Journal of Electrical Engineering and Information Technologies, Vol. **7,** No. 1, pp. 23–31 (2022)
Article 193
Received: Јануарс 21, 2022
Accepted: March 2, 2022
In print: ISSN 2545–4250
On line: ISSN 2545–4269
UDC: 004.75:004.455
https://doi.org/10.51466/JEEIT2271188023s

Original scientific paper

A MOBILE EDGE COMPUTING SERVICES WITH QoS SUPPORT MODEL FOR NEXT GENERATION MOBILE NETWORKS

Tomislav Šuminoski[1], Bojana Veličkovska[2], Toni Janevski[3]

Faculty of Electrical Engineering and Information Technologies,
"Ss. Cyril and Methodius" University in Skopje,
Rugjer Bošković bb, P.O. box 574, 1001 Skopje, Republic of North Macedonia
tomish@feit.ukim.edu.mk

A b s t r a c t: This paper presents a novel overview in intelligent multi-access QoS mobile edge computing (MEC) for beyond 5G networks and services. There are many challenges faced by the expansion of Cloud networks and Mobile networks, which can be solved by providing connectivity at the edge of the network, i.e. with Mobile Edge Computing networks. The MEC improves overall network performance and reduces end-to-end service delay. Also, the improved advanced QoS model including Machine Learning (ML) algorithm within for next generation of mobile networks and services are proposed. The purpose of the ML algorithm is to understand the traffic activity and determine how the traffic schedule should be made. Given a set of machines and a set of jobs, the model should compute the processing schedule that minimizes specified metrics. The proposed model combines the most powerful features of both Cloud and Edge computing, independent from any existing and future Radio Access Technology, leading to possible better performance utility networks, lower service delay with high QoS provisioning for many used multimedia service. Finally, this paper gives an overview of the existing Mobile Edge Computing technologies and several existing use cases. Undoubtedly, MEC with QoS support is an innovative network paradigm going in 6G, which can essentially answer many of the existing Mobile Networks' challenges.

Key words: aggregation; machine learning; mobile edge computing; quality of service

СЕРВИСИ ЗА МОБИЛНИ ПРЕСМЕТКИ НА РАБОТ ОД МРЕЖАТА СО МОДЕЛ ЗА QoS ПОДДРШКА ЗА СЛЕДНА ГЕНЕРАЦИЈА МОБИЛНИ МРЕЖИ

А п с т р а к т: Овој труд претставува нов поглед на интелигентното мобилно пресметување на рабови од мрежата со повеќе пристапи кон квалитетот на услугата, и за сегашните 5G мрежи и услуги и за тие што следат после нив. Постојат многу предизвици со кои се соочуваат мрежите во облак (Cloud) и мобилните мрежи, кои можат да се решат со обезбедување на интелигентно поврзување на работ на мрежата, т.е. со мобилно пресметување на работ од мрежата (англ. Mobile Edge Computing, скратено MEC). MEC ги подобрува вкупните перформанси на мрежата и го намалува доцнењето на услугата од крај до крај. Исто така се предлага подобрен напреден модел на квалитетот на услуга (QoS) со вклучен алгоритам за машинско учење (МУ) за системи и услуги за следната генерација мобилни мрежи. Целта на алгоритмот за МУ е да го разбере редоследот на активностите и да го одреди нивното распределување. Доколку е дадено множество од машини и множество од работи, модулот ќе ја пресмета процесната распределба која ги минимизира специфичните метрики. Предложениот модел ги комбинира најмоќните карактеристики и на Cloud и на Edge-компјутерите, независни од која било постојна и идна радио-пристапна технологија, што доведува до можни комуникациски мрежи со подобрени перформанси, помало доцнење на услугите со високо обезбедување на квалитетот на услугата за многу поллзувани мултимедиски сервиси. Конечно, овој труд дава преглед на постојните технологии за мобилно претсметување на работ на мрежата (Mobile Edge Computing) и на неколку постојни случаи на употреба. Несомнено, MEC со поддршка на QoS е иновативна мрежна парадигма што се случува во 6G, која всушност може да одговори на повеќето предизвици што постојат во мобилните мрежи.

Клучни зборови: агрегација; машинско учење; мобилни пресметки на работ; квалитет на услуга

1. INTRODUCTION

NOW the technology is at the point of interesting era when the peak of 5G implementation is reached, the period of the first 6G researches and the period of implementation of the first pilot 6G algorithms and machine learning (ML) networks are starting.

Above all, the key goal is provisioning high Quality of Service (QoS) support, as well as faster computing features and longer battery life of mobile terminals (MTs). The exponential growth in the amount of traffic carried through mobile networks and cloud computing is followed by a novel research works towards the advanced computing capabilities of the core part of the networks. Recent years there is an increased interest in transferring computing from Clouds towards the network edges or Mobile Edge Computing (MEC).

Back then years, the key trend was to push the computing, control, data storage and processing in the cloud computing (CC) [1], [2]. However, in order to meet the intelligent networking and computing demands in 5G network, the cloud alone encounters too many limitations, such as requirements for reduced latency, high mobility, high scalability and real-time execution.

To help in solving some problems with CC appears the new paradigm called Fog Computing has emerged to overcome these limitations [3], [4]. Fog distributes computing, data processing, and networking services to the edge of the network, closer to end users, or in the MTs. It is an architecture where distributed edge and user devices collaborate with each other and with the clouds to carry out computing, control, networking, and data management tasks [5].

Rather than concentrating data and computation in a small number of large clouds, many fog systems would be deployed at the proximity to end users, or where computing and intelligent networking can best meet user needs. The main idea is to take full advantages of local radio signal processing, cooperative radio resource management, and distributed storing capabilities in edge devices, which can decrease the heavy burden on front haul and avoid large-scale radio signal processing in the centralized baseband unit pool.

Therefore the MEC concepts appeared to resolve all above mentioned challenges [6]. The MEC concepts distribute computing, data processing, and networking services close to the end users, where computing and intelligent networking can best meet user needs. MEC and Fog Computing provide an infrastructure where distributed edge and user devices collaborate with each other, as well as, with the CC centers, in order to carry out computing, control, networking, and data management tasks. Also, there are significant disparities between MEC and CC systems in terms of computing, data storage, distance to end users and end-to-end latency. MEC has the advantages of achieving lower latency due to the shorter distances, saving energy for the mobile devices, supporting context aware computing and enhancing the privacy and security for mobile applications. Despite Fog Computing, both MEC and CC, in the core of their networks are using Network Virtualization, which is a powerful combination of SDN (Software-defined networking) and NFV (Network functions virtualization) infrastructure.

In this paper, the user-centric approach is accepted as a basis for our work on Mobile Edge Computing system model, where the future the MTs would have access to different radio access technologies at the same time and should be able to combine different flows from different technologies using advance QoS algorithms within the Cloud orchestrator for used multimedia services, using vertical multi-homing and multi-streaming performances [7], [8].

The remainder of this article is organized as follows. Section 2 gives an overview and key MEC concepts and MEC requirements. Section 3 presents MEC in 5G. Section 4 provides description of a system model for our proposed framework and model. Finally, the last section conclude the paper and gives some future works.

2. MEC FUNDAMENTALS AND REQUIREMENTS

The tremendous interest and developments of mobile broadband Internet networks, undoubtedly lead to intensive research works towards advanced mobile and cloud computing algorithms and frameworks for high level of QoS provisioning in each core and access network. At the first place, the main motivation for our proposed intelligent multi-access QoS provisioning framework could be found in [7–11]. Device-centric multi-RAT architectures, native support of machine-to-machine communications and smarter devices are part of the main features of 5G [12]. Moreover, our framework and design of a novel MT with Mobile Fog, CC support is a next step from previous works on adaptive QoS provisioning in heterogeneous wireless and mobile

IP networks [10, 11]. Those papers were introducting a framework adaptive QoS provisioning model that provides the best QoS and lower cost for a given multimedia service by using one or more radio access technologies (RATs) at a given time. A key concept that allows highlighting the potential of CC environment is orchestration that aims to coordinate the execution of a set of virtualized services within the same process.

Furthermore we are giving an overview of fundamentals for MEC.

The key idea of MEC is in providing an Internet broadband service environment and cloud-computing capabilities at the edge of the mobile network part, within the RAN and in close proximity to MTs. In the foreseeable future, MEC will open up new markets for different industries and sectors by enabling a wide variety of 5G use cases, e.g., Internet of Things / Internet of Everything, Industry 4.0, Vehicle-to-everything (V2X) communication, smart city, Tactile Internet and etc. According to the ETSI [13], [14] white paper MEC can be characterized by some features, namely on-premises, proximity, lower latency, location awareness, and network context information.

Furthermore, there are several existing use cases for MEC [15]:

- **RAN-Aware Video Optimization**: Video is currently taking half of mobile network traffic and set to exceed 70% of traffic over the next couple of years. Providing throughput guidance information is one of the MEC use cases. The proposed solution is to use MEC technology to inform the video server on the optimal bit rate to use given the radio conditions for a particular stream.
- **Video Analysis Service**: Many recognition type application could benefit from the MEC architecture, mostly by the proximity of the computation that is executed at the edge devices. Whenever some video data needs to be analyzed, it can be sent to the MEC server and only needed data can be sent to the centralized cloud. The system benefits of low latency and avoids the problem of network congestion.
- **Augmented Reality (AR)** and **Virtual Reality (VR) Services**: AR is a live view of a real world environment whose elements are supplemented by computer generated inputs such as sound, video, graphics or other data, A MEC based AR and VR application systems should be able to distinguish the requested contents by correctly analyzing the input data and then transmit back the AR/VR data back to the end user.
- **Enterprise and Campus Networks**: In large enterprise organizations, there is a desire to process users locally rather than backhaul traffic to centralized mobile core just so that it can send the data back again. This could be for services as simple as access to corporate intranet (4k video training to a mass of employees at the same time), or more advanced services such as security policy, location tracking and asset tracking services.
- **IoT Applications for Smart Home, Smart Grid and Smart City**: MEC can be used to process and aggregate the small packets generated by IoT services before they reach the core network. Much of the data generated in a smart building is inherently local and involves D2D communication, so the benefits of this would be from moving the local computing and security, tracking, climate control to the edge servers and process and work with that data closer to the user without significant latency.

The newly emerging applications corresponding to mobile AR, VR, and wearable devices, e.g. smart glasses and watches, are anticipated to be among the most demanding applications over wireless networks so far, but there is still lack of sufficient capacities to execute sophisticated data processing algorithms. To overcome such challenges, the emergence of MEC and 5G techniques would pose the longer battery lifetime, powerful set of computing and storage resources, and low end-to-end latency. Sharing this view, [16] presented Outlet system to explore the available computing resources from user's ambience, e.g., from nearby smart phones, tablets, computers, WLAN APs, to form a MEC platform for executing the offloading tasks from wearable devices. Promising performance achieved by Outlet, e.g., mostly within 97.6% to 99.5% closeness of the optimal performance, has demonstrated the advantage of enabling edge computing technique into wearable IoT systems. Applying MEC on VR devices, [17] presented an effective solution to deliver VR videos over wireless networks minimizing the communication-resource consumption under the delay constraint. This work also demonstrated the interesting trade-offs among communications, computing, and caching. In [18], a novel delivery framework enabling field of views caching and post-processing procedures at the mobile VR device was proposed to save communication band width while meeting low latency requirement. Impressively, an implementation of MEC concepts over Android OS and Unity VR application engine in [19] enabled to reduce more than 90% computation burden, and more than 95% of the VR frame data being transmitted to MTs by letting MEC servers adaptively store the previous results of VR frame rendering of each user and considerably reuse them for others to reduce the computation load.

On the other side, the Tactile Internet (TI) is defined by the ITU as the next evolution of IoT that combines ultra-low latency with extremely high availability, reliability and security. Encompassing human-to-machine and machine-to-machine interaction, Tactile Internet will combine multiple technologies including 5G and MEC, i.e., 5G may be employed for the data transmission with low delay and high reliability while MEC efficiently exploit computing resources close to the end users for better QoE. The applications related to Tactile Internet can be automation, robotics, tele-presence, tele-operation, AR, VR. The following summarizes the recent works focusing on the technical aspects involving to the MEC implementation in Tactile Internet. An energy-efficient design of fog computing networks will support low service response time of end-users in Tactile Internet applications and efficiently utilize the power of fog nodes. The trade-off between the latency and required power was presented and then extended to fog computing networks leveraging cooperation between fog nodes. We can exploit the MEC systems including cloud, decentralized cloudlets, and neighboring robots equipped with computing resource collaborative nodes for computation offloading in support of a host robot's task execution. MEC based collaborative task execution scheme outperforms the non-collaborative scheme in terms of task response time and energy consumption efficiency. Recently, in [20] designed a hybrid edge caching scheme for Tactile Internet which can reduce latency and achieve better performance in overall energy efficiency than existing ones.

Moreover, the Multi-access Edge Computing will enable new vertical business segments and services for consumers and enterprise customers.

On the other side, there are also many benefits in cooperation of MEC with SDN. The benefits of programmable networks align with the MEC requirements, and the recent form of SDN has the ability to mitigate the barriers that prevent MEC to reach its full potential and achieving high reliability. All data flow management, service orchestration and other management tasks are done by the central SDN controller that is transparent to the end-user. Moreover, MEC will fit into the 5G concepts and what specifications have been developed based on the industry consensus. The 3GPP clarified how to deploy MEC in and seamlessly integrate MEC into 5G, which can be illustrated in [13]. Actually, the architecture comprises two parts: the 5G service-based architecture (SBA) and a MEC reference architecture. The network functions defined in the 5G architecture, and their roles can be briefly summarized as: Access and Mobility Management Function (AMF); Session Management Function (SMF); Network Slice Selection Function (NSSF); Network Repository Function (NRF): supports the discovery of network functions and services; Unified Data Management (UDM); Policy Control Function (PCF); Network Exposure Function (NEF); Authentication Server Function (AUSF); User Plane Function (UPF).

The MEC orchestrator (MECO) is the core component of the MEC system level, which maintains information on deployed MEC hosts (i.e., servers), available resources, MEC services, and topology of the entire MEC system. The MECO is also responsible for selecting of MEC hosts for application instantiation, onboarding of application packages, triggering application relocation, and triggering application instantiation and termination. The host level management consists of the MEC platform manager and the virtualization infrastructure manager (VIM). The MEC platform manager carries out the duties on managing the life cycle of applications, providing element management functions, and controlling the application rules and requirements. The MEC platform manager also processes fault reports and performance measurements received from the VIM. Meanwhile, the VIM is in charge of allocating virtualized resources, preparing the virtualization infrastructure to run software images, provisioning MEC services, and monitoring application faults and performance.

Moreover, we summarized the MEC requirements from user-side:

- **Low latency:** The sensor data can be collected in near real-time by an MEC server. The closer to the edge the server is located, the lower the latency. For services such as drone control or image recognition, edge servers can be located in very close proximity to the device, meeting user's low latency requirements whilst offering the same level of control as centralized services.

- **Network independence:** IoT services do not care for the delivery mechanism of data. The end-users just need the data to be made available by the most effective means possible. This in many cases will be mobile and wireless networks, but in some scenarios such as smart buildings or smart homes, WLAN or local mesh networking may be the most effective mechanism of collecting data to ensure latency and other collection requirements can be met.
- **Security and privacy:** Data security and data privacy are requirements that must be met with the same rigour at the edge as they are at the core. However, data security at the edge of the access networks has different challenges to the core, not least that data is spread across many more locations at the same time, and customers need control of security at each of these points. Additionally, physical security requirements are more prevalent at the MEC, as servers may be located outside of highly secure operator data centres. Both end-users and cloud providers will need security assurances backed up by tools such as constant monitoring of edge nodes, reliable access logs and appropriate authentication and others. The integration of equipment running proprietary or legacy protocols with other data sets from modern equipment. A secure edge can offer a point of protocol translation, allowing data from multiple sources to be combined and analyzed at the edge.
- **Flexibility in future enhancements:** Additional IoT sensors can be added and managed at the edge as requirements change. Sensors such as accelerometers and cameras can be added to equipment, with seamless integration and control at the edge.

In reality, end-users may have little say over the shape of the edge unless they have very specific requirements that will dictate dedicated and bespoke edge resources to support their operations.

Most services will utilize shared infrastructure that is available across a range of use cases. There are several core requirements that need to be met regardless of the class of infrastructure deployed at the edge

On the other side, looking from the Telcos-side, there are multiple business models that can be applied to capture the emerging commercial opportunities in MEC. Furthermore, there are some key MEC challenges that need to be fulfilled:

- **Security challenges**: As cloud becomes distributed, this presents a potential security challenge. Stricter rules and regulations surrounding data protection and sovereignty may reduce the perceived attractiveness of MEC, as the security challenges would need to be overcome.
- **Commercialization**: It is unclear which telco MEC use cases will deliver significant value to telcos and their customers. Telcos need to identify a segment in value chain where they want to play in, based on their existing capabilities and propositions.
- **Operationalization**: Different parts of telco organization perceive edge computing in a different manner, for both internal/external use cases as well as wider efforts related to NFV/SDN and 5G core. Some telcos are using edge capabilities internally as a way of supporting 5G rollouts, while others perceive it as a by-product of 5G, for which distributed compute needs to be deployed anyway.

Finally, the MEC host comprises a MEC platform and a virtualization infrastructure. New functional enablers were defined in [21] to integrate MEC into the 5G. The MEC turns this paradigm on its head by collecting, storing and processing data closer to its source.

3. MEC IN 5G AND BEYOND

For supporting the large scale of network connections, 5G uses the tremendous computation and storage resources from remote datacenter and utilizes NFV and SDN technologies to virtualize the network resources for achieving an end-to-end optimized system for service provisioning. However, one issue that 5G network suffers from is the high latency, which could not meet the requirements of the emerging IoT applications [22]. For solving this issue, MEC can be deployed in 5G gNB to eliminate the latency in the core network transmission, enhancing the service provisioning capability of 5G network for small-scale and ultra-low-latency services and application scenarios.

To emphasized, there are many benefits of employing MEC into IoT systems, including but not limited to, lowering the amount of traffic passing through the infrastructure and reducing the latency for applications and services.

Among these, the most significant is the low latency introduced by MEC which is suitable for 5G Tactile Internet applications requiring round-trip latency in the millisecond range. MEC technologies are envisioned to work as gateways placed at the middle layer of IoT architecture which can aggregate and process the small data packets generated by IoT services and provide some additional special edge functions before they reach the core network; hence, the end-to-end latency can be reduced.

In addition, based on the context and platforms of MEC, ML (Machine Learning) on the edge can gain the huge benefit to realize distributed IoT applications [23, 24] and intelligent system management, which is now considered as a part of beyond 5G standardization. Since ML relies on building models from input data in order to automate the decision-making process and allows saving and storing the resulting models at the end of the learning process, it is the perfect tool to use when the knowledge, which the models have, has to be reapplied on new data in a widely-distributed manner. An additional benefit is that the requirements of the model can be easily met, since it requires: the new data be processed in the same was as the data used for the training of the model, and the memory and processing power which is dependent on the purpose of the model itself (which, for “simpler” ML tasks can be a minimal requirement). Inversely, IoT also energizes MEC with mutual advantages. In particular, IoT expands MEC servies to all types of smart objects ranging from sensors and actuators to smart vehicles. Integrating MEC capabilities to the IoT systems come with an assurance of better performance in terms of quality of service and ease of implementation.

Finally, the existing 5G use cases for MEC are requiring: low latency, local data processing and local data storage.

4. SYSTEM MODEL FOR MEC WITH QoS SUPPORT

The Figure 1 depicts the system architecture and usage scenario for our proposed intelligent multi-access QoS mobile edge computing framework for next generation mobile network and services, using heterogeneous environment orchestrated services. First, the main characteristics of our proposed edge MT with incorporated advanced QoS user-centric ML module (AQUML) with vertical multi-homing and multi-streaming features are illustrated in [11], and with ML being an essential tool for data intelligence it guarantees improvement of services [25].

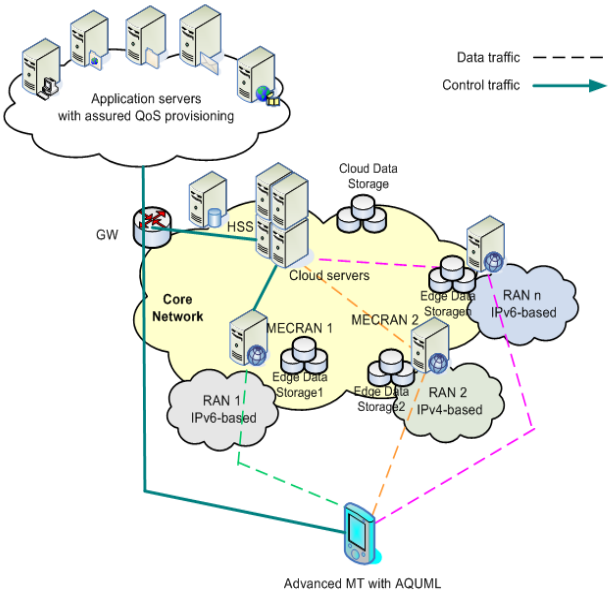

Fig. 1. System Model for MEC model for future generation mobile networks

The Cloud server placed in the core part of the network is in constant communication with the MEC Radio Access Network (MECRAN) servers in which are placed the multimedia broadband orchestrators which orchestrates the MECs. Moreover, each MT used in the above scenario is multi-RAT node, with several (n) RAT interfaces. The advanced QoS routing algorithm with ML is set within the AQUML module on IP layer in both MT as edge device in one side, and the MECRAN server in the another side. Also in the edge devices (MTs) there is an orchestrator agent which collect the QoS parameters of interest and sends to the orchestrator manager in the MECRAN. Here, the QoS parameters of interests are: processing speed, service price per RAT, MT velocity, MT battery level, MT delay (from MT to MECRAN), detected signals strength, response time, availability, number of jobs, maintainability etc. If the MECRAN orchestrated-service manager is overloaded, he can send part of his work for processing to the local edge agents in the heterogeneous environment or to the Cloud server in the core or to global Cloud Server Farm. However, part of the optimizations in selecting the most appropriate RAT for a given services/service are done in the service orchestrator agent, but mostly all those optimizations are done in the service orchestrator manager in MECRAN, by starting the AQUML module with ML within.

The purpose of the ML algorithm is to understand the traffic activity and determine how the distribution of service flows should be made. This can be looked at as a classification problem or as a scheduling problem, and in both cases ML can provide quality algorithms and ML-enabled Management System for investigating traffic features in determining an outcome [26]. ML has experimentally established algorithms to address data divisions into recursive branches according to feature values, as are Decision Tree (where the branching ends with a predicted outcome), Random Forest (where there is an ensemble of decision trees making the predictions, in so that each tree has an independent result, but the final decision is a contribution from all the outcomes), and XGBoost (which is an implementation of gradient boosted decision trees, and is one of the most popular and best performing ML algorithms). Mathematically scheduling decisions in a multi-categorical services distribution, so a sequence of actions can be optimized based on a specific metric, can be done with Markov Decision Process (MDPs) [27]. On a transport layer, the most suitable protocols which are used here are: Stream Control Transmission Protocol (SCTP), Datagram Congestion Control Protocol (DCCP). Also, on the other end of the connection must be installed SCTP/DCCP on its transport layer in order to have successfully established SCTP/DCCP association.

Furthermore, one of the advantages of our framework is the following: it is defined independently from different RATs, implemented on IP Network Layer. This advanced MT with AQUML is using Multi-RAT interfaces and is able to provide intelligent QoS management and routing over variety of heterogeneous RATs at the same time. Moreover, the services orchestrator with help of AQUML module is able to combine simultaneously several different traffic flows from different multimedia services transmitted over the same or different RAT channels, ML to understand traffic activity, and therefore, improve upon existing services, as well as optimize the chosen traffic flows accordingly to resources. The final aim is in achieving higher throughput, higher access probability ratio and optimally (regarding the resources) using the heterogeneous RAT resources. Our proposed architecture incorporates the above model, which consists of several levels: Regional Service Orchestrator (RSO), Domain Service Orchestrator (DSO), and Global Service Orchestrator (GSO). The RSOs are located at the edges of the MECRANs and at the advanced MTs, enabling semi-autonomous operation of the different regions. Due to the advanced MT proximity, this provides quicker distribution of the load, lower latency and higher scalability. The DSOs are located in the cloud computing data centers, in the core networks. Each DSO is responsible for their domain/s and supervises the RSOs below. Like that global mechanisms are provided in order to enable intra-domain cooperation between different regions.

At the top of the architecture are located the GSOs, which allow a fruitful interaction between different cloud and fog domains. The GSO enables the management functionality between different cloud and fog domains and, similarly to the DSOs, it should be properly adapted to operate in a global Cloud environment. GSO communicate with other GSOs and like that global mechanisms are provided that enable cooperation among different cloud computing domains (e.g. under the administration of different authorities). These global mechanisms also enable the creation of a Multi-Domain Mobile Cloud Environment able to support service ubiquity.

The process of establishing a tunnel to the Cloud server in the core, for routing based on the QoS policies and QoS requirements per service; are carried out immediately after the establishment of peer-to-peer connection between the MT services orchestrator agent with MEC features and MEC-RAN server on the other side. The MT and MEC-RAN/Cloud server with vertical multi-homing and multi-streaming features and service orchestrator, with ML within, are able to handle simultaneously multiple radio network connections and speed up the transfer of the multimedia services. Moreover, by transmitting each object of each service in a separate stream, the highest level of satisfied end-users is achieved. In that way, by using our proposed MT with AQUML with ML algorithm within, instead of creating a separate connection for each object as in TCP, makes use of network capacity aggregation, multi-streaming and multi-homing feature to speed up the transfer of the target multimedia service over separate streams over different RATs. So, all mobile broadband services are going over MECRAN and MEC agent in the user's MT (in the downstream direction) and vice versa (in the upstream direction). Also, in comparison with all related works, we must to emphasize that our advanced QoS framework for mobile broadband with MEC and ML is implemented on IP level in the Cloud servers, MECRAN servers and in the edge (MT) sides.

Here we are emphasizing two important QoS parameters, when using AQUML: The end-to-end service delay and the quantity of energy saved (QES) during MEC.

The end-to-end service delay is given with the equation:

$$d_{total} = d_{claud} + d_{MEC} + 2(d_{prop} + d_{core}), \quad (1)$$

where d_{total} is sum of the two times of propagation delay d_{prop} (in wireless part), d_{cloud} delay for cloud processing, d_{MEC} delay for MEC processing, and two times delay in the core part of the network d_{core}. We can optimized the end-to-end service delay, and achieved minimal values for it, by processing most of the services in MEC servers (MECRANs). Those, the processing time in the cloud and the delay in the core, will tend to be zero, so d_{total} will tend to its minimal value.

Moreover, the QES in $\mathrm{MECRAN_i}$ can be calculated as:

$$QES_{MERCAN_i} = \frac{I}{M}\left(E_c - \frac{E_i}{\frac{R_{MEC}}{R_{MT}}}\right) - E_{TR}\left(\frac{U}{B}\right), \quad (2)$$

where I is number of instructions per calculation (assuming that the MT and the MEC server have equal number of instructions), the M is the processing speed of the MT. The energy consumption in the MT during the idle, computing and transmitting states are given with E_i, E_C and E_{TR}, respectively. The processing speed of the MECRAN server is R_{MEC}, and the R_{MT} is processing speed of the MT. The U is the number of bytes which should be transferred to the MECRAN server for processing, and the B is the bandwidth of the i-th radio access network.

The ML within AQUML algorithm will take into consideration end-to-end service delays and QES values for each RAN and for each traffic flow (stream), and based on those training values will determine which traffic flow to go over which RAN.

5. CONCLUSION

In this paper overviewed MEC essence, provides existing use cases of MEC, and proposes a novel beyond 5G framework for MEC for mobile broadband orchestrated services in heterogeneous RATs. According to the analysis, our proposed framework with MEC orchestrated services is expected to perform fairly well under a variety of network conditions and optimally utilized the resources due to the application of the ML algorithm and MEC processing. The benefits of the ML algorithm are manifold, in that edge ML solves security concerns in sending and storing personal user information on the cloud, while at the same time carrying improvement to the QoS for the user timewise. Additionally, edge ML reduces the demands made to the cloud networks by distributing the needed services in real-time. In that manner, efficient and QoS-based usage of available mobile resources, plus efficient MEC orchestrated services performances are most essential for provision of seamless mobile broadband Internet services.

Undoubtedly, MEC is an innovative network paradigm to cater for the unprecedented growth of computation demands and the ever-increasing computation quality of user experience requirements. It aims at enabling cloud computing capabilities and telecommunication services in close proximity to end users, by pushing abundant computation and storage resources towards the network edges. The direct interaction between MTs and edge servers through wireless and mobile communications brings the possibility of supporting applications with ultra-low latency requirement, prolonging device battery lives, energy saving and facilitating highly efficient network operations.

In the future we are planning to simulate our proposed model and to do proof of concept simulation and analysis using ML. We are strongly confident that the proposed model combines the most powerful features of both Cloud and Edge computing, independent from any existing and future Radio Access Technology, leading to better performance utility networks with high QoS provisioning.

REFERENCES

[1] Armbrust, M. et al.: A view of cloud computing, *Communications of the ACM*, Vol. **53**, No. 4, pp.50–58 (April 2010).

[2] Zhang, S., Zhang, S., Chen, X., Huo, X.: Cloud computing research and development trend, *Proceedings of the Second IEEE International Conference on Future Networks (ICFN 2010)*, Sanya, Hainan, 2010, pp. 93–97.

[3] Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the Internet of Things, *Proceedings of the First Edition of the ACM SIGCOMM Workshop on Mobile Cloud Computing (MCC 2012)*, Helsinki, Finland, 2012, pp. 13–16.

[4] Mouradian, C., Naboulsi, D., Yangui, S., Glitho, R. H., Morrow, M. J., Polakos, P. A.: A Comprehensive survey on fog computing: State-of-the-art and Research Challenges, *IEEE Communications Surveys and Tutorials*, Vol. **20**, Issue 1, pp. 1–51 (2017). DOI: 10.1109/COMST.2017.2771153.

[5] Liu, Y., Fieldsend, J. E., Min, G.: A Framework of fog computing: architecture, challenges and optimization, *IEEE Access*, Vol. **5**, pp. 25445–25454 (2017).

[6] Yu, Y.: Mobile edge computing towards 5G: vision, recent progress, and open challenges, *China Communications*, Vol. 13, Supplement 2, pp. 89–99 (2016). DOI: 10.1109/CC.2016.783346

[7] Recommendation ITU-T Y.2052: Framework of multi-homing in IPv6-based NGN (02/2008).

[8] Recommendation ITU-T Y.2056: Framework of vertical multihoming in IPv6-based Next Generation Networks (08/2011).

[9] Parkvall, S., Furuskog, J., Kishiyama, Y., Harada, A., Nakamura, T., Naucler, P., Halvarsson, B.: 5G wireless access – trial concept and results, *IEEE Global Communications Conference* (*GLOBECOM*), 2015, pp. 1–6.

[10] Chang, B. J., Chang, W. T.: Cost-reward-based carrier aggregation with differentiating network slicing for optimizing radio RB allocation in 5G new radio network. *IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference* (*IEMCON*), 2019, pp. 0813–0819.

[11] Shuminoski, T., Janevski, T.: 5G mobile terminals with advanced QoS based user-centric aggregation (AQUA) for heterogeneous wireless and mobile network, *Wireless Networks*, **22** (5), pp. 1553–1570, 2016.

[12] Boyd Bangerter, Talwar, S., Arefi, R., Stewart, K.: Networks and devices for the 5G era, *IEEE Communications Magazine*, Vol. **52**, No. 2, pp. 90–96 (2014). DOI: 10.1109/MCOM.2014.6736748

[13] Kekki, S. et al.: MEC in 5G networks, *ETSI White Paper*, No. 28, pp. 1–28, Jun. 2018. ISBN No. 979-10-92620-22-1.

[14] Taleb, T., Dutta, S., Ksentini, A., Iqbal, M., Flinck. H.: Mobile edge computing potential in making cities smarter, *IEEE Communications Magazine*, Vol. **55**, No. 3, pp. 38–43 (Mar. 2017).

[15] Pham, Q., et al.: A Survey of multi-access edge computing in 5G and beyond: fundamentals, technology integration, and state-of-the-art, *IEEE Access*, Vol. **8**, pp. 116974-117017, 2020.

[16] Tao, L., Li, Z., Wu, L.: Outlet: Outsourcing wearable computing to the ambient mobile computing edge, *IEEE Access*, Vol. **6**, pp. 18408–18419, Mar. 2018.

[17] Yang, X., Chen, Z., Li, K., Sun, Y., Liu, N., Xie, W., Zhao, Y.: Communication-constrained mobile edge computing systems for wireless virtual reality: Scheduling and tradeoff, *IEEE Access*, Vol. **6**, pp. 16665–16677 (0Mar. 2018).

[18] Sun, Y., et al.: Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff, *IEEE Transactions on Communications*, Vol. **67**, No. 11, pp. 7573–7586 (Nov. 2019).

[19] Li, Y., Gao, W.: MUVR: Supporting multi-user mobile virtual reality with resource constrained edge cloud, *2018 IEEE/ACM Symposium on Edge Computing* (*SEC*), Seattle, WA, USA, Oct. 2018. pp. 1–16, DOI: 10.1109/SEC.2018.00008.

[20] Xu, J., Ota, K., Dong, M.: Energy efficient hybrid edge caching scheme for tactile internet in 5G *IEEE Transactions on Green Communications and Networking*, Vol. **3**, No. 2, pp. 483–493 (Jun. 2019).

[21] "3GPP technical specification group services and system aspects; system architecture for the 5G system," Jun. 2019, 3GPP TS 23.501 v16.1.0.

[22] Khan, U. Y., Soomro, T. R.: Applications of IoT: mobile edge computing perspectives, *12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, 2018, pp. 1–7, 2018.

[23] Yassine, A., Singh, S., Hossain, M. S., Muhammad, G.: IoT big data analytics for smart homes with fog and cloud computing, *Future Generation Computer Systems*, Vol. **91**, pp. 563–573 (Feb. 2019).

[24] Rahman, M. A., Rashid, M. M., Hossain, M. S., Hassanain, E., Alhamid, M. F., Guizani, M.: Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city, *IEEE Access*, Vol. **7**, pp. 18611–18621 (Jan. 2019).

[25] Bin Qian, et al. Orchestrating the Development Lifecycleof Machine Learning-Based IoTApplications: A Taxonomy and Survey", available: https://arxiv.org/pdf/1910.05433.pdf

[26] Archana, T., Deepika, V., Arjun Kumar, A., Ramachandran, M., Sivalingam, K. M., Babu Narayanan Koonampilli J.: CygNet MaSoN: Analytics and Machine Learning Enabled Management System for 5G Networks, *International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 2021, pp. 94–96.

[27] Krishnamurthy, V.:Algorithms for optimal scheduling and management of hidden Markov model sensors, *IEEE Transactions on Signal Processing*, Vol. **50**, No. 6, pp. 1382–1397 (June 2002), DOI: 10.1109/TSP.2002.1003062.