# Energy Driven Avatar Migration in Green Cloudlet Networks

Qiang Fan, Nirwan Ansari, and Xiang Sun

***Abstract*— Fully utilizing green energy can remarkably decrease the operational cost of cloudlet providers in provisioning green cloudlet networks (GCNs), which are powered by both green and brown energy. Owing to the spatial and temporal dynamics of energy demands and green energy generation, migrating Avatars (i.e., virtual machines) from green energy deprived cloudlets into green energy over-provisioned cloudlets can reduce the total on-grid energy consumption of GCN. However, Avatar migration itself consumes non-negligible energy consumption. In this letter, we propose the Energy driven AvataR migration (EARN) scheme to reduce the total on-grid energy consumption of GCN by considering the energy consumption of Avatar migrations. The performance of EARN is demonstrated by extensive simulations.**

***Index Terms*— Cloudlet, edge computing, virtual machine migration, green energy.**

## I. Introduction

MOBILE applications are increasingly computation-intensive while the computational capacity of battery powered user equipments (UEs) remains limited. Mobile Cloud Computing (MCC) enables UEs to offload some tasks to high performance Virtual Machines (VMs) in remote clouds, thus reducing the task execution time and energy consumption of UEs. Existing researches mostly consider the remote cloud as the offloading destination, owing to its abundant resources. However, the long end-to-end (E2E) delay between a UE and its VM far away imposes a detrimental impact on the quality of service of applications, such as augmented reality and online gaming, where a low E2E delay is required. Thus, the concept of cloudlets is employed to reduce the E2E delay between a UE and its VM. Cloudlets, tiny versions of data centers, are generally placed at the network edge that are close to UEs. The physical proximity between UEs and cloudlets leads to a low E2E delay [1]. Meanwhile, as green energy technologies advance, green energy can be readily employed to reduce the on-grid energy cost. Energy generated from solar panels can be used to power distributed cloudlets, with on-grid energy as a backup. Recent research works have already shown that distributed cloudlets can remarkably reduce the E2E delay between UEs and VMs in the cloudlet. Sun and Ansari [2] proposed a profit maximization Avatar placement for mobile edge computing, referred to as PRIMAL, which makes a tradeoff between the E2E delay reduction and migration overheads by selectively migrating the VMs to their optimal locations. Rather than considering the E2E delay between UEs and their VMs, some works have focused on minimizing the energy consumption of cloudlet networks and clouds. Giacobbe *et al.* [3] aimed to minimize the cost of Internet Data Centers (IDCs), by migrating the workloads from IDCs of high electricity cost to IDCs of low electricity cost. While the introduction of green energy leads to the reduction of on-grid energy consumption, matching the dynamic green energy generation and dynamic energy demands of data centers is a great challenge [4]. Sun *et al.* [5] proposed a green energy aware Avatar migration strategy, referred to as GEAR, to migrate UEs' virtual machines from the cloudlets with insufficient green energy generation to the cloudlets with excess green energy generation in order to reduce the total on-grid energy consumption in the network. However, Avatar migration itself consumes non-negligible energy, which may affect the total energy consumption in the network. Specifically, although migrating UEs' Avatars (i.e., VMs) to cloudlets with higher green energy and lower energy demands can reduce the on-grid energy consumption, Avatar migration between cloudlets incurs significant traffic, and thus incurs migration cost in terms of the on-grid energy consumption of both source and destination cloudlets. For instance, if Avatar-1 migrates from cloudlet-A (source) to cloudlet-B (destination), this migration consumes non-negligible energy of cloudlet-A and cloudlet-B. To tackle this problem, we propose the Energy driven AvataR migratioN (EARN) scheme to minimize the total on-grid energy consumption by considering the migration cost (in terms of the energy consumption incurred by Avatar migrations), while ensuring the service level agreement (SLA) for each UE in terms of the maximum E2E delay for each UE.

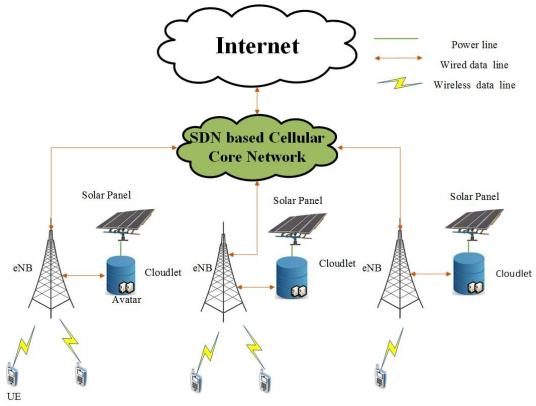


Fig. 1. GCN architecture.

## II. System Model

A Green Cloudlet Network (GCN) architecture is illustrated in Fig. 1 in which each cloudlet is collocated with

Manuscript received January 29, 2017; accepted March 12, 2017. Date of publication March 20, 2017; date of current version July 8, 2017. The associate editor coordinating the review of this letter and approving it for publication was D. Calin. *(Corresponding author: Qiang Fan.)*

The authors are with the New Jersey Institute of Technology, Newark, NJ 07102-1982, USA (e-mail: qf4@njit.edu; nirwan.ansari@njit.edu; xs47@njit.edu).

Digital Object Identifier 10.1109/LCOMM.2017.2684812

an eNB. Distributed cloudlets are able to transfer data to each other via the cellular core network. Software Defined Network (SDN) based cellular network is employed to provide efficient and flexible communications paths between eNBs. Meanwhile, LTE providers offer seamless wireless communications between a UE and its eNB, and thereby each UE can connect to a nearby cloudlet to minimize the E2E delay. In GCN, each UE can be mapped to a specific Avatar (i.e., one VM in the cloudlet), which runs tasks offloaded from its corresponding UE [5]. An Avatar is a software clone of a UE and always offers service to the UE wherever it moves. Moreover, in order to reduce on-grid energy consumption, cloudlets can be powered by solar energy.

We assume every UE's Avatar is homogeneous (i.e., the configuration of Avatars are the same) although the workloads of different Avatars are different. Also, all servers in cloudlets are homogeneous, i.e., the configuration of every server is the same. Therefore, each server is assumed to host the same number of Avatars. Since the E2E delay between a UE and its Avatar has a vital impact on the performance of delay sensitive applications, the cloudlet provider needs to ensure the SLA for UEs in terms of the maximum E2E delay.

Denote $\mathcal{I}$ as the set of cloudlets in the network and $\mathcal{K}$ as the set of UEs/Avatars (note that one UE is associated with one specific Avatar, and thus UEs and Avatars share the same set). $N_i$ is the total number of servers in cloudlet $i$. To identify the location of a UE's Avatar, we introduce two binary variables $x_{i,k}$ and $\eta_i\ (j,k)$, where $i$ is the index of cloudlets, $j$ is the index of servers, and $k$ is the index of Avatars. Here, $x_{i,k}$ means whether Avatar $k$ is in cloudlet $i$; $\eta_i\ (j,k)$ indicates whether Avatar $k$ is in cloudlet $i$'s server $j$. The relationship between them can be expressed as follows:

$$x_{i,k} = \sum_{j=1}^{N_i} \eta_i(j,k). \tag{1}$$

Then, we can get the number of active servers in cloudlet $i$:

$$n_i = \left\lceil \frac{\sum\limits_{k\in\mathcal{K}} x_{i,k}}{\tau} \right\rceil, \tag{2}$$

where $\tau$ is the maximum number of Avatars inside a server.

### A. Energy Model of Cloudlets

The power consumption of a cloudlet is drawn from active servers as follows [6]:

$$P_{i,j} = P^s + \alpha u_{i,j}, \tag{3}$$

where $P_{i,j}$ is the power consumption of server $j$ in cloudlet $i$, $P^s$ is the idle power of an active server, $\alpha$ is a coefficient in mapping the CPU utilization into the power consumption, and $u_{i,j}$ is the CPU utilization of server $j$ in cloudlet $i$ for running Avatars.

If the CPU utilization for running Avatar $k$'s applications is $u_k$, the CPU utilization of server $j$ in cloudlet $i$, denoted as $u_{i,j}$, can be expressed as a function of $\eta_i(j,k)$.

$$u_{i,j} = \sum_{k\in\mathcal{K}} \eta_i(j,k) \times u_k. \tag{4}$$

By substituting Eq. (4) into Eq. (3), we get

$$P_{i,j} = P^s + \sum_{k\in\mathcal{K}} \eta_i(j,k) \times \alpha u_k. \tag{5}$$

After aggregating the power consumption of all active servers, we achieve the total power consumption of cloudlet $i$ as follows:

$$P_i = \sum_{j=1}^{n_i} P_{i,j} = n_i P^s + \sum_{j=1}^{n_i}\sum_{k\in\mathcal{K}} \eta_i(j,k) \times \alpha u_k. \tag{6}$$

Since $n_i \approx \frac{\sum_k x_{i,k}}{\tau}$, Eq. (6) can be transformed into:

$$P_i = \sum_{k\in\mathcal{K}} \left[ x_{i,k} \times \left(\frac{P_s}{\tau} + \alpha u_k\right)\right]. \tag{7}$$

We need to maximize the utilization of available green energy of each cloudlet in each time slot by migrating Avatars in order to minimize on-grid energy consumption of cloudlets. Furthermore, the on-grid energy consumption of cloudlet $i$ is expressed as : $\rho_i = max((P_i T + E_i^{mig} - G_i T), 0)$ [7], where $T$ is the length of a time slot, $P_i$ is the energy consumption for running Avatars in cloudlet $i$, $E_i^{mig}$ is the energy consumption of cloudlet $i$ incurred by Avatar migrations, and $G_i$ is the green energy generation rate of cloudlet $i$.

### B. Migration Cost Model

When migrating an Avatar, traffic is generated from the source cloudlet to the destination cloudlet, thus incurring extra energy consumption on the source cloudlet and destination cloudlet, which is defined as the Avatar migration cost.

Denote $x_{i,k}^t$ and $x_{i,k}^{t+1}$ as two indicators on whether Avatar $k$ is located in cloudlet $i$ in time slot $t$ and $t+1$, respectively. Therefore, $(x_{i,k}^t - x_{i,k}^{t+1})^2$ indicates whether Avatar $k$ is migrated to or from cloudlet $i$ in time slot $t+1$. Furthermore, when an Avatar migration occurs at the destination cloudlet $i$, $(x_{i,k}^t - x_{i,k}^{t+1})^2 x_{i,k}^{t+1}$ equals to 1, otherwise 0. Similarly, if cloudlet $i$ acts as the source cloudlet, $(x_{i,k}^t - x_{i,k}^{t+1})^2(1 - x_{i,k}^{t+1})$ equals to 1, otherwise 0.

According to the energy consumption model for conducting migrations [8], the migration cost of source and destination cloudlet can be expressed as

$$E_k^s = \sigma^s V_k + \beta^s, \tag{8}$$

$$E_k^d = \sigma^d V_k + \beta^d, \tag{9}$$

respectively, where $V_k$ is the data volume incurred by migrating Avatar $k$, and $\sigma^s$,$\sigma^d$, $\beta^s$, and $\beta^d$ are the coefficients that map the data volume into energy consumption, which can be trained based on different platforms. Consequently, the migration cost of cloudlet $i$ in slot $t+1$ is expressed as

$$E_i^{mig} = \sum_{k\in\mathcal{K}} (x_{i,k}^t - x_{i,k}^{t+1})^2 x_{i,k}^{t+1} E_k^d + (x_{i,k}^t - x_{i,k}^{t+1})^2 (1 - x_{i,k}^{t+1}) E_k^s, \tag{10}$$

where the first term is the total migration cost of cloudlet $i$ when cloudlet $i$ acts as the destination cloudlet and the second term is that when cloudlet $i$ acts as the source cloudlet.

### C. E2E Delay Model

When UEs move in the network, their Avatars tend to be migrated to the optimal cloudlets in order to minimize the provider's cost. In this case, the communications between a UE and its Avatar may transverse the SDN-based cellular core. Therefore, the E2E delay between a UE and its Avatar consists of three parts: first, the E2E delay between a UE and its eNB; second, the E2E delay between the UE's eNB and its cloudlet where its Avatar is located; third, the E2E delay within the cloudlet. Changing the locations of UEs' Avatars does not significantly affect the first and third parts. Thus, we only consider the E2E delay between a UE's eNB and its cloudlet, which is the most important factor affecting the E2E delay between a UE and its Avatar. When an Avatar is migrated among cloudlets, the SLA in terms of maximum E2E delay between the UE's eNB and the cloudlet should be satisfied.

By taking advantage of the SDN network, the SDN controller is used to measure the E2E delay between UE $k$'s eNB and cloudlet $i$ in each slot, denoted as $d_{i,k}$. Thus, the E2E delay between UE $k$'s eNB and its cloudlet in time slot $t$ is:

$$D_k = \sum_{i \in I} d_{i,k} x_{i,k}^t. \tag{11}$$

## III. Problem Formulation and Analysis

Owing to the spatial dynamics of energy demands among cloudlets, some cloudlets are able to satisfy the energy demand by its available green energy while others need to consume on-grid energy. The on-grid energy consumption of cloudlets can be minimized through migrating Avatars to the cloudlets with excessive green energy. However, Avatar migrations consume significant energy of corresponding cloudlets. We should design an optimal Avatar migrations strategy, i.e., EARN, by considering the energy consumption for running Avatars and that for migrating Avatars in each cloudlet such that the total on-grid energy consumption is minimized in time slot $t+1$. We formulate EARN as follows:

$$P1: \min_{x_{i,k}^{t+1}} \sum_{i \in I} \rho_i \tag{12}$$

$$s.t.\ \rho_i \geq \sum_{k \in \mathcal{K}} x_{i,k}^{t+1} \left(\frac{P^s}{\tau} + \alpha u_k\right) T + E_i^{mig}(x_{i,k}^{t+1}) - G_i T, \quad \forall i \in I, \tag{13}$$

$$\rho_i \geq 0, \quad \forall i \in I, \tag{14}$$

$$\sum_{i \in I} d_{i,k} x_{i,k}^{t+1} \leq \varepsilon, \quad \forall k \in \mathcal{K}, \tag{15}$$

$$\sum_{i \in I} x_{i,k}^{t+1} = 1, \quad \forall k \in \mathcal{K}, \tag{16}$$

$$\frac{\sum_{k \in K} x_{i,k}^{t+1}}{\tau} \leq m_i, \quad \forall i \in I, \tag{17}$$

where $\epsilon$ is SLA given by the provider, and $m_i$ represents the maximum number of servers in cloudlet $i$. Constraints (13) and (14) ensure $\rho_i = max(P_i T + E_i^{mig} - G_i T, 0)$. Constraint (15) represents that the E2E delay for each UE should satisfy SLA. Constraint (16) ensures that each Avatar is assigned to a specific cloudlet. Constraint (17) imposes the number of the required servers not to be more than the maximum number of servers in each cloudlet.

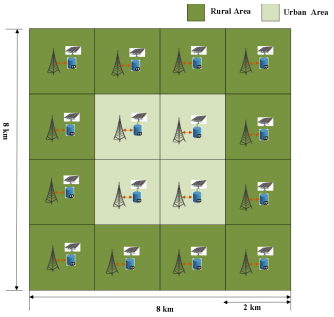


Fig. 2. Network topology.

*Theorem 1:* The problem of P1 is an NP-hard problem.

*Proof:* Suppose there are two cloudlets; the capacity of each cloudlet is infinite, Therefore, every Avatar can be severed by any of the two cloudlets without violating their capacity constraints. We assume that the green energy generation rate of each cloudlet is the same, which equals to $G$, while the total power consumption of running all Avatars is equal to the total green energy generation in the network. Moreover, $\sigma^s$,$\sigma^d$, $\beta^s$, and $\beta^d$ equal to zero; consequently, $E_i^{mig} = 0$. So, the original problem P1 can be transformed to

$$R1: \min \sum_{i=1}^{2} \max\{P_i T - GT, 0\} \tag{18}$$

$$s.t. \sum_{i=1}^{2} P_i T = 2GT. \tag{19}$$

Here, the optimal solution of R1 is to assign the energy demands into the two cloudlets equally, i.e., this becomes a partition problem, which is NP-hard. Thus, the partition problem is reducible to P1, i.e., P1 is NP-hard. ■

Since P1 is NP-hard, we propose to use the Mixed-Integer Linear Programming (MILP) toolbox in the CPLEX solver to solve the problem. The branch-and-cut algorithm is applied by the MILP toolbox to search for the suboptimal solution of the MILP problem. The branch-and-cut algorithm executes the branch and bound process and applies cutting planes to reduce the number of branches required to solve the problem.

## IV. Simulation Results

We set up the simulation to demonstrate the performance of EARN. For comparisons, we select the other two schemes, i.e., GEAR [5] and Follow me AvataR (FAR). GEAR is to minimize the on-grid energy consumption without considering the migration cost. FAR tries to minimize the E2E delay by choosing the closest cloudlet to host the Avatar. The network topology is shown in Fig. 2, which includes 16 cloudlet-eNB pairs within an area of 64 $km^2$, which is divided into two areas: urban area and rural area. The cloudlet's capacity can be randomly chosen from 10 to 30 servers, while each server can host 16 Avatars at most. The idle power of each active server (i.e., $P^s$) is 80 Watt, and the maximum power

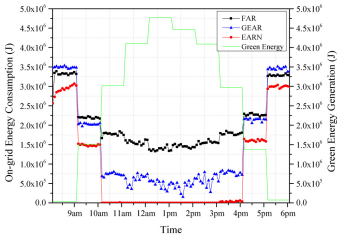


Fig. 3. On-grid Energy Consumption during a day.

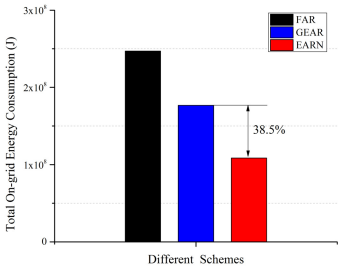


Fig. 4. Total on-grid energy consumption in one day.

consumption of a server (i.e., the power consumption of a server when its CPU utilization is 100%) is 160 Watt. The CPU utilization of each Avatar is randomly chosen from 30% to 100%. The data volume incurred by an Avatar migration is chosen from 1 to 3 Gbit randomly. Based on the experiments in [8], $\sigma^s$ and $\sigma^d$ are set to be 0.256, and $\beta^s$ and $\beta^d$ are set to be 10.08. In every time slot (5 $mins$), each UE randomly selects a moving speed between 0 and 10 $m/s$, and moves towards its destination. The location of UEs' destinations are determined according to a normal distribution $N(4\ km,$ $1.4\ km)$. In addition, the solar panel size of each cloudlet is randomly selected between 4 and 6 $m^2$. The local daily solar radiation data trace (Millbrook, NY in Jan. 1st. 2015) is adopted as the solar radiation within one day [9], while the efficiency for converting solar radiation into electricity is 46%. Denote $A_i$ as the panel size of cloudlet $i$, $g$ as the solar radiation, and $\gamma$ as the converting efficiency from solar radiation to electricity, and thus the green energy generation rate of cloudlet $i$ can be expressed as $G_i = A_i g \gamma$.

Fig. 3 shows that EARN remarkably saves on-grid energy as compared to FAR and GEAR within one day. When the green energy generation rate is low, by considering the migration cost of Avatars, EARN enables less Avatar migrations as compared to FAR and GEAR, thus leading to less on-grid consumption of GCNs. As the green energy generation rate increases, EARN determines the Avatar migration based on the green energy utilization of different cloudlets and the migration cost, and thus the on-grid energy consumption is still much lower than that of either FAR or GEAR. Fig. 4 illustrates the total on-grid energy consumption in one day. EARN saves 38.5% on-grid energy as compared to GEAR, and saves 56% as compared to FAR. Fig. 5 compares the total on-grid energy consumption of the three schemes in one day by varying the

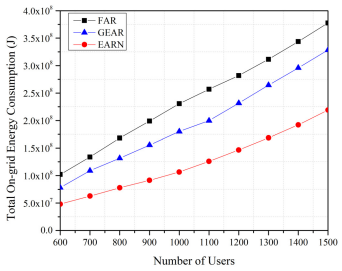


Fig. 5. Total on-grid Energy vs. number of UEs.

number of UEs in the network. We can see that the on-grid energy consumption of the three schemes grows as the number of UEs increases. EARN only migrates the Avatar whose on-grid energy reduction is higher than its migration cost, in order to minimize the on-grid energy consumption of GCNs. In contrast, GEAR and FAR lead to the high migration cost when Avatars are migrated frequently. Therefore, when the number of UEs varies, EARN achieves much less on-grid energy consumption of GCNs than that of GEAR and FAR.

## V. Conclusion

In this letter, we have proposed the Energy driven AvataR migratioN (EARN) scheme for GCNs. EARN is to minimize the total on-grid energy consumption by considering the migration cost (in terms of the energy consumption introduced in the Avatar migration), while ensuring the service level agreement (SLA) for each UE. Simulation results have verified the performance of the proposed EARN scheme.

## References

[1] X. Sun and N. Ansari, "EdgeIoT: Mobile edge computing for the Internet of Things," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 22–29, Dec. 2016.

[2] X. Sun and N. Ansari, "PRIMAL: Profit maximization avatar placement for mobile edge computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[3] M. Giacobbe *et al.*, "An approach to reduce carbon dioxide emissions through virtual machine migrations in a sustainable cloud federation," in *Proc. Sustain. Internet ICT Sustain. (SustainIT)*, Madrid, Spain, Apr. 2015, pp. 1–4.

[4] L. Gkatzikis and I. Koutsopoulos, "Migrate or not exploiting dynamic task migration in mobile cloud computing systems," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 24–32, Jun. 2013.

[5] X. Sun, N. Ansari, and Q. Fan, "Green energy aware avatar migration strategy in green cloudlet networks," in *Proc. IEEE 7th Int. Conf. Cloud Comput. Technol. Sci. (CloudCom)*, Vancouver, BC, Canada, Nov./Dec. 2015, pp. 139–146.

[6] S. Pelley, D. Meisner, T. F. Wenisch, and J. W. VanGilder, "Understanding and abstracting total data center power," in *Proc. Workshop Energy Efficient Design (WEED)*, Austin, TX, USA, Jun. 2009, pp. 1–6.

[7] T. Han and N. Ansari, "A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1038–1051, Apr. 2016.

[8] H. Liu, H. Jin, C.-Z. Xu, and X. Liao, "Performance and energy modeling for live migration of virtual machines," *Cluster Comput.*, vol. 16, no. 2, pp. 249–264, Jun. 2013.

[9] National Climatic Data Center. *National Climatic Data Center Daily Solar Radiation Data Trace*, accessed on Mar. 21, 2017. [Online]. Available: https://www1.ncdc.noaa.gov/pub/data/uscrn/products/hourly02/2015/CRNH0203-2015-NY_Millbrook_3_W.txt.