Slice Allocation Management Model in 5G Networks for IoT Services with Reliable Low Latency

Mohammed Dighriri[1], Abayomi Otebolaku[2], Ali Saeed Dayem Alfoudi[3], Gyu Myoung Lee[4]

1.Department of MIS, University of Business and Technology, Jeddah, Saudi Arabia

2. Department of Computing, Sheffield Hallam University, Sheffield, UK

3. Department of Computer Science, College of Computer Science and Information Technology, University of Al-Qadisiyah, Al Diwaniyah, Iraq

4. Department of Computer Science, Liverpool John Moores University, Liverpool, UK

Abstract

Network slicing is a promising technology for 5G networks in which operators can sell customized services to different tenants at various prices and Quality of Services (QoS) demands. Thus, the latest 4th Generation (4G) and upcoming 5th Generation (5G) mobile technologies are expected to offer massive connectivity and management of high volume of data traffic in the presence of immense interferences from mobile networks of IoT devices. Further, it will face challenges of congestion and overload of data traffic due to humongous number of IoT devices. Nevertheless, these devices are likely to demand high throughput, low latency and high level of reliability especially for critical real-time applications such as in Vehicular Communication System (VCS). To address these issues in 5G mobile networks, this paper proposes a Slice Allocation Management (SAM) Model based on critical services of smart systems such as VCS to satisfy QoS demands. The proposed model aims at providing dedicated slices on the basis of service requirements such as expected throughput and latency for VCS. To ensure such performance and provide data traffic priorities of IoT devices in uplink of Relay Nodes (RNs) cells in the 5G mobile networks, we have sliced the Radio Access Networks (RAN), along with assignment of the nearest Mobile Edge Computing (MEC) with isolated slices based on the priorities for each IoT nodes to reduce latency level. The proposed model was simulated and validated using the OPNET simulator. The results obtained demonstrate that SAM Model is able to achieve improvement of end to end delays and uplink throughputs of the networks in high-density networks of IoT devices.

Keywords 5G, IoT, SAM Model; Vehicular Communication System.

1 Introduction

In the next few years, the fast growth of Internet of Things (IoT) with billions of devices, and the 5th Generation (5G) mobile networks will be required to offer massive connectivity for IoT devices and meeting the demand of Quality of Service (QoS) such as low latency. These QoS requirements are extremely important for Vehicular Communication System (VCS), where the communication system of connected vehicles: Vehicle-to-Vehicle (V2V), Vehicle-to-Roadside (V2Rs), Vehicle-to-Infrastructure (V2I), and Vehicle-to-everything (V2X) are the basis of intelligent and connected transportation systems where all vehicles and infrastructure systems are interconnected with one another. [1][2]. In addition, this massive number of IoT devices and services are connected through various Heterogeneous Networks (HetNets) such as RNs and Donor eNB (DeNB).

Although many of these devices will only be sending and receiving relatively small amount of data, they will make new demands, in terms of managing the total accumulation of data and number of physical connections. However, current 5G networks lack the ability to connect a higher number of users and establish and maintain a healthy transmission at the same time due to inherited control plane limitations and scheduling limitations respectively [3][4]. Therefore, new scheduling and access control mechanisms are required to reduce the amount of control plane signaling for IoT users. Nonetheless, emergency service providers like Smart Healthcare System (SHS) need a real-time data availability and a higher level of reliability in order to deal with critical situations more effectively. Police, fire, and ambulance services must have highly reliable voice links without having issues like call dropping and unresponsive networks. Today, some of these are provided

using dedicated networks, but they have limited data capacity (narrow bandwidth) and require high investment just to provide reasonable coverage [5][6]. Moreover, these systems do not guarantee to fulfill futuristic service requirements in terms of high data rates and real-time interactions. Therefore, new network technologies and innovations are required for "ultra-reliable" scenarios, where the ability to connect and operate in severely degraded or complete lack of infrastructure must be assured. This is based on using a device to device direct communication, ad-hoc backhaul and networking, and flexible re-configuration of networks[3][7].

Besides, mobile networks are estimated to face challenges as a result of extremely demanding QoS requirements of futuristic IoT services such as a provision of radio resources to a massive number of IoT devices, prioritization, and inter-device communication [8]. On the one hand, the existing mobile systems such as 4G and upcoming 5G might run out of capacity due to the increasing IoT traffic, resulting in the performance degradation of regular mobile data traffic [9]. On the other hand, IoT devices demand various type of QoS levels to facilitate the seamless delivery of different services. For example, SHS devices convey big sized data that are sensitive to delay [10]. The Radio Access Networks (RAN) is the smallest radio resource unit called Physical Resource Block (PRB), which is allocated to a device for data transferred in 5G mobile networks. In smart systems, there are different devices transmitting numerous sizes of data, where some transmit small size of data traffic. Therefore, the capacity of the PRB is not fully used and without radio allocation in the shaping of slicing, resulting in significant poor performing of the smart systems.

However, to address these problems, there are several existing slice allocation models, which have been reported in literature. For example, the models in [1] and [3] are based on Network Virtualization (NV), whereas [8] is based on a central broker in which slicing allocation is based on Network Virtualization (NV), [8] is based on a central broker. In [11], the proposed is based on a radio resources scheduling model, that relies on the demanded Service Level Agreement (SLA). Moreover, 5G mobile networks have emerged as an attractive research topic in terms of supporting IoT devices. In the 5G network slicing communication, existing works have played important roles in interconnecting physical systems with the internet. For example, in [12] and [13] the authors have developed solutions that model data traffic of IoT devices as a simple Poisson process where numerous machines are assigned to one server. In [14], the authors have discussed the difficulty of capturing diverse statistical patterns of application flow of IoT devices. Also, in [15] and [16], the authors have proposed architectures for RAN source and aggregated data traffics that provide IoT devices with data traffic integrity, and energy efficiency.

To address these gaps, the contribution of this paper is a Slice Allocation Management (SAM) model for efficient utilization PRBs via HetNets cells such as RNs and DeNB, and the customization of dedicated slices to specific IoT devices and smart systems. Therefore, 5G radio resources will be efficiently exploited to manage the data of different IoT devices for each slice individually, it also exploits the Mobile Edge Computing (MEC) for each slice to reduce latency and deliver improved QoS. To the best of our knowledge, we are the first to propose a novel SAM Model over 5G mobile networks based on wireless layer 3 in band RN which is used for improving coverage and managing uplink data traffic of IoT devices such as VCS by considering a separate slice. The SAM Model can be executed to improve communication reliability and network utilization, especially for smart systems with critical services. Moreover, we have used OPNET simulator to assess the performance of the proposed Model in case of QoS for each data traffic slices separation. The simulated 5G mobile network of IoT devices and applications include Simple Mail Transfer Protocol (SMTP), File Transfer Protocol (FTP), and Voice over IP (VoIP) and Real-time Transport Protocol (RTP). Slices of these scenarios are classified depending on the popularity, sensitivity, and volume of the IoT data traffic. The results show the impact of the proposed model in terms of assuring different QoS characteristics for the different types of data traffic of 5G networks. The end-to-end (E2E) performance of the network is tested by managing data of different IoT devices in each slice in cases of average cell throughput, SMTP and FTP average upload response time, and FTP average packet E2E delay. Simulation results of the proposed SAM Model scenarios in comparison with other scenarios show a significant improvement in IoT nodes packets transmission via RNs and DeNB cells. in the Results also show improved End-to-End (E2E) delay, reaching 1ms, in the FTP nodes when loading in VoIP nodes by 80% and throughput of all nodes in uplink side of the network by 66%. The rest of this paper is organized as follows.

Section 2 presents related works on slicing allocation models and IoT devices supporting models for HetNets technology. In section 3, the SAM model is presented in detail, followed by the detailed elaboration of the vehicular communication use case scenario and service slicing algorithms in section 4. In section 5, we present the SAM Edge cloud model. Then we in section 6, we present OPNET simulation environment of the solution including simulation parameters. In section 7,

we present the simulation scenarios. Section 8 presents the simulation results, analysis, and discussion. Section 9 concludes the paper and outlines our future works.

2 Related Work

5G mobile networks remain a hot research topic in terms of supporting IoT devices. 5G network slicing communication has played a significant role in interconnecting the physical systems with the internet world. IoT devices generate bursty traffic compared to traditional mobile devices. In [12] and [13] the authors have developed models for IoT device data traffic as a simple Poisson process where numerous machines are assigned to one server. In [14], the authors state that the data streams in various IoT applications follow different statistical patterns, which are difficult to capture. According to [15] and [16], IoT data traffic is classified into the source and aggregated traffic. In [17], on the one hand, the authors proposed a novel M2M system architecture that integrates smart road with unmanned vehicle and includes many challenges, such as data integrity, and privacy for IoT devices are aggregated. On the other hand, in [17], the authors proposed a data aggregation architecture with an energy efficient technique in Wireless Sensor Networks (WSNs). But due to high node density in sensor networks, same data is sensed by IoT devices, resulting in redundancy. This redundancy can be eliminated by using data aggregation approach while routing packets from source nodes to base the stations. To address such limitations, authors in [12] proposed a model that understands the behaviour of the IoT devices data traffic more accurately, where source traffic model is required (i.e., addressing each IoT device on its own with their respected energy consumption). However, in the IoT device source of data traffic model, there are many challenges that need to be addressed. For example, it is very hard to model the behaviours of the traffics being produced by massive numbers of IoT devices in parallel and in the existence of strong spatial and temporal correlation among the devices.

In addition, an efficient wireless network virtualization for Long Term Evolution (LTE) systems has been proposed in [12] and [18], with a slicing structure model to efficiently allocate physical resource blocks to Diverse Service Providers (SPs) to maximize the utilization of resources. The mechanism is dynamic and flexible for addressing arbitrary fairness requirements of different SPs. The authors in [19] and [20] have proposed models for wireless resource virtualization in the LTE systems to allow slicing of radio resources among mobile network operators. An iterative algorithm was proposed to solve the Binary Integer Programming (BIP) with less computational overhead. However, the above-considered models do not the prioritize different slices. Besides, they do not consider the priority among the users of the same slice. Moreover, in [19] and [21], authors have described a model for slicing downlink network resources. The model accepts a service only if the provisioning of this new service does not affect the throughput of the services in the cell. Consequently, this work does not take into consideration the dynamic modification of the Quality of Experience (QoE) of mobile users in order to increase network capacity and resource utilization. Authors in [22] proposed a combined resource provisioning and resource allocation model targeting to maximize the total rate of virtualized networks based on their channel state information. An iterative slice-provisioning algorithm was proposed to adjust minimum slice requirements based on channel state information but without considering global resource utilization of the network and inter- and intra-slice priority.

The Network Virtualization Substrate (NVS) was presented by the authors in [23]and [3], allowing the infrastructure providers to manage the resource allocation on the way to each virtual instance of a DeNB before each virtual operator customizes scheduling within the allocated resources. In [24] and [25], the authors presented the related technologies for network slicing with a specific focus on synchronous purposes, such as multi-dimensional resource management, dynamic traffic steering, and resource abstraction. An architecture for network slicing has also been presented and elaborated in the 5G NORMA project [24]. Additional network slicing solution based on a gateway-based method is presented in [18] and [26]. The authors used a controller to offer application-oriented resource concept of the fundamental RAN. An ability broker for resource slices was presented first by the Third Generation Partnership Project (3GPP) and widely assessed in [27] by allowing on-demand slice resource allocation. The infrastructure provider instantiates a network slice by allocating precise resources to a Mobile Virtual Network Operator (MVNO), service providers, and vertical parts for a stated time duration. Another valuable proposal that discovers the diverse options of network allocation and relies on a central broker is provided in [28]. The presented solution provides the means of mobility for transmitting users to other networks, spectrum transfer policies, and the application of resource virtualization. In [11], authors present a dynamic and flexible slicing model that schedules radio resources relying on the demanded SLA, whilst maximizing the user rate and applying fairness criteria.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

In our work, we first present a basic network slice concept in terms of slice customization to specific users, with RAN parts completely reserved to certain services such as an "isolated slice." However, with the emergence of progressive network virtualization techniques, the perception of network slicing in 5G has developed to more flexible allocation, targeting to attain an important multiplexing gain while still assuring isolation and separation. In addition, we have presented other studies in 5G slicing technology to support IoT devices.

3 System Model

In this section, we describe the proposed SAM model, which relies on the SAM functional architecture, which is presented in section 3.2, to support specific network slicing, concentrating on categorization and dedication of QoS demands of IoT devices such as Smartphone, VCSs, and SHS, but first we elaborate more on the challenges being addressed.

3.1 Heterogeneous IoT Service Challenges

The main challenge of the IoT devices in near future with high densification and heterogeneity of wireless networks, particularly, when these IoT devices increase to reach more than twenty billion in 2020, which causes overload and congestion of incoming data traffic. In the world of cellular technology the fast growth of wireless network technologies (e.g. 5G) and ever-increasing demand for services with high QoS request, the management of network resources becomes a permanently more challenging problem that requires to be correctly designed in order to advance network performance. Nevertheless, in this situation, network slicing is getting an always-increasing importance as an effective approach to introduce flexibility in the management of network resources. A slice is a gathering of network resources, selected in order to satisfy the demands (e.g., in terms of QoS) of the service(s) to be delivered by the slice. The aim of slicing is to introduce flexibility and higher utilization of network resources by offering only the network resources necessary to fulfil the requirements of the slices enabled in the smart systems. As our approach here is designed to exploit and manage the RAN capacity of RNs and DeNB cells in slicing form, which allows to customize and reduce the PRBs wastage in each slice in terms of technical requirements such as mobility management and priorities, also QoS requirements such as latency, throughput and loading. In addition, we have designed the MEC Model to reduce the latency of IoT devices based on finding the closed edge cloud to these IoT devices, which also set up the appropriate priority for each IoT nodes. Therefore, we have used smart city use case as the smart system, which forms the network slices by differentiating the data traffic smartly in terms of QoS requirements of each slice such as in VCS, Smartphone, and smart healthcare system. In the next section we present the functional architecture of our proposed solution. Smartphone, and smart healthcare system. In the next section we present the functional architecture of our proposed solution.

3.1 SAM Functional Architecture

From a functional perspective, the 5G-RAN consists of two types of network functions (NFs), each distributing the full radio access functionality to interact with the UE over the radio interface: gNBs, using the 5G New Radio (NR) interface; and ng-eNBs, and an evolution of the LTE interface. Focusing on 5G NR access, gNBs are linked to the 5G Core network (5GC) by means of NG interfaces and may be interconnected with other gNBs and ng-eNBs over Xn interfaces. To present modularity and support different deployment options, 3GPP has also standardised the F1 interface that functionally splits a gNB into a gNB Central Unit (gNB-CU) for upper protocol layer processing and a gNB Distributed Unit (gNB-DU) for lower protocol layer processing[29]. A single gNB, regardless of whether it is divided into gNB-CU/ gNB-DU or not, handles the operation of one or more 5G (RN) cells. Each 5G cell, individually identified by a *cell ID*, is allocated with specific radio resources such as RF carriers which are operated under a common set of control channels (e.g. synchronisation, broadcast). The 5G (RN) cell interface is being designed with high flexible OFDM based waveforms with different numerologies (e.g. different subcarrier spacing and cyclic prefix lengths and adaptable time frequency frame structures such as selectable slot durations and dynamic assignment of downlink/uplink transmission direction) [29]. Furthermore, the 5G (RN) cell interface allows for UEs served via the same 5G (RN) cell to be instructed to receive or transmit using only a subset of the cell resource grid. Eventually, this flexibility of the 5G cell interface allows UEs with diverse access types such as enhanced Mobile Broadband [eMBB], massive Machine Type Communications [mMTC], and Ultra Reliable Low Latency Communications [URLLC] to be concurrently multiplexed over the same 5G cell, as shown in Figure 1. From the service perspectives, the overall 5G network RAN and 5GC are considered to support a PDU Connectivity Service, such as a service that provides exchange of Protocol Data Units (PDUs) such as IPv4, IPv6, Ethernet or Unstructured data packets between a UE and an external data network reachable from the 5GC. The PDU Connectivity Service is realized via the establishment of one or multiple PDU sessions, which are the logical associations created

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

between the 5GC and the UE to handle the data packet exchanges [30]. On this basis, the understanding of network slicing relies on the principle that each PDU session is associated with a specific *Network Slice*.

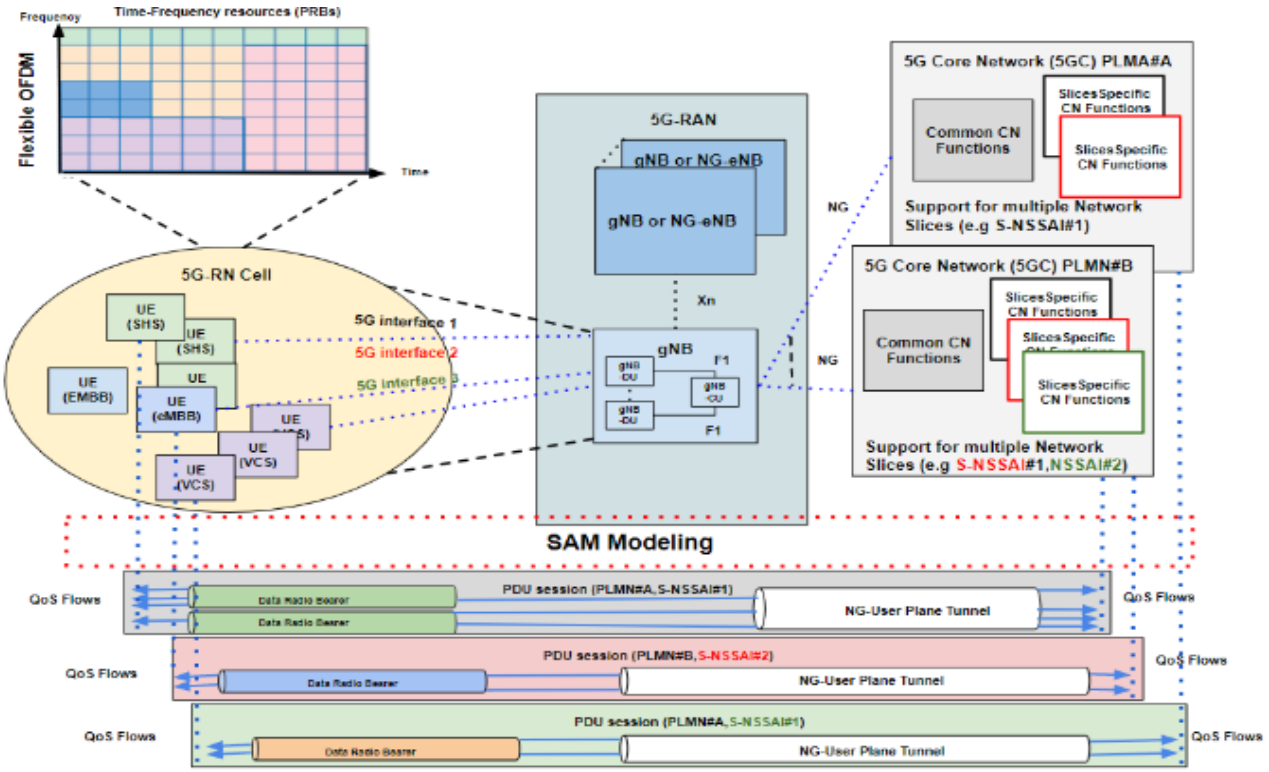


Fig.1. SAM Functional Architecture

In fact, a network slice, which is defined as a logical network that delivers specific network abilities and network characteristics allows, providing a differentiated network behaviour to UEs that are attached to the same 5G network such as to the same 5GC ,uniquely identified by a Public Land Mobile Network [PLMN] identity ,but have PDU sessions connected with different delivered *Network Slices* [31] .Moreover ,to differentiate traffic treatment ,*Network Slices* can also be used to serve diverse customers separately as per an agreed Service Level Agreement (SLA .(Therefore ,a *Network Slice* is officially identified in 3GPP specifications [30] by a Single Network Slice Selection Assistance Information (S-NSSAI) identifier ,which is unique within a PLMN and is comprised of a Slice/Service type (SST .(The SST denotes the expected network behaviour ,and a Slice Differentiator (SD ,(differentiating amongst multiple network slices of the same Slice/Service type .Therefore ,each PDU session activated between a UE and a 5GC/PLMN network is associated with one and only one S-NSSAI so that the corresponding traffic flows, denoted as QoS flows in the 5G network are handled according to whatever behaviour is pre-established for the allocated S-NSSAI .The allocation of the serving S-NSSAI is decided between the UE and 5GC based on such as subscription rights and communicated to the 5G-RAN through signalling .Within the 5G-RAN ,the pre-established behaviour associated with the S-NSSAI can then be enforced by the proper handling of Data Radio Bearers (DRBs ,(which are the delivery services provided by the 5G-RAN over the radio interface such as specific scheduling rules and/or radio protocol stack configuration for the corresponding DRBs . Furthermore ,to support multiple S-NSSAIs of a particular 5GC/PLMN ,the 5G-RAN could also serve multiple 5GC/PLMN networks by leveraging the sort of RAN sharing solutions introduced for legacy technologies such as 3GPP Multi-Operator Core Network (MOCN .(Hence ,gNBs could be linked to several 5GCs and the shared 5G cells could broadcast information about the reachable 5GC/PLMN networks as well as support flexible access control mechanisms per PLMN/SNSSA such as 5G Unified Access Control mechanisms

3.2 Slicing Allocation Management Model

We propose a novel SAM model that operates and manages the slice allocation by reducing the loss of multiplexing. If the number of slices is low, the wastage for each slice is also low. In the proposed SAM model, virtualized links and nodes are managed by the Service-Based Slice Allocator (SBSA). In addition, an Operation Support System (OSS) and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

Business Support System (BSS) are expanded to implement the SAM model as showed in Figure 2.

The SAM model starts when the service operator needs to execute a service to the user demand (e.g., smart healthcare system, smartphone and VCS, (1) the service operator sends an inquiry to SAM requesting service release admittance accompanied by a list of the service technical required (e.g., priority management, access area range, etc.) and their QoS requirements (e.g., lower latency limit, high bandwidth limit, service specification protocol). (2) SAM calculates the required resources and chooses one of three options: (A) allocate current slice, (B) allocate current slice after expansion, and (C) create service-dedicated slice on basis of service requirements and current slice's utilization [31].

If option (B) or (C) is selected, (3) SAM instructs the Network Functions Virtualization Orchestrator (NFVO) to expand the current slice or create a new slice. Then, (4) SAM sends the SBSA the service-related information and the forwarding destination of the service. Then, (5) SAM notifies the service operator of the results (which option was selected) and the access point. (6) The service operator embeds the access point into the background application of the service or its IoT devices (7) so that when someone uses the services, the service information is provided to the user. Finally, (8) the service user accesses the SBSA with the information, and (9) the SBSA transmissions the service traffic to the assigned slice along with various of technical and QoS requirements based on equations (1) and (2) below, which consist of the priority or latency of the packets, the highest priority and lower latency need is conveyed on the output port first and then the packets with lower priority with high latency and so on as illustrated in service slices algorithm as showed Figure 3. Therefore, we design our smart systems environment in three technical and QoS requirements high *(slice1)*, medium *(slice2)*, and low *(slice3)*, rely on the data traffic types as follow:

- Vehicular Communication Systems (VCS)as sensitive data traffics
- Smart Healthcare System (SHS) as heavy data traffics
- Smartphone as popular data traffics

These data traffic will work in slicing over the 5G mobile network in the uplink path between RNs and DeNB based on user plane interface. This Model and slice operating method reduce the number of slices to the minimum, thus improving multiplexing gain by accommodating more service traffic in a slice. Compared to current monolithic EPC architecture, since service traffic's time-varying resource demand patterns with bursts of high demand periods and low- utilization services are complementary, the more multiplexing service traffic in a slice, the less total capacity required to satisfy the demand of all the services.

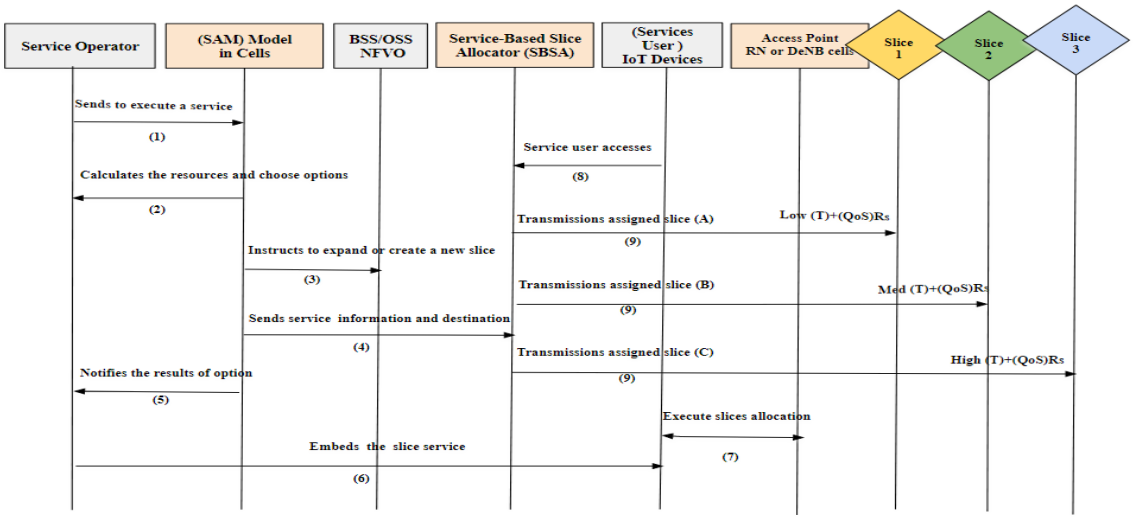


Fig.2. Slicing Allocation Management Model

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

4 The vehicular communication system (VCS) and SAM Solution

4.1 4.1 The vehicular communication use case

In this section, we present VCS as the main use case of the proposed solution. The VCS enables the exchange of information between vehicles infrastructure and VCS applications through communication methods and technologies. In VCS, vehicles communicate with other vehicles (V2V) or communicate with the VCS server's infrastructure (V2I). The VCS applications include collision avoidance and safety, parking time, the Internet connectivity, transportation time, fuel consumption etc. [4] . Several research efforts were made to investigate the support of IoT communication in VCS [4]. In order to explain the use of IoT in VCS, few VCS applications such as collision avoidance and onboard security are the most significant applications of the VCS. When driving a vehicle which is not networked with the VCS server, the decisions are made depending upon the information within the Line-of-Sight (LoS) of the vehicle. According to [4], the main purpose of using wireless IoT communication is to deal with such LoS limitations in order to avoid an accident. In the case of emergency, the information from devices positioned to monitor emergency situations are transmitted to other networked vehicles within the communication range. To avoid any further accidents, the communication between the server and vehicles must be very fast for the detection of emergency messages and delivering of warning messages. Since the response time against the warning messages is very small, so collision avoidance services demand high QoS services and low latency. According to [4], the warning messages are small and should only be sent in critical situations for efficient utilization of the communication network bandwidth. Traffic and infrastructure management plays a significant role in controlling the problem of road congestion. All over the world, every day the drivers face the problem of road congestion that not only increase fuel consumption which lead to more emission and causes an increase in pollution but also causes high tensions for the drivers[4]. A better-managed infrastructure improves productivity and reduces the factors of costs and pollution in society. VCS tackle the problem by providing a bidirectional IoT communication. Such applications do not demand high data rates as few parameters like time, speed and vehicle identification are required. As a result of the low latency which will be delivered by the fundamental access network, 5G network emergency services depend on massive IoT and device-to-device connections will be categorised by higher throughput, QoE, higher QoS and low buffer demands for the IoT devices [4]. Based on the VCS use case, in the next section we present the service slicing algorithm and the MEC data traffic placement algorithms.

Table 1: Use case performance metrics

Services	Traffic types	No of devices	Loading	Priority	Latency	Throughput
Vehicular Communication System (VCS)	Sensitive	Thousands	High	1ms	Low	Very High
Smart Healthcare System (SHS)	Heavy	Thousands	Very High	5ms	Low	High
Smartphone	Popular	Billions	Medium	10ms	High	Medium

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

4.2 Service Slicing algorithm

We consider t_n and $q_n$for each slice n (n = 1, 2, 3.) where t_n represents technical requirements such as mobility management, tunnelling while q_n represents QoS criteria such as maximum bandwidth, minimum latency stated in Table.1. The algorithm used by SAM Model to decide whether to (A) allocate existing slice, (B) allocate existing slice after expansion, or (C) create service-dedicated slice on basis of service requirements and current slice's utilization is diagrammed in Fig 4. When the service operator S requires a slice, which characterized by $t_s$andq_s. It will send $t_s$and q_s to SAM Model and slice allocation will be calculated at SAM Model according to equation (1) and (2). $d_{t\,(n)}$ and $d_{q\,(n)}$represent the difference between the required (t_s, q_s) and current slice's (t_n ,q_n).

$$d_{t\,(n)} = t_n - t_s \qquad (1)$$

$$d_{q\,(n)} = q_n - q_s \qquad (2)$$

First, it determines whether S can be accommodated in a slice by calculating the parameters $d_t\,(n)$ and $d_q\,(n)$. For every slice n, if one or both parameters are always negative, accommodating S by using a current slice is impossible because no slice meets S's technical requirements or/and QoS criteria. Then, for every slice n, SAM Model calculates C_{en} , which is the cost of expanding the current slice, C_{oen} , which is the cost of operating the current slice after expansion, C_C , which is the cost of creating a service-dedicated slice, C_O , which is the cost of operating a created service-dedicated slice, and I, which is the loss of multiplexing gain. Next, SAM Model calculates d_{ctech} , which is the difference between cost of expanding a slice and the cost of creating a dedicated slice.

$$d_{ctech\,(n)} = C_{en+}\ C_{oen-}\ (C_{C+}\ \ C_o\ )+I) \qquad (3)$$

If some $d_{ctech}\ (n)$ is negative, SAM Model decides that the expansion cost is lower than the slice creation cost and expands the slice with the lowest $d_{ctech}\ (n)$. The service example, in this case, would fall within non-emergency communication such as Smartphone with (10 ms) priority at *Max*T as shown in Table 1. It could be high-resolution RTP streaming, which is feasible only if some priority parameters are expanded. Compared to a physical architecture, the proposed architecture expands virtually, so it is easier and more cost-effective to scale-up or scale-down.

If all $d_{ctech}\ (n)$ is positive, SAM Model decides that the lowest slice expansion cost is higher than the slice creation cost and creates a new slice of the service. The service example, in this case, would fall within emergency communication systems such as SHS and VCS (5ms) priority at *Max*T. It could be remote surgery service, which has service requirements (e.g., throughput, latency, and topology) that are very stringent. With the proposed architecture, the network provisioning cost for a service with a small number of users and tough requirements would be reduced greatly because of the reduced hardware cost and higher multiplexing gain.

If $d_t\,(n)$ and $d_q\,(n)$ are both positive, SAM Model decides that there is no technical problem with accommodating the service in a current slice. However, for a service that uses much more than enough node functions such as priority management at the slice, creating a slice with minimum functionality would reduce operation cost. Therefore, SAM Model decides whether to create a slice from the commercial phase. It calculates the cost of operating the service C_n for each slice with positive $d_t\,(n)$ and $d_q\,(n)$. Then, it calculates $d_{Ccomm}\,(n)$, which is the difference between the slice expansion cost and the slice creation cost.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

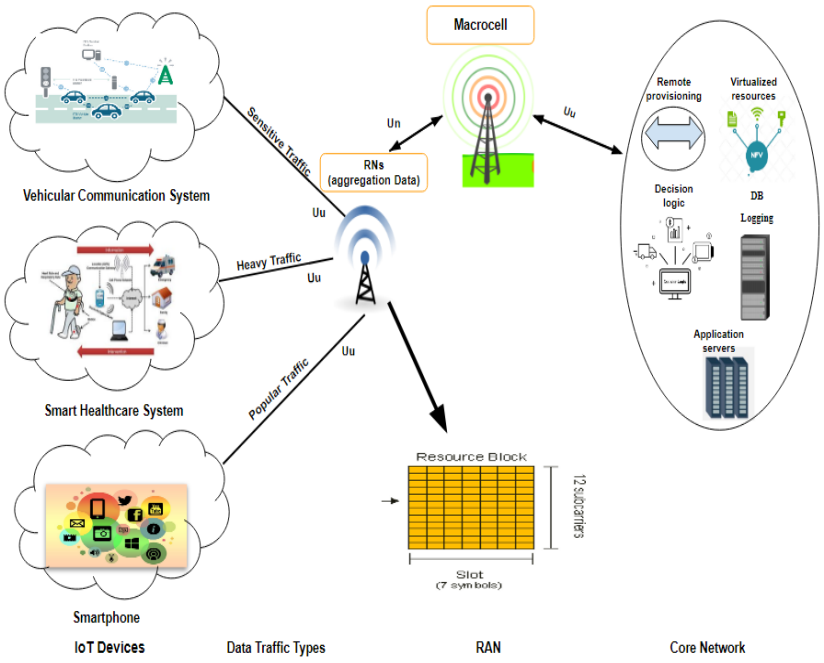


Fig.3. SAM Mechanism Environment

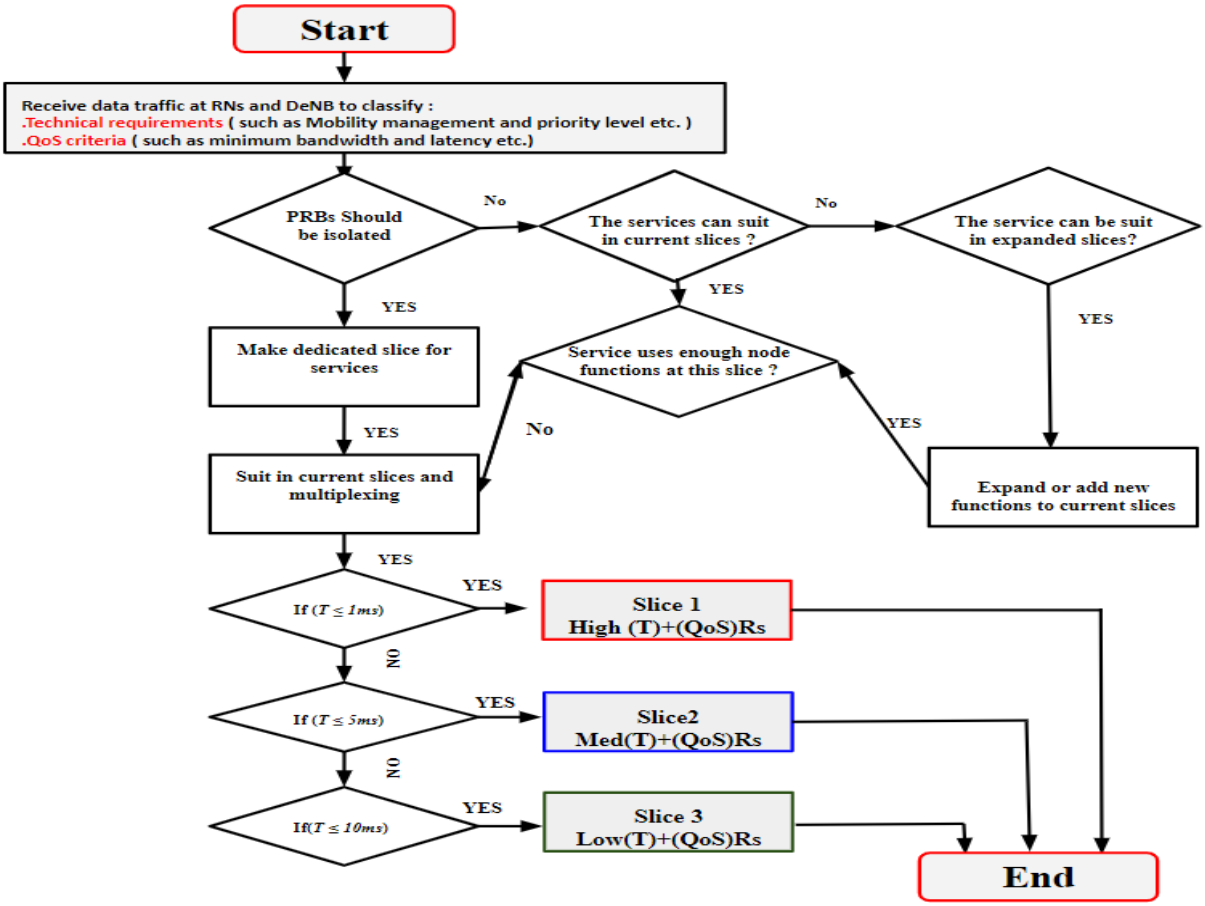


Fig.4. Service Slices Algorithm

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

$$d_{Ccomm}(n) = C_n - (C_c + C_o) + I) \quad (4)$$

If $d_{Ccomm}(n)$, is negative, SAM Model decides to accommodate the service by using an existing slice. If some slices have a negative $d_{Ccomm}(n)$, SAM Model accommodates the service in the slice with the lowest $d_{Ccomm}(n)$. The service example, in this case, would fall within middle emergency communication systems such as VCS (1ms) priority at *Max*T as shown in Table 1. It could be voice communication or browsing with requirements that fit within the current slice's capacity. If $d_{Ccomm}(n)$ is positive, the cost of operating the current slice's excessive node functions is higher than the slice creation cost. SAM Model thus creates a service dedicated slice for the service. The service example, in this case, would fall within non-emergency communication such as Smartphone. It could be a Smartphone service, which has a massive number of devices and low functional requirements.

5 SAM Edge Cloud Model

In this section, we have presented novel algorithms that can place Mobile Edge Computing (MEC) for direct placement of IoT device data traffic near the edge cloud with assigned priorities. These algorithms have been developed to reduce bandwidth consumption by placing caches at the edge of the targeted IoT nodes in the smart systems. Edge or fog cloud facilities will also participate in the management of network latency, routing, and load balancing, becoming the ingress points for the data coming from multiple heterogeneous sources and deciding if it must be analyzed locally or conveyed through a specific path to the cloud for further processing. Such a complex ecosystem is referred to as edge-fog cloud computing. Its scalability, flexibility, and performance characteristics represent a driving force for a new type of applications that involve effective and efficient data management and analytics, such as VCS or SHS applications with needing reliable low latency and data traffic management connected as illustrated in Figure. 5.

5.1 MEC Placement

Several applications and systems, including VCS, SHS, RTP streaming, and machine control can benefit from placing MECs close to the user. MECs running application layer services such as control processes, data pre-processing, and caching. The MEC can create significant performance improvements when being run at the network edge. The benefits of edge processing include low latency, caching at the edge, and reduction of data transfer to the core and local significance of data (including device-to-device communication).

With the term edge node, we mainly refer to RNs/DeNB cells, but the network may contain several hierarchical levels of edge nodes between the cells and the central data centre, such as regional nodes. The goal of the edge processing is to select those MECs that most benefit from being close to the IoT device and, based on their priority, place them on the available nodes. In case of all capacity of the edge nodes are already allocated, the MECs with the lowest need for being at the edge should be moved farther away from the edge. MECs implementing common functions, i.e. functions applicable to most of the users in the slice, should be placed on every edge node that is part of a slice. Thus, when the slice is created, MECs are created at all identified edge nodes. When a new edge node becomes part of a slice that an IoT device belongs to, the MECs will be installed on that node.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

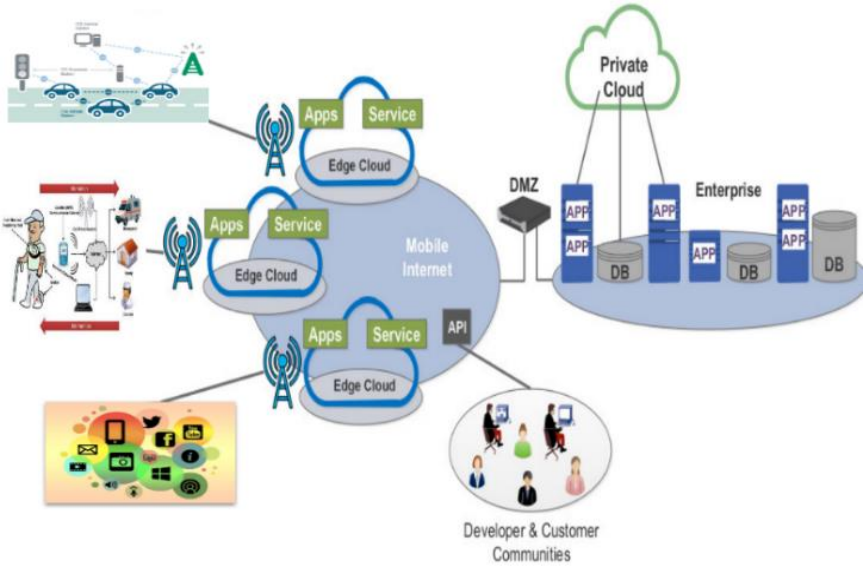


Fig 5: SAM Edge Cloud Model

Each MEC is assigned a priority for being located at the edge. The priority is increased by factors such as the need to minimize the latency or reduce bandwidth consumption by placing caches at the edge. On the other hand, each MEC has a cost in terms of the consumed resources. If MECs are placed at the edge, they have to be replicated to more nodes than in a centralized case. Although the needed capacity of the MEC is lower in the edge case, there is still overhead associated with each new MEC. Therefore, the priority also needs to consider the cost. Thus, if the cost is high in comparison to the benefits offered by edge placement of MECs, the priority should be low. As we have specified each slice with different priorities.

In addition, in terms of placing IoT devices closed to MEC a configuration is required. The network must be configured to assign the IoT devices to use the appropriate (closest) MECs. The choice of MEC software to install also depend on the nearest network. A MEC at the edge typically does not need to handle the same amount of traffic as the corresponding MEC more centrally located. Therefore, the MEC size, the image to be deployed or the scaling must be considered. This also affects the placement algorithm in that the capacity required by the MEC is a function of the location or the number of users served by the MEC. A MEC flow chart for this MEC service is provided in Figure 6. Precisely, radio conditions are monitored through a Radio Network Information Service (RNIS) specified per user (or per network slice). Different actions might be activated: the MEC controller may directly adjust network resources allocated to different network slices in order to efficiently handle slice SLA violations.

5.2 MEC Placement algorithms

In the placement MECs algorithm, we start from the cells as users traverse the networks toward the core data centre while looking for an available location for the edge MECs. We place the MEC near to the first cell with enough capacity for the user or IoT device. If they encounter the IoT device or user with the same MEC already installed, we stop. This may happen at the edge of the cell already if there is another user part of the slice. Along the path, we may need to move any of the existing MECs with a lower priority far away from the edge placement. These are moved toward the core using the same algorithm. Thus, when we encounter a MEC with a lower preference of the priority user at the edge, we remove the lower preference user and place the higher preference on the current MEC, and finally continuing the algorithm for the placement of lower priority user or IoT device. The algorithm is presented in Table .2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

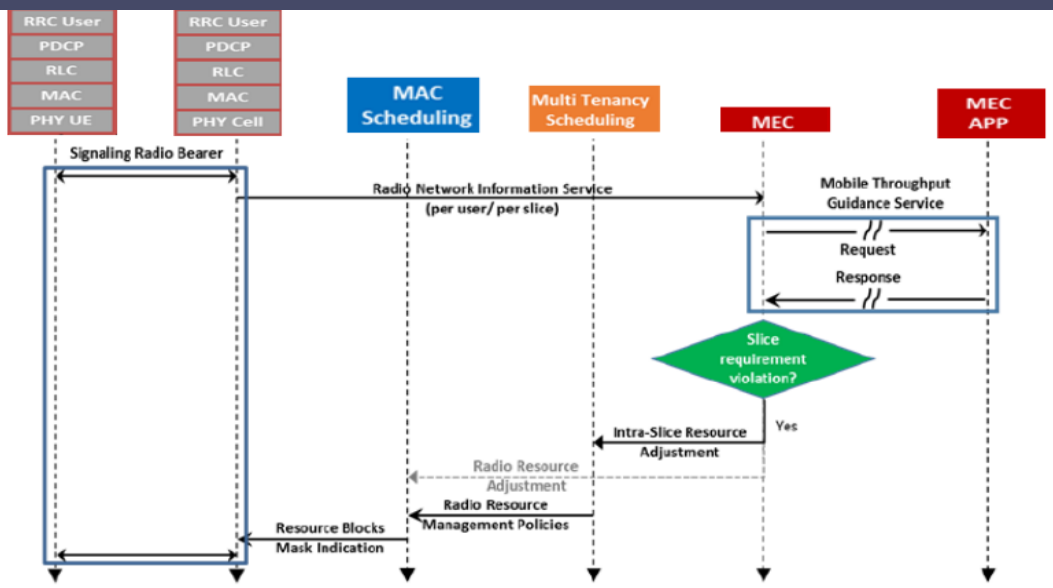


Fig 6: SAM Flow chart placement close to the edge

Table 2: Basic Algorithm for Placing MECs at Edge

Algorithm1 aim, Steps
Aim: Placing MECs at Edge
Steps:
1. $n :=$ **starting node**
2. $m :=$ **MEC instance to place**
3. **if** $T_m \in \{T\omega \mid \forall \omega \in m_n\}$ **then stop**
4. **if** $C_n \geq C_m$ **then** $M_n := M_n \cup \{m\}$; **stop**
5. $\omega := arg_\omega \, min \{P\omega \mid \forall \omega \in Mn\}$
6. **if** $P\omega < Pm$ **and** $C_n + C_\omega - C_m > 0$ **then** $M_n := M_n \setminus \{\omega\} \cup \{m\}$; $M = \omega$
7. **if** $P_n = \emptyset$ **then stop with failure**
8. $n := P_n$
9. ***Go to* 3**

In this algorithm, we denote C_n as the available capacity of node (users) n and C_m as the capacity required by MEC m. In practical applications, C_n and C_m are vectors consisting of multiple properties such as CPU power, memory, disk space, etc. each represented as an element. To consider the dependency of the location (e.g. the number of users served) C_m can be replaced by a function C_m (d, u), where d denotes the distance from the edge and u denotes the estimated number of users (or bandwidth) served. We further denote M_n as the set of MECs currently allocated to node n, T_m as the type or class of the MEC m, P_n as the parent node of node n in the hierarchical network topology, and Pm as the edge priority of MEC m. The edge priority indicates how important it is for the MEC to run at the edge.

We run the algorithm once per cell included in the slice. Thus, when the algorithm starts, n is the cell to which the user is connected. If later a new cell is added to the slice, the algorithm is run again. The above algorithm, in its simplicity, can only replace a single existing MEC with a lower priority MEC. The algorithm can be extended into a version that moves several of the existing MECs as needed, shown in Table 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

Table 3. Algorithm for Moving Lower Priority

Algorithm2 Aim, Steps
Aim: Moving Several Lower Priority MECs at Edge
Steps:
10. $n :=$ **starting node**
11. $m :=$ **MEC instance to place**
12. **if** $T_m \in \{T_\omega \mid \forall_\omega \in M_n\}$ **then stop**
13. $M_{low} := \{\omega \mid \forall_\omega \in V_n, P_\omega < P_m\}$
14. $C_{low} := \sum_\omega \in M_{low} C_\omega$
15. **if** $C_n + C_{low} - C_m < 0$ **then** $n := P_n$ **; go to 3**
16. $M_n := M_n \setminus M_{low} \cup \{m\}$
17. **For each** $m \in M_{low}$ **run this algorithm with** $n := P_n$

Depending on the policy, moving MECs can consider priorities across network slices. Thus, a higher priority MEC in one slice might cause a lower priority MEC in another slice to be relocated. This requires that priorities are specified in a uniform way. Using such a policy may improve the use of processing resources and the overall QoE across slices, but on the other hand, causes undesired dependency between slices where the slice performance may be impacted by events in other slices. Mobility also affects deciding which MECs are located at the edge. MECs that are affected negatively by mobility need to be farther away from the edge. These are the MECs that are specific to a given user or groups of users. To consider for these cases, a lower priority (possibly even negative) can be assigned to these MECs. In a multi-level hierarchical network, an alternative approach would be to start the algorithm from a higher layer starting node, e.g. $P_{(n)}$, for these MECs. For removed MECs or the case of the last user of a slice leaving the cell, a similar process is started in reverse. The aim is to optimize the use of edge nodes in this new situation as shown in table 4.

Table 4. Major Symbols Use

Slice Allocation Algorithms Symbols			
Symbol.	Meaning	Symbol.	Meaning
n	Each Slice	$Coen$	The cost of operating the current slice after expansion
t	Technical requirements	Cc	cost of creating a service-dedicated slice
q	QoS criteria	Co	The cost of operating a newly created service-dedicated slice.
S	Service operator	$dctech$	The difference between cost of expanding a slice and the cost of creating a slice
d	Difference between the required and current slice's	l	The loss of multiplexing gain
Edge Cloud Allocation Algorithms Symbols			
n	Node (user)	d	Distance from the edge
m	MEC (edge cloud)	u	Estimated number of users (or bandwidth) served
C_n	Capacity of node	T_m	The type or class of the MEC
C_m	Capacity required by MEC	P_n	The parent node of node
M_n	The set of MECs currently allocated to node	Pm	The edge priority of MEC
Cen	The cost of expanding the current slice	$dCcomm$	The difference between the slice expansion cost and the slice creation cost

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

6 Simulation environment

In this section, we are using a simulation tool was OPNET version 18.5, we have considered the LTE-A nodes with LTE-A protocols, which were modified to be suitable along with 5G features. The remote server includes SMTP, FTP, VoIP and RTP applications among all smart systems. The remote server and the aGW are connected with an Ethernet link with an average delay of 20 ms. the aGW node protocols contain Internet Protocol (IP) and Ethernet. The aGW and eNB nodes (eNB1, and RNs1, 2, 3 ...) connect over IP edge cloud (1, 2, 3 and 4). QoS Parameters at the Transport Network (TN) guarantees QoS parameterization and traffic difference as seen in Figure 7 and Table 5. The user mobility in a cell is coordinated by the mobility model by updating the location of the users at every single interval. The user's mobility data is saved on the Global Server (Global-UE-Server). The channel Model parameters for the air interface cover slow fading, fast fading models, and path loss. The simulation emphasises on the user plane to execute E2E performance assessments [32]. The several traffic QoS have been established in relation to the 3GPP standardization.

6.1 Simulation Setup

The Optimized Network Engineering Tool (OPNET) is a simulation used to assess the performance of the proposed scheme. Several scenarios are simulated to evaluate the impact of smart devices data traffic on regular 4G and 5G mobile networks data traffic. The simulated 4G and 5G data traffic applications include FTP, VoIP SMTP and RTP. The scenarios are categorized into four scenarios the first one was designed for 4G mobile networks without density connection of devices and the other three were designed for supporting 5G mobile networks the first two 5G mobile networks in density connection of IoT devices without small cells and edged clouds, the last one 5G mobile networks with density connection of IoT devices in the form of slicing based on our smart systems use case requirement and it was supporting of small cells and edge clouds. The results show the significant impact of smart devices data traffic on high priority data traffic. The E2E network performance has been improved by allocated data of several IoT devices, which is determined by simulating several scenarios. Considerable performance improvement is achieved in terms of average IoT device and cell throughput, average upload response time, average packet end-to-end delay and radio resource utilization in the SMTP, VoIP, FTP and RTP applications. [29].

Table 5: Simulation parameters

Parameters	Setting
Simulation length	2000 sec
Cell layout	1 Enb
eNB coverage radius	350 m
Min. Enb-UEs	35 m
Max. terminal power	23 dBm
5G Parameters	
5G cell	8*8 antennas
MEC	Edge could difference latencies
Capability	Enabled
RN Parameters	
PRBs are allocated to Cells	RN 1 = 50 PRBs, RN 2= 25 PRBs, RN3 = 15 PRBs and DeNB 50 PRBs to evaluate PRB utilization.
Type of RN	Fixed
RN 1	Supported by 6 antennas, 10 MHz TDD
RN 2	Supported by 3 antennas, 5 MHz TDD
RN 3	Supported by 1 antenna, 3 MHz TDD
TBS capacity	1608 bits against MCS 16 and PRBs 5. Available service rate TBS-overhead (bits/TTI), 1608 (TBS)-352(overhead) =1256 bits/TTI.
General Parameters	
Terminal speed	120 km/h
Mobility model	Random Way Point (RWP)
Frequency reuse factor	1
System Bandwidth	5 MHz
Path loss	128.1+37.6 *log* 10 *(R)*. R in km
Slow Fading	Log-normal shadowing, correlation 1, deviation 8 Db
Fast Fading	Jakes-like method
UE buffer size	∞
RN PDCP buffer size	∞
Power control	Fractional PC, α= 0.6, *Po*= -58 dBm
Applications	SMTP, VoIP, RTP and FTP.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

6.2 QoS of Radio Bearers

The LTE QoS has gained considerable importance in the designing and planning of the networks. There are possibilities to use the LTE network for various operations. For example, some subscriber uses the network services for emergency cases, while others use the services for entertainment purposes. QoS explains how a network serves subscribers due to the enclosed network architecture and protocols. In LTE, the term bearer can be defined as the flow of an IP packet between the UE and P-GW. Each bearer is linked with a particular QoS parameter. The network provides almost the same services to the packets which are linked to individual or same bearer. For establishing a communication path between UE and PDN, UE attempts to generate a bearer by default. Such bearers are called default bearers. The other bearers are named as dedicated bearers which are established to the PDNs. Establishing more than one bearer is possible.

This is because one user demands several services and each service demands specific bearer. For example, if a bearer is established, it is possible to generate more bearers in the presence of an existing bearer. Moreover, the QoS value of an existing and newly created bearer is possible to vary. The bearer can be classified into Guaranteed Bit Rate (GBT) and Non-Guaranteed Bit Rate (Non-GBR). • The GBR bearer has a minimum bandwidth which is allocated by the network for various services such as voice and RTP communication, regardless of that are used or not. Due to dedicated system bandwidth, the GBR bearer does not undergo any packet loss due to congestion and are free from latency.

• Non-GBR bearer is not allocated a specified bandwidth by the network. These bearers are used for best-effort services such as web browsing, SMTP, etc. These bearers might undergo packet loss due to congestion.

• Quality Control Identifier (QCI) describes how the network treats the received IP packets. The QCI value is differentiated according to the priority of the bearer, bearer delay budget and bearer packet loss rate. 3GPP has defined several QCI values in LTE, which are summarized in Table.5.

6.3 Radio Resource scheduling

Packet scheduling is the distribution of radio resources between the radio bearers in a cell by the HeNB/DeNB. In 3GPP LTE standards, this task is performed by the MAC scheduler in the HeNB/DeNB. The allocation of the downlink and uplink radio resources by the HeNB/DeNB to the UEs depends upon the data present in the buffers of the HeNB/DeNB and the UEs respectively. If the data for a particular UE is present in the buffer of the HeNB/DeNB, then the HeNB/DeNB allocates radio resources to the UE for downlink transmission if eNB has enough available radio resources and the QoS requirements of the other UEs located in the coverage area of the HeNB/DeNB are fulfilled. Similarly, in uplink transmission, the UEs transmit Buffer Status Report (BSR) information to the eNB for granting radio resources if there is data present in the buffer of the UEs. UE BSR information also identifies the types of traffic in the UE buffer. The HeNB/DeNB allocates radio resources for downlink and uplink according to the radio bearers QoS requirements of the UE. Time Domain-Maximum Throughput (TD-MT) scheduler provides the radio resources to the UEs close to HeNB/DeNB and bears good channel conditions. The users at the cell-edge may not get radio resources. The TD-MT scheduler provides maximum throughput at the cost of fairness [33], which can be expressed simply as in as shown in table 6 and (Equation 5):

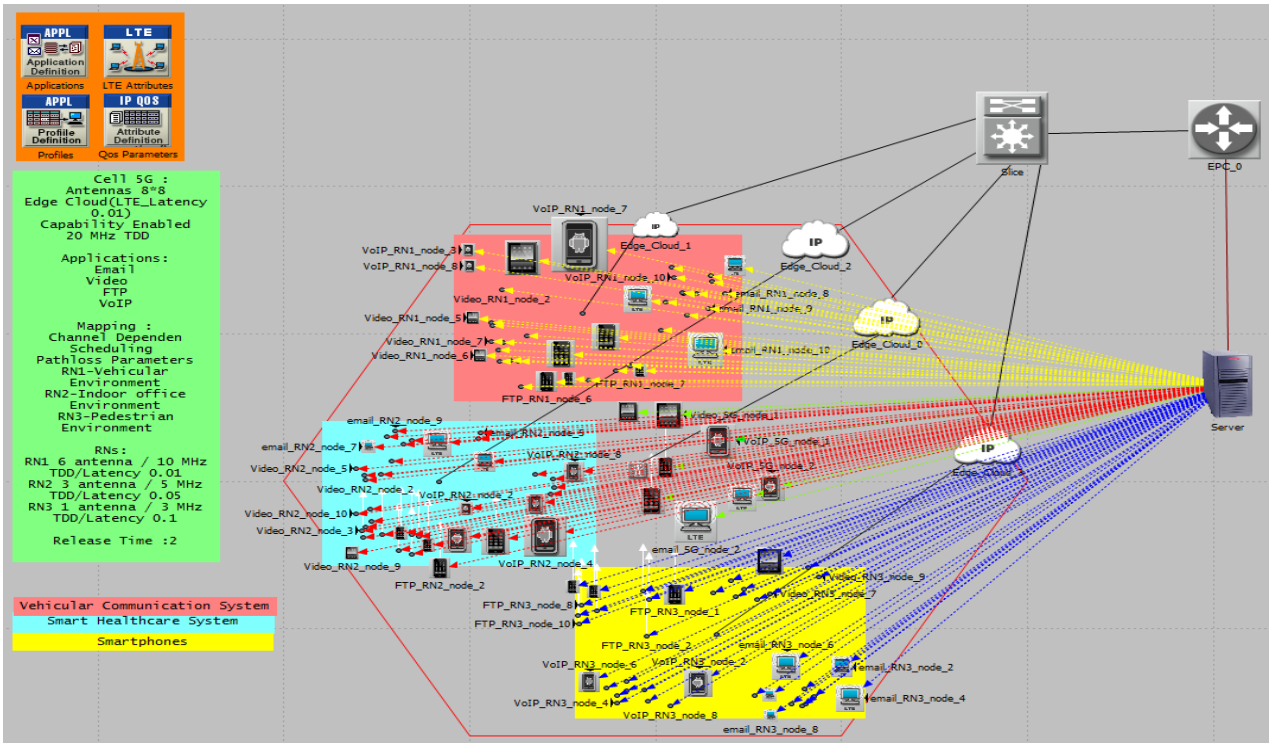
$$\boldsymbol{PkT\ D = rk(t)} \qquad (\mathbf{5})$$

Table 6: LTE QCI value[33]

QCI	Resource	Delay	Priority	Error	Services type
1	GBR	100 ms	2	10-2	Conversational (VoIP)
2		150 ms	4	10-3	Conversational (Video)
3		50 ms	3	10-3	Real Time gaming
4		300 ms	5	10-6	Non-Conversational voice
5		100 ms	1	10-6	IMS signaling
6		300 ms	6	10-6	RTP Buffered streaming
7	Non GBR	100 ms	7	10-3	TCP based (SMTP, HTTP, FTP)
8		300 ms	8	10-6	Voice, RTP and interactive gaming
9		300 ms	9	10-6	RTP Buffering streaming

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

Fig.7. OPNET 5G Project



7 Simulation Scenarios

In this section, we evaluate and compare performance results of the proposed models based on four scenarios. These scenarios are based on the RNs and DeNB cells to assess the three slices of smart systems based on throughput, load and latency requirements within high density environment of IoT devices. The first scenario (scenario 1) is 4G mobile networks with traditional 4G smartphones without high density of IoT device connections. The second scenario (scenario 2) is 5G mobile networks with high density of IoT device connections without edge clouds. The third scenario (scenario 3) is 5G mobile networks with high density of IoT device connections with no support for small cells. The fourth scenario (scenario 4) is the proposed SAM model in 5G mobile networks with high destiny of IoT device connections, with edge clouds and small cell supports as illustrated in Table 7. The data packets from all the active smart devices, which are positioned in the nearness of the RNs and DeNB cells, are allocated slice services at the RN cell before being sent to the DeNB[33].

However, only the periodic per-hop control model is used in which the large allocation data packets are served to guarantee full utilization of RAN. The expiry timer is presented in order to limit the multiplexing delay particularly in the low loaded scenarios between RN and DeNB cells. In this situation, the allocated packet is served after Tmax at the latest. All the stated scenarios are further sub-categorized into three sub-scenarios. In the first sub-scenario, VCS, the IoT devices are placed near to RN1 cell, which is supported by 6 antennas and 10 MHz TDD with a low level of priority 5 ms in both the small cell and edge cloud. The second sub-scenarios, SHS, the IoT devices are placed near to RN2 cell, which is supported by 3 antennas and 5 MHz TDD with a medium level of priority 10 ms in both the small cell and edge cloud. The Third sub-scenario's smartphone, and devices are placed near to RN3 cell, which is supported by 1 antenna and 3 MHz TDD with a medium level of priority 15 ms in both the small cell and edge cloud.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

Table 7: Simulation Scenarios

Scenarios	DeNB Cell	Small Cells	(MEC)	SDN (Slicer)	Smart Systems	Application Types
4G Mobile Broadband No Density	4G	Yes	No	No	Mobile Broadband	SMTP, VoIP,FTP & RTP
5G Density No Edge Clouds	5G + MIMO	Yes	No	Yes	All	SMTP, VoIP,FTP & RTP
5G Density No Small Cells	5G + MIMO	No	Yes	Yes	All	SMTP, VoIP,FTP & RTP
SAM Density with Small Cells Edge Clouds	5G + MIMO	Yes	Yes	Yes	All	SMTP, VoIP,FTP & RTP

8 Simulation Results and Analysis

8.1 Evaluation IoT Nodes E2E delay

In this subsection, we compare the evaluation results in terms of IoT nodes end-to-end delay of the four scenarios and the three sub-scenarios as mentioned above: Figures 8 and 9 illustrate the average air interface packet E2E delay of FTP and SMTP nodes. The results show that the FTP and SMTP nodes have the diverse E2E delay variation in all four scenarios even when allocated together with GBR bearers. This is due to the proportional varieties distinguishing of priorities, which is characterized by SAM model in “scenario2, 3 and 4”. Meanwhile, the VoIP bearer has a relatively low level of packets E2E delay in the “scenario 4” compared to “scenario1, 2, and 3”. As a result of the support by SAM and MEC placing models within the small cells and edge clouds with RAN allocation in the shape of slicing, it tends to get higher priority feature and will permanently be scheduled first. Then, the average packets E2E delay of VoIP and RTP node are shown in Fig10 and 11. It can be seen that “scenario 4" has somewhat better packets E2E delay compared to “scenario1, 2 and 3". However, I can see an example "scenario1", in RTP node has better performance compared to "scenario 2, and 3" .This result is due to the fact that there is low density of IoT device connections, which would have caused low level packet overload and congestion. In addition, scenario 4 is a product of SAM and MEC placement models, allocating RAN to VOIP and RTP node bearers in the RN1. This helps to customize the slice to support SHS with higher MAC QoS class for VoIP SMTP IoT devices in this sub-scenario.

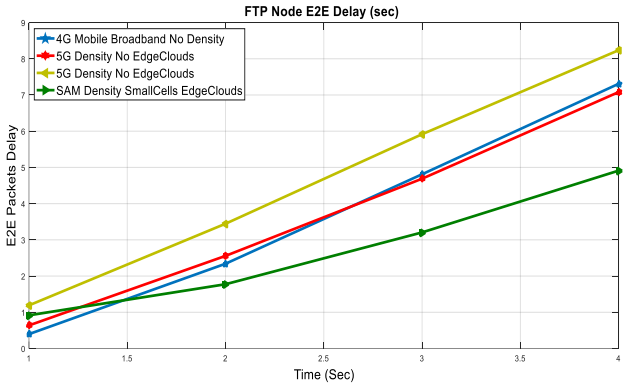


Fig 8. FTP Node E2E Delay

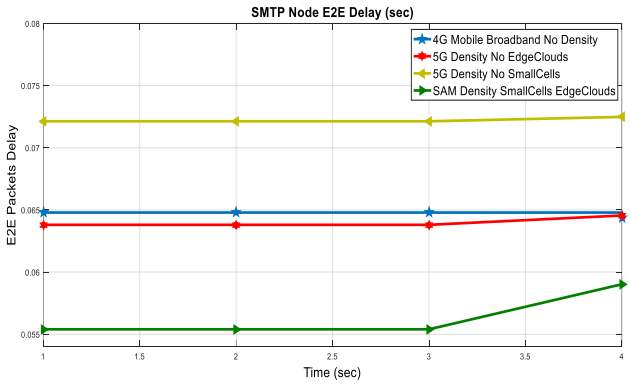


Fig 9. SMTP Node E2E Delay

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

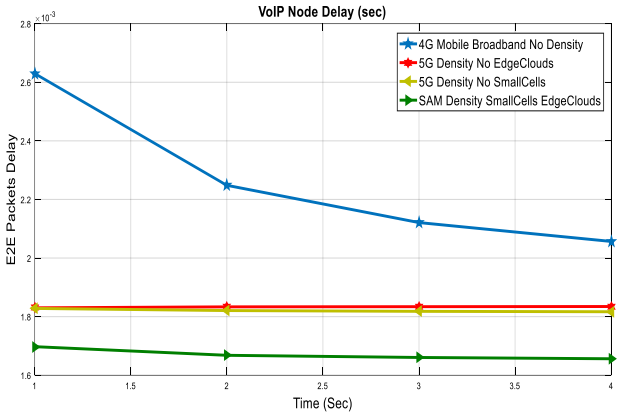


Fig 10. VoIP Node Delay

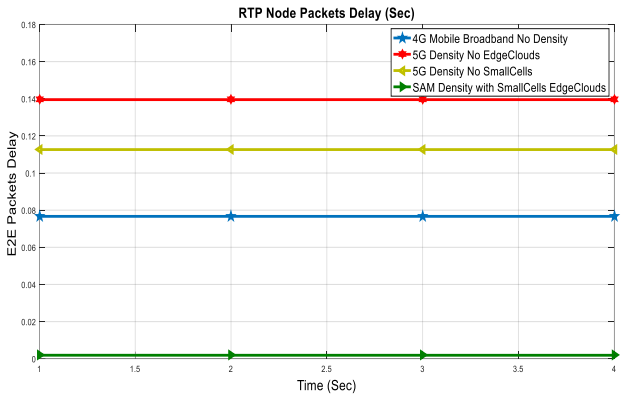


Fig 11. RTP Node Packets Delay

8.2 Evaluation IoT Nodes Loading

The performance of packet load of the VoIP nodes of scenarios 1 and 4 is shown in Fig. 12. The VoIP node has a significantly higher ratio of loading performance compared to "scenario 2, and 3". This is result is due to scenario 1 having low density and low connectivity of IoT devices as well as scenario 4 having customized slice due to the SAM and MEC placement models. On the other hand, the SMTP node in Fig 13, in "scenario 4" has best loading compared with the other scenarios, where the SMTP node has served with specific priority requirement in "scenario 4" in the form of slicing based on SAM and MEC placing models with existing small cells and edge clouds or without, mostly when it is not mixed with the FTP and RTP nodes and is allocated to a lower MAC QoS class than FTP. Therefore, the QoS is customized for different slices same as in "scenario 4" but not customized in the "scenario1".

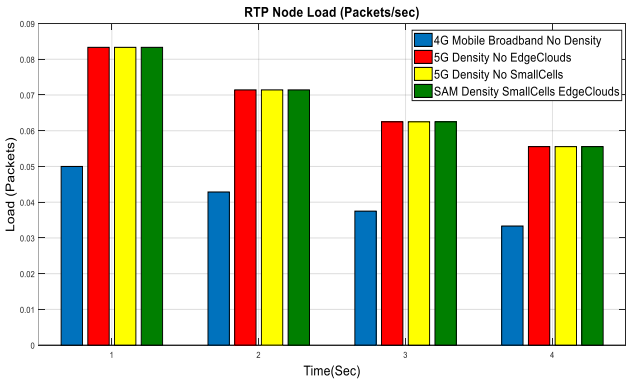


Fig 12. VoIP Node Load per Packets

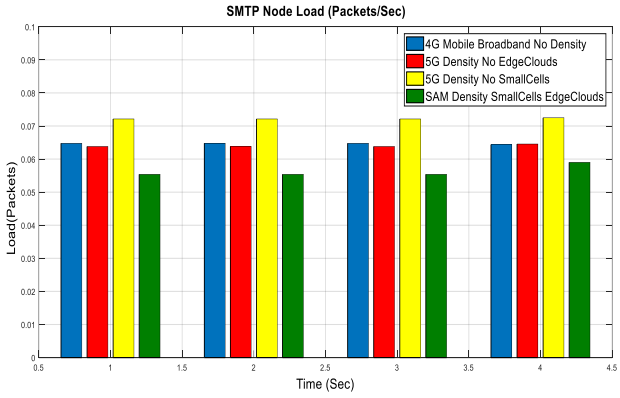


Fig 13. SMTP Node Load per Packets

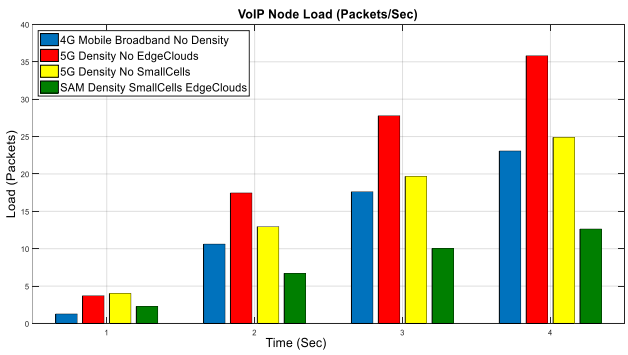


Fig 14. RTP Node Load per Packets

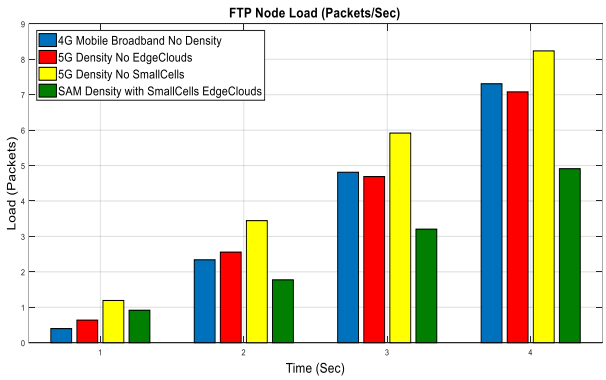


Fig 15. FTP Node Load per Packets

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

In addition, priority at different levels based on the needs of applications and smart systems, DeNB cells and edge clouds in scenarios 2,3 and 4, using SAM and MEC placement models, the results of FTP ands RTP nodes packet loads are as shown in Fig. 14 and Fig. 15. As expected, the RTP node performance decreases from mixed scenario to fully separated scenario as FTP loading time improves in scenario 4. This is due to the FTP node being allocated to the lower level of MAC QoS class and is supported with low priority in the RNs, DeNB cells and edge clouds as compared to the other applications. However, offering the FTP node lower priority is realistic since FTP is not a real-time application and in real life, it is acceptable for the FTP IoT devices to wait a couple or more seconds for their files to be sent, whereas the same is not acceptable when it comes to real-time applications such as RTP or VoIP.

8.3 Evaluation IoT Nodes Throughput

In section, we present the performance of the throughput with different protocols, we have started via the throughput of FTP and VoIP nodes as shown in Fig 16 and 17 the throughput for the FTP and VoIP IoT devices, the result describes that the FTP and VoIP IoT devices have worse performance in the “scenario1” compared to “scenario2,3 or 4” where the FTP nodes are allocated into the GBE MAC classes. In the “scenario 4”, the FTP and VoIP nodes do not share the same non-GBR MAC QoS class with SMTP, RTP nodes in the same slice, which has different priority in the RNs, DeNB cells and edge clouds, since I customize the slices and IoT device such as the data rate of the FTP node. It was observed that in all nodes, downlink and uplink throughput has a significantly higher ratio in the “scenario 4” compared to “scenario1, 2 or 3” as results seen in Fig 18 and 19 demonstrate. This is due to the SAM and MEC placing Models allocating the packets to the different levels of MAC QoS class and being supported with low priority in the RNs, DeNB cells and edge clouds.

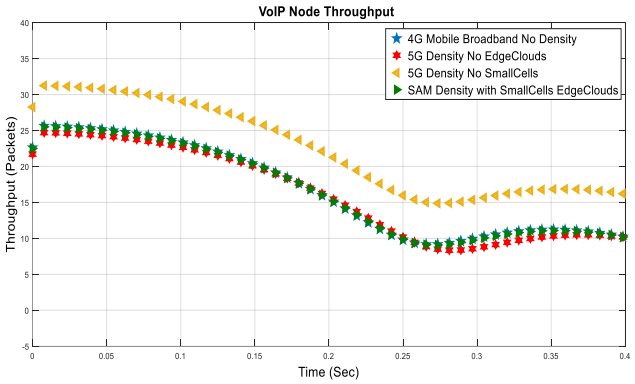


Fig 16. FTP Node Throughput

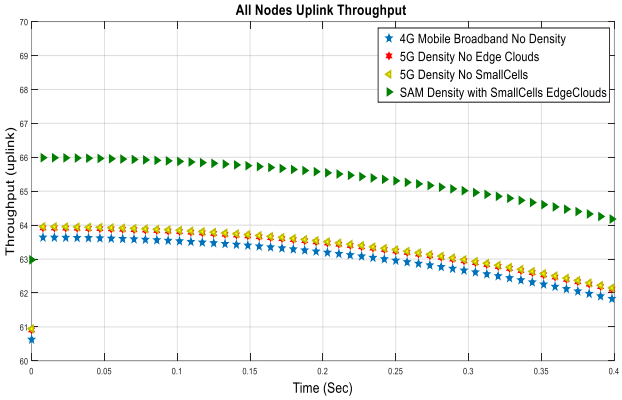


Fig 17. VoIP Node Throughput

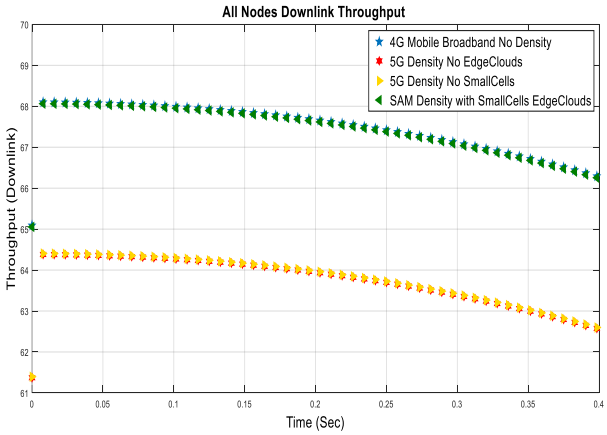


Fig 18. All Nodes Uplink Throughput

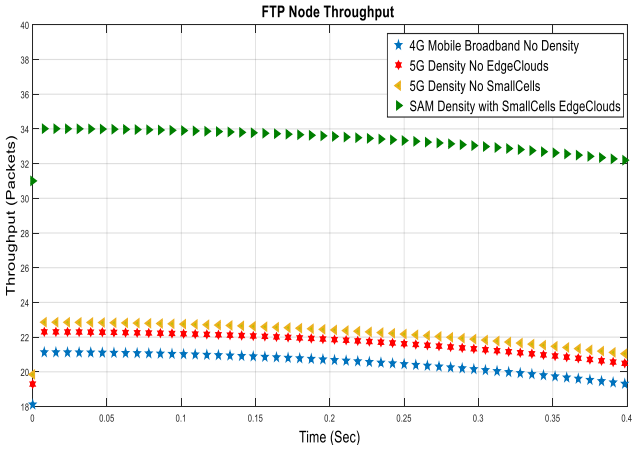


Fig 19. All Nodes **Downlink** Throughput

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

8.4 Evaluation RNs and DeNB cells Performance

In this section we present the evaluation results of the RNs and DeNB cells load, delay and throughput as shown in Fig 20, 21, and 22. It can be seen that overall load, delay and throughput of cells are evaluated in just one scenario, which is "scenario 4" (SAM Density with Small Cells and Edge Clouds), where the IoT devices are connected directly to DeNB while some other IoT devices placed close to cell edge communicate with DeNB through the RNs, therefore, this part is focused on the individual cell to evaluate the overall performance and assess the SAM Model slicing process, when the single IoT devices are communicating individually via RNs or DeNB cells with the core networks. SAM Model has a positive impact on the network by freeing the network resources, which ultimately increases the DeNB cell load, throughput and improving the networks E2E packets delay. The overall DeNB cell load and throughput are lower levels compare with small cells as a result of enhancing slicing of the smart systems and separating the load and throughput among RN1, RN2 and RN3 cells with the result that, for example, the load and throughput in the RN1 has a significantly higher performance for VCS compare with RN2 SHS and RN3 (smartphone).

Therefore, in the density environment of IoT devices in which I have considered both the VCS and SHS devices in the RN1 and RN2 cells, along with normal Smartphone users in RN3 cell, the performance evaluation showed there are RAN allocated in each slice based on smart systems QoS requirements, the single IoT devices are communicating individually via RNs or DeNB cells with the core networks. Moreover, the performance evaluation showed there are RAN allocated in each slice based on smart systems QoS requirements, the cells have different levels of E2E delay based on the slices close to RNs cells .The RN1 has lower level of E2E delay can offer to the IoT devices closed to this cell such as VCS, compared with the RN3 has the high level of E2E delay can serve the IoT devices closed to this cell such as smartphone.

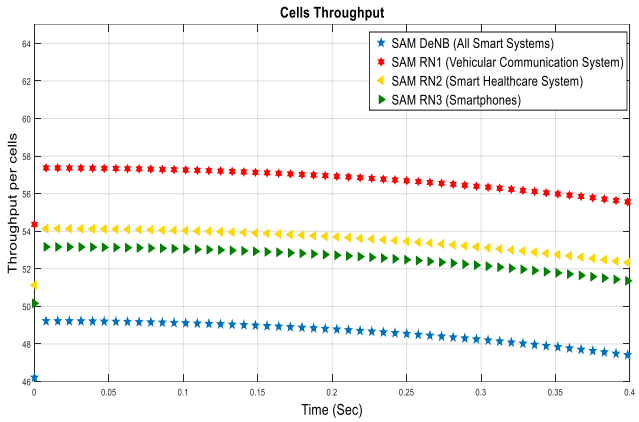


Fig 20. Cells E2E Delay

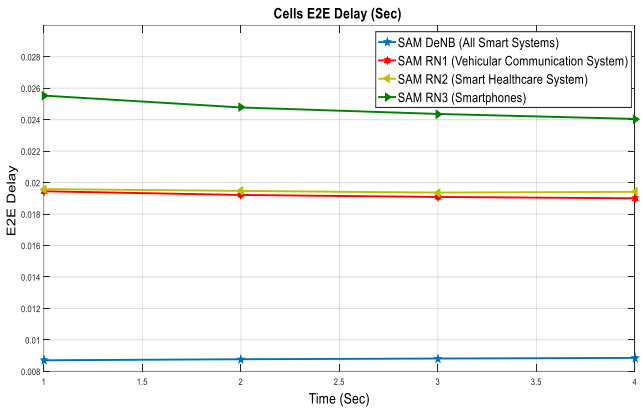


Fig 21. Cell Load per Packets

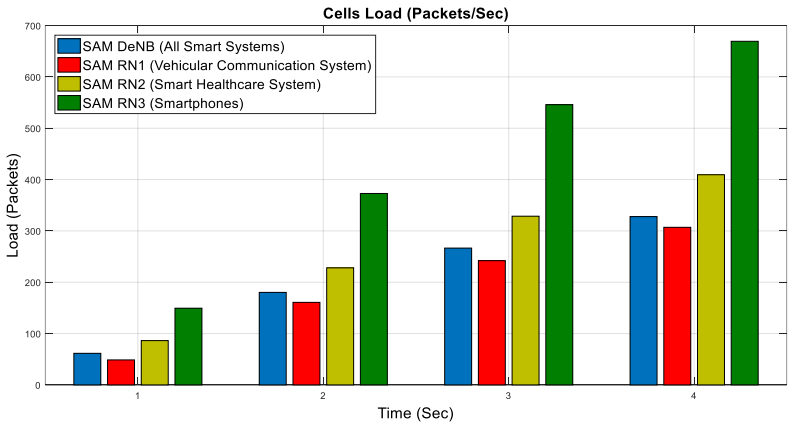


Fig 22. Cells Throughput

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2020 doi:10.20944/preprints202007.0536.v1

8.5 Discussion

In this article, SAM and MEC models have been proposed as an efficient solution to the problem of traffic congestion and data overload providing reliable and low latency for real time IoT applications in 5G networks. In the proposed solution, unique slices are allocated to each smart system application such as VCS, SHS, and smartphone in the 5G mobile networks, based on the technical and QoS requirements. It uses well-recognized and promising network slicing technology, enabling cost-effective service deployment and an effective operational model. The proposed models were designed and developed for high density IoT environment, considering VCS and SHS devices in RN1 and RN2 cells. It also considers normal smartphone users in RN3 cells. The performance evaluation demonstrates there are RAN allocated in each slice based on smart systems QoS requirements, the single IoT devices are communicating individually via RNs or DeNB cells with the core networks. In addition, when RAN is allocated to each slice based on the QoS requirements of the smart systems with different levels of E2E delays considering the closeness of the slices to the RN cells, RN1 produced lower level of E2E delays for IoT devices. Compared to RN1, the RN3 has higher level of E2E delays for IoT devices such as smartphones, which are closer to the cell.

9 Conclusions

The latest 4th Generation (4G) and the upcoming 5th Generation (5G) mobile technologies are expected to offer massive connectivity and management of high volume of data traffic in the presence of immense interferences from mobile networks of IoT devices. This development comes with various challenges, especially that of congestion and overload of data traffic due to humongous number of IoT devices. This means there will be demand for high throughput, low latency and high level of reliability especially for critical real-time applications such as in Vehicular Communication System (VCS). To address these issues in 5G mobile networks, this paper proposed a Slice Allocation Management (SAM) Model based on critical services of smart systems such as VCS to satisfy QoS demands. The main contribution of this article is the development of the SAM and MEC models, which are able to work in high density environment of connected IoT devices, with specific focus and application in both the VCS and SHS devices in the RN1 and RN2 cells, along with normal Smartphone users in RN3 cell.

The proposed models have been evaluated using the OPNET simulator to measure their performance. The simulation considers IoT devices in several smart systems such as VCS, SHS and smartphones also, diverse protocols contain SMTP, FTP, VoIP and RTP. The performance evaluation of the models showed that with the RAN allocated in each slice based on smart systems QoS requirements, every single IoT device can communicate individually via RNs or DeNB cells with the core networks. In addition, the performance evaluation also showed with RAN allocated to each slice based on smart systems QoS requirements, the cells have different levels of E2E delay based on the closeness of the slices to RN cells. The results also show that the RN1 has lower level of E2E delay for IoT devices closer to the cells such as VCS, compared with the RN3 that has the high level of E2E delay serving the IoT devices closer to the smartphone cells.

Simulation results demonstrate an important improvement of IoT node packets transmission through RNs and DeNB cells, in the proposed SAM model scenario when compared with other scenarios with or without density area. The models have enhanced E2E delays in FTP node by reaching 1ms, loading in VoIP node by 80% and throughput of all nodes in the uplink side of networks by 66%. Moreover, this work presents a service-oriented network in a top-down manner. we have improved the uplink network infrastructure of the 5G heterogeneous network by applying RNs as small cells and MEC as Edge cloud. These improvements can be useful for 5G network slicing operators and tenants, utilizing the appropriate slices in both QoS and technical support.

For the future works, we will consider the core 5G network infrastructure of 5G such as SDN and NFV in terms of managing programmable data and user plans, which have relations to improving the downlink side of the 5G networks. In addition, we would investigate the measurement radio of resource allocations in terms of Signal to Interference plus Noise Ratio (SINR) in heterogeneous networks, which is a result of transmission of small-cells users to the macro-cell users, since interference within small-cells is considered detrimental to performance improvement.

Acknowledgements

This work is supported by Liverpool John Moores University (LJMU), Saudi cultural bureau in London and the University of Technology and Business (UBT) Jeddah Saudi Arabia.

References

[1] NGMN Alliance, "Description of Network Slicing Concept," *ngmn*, 2016.

[2] Y. Shoji, M. Ito, K. Nakauchi, L. Zhong, Y. Kitatsuji, and H. Yokota, "Bring your own network - A network management technique to mitigate the impact of signaling traffic on network resource utilization," in *2014 IEEE 11th Consumer Communications and Networking Conference, CCNC 2014*, 2014, pp. 182–187, doi: 10.1109/CCNC.2014.6866568.

[3] NGMN, "Description of Network Slicing Concept by NGMN Alliance," 2016.

[4] M. Iwamura, "NGMN view on 5G architecture," in *IEEE Vehicular Technology Conference*, 2015, vol. 2015, doi: 10.1109/VTCSpring.2015.7145953.

[5] A. Gupta and R. K. Jha, "A Survey of 5G Network: Architecture and Emerging Technologies," *IEEE Access*. 2015, doi: 10.1109/ACCESS.2015.2461602.

[6] 5GPPP Architecture Group, "5G PPP Architecture Working Group: View on 5G Architecture," 2016.

[7] M. Jiang, M. Condoluci, and T. Mahmoodi, "Network slicing management & prioritization in 5G mobile systems," 2016.

[8] F. Ghavimi and H. H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 2, pp. 525–549, 2015, doi: 10.1109/COMST.2014.2361626.

[9] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 27–35, 2013, doi: 10.1109/MCOM.2013.6553675.

[10] G. Casale, E. Z. Zhang, and E. Smirni, "Trace data characterization and fitting for Markov modeling," *Perform. Eval.*, vol. 67, no. 2, pp. 61–79, 2010, doi: 10.1016/j.peva.2009.09.003.

[11] M. Chen, J. Wan, and F. Li, "Machine-to-machine communications: Architectures, standards and applications," *KSII Transactions on Internet and Information Systems*, vol. 6, no. 2. pp. 480–497, 2012, doi: 10.3837/tiis.2012.02.002.

[12] Y. E. Massad, M. Goyeneche, J. J. Astrain, and J. Villadangos, "Data Aggregation in Wireless Sensor Networks," *2008 3rd Int. Conf. Inf. Commun. Technol. From Theory to Appl.*, vol. 2, no. June, pp. 1040–1052, 2008, doi: 10.1007/978-3-642-13965-9_3.

[13] M. T. Lazarescu and L. Lavagno, "Wireless sensor networks," in *Handbook of Hardware/Software Codesign*, 2017.

[14] M. M. Rahman, C. Despins, and S. Affes, "HetNet Cloud: Leveraging SDN & Cloud Computing for Wireless Access Virtualization," 2015, doi: 10.1109/ICUWB.2015.7324454.

[15] M. Dighriri, G. M. Lee, and T. Baker, "Measurement and classification of smart systems data traffic over 5g mobile networks," in *Technology for Smart Futures*, 2017.

[16] M. Hasan, E. Hossain, and D. I. Kim, "Resource allocation under channel uncertainties for relay-aided device-to-device communication underlaying LTE-A cellular networks," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 4, pp. 2322–2338, 2014, doi: 10.1109/TWC.2014.031314.131651.

[17] M. Hasan, E. Hossain, and D. I. Kim, "Resource allocation under channel uncertainties for relay-aided device-to-device communication underlaying LTE-A cellular networks," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 4, pp. 2322–2338, Apr. 2014, doi: 10.1109/TWC.2014.031314.131651.

[18] S. Muppala, G. Chen, and X. Zhou, "Multi-tier service differentiation by coordinated learning-based resource provisioning and admission control," *J. Parallel Distrib. Comput.*, vol. 74, no. 5, pp. 2351–2364, May 2014, doi: 10.1016/j.jpdc.2014.01.004.

[19] W. S. E. Architecture, "Description of Network Slicing Concept by NGMN Alliance," vol. 1, no. September, pp. 1–11, 2016.

[20] 5G PPP Architecture Working Group, "View on 5G Architecture," *White Pap.*, no. July, 2016, doi:

10.13140/RG.2.1.3815.7049.

[21] H. Zhang, N. Liu, X. Chu, K. Long, A. H. Aghvami, and V. C. M. Leung, "Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, 2017, doi: 10.1109/MCOM.2017.1600940.

[22] E. Van Glabeke, P. Philippe-Chomette, and C. Grapin, "App RAN Application Oriented Radio Access Network Sharing in Mobile Networks," *Arch. Pediatr.*, vol. 5, no. 7, pp. 783–784, 1998, doi: 10.1016/0929-693X(96)89900-3.

[23] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017, doi: 10.1109/COMST.2017.2705720.

[24] J. S. Panchal, R. D. Yates, and M. M. Buddhikot, "Mobile network resource sharing options: Performance comparisons," *IEEE Trans. Wirel. Commun.*, 2013, doi: 10.1109/TWC.2013.071913.121597.

[25] 3GPP, "NR; NR and NG-RAN Overall Description; Stage 2," 2018.

[26] 3GPP SA2 Chairman Frank Mademann, "System architecture milestone of 5G Phase 1 is achieved," *3GPP SA2 Chairman*, 2017. .

[27] C. Mannweiler *et al.*, "5G NORMA: System architecture for programmable & multi-tenant 5G mobile networks," 2017, doi: 10.1109/EuCNC.2017.7980662.

[28] STANDARDS INSTITUTIONS, "GS NFV 002 - V1.2.1 - Network Functions Virtualisation (NFV); Architectural Framework," *Tbd*, vol. 1, pp. 1–21, 2014.

[29] D. Sattar and A. Matrawy, "Optimal Slice Allocation in 5G Core Networks," *IEEE Netw. Lett.*, 2019, doi: 10.1109/lnet.2019.2908351.

[30] M. Dighriri, A. S. D. Alfoudi, G. M. Lee, and T. Baker, "Data Traffic Model in Machine to Machine Communications over 5G Network Slicing," in *Proceedings - 2016 9th International Conference on Developments in eSystems Engineering, DeSE 2016*, 2017, pp. 239–244, doi: 10.1109/DeSE.2016.54.

[31] M. Dighriri, G. M. Lee, T. Baker, and L. J. Moores, "Measuring and Classification of Smart Systems Data Traffic Over 5G Mobile Networks," in *Technology for Smart Futures*, A. B. Dastbaz M., Arabnia H., Ed. Springer, Cham, 2015.

[32] A. Pokhariyal *et al.*, "HARQ aware frequency domain packet scheduler with different degrees of fairness for the UTRAN long term evolution," 2007, doi: 10.1109/VETECS.2007.567.

[33] "Handbook of simulation: Principles, methodology, advances, applications, and practice," *J. Manuf. Syst.*, 2008, doi: 10.1016/s0278-6125(99)90111-5.