Fog Computing Service Orchestration Mechanisms for 5G Networks

Stojan Kitanov[1], Toni Janevski[2]

[1] Faculty of Informatics, Mother Teresa University, Republic of Macedonia
[2] Faculty of Electrical Engineering and Information Technologies, Ss Cyril and Methodius University,
Republic of Macedonia
stojan.kitanov@unt.edu.mk, tonij@feit.ukim.edu.mk

Abstract

5G network will enable new future Internet of Services paradigms such as Anything as a Service, where devices, terminals, machines, also smart things and robots will become innovative tools that will produce and will use applications, services and data. However, the emerging applications in the context of the Internet of Everything introduce high mobility, high scalability, real-time, and low latency requirements that raise new challenges on the services being provided to the users. Fortunately, Fog Computing or briefly Fog, which extends Cloud Computing to the edge of the network, with its service orchestration mechanisms offers virtually unlimited dynamic resources for computation, storage and service provision, that will effectively cope with the requirements of the forthcoming services. 5G in the fog computing environment will create opportunities for companies to deploy many new real-time services that cannot be delivered over current mobile and wireless networks. This paper evaluates fog computing service orchestration mechanisms in 5G network in terms of throughput, round trip time latency and the product latency - throughput.

Keywords: 5G, Cloud computing, Fog, Fog computing, Mobile cloud computing

1 Introduction

Mobile and wireless networks have made tremendous growth in the last decade. This growth is due to the support of a wide range of applications and services by the smart mobile devices such as laptops, smartphones, tablets, phablets, etc. This resulted with an increased demand for pervasive mobile broadband services, that demand high data rates, high mobility, low latency, broadband spectrum and high energy consumption, comparable to the fixed broadband Internet [1].

In this direction many global research and industrial initiatives have started to work on the building blocks of the next generation of mobile and wireless networks, usually referred to as 5G – the Fifth Generation [2-3]. 5G will enable the future Internet of Services (IoSs) paradigms such as Anything as a Service (AaaS), where devices, terminals, machines, and also smart things and robots will become innovative tools that will produce and use applications, services and data [4].

However, 5G will have to support huge mobile traffic volumes, 1000 times larger than those today in the order of multiples of gigabits per second [1-2]. It will also have to deal with a proliferation of new and complex applications and services, many of which are unknown today. The emerging applications in the context of the Internet of Everything (IoE) will introduce high mobility, high scalability, real-time, and low latency requirements that raise new challenges on the services being provided to the users. These demands can only be partially fulfilled by existing cloud computing solutions [5].

Fortunately, Fog Computing, or briefly Fog offers virtually unlimited dynamic resources for computation, storage and service provision that will overcome the constraints in the smart mobile devices [6]. Fog computing extends cloud computing and services to the edge of the network. With its service orchestration mechanisms, it provides data, computing, storage, and application services to end-users that can be hosted at the network edge or even end devices such as set-top-boxes or access points. The main features of fog are its proximity to end-users, its dense geographical distribution, and its support for mobility [7]. By implementing fog computing in 5G, it will create opportunities for companies to deploy many new real-time services that cannot be delivered over current mobile and wireless networks.

The move from cloud to fog in 5G brings out several key challenges, including the need for supporting the on-demand orchestration and runtime adaptation of resilient and trustworthy Fog Services. This is essential for the success of the future Internet of Everything (IoE).

This paper is about fog computing service

---

[*] Corresponding Author: Stojan Kitanov; E-mail: stojan.kitanov@unt.edu.mk

DOI: 10.3966/160792642018011901029

orchestration mechanisms as a support for 5G network. The paper is organized as follows. Section 2 provides an overview of Mobile Cloud Computing in 5G network. Section 3 provides an overview of Fog Computing. Section 4 proposes a Hybrid Environment Service Orchestrator model for 5G network in the fog. Section 5 performs evaluation of this model. Finally, Section 6 concludes the paper and provides future work directions.

2 5G Network

The Fifth Generation, or 5G is a name which is used in some research papers and projects to denote the next major phase of mobile telecommunications standards [2, 8]. 5G network will include support of a large number of connected devices and flexible air interfaces, different interworking technologies that are energy efficient, and will possess always on-line capabilities [9]. This will require not only upgrade of existing systems, but also innovation of new protocols and new access technologies altogether. There are three possible migration paths to 5G network:

(1) step-by-step evolutionary path focusing on further enhancements of existing technologies;

(2) a revolutionary path using a brand new innovative technologies; or

(3) a symbiotic integration and convergence of existing or new technologies such as communication, information systems and electronics, multi radio access technologies, computing techniques, device-to-device communications, bands, links, layers, services, multiplexing, etc.

5G network will be an all IP network that will provide a continued evolution and optimization of the system concept in order to provide a competitive edge in terms of both performance and cost [2, 8-9]. This will lessen burden on aggregation point and traffic will directly move from base station to media gateways. 5G network architecture will consist of several levels as it is shown in Figure 1 [10].

At the top level of the architecture are the innovative service and content providers that will accommodate 5G requirements and will provide new ubiquitous and pervasive user experience. New applications and services such as augmented and virtual reality, hologram, mobile ultra-high definition will be provided.

The next level is the enabling software-based core platform. At this level flexible a reliable flat IP architecture will converge into different technologies to form a single 5G core, where a range of intelligent complex telecommunication network functions can be efficiently implemented. All network operators will be connected to one single network core with massive capacity that will eliminate all interconnecting charges and complexities, in which right now the network operator is facing, and will reduce the number of network entities in an end-to-end connection, thus reducing latency significantly.

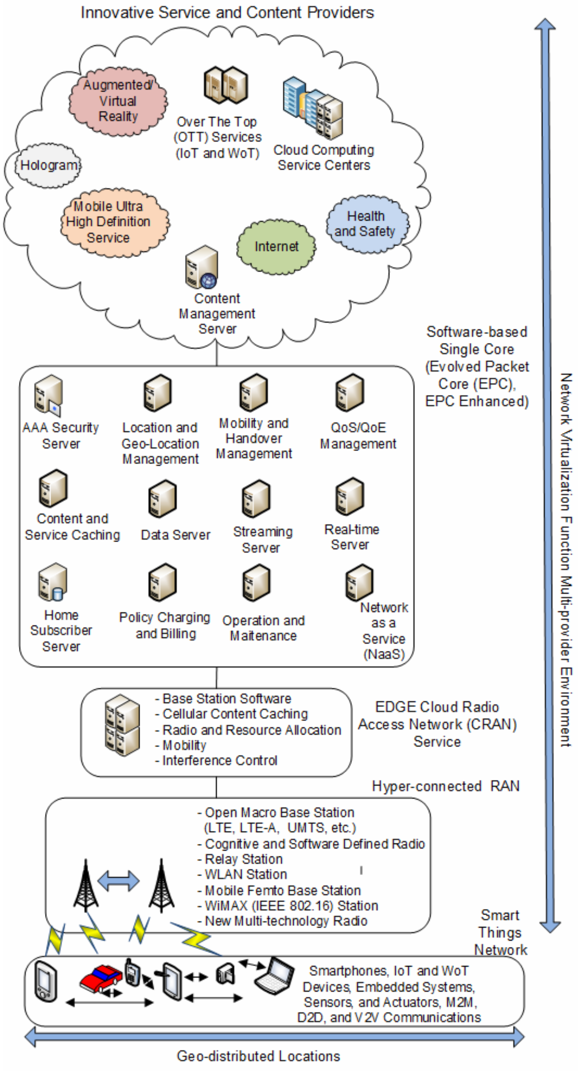

Figure 1. 5G network architecture

The third level is the Radio Access Network (RAN) that consists of EDGE Cloud RAN (CRAN) service and Hyper-connected RAN. EDGE CRAN moves the cloud from the innovative service content provider to the radio access network. By placing storage and computing resources at, or close to, the cell site, operators can improve response times, making services feel “snappier” and, uniquely, more responsive to prevailing radio conditions. Hyper-connected Radio Access Network (RAN) infrastructure acts as a data pipe supporting the massive and ultra-high speed connectivity over ultra-dense networks (UDN) of wireless access with heterogeneous cells arrangement. Cognitive radio and software defined networking together with existing, or new modulation and transmission methods can be used for deploying new applications and services that will improve the utilization of the congested RF spectrum. The RAN infrastructure will be responsible for cellular content caching, radio and resource allocation, mobility, interference control, etc.

Finally, the lowest level of the 5G network architecture consists of a smart things network with different smart mobile, Internet of Things (IoT) and Web of Things (WoT) end user devices. This includes various devices from laptops, smartphones, tablets, phablets, and wearable devices to connectivity embedded devices and sensors in cars, trucks, bikes, etc.

The modules in 5G will be defined primarily by the service requirements, and the following technologies: nanotechnology, quantum cryptography, and (mobile) cloud computing. 5G QoS requirements and mobile cloud computing in 5G are described in the following subsections.

2.1 5G QoS Requirements

5G QoS requirements should be defined in multiple dimensions such as user perspective, network and service. Two key traffic models should be considered: high-speed video flow "server-subscriber" and massive Machine-to-Machine (M2M), or Device-to-Device (D2D) communications [11].

5G will be a set of telecommunication technologies and services that support 1000 times more data capacity than today, and should provide ultra-low latency response of less than few milliseconds. The network should provide a capacity of 50 Gbps per cell, and guarantee anywhere more than 1 Gbps per user through superdense networking, while supporting ultralow latency response of less than 1 ms in data plane [11-12]. It should provide 4A (Anytime, Anywhere, Anyone, Anything) massive connectivity that will accommodate 1000 times more different mobile devices. It will have flexible and intelligent network architecture with software based structure that will analyze data in real time and will provide intelligent and personalized services [2].

5G will provide reliable secure operation with more than 99% network availability. There will be possibility for self-healing reconfiguration and self-optimization. The battery life of the mobile devices will be increased to 10 times. Finally, 5G will have low cost for infrastructure and devices and will be 50 to 100 times more efficient in terms of energy [12].

The QoS management mechanisms in 5G networks should provide video and VoIP traffic prioritization towards web-search traffic and other applications tolerant to quality [11]. Some of the QoS parameters are Packet Delay Budget (PDB), or maximum packet delay and Packet Loss Ratio (PLR). Table 1 summarizes some of the 5G QoS requirements.

Table 1. 5G QoS requirements

Parameter	Value
Air Link User Plane Latency	1 ms
Air Link Control Plane Latency	50 ms
End-to-end Latency (device to core)	5 to 10 ms
Connection density	100 x compared with LTE
Area Capacity Density	1 Tbit/s/km^2
Uplink Cell Spectral Efficiency	5 bit/s/cell
Downlink Cell Spectral Efficiency	10 bit/s/cell
System Spectral Efficiency	10 bit/s/Hz/cell
Peak Throughput (Downlink) per Connection	10 to 50 Gbit/s
Energy Efficiency	90% improvement over LTE
PDB with guaranteed quality	1 ms
PLR for video broadcasting (8k UHD)	10^{-9}
PLR for M2M services (without quality assurance)	10^{-4}
PLR for M2M services (with guaranteed quality)	10^{-7}

2.2 Mobile Cloud Computing in 5G

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [5, 13-15]. Mobile cloud computing on the other hand is an integration of cloud computing technology in mobile environment and provides all the necessary resources to overcome the obstacles of the mobile devices [16-18]. It is an infrastructure used by mobile applications where both the data storage and data processing are moved away from the mobile device to powerful, centralized and high performance computing platforms located in the clouds.

Full Network Function Virtualization (NFV) will take place in 5G, in order to satisfy the service requirements. NFV pools the underlying physical resources or logical elements in a network, by using the current technologies such as cognitive and software defined radios in the 5G RAN, software defined networking and cloud services in 5G core [19].

The NFV functions should cover the control and management of QoS, the service policy and prioritization of traffic. The network functions like signal processing and path computation will be virtualized and offloaded to network management clouds. This will make network operating easier, reduce end-to-end energy consumption and open the way to new flexible communication services. The main cloud computing functionality in 5G core will be Network as a Service (NaaS) platform that will allow configuration of all telecommunication and service functions with virtualized software on a programmable hardware [10].

5G will include a mix of centralized cloud and a distributed cloud using access network and RAN elements [10]. By moving the cloud in radio access network, known as Cloud RAN (CRAN), i.e. by moving the base station from the cell site into cloud, where storage and computing resources are placed at, or close to the cell site, operators will be able to guarantee the necessary Service Level Agreement (SLA) and can better support delay-sensitive applications such as virtual desktops or electronic programming guides. They will also improve response times, making services feel "snappier" and, uniquely, more responsive to prevailing radio conditions. This might be useful for congestion control, or rate adaptation for video streams. On the network side, cell site caching can reduce demand on the backhaul network, and potentially play a role in limiting signaling to the core network.

Since 5G RAN consists of a dense deployment of micro, pico and femto cells, mobile cloud computing would significantly reduce latency over the wireless communication channel, as well as the transmit power necessary for computation offloading. Moreover, mobile device can easily locate a cloud access point. Millimeter wave links, massive MIMO, and multi-cell cooperation can be used to improve the spectral efficiency, by reducing the time necessary to transfer the users' offloading requests to the cloud.

Mobile cloud computing in 5G network should support context-rich support services such as context extraction service, recommendation service and group privacy service [18]. Particularly important is the context extraction service that performs data mining analysis of mobile data combined with other forms of data such as social networking data, and sensor network data in order to extract contextual clues relevant to the user. Data mining services should be able to scale and analyze large group of people and large quantities of data (big data) in order to extract collective trends among the population of users in real time. Additionally, recommendation services based on collective group context rather than individual context need to be created and scaled. By using these clues, a layer of cloud recommendation services can be built that creates output which is adjusted to a user, or set of users with those contextual characteristics.

3 Fog Computing

Although mobile cloud computing is a promising solution for 5G, still it cannot deal with all future Internet services and applications. This is because future Internet will exacerbate the need for improved QoS/QoE, supported by services that are orchestrated on-demand and are capable of adapt at runtime, depending on the contextual conditions, to allow reduced latency, high mobility, high scalability, and real time execution. The emerging wave of Internet of Things (IoTs) would require seamless mobility support and geo-distribution in addition to location awareness and low latency. These demands can only be partially fulfilled by existing cloud computing solutions [5].

A new paradigm called Fog Computing, or briefly Fog has emerged to meet these requirements [6]. Fog extends cloud computing and services to the edge of the network. It provides data, computing, storage, and application services to end-users that can be hosted at the network edge or even end devices such as set-top-boxes or access points. Its main features are proximity to end-users, dense geographical distribution, and support for mobility [7]. A comparison between fog computing and cloud computing is given in [20], which is summarized in Table 2.

Table 2. Fog computing vs cloud computing

	Fog Computing	Cloud Computing
Target Type	Mobile users	General Internet users
Service Type	Limited localized information services related to specific deployment locations	Global information collected from worldwide
Hardware	Limited storage, compute power and wireless interface	Ample and scalable storage space and compute power
Distance to Users	In the physical proximity and communicate through single hop wireless connection	Faraway from users and communicate through IP networks
Working Environment	Outdoor (streets, parklands, etc.) or indoor (restaurants, shopping malls, etc.)	Warehouse-size building with air conditioning systems
Deployment	Centralized or distributed in regional areas by local business (local telecommunication vendor, shopping mall retailer, etc.)	Centralized and maintained by Amazon, Google, etc.

By deploying reserved computing and communication resources at the edge, fog computing absorbs the intensive mobile traffic using local fast-rate connections and relieves the long back and forth data transmissions among cloud and mobile devices [7, 20]. This significantly reduces the service latency and improves the service quality perceived by mobile users, and more importantly, greatly saves both the bandwidth cost and energy consumptions inside the Internet backbone. Fog computing represents a scalable, sustainable and efficient solution to enable the convergence of cloud-based Internet and the mobile computing. Therefore, fog paradigm is well positioned for real time big data analytics, 5G network, and IoT. An overview of three layered Fog Computing architecture is given in Figure 2.

The intermediate fog layer consists of geo-distributed intelligent fog computing servers which are deployed at the edge of networks, e.g., parks, bus terminals, shopping centers, etc. Each fog server is a highly virtualized computing system and is equipped

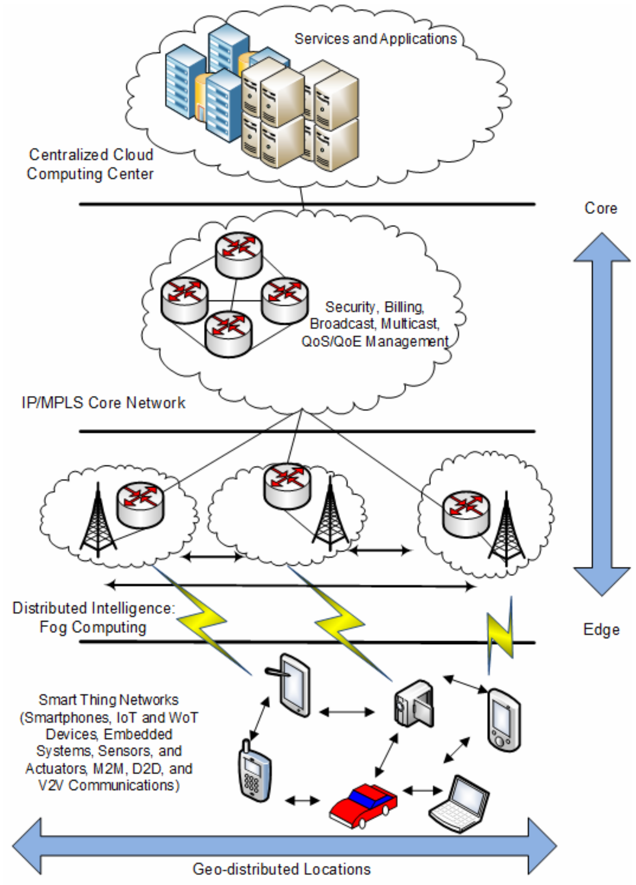

Figure 2. Fog computing architecture

with the on-board large volume data storage, compute and wireless communication facility [20].

The role of fog servers is to bridge the smart mobile device things and the cloud. Each smart thing device is attached to one of fog servers that could be interconnected and each of them is linked to the cloud [7]. The geo-distributed intelligent fog servers directly communicate with the mobile users through single-hop wireless connections using the off-the-shelf wireless interfaces, such as, LTE, WiFi, Bluetooth, etc. They can independently provide pre-defined service applications to mobile users without any assistance from cloud or Internet. In addition, the fog servers are connected to the cloud in order to leverage the rich functions and application tools of the cloud.

In a fog the distributed Peer-to-Peer (P2P) mobile cloud approach can be used [21], where a group of mobile devices acts as a cloud and provides cloud services to other mobile devices with a guaranteed quality of certain level of service agreements. The peers have 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. P2P architectures have ability to adapt to network failures and dynamically changing network topology with a transient population of nodes/devices, while ensuring acceptable connectivity and performance. Thus, P2P systems exhibit a high degree of self-organization and fault tolerance.

4 Fog Service Orchestration for 5G

Traditional service orchestration approaches that have been applied to cloud services are not adequate to the forthcoming large-scale and dynamic 5G services, since they cannot effectively cope with reduced latency, high mobility, high scalability, and real time execution. A new Hybrid Environment Service Orchestrator (HESO) is needed, in order to ensure resilience and trustworthiness of open, large scale, dynamic 5G services in the fog [10].

The Orchestrator will be responsible for the composition of service elements available in the Fog environment (e.g. sensing, connectivity, storage, processing, platform services, and software services) into more complex fog services (e.g. traffic crowd sensing and trip planning services) to be offered to the 5G users in the Fog. The orchestrator also need to synchronize and combine the operation of the different service elements in order to meet the specifications of the composed fog services.

Our proposed HESO model is given in Figure 3. The HESO should operate in a loosely coupled mode, resulting in a solution with several levels: Regional Service Orchestrator (RSO), Domain Service Orchestrator (DSO) and Federated Service Orchestrator (FSO).

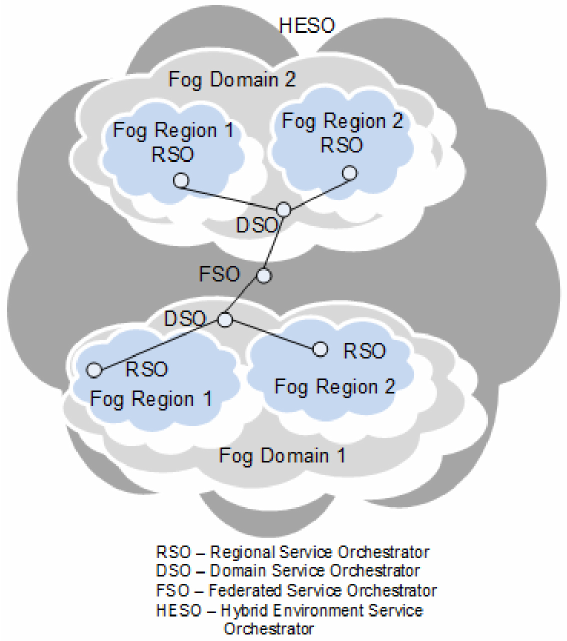

Figure 3. Hybrid environment service orchestrator

The RSOs are located at the edges of the Fog environment and enable semi-autonomous operation of the different fog regions. This allows the distribution of the load which provides scalability and a much higher proximity to the end users. Therefore, lower latencies are provided.

The DSOs are responsible for the Fog domains and supervises the RSOs below. This level supports federation mechanisms to enable intra-domain cooperation between different regions within one domain.

The FSO allows a fruitful interaction between different fog domains. It is responsible for the management between different fog domains and, similarly to the DSOs, it should be properly adapted to operate in a federated cloud environment. The FSOs will support federation mechanisms to enable cooperation among different fog domains (e.g. belonging to different entities or under the administration of different authorities) and the creation of a multi-domain fog environment able to support service ubiquity.

HESO model is flexible and scalable and can be implemented in any network technology standard. In particular, its application is important for critical usage cases of IoT devices and Tactile Internet that requires 1 ms end-to-end latency in order to 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 [22].

5 HESO Model Evaluation

The performances of HESO model can be explored in terms of Round Trip Time (RTT) latency, throughput, and the product latency - throughput. To our best knowledge, so far no similar evaluation is performed.

The simulation scenario is given in Figure 4. There is a single region in which are located a group of smart user devices, which are simultaneously served by 3G, 4G and 5G RAN network. Each RAN is connected to ten clouds. First five clouds are in the same region with the RANs, and the other 5 clouds are in a different region with the RANs. The smart user devices are assumed to be equally capable and are simultaneously served by the RANs and the clouds. Here a single user equipment will be considered. The simulation scenario can be easily constructed with Matlab.

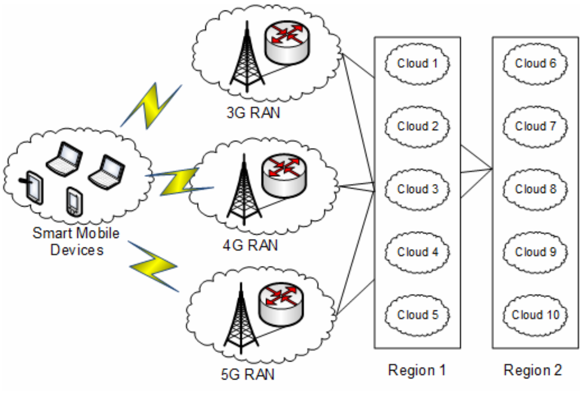

Figure 4. Simulation scenario

5.1 RTT Latency

RTT latency is the time which takes a single data transaction to occur, meaning 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 to any cloud via any RAN is equal to:

$$RTT = RTT_{RAN} + RTT_{RAN-CLOUD} \quad \textbf{(1)}$$

where RTT_{RAN} represents RTT RAN latency for 3G, 4G and 5G network, and $RTT_{RAN-CLOUD}$ represents the RTT Latency between the RAN network and the cloud. The average RTT_{RAN} values for 3G, 4G and 5G are given in Table 3, and the average values of $RTT_{RAN-CLOUD}$ varies from 50 ms to 500 ms depending whether the cloud is in the same or different region with the RAN network. For the simulation purposes randomly generated values were used for $RTT_{RAN-CLOUD}$. Cloud 1 is the least distant from the RANs, with the lowest RAN – CLOUD latency, while cloud 10 is the most distant from the RANs, with the highest RAN – CLOUD latency.

Table 3. Peak Data Rate, RTT Latency for 3G, 4G and 5G RAN and the Distance between the User Equipment and Each RAN

Parameter	Network Type		
	3G	4G	5G
RTT_RAN Latency [ms]	70	20	5
Peak Data Rate [Gbps]	0.021	3	40
Distance d [m]	1200	800	400

The simulation results for the RTT latency for different RANs are given in Figure 5. RTT latency between the user equipment and any cloud is the lowest for 5G RAN, and the highest for 3G RAN. The RTT latency for any RAN exponentially increases from cloud 1 to cloud 10, because cloud 1 has the lowest latency to the RANs, and cloud 10 has the highest latency to the RANs.

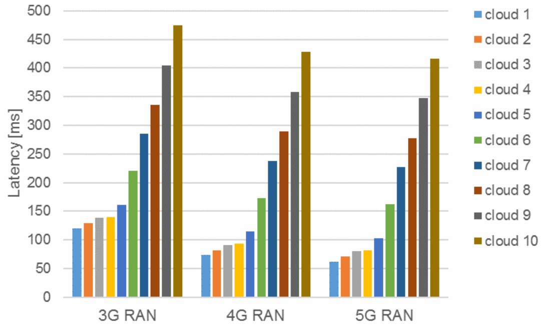

Figure 5. RTT Latency in a 5G Network in the Fog

5.2 Throughput

Throughput is the quantity of data that can pass from source to destination in a specific time. The total throughput of any user served by L RANs (L=3) is:

$$T = \sum_{i=1}^{L} r_i T_{RAN_i} \qquad \textbf{(2)}$$

Here r_i is the weight coefficient that identifies the contribution of each RAN in the total throughput. These values depend primarily from the distance between the mobile user and each RAN. They can be configured by the RAN operators and one such possible configuration is given in Table 4. The distances between each user equipment and each RAN are given in Table 3.

Table 4. Values for r_i as a Function of the Distance between the User Equipment and the RAN *i*

Distance d [m]	r_i
< 500	1
$500 \le d < 1000$	0.9
$1000 \le d < 1500$	0.8
$1500 \le d < 2000$	0.7
$2000 \le d < 2500$	0.6
$2500 \le d < 3000$	0.5
$2500 \le d < 3000$	0.4
$3000 \le d < 3500$	0.3
$3500 \le d < 4000$	0.2
$4000 \le d < 4500$	0.1
$4500 \le d$	0

The user throughput for each RAN T_{RAN_i} can be calculated as a ratio of the peak data rate R in the RAN given in Table 3, and the number of users N served by all clouds connected to that RAN is:

$$T_{RAN_i} = \frac{R}{N} \qquad \textbf{(3)}$$

T_{RAN_i} is also equal to sum of the flows of all clouds M (M= 10):

$$T_{RAN_i} = \sum_{j=1}^{M} T_{RAN_{ij}} \qquad \textbf{(4)}$$

where

$$T_{RAN_{ij}} = c_j T_{RAN_i} = \frac{c_j R}{N} \qquad \textbf{(5)}$$

where c_j is the weight coefficient that identifies the contribution of each cloud in the user throughput in RAN *i*. Here the first 5 clouds (cloud 1 to 5) are in the same region with the RANs, and the other 5 clouds (cloud 6 to 10) are in a different region. Therefore the first 5 clouds will have weight coefficients equal to c_1, and the other 5 clouds will have weight coefficients equal to c_2. Because of the geographic location of the clouds it will be assumed the coefficient c_1 to be twice the coefficient c_2. Like that the following values are obtained: $c_1 = \frac{2}{15}$, and $c_2 = \frac{1}{15}$.

The user throughput results, when the user receives data from all clouds, are shown in Figure 6. 5G offers much higher user throughput than 4G and 3G. This means that much higher quantity data can pass through 5G RAN, compared to 3G and 4G RANs.

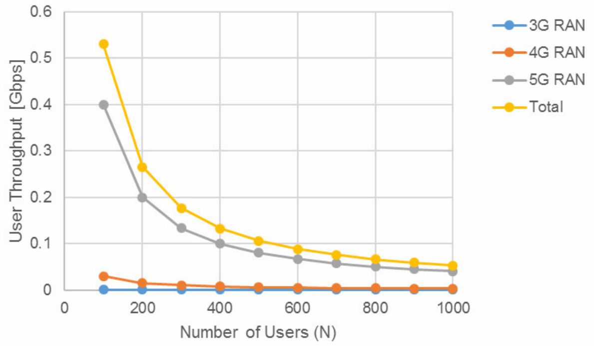

Figure 6. User throughput in a 5G network in the fog (user receives data from all clouds)

In 5G the user equipment will be simultaneously connected to several RANs, and several clouds i.e. it will combine the flows from several RANs, and several clouds, for a single application or service. This is the total user throughput which is also shown in Figure 6. Therefore, the user throughput is significantly increased.

The user throughput results, when the user receives data from a single cloud (cloud 1, or cloud 10) are shown in Figure 7 and Figure 8. Here 5G again offers much higher user throughput than 4G and 3G, and the total throughput is the highest. However, when a single cloud is used the throughput is significantly reduced. If the cloud is in the same region with the RANs the multiplication factor is 2/15, and if the cloud is in a different region the multiplication factor is 1/15.

5.3 Product Latency - Throughput

The product *RTT x User Throughput* and *RTT x Peak Data Rate* correspond to the bandwidth delay product, i.e. to the maximum amount of data on the network at any given time, that has been transmitted but not yet acknowledged. The simulation results for these products for cloud 1 and 10 are given on Figure 9, Figure 10 and Figure 11. Here 5G RAN again outperforms 3G and 4G RANs.

5.4 Discussion of the Results

The results clearly demonstrate the benefits of fog computing HESO Mechanisms in 5G network, especially if the flows from different RANs and clouds for a particular application or service requested by the user are combined. This will primarily depend, whether the user requests a service that requires high throughput, low latency, optimal product latency - throughput, energy efficiency of clouds and RANs, etc.

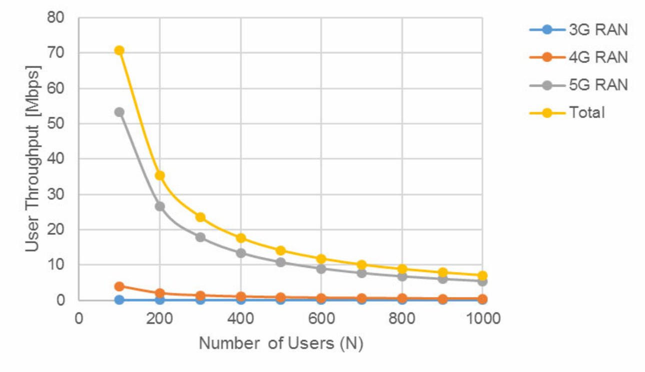

Figure 7. User throughput in a 5G network in the fog (user receives data from cloud 1)

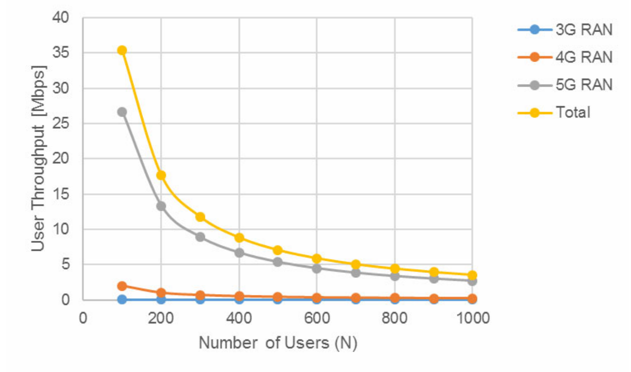

Figure 8. User throughput in a 5G network in the fog (user receives data from cloud 10)

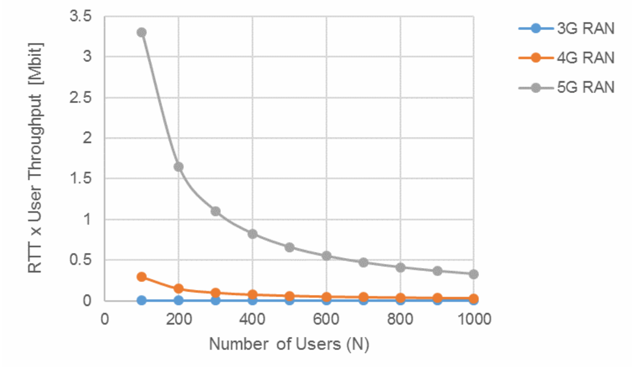

Figure 9. Product RTT x user throughput for cloud 1

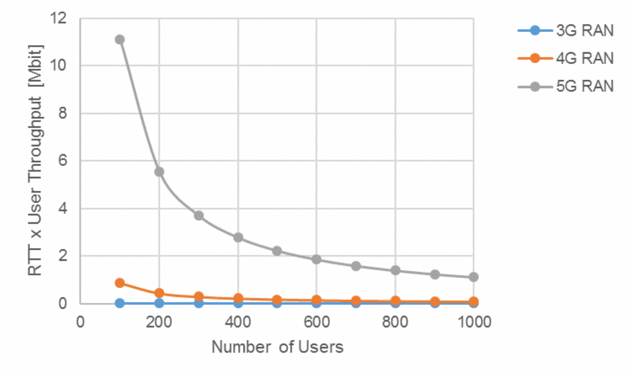

Figure 10. Product RTT x user throughput for cloud 10

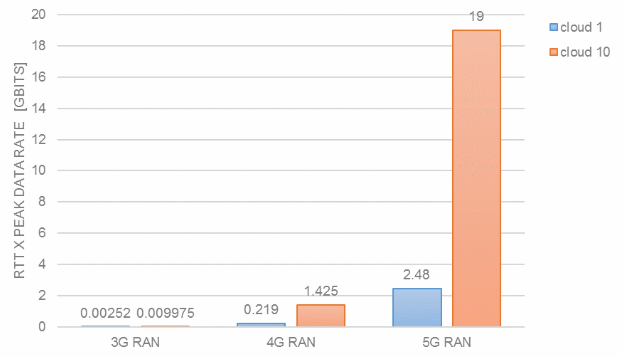

Figure 11. Product RTT x peak data rate for cloud 1 and 10

6 Conclusion

This paper proposed a Hybrid Environment Service Orchestrator (HESO) model for resilient and trustworthy fog computing services for 5G network. It also performed evaluation HESO model in terms of latency, throughput, and the product latency - throughput. The results clearly show that 5G will have a great benefit of using the HESO model, because its service orchestration mechanisms will effectively cope with the forthcoming services that require reduced latency, high mobility, high scalability and real-time execution, such as Tactile Internet and real time big data analytics.

In 5G network, HESO will act as a nervous system of the digital society, economy, and everyday people's life. The cloud in 5G networks will be diffused among the client devices often with mobility too, i.e. the cloud will become fog. More and more virtual network functionality will be executed in a fog computing environment, and it will provide *mobiquitous* service to the users. This will enable new *AaaS* service paradigms, where devices, terminals, machines, and also smart things and robots will become innovative tools that will produce and use applications, services and data. 5G in the fog will use the benefits of the centralized cloud, CRAN cloud and the distributed P2P mobile cloud which will create opportunities for companies to deploy many new real-time services that cannot be delivered over current mobile and wireless networks.

Finally, the selection of RAN and cloud flows that will be combined and used for a single service or application by the mobile user will depend from the low latency requirements, high throughput requirements, and optimal latency throughput, or bandwidth delay, or energy efficiency of the clouds and the RANs. The algorithm for selecting the RAN and cloud flows will be the direction of our future work.

References

[1] T. Janevski, *Internet Technologies for Fixed and Mobile Networks*, Artech House, 2015.

[2] T. Janevski, *NGN Architectures Protocols and Services*, John Wiley & Sons, 2014.

[3] C. Wang, F. Haider, X. Gao, X. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, E. Hepsaydir, Cellular Architecture and Key Technologies for 5G Wireless Communication Networks, *IEEE Communication Magazine*, Vol. 52, No. 2, pp. 122-130, February, 2014.

[4] X. Wang, M. Chen, T. Haleb, Cache in the Air: Exploiting Content Caching and Delivery Techniques for 5G Systems, *IEEE Communication Magazine*, Vol. 52, No. 2, pp. 131-139, February, 2014.

[5] S. Zhang, S. Zhang, X. Chen, X. Huo, Cloud Computing Research and Development Trend, *Proceedings of the Second IEEE International Conference on Future Networks (ICFN 2010)*, Sanya, Hainan, 2010, pp. 93-97.

[6] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, 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.

[7] L. M. Vaquero, L. Rodero-Merino, Finding your Way in the Fog: Towards a Comprehensive Definition of Fog Computing, *ACM SIGCOMM Computer Communication Review Newsletter*, Vol. 44, No. 5, pp. 27-32, October, 2014.

[8] T. Janevski, 5G Mobile Phone Concept, *Proceedings of 6th IEEE Consumer Communications and Networking Conference-CCNC 2009*, Las Vegas, Nevada, 2009, pp. 1-2.

[9] A. Tudzarov, T. Janevski, Functional Architecture for 5G Mobile Networks, *International Journal of Advanced Science and Technology (IJAST)*, Vol. 32, pp. 65-78, July, 2011.

[10] S. Kitanov, E. Monteiro, T. Janevski, 5G and the Fog–Survey of Related Technologies and Research Directions, *in Proceedings of the 18th Mediterranean IEEE Electrotechnical Conference MELECON 2016*, Limassol, Cyprus, 2016, pp. 1-6.

[11] V. Tikhvinskiy, G. Bochechka, Prospects and QoS Requirements in 5G Networks, *Journal of Telecommunications and Information Technologies*, Vol. 1, No. 1, pp. 23-26, January, 2015.

[12] GSA, *The Road to 5G: Drivers, Applications, Requirements and Technical Development*, 2015.

[13] Recommendation ITU-T Y.3501, *Cloud Computing Framework and High-Level requirements*, 2013.

[14] Recommendation ITU-T Y.3510, *Cloud Computing Infrastructure Requirements*, 2013.

[15] NIST Special Publication 500-291, *NIST Cloud Computing Standards Roadmap*, Version 2, 2013.

[16] T. H. Dihn, C. Lee, D. Niyato, P. Wang, A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches, *Wireless Communications and Mobile Computing*, Vol. 13, No. 18, pp. 1587-1611, December, 2013.

[17] D. Huang et al., Mobile Cloud Computing, *IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter*, Vol. 6, No. 10, pp. 27-31, October, 2011.

[18] S. Kitanov, T. Janevski, State of the Art: Mobile Cloud Computing, *Proceedings of the Sixth IEEE International Conference on Computational Intelligence, Communication Systems and Networks 2014* (CICSYN 2014), Tetovo, Macedonia, 2014, pp. 153-158.

[19] J. Marinho, E. Monteiro, Cognitive radio: Survey on Communication Protocols, Spectrum Decision Issues, and Future Fesearch Directions, *Springer Wireless Networks*, Vol. 18, No. 2, pp. 147-164, February, 2012.

[20] H. T. Luan, L. Gao, Z. Li, L. X. Y. Sun, Fog Computing: Focusing on Mobile Users at the Edge, *arXiv:1502.01815v3 [csNI]*, 2016.

[21] H. Kavalionak, A. Montresor, P2P and Cloud: A Marriage of Convenience for Replica Management, *Proceedings of the 6th IFIP TC 6 international conference on Self-Organizing Systems,* Delft, Netherlands, 2012. pp. 60-71.

[22] The Tactile Internet, *ITU Technology Watch Report*, 2014.

Biographies

Stojan Kitanov is a Teaching and Research Assistant at Mother Teresa University, Faculty of Informatics, Skopje, Macedonia. His interests include Internet Technologies, Next Generation Mobile/Wireless Networks and Services, Quality of Service, Cloud and Fog Computing.

Dr. Toni Janevski is a Full Professor at the Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia. His interests include Internet Technologies, Mobile/Wireless Networks and Services, Traffic Engineering, Quality of Service, Design of Telecommunication Networks, Next Generation Networks, Cloud Computing, and Future Networks.