Journal of Electrical Engineering and Information Technologies, Vol. **7,** No. 1, pp. 43–52 (2022)
Article 195
Received: May 15, 2022
Accepted: June 2, 2022
In print: ISSN 2545–4250
On line: ISSN 2545–4269
UDC: 004.8:621.39
https://doi.org/10.51466/JEEIT2271195043k

Original scientific paper

QUALITY OF ARTIFICIALLY INTELLIGENTLY ORCHESTRATED SERVICES IN 6G MOBILE NETWORKS

Stojan Kitanov[1], Vladimir Nikolikj[2], Toni Janevski[3]

[1]*Faculty of Informatics, Mother Teresa University, Skopje, Republic of North Macedonia*
[2]*Faculty of Informatics, European University, Skopje, Republic of North Macedonia*
[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
tonij@feit.ukim.edu.mk

A b s t r a c t: With the introduction of 5G mobile and wireless networks, significant improvements have been achieved in terms of latency, data speeds, spectral efficiency, mobility, as well as, the number of connected devices. This makes the emergence of a true digital society a reality. However, although 5G networks offer a wide range of applications and services, they are still unable to meet the demands of rapidly increasing data traffic demands. Therefore, at this time the main research and development activities are focused on the next 6G mobile and wireless networks, which are expected to be commercially available around 2030. In this direction, this paper evaluates the performance quality of artificially intelligently orchestrated services in 6G mobile networks in terms of latency delay, user data throughput and energy efficiency.

Key words;; 5G; 6G; artificial intelligence (AI); Internet of Things (IoT); mobile networks; mobile technology

КВАЛИТЕТ НА ВЕШТАЧКОИНТЕЛИГЕНТНИ ОРКЕСТРИРАНИ СЕРВИСИ КАЈ МОБИЛНИТЕ 6G МРЕЖИ

А п с т р а к т: Со воведувањето на мобилните и безжични 5G-мрежи се постигнати значителни подобрувања во поглед на латентноста, податочните брзини, спектралната ефикасност, мобилноста и бројот на поврзани уреди. Со тоа појавата на вистинско дигитално општество станува реалност. Сепак, иако 5G-мрежите нудат широк спектар на апликации и сервиси, тие не се во можност да ги задоволат барањата на рапидно зголемениот проток на податочни побарувања. Затоа во овој момент главните истражувачки и развојни активности се насочени кон идните мобилни и безжични 6G-мрежи, кои се очекува да бидат комерцијално достапни околу 2030 година. Во таа насока, овој труд прави евалуација на перформансите на квалитетот на вештачкоинтелигентните оркестрирани сервиси кај мобилните 6G-мрежи во поглед на доцнењето, корисничкиот проток на податоци и енергетската ефикасност.

Клучни зборови: 5G; 6G; вештачка интелигенција (ВИ); Интернет на нештата (IoT); мобилни мрежи; мобилна технологија

1. INTRODUCTION

5G mobile and wireless networks are already commercially available all over the world and they have introduced the veritable digital society. Compared to the legacy networks they demonstrated much better improvements in terms of data rates, spectral efficiency, delay, mobility, handover, energy efficiency, as well as in the number of connected smart mobile devices [1–4].

However, the enormous data volumes generated by the users and by Machine to Machine (M2M) traffic within the Internet of Things (IoT) world, are expected to further grow in exponential manner. The International Telecommunication Union (ITU) predicts that the overall mobile data traffic would reach about 607 exabytes per month in 2025 and even more than 5,000 exabytes per month in 2030 [5–6].

Therefore, 5G network may not be able to meet the increased data traffic demands of the existing and new applications. For example, the holographic communication would require a data rate up to terabits per second (Tb/s), which is three times higher than the 5G's data rate, as well as, a massive low latency (hundreds of microseconds), which is three time less than 5G's latency [7–9]. In addition, due to the ever-increasing growth of the deployment of Internet of Things (IoT) and future Internet of Everything (IoE) devices, it is necessary further to increase the connection density and coverage of 5G enabled IoT networks [4, 10]. In addition, the future mobile networks are expected to be ultra-large-scale, highly dynamic, and incredibly complex system. Furthermore, the manual optimization and configuration tasks is no longer applicable in ultra-large-scale, highly dynamic, and incredibly complex system in the next generation mobile networks [11 – 14]. Finally, the new services such as extended reality (XR), telemedicine systems, mind-machine interface (MMI), and flying cars would demand high transmission rates, high reliability, and low latency, which significantly exceeds the original goals of the 5G networks [15 – 17].

Therefore, many 6G network research initiatives have gained significant attention in both academy and industry. 6G is the next generation of mobile and wireless networks that would allow subscribers to communicate with one another everywhere with a high-speed data rate speed due to the large THz bandwidth, and the artificial intelligence (AI).

The main driving force in designing and optimizing 6G architectures, protocols, and operations is expected to be AI, which would transform the wireless evolution from "connected things" to "connected intelligence everywhere" [18].

In this direction, this paper evaluates the performance quality of artificially intelligently orchestrated services in 6G mobile networks in terms of latency delay, user data throughput and energy efficiency. The analysis is very high-level and indicative without any practical implication for wireless networks. Only the AI/ML for 6G network is considered, and the already envisioned concepts such as network densification, cell-free MIMO etc., are not discussed.

The remaining of the paper is organized in the following manner. Section 2 provides an overview of 6G network. Section 3 proposes an AI network model for 6G network. The 6G network architecture is discussed in Section 4. Section 5 performs an evaluation of Edge AI services in 6G network in terms round trip time latency, user data throughput and energy efficiency. Finally, Section 6 concludes the paper, and provides directions for future work.

2. OVERVIEW OF 6G NETWORK

6G mobile and wireless networks would provide a large coverage that allows subscribers to communicate with one another everywhere with a high data rate speed. In order to address the challenges and the issues that 5G network currently faces, it is necessary in 6G network to be developed and used new operations in shared spectrum bands among the network operators, cooperation strategies in heterogeneous networks, and leasing networking slices on-domain. In addition, 6G network would require higher frequency bands in the terahertz spectrum, a very high and opportunistic data rate in order to support new immersive multimedia applications [19]. The end-to-end delay in 6G network should be less than 1 millisecond, in order augmented reality, telepresence, and other delay sensitive services to be supported. Furthermore, 6G network should provide very high reliability, in order mission and safety-critical applications to be enabled.

A comparison between 5G and 6G network parameters and requirements is given in [20, 21]. All parameters such as traffic capacity, data rate, end-to-end delay, processing delay, spectral and energy efficiency, etc. are expected to be improved several times over the value provided by 5G. A comparison between 5G and of 6G parameter requirements is given on Table 1.

Table 1

A comparison between 5G and 6G parameter requirements

Requirement parameter	5G	6G
End-to-end delay	1 ms	0.1 ms
Traffic capacity	10 Mbit/s/m^2	10 Gbit/s/m^3
Latency	Fair	Slightly annoying
Localization precision	10 cm on 2D space	1 cm on 3D space
User experience	50 Mbit/s everywhere on 2D space	10 Gbit/s everywhere on 3D space
Downlink peak data rate	100 Gbit/s	1 Tbit/s
Uplink peak data rate	50 Gbit/s	~ 1 Tbit/s
Frame error rate (FER) reliability	10^{-5}	10^{-9}
Spectral and energy efficiency compared to today's network	10 up to 100 times in bits/s/Hz/m^2/J	1000 times in bits/s/Hz/m^3/J (volumetric)

According to the 6G network roadmap given in [21] currently the academic and industrial researchers all over the world are working on the vision and the requirements of 6G network. The deployment of the first commercially available 6G network is expected to in 2030.

6G network would provide many new use cases, which cannot be completely supported by 5G [21]. Some of them are holographic telepresence, industrial automation (industry 4.0 transform), e-health, tactile internet, augmented, and virtual reality.

As a result, many new 6G services would appear. Most of the 6G services would a hybrid combination of two 5G services. The services in 6G network would require low latency, high reliability, high data rate, massive connectivity, and full mobility. Some of the possible 6G services are massive URLLC (mURLLC), enhanced mobile broadband URLLC (eURLLC), and massive eMBB (meMBB) [21].

3. ARTIFICIAL INTELLIGENT NETWORK MODEL FOR 6G

6G network would not be able to provide the necessary complex services demanded by the users with the guaranteed QoS and QoE parameters only with the involvement of the humans in performing the necessary configuration and optimization of the network. This can be achieved only with the involvement of **Artificial Intelligence (AI)**, which can reconfigure and optimize the network quickly enough to maximize traffic routing and spectral efficiency to meet the service requirements. AI represents the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.

Therefore, AI is the most influential and recently proposed enabling technology for the 6G network [23]. The conventional approach is to place the AI in the 6G core in all TCP/IP layers. This would introduce descriptive, diagnostic, predictive and prescriptive AI data analytics would analyze the collected historical data to get insights of the network status especially of the PHY, MAC, Network and Transport layer [22–23].

However, if the AI is placed only in the core part of the network, 6G would not be able to deal with the future Internet services and applications. This is because the conventional AI core service orchestration approaches that have been applied to services are not adequate to deal with the forthcoming large-scale and dynamic services and applications, since they cannot effectively cope with reduced latency, high mobility, high scalability, and real-time execution.

Therefore, another promising computing paradigm that recently started to gain enormous interest is the **Edge Intelligence (EI)** or edge AI located at the edge of the network [24–25]. Moreover, big data sources as an enabling technology for learning based solutions have recently represented a significant shift from the cloud data centers to the ever-increasing edge devices, e.g., smartphones and industrial IoT devices. It is evident that these edge devices would push the AI solutions to the edge of the network to exploit the edge big data sources' potential entirely.

In other words, just like cloud computing is distributed to the edge of 5G network, the artificial intelligence would also be distributed to the edge of 6G network. Therefore, an improved QoS and QoE would be guaranteed to the end users in terms of delay, user throughput and energy efficiency.

Following this direction here is proposed a new **Hybrid AI Services Orchestrator** (**HAISO**) model, which would ensure resilience and trustworthiness of open, large scale, dynamic services. To our best knowledge we did not find a similar model to be proposed. The HAISO would be primarily responsible for the composition of service elements available in the edge AI environment, such as, data analytics and data processing into more complex AI services, which could be offered to the end users. For some of the services may include sensing the traffic crowd sensing, or planning the trip. The execution of such services is performed through multiple different components and entities that are spread in a wide area. This would increase the complexity in terms of decision-making process, particularly in the allocation of 6G network resources in order to achieve the QoS/QoE levels desired by the users. In order the execution of the AI services to be coordinated, it is necessary the orchestration mechanisms to be synchronized and combined from different service elements. Like that the QoS/QoE levels of a particular service, such as low latency, high user throughput, improved energy efficiency.

The HAISO would operate in a loosely coupled mode, which would consist with several levels: Edge AI Service Orchestrator (EAISO), Regional

Domain AI Service Orchestrator (RDAISO), and Federated AI Service Orchestrator (FAISO), as it is shown in Figure 1.

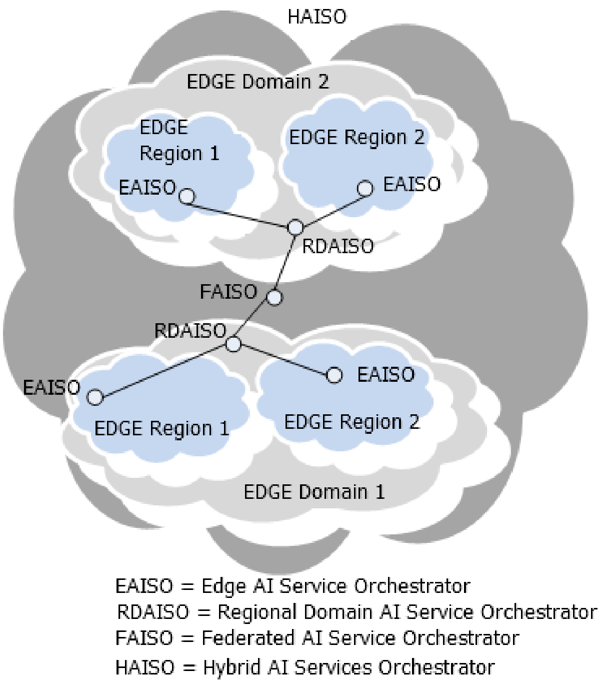

Fig. 1. Hybrid AI services orchestrator (HAISO) model for 6G network

The EAISOs are located at the edge of the network Fog environment and enable semi-autonomous operation in different local Edge Regions. This allows the distribution of the load which provides scalability and much higher proximity to the end users with lower latencies.

The RDAISO is responsible for the and supervision of the EAISOs, within one edge domain. This level supports mechanisms that enable intra-domain cooperation between different local Edge regions.

The FAISO is responsible for the management between different edge domains, and allows a fruitful interaction among RDAISO modules. Such cooperation is enabled through various federation mechanisms implemented in FAISO module. Like that it is created a Multi-Domain AI Environment in 6G which would provide support of service ubiquity.

HAISO model provides flexibility and scalability, and it can be independently implemented in any network technology standard. It would possess its own federation machine learning mechanisms which would be implemented on all entities. In particular, the application of this model could be important for critical usage cases of IoT devices and Tactile Internet that requires 1 ms end-to-end latency in order to provide virtual-reality-type interfaces between humans and machines, and big data analytics that requires real time processing with stringent time requirement that can only be carried out in the fog.

4. OVERVIEW OF ARTIFICIAL INTELLIGENT 6G NETWORK ARCHITECTURE

On the basis of the HAISO model for 6G, the proposed architecture of 6G network (Figure 2). The top two levels are located in 6G core, while the two levels at the bottom are located in 6G radio access network.

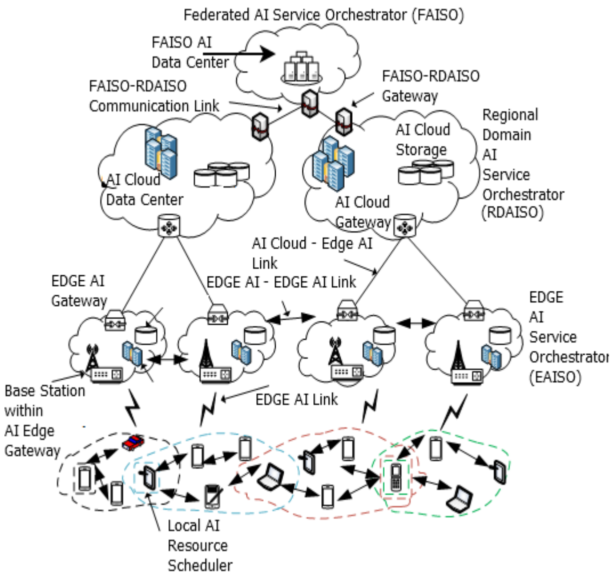

Fig. 2. 6G network architecture

The top level is the Federated AI Service Orchestrator (FAISO) which allows a fruitful interaction between different Regional Domain Service Orchestrators (RDAISOs). The FAISO with its federation mechanisms and with its FAISO Data Center would enable cooperation and exchange of data among communication links. The domains may belong to different entities, and can be administered by different authorities. Like that a multi-domain AI environment would be created, that would support service ubiquity.

The next level contains the AI cloud computing data centers, and AI cloud storage together with the Regional Domain AI Service Orchestrators (RDAISOs). Each RDAISO is responsible for a single edge domain consisted of several edge regions by supervising the correspondent Edge AI Regional Service Orchestrators (EAISOs) below.

The RDAISO would support federation mechanisms to enable intra-domain cooperation and exchange of data between different edge regions. This is enabled through the AI cloud gateways, AI cloud – AI edge communication links and edge gateway. Each AI cloud computing data center contains multiple high-end high performance computing AI cloud servers, and AI cloud storage that are capable of processing and storing an enormous amount of data.

The third level is the edge computing and networking layer, that consists of the 6G Radio Access Network (6G RAN), or 6GRAN Edge AI Service Orchestrators (EAISOs). Each EAISO enables semi-autonomous operation a particular local cloud region. Each EAISO could be interconnected with other EAISOs and each of them is linked to the cloud. This allows the distribution of the load which provides scalability and a much higher proximity to the end users, i.e. lower latencies.

The AI edge computing and networking layer comprises of geo-distributed AI edge devices, deployed at the edge of the network, such as Edge AI Data Center and AI Edge Storage, that are intelligent enough to process, compute, and temporarily store the received information. The AI edge devices directly communicate with the mobile users through edge gateway and single-hop wireless connections using the off-the-shelf wireless interfaces, such as, LTE-A, LTE, WiFi, Bluetooth, or any wireless access network. They can independently provide predefined service applications to mobile users without assistances from the AI located in the 6G core. In addition, the Edge AI servers are connected to the the AI cloud servers in the 6G core in order to leverage the rich functions and application tools of the cloud.

Finally, the lowest level consists of local AI regions, formed by a group of smart devices, such as smartphones, IoT, sensors, which sense multitude of events and transmit the sensed data to the upper fog computing and networking layer, for further processing if necessary. Within this region the smart devices can be either resource users, or resource providers. The devices form so called locally distrbuted peer-to-peer mobile AI cloud, where each device shares the resources with other devices in the same local AI cloud. The devices in each local cloud elect a Local AI Resource Scheduler, that performs management on the resource requests and allocates tasks to the devices in the local AI region or Edge AI Data Center if necessary. The decision about the election of the Local AI Resource Scheduler is done according to the connectivity to the local network, CPU performance battery life time, etc. In addition, one device can be served by several AI edge RANs, and one device can be elected as local AI resource scheduler for several local AI regions.

The locally distributed peer-to-peer mobile AI region has its own strong capacities such as storage space, computational power, online time, and bandwidth. The workload of the application is managed in a distributed fashion without any point of centralization. The lack of centralization provides scalability, while exploitation of user resources reduces the service cost. The local AI has ability to adapt to network failures and dynamically changing network topology with a transient population of devices, while ensuring acceptable connectivity and performance.

At the end it worth to point out that this is high level 6G network architecture. The low level details can be discussed once the 6G network would enter in the phase of standardization.

5. PERFORMANCE EVALUATION OF ARTIFICIAL INTELLIGENT SERVICES IN 6G NETWORK

There are many ways to evaluate the quality of AI orchestrated services in 6G mobile networks. The evaluation can be performed in terms of Round Trip Time (RTT) latency, or end-to-end delay, user data throughput and energy efficiency for the used bits per power consumption for the user device, or vice versa.

5.1. *Round Trip Time latency delay*

Round Trip Time (RTT) latency, or end-to-end delay is the time needed to happen a single transaction. It is the time it takes for the packet of data to travel to and from the source to the destination, and back to the source. The RTT latency between the user equipment towards any AI cloud data cloud is equal to:

$$RTT = RTT_{edge} + RTT_{core-edge}, \qquad (1)$$

where RTT_{edge} represents the latency, i.e. delay at the edge for 6G AI radio access networks, while $RTT_{core-edge}$ is round trip time latency (delay) between the radio access network and the 6G core network.

If there is an edge AI computing node in the radio access network and if information required by the user is contained in the Edge AI computing node, then the RTT latency delay for any user device is equal to: any AI cloud data cloud is equal to:

$$RTT = RTT_{edge}. \quad (2)$$

The total typical value for RTT is around 0.1 ms, and it was already specified in [21]. For the simulation purposes here it is assumed, the end-user device to be served by any of the 5 AI data centers, and the values for the AI entities are randomly generated for both RTT_{edge} and $RTT_{core-edge}$. Here the AI data center 1 is considered to be at the closest proximity to the mobile device, i.e., it is located at the edge of the network, and therefore it has the lowest RTT latency. On the other hand, the AI data center 5 is located in the 6G core, which is at the greatest distance from the mobile device, and because there are several number of hops of transmitting data, therefore it has the highest RTT latency. The user may randomly service from any AI entity.

The simulation results for the RTT latency for different AI cloud data centers in 6G network are given on Figure 3. It can be noticed that RTT latency is increased from the AI data center 1 to the AI data center 5, because the AI cloud data center 1 has the least distance to the mobile user device, and AI cloud data center 5 has the greatest distance to the mobile user device. This means, the RTT latency at the edge of 6G network has significantly lower values than the RTT latency in 6G core.

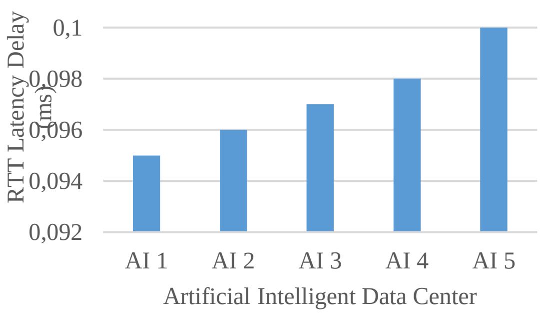

Fig. 3. A comparison of RTT latency delay for different AI data centers in 6G network

5.2. *User data throughput*

Throughput is the quantity of data that can pass from source to destination in a specific time. The user throughput of a particular smart device R for network can be calculated as a ratio between the peak data rate R_{max} of the network and the number of the user devices N, and proportional to some weight coefficient μ:

$$R = \mu \cdot \frac{R_{max}}{N} \quad (3)$$

Here μ is a weight coefficient that models the bottleneck problem for the data that carry services from the AI computing data centers. The weight coefficient μ may receive any positive value between 0.7 and 1, and its value depends how much the AI data center is far away from the radio access network. If the AI data center is closer to the base station of the radio access network, then the coefficient μ has higher value close to 1, and if the cloud is at a greater distance from the base station of the radio access network, then the coefficient μ would have lower value. If the smart mobile device uses a service from edge AI networking intelligence, i.e. in the radio access network, then the weight coefficient μ is equal to 1.

In order to obtain the results for the user throughput the following values are used. The peak data rate 6G network in downlink is equal to be 1 Tbps in both downlink and uplink direction [21]. The number of the users is varied from 100 to 1000. The weight coefficient μ is randomly taken to be 0.85 in downlink direction and 0.75 in uplink direction.

The user throughput results are shown in Figures 4 and 5. It can be noticed that 6G offers much higher user throughput in both downlink and uplink direction by using the Edge AI, rather than the AI in the core part. This means that much higher quantity of data can pass through 6G network by using the Edge AI.

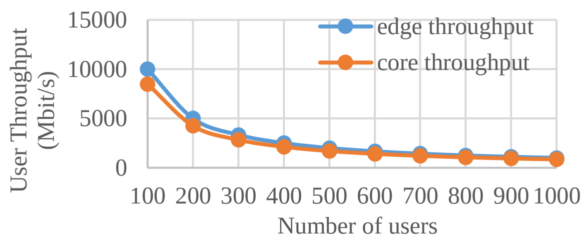

Fig. 4. A comparison of downlink user throughput in 6G network in both core and edge AI environment

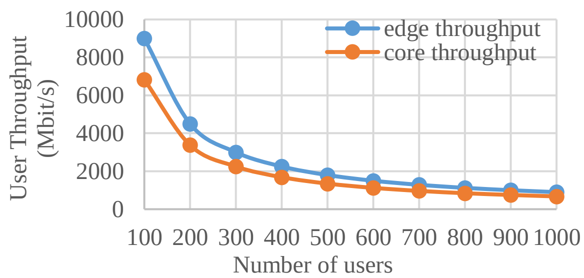

Fig. 5. A comparison of uplink user throughput in 6G network in both core and edge AI environment

In addition, Figures 6 and 7 model the user throughput for 100 user devices, located at various distance from AI data center. At every 2 km different modulation coding schemes, such as probabilistic constellation shaping, and QAM are applied. In other words, if the mobile device is closer to the AI data center than the throughput is higher, because higher modulation coding scheme is applied, and vice versa if the mobile device is getting more distant from the AI data center, then the user throughput is lowered and therefore lower modulation coding scheme is being applied. Here the user throughput is also higher if the service is being used by the edge AI, rather than the AI in the 6G core.

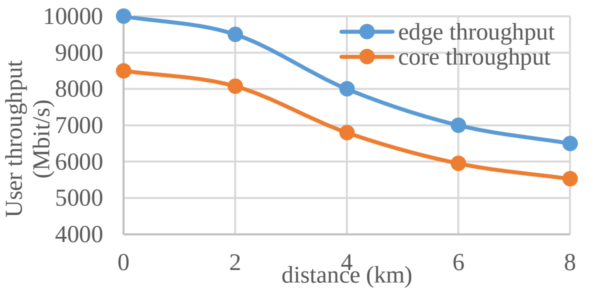

Fig. 6. A comparison of downlink user throughput in 6G-network in both core and edge AI environment, for 100 users

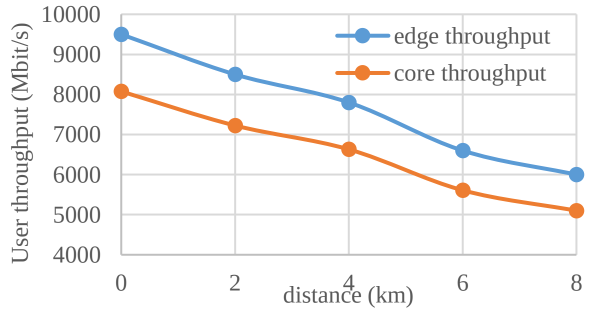

Fig. 7. A comparison of uplink user throughput in 6G network in both core and edge AI environment, for 100 users

5.3. *Energy efficiency*

The reduction of the power and power consumption by the networks and the devices is of vital importance for the economic and ecological sustainability in the industry. The general principle for minimizing of the power consumption at the network and the device should include all technology generations. This principle is recognized as an ecological goal, and is very important for the reduction of operating expenses in the network management [26]. In addition, the reduction of the power consumption would result to a longer battery life, which would contribute to a greater satisfaction at the mobile device users.

One of the possible methods for the reduction of power consumption in 6G mobile networks is by implementing the edge Artificial Intelligence. This can be evaluated through the energy efficiency.

The energy efficiency EE represents the amount of data that can be transfered through the power consumed per user, usually on a single cell, and is the ratio between the user throughput R and the power P:

$$EE = \frac{R}{P}\left[\frac{[\text{bit/s/cell}]}{[\text{J/s/cell}]}\right] = \frac{R}{P}\left[\frac{\text{bit}}{\text{J}}\right] \tag{4}$$

In the relation (4) R is the user throughput which was already discussed in the subsection 5.2. The consumed power P on the other hand, can be expressed through the user throughput R with the following linear equation [27, 28]:

$$P = \alpha R + \beta, \tag{5}$$

where α is the coefficient that gives the power necessary for data transfer (in downlink or uplink direction), and β is a coefficient that represents the idle power [29]. According 6G requirements given in [21] the energy efficiency is about to increase by a factor of 1000, and therefore the typical values for the coefficient α is taken to be 10^{-6} W/Mbit, and the value for the coefficient β is taken to be 10^{-5} W.

One of the possible methods for the reduction of power consumption in 6G mobile networks is by implementing the edge Artificial Intelligence. This can be evaluated through the energy efficiency. The simulation results are presented on Figures 8 and 9. It can be noticed that 6G offers much higher energy efficiency in both downlink and uplink direction by using the Edge AI, rather than the AI in the core part. This means that much higher quantity of data can pass through 6G network by using the Edge AI, for lower power consumption.

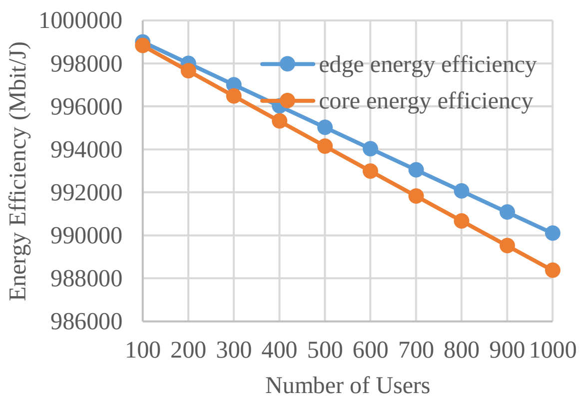

Fig. 8. A comparison of downlink energy efficiency in 6G network in both core and edge AI environment

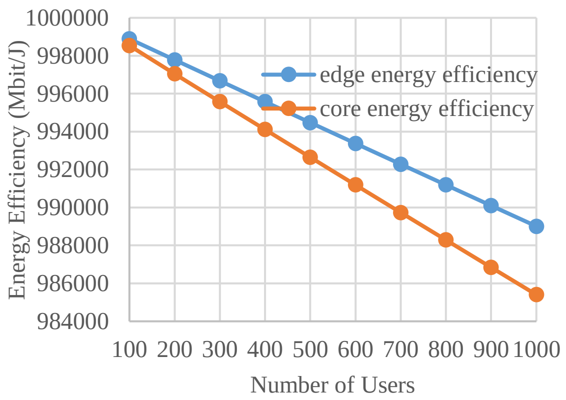

Fig. 9. A comparison of downlink energy efficiency in 6G network in both core and edge AI environment

In addition, Figures 10 and 11 model the energy efficiency for 100 user devices, located at various distance from AI data center. Again, at every 2 km different modulation coding schemes, such as probabilistic constellation shaping, and QAM are applied. In other words, if the mobile device is closer to the AI data center than the energy efficiency is higher, because higher modulation coding scheme is applied, and therefore power consumption is lower and vice versa if the mobile device is getting more distant from the AI data center, then the energy efficiency is lowered because lower modulation coding scheme is being applied, and the power consumption is higher. Here again the energy efficiency is also higher if the service is being used by the edge AI, rather than the AI in the 6G core. Moreover in [30–32] it is stated that a basic goal of 6G communication is to operate battery-free whenever and wherever possible, with a target efficiency of 1 pico-joule per bit. And the simulation results given in Figures 8, 9, 10 and 11 nearly achieve this energy efficiency.

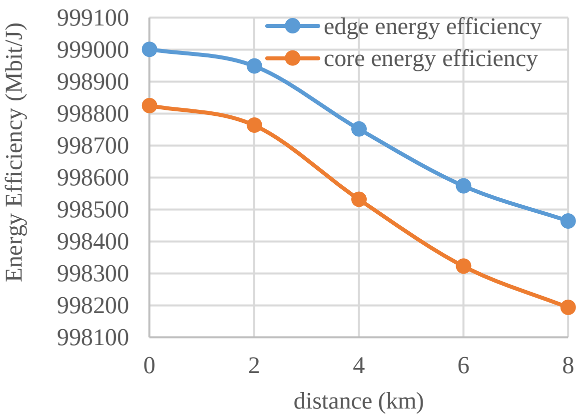

Fig. 10. A comparison of downlink energy efficiency in 6G network in both core and edge AI environment, for 100 users

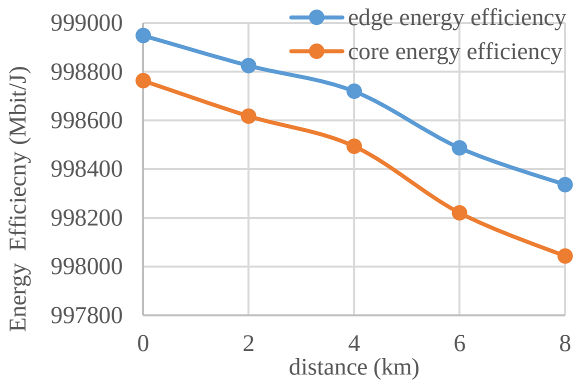

Fig. 11. A comparison of uplink energy efficiency in 6G network in both core and edge AI environment, for 100 users

6. DISCUSSION OF THE RESULTS

The obtained results of RTT latency delay, user throughput and energy efficiency clearly demonstrate the benefits of implementing the edge AI in 6G network. In particular, the big data analytics that requires real time processing and very often has stringent time requirement can only be carried out in the edge AI.

This is essential for critical usage cases of IoT devices and Tactile Internet where 1 ms end-to-end latency is required in the network in order to provide virtual-reality-type interfaces between humans and machines (human-machine interaction and machine-machine interaction). 6G network would provide a more intelligent human-to-machine type of communication for real-time controlling IoT devices [23]. The tactile internet would enable humans and machines to exchange control, touch, and sense data in a real-time manner, which would provide support for haptics interface, as well as, possible visual feedback and remote response behavior that would be used in the industry, e-commerce, and many other possible applications.

Another critical application in 6G network is the holographic telepresence, which would enable users to enrich their traditional audiovisual communication with the sense of touch, while they are in different geographical locations. Holographic telepresence has very strict requirements such as terabits data rate (up to 4 Tb/s), ultra-low latency (less than 1 ms), and reliable communications, which cannot be supported by 5G networks.

In addition, new augmented and virtual reality applications such as haptic technology and virtual meeting room (VMR) which would transmit a large amount of real-time data, and would require very low end-to-end latency, which can be accomplished by implementing edge AI intelligence in 6G.

7. CONCLUSION

Artificial intelligence deployed on the core and edge part of the network would be the driving force in designing and optimizing 6G architectures, protocols, and operations. Our results demonstrated that 6G network offers much higher user throughput in both downlink and uplink direction by using the Edge AI, rather than the AI in the 6G core. The analysis in this paper is very high-level and indicative without any practical implication for wireless networks. Only the AI/ML for 6G network was considered, and the already envisioned concepts such as network densification, cell-free MIMO etc., were not discussed.

AI can significantly optimize the 6G network performance based on its powerful learning ability and strong reasoning ability. Artificial Intelligence is the important characteristic of 6G networks, where with the application of AI, 6G networks can learn to achieve self-configuration, self-optimization, self-organization and self-healing, finally increasing the feasibility level.

This paper proposed a **Hybrid AI Services Orchestrator** (**HAISO**) model, which would ensure resilience and trustworthiness of open, large scale, dynamic services. This model possesses its own federation machine learning mechanisms which would be implemented on all entities. On the basis of that model a high level 6G network architecture was proposed. Finally, it was demonstrated the implementation of edge AI in 6G network. The simulation results demonstrated that lower latency, higher user throughput and higher energy efficiency are achieved by using the artificial intelligence at the network's edge in 6G, rather than in 6G core.

Despite the benefits of AI learning algorithms in terms of learning and recognition ability, they usually require high computational complexity, power consumption, and sufficient computing resources. Hence, collaboration among hardware components and AI learning algorithms need to be advocated, which needs significant research efforts. In addition, great robustness, scalability and flexibility of learning frameworks are crucial aspects for supporting the potential unbounded number of interacting entities and providing high-quality services in 6G networks. Thus, how to design robust, scalable, and flexible learning frameworks for 6G networks is still an open issue.

REFERENCES

[1] Kitanov, S., Monteiro, E., Janevski, T.: 5G and the fog – survey of related technologies and research directions, In: *Proceedings of the 18th Mediterranean IEEE Electrotechnical Conference MELECON 2016*, Limassol, Cyprus, pp. 1–6 (2016),

[2] Nikolikj, V., Janevski, T.: Cost-effectiveness assessment of 5G systems with cooperative radio resource sharing, *Telfor Journal*, Vol. **7**, No. 2, pp. 68–73 (2015).

[3] Nikolikj, V., Janevski, T.: Incremental deployment aspects of beyond 4G and 5G mobile HetNets, *International Journal of Future Generation Communication and Networking*, Vol. **8**, No. 6 (2015).

[4] Jiang, D., Liu, G.: An overview of 5G requirements, In: *5G Mobile Communications*, Springer, Oct. 2017, pp. 3–26.

[5] ITU-R M.2370-0, *IMT traffic estimates for the years 2020 to 2030*, Jul. 2015.

[6] Tariq F., et. al.: A speculative study on 6G, *IEEE Wireless Communications*, Vol. **27**, issue 4, (Aug. 2020).

[7] Clemm, A., Vega, M. T., Ravuri, H. K., Wauters, T., De Turck, F.: Toward truly immersive holographic-type communication: challenges and solutions, *IEEE Communications Magazine*, Vol. **58**, No. 1, pp. 93–99 (Jan. 2020),.

[8] Du, J., Jiang, C., Wang, J., Ren, Y., Debbah, M.: Machine learning for 6G wireless networks: Carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service, *IEEE Vehicular Technology Magazine*, Vol. **15**, No. 4, pp. 122–134 (Sep. 2020),.

[9] Strinati, E. C., Barbarossa, S., Gonzalez-Jimenez, J. L., Ktenas D., Cassiau, N., Maret, L., and Dehos, C.: 6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication, *IEEE Vehicular Technology Magazine*, vol. **14**, No. 3, pp. 42–50 (Aug. 2019).

[10] Swama Priya, R. M., Bhattacharya, S., Praveen Kumar Reddy, M., Siva Rama Krishnan. S., Lakshmanna, K., Kaluri, R., Hussien, A., Reddy Gadekallu, T.: *Journal of Parallel and Distributed Computing*, Vol. **142**, pp 16–26 (Aug. 2020).

[11] Zhang, L., Liang, Y.-C., Niyato, D.: 6G visions: Mobile ultrabroadband, super internet-of-things, and artificial intelligence, *China Communications*, Vol. **16**, No. 8, pp. 1–14 (Aug. 2019).

[12] Nayak, S., Patgiri, R.: 6G: Envisioning the key issues and challenges, *arXiv preprint arXiv:2004.04024*, Jun. 2020.

[13] Zhang, S., Liu, J., Guo, H., Qi, M., Kato, N.: Envisioning device to-device communications in 6G, *IEEE Network*, Vol. **3**4, No. 3, pp. 86–91 (Mar. 2020).

[14] Tang, F., Kawamoto, Y., Kato, N., Liu, J.: Future intelligent and secure vehicular network toward 6G: Machine-learning approaches, *Proceedings of the IEEE*, Vol. **108**, No. 2, pp. 292–307 (Dec. 2019).

[15] Gupta, R., Shukla, A., Tanwar, S.: BATS: A blockchain and AI-empowered drone-assisted telesurgery system towards 6G, *IEEE Transactions on Network Science and Engineering*, Dec. 2020.

[16] Khan, L. U., Yaqoob, I., Imran, M., Han, Z., Hong, C. S.: 6G wireless systems: A vision, architectural elements, and future directions, *IEEE Access*, Vol. **8**, pp. 147029–147044 (Aug. 2020),.

[17] Lv, Z., Qiao, L., You, I.: 6G-enabled network in box for internet of connected vehicles, *IEEE Transactions on Intelligent Transportation Systems*, Nov. 2020.

[18] Letaief, K. B., Chen, W., Shi, Y., Zhang, J., Zhang, Y. A.: The roadmap to 6G: AI empowered wireless networks, *IEEE Communications Magazine*, Vol. **57**, No. 8 (Aug. 2019).

[19] Sliwa, B., Falkenberg, R., Wietfeld, C.: Towards cooperative data rate prediction for future mobile and vehicular 6G networks, In: *Proc. 6G Wireless Summit (6G SUMMIT)*, Levi, Finland, Mar. 2020, pp. 1–5.

[20] Saad, W., Bennis, M., Chen, M.: A vision of 6G wireless systems: applications, trends, technologies, and open research problems, *IEEE Network*, Vol. **34**, No. 8, pp. 134-142 (2020).

[21] Kitanov, S., Petrov, I., Janevski, T.: 6G mobile networks: research trends, challenges and potential solutions, *Journal of Electrical Engineering and Information Technologies*, Vol. **6**, No. 2, pp. 67–77 (2021).

[22] Chen, M., Challita, U., Saad, W., Yin, C., Debbah M.: Artificial neural networks-based machine learning for wireless networks: A tutorial, *IEEE Communications Surveys & Tutorials*, Vol. **21**, No. 4, pp. 3039–3071 Jul. (2019),.

[23] Balali, F., Nouri, J., Nasiri, A., Zhao, T.: Data analytics. In: *Data Intensive Industrial Asset Management*. Springer, 2020, pp. 105–113.

[24] Kibria, M. G., Nguyen, K., Villardi, G. P., Zhao, O., Ishizu, K., Kojima, F.: Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks, *IEEE Access*, Vol. **6**, pp. 32328–32338 (May 2018).

[25] Chen, M., Yang, Z., Saad, W., Yin, C., Poor, H. V., Cui, S.: A joint learning and communications framework for federated learning over wireless networks, *IEEE Transactions on Wireless Communications*, 2020.

[26] Heddeghem, W. V., Lambert, S., Lannoo B., Colle, D., Pickavet, M., Demeester, P.: Trends in worldwide ICT electricity consumption from 2007 to 2012, *Journal of Computer Communications*, Vol. **50**, pp. 64–76 (2014),

[27] Huang, J.; Qian, T.; Gerber, A.; Mao, Z. M.; Sen. S.; Spatscheck, O.: A close examination of performance and power characteristics of 4G LTE Networks. In: *Mobile Systems, Applications and Services (MobiSys 2012) Conference*, Low Wood Bay, Lake District, UK, 2012.

[28] Kitanov, S., Janevski, T.: Energy-efficiency of fog computing and networking services in 5G networks, In: *XVII IEEE International Conference on Smart Technologies EUROCON 2017*, Ohrid, Macedonia, 2017.

[29] Ragona, C., Granelli, F., Fiandrino, C.: Energy-efficient computation offloading for wearable devices and smartphones in mobile cloud computing. In: *IEEE Global Communications Conference (GLOBECOM)*, 2015.

[30] Mahdi, M. N., Ahmad, A. R., Qassim, Q. S., Natiq, H., Subhi, M. A., Mahmoud, M.: From 5G to 6G technology: meets ernergy, Internet-of-things and machine learning: A Survey, *Applied Sciences.*, Vol. **11**, No. 8117 (2021).

[31] Huang, C., Hu S., Alexandropoulos, G. C., Zappone, A., Yuen, C., Zhang, R., Renzo, M. D., Debbah, M.: Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends, *IEEE Wireless Communications*. Vol. **27**, No. 5, pp. 118–125 (Jul. 2020),.

[32] Liu, Q., Wu, J., Xia, P., Zhao, S., Chen, W., Yang Y., Hanzo, L.: Charging unplugged: Will distributed laser charging for mobile wireless power transfer work?, *IEEE Vehicular Technology Magazine*. Vol. **11**, pp. 36–45 (2016).