A Device-to-device enabled Fog Computing Architecture for 5G and IoT in Underserved Areas

Nosipho Khumalo[#], Olutayo Oyerinde[#], Luzango Mfupe[*]

[#]*School of Electrical and Information Engineering, University of the Witwatersrand*

[1]nkhumalo3@csir.co.za

[2]Olutayo.Oyerinde@wits.ac.za

[*]*Council of Scientific and Industrial Research, South Africa*

[3]lmfupe@csir.co.za

Abstract— **The fifth generation (5G) of wireless network ecosystem which is expected to be deployed in 2020 promises to provide higher speeds, higher capacity and lower latency than the current mobile networks. 5G also promises to revolutionise the telecommunications industry in unprecedented ways by enabling new applications and services that could change the way we live and do things. However, it is not economically feasible to deploy 5G wireless solutions in low Average Revenue Per User (ARPU) areas because of the low revenue potential presented by these regions. One of the contending technologies for low cost computing networks is fog computing, which selectively moves resources and services of computing, storage, control and networking at the edge of the network closer to the users, thereby improving the speed of decision-making, network cost, and the performance of the system. In this paper, we aim to investigate how fog computing can enable cost-efficient solutions in underserved areas to counteract the economic barrier of low ARPU, while providing good quality of service for users that meets key performance requirements. We propose a fog-based architecture that exploits device-to-device communication as well as local computation, storage, and communication as a means to reduce communication costs and thus overcome the financial constraint in 5G deployment. Preliminary simulation results indicate that the proposed architecture shows a twofold improvement in throughput and reduces round-trip delay up to a factor of four.**

Keywords— **Fog computing, Internet of Things (IoT), 5G, Device-to-device (D2D), Machine-to-machine (M2M)**

I. Introduction

Fifth generation (5G) mobile networks are the next step of revolution in mobile communication and will be a fundamental enabler of a better-connected networked society. By supporting new protocols, new users and devices, and a wide range of unprecedented services and applications ranging from enhanced mobile broadband (eMBB) with higher data speeds, massive machine-type communications (mMTC)- also called the Internet of Things (IoT)- and ultra-reliable and low-latency communications (URLLC) applications, 5G promises an evolved platform for information sharing to anyone/anything anytime and anywhere [1], [2]. The new use cases that could be enabled by 5G are illustrated in Fig. 1 along with their key requirements and application areas.

Although 5G promises to revolutionise the telecommunications industry in unparalleled ways, developed markets are likely to reap the benefits of 5G more than their developing counterparts due to the lack of adequate broadband infrastructure in developing areas. In these developing regions, which are mainly very low Average Revenue Per User (ARPU) areas e.g., rural areas with very low densities of populations, and rural/suburban areas not yet connected to the Internet because of economic constraints, mobile network operators (MNOs) are reluctant to extend coverage due to the low return on investment these areas present [3]. The challenge, thereafter, for MNOs becomes the question of "*How to extend coverage to underserved areas with sparse network infrastructure while providing good quality of service for users?*" To this end, fog computing has emerged as a promising solution.

Fog computing, or simply fog, complements and extends the traditional cloud-based models by selectively moving cloud computational and communication functions close to the network edge. In this way, fog can effectively meet the demands of latency-sensitive applications and in turn reduce network bandwidth bottlenecks, thus addressing the needs of future 5G networks. In addition, the distribution and sharing of resources can reduce capital and operational network expenses [4].

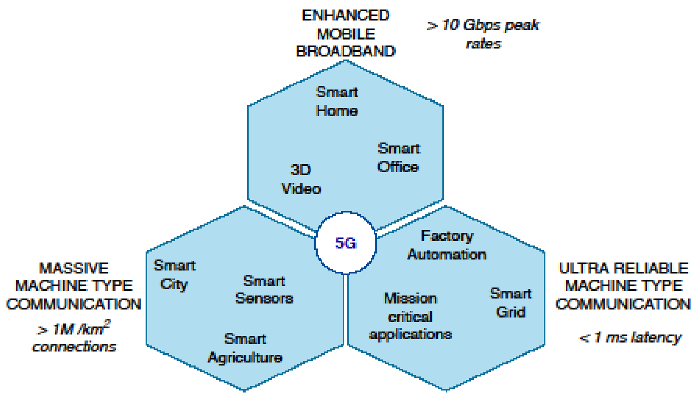

Figure 1: 5G use cases and applications

Fog computing and 5G are both relatively innovative research topics with growing popularity in the research community and telecommunications industry at large. Much attention has been paid to using fog computing as an enabling technology for IoT applications that can deliver 5G capabilities. However, the performance-cost trade-off, which is a significant issue especially in low ARPU regions, is studied to a much lesser extent. There is a lack of studies in the area of utilising fog computing in underserved regions to enable cost efficient 5G solutions.

In this paper, we are motivated to make an effort to offer insight into tackling this issue. Our objective is to investigate how fog computing can enable cost-efficient solutions in underserved areas to counteract the financial obstacle presented by low ARPU while providing good quality service for users that addresses key performance requirements. The main contributions of this paper include:

- We present a network architectural model for IoT applications in 5G networks based on the fog-computing environment, in which device-to-device (D2D) communication is exploited as a means to address the capacity demand of 5G and IoT.
- We measure the performance of the proposed architecture by evaluating the advantages of D2D communication in contrast to the traditional infrastructure mode (two-hop path through the 5G base station).

The rest of the paper can be outlined as follows. After reviewing the related work in Section II, we describe the proposed architecture in Section III. The simulation details and performance evaluation are presented in Section IV. Finally, we conclude the paper and discuss future work.

II. Related Work

The early work in [5] sought to address the issue of cost-efficient wireless Internet connectivity in ARPU regions of developing countries by designing a network architecture. However, the proposed solution is based on the architecture of 4G technology and therefore would be unable to address the ever-increasing capacity and data rate requirements of forthcoming mobile networks. Furthermore, the architecture is prone to the inherent challenges of the conventional cloud-computing approach, thus rendering it ineffective to meet the computational needs of emerging applications like the hyper-connected IoT. In rural areas, the cloud-computing model with a shared cloud data centre in a remote location is not feasible because of bandwidth limitations and network uncertainty along the connection to the public cloud servers.

In more recent literature, there have been some attempts to deploy 5G in underserved communities, as presented by the works in [6]–[8]. However, these approaches are cloud-based and therefore poorly suited to delay sensitive IoT applications.

Very little work has been done in the area of integrating fog computing into the 5G network architecture as a means to aid 5G deployment in underserved areas. However, the application of fog computing into 5G networks for urban areas with adequate telecommunications infrastructure point out its great potential. In addition to low latency, when coupled with the significantly faster transmission speeds of 5G technologies, fog computing is able to provide resource-sensitive applications to end users with limited access to resources [9].

However, to the best of our knowledge, enabling 5G services and applications to underserved areas by integrating fog computing into the 5G network architecture has been studied to a much lesser extent. To this end, a fog-enabled 5G network architecture is proposed. Differing from the abovementioned methods for deploying 5G in rural regions, our proposed approach exploits the fog computing paradigm. The advantages of our method are twofold. Firstly, we emphasise local data processing and storage as a means to counteract the unreachability of the remote SDN-based 5G core network. In our model, all data is processed in the local fog network and transmitted to the 5G core network for long-term storage and batch analytics. Secondly, we utilise device-centric connections in an effort to facilitate Machine-to-machine (M2M) communication between devices and further reduce latency.

III. Proposed System Model

The three-layered architecture of a 5G network with fog computing is considered, shown in Fig. 2, which is composed of the IoT, fog computing, and cloud computing layers. At the top of the hierarchical structure, the cloud-computing layer is composed of multiple high performance cloud virtual machines with intensive computation and storage capabilities. The cloud-computing layer, which can be viewed, theoretically, as having infinite scalability, is also responsible for creating multiple dedicated virtual networks that are then customised to meet the specific needs of applications and services. In this system model, three kinds of generic services are illustrated, namely enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable and low latency communications (URLLC). Each service is characterised by a wide variety of application areas. For example, mMTC use cases include industrial automation and control, smart water supply, and smart agriculture. We focus on smart agriculture. The illustrated fog architecture represents one segment of the agriculture network slice, which operates independently from other farms.

At the level below the cloud computing layer, is the fog computing layer formed by multiple fog networks linked to farms. Therefore, each farm has exactly one fog network composed of several edge devices with compute and storage capacity presented in terms of virtual (abstracted) computing units. These edge devices, which constitute the processing offloading points, are referred to here as fog nodes. To ease the burden on the fronthaul and the 5G core network, and to overcome the challenge of the increasing number of IoT devices and low-latency applications, fog nodes are empowered with capability to deliver network functionalities at the edge. Hence, they are equipped with caching capacity, computing and signal processing capabilities.

Finally, the IoT layer includes many smart and small-sized heterogeneous sensing devices to collect data from the field. For simplicity, any standalone sensor or device with sensing capabilities is referred to as a sensor. There can be different types of sensors in the field, such as location sensors that use signals to determine latitude, longitude, and altitude of any position within required area, optical sensors that use light to measure soil properties, electrochemical sensors, mechanical sensors, etc. The sensors can designed with an extremely low duty cycle that is also related to their density. During the active mode, each sensor gathers data from the physical environment and transmits these data to the fog nodes which are connected to the sensors and the remote 5G core network via wireless technologies and fibre wired links, respectively. The fog nodes then perform processing on the data, such as anomaly detection and offload the data for long-term storage and batch analytics to the SDN-based 5G Core Network.

In this architecture, various user equipment (UEs) in close proximity to each other form a cluster, along with the fog server serving that cluster. Cluster formation is determined according to the geographical locations of IoT devices and fog nodes, therefore mobile UEs can dynamically enter and leave clusters. In our approach, IoT devices sending application data can communicate directly with each other and the fog nodes using D2D links instead of relying solely on cellular links, with the base station handling only control information. This allows inter-device communication to take place without burdening the base station, thus enabling scalability of handling numerous devices interacting with each other. The fog server, which allocates computational resources on demand to users requesting a particular service or task, connects to the 5G radio base station using cellular communication in order to obtain valuable information on the status of the radio network and its users.

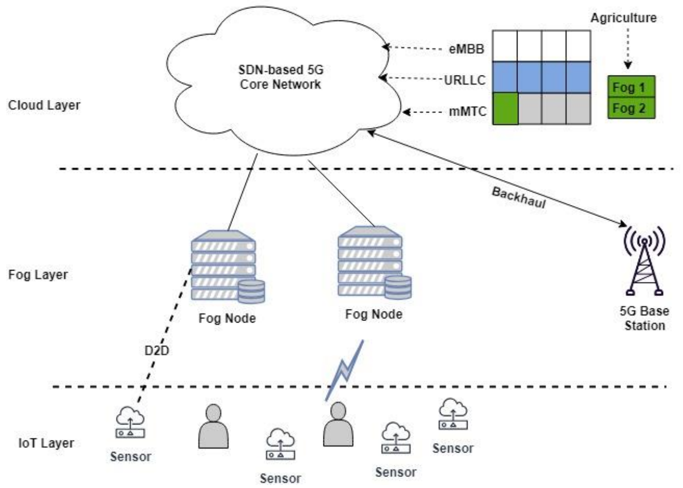

Figure 2: Proposed Network Model

A. *D2D Communication*

In our proposed solution, D2D communication is employed to support efficient communication between fog nodes and smart IoT devices. Using D2D communication, a large amount of data can be transferred quickly between devices in short range. D2D connectivity will make operators more flexible in terms of offloading traffic from the SDN-based 5G core network, increase spectral efficiency and reduce the energy and the cost per bit [10]. In this work, we leverage D2D communication to harness the following benefits:

1. Data and computation offloading

A fog node with a good Internet connectivity can act as a hotspot to which data is offloaded from the BS and from which other nodes may download data using D2D links. Fog nodes with constrained processing power may also offload computation-heavy tasks to nearby more capable fog nodes through D2D links [11].

2. Machine-to-machine (M2M) communication

M2M communication is an enabling technology for Internet of-Things (IoT). It involves independent connectivity and communication among devices ranging from embedded low-power devices to powerful compute-rich devices. D2D connections can be used to establish M2M communication in IoT because of their ability to provide ultra-low latency and thus, real-time responses [12], [13].

IV. Performance Evaluation

To evaluate the performance of the architecture, a communication network scenario is modelled based on a smart agriculture system. In a smart agriculture system, a farm may have hundreds of sensors distributed throughout its fields gathering information about soil moisture, growing conditions, and crop health, and watching for signs of pests. Since we are using multiple sensors like temperature, humidity, PH level, water level, soil nutrient to gather agri-related information, the data is raw and unstructured which can lead to high latency. To overcome this problem, fog computing can be used to process the raw data obtained from end devices locally and can periodically push the data to the central data centre, thus ensuring the most optimal use of network bandwidth. There are also farmers using mobile devices that can generate a request for computing resources to perform a search or prediction.

In the simulation scenario, all the fog nodes are allocated the same amount of computational resources by the SDN-based 5G core network, which is measured in CPU cycles per set of data per second. The resource allocation remains fixed regardless of the workload. Therefore, when a fog node is unable to handle a particular service request and the neighbouring nodes are also occupied, the request is added to a queue to await processing when resources become available. Alternatively, if a neighbouring fog node is available, the request is offloaded to that node for processing.

A. *Design Requirements*

- High speed communication

Due to the number of devices connected in a smart farm ecosystem, there is a need for an efficient M2M communication mechanism that incurs a very small delay. For better user experience, it is necessary for the devices to coordinate and perform in real-time, thus requiring high-speed M2M communication.

- Handling high data volume

Smart farms generate massive volumes of data, particularly due to the number of devices connected. Therefore, the devices processing the data as well as connecting the network should be able to handle such an enormous volume of data.

- Scalability

Fog networks will be required to serve many different requirements and grow with minimal disruption to have sufficient capacity for addressing growing application demands. Fog networks should be architected with scalability in mind, which can have many dimensions including capacity, performance, and reliability.

In this work, we analyzed the network architecture based on the above mentioned scenario using the OMNet++ simulation framework [14] integrated with the INET network simulation package [15]. The simulation parameters, which were configured based on the LTE-A standard, are listed in Table 1. The architecture was simulated on an Intel ® Core (TM) i7 CPU at 2.70 GHz, with 8 Gb of RAM, a Linux Ubuntu 16.04 operating system, OMNeT++ version 5.1, simuLTE version 0.9.1 and INET version 3.6.4

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Channel Parameters	
Maximum transmission power	100mW
Saturation	-90 dBm (minimum signal attenuation threshold)
Alpha	1.0 (minimum path loss co-efficient)
Carrier frequency	2100e+6Hz
Physical layer parameters	
Thermal noise	-104.5 dBm
UE noise	7 dBm
5G base station noise	5 dBm
Sensitivity	-90 dBm
Maximum power transmission	100 mW
Physical header length	0 bit
MAC layer parameters	
Queue size	1MiB
Max bytes per tti	1KiB
Application parameters	
User mobility	Random way point mobility
User speed	1.5 mps
System transport layer protocol	TCP
Simulation time	30 seconds

We simulated one scenario comparing D2D communication with the traditional cellular network, and another comparing scheduling mechanism. Each experiment was conducted ten times and the mean values were calculated. For each scenario, we will show the topology setup and the performance results.

B. Scenario 1

In the first simulation scenario, we simulated the wireless communication between mobile UEs and fog nodes attached to a 5G base station. If at the time of the request the distance between a UE and a fog node is within the maximum allowable distance for D2D communication, a direct communication link is created. Alternatively, the UE and fog node connect to the 5G base station. We consider the maximum D2D range to be 62 m [16].

1) Topology

The topology of this simulation scenario includes three types of nodes, UEs to represent users and fog nodes, and one 5G base station. In the traditional cellular network, we connect the UEs and fog nodes to the 5G base station, however they have no direct connection with each other. Under the D2D communication mode, direct communication between UEs and fog nodes is enabled. Fig. 3 below shows the traditional cellular network topology and the D2D communication network topology.

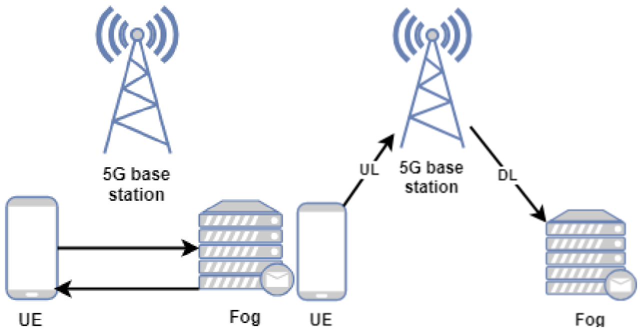

Figure 3: (a) D2D communication (b) Basic access through 5G base station

2) Results

The simulation was executed to measure the round-trip delay time of requested resource assignments by varying the number of users. In the conventional cellular communication, data packets are first uploaded to the 5G base station using the uplink resources, and then routed to the intended destination along the downlink. The average round-trip times measured are shown in Fig. 4. This delay includes the time when the data packet is sent to the fog node until a response is received. The communication delay in fulfilling the request increases with the number of users, but in the D2D communication deployment the times are still significantly lower than that of traditional cellular links. The simulation is conducted where up to 100 users generate a request concurrently and the maximum round-trip times with D2D enabled and traditional mode are 9.67 ms and 38.097 ms respectively. Therefore, offloading data traffic from the 5G base station reduces communication delay and thus improves network performance.

The decision to establish direct links was affected by mobility of users, while the availability of computational resources had a significant impact on the observed latency. Therefore, regarding the latter, there is a need for a dynamic resource allocation method that will optimise each fog node's resources based on the workload of the IoT devices.

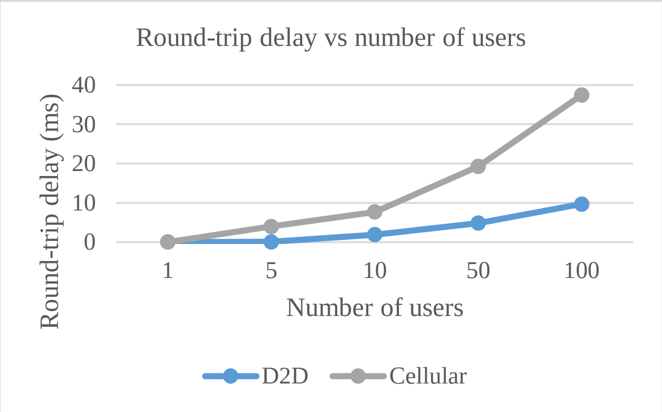

Figure 4: Round-trip times by varying the number of users

We also measured effect of increasing system workload on throughput. In this experiment, the load of the system was varied by the number of UEs. The measured throughput is shown in Fig. 5 below. From this graph, we can see that with D2D communication enabled, the throughput is marginally higher with an increase in the number of UEs.

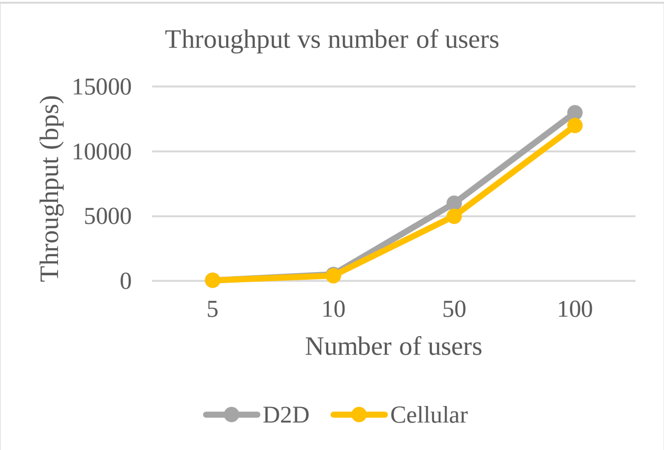

Figure 5: Throughput vs number of UEs

From this scenario, we can see that D2D communication is capable of enhancing network performance. In particular, we observe reduced round-trip delay and improved throughput in the network.

C. Scenario 2

The typical fog paradigm consists of limited fog nodes with limited computing capabilities. Thus, efficient management of resources is essential. As part of this experiment, we compared three well-known packet-scheduling algorithms, namely Proportional Fair Scheduler (PF), MaxCI and Deficit Round Robin (DRR), in a scenario with real-time traffic.

1) Topology

The topology of this simulation scenario includes UEs to represent users and fog nodes, one 5G base station and a 5G core network. The UEs and fog nodes in close proximity communicate directly between each other using peer-to-peer links, while longer-range communication relies on the radio network infrastructure. The fog network is connected to the 5G core network through point-to-point links with a data rate of 10 Gbps and a propagation delay of 2 ms. The abovementioned topology is illustrated in Fig. 6 below.

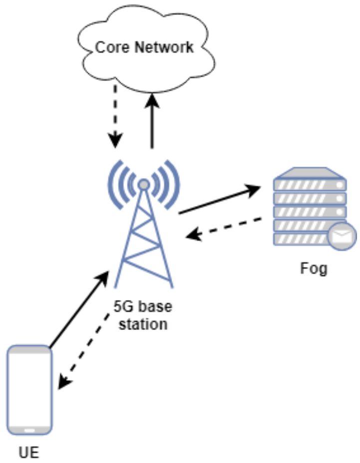

Figure 6: Topology of scheduler scenario

2) Results

As part of this simulation scenario, the system load is varied by packet size and the number of UEs to measure the downlink throughput for real-time data. Fig. 7 below shows the measured throughput for a fixed number of UEs (five) and varied packet sizes (30, 40, 50, and 60 bytes). From the graph, it can be noticed that the behaviour of throughput performance is similar for DRR and PF. Throughput increases with the increase in packet size, with DRR achieving higher measurements than PF. Using MaxCI, maximum throughput is achieved with medium-sized packets (50 bytes) and deteriorates with larger packet sizes. From this experiment, it is clear that DRR outperforms both MaxCI and PF in a system with a varied load of packet sizes.

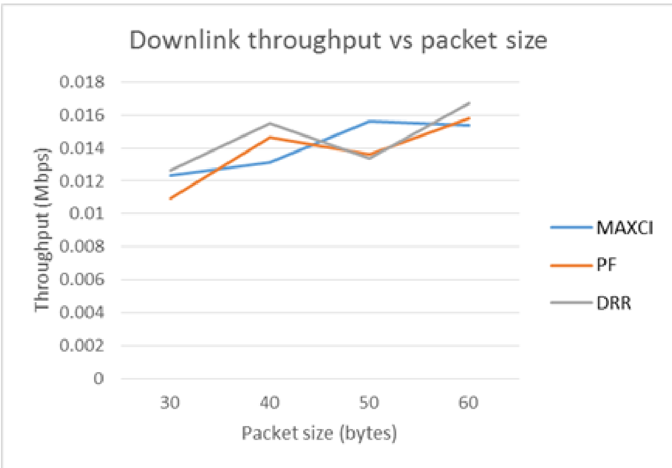

Figure 7: Downlink throughput by varying packet size

We then measured the effect of varying system load by means of the number of UEs, as illustrated in Fig. 8. From this graph, we can see that the throughput performance of all three algorithms is similar for small and medium number of

UEs, however DRR is significantly inferior for a system load with a large number of UEs. Therefore, in a system with many UEs, MaxCI and PF perform well.

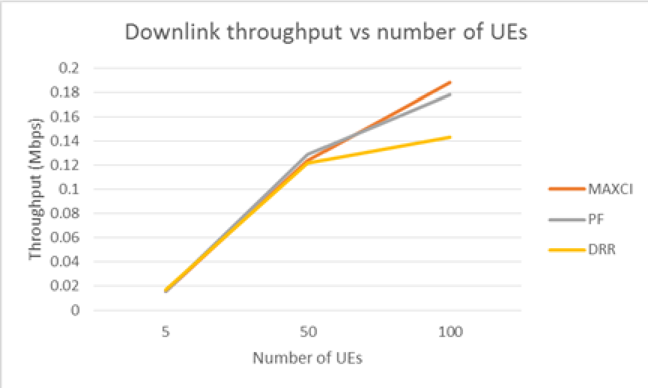

Figure 8: Downlink throughput by varying number of UEs

From this scenario, we can see that for a large number of UEs, MaxCI yields superior performance gains in terms of downlink throughput for real-time data, while DRR outperforms both MaxCI and PF in a system workload of big packet sizes. While none of the algorithms studied achieved good performance under high system load conditions (large packet sizes and high number of UEs), PF is the most consistent algorithm because it achieves fair results under both conditions. Nonetheless, there is a need to employ intelligent and adaptive scheduling capabilities.

V. Conclusions

In this work, we investigated how fog computing can enable cost-efficient solutions in underserved areas to counter the financial drawback of low ARPU while providing good quality service for users that meets key requirements. We propose a fog computing based architecture that capitilises on D2D communication and local computation, storage, communication, control, and decision making as a means to reduce communication costs. We observed that the proposed architecture enhances network performance with regards to improved throughput and reduced round-trip delay. Furthermore, we witness the benefits of D2D communication over the conventional cellular infrastructure mode. However, further improvements should be considered.

The key findings of this paper can be summarised as:

- Any approach of offloading tasks to neighbouring nodes needs to account for the uncertainty of fog node neighbour locations.
- The user location and the mobility model used for each node affect the round-trip delay. Thus, more processing nodes distributed closer to the users can enhance system performance.
- The performance-cost trade-off can be affected by the number of devices (including sensors and fog nodes) and the amount of tasks assigned for each fog node. The challenge is finding the optimal number of fog nodes and sensors in the network.
- There is a need for packet scheduling algorithms that can allocate resources dynamically. In this domain, machine learning based algorithms and policies are a promising candidate.

Future Work

As part of our future work, we are motivated to extend the architecture by integrating machine-learning techniques to optimize network operations. In particular, it would be interesting to focus on the dynamic allocation of computational resources from the SDN-based 5G core to the fog network

References

[1] S. Kitanov, E. Monteiro, and T. Janevski, "5G and the Fog—Survey of related technologies and research directions," presented at the Electrotechnical Conference (MELECON), 2016 18th Mediterranean, 2016, pp. 1–6.

[2] "IEEE 5G and Beyond Technology Roadmap White Paper," IEEE, Oct. 2017.

[3] D. Pouillot, "The Dynamics of Broadband Markets in Europe: Realizing the 2020 Digital Agenda," *Communications & Strategies*, no. 99, p. 183, 2015.

[4] OpenFog Consortium Architecture Working Group, "OpenFog reference architecture for fog computing," *OPFRA001*, vol. 20817, p. 162, 2017.

[5] Y. Gourhant, E. Lukashova, M. R. Sama, S. A. Wahed, D. Meddour, and D. P. Venmani, "Low-cost wireless network architecture for developing countries," in *2014 14th International Conference on Innovations for Community Services (I4CS)*, 2014, pp. 16–21.

[6] A. Markhasin, V. Belenky, V. Drozdova, A. Loshkarev, and I. Svinarev, "Cost-effective ubiquitous IoT/M2M/H2H 5G communications for rural and remote areas," in *2016 International Conference on Information Science and Communications Technologies (ICISCT)*, 2016, pp. 1–8.

[7] A. Karlsson, O. Al-Saadeh, A. Gusarov, R. V. R. Challa, S. Tombaz, and K. W. Sung, "Energy-efficient 5G deployment in rural areas," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016, pp. 1–7.

[8] A. Markhasin, "Fundamentals of the extremely green, flexible, and profitable 5G M2M ubiquitous communications for remote e-healthcare and other social e-Applications," in *2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, 2017, pp. 292–297.

[9] M. Tao, K. Ota, and M. Dong, "Foud: Integrating Fog and Cloud for 5G-Enabled V2G Networks," *IEEE Network*, vol. 31, no. 2, pp. 8–13, Mar. 2017.

[10] U. N. Kar and D. K. Sanyal, "An overview of device-to-device communication in cellular networks," *ICT Express*, vol. 4, no. 4, pp. 203–208, Dec. 2018.

[11] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 104–112, Apr. 2013.

[12] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 86–92, May 2014.

[13] O. Bello and S. Zeadally, "Intelligent Device-to-Device Communication in the Internet of Things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1172–1182, Sep. 2016.

[14] A. Varga, "OMNeT++," in *Modeling and Tools for Network Simulation*, K. Wehrle, M. Güneş, and J. Gross, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 35–59.

[15] "INET Framework - INET Framework." [Online]. Available: https://inet.omnetpp.org/. [Accessed: 10-May-2019].

[16] C. Singhal and S. De, *Resource Allocation in Next-Generation Broadband Wireless Access Networks*. IGI Global, 2001.

Nosipho Khumalo received a BSc (Honours) degree in Computer Sciences from the University of Cape Town, South Africa, in 2017. She is currently pursuing her MSc (Eng) degree in Electrical and Information Engineering from the University of the Witwatersrand. She is currently involved in developing Research and Development applications at the Council for Scientific and Industrial Research (CSIR) under the Future Wireless Networks research group. Her interests include, but not limited to, Edge Computing, the Internet of Things, and ICT for Development.