

Multi-level edge computing framework for 5G cellular system with D2D enabled communication

Abdelhamied A. Ateya^{1,*}, Ammar Muthanna*, Andrey Koucheryavy*

¹Electronics and Communications Engineering, Zagazig University, Zagazig, Egypt,

*State University of Telecommunication, 22 Prospekt Bolshevnikov, St. Petersburg, Russia

a_ashraf@zu.edu.eg, ammarexpress@gmail.com, akouch@mail.ru

Abstract— There is no doubt that the release of the fifth generation of the mobile cellular system (5G) becomes a great demand year by year, especially with the massive increase of heterogeneous Wireless devices. The 5G network will be a platform for wide variety of industries. Designing of 5G cellular system faces various challenges related to the capacity and traffic. One way to solve these challenges is to employ device-to-device (D2D) communication and mobile edge computing (MEC). Employing these technologies offload the core network and increase the capacity of the system. In this work, we propose a frame work for the 5G cellular system based on D2D communication and multi-level cloud units employed at the edge of the cellular network. The system employs four levels of cloud units with various hardware capabilities. The D2D communication is used as the communication technology in the first level of clouds. Employing D2D together with multi-level edge cloud units achieves various benefits to the system as the system level simulation provides.

Keywords— 5G, D2D, Offloading, Mobile edge computing, Latency, Traffic.

I. INTRODUCTION

With the massive increase of heterogeneous wireless devices (e.g. smart phones, MTC devices and wearable devices), connectivity and traffic volume become a design challenge for the 5G cellular system. It is expected that the data traffic in 2020 will be 200 times much higher than that in 2010, and for 2030 data traffic will be 20,000 times higher than that of 2010 [1]. One of the main design goals of the 5G cellular system is to achieve a data rate up to ten Gbps [2]. Moreover, 5G cellular system is required to achieve high throughput, ultra-low latency, high reliability, high connectivity and high mobility [3].

To overcome the traffic load problem and the connectivity problem; and to achieve the expected data rate with ultra low latency, 5G cellular system should employ new technologies at various locations of the network. Some technologies should be employed at the core network such as the software defined networking (SDN) and network function virtualization (NFV) [4], [5]. Other technologies are employed at the edge of the network to offload the core network and also to provide a low latency. These technologies are like MEC and D2D communication [6], [7].

MEC is a recent technology that moves IT and cloud computing capabilities at the edge of the cellular network

within the Radio Access Network (RAN). MEC provides cloud computing near to users, approximately one hop away [8]. Employing MEC means, moving from centralized huge data centers to a distributed small data centers with limited capabilities compared to the centralized units. This can achieve the various benefits [9].

Mobile cloud computing is also an important trend in cloud computing, that moves the computing to the mobile device. The main goal of mobile cloud computing is to offer the execution of mobile application with the user experience [10]. Mobile cloud computing can be classified based on the service model into four types mobile as a service consumer (MaaSC), mobile as a service provider (MaaSP), mobile as a service broker (MaaSB) and mobile as a service representer (MaaSR) [11]. Each user in the cellular cell can be viewed as a type of the four types. In MaaSC, the mobile doesn't handle any tasks and use the computation and storage from the cloud or other devices. This type is the most common one. If the device has the ability to share a part of its processing and storage hardware, the mobile belongs to the MaaSP type. An example of this type is a mobile device collects data with its embedded sensors (e.g. camera, GPS) and shares these data with other devices. MaaSB type acts as MaaSP with the ability of networking and forwarding. This means that the devices in MaaSB can act as a gateway to other devices [12].

In the other hand, D2D communication is also an enabling technology for the 5G cellular system [13]. D2D communication technology refers to the direct communication between proximate wireless devices over a licensed or unlicensed spectrum (e.g. WiFi-Direct, Bluetooth) without relaying traffic to the base station (BS) [14]. The BS may assist and control the D2D communication or stay away based on the operated mode of D2D communication, various modes of D2D communications are discussed in [15].

Based on the Wi-Fi alliance watch report [16], the number of the certified wireless devices support D2D communication in 2017 is 4 million and in 2018 is predicted to reach to 10 million. This pushes the designers of the next generation cellular system to use the D2D as an enabling technology [17]. Employing D2D communication in cellular cell achieves higher throughput, increase the cell coverage, reduces traffic congestion and achieves higher spectral efficiency [18]. Thus, D2D communication presents a flexible efficient paradigm for 5G cellular system.

In this paper, we introduce a framework for 5G cellular system based on multi-level hierarchical cloud units and D2D communication. In (Sec. II) the related works to the proposed work is discussed. In (Sec. III) the framework is presented. In (Sec. IV) the simulation results are presented.

II. RELATED WORKS

Recently, the research in the area of edge computing and D2D communications as a part of future 5G cellular system becomes more attractive. Both technologies offload the core network and reduce the network congestion. D2D communication increases the system coverage by enabling uncovered devices to relay their data through other covered devices. There are a lot of literatures in both technologies but separately and little works discuss the employment of both technologies together in 5G system [19], [20]. We provide a brief description for the most related works to our proposed work.

In [21] authors provide a framework for internet of things (IoT) based on D2D communication. The work is based on a forwarding scheme used when an IoT device become in the coverage of neighbouring cellular edge nodes. IoT device can then use the D2D communication to communicate with the edge neighbouring cellular nodes. Authors, prove that using D2D communication in this way achieves better latency efficiency, reduces the traffic load through the network and improves the cooperation of IoT devices at the network edge.

In [11] authors provide a network assisted architecture for the long term evolution (LTE) cellular system based on hierarchical clouds. The main idea behind this work is to build a mobile dynamic clouds using either licensed or unlicensed D2D communication. The system improves the channel utilization and offloads traffic for the LTE system.

In [9] we present a frame work for the Tactile Internet based on the multi-level edge clouds. The system employs a small data centers at the edge of the network connected to each BS and referred as Micro-cloud unit. Each group of Micro-cloud units is connected to a large data center with higher processing and storage capabilities marked as Mini-cloud unit. The core network employs a huge central cloud unit with powerful hardware capabilities. All Mini-cloud units are connected through a fiber network with the core network cloud. The system achieves high performance in terms of round trip latency.

III. PROPOSED SYSTEM

Employing D2D communication together with the MEC achieves various benefits to the cellular system. We present a frame work for the 5G cellular system based on D2D communication and Multi-level cloud units at the edge of the network. In this section, we discuss the proposed structure.

Table 1. System devices.

parameter	Definition
OD	Ordinary Device
SD	Sensing Device
MD	Master Device
GD	Gateway Device

The proposed system consists of four main parts from the cloud computing point of view, based on the computing capabilities of the network parts in comparable to the required task. Figure 1 illustrates the proposed system structure. Table 1 point out all abbreviations used.

A. Part one: Cloutlet Tire

GDs with MDs represent the cloutlet tire. Figure 2 illustrates such example for the structure of the cell under this tire. GD is a wireless device with a powerful hardware capable of providing some computing processes, networking and data forwarding to other wireless devices. An example of a GD is a mobile device with a powerful hardware that can employ its sensors to collect data from the surrounding environment (even social or natural). This device can process and analyze the collected data and share with the ODs. Also, GD can use its storage units to handle data for ODs and SDs. If the GD device receives tasks higher than its capabilities (e.g. can't perform a task), the GD forwards these undelivered tasks to a MD that has higher capabilities.

MD is a wireless device with extremely powerful hardware capable of processing tasks that can't be handled by the GD. A wireless device; capable of supporting most of available air interfaces, has a powerful multi-core processor, with a wide storage unit and latest reliable software can be employed as a MD.

MD can handle different tasks for GDs, ODs and SDs. If the received tasks are higher than the capabilities of the MD and required more powerful processing or storage, MD forwards these tasks to the Micro-cloud unit at the base station (BS).

Figure 3 illustrates an example of a structure with heterogeneous devices and services. Device A acts as an ordinary device which require a multimedia task, device B and C are also ODs which ask for services from the GD D.

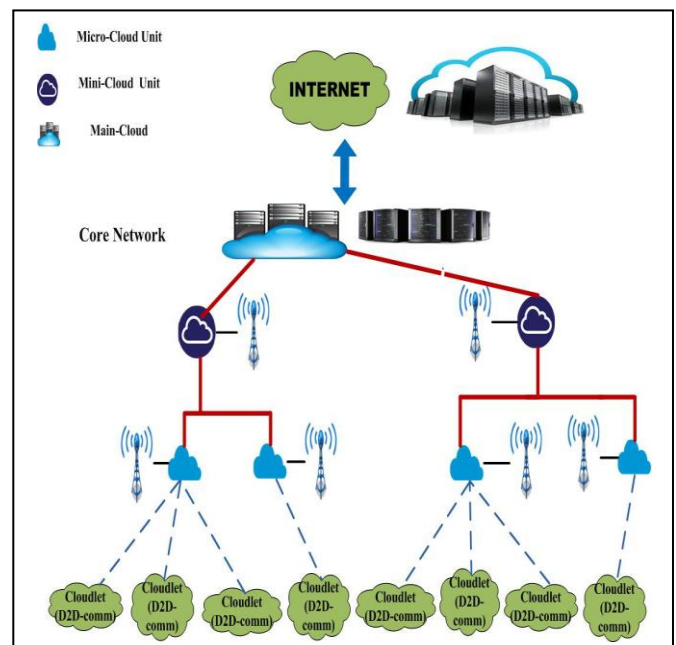


Figure 1. System structure in terms of cloud units.

The GD D can handle tasks for ODs B and C, but the multimedia task of OD A can't be handled due to hardware limitations of device D. In order to handle this task, device D forwards the task to the MD G which can perform it

B. Part two: Micro-cloud Tire

We built a multi-level cloud system in [9], [22] based on Micro-cloud and Mini-cloud units. Here, we use them with the cloud let level. Each eNB is connected to a cloud unit serves for a cellular cell, and the cloud unit will be a Micro or a Mini- cloud unit.

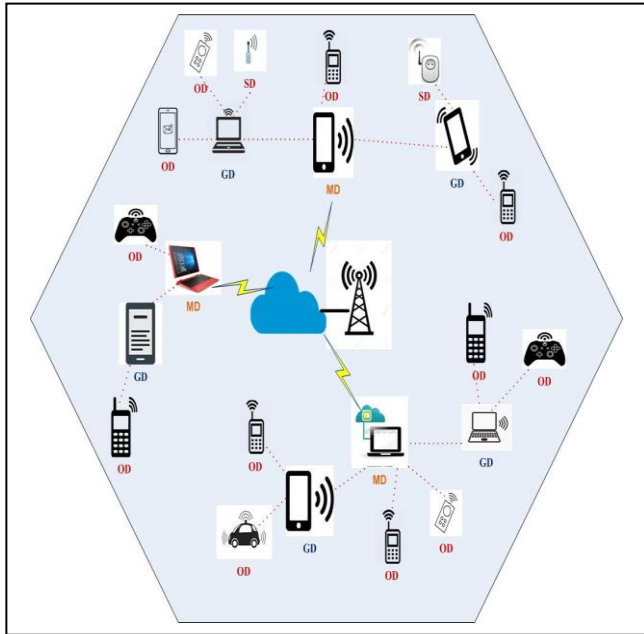


Figure 2. Cell example under Cloudlet tire.

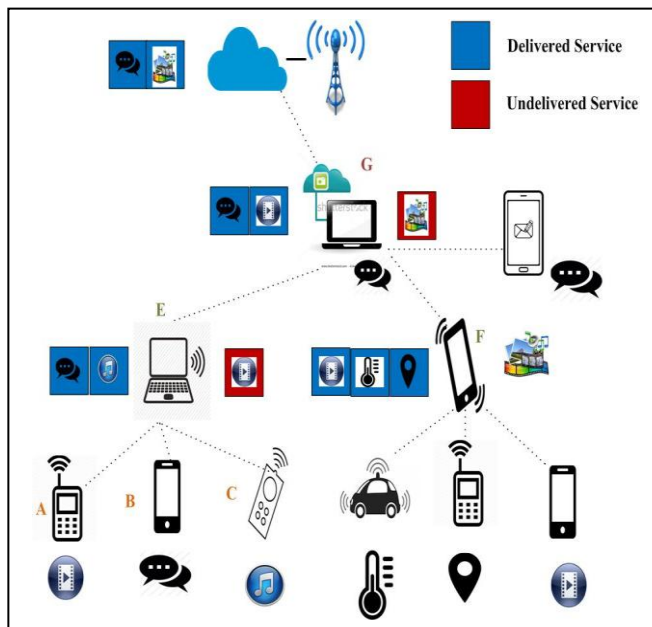


Figure 3. Service handling in Cloudlet tire.

The 5G system will employ heterogeneous types of cellular cells, thus cells with predictable high density of users employ Mini-cloud unit and the less density employ Micro-cloud unit. Each group of Micro-cloud units is connected directly with a Mini-cloud unit.

A Micro-cloud unit is a small data centers with limited processing and storage capabilities connected to the eNB through a very high speed fiber cables. Micro-cloud unit receive and process tasks from MDs. Tasks delivered to MDs that need higher processing or storage capabilities than that of MDs are forwarded to the Micro-cloud unit connected to the eNB that serve for the cell. Moreover, MDs forward their own task to Micro-cloud unit at the eNB. Micro-cloud unit provides an offloading way for user workload and thus reduce the load and congestion in the core network

In fig.3 device G is a MD and can't handle the multimedia task for the device F. Thus, device G forward the task to the Micro-cloud unit connected to the eNB. Moreover, device G moves its own task to the Micro-cloud unit.

C. Part three: Mini-cloud Tire

Mini-cloud unit has higher processing and storage capabilities than Micro-cloud unit. It also serves for a cell as Micro-cloud unit. Mini-cloud unit consists of powerful data centers that can serve, monitor and control the connected Micro-cloud units. All information about the connection of Micro-cloud units to Mini-cloud unit and their numbers are discussed in [22].

Tasks that can't be processed by a Micro-cloud unit are forwarded to the Mini-cloud unit that connected to it. This is because the higher capabilities of Mini-cloud units compared to Micro-cloud units. Without the existence of Mini-cloud unit, the unperformed tasks at Micro-cloud units would be shifted far to the main central clouds passing through the core network. This will increase the overall delay for handling the task, overload the core network with extra traffic and increase the congestion of the network. Thus, presenting Mini-cloud unit in the way between Micro-cloud units and main central cloud units is helpful in terms of latency and offloading.

D. Part Four: Main cloud Tire

All Mini-cloud units are connected to the main central cloud unit employed at the core network through very high speed fiber connections. This main cloud unit acts as the gateway to the Internet public clouds (e.g., Microsoft Azure, Amazon Elastic Compute Cloud (EC2) and Google AppEngine). Mini-cloud unit is a group of powerful very high speed data centers. The main function of this tire is to act as a gateway and another function is to process and handle the received tasks that can't be handled by Mini-cloud units (either by process or forward to the public clouds).

Finally, we should point out the overall benefits achieved by the proposed system structure. Employing D2D communication technology together with the multi-level of cloud units provides variety of added values to the 5G cellular system. These benefits can be summarized in the following points:

- 1- The system structure provides an offloading way for the core of the cellular network. This can be achieved by D2D communication which can perform tasks without the involvement of the cellular network. Also multi-level clouds provide another way of offloading as they handle tasks as much as possible far from the core network. Thus, D2D together with multi-level clouds reduce the traffic load in the core network.
- 2- Reduction of round trip latency, as the communication distance is reduced either by D2D communication or multi-level clouds.
- 3- Reduction of network congestion,
- 4- Achieving higher throughput,
- 5- Increasing energy efficiency,
- 6- Better cell coverage, and
- 7- Increasing the spectral efficiency.

IV. SIMULATION AND RESULTS

In this section, the proposed system structure is simulated to evaluate the system performance. The system level simulation is built with Java language and implemented using Matlab R2015b. The simulation parameters are illustrated in table2 [22]. The users are uniformly distributed and the source nodes are assumed to generate heterogeneous types of events and workloads. The simulation is hold for different values of service latency. The maximum allowable service latency is that time, after which the service is blocked if it wasn't handled. We start with a small value of service latency then get this time increase, and measure the blocking rate for each time.

In order to evaluate the system performance, we consider a four simulation scenarios. Scenario (A): the network is simulated without D2D communication and the multi-level clouds. This means that all SDs, ODs, GDs and MDs have only communication with the eNB and the core network is the way to handle their workload. This is the worst scenario and the tasks blocking should be high especially for latency sensitive services.

Scenario (B): the network is simulated with the employment of D2D communication and thus employing the cloudlet tire. In this scenario, SDs and ODs can communicate directly with GDs and GDs can communicate with MDs. Employing D2D communication should reduce blocking probability of tasks especially for latency sensitive applications. This is because of the reduction of communication path.

Scenario (c): through this scenario we employ the Micro-cloud tire. Tasks that can't be handled by D2D (cloudlet) are forwarded to the Micro-cloud unit instead of the core network. The last scenario is (D) which employ more level of clouds. In this scenario we simulate the system with the existence of Mini-cloud unit connected to Micro-cloud unit. This is helpful for tasks that can't be handled by Micro-cloud unit and instead of moving them to the core network; they can forwarded to Mini-cloud unit.

Table 1. Simulation parameters.

parameter	Description	Value
N_{SD}	Number of sensing devices	10
N_{CD}	Number of ordinary devices	20
N_{GD}	Number of gate way devices	8
N_{MD}	Number of master devices	5
N_{maxGD}	Maximum number of SDs or CDs connected to a GD.	3
N_{maxMD}	Maximum number of SDs, CDs or GDs connected to a MD.	3
W_{mmax}	Maximum work load of the Mini-cloud unit per second	100 events/s
W_{cmax}	Maximum work load of the Micro-cloud unit per second	30 events/s
W_{MDmax}	Maximum work load of the MD	5 events/s
W_{GDmax}	Maximum work load of the GD	2 events/s
λ_i	Arrival rate of the Micro-cloud unit	15
μ_i	Service rate of the Micro-cloud unit	5 Mbps
μ_j	Service rate of the Mini-cloud unit	8 Mbps
d_{cell}	The communication latency inside the cellular cell	1 ms/hop
$d_{Cmicro,Cmini}$	The communication delay between Micro-cloud unit and Mini-cloud unit	1.5ms
S_{dmax}	Maximum allowable service latency	(5ms -15ms)
α	Gradient of computing function	10
β	Constant of computing function	0
RAM,HDD	Micro-cloud RAM, Storage Mini-cloud RAM, Storage	1024Mb,1Gb 2048Mb,5Gb

The performance metric used in all scenarios is the blocking probability (P_b), which is presents the ratio of the overall rejected work load to the total workload.

Figure 4 illustrates the system performance for all scenarios in terms of blocking probability for different values of maximum allowable latency of services. Comparing the four curves, we can point out that employing D2D communication achieves better blocking probability. Moreover employing Micro-cloud and Mini-cloud units add more efficiency in terms of blocking probability.

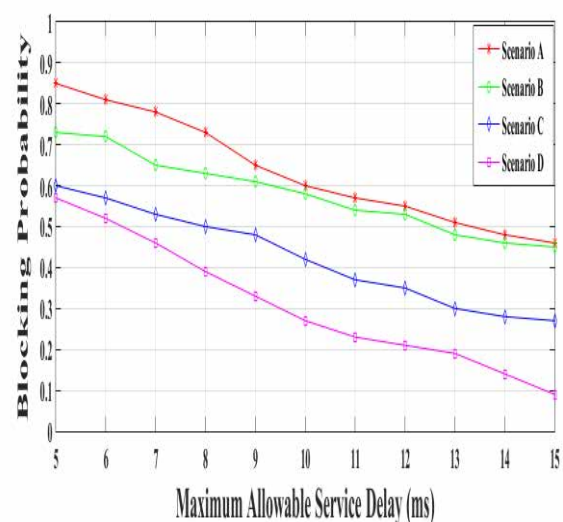


Figure 4. Blocking probability vs. Maximum allowable service delay (ms).

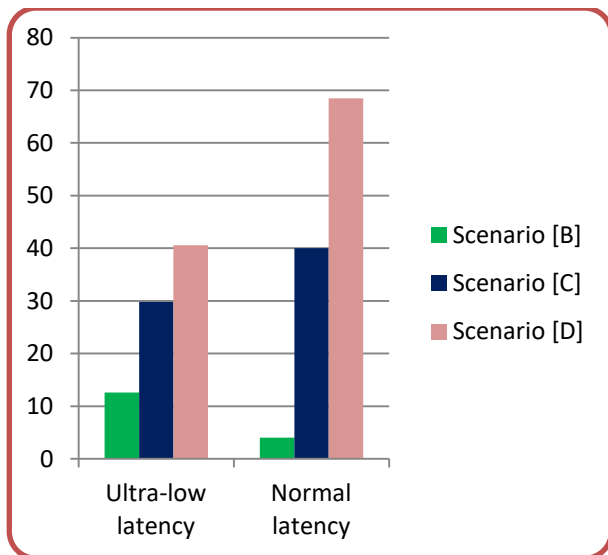


Figure 5. Four scenarios comparison in terms of blocking probability.

Figure 5 provides the percentage increase of efficiency in terms of blocking achieved in each simulation scenario compared with the first scenario (A), for two types of services. The first type is the high latency sensitive application which represents the first third part of the fig 4. The other type is the low latency services which represents the past third part of the figure 4.

V. CONCLUSIONS

The proposed frame work can be used as a structure for the 5G cellular system. The system employing four levels of heterogeneous cloud units. The first cloud tire is formed based on the D2D communication between various devices. The higher two levels of cloud units are edge cloud units employed at the edge of the RAN to offload the core network. The system achieves higher efficiency in terms of offloading, congestion, coverage and latency. The system is simulated in terms of blocking probability, for four scenarios each of them represents a cloud level. Simulation results illustrate that the worst case is the employment of the network without D2D and cloud levels. The best case is when employing D2D with all levels of cloud units.

FUTURE WORK

Our future vision is to comprise the SDN and NFV to the core of the network of the proposed system.

REFERENCES

- [1] D. Jiang, and G. Liu, "An Overview of 5G Requirements," 5G Mobile Communications, Springer, pp. 3-26, 2017.
- [2] K. Zheng, L. Zhao, J. Mei, M. Dohler, W. Xiang and Y. Peng, "10 Gb/s hetsnets with millimeter-wave communications: access and networking-challenges and protocols," IEEE Communications Magazine, Vol. 53(1), pp.222-231, Jan. 2017.
- [3] A. Tudzarov and S. Gelev, "Requirements for next generation business transformation and their implementation in 5G architecture," International Journal of Computer Applications, Vol. 162(2), pp.31-35, 2017.
- [4] A. Ateya, A. Muthanna, I. Gudkova, A. Vybornova and A. Koucheryavy, "Intelligent core network for Tactile Internet system," International Conference on Future Networks and Distributed Systems, ACM, P.15, Cambridge, Jul.2017.
- [5] R. Kirichek, A. Vladyko, M. Zakharov and A. Koucheryavy, "Model networks for internet of things and SDN," IEEE-ICACT2016 international conference, Korea, Jan.2016.
- [6] B. Rimal, D. Van and M. Maier, "Mobile edge computing empowered fiber-wireless access networks in the 5G era," IEEE Communications Magazine, Vol. 55(2), pp.192-200, Feb.2017.
- [7] G. Sim, A. Loch, A. Asadi, V. Mancuso and J. Widmer, "5G millimeter-wave and D2D symbiosis: 60 GHz for proximity-based services," IEEE Wireless Communications, Vol. 24(4), pp.140-145, 2017.
- [8] "Mobile Edge Computing A key technology towards 5G," ETSI White Paper, No. 11, September 2015.
- [9] A. Ateya, A. Vybornova, R. Kirichek and A. Koucheryavy, "Multilevel cloud based Tactile Internet system", IEEE-ICACT2017 international conference, Korea, Feb.2017.
- [10] A. Paranjothi, M. Khan and M. Nijim, "Survey on Three Components of Mobile Cloud Computing: Offloading, Distribution and Privacy," Journal of Computer and Communications, Vol. 5(06), p.1, April 2017.
- [11] M. Jo, T. Maksymuk, B. Strykalyuk and C. Cho, "Device-to-device-based heterogeneous radio access network architecture for mobile cloud computing," IEEE Wireless Communications, Vol.22(3), pp.50-58, June 2015.
- [12] Y. Wang, R. Chen, and D. Wang, "A survey of mobile cloud computing applications: perspectives and challenges," Wireless Personal Communications, Vol. 80(4), pp.1607-1623, Feb.2015.
- [13] A. Muthanna, P. Masek, J. Hosek, R. Fudjak, O. Hussein, A. Paramonov and A. Koucheryavy, "Analytical evaluation of D2D connectivity potential in 5G wireless systems", International Conference on Next Generation Wired/Wireless Networking, Springer, pp. 395-403, Sep.2017.
- [14] M. Ali, S. Qaisar, M. Naeem, S. Mumtaz and J. Rodrigues, "Combinatorial resource allocation in D2D assisted heterogeneous relay networks", Future Generation Computer Systems, Aug 2017.
- [15] A. Ateya, A. Muthanna, A. Vybornova and A. Koucheryavy, "Multi-level Cluster Based Device-to-Device (D2D) Communication Protocol for the Base Station Failure Situation", Internet of Things, Smart Spaces, and Next Generation Networks and Systems, Springer, Cham, pp. 755-765, Aug.2017.
- [16] Alliance, Wi-Fi, "Wi-fi certified wi-fi direct: Personal, portable wi-fi.," Wi-Fi Alliance, 2014.
- [17] L. Ignatova, A. Khakimov, M. Albahri, A. Muthanna, "Analysis of the internet of things devices integration in 5g networks.," IEEE, Systems of signal synchronization, generating and processing in telecommunications (SINKHROINFO), Kazan, 2017.
- [18] A. Asadi and V. Mancuso, "Network-assisted Outband D2D-clustering in 5G cellular networks: theory and practice," IEEE Transactions on Mobile Computing, Vol. 16(8), pp.2246-2259, Aug. 2017.
- [19] H. Zhang, N. Liu, X. Chu, K. Long, A. AghvamiH and V. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," IEEE Communications Magazine, Vol. 55(8), pp.138-145, 2017.
- [20] S. Singh, Y. Chiu, Y. Tsai and J. Yang, "Mobile Edge Fog Computing in 5G Era: Architecture and Implementation", IEEE International Computer Symposium (ICS), pp. 731-735, Dec. 2016.
- [21] A. Orsino, I. Farris, L. Militano, G. Araniti, S. Andreev, I. Gudkova, Y. Koucheryavy and A. Iera, "Exploiting D2D Communications at the

Network Edge for Mission-Critical IoT Applications", 23th European Wireless Conference, pp. 1-6, May 2017.



Abdelhamied A. Ateya received the BSc. and MSc. in Electrical Engineering from Zagazig University, Egypt, in 2010 and 2014, respectively. He is working as an Assistant Lecturer at the Faculty of Engineering, Zagazig University. He is currently a PhD student at St. Petersburg State University of Telecommunication, in St. Petersburg, Russia. Mr. Ateya's main area of research is wireless communication and currently, he is working on future Tactile Internet and its applications.



Ammar Muthanna is an Associate Professor at the Department of Telecommunication networks and Head of SDN Laboratory (sdnlab.ru), Saint - Petersburg State University of Telecommunications, Russia. He received his B.Sc. (2009), M.Sc. (2011) and as well as Ph.D. (2016) degrees from Saint - Petersburg State University of Telecommunications. In 2012/2013 he took part in Erasmus student program in the University of Ljubljana in Faculty of Electrical engineering. Area of research: wireless communications, 4G/5G cellular systems, IoT applications, and software defined networking.

- [22] A. Ateya, A. Vybornova, K. Samouylov and A. Koucheryavy, "System Model for Multi-level Cloud Based Tactile Internet System", Springer International Conference on Wired/Wireless Internet Communication, pp. 77-86, March 2017.



Andrey Koucheryavy was born in Leningrad 02.02.1952. After graduated from Leningrad University of Telecommunication in 1974 he going to Telecommunication Research Institute named LONIIS, where A.Koucheryavy working up to October 2003 (from 1986 up to 2003 as the First Deputy Director). He became the Ph.D. and Dr.Sc. in 1982 and 1994 respectively. A.Koucheryavy is the St. Petersburg State University of Telecommunication (SUT) professor from 1998. He is Chaired professor of the department "Telecommunication Networks and Data Transmission" from 2011. He is honorary member of A.S.Popov's society. Prof. A.Koucheryavy was the Chairman Study Group 11 ITU-T (Study periods 2017-2020). His scientific areas of interest are the network planning, teletraffic theory, IoT and its enablers.