means that they will be served with more resource blocks in the next resource allocation phase in order to reduce their blocking probability. The ordered requests will be then sent to the slice manager to start the resource allocation process. At the same time, the slice providers will compute the number of available resource blocks in each slice and send this information to the slice manager. The slice manager will then decide which slice provider will be selected in order to serve the first request from the pool.

Network Slicing

The transportation of the network is used by a significant amount of clients demanding various services. The 5G network implements the network slicing concept, which is used to provide different and distinct dedicated logical networks that are customized to various services. Network slicing exemplifies a fundamental characteristic essentially for 5G networks that support providing services with different demands related to time response and reliableness. 5G network slicing for network operators has a distinct advantage, which is mainly deploying the functions needed to sustain selective clients and specific market segments. A secondary advantage is the capability to use 5G systems faster because fewer functions are required to deploy. All the slices in the infrastructure are sat up using diverse latency and bandwidth. This analysis evaluates the benefits of static allocation and network slicing in efficiently allocating resources to the various clients. In this research, the resources will be located in the base station. The base stations and clients can be used to simulate different 5G scenarios to evaluate the different 5G concepts. In addition, the purpose is to ensure the efficiency and performance of the network system while guaranteeing the clients' demand.

EXPERIMENTAL RESULTS

Our approach is to compare our methodology with the existing solutions to discover which mechanism did a better performance in reducing the overall blocking probability of the system. The assumptions that were made while implementing the proposed methodology are the followings:

1. There are three requesters and one slice provider, and each requester has its arrival rate or Lambda. Requester 1 has a variable arrival rate from 2 till 8, while requester two and requester three have a constant arrival rate of 2 and 1. The slice provider has 100 resource blocks that are free to be served to any slice tenant.
2. Priority definition was based on the previously mentioned three applications that were ordered as the following: uRLLC has the highest priority, then mMTC, and finally eMBB, which means that request priority with number 1 means that it needs the resource blocks for an uRLLC application. The reason behind this ordering is that uRLLC application does not tolerate latency, so it needs to be served as fast as possible.
3. In the static allocation process, 30 resource blocks were distributed among the three requesters, where each requester had ten resource blocks to serve their users. Once one of the requester's resource blocks pool is empty, the dynamic allocation process takes place based on its priority and the blocking probability value.

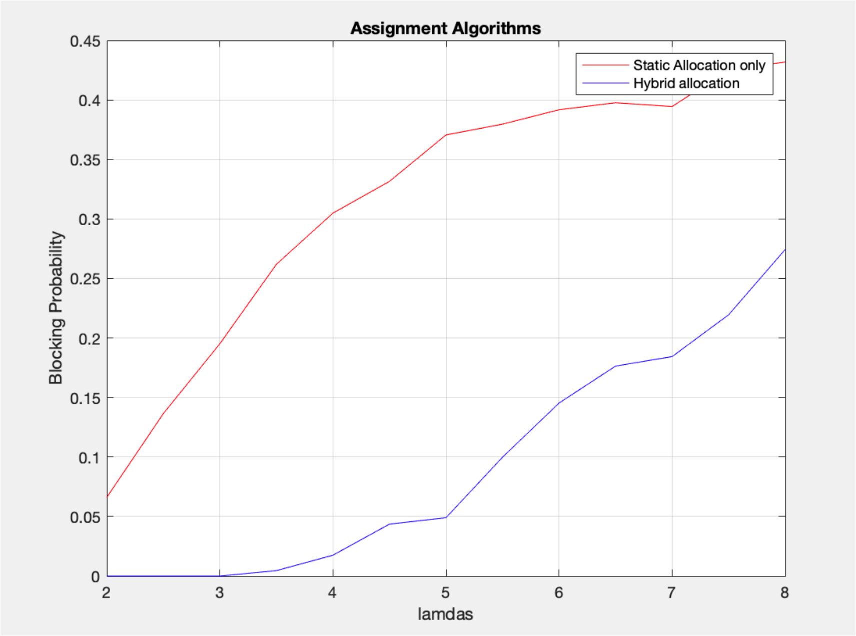

Figure 2. Comparison between Static & Hybrid Solution.

As shown in Figure 2, there is the performance of our hybrid allocation solutions with the static allocation solution. It is clearly shown that the hybrid methodology did a great enhancement in reducing the overall blocking probability among different values of arrival rates, or Lambdas. The static allocation mechanism was based on distributing all the 100 resource blocks among the 3 requesters, and if any requester needs extra resource blocks, it will need to wait until one of the RBs becomes free; otherwise, it will be dropped since all the resource blocks are allocated to other tenants, while in the hybrid allocation mechanism was based on partially allocating some of the resource blocks statically among the tenants to avoid starvation, and dynamically allocating the other RBs to the tenants that are in need of resource blocks more than the others.

Table 1. Blocking Probability Enhancement compared between the two mechanisms.

Lambda	**Static Allocation**	**Hybrid Allocation**	**Enhancement BP %**
2	0.0689	0	6.89 %
4	0.2500	0.0132	23.68 %
5	0.3080	0.0497	25.83 %
6	0.3809	0.1446	23.63 %
7	0.3939	0.1813	21.26 %
8	0.4277	0.2712	15.65 %

Table 1 shows the blocking probability enhancement rate that was shown in Figure 2 between the proposed methodology and the static allocation mechanism. It illustrates that the proposed methodology enhanced the number of dropped requests by an average rate of 19.85 % among different values of lambdas. It is noticed that as we increased the value of lambda, the hybrid allocation did a great performance in terms of the number of dropped requests as compared to the static allocation mechanism. However, this difference between the two mechanisms in terms of the BP is decreased at lambda = 6 and started to decrease until it reached an enhancement of 15.65 only.

The reason behind this result is that, at lambda =8, the number of generated requests is too huge as compared to lambda 5 for example, which has the highest enhancement rate of 25.83%, so this can lead to having more dropped packets than usual due to the limited number of available resource blocks. Furthermore, it shows that it can still serve many requests with a small blocking probability of lambda = 8. The enhancement BP % kept on increasing starting from lambda 2 until it reached to its peak at lambda =5.5. Then, the static allocation performance was somehow constant, while the hybrid mechanism kept on increasing, which led to a decrease in the enhancement BP factor.

In addition, the number of rejected requests for each RB type in 200 iterations is shown in Figure 3. This simulation result was computed in order to validate the correctness of the allocation concept. Since it is most likely expected that the requests with a higher resource blocks will most likely be blocked after several iterations because most of the RBs are served for other tenants and the slice provider does not have enough RBs to serve more requests that needs 4 sequential RBs, it shows that most of the requests with only 1 RB were served, with an average number of 50 dropped requests among the 200 iterations for each lambda value, which is considered to be a very accepted number, while for the requests that need 3 or 4 RBs, it had a much higher dropped requests that range from 200 to 1300.

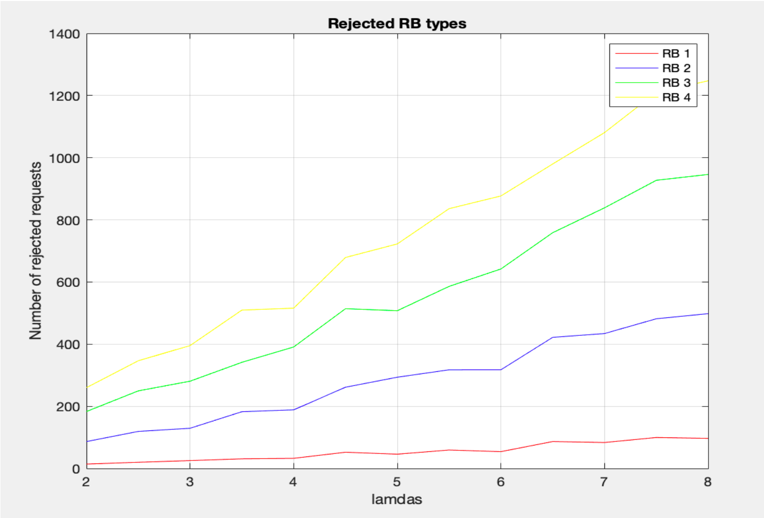

Figure 3. Rejected Requests in Hybrid methodology for each RB type.

SIMULATION

Network slicing was completed in Python using the SliceSim simulation Software (python –m slicesim <inpu.file.yml>). Key definitions: Client: these are the simulation customers that will generate the requests via distribution parameters, and Base Stations (as shown in Figure 4): the slices of the simulation resources.

Pseudocode

```
slices:
 Name of the Slice:
  toleranceofdelay: #
  quality of service: #
  ensuredbandwidth: 0 # in bps
  maxumbandwidth: 100000000 # in bps
  weightofclient:  # [0,1]
  threshold: #
  #  represents  the  bit  routine  pattern  for  client
contributed to this part
   allocation: # allocation name
   parameters: # allocation parameters
    # minimum value
    # maxum value
```

Structure

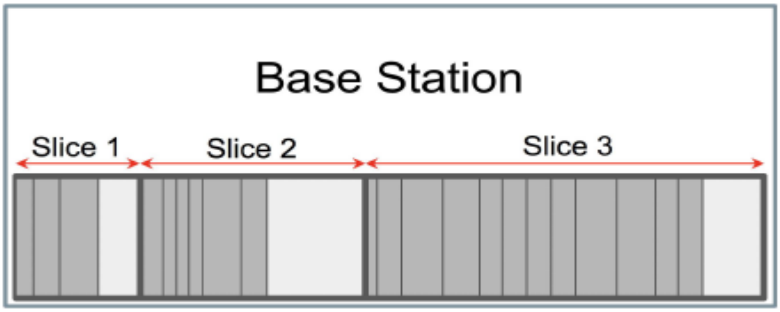

Figure 4. The Base station.

The bases stations in the simulations are tailored for different requirements such as Allocated throughput for a slice, QoS class, Maximum bandwidth for each client, Guaranteed bandwidth for each client, and Delay Tolerance. Tables 2, 3, 4, and 5 demonstrate the input data for the simulation of the base station, the input data for the simulation of the mobility patterns, the input data for the simulation of the client information, and the input data for the simulation of the slices, respectively.

Table 2. The input data for the simulation of the base station.

ID	Throughput	Coverage (x,y), r (meters)	Slices
1	20Gbps	(181,1414), 224	…
5	30Gbps	(126,1016),384	…
20	50 Gbps	(44,1916),368	…

Table 3. The input data for the simulation of the mobility patterns.

Name	**Parameters**	**Distribution**	**Client Weight**
Stationary	Normal Dist.	$\mathrm{N}(\mu = 0, \sigma = 0.1)$	0.2
Slack Person	Random Integer	min = 0, max=1	0.2
Walk	Random Integer	min =0, max=7	0.4
Tram	Random Integer	min=-4, max=4	0.1
Car	Normal Dist.	$\mathrm{N}(\mu = 0, \sigma = 7)0.1$	

Table 4. The input data for the simulation of the client information.

Name	**Parameter**	**Distribution**
X	min=0,max=1980	Random Integer
Y	min=0,max=1980	Random Integer
Usage Frequency	min=0,max=0.1	Random

Table 5 The input data for the simulation of the slices

Name	**Guaranteed Bandwidth**	**Maximum Bandwidth**	**QoS Class**
x_eMBB	-	100 Mbps	5
x_mMTC	1 Mbps	10 Mbps	2
x_URLLC	5 Mbps	10Mbps	1
x_voice	500 kbps	1 Mbps	3
y_eMBB	-	100 Mbps	5
y_eMBB_p	100 Mbps	1 Gbps	4
y_voice	500 Kbps	1 Mbps	3

In this simulation, as shown in Figures 5, 6, and 7, we evaluated the allocation of resources using network slicing concept. The network topology consisted of a single BS, IoT devices, and mobile devices. The results obtained here were related to connected clients, the total bandwidth usage, coverage ratio, block ratio, Handover ration, and client count ration per slice. The base stations increased the number of clients, which in turn caused the rise in the number of coverages. Moreover, the coverage ratio remained relatively constant, while the total bandwidth usage decreased with an increase in the coverage of the BS. The network slicing technique had consistency in the block ratio, which was approximately 0. 1902.

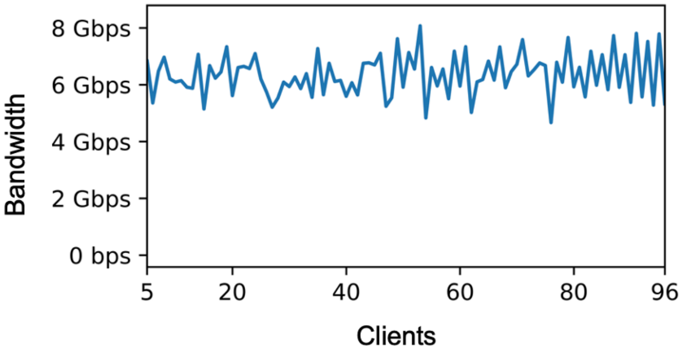

Figure 5. Total Bandwidth Usage.

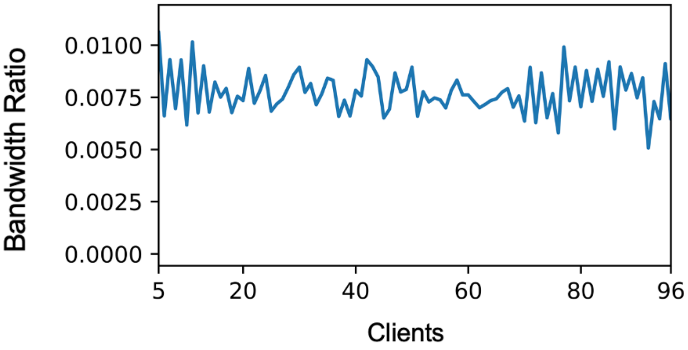

Figure 6. Bandwidth Usage Ratio in Slices.

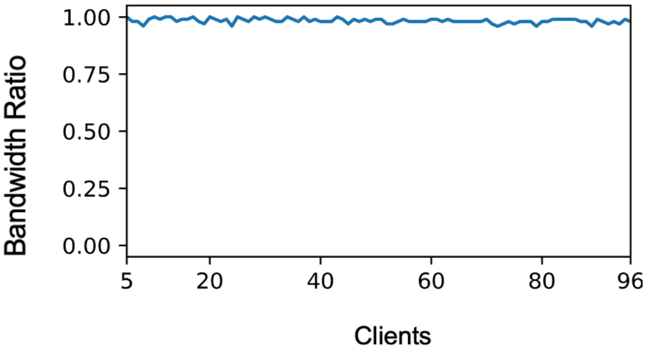

Figure 7. The Average Connected Clients.

Additionally, the handover ration fluctuated considerably. The average connected clients in this simulation were 0.79, while the average load factor of slices was 0.31. Moreover, the BS was the slice owner, and the resources were passed through the slice controller to the IoT and mobile devices. Similarly, the request flowed from the device through the controller to the IoT devices. The devices request for bandwidth and processing resources from the loud provider, that is, the resource allocator. The BS assigned end-to-end paths and provided bandwidth. Under network slicing, it was easy to design the resource allocating algorithm for distributed slice resource allocation that allowed network slices to act as a provisioned and auto-scaled in real-time to ensure efficiency and optimality from the traffic-fair perspective. For simplicity in the simulation, we modeled the network slices to have a single source-destination pair. Moreover, decision variables were used to control the weight of the resources in the slices. The allocation of resources to the network slices had to satisfy the additional constraints, such as the delays' isolation and satisfaction. The initial constraints were in the form of box constraints to ensure that they are handled efficiently in the slices, while the second set of constraints addressed how the weighted hop-count metric could operate with the bias of the paths to the proximal cloud resource. Based on the results, we can conclude that an increase in the number of clients increased the used bandwidth. Furthermore, the network slicing approach is better at blocking, and it was evident that the block ratio also increased the usage of the bandwidth.

CONCLUSION

The proposed methodology satisfied all tenants by allocating RBs to all tenants in the static phase and then allocating RBs to tenants that are in need for more RBs than the others. In this way, it helped utilize the RBS to the maximum since there will be no RBs in an idle state. The simulation results demonstrated that the proposed mechanism had a better result with the static allocation mechanism in terms of the number of dropped requests or the blocking probability. The network slicing methodology is enhanced at blocking. Further, it was obvious that the block ratio correspondingly increased the usage of the bandwidth. The encouraging performance of network slicing based 5G networks has been shown through simulations. Future plans for this work will be adding more statistical measures and enriching this research with the latest literature in the field.

REFERENCES

Afolabi, I., Taleb, T., Samdanis, K., Ksentini, A. and Flinck, H., 2018. Network Slicing and Softwarization: A Survey on Principles, Pnabling Technologies, and Solutions. IEEE Communications Surveys & Tutorials, 20(3), 2429-2453.

Agiwal, M., Navrati, S. and Abhishek, R., 2019. Towards Connected Living: 5G Enabled Internet of Things (IoT). IETE Technical Review, 36(2), 190-202.

Alexander, N., (2018). A Markov Process Approach to the Asymptotic Theory of Abstract Cauchy Problems Driven by Poisson Processes. arXiv preprint arXiv:1801.05726 .

Bursalioglu, O.Y., Li, Z., Wang, C. and Papadopoulos, H., 2018. Efficient C-RAN Random Access for IoT Devices: Learning Links via Recommendation Systems. In 2018 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1-6. IEEE.

Casellas, R., Giorgetti, A., Morro, R., Martinez, R., Vilalta, R. and Munoz, R., 2020. Virtualization of Disaggregated Optical Networks with Open Data Models in Support of Network Slicing. Journal of Optical Communications and Networking, 12(2), A144-A154.

Costanzo, S., Fajjari, I., Aitsaadi, N. and Langar, R., 2018. Dynamic Network Slicing for 5G IoT and eMBB Services: A New Design with Prototype and Implementation Results. In 2018 3rd Cloudification of the Internet of Things (CIoT), pp. 1-7. IEEE.

Gebremariam, A.A., Chowdhury, M., Usman, M., Goldsmith, A. and Granelli, F., 2018. SoftSLICE: Policy-based Dynamic Spectrum Slicing in 5G Cellular Networks. In 2018 IEEE International Conference on Communications (ICC), pp. 1-6. IEEE.

Kapassa, E., Touloupou, M., Stavrianos, P. and Kyriazis, D., 2018. Dynamic 5G Slices for IoT Applications with Diverse Requirements. In 2018 Fifth International Conference on Internet of Things: Systems, Management and Security, pp. 195-199. IEEE.

Kiss, P., Reale, A., Ferrari, C.J. and Istenes Z., 2018. Deployment of IoT Applications on 5G edge. In 2018 IEEE International Conference on Future IoT Technologies (Future IoT), pp. 1-9. IEEE.

Ksentini, A. and Nikaein, N., 2017. Toward Enforcing Network Slicing on RAN: Flexibility and Resources Abstraction. IEEE Communications Magazine, 55(6), 102-108.

Ordonez-Lucena, J., Ameigeiras, P., Lopez, D., Ramos-Munoz, J.J, Lorca, J. and Folgueira, J., 2017. Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges. IEEE Communications Magazine, 55(5), 80-87.

Popovski, P., Trillingsgaard, K.F., Simeone, O. and Durisi, G., 2018. 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-theoretic View. IEEE Access 6 : 55765-55779.

Rittinghouse, J. and Ransome, J., 2016. Cloud computing: implementation, management, and security. CRC press.

Rost, P., Mannweiler, C., Michalopoulos, D.S., Sartori, C., Sciancalepore, V., Sastry, N., Holland, O., Tayade, S., Han, B., Bega, D., Aziz, D. and Bakker, H., 2017. Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks. IEEE Communications magazine, 55(5), 72-79.

Samdanis, K., Costa-Perez, X. and Sciancalepore, V., 2016. From Network Sharing to Multi-tenancy: The 5G Network Slice Broker. IEEE Communications Magazine, 54(7), 32-39.

Wang, D., Chen, D., Song, B., Guizani, N., Yu, X. and Du, X., 2018. From IoT to 5G I-IoT: The Next Generation IoT-based Intelligent Algorithms and 5G Technologies. IEEE Communications Magazine, 56(10), 114-120.

Wu, D., Zhang, Z., Wu, S., Yang, J. and Wang, R., 2018. Biologically Inspired Resource Allocation for Network Slices in 5G-enabled Internet of Things. IEEE Internet of Things Journal, 6(6), 9266-9279.