

Optimizing Location of Edge Clouds with Baseband Units in Cloud Radio Access Network

Jackline Nakazibwe*, Jonathan Serugunda*, Roseline Akol*, Stephen Mwanje*

*Department of Electrical and Computer Engineering
Makerere University Kampala,
Uganda.

E-mail: jac14lyn@gmail.com, {jserugunda, rnakol}@cedat.mak.ac.ug
, stephen.mwanje@gmail.com.

Abstract—Cloud radio access network architecture has become increasingly important in meeting the growing demand of high data rate services as well as managing interference among base station sites. However, this architecture is associated with high fronthaul link latencies that result into increased overall network latencies. This paper proposes a placement method for baseband units so as to lower the fronthaul link latencies. This is achieved by formulating the problem as a nonlinear optimization problem that is solved by fuzzy c-means clustering and a heuristic genetic based algorithm. The paper further investigates the user response time comparing it with the scenario before optimizing location of baseband units in edge clouds. Simulation results show that the proposed scheme greatly reduces overall fronthaul link latencies and cost of ownership in cloud radio access networks. We also show that there is an optimal number of edge clouds with baseband units for different cloud radio access network sizes. On the side of the users, the response time is greatly reduced when the baseband units are optimally placed.

Keywords—Baseband unit, cloud radio access network, edge cloud, latency, optimization.

I. INTRODUCTION

As mobile network users increasingly demand mobile network services at extremely high data rates and ultra-low latencies, mobile network operators (MNO) have deployed the cloud radio access network (CRAN) to meet the user demands at low cost of ownership [1]. In CRAN, several baseband units (BBUs) are pooled at a location far away from the remote radio heads (RRHs) and centralized in one shared location called the BBU pool. CRAN has many advantages such as improved spectral efficiency, network capacity and reduced total cost of ownership (TCO). However, the fronthaul link separating the RRH functions from the BBU functions can be long up to 40 km which affects RRHs-BBU synchronization and leads to high fronthaul link latencies thus increased overall network latencies [2]–[4]. This is a challenge to applications that require near zero latencies as will be required in 5G [5]–[7]

Latency is an important metric in telecommunication systems because it determines the user quality of experience (QoE). According to 3GPP specifications, it is an industry standard that a base station (BS) needs to prepare a hybrid automatic repeat request (HARQ) acknowledgment (ACK) / non-acknowledgment (NACK) within 3 ms: decode uplink (UL) data, prepare ACK/NACK and create a downlink (DL)

frame with ACK/NACK. Part of the 3ms, 100-200 μ s delay budget needs to be spent on fronthaul link one-way delay otherwise throughput will be lowered [2], [8], [9].

Position of the BBU pool within CRAN is crucial as it determines the length of the fronthaul link which in turn determines fronthaul link latency. In [10], the authors provided a BBU placement heuristic named SWAN for placing BBUs in the network collocated with macro cells. An integer linear programming (ILP) based algorithm was introduced to solve the placement problem for small networks and a BBU placement heuristic for larger networks. The results showed that fewer BBU pool resources are needed in cases of long fronthaul links. The authors focused on minimizing the number of substrate links and microwave interfaces and did not minimize distance between the BBU and the RRH.

For distributed network nodes, research in [11] focused on determining subsets of available distributed network nodes which can be used to host BBU servers and to which of the available BBUs each RRH should be assigned so as to minimize the server and fronthaul link setup costs as well as minimizing the delay of function processing at the selected nodes. The problem was formulated as a binary ILP and a tabu search method was used for placement of BBUs in the network. The algorithms proposed in this paper do not consider the links between physical nodes, and consequently, the link delays for transferring a given function from one node to another yet it would be interesting to evaluate the effect of link delays. The algorithms considered in this paper are also static in that after the initial mapping and scheduling, the functions do not apply for changing network conditions.

A heuristic optimization algorithm based on a greedy scheme was proposed in [12] and used to place BBUs with an aim of minimizing the number of BBU hotels needed. Results therein show that BBU consolidation factor rapidly decays when the maximum fronthaul length increases, because higher values of fronthaul link lengths allow BBUs to be placed at highest nodes closer to the point of presence (PoP), thus enabling a bigger number of cell sites to be controlled by a single hotel site. Results clearly indicate that BBU consolidation into fewer hotel nodes is the key principle for reducing the overall cost/energy while placing the BBU pools in CRAN.

While determining locations for BBU pool, it is necessary that setup costs are maintained to as low as possible. This was

achieved in [13] by assuming that the system cost incurred by a C-RAN is the total virtual machine cost in the BBU pool and the power consumption incurred by the RRHs. They minimized the system cost by optimizing the number of active virtual machines (VM) in the BBU pool and the beamformers at the RRHs. The problem was formulated as a mixed-integer nonlinear programming by minimizing the overall cost, which consisted of two parts: the cloud processing cost (in the BBU pool) with respect to the number of active VMs, and the wireless transmission cost (in the fronthaul and RRHs) with respect to the transmit beamformers. They proposed a price adjusting algorithm and used integer search and joint optimization to find out the optimal number of virtual machines.

An algorithm for reducing total costs of BBU pool planning in CRAN was proposed in [14]. The authors formulated the problem as an NP-hard integer programming optimization task and addressed it using a local search algorithm while considering BBU processing capacities, traffic demands of RRHs and the signal synchronization between RRHs and BBUs. Numerical results therein show that less costs are incurred when BBU pool planning is done than when it is not done. The proposed algorithm converges quickly and always yields a performance-guaranteed solution to the BBU pool planning. The authors aimed at minimizing the deployment costs of the BBU pool. Numerical results also showed that the proposed scheme reduces the deployment cost of the BBU pools planning in the C-RAN significantly by about 27% and 53% for the cases of 20 and 40 candidate BBU pools respectively.

As discussed above, BBU placement problem has been tackled by different researchers with varying objectives and different approaches to achieve their objectives. No work has been done so far to place BBUs in optimally placed ECs with an aim of minimizing fronthaul link latencies. In this work, we aim to optimally place ECs running virtualized BBUs so as to reduce fronthaul link latencies. EC computing infrastructure is a resource that extends the capabilities of a data center-based cloud service to the network edge [15], [16]. Furthermore, optimization of the ECs running the virtualized BBU pools is necessary to minimize costs.

The main contributions of this work can be summarized as follows: Formulation and solving of an EC location optimization problem using both fuzzy c-means clustering and a heuristic genetic based algorithm. Moreover, simulation results verify the effectiveness of this scheme.

The rest of the paper is organized as follows. In section II network model is presented. Section III presents the simulation results and discussion. Finally, the conclusion is given in section IV.

II. NETWORK MODEL

The network under consideration is shown in figure 1 and it covers an area of 2 km by 2 km. It consists of RRHs generated as a Poisson point process (ϕ) with homogeneous intensity λ_r of 0.000025 and candidate EC locations generated as a Poisson point process (Ω) with homogeneous intensity λ_e of 0.000005. Poisson point process (PPP) was used in this case because in reality the RRHs and EC positions are not uniform

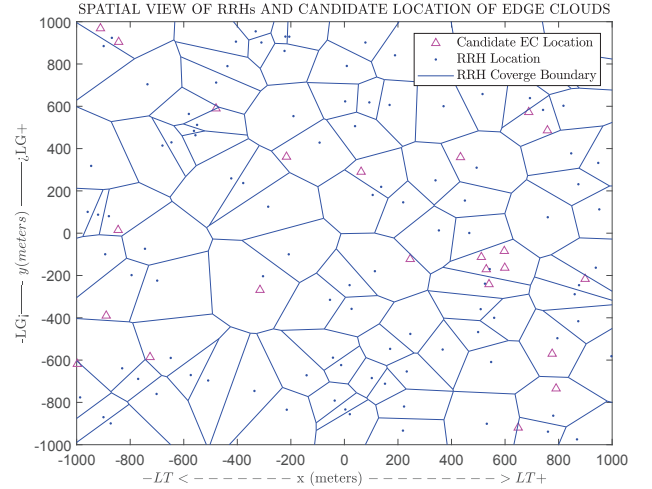


Fig. 1: Spatial view of RRHs and candidate location of ECs.

but random and follow a poisson distribution. PPP thus uses the RRH density, candidate EC density and total network area to determine the location of the RRHs and candidate ECs. Cell shapes and sizes are generated by voronoi algorithm which shows the real-life irregularity of network cells.

From the network shown in figure 1, the task is to select a subset of the candidate EC locations $N_0 \subset N$. Where, N is the set of n ECs, to which m RRHs can connect to, under stringent constraints of fronthaul latency and EC capacity. Let M_i represent the RRH locations in the network, D_i be the Euclidean distance between $N_0(x_0, y_0)$ and $M_i(x_i, y_i)$ and $f(D) \rightarrow f(x, y)$ be the function such that: $f(D) = \sum_{i=1}^m D_i$. The search for the optimal location of EC is a constrained non-linear optimization problem formulated as follows;

$$\min_{x,y} f(x, y) = \sum_{i=1}^m ((x - x_i)^2 + (y - y_i)^2)^{\frac{1}{2}}, \quad (1)$$

Subject to :

$$C1 : \sum_{i=1}^m C_f \leq C_{EC}, \quad (2)$$

$$C2 : ((x - x_i)^2 + (y - y_i)^2) \leq \left(\frac{1}{2} \frac{L_{max}}{K}\right)^2, \quad (3)$$

where C_{EC} is the Edge cloud capacity, C_f is the fronthaul link capacity, L_{max} is the maximum allowable fronthaul link latency, K is the proportionality constant (propagation speed of signals through fiber: $2 * 10^8 m/s$, x_c, y_c is the center of the EC coverage area.

The objective function in (1) minimizes the sum of fronthaul link lengths between a set of RRHs and a particular EC. Constraint C1 in (2) ensures that capacity of each EC is not exceeded. I.e. the total capacity over all fronthaul links converging to an EC should not exceed capacity of that particular EC. Thus, if each RRH site provides service to user equipment whose combined average processing requirement is ∂_m Mbps, then; $\partial_m \leq C_f$ and C1 has to be fulfilled. Constraint C2 in (3) ensures that each EC covers a circular area of RRHs whose diameter is equivalent to the maximum allowable fronthaul link

length D_{max} at which L_{max} is obtained.

A. Placement Procedure

In order to determine the optimal locations two techniques are applied; firstly, fuzzy c-means clustering algorithm is applied to group RRHs in clusters. RRHs close to each other are likely to be grouped in the same cluster, the closer a RRH is to a cluster center, and the higher its membership to that cluster thus also partially minimizing the fronthaul link latency between EC and the RRHs. Secondly, genetic algorithm is applied to select the optimal candidate EC location for each cluster. In cases when optimal location is not among the candidates, it determines the candidate closest to the optimal value and considers that as the optimal EC location in the cluster. A snap shot of the optimal locations determined is shown in figure 3.

B. System parameters

- User response time: Is the time delay between the time a UE submits a request to the network and the time of getting the response given by

$$t_{user} = t_r + t_{RRH,BBU} + t_{cons}, \quad (4)$$

where t_{user} is the user response time, t_r is transmission delay between UE and the RRH, $t_{RRH,BBU}$ is the fronthaul link delay, t_{cons} is the BBU and RRH processing delay.

- Cost of ownership: The cost of the EC and the fronthaul link is given by

$$C_O = (N) \times (Cost_{EC}) + L_f \times Cost_{f/km}, \quad (5)$$

where N is number of base band unit edge clouds in the network, $Cost_{EC}$ is cost of each EC in USD, L_f is length of the Fronthaul link in km, $Cost_{f/km}$ is cost of 1km of fiber

C. Flowchart

The flow chart is shown in figure 2.

D. Pseudo code

The steps followed in the proposed algorithms, are listed in the pseudo code presented in table 1.

E. Determined optimal locations

Optimal locations determined by the heuristic genetic based algorithm are shown in figure 3. RRHs are connected to the BBUs in optimally located ECs. Moreover, each cluster consists of one serving EC. Other ECs have been left out by the genetic algorithm since their locations are not optimal.

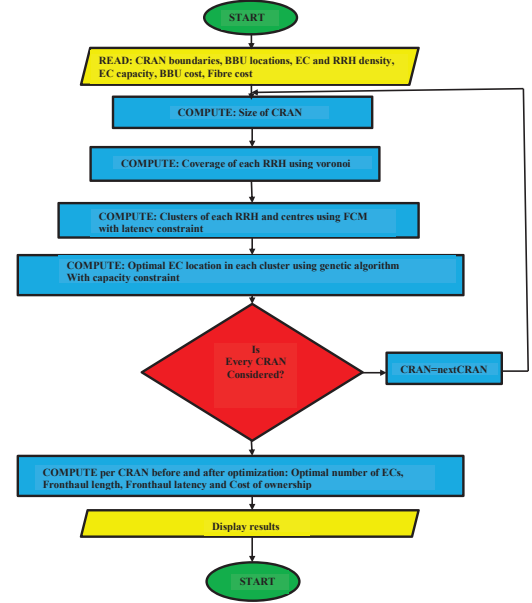


Fig. 2: Simulation flow chart.

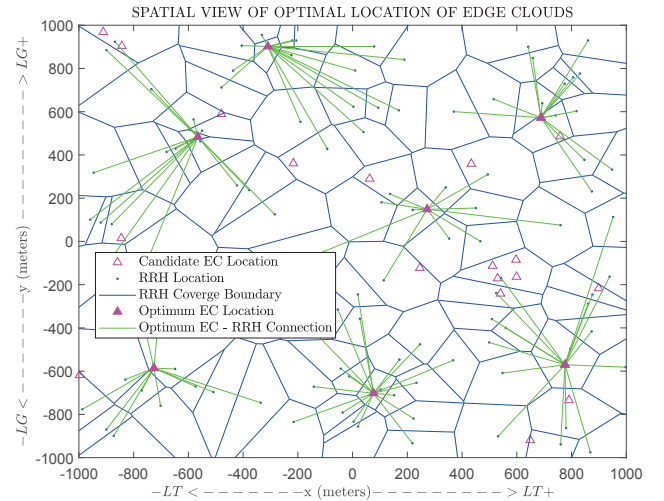


Fig. 3: Spatial view of optimal location of ECs.

To analyze the user response time, one cluster served by one EC is randomly chosen and users are randomly placed in it, then one cell in this cluster (the active cell) is randomly selected as shown in figure 4.

```

1 Declare  $\{\mathbf{CRAN}\} : \mathbf{cran} \in \mathbf{CRAN}$ 
2 Declare  $\{\mathbf{H}\} : \mathbf{m}$  RRH locations in Euclidean space  $\mathbb{R}^2$  as PPP ( $\varphi$ ) of homogeneous intensity  $\lambda_r$ 
3 Declare  $\{\mathbf{E}\} : \mathbf{n}$  EC locations in Euclidean space  $\mathbb{R}^2$  as PPP ( $\Omega$ ) of homogeneous intensity  $\lambda_e$ 
4  $\forall \mathbf{H}$  define capacity  $\mathbf{C}$ 
5 Define latency radius  $I_r$ 
6  $\forall \mathbf{H}$  do: voronoi ( $\mathbf{H}$ ) to obtain coverage
7  $\forall \mathbf{H}$  do: fuzzy CMC ( $\mathbf{H}, \mathbf{C}$ ) to obtain  $\{\mathbf{CLUSTERS}\}$  and corresponding  $\{\mathbf{CENTERS}\}$ 
8 Declare:  $cluster \in \mathbf{CLUSTERS}, center \in \mathbf{CENTERS}, e \in E$ 
9  $\forall H: \mathbf{H}_{cluster} \leftarrow \{cluster \cap \mathbf{H}\}$ 
10 do GA on cluster constrained by latency, circular at center
11  $\forall \mathbf{E}$  compute: Euclidean distance  $\mathbf{L}(e, center)$ 
12  $l \in \mathbf{L}$ : if  $\forall \mathbf{L}, l > 2\mathbf{L}_r$ 
13  $cluster\_no\_ec \leftarrow TRUE$ 
14 end if
15  $\mathbf{FMIN}_{solution} \leftarrow \mathbf{GA}$ 
16 If  $cluster\_no\_ec = TRUE$ 
17  $e_0 \leftarrow \mathbf{FMIN}_{solution}$ 
18 end if
19 if  $cluster\_no\_ec \sim = TRUE$ 
20  $\forall \mathbf{E}$  compute: Euclidean distance  $D(e, \mathbf{FMIN}_{solution})$ 
21 end if
22  $\forall \mathbf{D}: d_0 \leftarrow \mathbf{minimum}(\mathbf{D})$ 
23 Display results
24 End
    
```

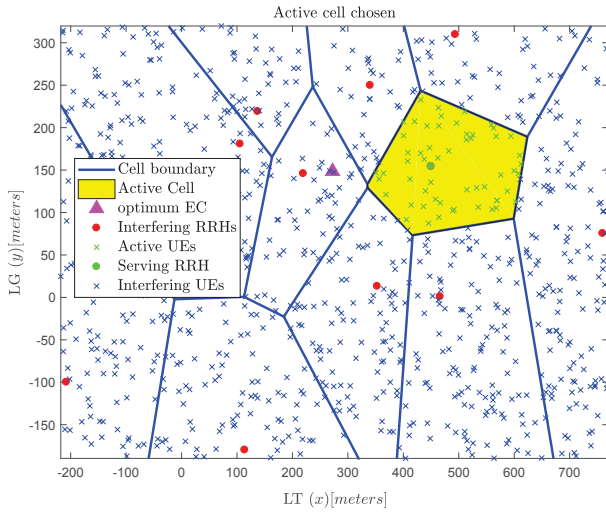


Fig. 4: Cluster under consideration.

III. SIMULATION RESULTS AND DISCUSSION

In this section, some simulation results are presented for two scenarios, namely: Before and after optimizing the location of ECs that run the BBUs. Figure 5 shows how the overall fronthaul link length varies with the size of the CRAN. From this figure, it can be observed that for both scenarios, the overall fronthaul link length increases with increase in the CRAN size. This can be explained that as the size of CRAN increases, more RRHs are needed because coverage of a RRH is limited. Thus, increase in need for more ECs because of their limited capacity. This causes the number of fronthaul links to increase and likewise the overall fronthaul link lengths. There is a big gap in fronthaul link length between the two scenarios considered. This is because ECs bring BBU functionalities closer to the RRHs, thus shorter fronthaul links. Considering CRAN of size $0.4 \times 10^7 m^2$, before ECs are used the fronthaul link length is $0.8 \times 10^5 m$ and after, the value is $0.1 \times 10^5 m$.

With no proper planning of BBU placements, some far away RRHs will connect to ECs that are distant which increases the fronthaul link lengths.

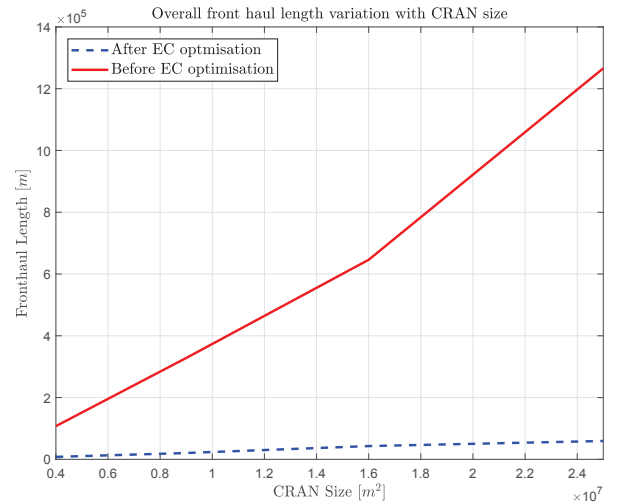


Fig. 5: Overall Fronthaul length variation with CRAN size.

Figure 6 depicts how the overall fronthaul link latency varies with CRAN size. It is observed that increase in the CRAN size increases the overall fronthaul link latency. This is because big CRAN sizes require more fronthaul links and lengths as shown in figure 5 thus increase in overall fronthaul link latency in both scenarios. After ECs have been optimally placed to run the BBU functions in the network, we observe that fronthaul link latency is much less compared to when BBUs are not in optimally placed ECs. This is because with ECs, the shorter fronthaul link lengths are associated with shorter round-trip time of signals thus low latency values. Ie. Considering CRAN of size $0.4 \times 10^7 m^2$, before ECs are used the fronthaul link latency is 0.4 ms and after the value is 0.05

ms.

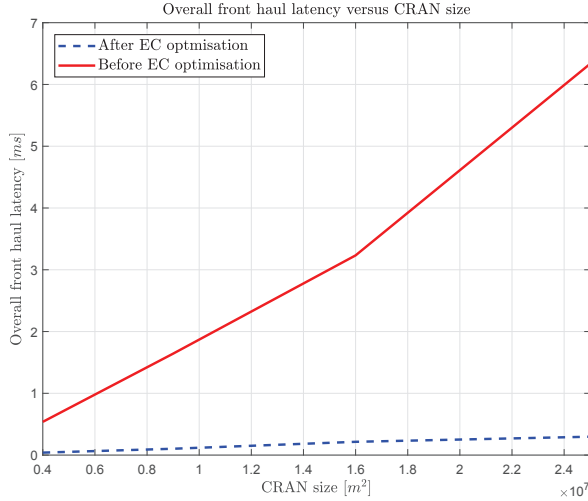


Fig. 6: Overall Fronthaul latency versus CRAN size.

The effect of how cost of ownership varies with CRAN size is shown in figure 7 shows. Herein, the cost of BBU pool and fiber are considered. The BBU cost is assumed to be 1000 USD and cost of fiber per km is assumed to be 1500 USD, installation costs are not considered. In both scenarios, we observe that increase in CRAN size increases cost of ownership. This is because bigger CRAN sizes require more ECs and fronthaul links as shown previously thus more BBUs and fiber. However, we observe that after using BBUs in optimally placed ECs, the cost of ownership greatly reduces because BBUs that run ECs are cheaper than those at centralized locations. Furthermore, the shorter fronthaul links require shorter lower cost fiber. Considering CRAN of size $0.4 \times 10^7 m^2$, before ECs are used the CO is 1.1×10^5 USD and after the value is 0.1×10^5 USD.

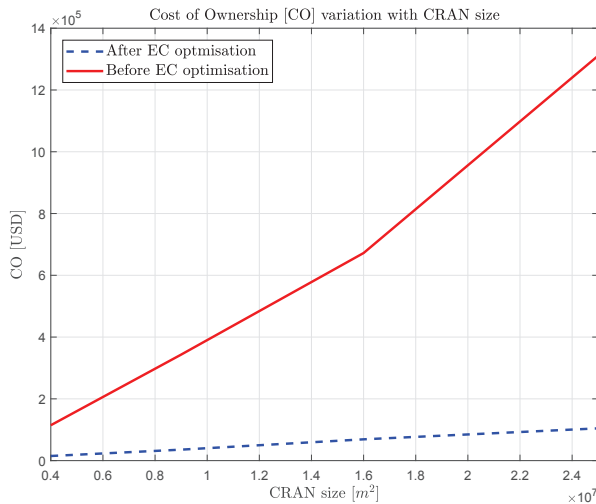


Fig. 7: Cost of ownership variation with CRAN size.

In figure 8 we study how optimal number of ECs varies with CRAN size. It can be observed that optimal number of ECs increases as the size of the CRAN increases. This is because as CRAN size increases, the number of RRHs required to provide coverage of the entire network increases and since capacity of the ECs is fixed, more of them to which the RRHs

will connect are required. The optimal number of edge clouds is determined by the heuristic based genetic algorithm and it is observed that even in a very small CRAN we still need an EC as RRHs have to connect to it. For example, 7 ECs are required for a 0.4×10^7 CRAN size with 100 RRHs.

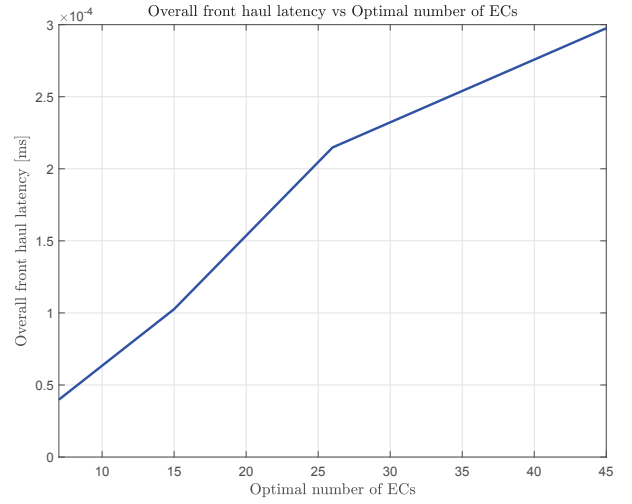


Fig. 8: Optimal number of ECs versus CRAN size.

In figure 9, user response time experienced versus distance between the UE and the BBU is presented. It can be observed that for smaller distances between the BBU and UE, user response times are smaller; an indication of improved user QoE since the user spends less time waiting for a response compared to when the BBU is far away from the UE, i.e. User response time of $4 \mu s$ is obtained when the BBU-UE distance is 1200 m and response time of $6 \mu s$ is obtained when the BBU-RRH distance is 1800 m.

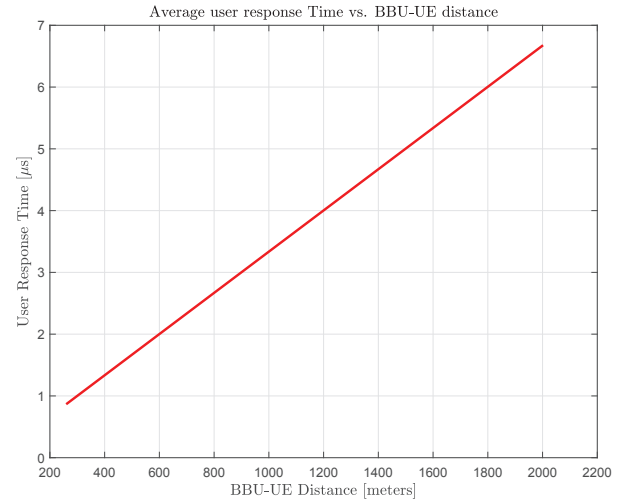


Fig. 9: User response time versus BBU-UE distance.

IV. CONCLUSION

In this paper, a nonlinear edge cloud placement optimization problem has been formulated and solved by fuzzy c-means clustering and heuristic genetic based algorithms. Results show that there is a greater reduction in the overall fronthaul link latency and cost of ownership when baseband unit functions

are run in optimally placed edge clouds. We also observed that user response time greatly reduces compared to when baseband units are not placed in optimally placed edge clouds. When the location of edge clouds running baseband units is not optimized, the results showed that fronthaul link latencies, cost of ownership and user response time remain higher. Therefore, we intend to analyze the utilization efficiency of baseband units and edge clouds in the cloud radio access network as an extension of this work.

ACKNOWLEDGMENT

This work was funded by Africa Center of Excellence in Materials, Product Development and Nanotechnology (MAPRONANO ACE).

REFERENCES

- [1] D. Sabella et al., "RAN as a service: Challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network," in *Proc. Future Netw. & Mobile Summit*, Lisboa, Portugal, 2013, pp. 1-8.
- [2] A. Checko et al., Cloud Radio Access Network Architecture. Towards 5G Mobile Networks, *Tech. Univ.*, Denmark, pp. 1-170, June 2016.
- [3] R. I. Tinini, L. C. M. Reis, D. M. Batista, G. B. Figueiredo, M. Tornatore and B. Mukherjee, "Optimal Placement of Virtualized BBU Processing in Hybrid Cloud-Fog RAN over TWDM-PON," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM 2017)*, Singapore, 2017, pp. 1-6.
- [4] O. Vierimaa, "Cost Modeling of Cloud-Based Radio Access Network," 2017.
- [5] A. MAURYA, "5G Technology And Its Key Drivers: Cloud Computing And Cloud Radio Access Networks," Ghaziabad, India, 2015.
- [6] M. Peng, Y. Li, Z. Zhao, C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6-14, Mar./Apr. 2015.
- [7] Q. Liu, T. Han, N. Ansari and G. Wu, "On Designing Energy-Efficient Heterogeneous Cloud Radio Access Networks," *IEEE Trans. on Green Commun. and Netw.*, vol. 2, no. 3, pp. 721-734, Sept. 2018.
- [8] N. Nikaiein, "Processing radio access network functions in the Cloud: Critical issues and modeling", MCSMS, 2015.
- [9] A. Marotta, Optimisation of Radio Access Network Cloud Architectures Deployment in LTE-Advanced, July 2015.
- [10] R. Riggio, D. Harutyunyan, A. Bradai, S. Kuklinski and T. Ahmed, "SWAN: Base-band units placement over reconfigurable wireless front-hauls," in *Proc. 12th Int. Conf. on Netw. and Service Manag. (CNSM)*, Montreal, QC, Canada, 2016, pp. 28-36.
- [11] R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten, F. De Turck, S. Davy, Placement and scheduling of functions in network function virtualization, 2015.
- [12] N. Carapellese, "BaseBand Unit Hotelling Architectures for Fixed-Mobile Converged Next-Generation Access and Aggregation Networks," 2015.
- [13] J. Tang, W. P. Tay, T. Q. S. Quek and B. Liang, "System Cost Minimization in Cloud RAN With Limited Fronthaul Capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3371-3384, May 2017.
- [14] S. Xu and S. Wang, "Efficient Algorithm for Baseband Unit Pool Planning in Cloud Radio Access Networks," in *IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, Nanjing, China, 2016, pp. 1-5.
- [15] L. P. Garcia et al., "Edge-centric computing: Vision and challenges," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 3, pp. 37-42, 2015.
- [16] A. M. Khan and F. Freitag, "On Edge Cloud Service Provision with Distributed Home Servers," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. and Sci. (CloudCom)*, Hong Kong, China, 2017, pp. 223-226.