Energy Efficient Multi-Tenant Resource Slicing in Virtualized Multi-Access Edge Computing

Yan Kyaw Tun, Madyan Alsenwi, Shashi Raj Pandey, Chit Wutyee Zaw, and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, 171-04, Republic of Korea
Email: {ykyawtun7, malsenwi, shashiraj, cwyzaw, cshong}@khu.ac.kr

***Abstract*—With the help of multi-access edge computing (MEC) system, traditional mobile network operators (MNOs) can provide various services to their mobile users with the minimum delay by installing micro-datacenters at the base stations (BSs)(i.e., at the edge of the radio access network). However, the capital and operational expenditures become significant challenge for the MNOs. Fortunately, multiple MNOs can coexist on the same infrastructure and share the network resources with the help of upcoming technologies such as network virtualization (i.e., network slicing) and software defined networking (SDN). In this work, we introduce a virtualized MEC system in which an infrastructure provider (InP) deploys a BS integrated with a micro-datacenter and owns the wireless network resource i.e., bandwidth. Then, InP creates the virtual network by slicing its network resources including bandwidth and the computation resource of the MEC server, and shares these resource slices to multiple virtual network operators (MVNOs) where MVNOs provide specific services to their mobile users. To solve our proposed problem, we first decompose the original problem into two subproblems. Then, we apply the Karush-Kuhn-Tucker (KKT) conditions to solve each subproblem. Moreover, simulation results prove that our proposed algorithm for joint communication and computation resources sharing outperforms the existing schemes.**

***Index Terms*—Multi-access edge computing, virtualized wireless networks, communication and computation resource slicing, KKT conditions.**

I. Introduction

Computation intensive and delay sensitive applications such as online gaming, high-resolution video streaming, augmented reality (AR), virtual reality (VR), and autonomous driving etc, are going to be provided in the future wireless networks. However, mobile devices have limited power and computation capacity to run those applications. Recently, the deployment multi-access edge computing (MEC) servers at the macro base stations (MBSs), small-cell base stations (SBSs), and access points (APs) brings the computation services to the edge of the radio access networks i.e., near to the mobile users. In other words, MEC system can significantly reduce the service delay experienced by the mobile users [1], [2].

Recently, the new release of 3GPP for 5G introduced the new technology called wireless network virtualization i.e., wireless network slicing, that can logically decouple the current cellular networks into two entities: 1) infrastructure providers (InPs) and 2) mobile virtual network operators (MVNOs), where InPs deploy the infrastructure and possess the wireless resources i.e., bandwidth, and MVNOs are leasing these resources from the InPs and provide specific services to their mobile users with the help of software defined networking (SDN) [3], [4].

In this work, we combine these promising technologies for mobile communication industry and propose a virtualized multi-access edge computing (MEC) system.

A. Research Contribution

The major research contributions of this work are summarized as follows:

- Firstly, we propose the joint communication and computation resource allocation problem in the virtualized MEC system that minimizes the energy consumption of both mobile users of MVNOs and the base station. Then, we apply KKT conditions and solve two subproblems iteratively.
- Simulation results demonstrate that our proposed algorithm outperforms the equal resource allocation scheme for both communication and computation resources.

The rest of this paper is organized as follows: In section II, the related works are organized. Then, the proposed system model and problem formulation are described in section III. In section IV, we describe the solution approach and simulation results are shown in section V. Finally, section VI concludes the paper.

II. Related Works

1) Multi-access computing: [5] has presented energy-efficient resource allocation for MEC sytem in a single cell scenario and the computation resource allocation for MEC-based AR applications were proposed in [6]. However, both of the aforementioned research works consider a single base station for the simplification. In [7], [8], [9], authors has studied an efficient task offloading schemes in the heterogeneous wireless networks. Both of these works considered to minimize the energy consumption of the mobile users and did not count the energy consumption of BS.

2) Wireless Network Slicing: [10], [11] studied downlink power allocation and interference mitigation in the multi-cells virtualized wireless network but the subchannel allocation is neglected.

III. System Model and problem formulation

As shown in Fig. 1, a mobile base station (BS) attached with a mobile edge server is installed by the infrastructure

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2017R1A2A2A05000995). *Dr. CS Hong is the corresponding author.

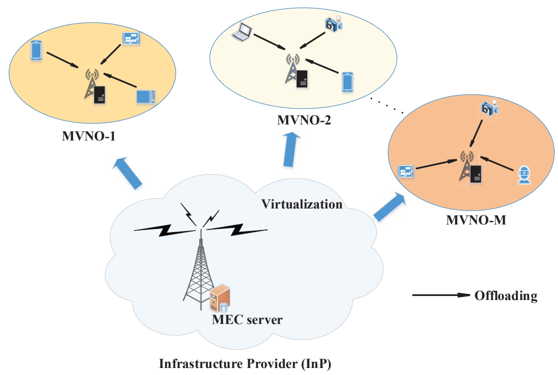

Fig. 1. A wireless virtualized network model

provider (InP) and providing the communication resources (i.e., bandwidth) and computation resources (i.e., CPU) to a set of mobile virtual network operators (MVNOs) $\mathcal{M} = \{1, 2, \ldots, M\}$ where BS is operating on the system bandwidth B and the mobile edge server has the maximum computation capacity f cycles/s. Then, each MVNO is providing a specific communication and computation services to their mobile users $\mathcal{U} = \{1, 2, \ldots, U\}$. The mobile user u of MVNO $m \in \mathcal{M}$ is generating a computation task k_{mu} which can be presented as a tuple $\{b_{mu}, c_{mu}, T_{mu}\}$, where b_{mu} is the total data size (bits) of the generating task, c_{mu} is the required computation capacity (i.e., CPU cycles) to execute one bit of data, and T_{mu} is the deadline of the task. Since, each user of MVNOs such as mobile device or sensor node, is the low power device and has limited computation capacity, it is difficult for the mobile user to execute the generated task locally within the deadline of the task. Therefore, each user u of MVNO m offloads their generated task to the edge server and the task is executed remotely.

A. Remote Computing Model

In this work, in order to avoid the intra-cell interference, we consider the orthogonal frequency division multiple access (OFDMA) system and the BS decides the amount of the system bandwidth allocated to each user of MVNOs to offload their computation task to the edge server and the computation resources allocation to each user to execute their computation task. The achievable data rate of the user u of the MVNO m can be obtained from the following equation:

$$R_{mu} = \alpha_{mu} B \log_2(1 + \frac{p_{mu} h_{mu}}{N_0}), \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \quad (1)$$

where α_{mu} is the fraction of the system bandwidth allocated to the user u of the MVNO m, p_{mu} is the uplink transmit power of the user u, h_{mu} is the channel gain of the user u of MVNO m, and N_0 is the Gaussian noise power. From there, we can formulate the time required for the user u in MVNO m to offload the generated task b_{mu} to the edge sever as follows:

$$t_{mu}^{Off} = \frac{b_{mu}}{\alpha_{mu} B \log_2(1 + \frac{p_{mu} h_{mu}}{N_0})}, \forall u \in \mathcal{U}, \forall m \in \mathcal{M}. \quad (2)$$

Furthermore, the energy consumption of offloading the task to the server is as follows:

$$E_{mu}^{Off} = \frac{p_{mu} b_{mu}}{\alpha_{mu} B \log_2(1 + \frac{p_{mu} h_{mu}}{N_0})}, \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \quad (3)$$

where p_{mu} is the uplink transmission power of the user u of MVNO m.

When the edge server receives all of the offloading task from the users of all MVNOs, the edge server allocates the computation resource to execute the offloading task of the users. Then, the execution delay at the edge server experienced by the user u of the MVNO m is as follows:

$$t_{mu}^{exe} = \frac{c_{mu} b_{mu}}{f_{mu}}, \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \quad (4)$$

where f_{mu} is the computation resource of the edge server (i.e., CPU capacity) allocated to the user u of MVNO m. Finally, the total delay experienced by the user u of MVNO m to complete the remote computing is as follows:

$$\begin{aligned} t_{mu}^{Re} &= t_{mu}^{Off} + t_{mu}^{exe} \\ &= \frac{p_{mu} b_{mu}}{\alpha_{mu} B \log_2(1 + \frac{p_{mu} h_{mu}}{N_0})} + \frac{c_{mu} b_{mu}}{f_{mu}}, \quad (5) \\ &\qquad \forall u \in \mathcal{U}, \forall m \in \mathcal{M}. \end{aligned}$$

Furthermore, the energy consumption at the edge server to execute the task of the user u of MVNO m is as follows:

$$E_{mu}^{exe} = k(f_{mu})^2 c_{mu} b_{mu}, \forall u \in \mathcal{U}, \forall m \in \mathcal{M} \quad (6)$$

where k is the energy consumption coefficient of the edge server.

In this work, the communication resource allocation based on the OFDMA and computation resource allocation in virtualized multi-access edge computing system is formulated as an optimization problem. The objective is to minimize the energy consumption of both mobile users of all MVNOs and the base station. Under the constraint of bandwidth and computation resource sharing at the edge server, and the latency of mobile users of MVNOs, we formulate the optimization problem as follows:

$$\min_{\alpha, \mathbf{f}} \sum_{m=1}^{M} \sum_{u=1}^{U} \left(\frac{p_{mu} b_{mu}}{\alpha_{mu} B \log_2(1 + \frac{p_{mu} h_{mu}}{N_0})} + k(f_{mu})^2 c_{mu} b_{mu} \right) \quad (7)$$

$$\text{s.t. C1}: \alpha_{mu} \in [0, 1], \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \quad (8)$$

$$\text{C2}: \sum_{m=1}^{M} \sum_{u=1}^{U} \alpha_{mu} \leq 1, \quad (9)$$

$$\text{C3}: \frac{b_{mu}}{\alpha_{mu} B \log_2(1 + \frac{p_{mu} h_{mu}}{N_0})} + \frac{c_{mu} b_{mu}}{f_{mu}} \leq T_{mu}, \quad \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \quad (10)$$

$$\text{C4}: \sum_{m=1}^{M} \sum_{u=1}^{U} f_{mu} \leq f, \quad (11)$$

where C1 and C2 represent the system bandwidth capacity constraint and C3 addresses the latency constraint of the task of the user u of MVNO m. Finally, C4 is the computation capacity constraint of the multi-access edge server. Straightforwardly, the two variables in the objective function are disjoint. Therefore, we decompose the proposed problem in (7) into two subproblems 1) communication resource allocation problem and 2) computation resource allocation problem.

IV. Efficient Algorithm for Joint Communication and Computation Resource Allocation in Virtualized MEC

Although the proposed optimization problem is non-convex, when we fix one variable among two decision variables, the proposed optimization problem becomes convex problem. Therefore, we decompose the original problem into two subproblems 1) communication resource allocation problem and 2) computation resource allocation problem. Moreover, the proofs on the convexity of two subproblems are presented in the next subsections.

A. Communication Resource Allocation Problem

For any fixed feasible computation resource allocation vector $\mathbf{f}$, the proposed problem in (7) can be rewritten as follows:

$$\text{P1}: \min_{\alpha} \; e(\alpha) \tag{12}$$

$$\text{s.t. C1}: \alpha_{mu} \in [0,1], \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \tag{13}$$

$$\text{C2}: \sum_{m=1}^{M}\sum_{u=1}^{U} \alpha_{mu} \leq 1, \tag{14}$$

$$\text{C3}: \frac{b_{mu}}{\alpha_{mu} B \log_2(1+\frac{p_{mu}h_{mu}}{N_0})} \leq T_{mu}, \tag{15}$$

where $e(\alpha) = \sum_{m=1}^{M}\sum_{u=1}^{U} \frac{p_{mu}b_{mu}}{\alpha_{mu}B\log_2(1+\frac{p_{mu}h_{mu}}{N_0})}$. Furthermore, the Lagrangian function of (12) is:

$$L(\alpha_{mu},\lambda,\nu_{mu}) = \sum_{m=1}^{M}\sum_{u=1}^{U} \frac{p_{mu}b_{mu}}{\alpha_{mu}B\log_2(1+\frac{p_{mu}h_{mu}}{N_0})} + \lambda\Big(\sum_{m=1}^{M}\sum_{u=1}^{U}\alpha_{mu} - 1\Big) + \sum_{m=1}^{M}\sum_{u=1}^{U}\nu_{mu} \left(\frac{b_{mu}}{\alpha_{mu}B\log_2(1+\frac{p_{mu}h_{mu}}{N_0})} - T_{mu}\right) \tag{16}$$

where $\lambda \leq 0$ and $\nu_{mu} \leq 0$ are the Lagrangian multipliers for constraints (14) and (15). Then, by using Karush-Kuhn-Tucker (KKT) conditions expressed in [12], we can get the optimal computation resource allocation to the user u of MVNO m as follows:

$$\alpha_{mu} = \left[T_{mu} - \frac{b_{mu}}{B\log_2(1+\frac{p_{mu}h_{mu}}{N_0})}\right]^{+}, \forall u \in \mathcal{U}, \forall m \in \mathcal{M}. \tag{17}$$

Algorithm 1 Joint communication and computation resource allocation in virtualized MEC

Initialize: $k=0$, $\alpha_{mu}^{(0)}$, $f_{mu}^{(0)}$, ϵ_1, ϵ_2;
while $||\alpha_{mu}^{(k+1)} - \alpha_{mu}^{(k)}|| \geq \epsilon_1$, $||f_{mu}^{(k+1)} - f_{mu}^{(k)}|| \geq \epsilon_2$ **do**
  for $m = 1$ to M **do**
    for $u = 1$ to U **do**
      Calculate the communication resource allocation $\alpha_{mu}^{(k+1)}$ at a given the computation resource allocation $f_{mu}^{(k)}$ by using (22)
      $\alpha_{mu} = \left[T_{mu} - \frac{b_{mu}}{B\log_2(1+\frac{p_{mu}h_{mu}}{N_0})}\right]^{+}$;
      Compute $f_{mu}^{(k+1)}$ from the computation resource allocation problem at a given $\alpha_{mu}^{(k+1)}$;
    end for
  end for
end while

B. Computation Resource Allocation Problem

For a given communication resource allocation, the proposed optimization problem described in (7) can be transformed to the computation resource allocation problem:

$$\mathbf{P2}: \min_{\mathbf{f}} \; e(\mathbf{f}) \tag{18}$$

$$\text{s.t. C3}: \frac{c_{mu}b_{mu}}{f_{mu}} \leq T_{mu}, \forall u \in \mathcal{U}, \forall m \in \mathcal{M}, \tag{19}$$

$$\text{C4}: \sum_{m=1}^{M}\sum_{u=1}^{U} f_{mu} \leq f, \tag{20}$$

where $e(\mathbf{f}) = \sum_{m=1}^{M}\sum_{u=1}^{U} k(f_{mu})^2 c_{mu}b_{mu}$. Then, we describe the Lagrangian function of the subproblem P2 as follows:

$$L(f_{mu},\phi,\zeta_{mu}) = \sum_{m=1}^{M}\sum_{u=1}^{U} k(f_{mu})^2 c_{mu}b_{mu} + \sum_{m=1}^{M}\sum_{u=1}^{U}\phi f_{mu} - \phi f + \sum_{m=1}^{M}\sum_{u=1}^{U}\zeta_{mu}\left(\frac{c_{mu}b_{mu}}{f_{mu}} - T_{mu}\right) \tag{21}$$

where $\phi \geq 0$ and $\zeta \geq 0$ and the Lagrangian multipliers of constraints (19) and (20). After applying KKT condition, we get the computation resource allocation to the user u of MVNO m is as follows:

$$\frac{\partial L}{\partial f_{mu}} = 2kc_{mu}f_{mu}b_{mu} + \phi - \zeta_{mu}\Big(\frac{c_{mu}b_{mu}}{f_{mu}^2}\Big) \leq 0, \tag{22}$$

Moreover, from (22), $2kc_{mu}f_{mu}b_{mu} > \zeta_{mu}(\frac{c_{mu}b_{mu}}{f_{mu}^2})$ so it can guarantee that the computation resource allocation to the user $u \in \mathcal{U}$ of MVNO $m \in \mathcal{M}$ is always positive.

V. Simulation Results

In this section, we present numerical results to illustrate the performance of the proposed joint communication and computing resource allocation scheme for the virtualized multi-access computing (MEC) system. In Fig. 2, we present a

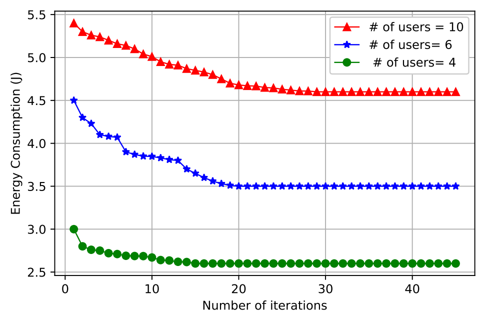

Fig. 2. Convergence of the proposed algorithm

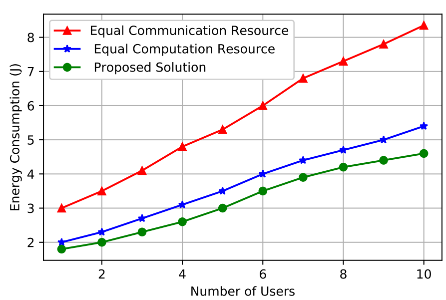

Fig. 3. Energy consumption (J) versus the number of users

convergence analysis of the energy function with our proposed algorithm for the number of users. The algorithm converges quickly for all use cases. Fig. 3 demonstrates total energy consumption to accomplish execution of tasks of users. Fig. 3 also compares the total energy consumption of users of MVNOs and BS by our proposed algorithm to the schemes of equal communication and computation resource allocation. It is clear that our proposed one outperforms both schemes. Moreover, in Fig. 4, we also compares the latency/delay experienced by our proposed algorithm with the equal communication and computation resource allocation schemes. It is clear that the delay is the highest in the equal communication resource allocation scheme and the lowest under our proposed algorithm.

VI. Conclusion

In this paper, we have proposed joint communication and computation resource allocation problem in the virtualized multi-access edge computing system. Then, we decomposed the proposed problem into two subproblems such as communication and computation resources allocation problems. The, we apply the KKT conditions to address each subproblem iteratively. Simulation results have confirmed that our proposed algorithm performs other schemes.

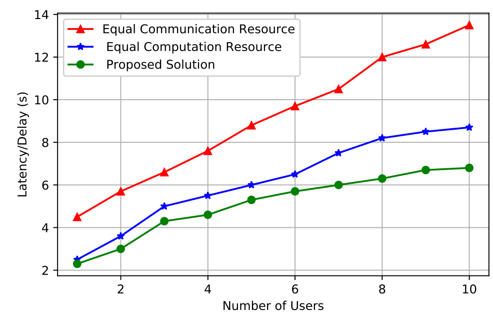

Fig. 4. Latency/Delay (s) versus the number of users

References

[1] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[2] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[3] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358–380, 2015.

[4] M. I. Kamel, L. B. Le, and A. Girard, "Lte wireless network virtualization: Dynamic slicing via flexible scheduling," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. IEEE, 2014, pp. 1–5.

[5] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, 2017.

[6] A. Al-Shuwaili and O. Simeone, "Energy-efficient resource allocation for mobile edge computing-based augmented reality applications," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 398–401, 2017.

[7] J. Zhang, X. Hu, Z. Ning, E. C.-H. Ngai, L. Zhou, J. Wei, J. Cheng, and B. Hu, "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2633–2645, 2018.

[8] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Energy-efficient offloading for mobile edge computing in 5g heterogeneous networks," *IEEE access*, vol. 4, pp. 5896–5907, 2016.

[9] C. W. Zaw, N. N. Ei, H. Y. R. Im, Y. K. Tun, and C. S. Hong, "Cost and latency tradeoff in mobile edge computing: A distributed game approach," in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2019, pp. 1–7.

[10] Y. K. Tun, C. W. Zaw, and C. S. Hong, "Downlink power allocation in virtualized wireless networks," in *2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2017, pp. 346–349.

[11] Y. K. Tun, S. R. Pandey, M. Alsenwi, and C. S. Hong, "Weighted proportional allocation based power allocation in wireless network virtualization for future wireless networks," in *2019 International Conference on Information Networking (ICOIN 2019)*. IEEE, 2019.

[12] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.