

# CCOMPASSION: A Hybrid Cloudlet Placement Framework over Passive Optical Access Networks

Sourav Mondal<sup>1</sup>, Goutam Das<sup>2</sup> and Elaine Wong<sup>1</sup>

<sup>1</sup>Dept of Electrical and Electronic Engineering, The University of Melbourne, VIC 3010, Australia.

<sup>2</sup>G. S. Sanyal School of Telecommunications, Indian Institute of Technology Kharagpur, 721302, India.

**Abstract**—Cloud-based computing technology is one of the most significant technical advents of the last decade and extension of this facility towards access networks by aggregation of cloudlets is a step further. To fulfill the ravenous demand for computational resources entangled with the stringent latency requirements of computationally-heavy applications related to augmented reality, cognitive assistance and context-aware computation, installation of cloudlets near the access segment is a very promising solution because of its support for wide geographical network distribution, low latency, mobility and heterogeneity. In this paper, we propose a novel framework, Cloudlet Cost Optimization over PASSive Optical Network (CCOMPASSION), and formulate a nonlinear mixed-integer program to identify optimal cloudlet placement locations such that installation cost is minimized whilst meeting the capacity and latency constraints. Considering urban, suburban and rural scenarios as commonly-used network deployment models, we investigate the feasibility of the proposed model over them and provide guidance on the overall cloudlet facility installation over optical access network. We also study the percentage of incremental energy budget in the presence of cloudlets of the existing network. The final results from our proposed model can be considered as fundamental cornerstones for network planning with hybrid cloudlet network architectures.

**Index Terms**—Cloudlet; low latency; non-linear mixed-integer programming; optical access network;

## I. INTRODUCTION

Riding on the shoulders of the gigantic technical advancements during the past few decades, the vision of *ubiquitous computing in the 21st century* published by Weiser in his seminal paper [1], has managed to land at a solid ground of reality of present times. In smart cities and smart homes, almost every object around us is expected to have some computational capability and connectivity to the Internet. According to the prediction from Gartner, when Internet of Things (IoT) becomes completely commercialized along with the standardization of 5G technologies, nearly 20 billion of devices will connect to the Internet by 2020 [2]. In conjunction with this, mobile data traffic has already increased by 63% in 2016, and by 2021, this is expected to increase by seven folds further [3].

The technology that provides distributed computation facilities on battery-powered, portable and mobile devices which are interconnected via mobile communication standards and protocols, is known as *Mobile Computing* [4]. However, along with several essential advantages, there exist some crucial limitations in mobile computing, e.g., resource constraints on battery, degradation of Quality-of-Service (QoS), limited

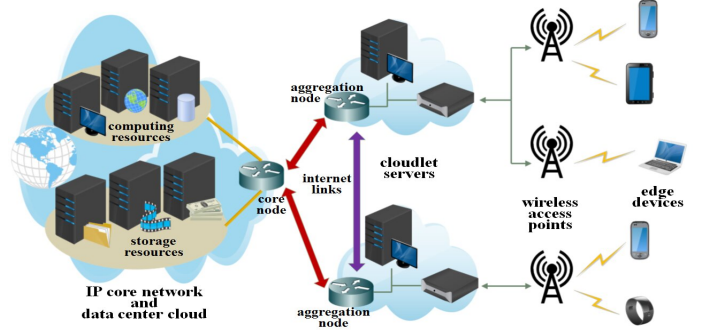


Fig. 1: A simplified top level architecture of cloud-cloudlet-edge devices network infrastructure [7].

bandwidth and connection latency. Some typical smart applications related to augmented reality, autonomous transport, and cognitive assistance, demand a network latency of 1-10 ms [5]. Moreover, mobile devices are designed to be lightweight and portable, which makes them poor in computational resources and memory. Therefore, most often meeting both the latency requirements as well extended battery life become a trivial challenge for the mobile devices and offloading the major computational tasks to cloud servers appeared to be highly essential [6].

The process of offloading the intensive computation to the cloud server by the mobile devices is known as *cyber foraging* [8]. However, cloud servers are installed at geographically secure locations, beyond the core backbone network, and hence the cloud server access should not be taken for granted. In disaster recovery scenarios or secure military operation areas, where basic network infrastructures have been disrupted, remote cloud access may become impossible [9]. Due to multi-hop network connectivity, not only the end-to-end network latency becomes large, as pointed out by authors in [6], maintenance of privacy, security and reliability of the offloaded data by users becomes challenging too. Nonetheless, when the end-to-end latency is less than QoS latency requirement, the best decision for the mobile devices is to offload [6].

To mitigate these aforementioned challenges, the authors of [7] proposed the idea of *cloudlets* as a small-scale but distributed cloud computation facility at close proximity of end-users in the access segment. A typical cloudlet-empowered cloud computation network consist of a *three-tier architecture*, as shown in Fig. 1.

The authors of [10] made a comparative study to understand the impact of distance on end-user experience especially with transient display applications. Both remote cloud and single-hop cloudlets were used to offload computational tasks in their experiment. Through their experimental results, the authors of [5] confirmed that execution of the offloaded task by a nearby computer is always more latency efficient over a distant server. The authors of [11] further demonstrated that by adopting efficient virtual machine (VM) provisioning algorithms, the computation latency can be reduced to a great extent for tactile applications.

Installation of cloudlet facilities at suitable locations is becoming a primary design challenge in satisfying the QoS latency requirements of latency-sensitive applications. The authors of [12] presented link-path mixed integer linear programming (MILP) formulations for *static planning* as well as *dynamic planning* with VM bulk migration and live migration over a 4G cellular network. The authors of [13] formulated an Integer Linear Programming (ILP) problem for cloudlet placement to minimize the *cloudlet access delay* in a Wireless Metropolitan Area Network (WMAN) with several wireless Access Points (WAPs) and proposed an efficient heuristic algorithm for faster convergence of their optimization model. On a similar note, the authors of [14] proposed a heuristic algorithm to minimize the *system response time* by optimally allocating cloudlets to end-users in a WMAN. Moreover, the authors of [15] proposed an architecture that minimizes the carbon footprint of cloudlet network by adopting harnessing and usage of green energy. Note that the authors in these works had considered only wireless access media.

On the contrary, the authors of [16] highlighted several benefits of using *optical access networks* for cloud computing applications due to its cost-effectiveness, high-bandwidth data transmission, efficient network virtualization and network scalability. Moreover, the authors of [17] outlined a novel cloud and cloudlet empowered fiber-wireless (FiWi)-heterogeneous network architecture for LTE-Advanced (LTE-A). The authors of [18] proposed a cloudlet-aware resource management scheme to reduce offload delay and prolong mobile-devices' battery life by incorporating offloading activities into the underlying FiWi dynamic bandwidth allocation process. In [19], the authors implemented a cloudlet framework with control server at the central office (CO) of a fiber-based access network for human-machine interactive applications.

However, to the best of our knowledge, the optimized design and planning of *hybrid placement of cloudlets based on TDM-PON support architecture* is an unexplored area. Optical access network standards like 10G(E)-PON has been the preferred technological solution of broadband access due to its low cost per bit and very high bandwidth support. Moreover, the evolution of radio-over-fiber (RoF) and FiWi makes the Internet accessible to a huge number of end users [20]. This motivates us to propose a hybrid cloudlet support architecture based on an existing 10G(E)-PON infrastructure and develop a framework that minimizes the capital expenditure (CAPEX), mainly subjected to the cloudlet computational capacity and

latency constraints.

In this paper, we propose the end-to-end hybrid cloudlet placement architecture of Cloudlet Cost OptiMization over PASSive Optical Networks (CCOMPASSION), with an existing tree-and-branch TDM-PON based network infrastructure. The cloudlets can be suitably placed in the field, remote node (RN) and optical line terminal (OLT) or CO locations depending upon computation task requirements. We formulate a mixed-integer non-linear programming (MINLP) problem to identify cost efficient cloudlet placement locations subjected to optimized connectivity with optical network units (ONUs). In this work, our main focus is on *static planning* of the network. Our model identifies optimal cloudlet placement locations as well as optimal computational resources, i.e., number of racks in each cloudlet. This helps to prevent under or over-utilization of the resources. We further *linearize the objective function* so that our design tool can compute the optimal solution with a much faster convergence rate. Our primary design constraint of the optimization problem is to *achieve a very low end-to-end latency* ideal for tactile applications, i.e., 1 ms, 10 ms and 100 ms. We validate our model against urban, suburban and rural scenarios and present the comparative deployment cost, workload distribution among field, RN and CO cloudlets, average number of racks per cloudlet and the increase of energy budget over a typical 10G(E)-PON, against the targeted latency figures of 1 ms, 10 ms and 100 ms.

The rest of this paper is organized as follows. In Section II, the details of TDM-PON based hybrid cloudlet placement network architecture are described. In Section III, the system model and the MINLP formulation are presented. The simulation results are presented for discussion in Section IV. Finally, Section V summarizes the key observations of the entire work.

## II. THE CCOMPASSION FRAMEWORK

The proposed end-to-end physical network infrastructure of hybrid cloudlet support network based on TDM-PON access, i.e., Cloudlet Cost OptiMization over PASSive Optical Network (CCOMPASSION) is illustrated in Fig. 2. With a conventional wireless access network, the base stations (BSs) or WAPs form a star connection with the aggregation node or CO [12]. Nevertheless, with optical access networks, the connection among the ONUs and cloudlets in the field, RN or CO need not necessarily follow a star topology. With a tree-and-branch network topology, as shown in Fig. 2, the split ratio of the passive splitter installed at RN is chosen according to number of end-users and their bandwidth requirements, e.g., 1:4 or 1:8 or 1:16 [20].

In a hybrid cloudlet placement scheme, cloudlets can be installed either in the field, RN or CO locations, so that close proximity with the end-user is maintained. Additionally, WAPs with WiFi or LTE/5G support can be integrated with ONUs to provide network coverage to a larger number of end-users. A typical cloudlet rack is expected to support around 2500 VMs and with a higher the number of end-user connectivity, higher number of racks are required [12]. The cloud servers are in

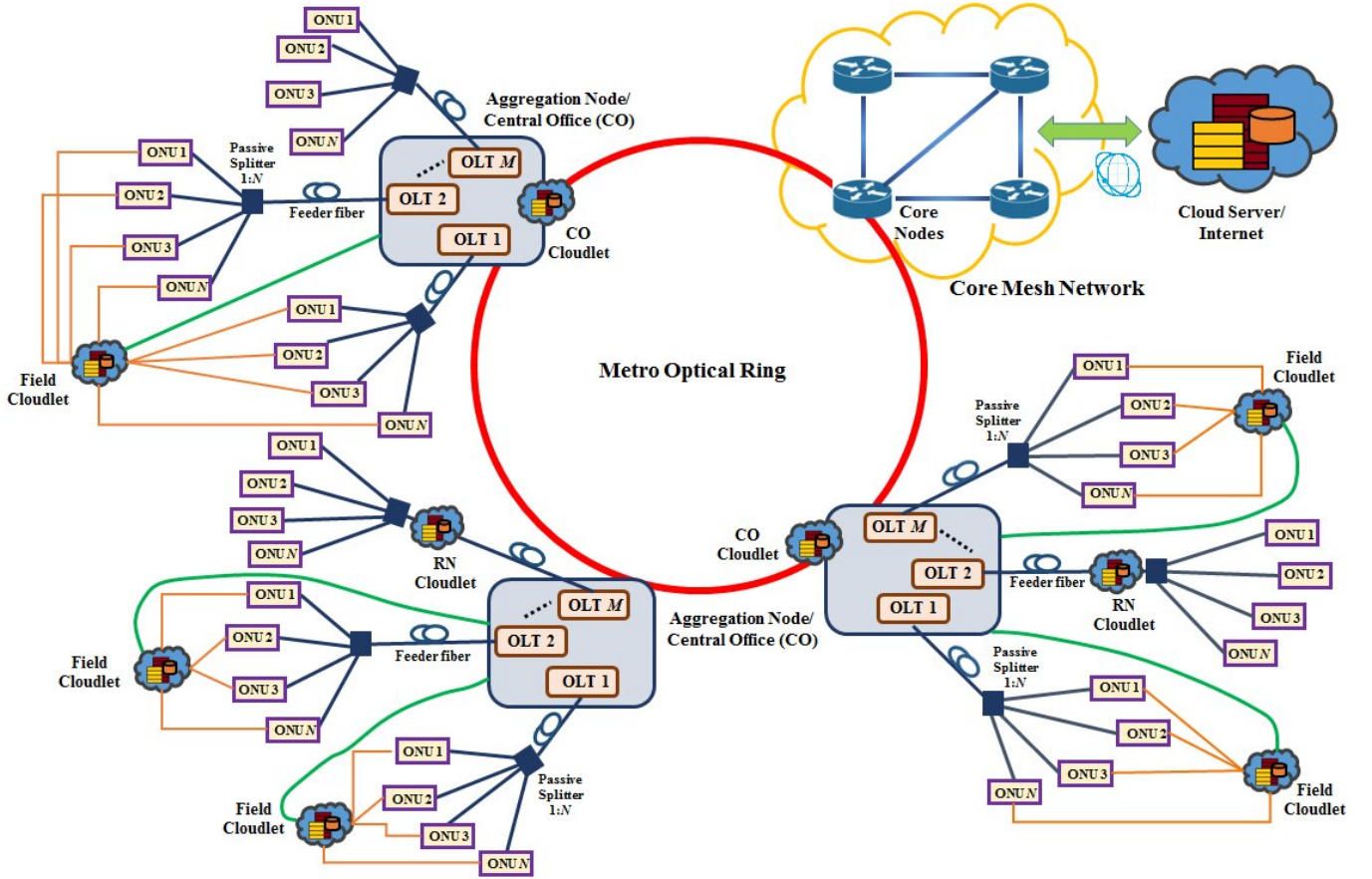


Fig. 2: Complete physical network infrastructure of the proposed CCOMPASSION framework showing cloudlet placement locations in the field, remote node and central office.

general placed at a geographically secured location, beyond backhaul metro networks, as shown in Fig. 2.

In the CCOMPASSION framework, for the *field cloudlets*, point-to-point fiber links (with data rate 1 Gbps) between a cloudlet and each ONU connected to it (brown links) are installed. An additional point-to-point fiber link between each cloudlet and the nearest CO (green links) is also installed to create the cloudlet-cloud connectivity.

However, the idea of *installing cloudlets at RN locations* is more beneficial over field cloudlets, especially in dense deployment scenarios, due to the fact that installation of a field cloudlet carries additional point-to-point new fiber link installation costs, whereas installing a cloudlet at the RN allows us to re-use the existing fiber links between ONUs and RNs. We can use just one or more additional wavelengths to be time-shared amongst the ONUs for communicating with the cloudlet placed at the RN.

Interestingly, *installing a cloudlet at the CO* is the most economical option amongst all three as the infrastructure setup cost is lowest, but the key drawback arises in terms of transmission latency and bandwidth. The COs are geographically farther from the ONUs, field and RN locations, and the ONUs use the default uplink and downlink channels for communication with cloudlets placed at the CO. In general,

the average background load for downlink and uplink can be considered to be around 70% and 50%, respectively [21] and hence, the ONUs can use the unused bandwidth for communication with cloudlet placed at CO. In this case, the transmission latency can potentially be higher under high load conditions.

### III. SYSTEM MODEL OF CCOMPASSION

We present the system model and MINLP formulation in this section. We assume that the CO, RN and ONU locations are already known to us irrespective of the deployment scenario. We pre-identify some potential field-cloudlet placement locations by using data clustering techniques on the ONU locations.

#### A. Hybrid Cloudlet Network Optimization Parameters

We assume each cloudlet contains a finite set of processors, e.g.,  $m \in K = \{1, \dots, 10\}$  racks with service rate of  $\mu$  for each rack. Such a system should be an M/M/m queueing system if it performs the incoming tasks sequentially and total processing delay can be computed using *Erlang-C formula* [14]. Nonetheless, here we model the system differently by assuming that service process of each task is independent of each other and are exponentially distributed with service rate  $\mu$ , if a task

TABLE I: Hybrid Cloudlet Network Optimization Parameters

Symbol	Definition
$A$	Set of possible field cloudlet locations
$B$	Set of possible RN cloudlet locations
$C$	Set of possible CO/OLT cloudlet locations
$D$	Set of locations of existing ONUs in the TDM-PON network
$K$	Set of number of racks in a cloudlet location [1,10]
$\alpha$	Cost of installing a single rack cloudlet at field, RN or CO
$\xi_a$	New infrastructure installation cost at field location $a \in A$
$\xi_b$	New infrastructure installation cost at RN location $b \in B$
$\xi_c$	New infrastructure installation cost at CO location $c \in C$
$\eta$	Cost of installing new optical fiber per kilometer
$L_{max}$	Maximum allowed fiber length between field cloudlet $a \in A$ and ONU $d \in D$
$L_{ad}$	Actual optical fiber length between field cloudlet $a \in A$ and ONU $d \in D$
$x_{bd}^{adj}$	Element of adjacency matrix denoting connectivity between ONUs $d \in D$ and RNs $b \in B$
$x_{cd}^{adj}$	Element of adjacency matrix denoting connectivity between ONUs $d \in D$ and COs $c \in C$
$BW_{dz}$	Bandwidth of the optical fiber link between cloudlet at $z \in \{a, b, c\}$ and ONU $d \in D$ in the uplink direction
$BW_{zd}$	Bandwidth of the optical fiber link between cloudlet at $z \in \{a, b, c\}$ and ONU $d \in D$ in the downlink direction
$D_{zd}$	Transmission latency between cloudlet at $z \in \{a, b, c\}$ and ONU at $d \in D$
$D_{QoS}$	Maximum allowed latency according to desired quality of service (QoS)
$\mu$	Average service rate of each rack in a cloudlet or cloud
$n_\lambda$	Number of wavelengths shared by ONUs to communicate to the corresponding RN cloudlet
$\lambda_z$	Total task arrival rate at cloudlet at $z \in \{a, b, c\}$ , from all $d \in D$ ONUs connected to it
$\lambda_d$	Task arrival rate from a single ONU at $d \in D$
$\mu_z$	Total service rate of a cloudlet at $z \in \{a, b, c\}$ containing $m \in K$ racks
$\Lambda$	Mean transmission latency to offload a service request to remote cloud
$\sigma_{ul}$	Average number of bits an ONU $d \in D$ sends to cloudlet $z \in \{a, b, c\}$ for processing
$\sigma_{dl}$	Average number of bits an ONU $d \in D$ receives from cloudlet $z \in \{a, b, c\}$ after processing
$\beta_{dl}$	Background load in the downlink of the considered TDM PON
$\beta_{ul}$	Background load in the uplink of the considered TDM PON

is serviced by a single rack. We also assume that only one task is performed at a particular instant, homogeneously by all the racks combined, i.e., parallel processing of a task by multiple racks. The task arrival process to any cloudlet from the corresponding ONUs are also assumed to be independent and Poisson distributed with an arrival rate  $\lambda$ . Based on these considerations, we model the cloudlets as M/M/1 queueing systems. In CCOMPASSION, the number of racks  $m$  may vary from cloudlet to cloudlet, depending on number of ONUs they are connected to. If a cloudlet location (either in the field or RN or OLT) is not chosen, then the total number of racks should also be zero for that particular location.

The total incoming service request rate to a cloudlet is calculated by adding all the service requests arriving from all ONUs connected to that cloudlet. The operational cloudlets may choose either to process the total incoming task request all by itself or offload some fraction to the remote cloud, while meeting the latency constraints. Offloading incoming tasks to remote cloud is obviously a cost minimizing opportunity for cloudlets, but meeting the latency requirements becomes challenging due to very high end-to-end latency from remote cloud. When a cloudlet chooses to process the total incoming

task request all by itself, the average processing time for any cloudlet in the field, RN or OLT location  $z \in \{a, b, c\}$  is expressed as follows [22]:

$$\tau_z^{cloudlet} = \frac{1}{(\mu_z - \lambda_z)}, \forall z \in \{a, b, c\} \quad (1)$$

The core cloud possesses huge computational and storage resources and hence can be assumed as an M/M/ $\infty$  queuing system. The total latency for a task when offloaded to cloud can be expressed as follows [14]:

$$\tau^{cloud} = \Lambda + \frac{1}{\mu} \quad (2)$$

A summary of network optimization parameters with their definitions are presented in Table I for convenience.

### B. The Decision Variables

To formulate the objective function and constraints of our MINLP problem, we choose a set of binary and fractional decision variables as follows. The parameters  $x_a$ ,  $x_b$  and  $x_c$  are binary variables, indicating the decision whether to install a cloudlet in a field site  $a \in A$ , in a RN site  $b \in B$  and in a CO site  $c \in C$ , respectively.

$$x_a = \begin{cases} 1; & \text{if a cloudlet is installed} \\ & \text{in the field location } a \in A \\ 0; & \text{otherwise} \end{cases}$$

$$x_b = \begin{cases} 1; & \text{if a cloudlet is installed} \\ & \text{at RN location } b \in B \\ 0; & \text{otherwise} \end{cases}$$

$$x_c = \begin{cases} 1; & \text{if a cloudlet is installed} \\ & \text{at CO location } c \in C \\ 0; & \text{otherwise} \end{cases}$$

The parameters  $n_{am}$ ,  $n_{bm}$  and  $n_{cm}$  are binary variables to indicate whether  $m$  number of racks are chosen to be placed in cloudlet location  $a \in A$ ,  $b \in B$  and  $c \in C$ , respectively.

$$n_{am} = \begin{cases} 1; & \text{if the cloudlet at } a \in A \\ & \text{contains } m \text{ racks} \\ 0; & \text{otherwise} \end{cases}$$

$$n_{bm} = \begin{cases} 1; & \text{if the cloudlet at } b \in B \\ & \text{contains } m \text{ racks} \\ 0; & \text{otherwise} \end{cases}$$

$$n_{cm} = \begin{cases} 1; & \text{if the cloudlet at } c \in C \\ & \text{contains } m \text{ racks} \\ 0; & \text{otherwise} \end{cases}$$

We define the following binary variables as the product of some particular combinations of the aforementioned binary variables, i.e.,  $\bar{x}_a = x_a n_{am}$ ,  $\bar{x}_b = x_b n_{bm}$  and  $\bar{x}_c = x_c n_{cm}$ .

The parameters  $x_{ad}$ ,  $x_{bd}$  and  $x_{cd}$  are binary variables to indicate whether the connectivity between an ONU at  $d \in D$

and a cloudlet at  $a \in A$ ,  $b \in B$  and  $c \in C$  respectively, is created or not.

$$x_{ad} = \begin{cases} 1; & \text{if ONU } d \in D \text{ is connected} \\ & \text{to field cloudlet } a \in A \\ 0; & \text{otherwise} \end{cases}$$

$$x_{bd} = \begin{cases} 1; & \text{if ONU } d \in D \text{ is connected} \\ & \text{to RN cloudlet } b \in B \\ 0; & \text{otherwise} \end{cases}$$

$$x_{cd} = \begin{cases} 1; & \text{if ONU } d \in D \text{ is connected} \\ & \text{to CO cloudlet } c \in C \\ 0; & \text{otherwise} \end{cases}$$

The parameters  $\varphi_a$ ,  $\varphi_b$  and  $\varphi_c$  are fractions lying within the range  $[0, 1]$  and indicate the portion of the total incoming task a cloudlet at  $a \in A$ ,  $b \in B$  and  $c \in C$  choses to process by itself.

### C. Objective Function and Constraints

We formulate the objective function to minimize the overall cloudlet installation expenditures (CAPEX) as follows:

$$\min \left\{ \sum_{a \in A} \sum_{m \in K} (m\alpha) \bar{x}_a + \sum_{b \in B} \sum_{m \in K} (m\alpha) \bar{x}_b + \sum_{c \in C} \sum_{m \in K} (m\alpha) \bar{x}_c \right. \\ \left. + \sum_{a \in A} \sum_{d \in D} \eta L_{ad} x_{ad} + \sum_{a \in A} \xi_a x_a + \sum_{b \in B} \xi_b x_b + \sum_{c \in C} \xi_c x_c \right\} \quad (3)$$

The above expression includes cost components arising from multiple sectors. The first, second and third terms denote the total cost of racks installed in the field, the RN and the CO cloudlet locations, respectively. The fourth term denotes the total cost of installing new fiber connections among ONUs and field cloudlets. Finally, the fifth, sixth and seventh terms denote the additional infrastructure setup costs in the field, the RN and the CO cloudlet locations, respectively.

Actually the first three terms of the objective function (3) are quadratic in nature due to  $\bar{x}_a = x_a n_{am}$ ,  $\bar{x}_b = x_b n_{bm}$  and  $\bar{x}_c = x_c n_{cm}$ , but we linearize them using three linear inequalities for each of them, as shown in (4)-(6) [23].

$$\bar{x}_a \leq n_{am}, \bar{x}_a \leq x_a, \bar{x}_a \geq n_{am} + x_a - 1, \forall a \in A, m \in K \quad (4)$$

$$\bar{x}_b \leq n_{bm}, \bar{x}_b \leq x_b, \bar{x}_b \geq n_{bm} + x_b - 1, \forall b \in B, m \in K \quad (5)$$

$$\bar{x}_c \leq n_{cm}, \bar{x}_c \leq x_c, \bar{x}_c \geq n_{cm} + x_c - 1, \forall c \in C, m \in K \quad (6)$$

The set of constraints (7)-(9) are the *capacity constraints* and are used to ensure that only one value of  $m$  (the number of racks in a particular cloudlet) is chosen. If a particular cloudlet is not chosen to be activated at either of field, RN or CO location, then no racks are placed at that location.

$$x_a \leq \sum_{m \in K} n_{am} \text{ and } \sum_{m \in K} n_{am} \leq 1, \forall a \in A, m \in K \quad (7)$$

$$x_b \leq \sum_{m \in K} n_{bm} \text{ and } \sum_{m \in K} n_{bm} \leq 1, \forall b \in B, m \in K \quad (8)$$

$$x_c \leq \sum_{m \in K} n_{cm} \text{ and } \sum_{m \in K} n_{cm} \leq 1, \forall c \in C, m \in K \quad (9)$$

Constraints described next are *connectivity constraints*. Constraint (10) denotes that any ONU can be chosen to be connected to a field cloudlet, only if their Euclidean separation is  $\leq L_{max}$ . Constraint (11) denotes that a field cloudlet will be activated if there exists at least one connected ONU.

$$x_{ad} \leq \max \left[ 0, \left( \frac{L_{ad} - L_{max}}{|L_{ad} - L_{max}|} \right) \right], \forall a \in A, d \in D \quad (10)$$

$$x_a \geq x_{ad}, \forall a \in A, d \in D \quad (11)$$

In a TDM-PON, any arbitrary ONU cannot be connected to any RN cloudlet, rather only those ONUs can be connected which already has an existing fiber connection to a particular RN and the same condition applies for connectivity between ONU and CO as well. These conditions are enforced by constraints (12)-(13). Furthermore, constraints (14)-(15) imply that a cloudlet in the RN and in the CO can be activated if there is at least one ONU connected to it, respectively.

$$x_{bd} \leq x_{bd}^{adj}, \forall b \in B, d \in D \quad (12)$$

$$x_{cd} \leq x_{cd}^{adj}, \forall c \in C, d \in D \quad (13)$$

$$x_b \geq x_{bd}, \forall b \in B, d \in D \quad (14)$$

$$x_c \geq x_{cd}, \forall c \in C, d \in D \quad (15)$$

The constraint in (16) ensures that every ONU is connected to one and only one cloudlet.

$$\sum_{a \in A} x_{ad} + \sum_{b \in B} x_{bd} + \sum_{c \in C} x_{cd} = 1, \forall d \in D \quad (16)$$

Finally, the most important constraint, i.e., the *QoS latency constraint* that provides the upper limit of total allowed latency for each ONU, is described in (17).

$$x_{ad} \left[ \varphi_a \left\{ \frac{1}{\mu_a - \lambda_a \varphi_a} + D_{ad} + \frac{\sigma_{ul}}{BW_{da}} + \frac{\sigma_{dl}}{BW_{ad}} \right\} \right. \\ \left. + (1 - \varphi_a) \left( \Lambda + \frac{1}{\mu} \right) \right] \\ + x_{bd} \left[ \varphi_b \left\{ \frac{1}{\mu_b - \lambda_b \varphi_b} + D_{bd} + \frac{\sigma_{ul} \sum_{d \in D} x_{bd}}{n_{\lambda} BW_{db}} + \frac{\sigma_{dl} \sum_{d \in D} x_{bd}}{n_{\lambda} BW_{bd}} \right\} \right. \\ \left. + (1 - \varphi_b) \left( \Lambda + \frac{1}{\mu} \right) \right] \\ + x_{cd} \left[ \varphi_c \left\{ \frac{1}{\mu_c - \lambda_c \varphi_c} + D_{cd} + \frac{\sigma_{ul} \sum_{d \in D} x_{cd}}{BW_{dc} - \beta_{ul}} + \frac{\sigma_{dl} \sum_{d \in D} x_{cd}}{BW_{cd} - \beta_{dl}} \right\} \right. \\ \left. + (1 - \varphi_c) \left( \Lambda + \frac{1}{\mu} \right) \right] \leq D_{QoS}, \forall a \in A, b \in B, c \in C, d \in D \quad (17)$$

Note that, the total task arrival rate to any cloudlet, depending on connectivity among ONUs and cloudlets, and the total service rate of any cloudlet is computed as follows:

$$\lambda_z = \sum_{d \in D} x_{zd} \lambda_d, \text{ and } \mu_z = \sum_{m \in K} m n_{zm} \mu, \forall z \in \{a, b, c\} \quad (18)$$

The constraint (17) consists of the weighted sum of total latencies when the incoming tasks are either executed in the field or RN or CO cloudlets. However, due to (10) each ONU can connect to only one cloudlet. Each of these



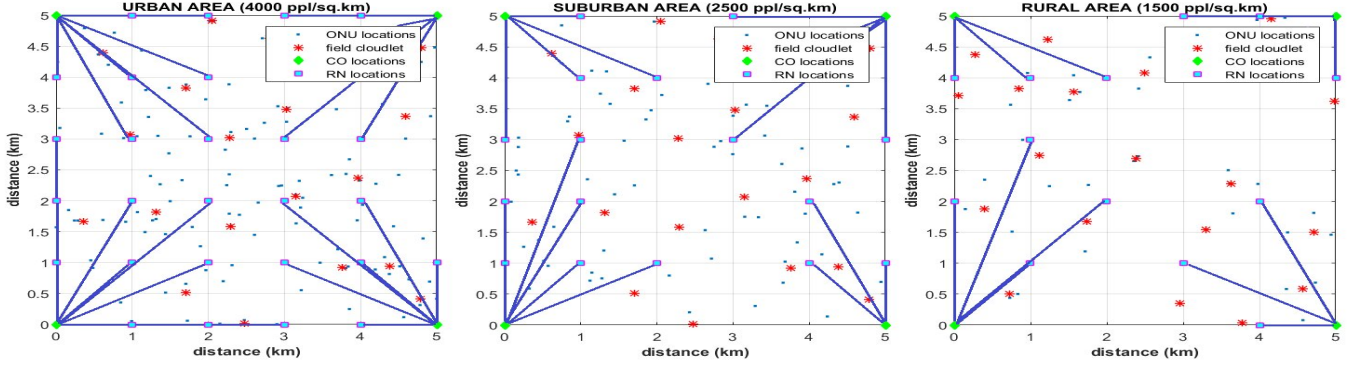


Fig. 3: Randomly generated urban, suburban and rural areas for cloudlet deployment with existing TDM-PON network infrastructure with 1:4 split ratio (The blue links from CO to RN locations represent the feeder fibers. Distribution fibers from RN to each ONU are not shown).

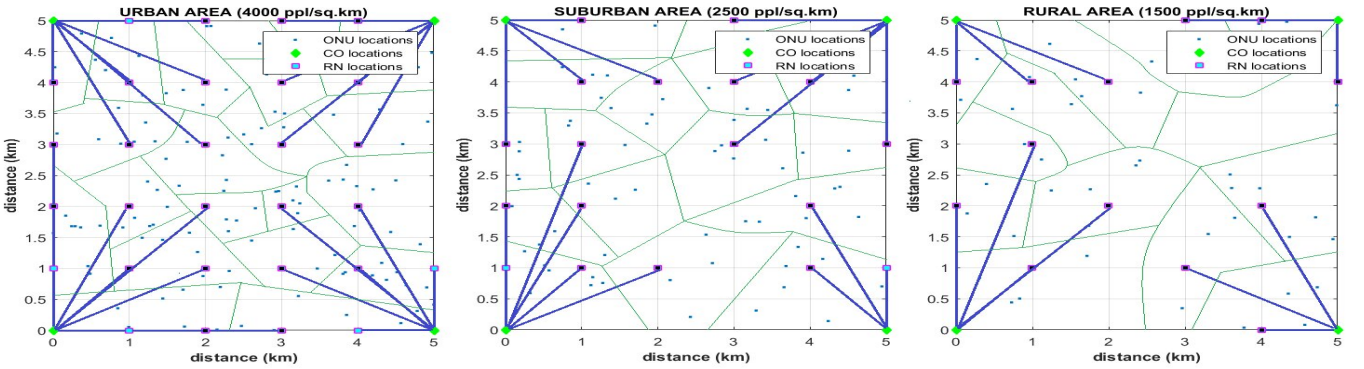


Fig. 4: Optimal cloudlet placement locations over the considered urban, suburban and rural areas for 1 ms latency constraint using CCOMPASSION (cloudlets are installed at darkened face RNs and all COs).

weighted latency terms consists of four components. The first component denotes the processing time, the second component denotes transmission time, the third component denotes the time to offload total bits for the task by ONU to cloudlet and the fourth component denotes the time to receive total bits post-processing by ONU from cloudlet. The third and fourth components are evaluated differently from each other, depending on the respective data transmission schemes from ONUs to cloudlets in the field, at RN or at CO.

#### IV. CCOMPASSION FRAMEWORK EVALUATION

We implemented the MINLP model described in Section III with the *A Modeling Language for Mathematical Programming* (AMPL) platform and solved the problem using the open-source solver COUENNE package for MINLPs [24]. However, we understand that MINLP problems are usually *NP-Hard* in nature and the COUENNE package uses some variations of *Branch-and-Bound algorithm* that may take several hours to days to compute the optimal solution, depending on volume of dataset and available computational capacity [25]. Hence, a time-scalable heuristic algorithm specific to CCOMPASSION needs to be developed, but during the

network planning stage, network operators can tolerate the required computational time.

##### A. Random Dataset Generation

To test the CCOMPASSION framework, we used stochastically-generated hypothetical 5 km x 5 km Australian city/towns to represent urban, suburban and rural deployment scenarios. Following the guidelines provided in [26], we consider the population densities of 4000, 2500 and 1500 per square kilometre for urban, suburban and rural areas, respectively. We used the *Poisson-point process* to distribute the total population across the considered area to identify the ONU locations and followed by used *k-means clustering algorithm* to identify potential field cloudlet placement locations [27]. The value of  $k$  needs to be sufficiently large enough to maintain the feasibility of the optimal solution and we arbitrarily choose it as 20. The ONUs are supported by the standard 10G(E)-PON infrastructure with 1: $N$ ,  $N \in \{4, 8, 16\}$  split ratio and 10 Gbps datarate in both downlink and uplink. Fig. 3 presents the considered datasets. The small blue dots represent the ONU locations, the red asterisks represent the potential field cloudlet locations, the cyan squares represent the RN locations and the green diamonds represent the CO

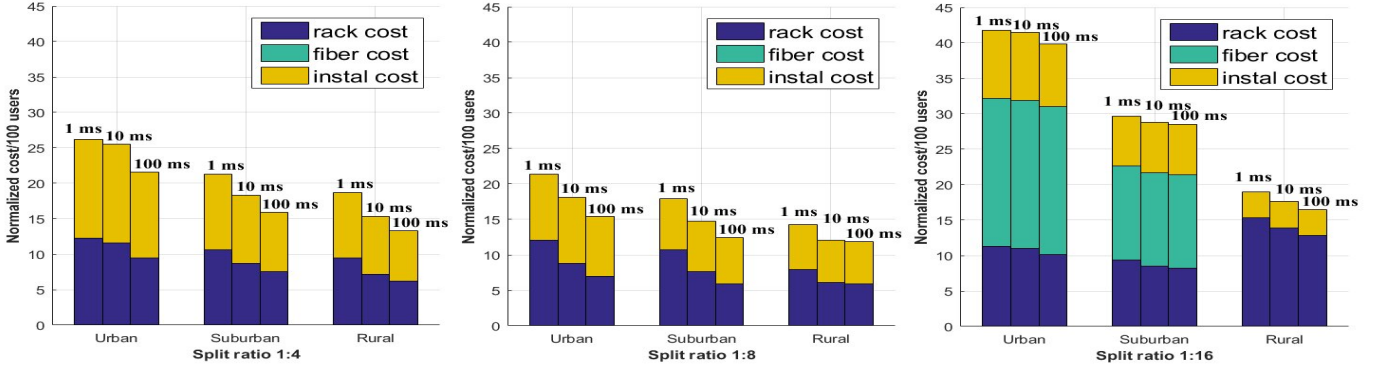


Fig. 5: Normalized cost/100 users in urban, suburban and rural scenarios against  $D_{QoS} = 1$  ms, 10 ms, and 100 ms.

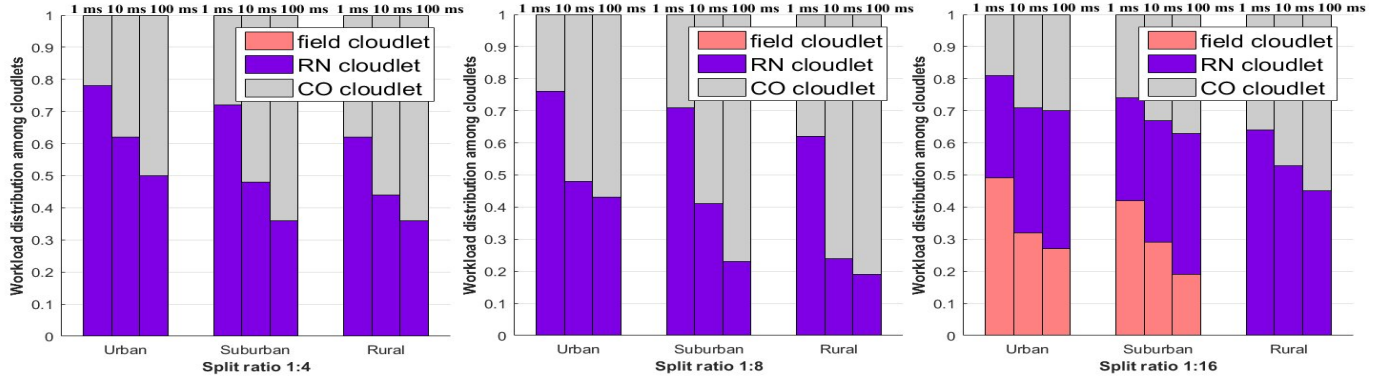


Fig. 6: Distribution of workload among cloudlets in the field, RN and CO locations in urban, suburban and rural scenarios against  $D_{QoS} = 1$  ms, 10 ms, and 100 ms.

locations (corner points of the 5 km x 5 km block). The solid blue lines indicate the connectivity amongst the RNs and COs. Note that feeder fiber and distribution fiber connectivity between RNs and ONUs is not shown.

### B. Values of Network Parameters and Cost components

The normalized costs of installing new point-to-point fiber per kilometer ( $\eta$ ) are 50, 35 and 20 for urban, suburban and rural, respectively [21]. The datarate of these new point-to-point fiber links are considered to be  $BW_{ad} = BW_{da} = 1$  Gbps and the maximum allowed distance among field cloudlets and ONUs is  $L_{max} = 4$  km, such that feasibility of optimal solution is maintained. We consider that  $n_\lambda = 1$  wavelength (with a maximum 10 Gbps datarate is shared among ONUs) to communicate with the cloudlet at their corresponding RN location and hence  $BW_{bd} = BW_{db} = 10$  Gbps. As ONUs are allowed to use the default uplink and downlink channels to communicate with cloudlets at their corresponding CO, therefore  $BW_{cd} = BW_{dc} = 10$  Gbps. We consider the approximate downlink and uplink background loads as  $\beta_{dl} = 7$  Gbps and  $\beta_{ul} = 5$  Gbps, respectively [21]. The normalized cost of installing one rack is considered as  $\alpha = 1$  [28]. We also consider the normalized costs of installing new cloudlet infrastructure at field and RN locations equal, i.e.,  $\xi_a = \xi_b = 4$  and less at CO locations, i.e.,  $\xi_c = 2$  [28]. We assumed that each ONU can serve a maximum

1000 users and hence the maximum task arrival rate from each ONU  $d \in D$  is  $\lambda_d = 1000$  VMs/sec. In [5], the authors have outlined a summary of average request and response sizes of some commonly used applications and thus we consider  $\sigma_{ul} = 1$  MB and  $\sigma_{dl} = 1$  KB. The service rate of each rack in cloud or cloudlet ( $\mu$ ) is 2500 VMs/sec [12]. The end-to-end latency to offload a task to cloud server ( $\Lambda$ ) is relatively large and considered to be 0.8 sec [14].

### C. Results and Discussions

In this section, we discuss different aspects of the optimal solutions obtained through evaluating CCOMPASSION framework on the urban, suburban and rural deployment scenarios against target latency 1 ms, 10 ms and 100 ms. Fig. 4 shows the optimal cloudlet placement locations in urban, suburban and rural scenarios with 1:4 split ratio against  $D_{QoS} = 1$  ms latency target. For this case, field cloudlets are not chosen, rather only some of the RN locations (darkened) and all CO locations are chosen. The green borders indicate approximate coverage area of each of the cloudlets installed. Fig. 5 shows the cost of cloudlet installation normalized to 100 users for the all three  $D_{QoS}$  values of 1 ms, 10 ms and 100 ms. The total cost consists of three components, e.g., cost of racks, cost of newly installed fiber and cost of new infrastructure setup. The bar graphs imply that the cost of installing cloudlets in the urban

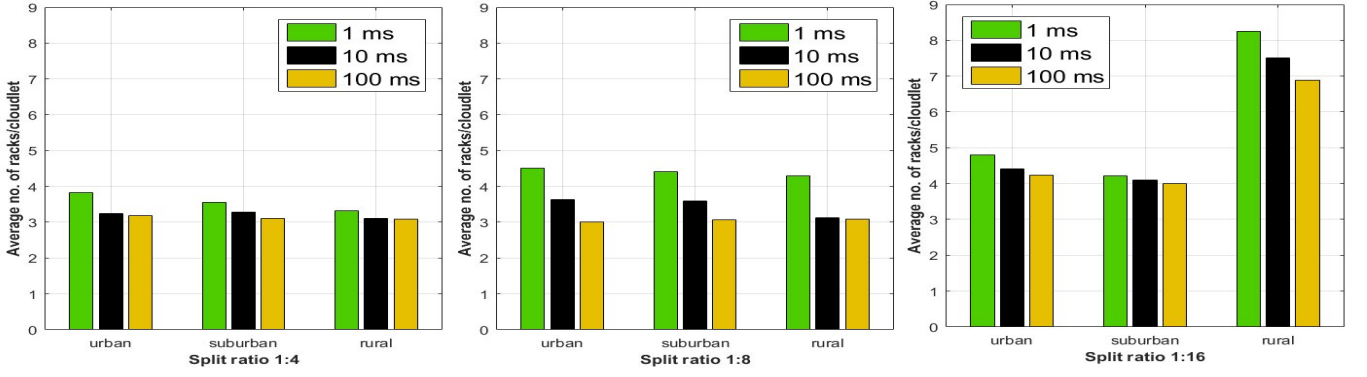


Fig. 7: Average number of racks/cloudlet in urban, suburban and rural scenarios against  $D_{QoS} = 1$  ms, 10 ms, and 100 ms.

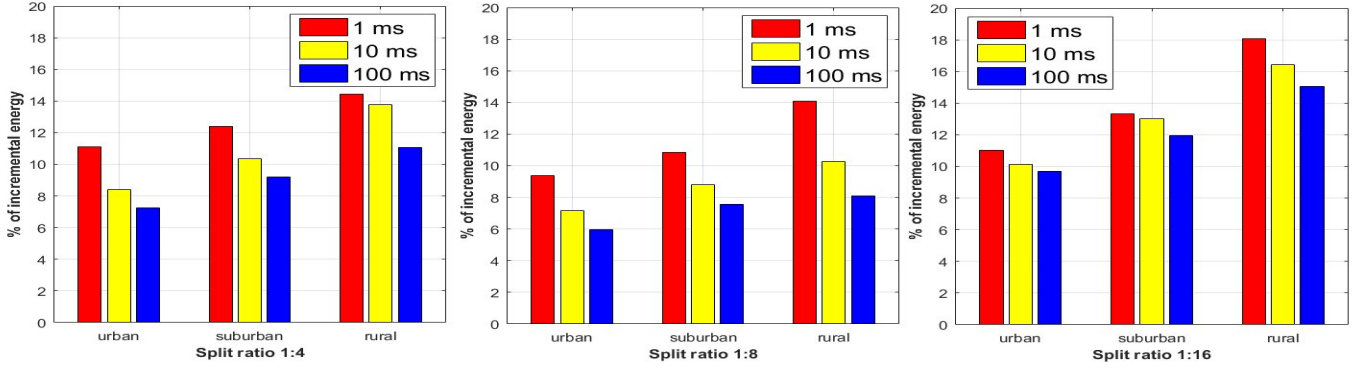


Fig. 8: Incremental energy consumption over existing TDM-PON in urban, suburban and rural scenarios against  $D_{QoS} = 1$  ms, 10 ms, and 100 ms.

scenario is the highest and the least in rural scenario with all split ratios. This is primarily due to the fact that more cloudlets are required in the urban and suburban scenarios compared to rural scenario to meet latency requirements. With a split ratio 1:16, a higher number of ONUs must share the same wavelength as compared to 1:4 and 1:8 split ratio scenarios and hence the per user average bandwidth is lower. As such, for urban and suburban scenarios with a 1:16 split ratio, cloudlets are required to be installed in the field, RN and CO locations, but for the rest of the cases only RN and CO cloudlets are required to be installed. Hence, the normalized cost of cloudlet deployment in urban and suburban scenarios with a 1:16 split ratio is relatively higher than other cases.

Fig. 6 shows the distribution of workload amongst the cloudlets installed in the field, RN or CO for urban, suburban and rural scenarios with  $1:N$ ,  $N \in \{4, 8, 16\}$  split ratios, for the all three  $D_{QoS}$  values of 1 ms, 10 ms and 100 ms. It is very interesting to observe that for all deployment scenarios, when  $D_{QoS} = 1$  ms, a majority of the computational tasks are performed by the field and RN cloudlets. However, as the QoS latency requirement is relaxed, i.e., with  $D_{QoS} = 10$  ms and 100 ms, a higher percentage of workload is handled by the CO cloudlets. Therefore, when the QoS latency requirement is stringent, expensive cloudlets are required to be installed in the field or RN depending on the deployment scenario. In contrast,

with a less stringent QoS latency requirement, installation of more economical CO cloudlets could be preferred.

The average number of racks required per cloudlet is shown in Fig. 7 for the urban, suburban and rural scenarios with  $1:N$ ,  $N \in \{4, 8, 16\}$  split ratios against  $D_{QoS}$  values of 1 ms, 10 ms and 100 ms. When  $D_{QoS} = 1$  ms, it can be interpreted from (1) that a higher number of racks are required, whereas when  $D_{QoS} = 10$  ms or 100 ms, a relatively lower number of racks are required to process the same number of VMs. This fact is also very much aligned with our observations from the results summarized in Fig. 7. We also note that the average number of racks per cloudlet in all scenarios with split ratio 1:8 is higher than with split ratio 1:4, because of the higher number of available RN locations with the 1:4 split ratio. Ideally, in a cost minimization framework a minimum number of cloudlets should be installed with an optimal number of racks. However, with a split ratio of 1:16, a few field cloudlets have to be installed in the urban and suburban scenarios. Thus in this case, the average number of racks per cloudlets in urban and suburban scenarios is lower as compared to the average number of racks in the rural scenario.

Installation of cloudlets over an existing 10G(E)-PON infrastructure increases the overall network energy budget. In Fig. 8, we present this percentage of increase in energy consumption for the urban, suburban and rural scenarios with



1:N,  $N \in \{4, 8, 16\}$  split ratios against  $D_{QoS}$  values of 1 ms, 10 ms and 100 ms. We considered the energy consumption of different elements of a 10G(E)-PON viz., OLT, ONU, wireless links etc. according to [21] and the energy consumption of cloudlets according to the technical specifications of Lenovo ThinkStation P900 [28]. Note that, albeit the urban and suburban scenarios have more cloudlets installed, the percentage of incremental energy budget is the highest for rural scenario against all  $D_{QoS}$  values of 1 ms, 10 ms and 100 ms. This is due to the small overall energy budget of rural deployment as compared to the urban and suburban deployment. Thus installing a single cloudlet in the rural scenario increases the energy budget more than urban and suburban scenarios. However, for all considered 5 km x 5 km urban, suburban and rural scenarios with 1 : N,  $N \in \{4, 8, 16\}$  split ratios, the percentage of incremental energy budget in the presence of cloudlets are less than 20%.

## V. SUMMARY

The idea of deploying distributed cloudlet facility solves the majority of the network bottleneck issues in the converged mobile computing and cloud computing environment and hence provisioning of many computation-intensive and interaction-intensive applications have become possible. Meeting the strict latency requirements of these applications has been a primary challenge due to latency in accessing the cloud services and optical access networks appeared as a promising solution in this aspect due to its low-cost per bit and huge bandwidth support. In this paper, we have proposed a hybrid cloudlet placement framework CCOMPASSION over existing TDM-PON infrastructures and have developed a MINLP model to identify cost optimal cloudlet placement locations. We tested this framework over hypothetical urban, suburban and rural deployment scenarios with 1:N,  $N \in \{4, 8, 16\}$  split ratios to achieve target latency values 1 ms, 10 ms and 100 ms. Through the evaluation of this framework, various seminal insights on the optimal deployment strategies of cloudlets in urban, suburban and rural deployment scenarios were provided.

## REFERENCES

- [1] M. Weiser, "The computer for the 21st century," *Scientific American*, vol. 265, no. 3, pp. 66–75, Sep. 1991.
- [2] R. V. D. Meulen, "Gartner says 8.4 billion connected "Things" will be in use in 2017, up 31 percent from 2016," *Gartner Newsroom*, Feb. 2017. [Online]. Available: <http://www.gartner.com/newsroom/id/3598917>
- [3] C. V. N. Index. (2017) Global mobile data traffic forecast update, 2016-2021. [Online]. Available: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf>
- [4] A. Helal, B. Haskell, J. L. Carter, R. Brice, D. Woelk, and M. Rusinkiewicz, *Introduction to Mobile Computing*. Boston, MA: Springer US, 2002, pp. 1–11.
- [5] K. Ha, P. Pillai, G. Lewis, S. Simanta, S. Clinch, N. Davies, and M. Satyanarayanan, "The Impact of Mobile Multimedia Applications on Data Center Consolidation," in *Proceedings of the 2013 IEEE International Conference on Cloud Engineering*, ser. IC2E '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 166–176.
- [6] K. Kumar and Y. H. Lu, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?" *Computer*, vol. 43, no. 4, pp. 51–56, April 2010.

- [7] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, Oct. 2009.
- [8] R. Balan, J. Flinn, M. Satyanarayanan, S. Sinnamohideen, and H.-I. Yang, "The Case for Cyber Foraging," in *Proceedings of the 10th Workshop on ACM SIGOPS European Workshop*, ser. EW 10. New York, NY, USA: ACM, 2002, pp. 87–92.
- [9] M. Satyanarayanan, "The Role of Cloudlets in Hostile Environments," in *Proceeding of the Fourth ACM Workshop on Mobile Cloud Computing and Services*, ser. MCS '13. New York, NY, USA: ACM, 2013, pp. 1–2.
- [10] S. Clinch, J. Harkes, A. Friday, N. Davies, and M. Satyanarayanan, "How close is close enough? Understanding the role of cloudlets in supporting display appropriation by mobile users," in *2012 IEEE International Conference on Pervasive Computing and Communications*, March 2012, pp. 122–127.
- [11] G. Lewis, S. Echeverria, S. Simanta, B. Bradshaw, and J. Root, "Tactical Cloudlets: Moving Cloud Computing to the Edge," in *2014 IEEE Military Communications Conference*, Oct 2014, pp. 1440–1446.
- [12] A. Ceselli, M. Premoli, and S. Secci, "Cloudlet network design optimization," in *2015 IFIP Networking Conference (IFIP Networking)*, May 2015, pp. 1–9.
- [13] Z. Xu, W. Liang, W. Xu, M. Jia, and S. Guo, "Efficient Algorithms for Capacitated Cloudlet Placements," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 10, pp. 2866–2880, Oct 2016.
- [14] M. Jia, J. Cao, and W. Liang, "Optimal Cloudlet Placement and User to Cloudlet Allocation in Wireless Metropolitan Area Networks," *IEEE Transactions on Cloud Computing*, vol. PP, no. 99, pp. 1–1, 2015.
- [15] X. Sun and N. Ansari, "Green Cloudlet Network: A Distributed Green Mobile Cloud Network," *IEEE Network*, vol. 31, no. 1, pp. 64–70, January 2017.
- [16] C. Develder, M. D. Leenheer, B. Dhoedt, M. Pickavet, D. Colle, F. D. Turck, and P. Demeester, "Optical Networks for Grid and Cloud Computing Applications," *Proceedings of the IEEE*, vol. 100, no. 5, pp. 1149–1167, May 2012.
- [17] M. Maier and B. P. Rimal, "Invited paper: The audacity of fiber-wireless (FiWi) networks: Revisited for clouds and cloudlets," *China Communications*, vol. 12, no. 8, pp. 33–45, August 2015.
- [18] B. P. Rimal, D. P. Van, and M. Maier, "Cloudlet Enhanced Fiber-Wireless Access Networks for Mobile-Edge Computing," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3601–3618, June 2017.
- [19] E. Wong, M. P. I. Dias, and L. Ruan, "Predictive Resource Allocation for Tactile Internet Capable Passive Optical LANs," *Journal of Lightwave Technology*, vol. 35, no. 13, pp. 2629–2641, July 2017.
- [20] E. Wong, "Next-Generation Broadband Access Networks and Technologies," *Journal of Lightwave Technology*, vol. 30, no. 4, pp. 597–608, Feb 2012.
- [21] E. Wong, C. M. Machuca, and L. Wosinska, "Survivable Hybrid Passive Optical Converged Network Architectures Based on Reflective Monitoring," *Journal of Lightwave Technology*, vol. 34, no. 18, pp. 4317–4328, Sept 2016.
- [22] L. Kleinrock, *Theory, Volume 1, Queueing Systems*. Wiley-Interscience, 1975.
- [23] U. Faigle, W. Kern, and G. Still, *Integer Programming*. Dordrecht: Springer Netherlands, 2002, pp. 173–196.
- [24] Convex Over and Under ENvelopes for Nonlinear Estimation (Couenne). [Online]. Available: <https://projects.coin-or.org/Couenne>
- [25] P. Bonami, M. Kılınç, and J. Linderoth, "Algorithms and software for convex mixed integer nonlinear programs," *Mixed integer nonlinear programming*, pp. 1–39, 2012.
- [26] A. Spencer, J. Gill, and L. Schmahmann, "Urban or Suburban? Examining The Density of Australian Cities in a Global Context," in *The State of Australian Cities Conference 2015*, Dec 2015, pp. 9–11.
- [27] S. Matthews and D. M. Parker, "Progress in Spatial Demography," *Demographic Research*, vol. S13, no. 10, pp. 271–312, 2013.
- [28] cnet. Lenovo ThinkStation P900 Workstation Specifications. [Online]. Available: <https://www.cnet.com/products/lenovo-thinkstation-p900-workstation/specs>