# SharedMEC: Sharing Clouds to Support User Mobility in Mobile Edge Computing

Wahida Nasrin and Jiang Xie
The University of North Carolina at Charlotte, Charlotte, NC 28223-0001
Email: {wnasrin, Linda.Xie}uncc.edu

***Abstract*—Recently, Mobile Edge Computing (MEC) is proposed to deploy with cellular base stations (BSs) to reduce the offloading delay and to provide computational resources and storage to nearby mobile users. However, the deployment of MEC with a BS introduces special challenges to mobility management. First, a MEC can only be accessed within the coverage area of its BS. Therefore, a user needs to perform both a radio handoff (HO) and a service migration when moving out of the cell coverage range. Second, as MECs have limited computational resource and storage, the target MEC may not have sufficient available resources to support new users. Therefore, a user with computation task may be forced to perform a radio HO to a different target BS in order to successfully migrate the service. These cause unnecessary HOs in cellular networks, especially when MECs are deployed with small cells, e.g., femtocells. These unnecessary HOs and service migrations incur a large signaling and migration cost. In this paper, we first propose a novel architecture, SharedMEC, to support user mobility. In addition, we propose a service HO decision algorithm and an analytical model to analyze the total cost which considers the total HO signaling cost, total migration signaling cost, and total migration cost. Simulation results show that our proposed architecture and service HO decision algorithm can significantly reduce the total cost. To the best of our knowledge, this is the first work that considers HO and migration issues together and analyzes the total cost in MEC systems.**

## I. Introduction

As mobile devices are resource constrained (limited computational resources and energy), offloading computation tasks to the remote cloud has become one of the most promising solutions to support energy-hungry mobile and Internet of Things (IoT) applications [1]. However, the remote execution of these offloaded applications to the cloud may not always fulfill the delay requirements. Therefore, mobile edge computing (MEC) is introduced recently to bring computational resources and storage closer to end users in order to reduce these delays [2], [3]. MECs are usually deployed with cellular base stations (BSs), e.g., macrocells and small cells.

However, the deployment of MECs with cellular BSs introduces new research challenges in mobility management. First, as MECs are deployed with BSs, a MEC can only be accessed within the coverage area of that BS. Therefore, when a user equipment (UE) moves out of the coverage area of a cellular BS, it needs to not only perform a radio handoff (HO), but also perform a migration of the offloaded service. Second, unlike the remote cloud, MECs have limited resources. As a result, sufficient resources might not always be available at the target MEC. Therefore, when a UE moves out of the cell coverage area, it needs to find a target BS with both sufficient radio and computational resources at the MEC. However, since the serving BS cannot get the information of the available resources of the target BS until the UE is connected to the target BS, and as the migration and HO issues are never considered together, the migration decision is made after the UE has completed a radio HO in the traditional system. Therefore, in the traditional system, the UE may again need to find a new BS after performing a radio HO to a target BS without any available computational resources at its MEC, which causes unnecessary HOs and adds additional delay at the user's end. Moreover, different offloaded services have different migration requirements. For example, an offloaded video streaming needs immediate service migration when the UE moves out of the coverage area of the MEC. All these issues are more severe when MECs are deployed with small cells, e.g., femtocells.

Existing work on mobility management in MEC only considers the issues of performing service migration and reducing the migration delay. These existing works are mainly focused on designing mathematical models and migration decisions based on the prediction of a user's future status. However, they may cause a large overhead in the network and the prediction of a user's future information is not practical to access. On the other hand, the radio HO and service migration issues are never considered together in MEC. Moreover, the extra signaling cost introduced by the unnecessary HOs are never analyzed. Therefore, a migration decision algorithm incorporating the radio HO decision is necessary to avoid extra signaling cost.

In this paper, we first introduce a novel architecture, *SharedMEC*, where a MEC is shared by a group of femto base stations (FBSs) to support user mobility. Then, we propose a service HO decision algorithm to make the decision whether to migrate a service or not, which can be used along with the radio HO decision algorithm. Later, we propose an analytical model to analyze the total cost (HO signaling, migration signaling, and migration cost) of a service HO. Our proposed architecture along with the service HO decision algorithm can reduce the total cost. To the best of our knowledge, this is the first work that considers the total cost during the radio HO and service migration process.

The rest of the paper is organized as follows. Related work, research motivation, and contributions are described in Section II. System model, the proposed HO decision algorithm, and the total cost analysis are presented in Section III. The performance evaluation is given in Section IV, followed by the conclusions in Section V.

This work was supported in part by the US National Science Foundation (NSF) under Grant No. 1343355, 1718666, and 1731675.

## II. Related Work, Research Motivation, and Contributions

### A. Related Work

Though various issues have been well investigated in the MEC system, the mobility management issue still needs further investigation to ensure low signaling and migration cost [4]. The solutions for supporting mobility in MEC are categorized in three groups: power control [5], virtual machine (VM) migration [6]–[17], and the selection of a new communication path between a UE and the MEC server [18], [19].

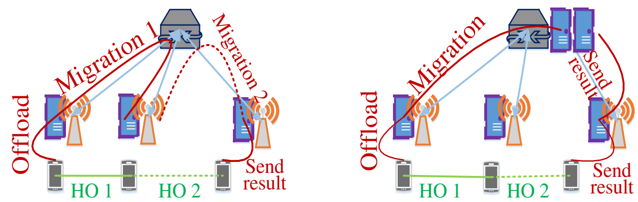
Power control solutions are considered for closed-access femtocells with computing capabilities. In these solutions, a power control algorithm is proposed to temporarily adjust the coverage of a femtocell in order to support mobile users [5]. The assumptions of these solutions are: femtocells can coordinate their transmission power and each femtocell supports only one user. Additionally, in these existing algorithms, the transmission power is controlled to adjust the coverage area in a way that the user is under the cell coverage until the offloaded computation is done and the result sends back to the user. However, FBSs cannot communicate with each other, and only considering one user per cell is not realistic. Moreover, existing solutions can only be applicable to low-speed indoor users who do not intend to leave home. However, other access modes of femtocells and users with an intention to leave home also need to be supported.

In VM migration solutions, existing works are mainly focused on designing mathematical models for different MEC systems, e.g., Follow-me cloud [6], [7] and software clone (Avatar) of cloudlets [8]. Furthermore, an optimal threshold is calculated in the threshold-based solutions [9], [12]. However, a complete framework or protocol design for VM migration is missing. On the other hand, prediction is considered in [10], [11], where the mobility path of a UE and the time window to travel a road segment are assumed to be known before making a migration decision. However, these prediction-based VM migration solutions require future information about a user's location, movement, etc. This information is random and not easy to obtain. Besides, using this information and a large calculation to make a migration decision may cause a large overhead at the network. Moreover, even a modest mobility can result in a significant network degradation [13], [14].

Path selection and/or VM migration solutions worked on the selection of the most appropriate way between a UE and an edge cloud [18], [19]. Though this is like the routing decision in wireless sensor networks (WSNs), in WSN, the nodes can communicate with each other. However, FBSs need to communicate via a backhaul in MEC, which will cause a delay in communication. On the other hand, the lowest delay path selection cannot always guarantee seamless service migration or fulfill the delay requirement of a service. In addition, the path is selected without any knowledge of the available computation resources, which can cause unnecessary HOs. After this discussion, we can observe that all the existing VM migration and path selection solutions are only focused on the migration issue. The HO and migration issues are not considered together.

### B. Research Motivation and Contributions

In the traditional MEC-deployed femtocell networks, when a UE moves out of the coverage area, it performs an HO and connects to a new FBS. As a result, the UE also moves out of the direct connection of the serving MEC, and the offloaded service also needs to be migrated to a new MEC associated with the new FBS. Each time a UE performs an HO, a service migration is required. However, it usually takes a long transfer time to perform a migration. Moreover, the migration delays vary based on the type of the service that needs to be migrated. Sharing a MEC can reduce the number of migrations. The migration scenarios for the traditional and the proposed (which is described in Section III-A) MEC-deployed femtocell networks are shown in Fig. 1.




(a) Traditional network (b) Proposed network

Fig. 1. Comparison of migration scenarios in the traditional MEC and the SharedMEC-deployed femtocell networks.

Additionally, simulation results showing the total number of migrations for both networks are presented in Fig. 2. In the figure, it is shown that the proposed model can reduce the number of migrations. Moreover, as the number of migrations is lower in the proposed network, it can also reduce the total cost. Furthermore, since the target MEC is predetermined, the necessary information for a service migration can be collected during the HO signaling period. Therefore, the migration signaling cost can also be reduced.

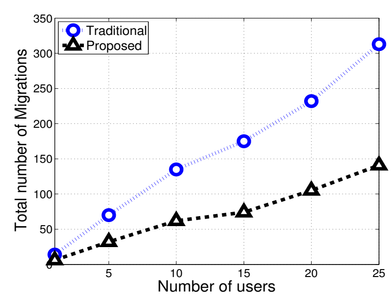



Fig. 2. Comparison of the number of migrations.

In this paper, we propose a service HO decision algorithm and an analytical model to calculate the total cost, which includes both radio and service HO signaling cost and the service migration cost. The contributions of this paper are summarized as follows.

- We propose an architecture, SharedMEC, by considering a shared-cloud to support mobility that can be deployed in any edge cloud-based network. Then, we propose a service HO decision algorithm that can make an HO decision based on the service type that needs to be migrated.

- We propose an analytical model for analyzing the total cost of a UE from triggering the radio HO to the completion of the service migration in the MEC-deployed femtocell network.
- We use a realistic simulation scenario to evaluate the performance of the proposed algorithm and compare it with the algorithm that does not consider service type and shared-MEC.

## III. Proposed Service HO Decision Algorithm and Total Cost Analysis

In this section, we propose a service HO decision algorithm in MEC systems based on our proposed SharedMEC architecture. An analytical model analyzing the total cost of an HO and a service migration for a mobile user is also presented.

### A. System Model

The architecture of the SharedMEC is presented in this section. In this architecture, each BS has a MEC with it, and a group of BSs share a MEC server which does not exist in the traditional MEC-deployed systems. Additionally, this SharedMEC can be implemented in any edge cloud networks, e.g., in MEC-deployed femtocell networks, in MEC-deployed macrocell networks, and in Cloudlets. Furthermore, though in this paper we consider this architecture in MEC-deployed femtocell networks, the proposed service HO decision algorithm and the analytical model can be applied to any edge cloud systems with or without considering this architecture. In the system model shown in Fig. 3, several FBSs form a cluster and all FBSs in the cluster share a MEC (which has larger computational resources and storage than the edge cloud at the FBS). The shared MEC may or may not be connected to the femto gateway (FGW). However, we propose to place the MEC with the FGW, so that it can be shared by all FBSs connected to the FGW. Therefore, the shared-MEC and FBSs are wire connected via the FGW. In addition, all wire connections in the proposed architecture are the same as in the traditional femtocell architecture. A number of such femtocell clusters can be within a macrocell coverage area.

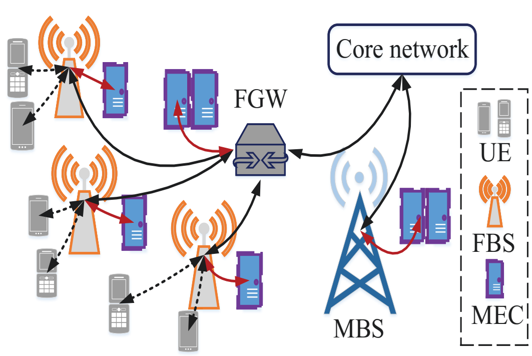



Fig. 3. The architecture of the SharedMEC.

The proposed architecture has several characteristics and advantages over the traditional MEC-deployed femtocell networks. First of all, this system can be easily deployed without any major changes in the traditional MEC-deployed networks. Second, both femto-MEC and shared-MEC are at most two-hops away from UEs. Third, sharing the information of available resources between the serving MEC and the target MEC is easier because they are wire connected, therefore, the migration cost can be reduced. Finally, a path selection between the serving MEC and the target MEC is not necessary, which reduces the delay of performing a migration.

In addition to the proposed architecture, to support the seamless service transfer between two MECs, we use the 3-layer container system as described in [20]. In the proposed SharedMEC, all MECs are deployed by the same service provider. Therefore, all MECs have the same "bins/libs" layer of the container. Now, our goal is to transfer the data instance when the application is available at the target MEC or transfer both the data instance and the application if the application is not available at the target MEC. Since we only focus on transferring the service between MECs, we named this process as *service HO*.

**Algorithm 1:** Service HO Decision Algorithm

**if** *Application availability = 1* **then**
- **if** $0MB < RAM\ utilization < 300MB$ **then**
  - Perform service HO;
- **else**
  - Find $T_{Time} = \frac{dataSize_{Remaining}}{Speed_{FBS-FGW}}$;
  - Find $P_{Time} = \frac{dataSize_{Remaining}}{Speed_{MEC}}$;
  - **if** $T_{Time} < P_{Time}$ **then**
    - Perform service HO;
  - **else**
    - Sent result after completing the task;

**else**
- **if** $0MB < RAM\ utilization < 100MB$ **then**
  - Perform service HO;
- **else**
  - Find $T_{Time} = \frac{dataSize_{Remaining}+appSize}{Speed_{FBS-FGW}}$;
  - Find $P_{Time} = \frac{dataSize_{Remaining}+appSize}{Speed_{MEC}}$;
  - **if** $T_{Time} < P_{Time}$ **then**
    - Perform service HO;
  - **else**
    - Sent result after completing the task;

End;

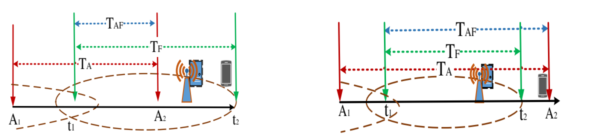
### B. Service HO Decision Algorithm

In the service HO decision algorithm, the service HO decision is made based on available resources, application type, and application availability. As the proposed SharedMEC architecture uses a MEC that is shared by a group of FBSs, it is not difficult to get this information before making a service HO decision. Therefore, the resource availability and the application availability can be obtained from the FGW. In addition, the type of an application can be determined based on the RAM utilization. Different application types have different memory requirements, e.g., a video streaming requires approximately 30MB and a game requires approximately 1MB RAM utilization, and the application that uses more memory needs longer migration time [20]. Our proposed service HO decision algorithm is shown in Algorithm 1.

In the algorithm, $T_{Time}$ represents the time to transfer a service, $P_{Time}$ represents the time to process a service at the serving MEC, $dataSize_{Remaining}$ is the size of the remaining computation, $appSize$ is the size of the application, $Speed_{FBS-FGW}$ is the speed of the wire connection between an FBS and an FGW, and $Speed_{MEC}$ is the CPU speed of a MEC. The application availability is represented by $\{0, 1\}$, where 1 means an application is available. The proposed algorithm is designed in a way that all real-time applications perform migration to avoid disconnection.

### C. Total Cost Analysis

In a MEC-deployed femtocell network, a service HO is triggered only when a radio HO is performed. Therefore, the mobility events that cause different radio HO scenarios are responsible for the service HOs. Now, in any homogeneous network, there are four mobility events that cause HOs. These events are shown in Fig. 4 and Fig. 5.

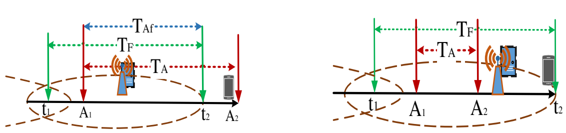


(a) A HO happens at $t_1$ (b) Two HOs happen at $t_1$ and $t_2$

Fig. 4. Timing diagram for mobility events in MEC.

The event in Fig. 4(a) represents an active UE performs an HO to a target cell, then becomes inactive. An active UE performs an HO to a target cell, then performs another HO before becoming inactive is the event shown in Fig. 4(b). Similarly, an inactive UE moves into a MEC coverage area, then becomes active and performs an HO to another cell is the event shown in Fig. 5(a). In Fig. 5(b), no HOs happen since the UE becomes active and finishes its task before moving out of the cell area. If the probability of the events in Fig. 4(a), 4(b), and 5(a) are considered as $P_{b1}$, $P_{b2}$, and $P_{b3}$. Then, we can get the probability of service HOs as:

$$P_{serviceHO} = P_{b1} + 2P_{b2} + P_{b3}. \quad (1)$$



(a) A HO happens at $t_2$ (b) No HOs happen

Fig. 5. Timing diagram for the mobility events in MEC.

In the timing diagrams, $T_A$ and $T_F$ are independent random variables. $T_A$ denotes the session duration which is exponentially distributed with mean $1/\eta$, and the probability density function of this session duration is $f_{TA}(t) = \eta e^{-\eta t}$. Similarly, $T_F$ is the duration of a UE being within the coverage area of a femtocell which is exponentially distributed with mean $1/\mu$, and the probability density function of this duration of stay is $f_{TF}(t) = \mu e^{-\mu t}$. $T_{AF}$ and $T_{Af}$ in the timing diagram follow the memoryless property of the residence times, $T_A$ and $T_F$, respectively. In addition, the probability density function of $T_{AF}$ is $f_{AF}$ (which is exponentially distributed with mean $1/\eta$) and the probability density function of $T_{Af}$ is $f_{Af}$ (which is exponentially distributed with mean $1/\mu$). Now, we can calculate $P_{b1}$, $P_{b2}$, and $P_{b3}$ as:

$$P_{b1} = P(A_1 < t_1 < A_1 + T_A) \cdot P(T_{AF} \le T_A), \quad (2)$$

$$P_{b2} = P(A_1 < t_1 < A_1 + T_A) \cdot P(T_{AF} > T_F), \quad (3)$$

and

$$P_{b3} = P(t_1 < A_1 < t_1 + T_F) \cdot P(T_A \ge T_{Af}). \quad (4)$$

Using the Laplace transform, we have

$$P_{b1} = \int_0^\infty \int_t^\infty \lambda t e^{-\lambda t} f_{TA}(x) dx dt \cdot (1 - \int_0^\infty \int_t^\infty \eta e^{-\eta t} f_{AF}(y) dy dt), \quad (5)$$

$$P_{b2} = \int_0^\infty \int_t^\infty \lambda t e^{-\lambda t} f_{TA}(y) dy dt \cdot (1 - \int_0^\infty \int_t^\infty \mu e^{-\mu x} f_{AF}(t) dx dt), \quad (6)$$

and $P_{b3} = \int_0^\infty \lambda t e^{-\lambda t} f_{Af}(t) dt \cdot$

$$\int_0^\infty \int_t^\infty \eta e^{-\eta y} f_{Af}(t) dy dt. \quad (7)$$

Solving (5), (6), and (7), we can obtain the probabilities as:

$$P_{b1} = \frac{\lambda \eta}{(\lambda + \eta)^2 (\mu + \eta)}, \quad (8)$$

$$P_{b2} = \frac{\lambda \mu}{(\lambda + \eta)^2 (\mu + \eta)}, \quad (9)$$

and

$$P_{b3} = \frac{\lambda \mu^2}{(\lambda + \eta)^2 (\mu + \eta)}. \quad (10)$$

Finally, the total cost can be calculated as:

$$C_{total} = P_{serviceHO} \cdot (\sum T_j^i + \sum P_i + \sum CT_j^i + \sum CP_i + \sum M_j^i), \quad (11)$$

Here, $T_j^i$ is the delivering cost of an HO message between node $i$ and $j$, $P_i$ is the processing cost of a message at node $i$, $CT_j^i$ is the delivering cost of a service HO message between node $i$ and $j$, $CP_i$ is the processing cost of a service message at node $i$, and $M_j^i$ is the migration cost of a service between node $i$ and $j$. The radio and service HO signaling procedure is given in Fig. 6. We can get $T_j^i$, $P_i$, and $CT_j^i$ from the HO signaling procedure as:

$$\sum (T_j^i) = 2T_{UE}^{FBS} + 10T_{FGW}^{FBS} + 2T_{FGW}^{MME}, \quad (12)$$

$$\sum (P_i) = P_{UE} + P_{FBS} + 2P_{FGW} + P_{MME}, \quad (13)$$

and

$$\sum (CT_j^i) = 2T_{FGW}^{FBS} + 5T_{MEC}^{FBS}. \quad (14)$$

However, $4T_{MEC}^{FBS}$ does not add extra cost in the process because they happen when another radio signaling is in action. Therefore,

$$\sum (CT_j^i) = 2T_{FGW}^{FBS} + T_{MEC}^{FBS}. \quad (15)$$

In addition, $CP_i$ and $M_j^i$ can be obtained from the HO signaling procedure as:

$$\sum(CP_i) = P_{FBS} \quad (16)$$

and

$$\sum(M_j^i) = T_{FGW}^{FBS} + 2T_{MEC}^{FBS}. \quad (17)$$

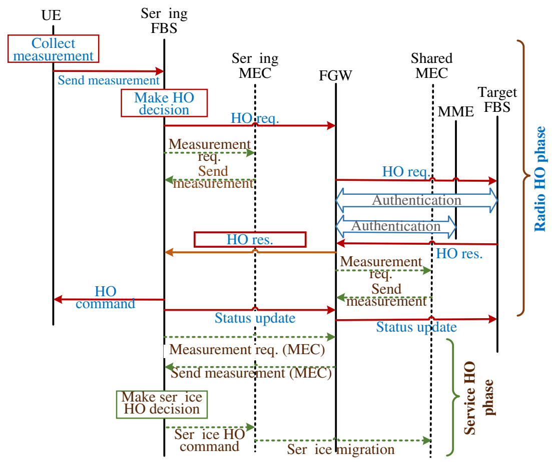



Fig. 6. Total HO and migration signaling.

Notations for different costs and their values are given in Table I [21]–[27] and Table II [20]. In Table II, AppType 1 (memory size: 0MB to 30MB) represents gaming applications, AppType 2 (memory size: 31MB to 99MB) represents audio or video streaming, AppType 3 (memory size: 101MB to 299MB) represents detection and editing applications, and AppType 4 (memory size: over 300MB) represents simulation and computation applications.

TABLE I
HO SIGNALING COST PARAMETERS

| | | |
|---|---|---|
| $T_{UE}^{FBS}$ | Transmission cost between a UE and an FBS | 2 |
| $T_{FBS}^{FGW}$ | Transmission cost between an FBS and an FGW | 2 |
| $T_{FGW}^{MME}$ | Trnsmission cost between an FGW and an MME | 4 |
| $T_{FBS}^{MEC}$ | Transmission cost between an FBS and an MEC | 1 |
| $P_{UE}$ | Processing time at UE | 40 |
| $P_{FBS}$ | Processing cost at FBS | 3 |
| $P_{FGW}$ | Processing cost at FGW | 2 |
| $P_{MME}$ | Processing cost at MME | 4 |

TABLE II
MIGRATION COST

| Application available at the shared-MEC | | | |
|---|---|---|---|
| **App Type** | **Data Transfered** | **Migration Time** | **Migration Cost** |
| AppType 1 | 1.6MB | 6.4s | 20 |
| AppType 2 | 7.4MB | 8.5s | 6 |
| AppType 3 | 10MB | 15.5s | 8 |
| AppType 4 | 97.1MB | 19.8s | 1 |
| **Application not available at the shared-MEC** | | | |
| AppType 1 | 2.7MB | 10.9s | 34 |
| AppType 2 | 184.6MB | 37.3s | 25 |
| AppType 3 | 365MB | 70.1s | 35 |
| AppType 4 | 97.6MB | 27.2s | 2 |

## IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed service HO decision scheme in terms of total signaling cost. We use NetLogo 6.0.1 [28] to simulate the SharedMEC environment. We deploy fifteen femtocells in a random manner and each FBS has a MEC that is deployed with it. We also deploy three shared-MECs, each of these MECs is shared by five FBSs. All users follow the Random Waypoint mobility model. We use 1 to 25 users in the system. The Okumura-Hata propagation model is used for the macrocell network, and the ITU-R P.1238-7 indoor path-loss model [29] is used for the femtocell network. The parameters used in our simulation are listed in Table III [30].

TABLE III
SIMULATION PARAMETERS

| | |
|---|---|
| Macrocell transmission power, $P_m$ | 45 dBm |
| Radius of macrocell | 1.2 km |
| Femtocell transmission power, $P_f$ | 10 dBm |
| Radius of femtocell | 15 m |
| Users speed | 0 to 10 km/hr |
| Threshold, $Th$ | -45 dB |
| $RSSI_{min}$ | -75 dB |
| $HM_{max}$ | 5 dB |

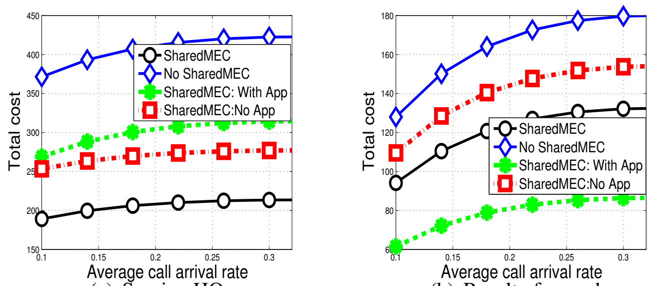
To evaluate the performance of the proposed model and algorithm, we investigate the following two performance metrics: 1) *Total cost of service HOs*: the summation of the total HO signaling cost, total migration signaling cost, and total migration cost when a user performs both radio and service HOs from the serving FBS to the target FBS; 2) *Total cost of results forward*: the summation of the total HO signaling cost, processing cost of the service at the service MEC, and the cost of result transfer when the serving FBS decides to process the service and send the result back to the UE via the shared-MEC. Additionally, we compare our proposed model and algorithm with the traditional MEC model and two other variations of our proposed model to analyze the worst and the best case scenarios. These algorithms are: 1) *No SharedMEC*: MECs are only deployed with each FBS and our proposed service HO decision algorithm is used; 2) *SharedMEC: With App*: our proposed system model and service HO decision algorithm are used, however, it is considered that applications are always available at the target MEC; 3) *SharedMEC: No App*: our proposed system model and service HO algorithm are used, however, it is considered that applications are never available at the target MEC.

### A. Total Cost of Service HOs

The performance of the total cost for service HOs is given in Fig. 7(a). The total cost is determined by considering all radio HO signaling cost, all migration signaling cost, and the migration cost. It is calculated for an exponential session duration (mean $1/\eta = 3$), an exponential residence time (mean $1/\mu = 10$), and the Poisson session arrival rate $\lambda$ (0.1 to 0.34). Then, the total HO signaling cost is calculated by multiplying the signaling cost of an HO by the rate of HOs, which also includes the rate of unnecessary HOs. From the figure, we can observe that the total cost in our proposed model is lower than that in the traditional MEC model. Furthermore, we can observe that our proposed model shows better performance than the traditional one even in the worst case (no applications are available at the target MEC) and the best case (all applications are always available at the target MEC) scenarios. The availability of applications at the target MEC encourage more services to be handed over to the target MEC, therefore, the migration cost is higher in this case than in the no application available case, which also forces the total cost to be higher.

### B. Total Cost of Results Forward

In Fig. 7 (b), the performance of results forwarding cost for our proposed and the traditional models is presented. Here, the total cost of results forwarding is determined by considering all radio HO signaling cost, all migration signaling cost, the processing cost of the service at the serving MEC, and the result forwarding cost. From the results in Fig. 7(b), we can observe that our proposed model presents better performance than the traditional MEC model. We can also observe that the total cost of results forwarding is lower in our proposed model than that in the traditional one even in the worst case and the best case scenarios. Additionally, the total cost of results forwarding is higher in SharedMEC: No App. This is because when applications are not available at the target MEC, all services that do not require live migration check whether the migration cost is higher than the result transfer cost. If the migration cost is higher, then these services are executed at the serving MEC.




(a) Service HOs (b) Results forward

Fig. 7. Total cost of service HOs and results forward.

## V. Conclusion

In this paper, a novel architecture along with a service handoff decision algorithm is presented in order to support mobile users in MEC-deployed cellular networks. We addressed the issue of extra signaling cost due to unnecessary handoffs caused by the resource unavailability at the target MEC. In addition, we proposed an analytical model to determine the total cost of radio handoff signaling, service handoff signaling, and total migration. Simulation results show that our proposed system model and service handoff decision algorithm can reduce the total cost due to both radio and service handoffs.

## References

[1] K. Kumar and Y.-H. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" *IEEE Computer*, vol. 43, no. 4, pp. 51–56, 2010.

[2] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal *et al.*, "Mobile-edge computing introductory technical white paper," *White Paper, Mobile-edge Computing (MEC) Industry Initiative*, 2014.

[3] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "Mobile edge computing: Survey and research outlook," *arXiv preprint arXiv:1701.01090*, 2017.

[4] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, 2017.

[5] ——, "Cloud-aware power control for real-time application offloading in mobile edge computing," *Transactions on Emerging Telecommunications Technologies*, vol. 27, no. 5, pp. 648–661, 2016.

[6] T. Taleb and A. Ksentini, "An analytical model for follow me cloud," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 1291–1296.

[7] A. Ksentini, T. Taleb, and M. Chen, "A Markov decision process-based service migration procedure for follow me cloud," in *Proc. of IEEE International Conference on Communications (ICC)*, 2014.

[8] X. Sun and N. Ansari, "Primal: Profit maximization avatar placement for mobile edge computing," in *Proc. of IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[9] S. Wang, R. Urgaonkar, M. Zafer, T. He, K. Chan, and K. K. Leung, "Dynamic service migration in mobile edge-clouds," in *Proc. of IFIP Networking Conference*, 2015, pp. 1–9.

[10] A. Nadembega, A. S. Hafid, and R. Brisebois, "Mobility prediction model-based service migration procedure for follow me cloud to support QoS and QoE," in *Proc. of IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[11] S. Wang, R. Urgaonkar, T. He, K. Chan, M. Zafer, and K. K. Leung, "Dynamic service placement for mobile micro-clouds with predicted future costs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 4, pp. 1002–1016, 2017.

[12] R. Urgaonkar, S. Wang, T. He, M. Zafer, K. Chan, and K. K. Leung, "Dynamic service migration and workload scheduling in edge-clouds," *Performance Evaluation*, vol. 91, pp. 205–228, 2015.

[13] K. Ha, Y. Abe, Z. Chen, W. Hu, B. Amos, P. Pillai, and M. Satyanarayanan, "Adaptive VM handoff across cloudlets," *Technical report, Technical Report CMU-C S-15–113*, 2015.

[14] S. Secci, P. Raad, and P. Gallard, "Linking virtual machine mobility to user mobility," *IEEE Transactions on Network and Service Management*, vol. 13, no. 4, pp. 927–940, 2016.

[15] M. Zhang, X. Du, and K. Nygard, "Improving coverage performance in sensor networks by using mobile sensors," in *Proc. of Military Communications Conference (MILCOM)*, 2005, pp. 3335–3341.

[16] X. Du and F. Lin, "Maintaining differentiated coverage in heterogeneous sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 5, no. 4, pp. 565–572, 2005.

[17] L. Huang, G. Zhu, and X. Du, "Cognitive femtocell networks: an opportunistic spectrum access for future indoor wireless coverage," *IEEE wireless communications*, vol. 20, no. 2, pp. 44–51, 2013.

[18] Z. Becvar, J. Plachy, and P. Mach, "Path selection using handover in mobile networks with cloud-enabled small cells," in *Proc. of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2014, pp. 1480–1485.

[19] J. Plachy, Z. Becvar, and P. Mach, "Path selection enabling user mobility and efficient distribution of data for computation at the edge of mobile network," *Computer Networks*, vol. 108, pp. 357–370, 2016.

[20] A. Machen, S. Wang, K. K. Leung, B. J. Ko, and T. Salonidis, "Live service migration in mobile edge clouds," *arXiv preprint arXiv:1706.04118*, 2017.

[21] H. Zhang, W. Ma, W. Li, W. Zheng, X. Wen, and C. Jiang, "Signalling cost evaluation of handover management schemes in LTE-advanced femtocell," in *Proc. of IEEE Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–5.

[22] W. Nasrin and J. Xie, "Signaling cost analysis for handoff decision algorithms in femtocell networks," in *Proc. of IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[23] J. Xie, "User independent paging scheme for mobile ip," *Wireless Networks*, vol. 12, no. 2, pp. 145–158, 2006.

[24] W. Nasrin and J. Xie, "A mobility management scheme to reduce the impact of channel heterogeneity in cognitive radio femtocell networks," in *Proc. of IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2016, pp. 1–9.

[25] J. McNair, T. Tugcu, W. Wang, and J. L. Xie, "A survey of cross-layer performance enhancements for Mobile IP networks," *Computer Networks*, vol. 49, no. 2, pp. 119–146, 2005.

[26] W. Nasrin and J. Xie, "Effects of heterogeneous frequency changes in cognitive radio femtocell networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[27] ——, "An effective target cell selection scheme in open-access femtocell networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2017.

[28] S. Tisue and U. Wilensky, "NetLogo: A simple environment for modeling complexity," in *Proc. of International Conference on Complex Systems*, 2004, pp. 16–21.

[29] P. Series, "Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 900 MHz to 100 GHz," *ITU-R Recs*, 2012.

[30] W. Nasrin and J. Xie, "A self-adaptive handoff decision algorithm for densely deployed closed-group femtocell networks," in *Proc. of IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2015, pp. 390–398.