Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6065119, 12 pages
https://doi.org/10.1155/2021/6065119

WILEY | Hindawi

Research Article

Remote and Rural Connectivity: Infrastructure and Resource Sharing Principles

Thembelihle Dlamini[1] **and Sifiso Vilakati**[2]

[1]*Department of Electrical and Electronic Engineering, University of Eswatini, Kwaluseni, Eswatini*
[2]*Department of Statistics and Demography, University of Eswatini, Kwaluseni, Eswatini*

Correspondence should be addressed to Thembelihle Dlamini; tldlamini@uniswa.sz

Received 19 July 2021; Accepted 27 August 2021; Published 13 September 2021

Academic Editor: Issa Elfergani

As mobile networks (MNs) are advancing towards meeting mobile user requirements, the rural-urban divide still remains a major challenge. While areas within the urban space (metropolitan mobile space) are being developed, i.e., small Base Stations (BSs) empowered with computing capabilities are deployed to improve the delivery of user requirements, rural areas are left behind. Due to challenges of low population density, low income, difficult terrain, nonexistent infrastructure, and lack of power grid, remote areas have low digital penetration. This situation makes remote areas less attractive towards investments and to operate connectivity networks, thus failing to achieve universal access to the Internet. In addressing this issue, this paper proposes a new BS deployment and resource management method for remote and rural areas. Here, two MN operators share their resources towards the procurement and deployment of green energy-powered BSs equipped with computing capabilities. Then, the network infrastructure is shared between the mobile operators, with the main goal of enabling energy-efficient infrastructure sharing, i.e., BS and its colocated computing platform. Using this resource management strategy in rural communication sites guarantees a quality of service (QoS) comparable to that of urban communication sites. The performance evaluation conducted through simulations validates our analysis as the prediction variations observed show greater accuracy between the harvested energy and the traffic load. Also, the energy savings decrease as the number of mobile users (50 users in our case) connected to the remote site increases. Lastly, the proposed algorithm achieves 51% energy savings when compared with the 43% obtained by our benchmark algorithm. The proposed method demonstrates superior performance over the benchmark algorithm as it uses foresighted optimization where the harvested energy and the expected load are predicted over a given short-term horizon.

1. Introduction

The evolution of the mobile and wireless communication networks into the fifth generation (5G) will play a significant role in improving the global economy. With the internet of things (IoT) dictating the way in which people communicate through information sharing and knowledge dissemination, internet coverage needs to be improved. The capacity to provide radio coverage over a wide geographic area is a prerequisite towards meeting the ultralow latency requirements demanded by mobile subscribers [1, 2]. Through the installation of a BS and the development of the mobile and wireless communications, continuous communications can be achieved. This constitutes a gigantic step towards solving the rural/remote connectivity problem since electricity might be unreliable and it is very costly to extend grid connection to remote areas. Therefore, the provisioning of communication services in remote areas entails the use of renewable energy. Using renewable energy, coupled with sustainable energy storage solutions, is a promising solution towards resolving the remote area energy predicament.

Despite the use of green energy as a potential solution, many rural and remote areas in developed or undeveloped countries around the world are facing the challenge of unreliable high-quality Internet connectivity [3]. This is because MN operators are still skeptical towards making information

and communications technology (ICT) infrastructure investments in remote areas—hence the digital divide. One of the essential reasons is low expected revenue, calculated as ARPU, which reduces companies' willingness to invest in these areas. However, with the current trends in battery and solar module costs showing a decrease, MN operators might be motivated to make investments in remote and rural areas and deploy connectivity networks. Moreover, the advent of open, programmable, and virtualized 5G networks will enable MN operators to overcome the limitations presented by the current MN [2, 4] and make the ease of deploying open and programmable MN a possibility.

To extend network coverage to remote/rural areas, the use of terrestrial or nonterrestrial networks is proposed in [5]. In parallel, Sparse Terrestrial Networks (STN) using high towers and large antenna arrays are being developed to deliver very long transmission ranges. Here, the systems are equipped with the latest emerging antenna technologies and designs such as reconfigurable phased/inflatable/fractal antennas realized with metasurface material. Towards this, the works of [5] study the feasibility of providing connectivity to sparse areas utilizing massive-MIMO where the existing infrastructure of TV towers was used. In that work, it is observed that higher frequencies provide larger area coverage, provided that the antenna array area is the same. Another strategy for achieving good coverage as well as high capacity in remote/rural areas is to utilize two frequency bands, one low band and one high band, in an aggregated configuration. Following this strategy, the authors of [6] combine the New Radio (NR) 3.5 GHz and LTE 800 MHz on a GSM grid. In addition, along the lines of long-range systems, the NR is expected to support high data rates with low average network energy consumption through its lean design and massive MIMO utilization. Also, the authors of [7] extend rural coverage with STNs. Here, the large cells are created by using long-range links between BS and UE, where the long range is achieved by high towers combined with large antenna arrays and efficient antenna techniques creating narrow beams with high gain with a line-of-sight (LoS) or near-LoS connection to the UE.

In order to end this digital divide, MNs have to relook the way in which they are operating and make the necessary adjustments. One workable solution is making use of the softwarization technologies such as SDN, NFV, and MEC, to be enablers for *resource sharing* and *edgefication* [4, 8]. Furthermore, the emergence of network slicing further avails new market opportunities [9] for MN to explore. In network slicing, the BS site infrastructure (*resource blocks*, *bandwidth*, and *computing resources*) can be shared *fairly* by two or more mobile operators in real-time. This is to effectively maximize the use of existing network resources while simultaneously minimizing the operational costs in remote sites. Also, the open and accessible shared infrastructure can enable more MN operators and Internet service providers to expand their footprint into low-income areas, increasing the availability of connectivity in these areas and contributing to bridging the digital divide. For continuous operation in the rural/remote communication sites, the BS empowered with computing capabilities can be colocated with EH systems for harvesting energy from the environment, storing it in EB (storage devices), and then powering the site.

There are several forms of infrastructure sharing cases already in existence [10], such as the roaming-based sharing where the MN operators share the cell coverage for a prenegotiated time period. For example, using this roaming-based sharing, a UE can employ the roaming procedure in order to connect to a foreign network. In these classical forms of sharing generally one MN operator still retains ownership of the mobile network. Under shared infrastructure, new entrants no longer need to incur the often-significant upfront cost of building their own infrastructure and can save time and resources that would otherwise be dedicated to administrative authorization and licensing. However, potential risks to competition, governance, and implementation need to be managed to achieve the greatest benefit from infrastructure sharing. In this article, the BS infrastructure sharing and its colocated computing platform (MEC server) are done only for handling delay-sensitive workloads in remote/rural areas. Here, MN operators still have control of the delay-tolerant workloads to their remote clouds. This entails bringing the notion of *coownership* of the communication sites in remote/rural areas, within the MEC paradigm, in which *two* MN operators pull together their capital expenditure in order to share the deployed infrastructure, thus saving precious (already limited) economic resources for other types of expenses. Then, in order to effectively manage the BS sites deployed in remote/rural areas, procedures for dynamic network control (*managing network resources when MN operators share fairly their network infrastructure*) and agile management are required. This will assist in efficiently delivering a comparable QoS in remote/rural areas to that of urban areas.

The work done in this article is an extension of [8], where BS sleep modes and VM soft-scaling procedures were employed towards energy saving in remote sites. In [8], energy savings were obtained through short-term traffic load and harvested energy predictions, along with energy management procedures. However, the considered energy cost model does not take the caching process, tuning of transmission drivers, and the use of container-based virtualization into account. In addition, the considered communication site belongs to *one* MN operator, i.e., the site infrastructure was not shared between multiple operators. Therefore, the computing-plus-communication energy cost model is the main motivation for this article, where the BS site is shared among multiple operators in order to handle delay-sensitive workloads only. One application of our model (strategy) corresponds to the current situation that has been caused by the new coronavirus (COVID-19) pandemic. The pandemic has reshaped our living preferences such that rural (remote) areas are now becoming more and more attractive. This can motivate MN operators to deploy networks in such areas and then share their communication infrastructure and the computing resources that are colocated. The contributions of this article are summarized as follows:

(1) A BS empowered with computing capabilities colocated with an EH system is considered, whereby

the MN operators share the BS site infrastructure (i.e., *bandwidth* and *computing resources*) for handling delay-sensitive workloads within a remote/rural area

(2) In order to enable foresighted optimization, a short-term future communication site workload and harvested energy is forecasted using a LSTM neural network [11]

(3) An online controller-based algorithm *called* DRC-RS for handling infrastructure sharing and managing the communication site located in remote/rural areas is developed. The proposed algorithm is based on the LLC approach and resource allocation procedures with the objective of enabling for infrastructure sharing (BS and its colocated computing platform) and resource management within remote and rural communication sites

(4) Real-world harvested energy and traffic load traces are used to evaluate the performance of the proposed optimization strategy. The numerical results obtained through simulation show that the proposed optimization strategy is able to efficiently manage the remote/rural site and also allows the sharing of the network infrastructure

In order to achieve these, the remainder of this article is organized as follows: Section 2 discusses previous research works related to the one undertaken in this article. Section 3 describes the proposed system model using detailed explanation on the operation of each network element. The mathematical problem formulation is given in Section 4 together with the details of the optimization problem and the proposed DRC-RS online algorithm. In Section 5, a performance evaluation of the proposed online algorithm is presented using simulation results and statistical discussions. The conclusions of this article are then given in Section 6.

2. Related Work

MN operators generally have complete ownership and control of their network, and their networks are characterized by an inflexible and monolithic infrastructure. Such a rigid status quo incapacitates networks of the required versatility; hence, they cannot cope with the dynamically changing requirements. As a result, in their current state, meeting the heterogeneity and variability of future MNs is an impossible task. As mobile and wireless networks evolve, MN operators are faced with the daunting task of keeping up and coping with the accelerated roll-out of new technologies. Due to these fast-paced technological advancements, large and frequent investments are made in order to cope with the new services and network management phases. This proactive network operation and management consequently increases the network operating costs, which reduces the intended profits. Thus, in order to reduce the per-MN operator investment cost, the sharing of network infrastructure between mobile operators is an attractive solution. To this effect, the authors in [12] proposed a RAN sharing scheme where MN operators share a single radio infrastructure while maintaining separation and full control over the backhauling and their respective core networks. In that paper, a mixed-integer linear programming (MILP) formulation is proposed for determining the sharing configurations that maximize the QoS, and a cooperative game theory concept is used to determine stable configurations as envisioned by the MN operator. The regulatory enforcement towards offering the best service level for the users and the greedy approach considered in that paper reduce the effectiveness of infrastructure sharing, as both approaches do not promote fairness among MN operators. In addition, the work of [13] employs an infrastructure sharing algorithm towards energy savings by exploiting the underutilization of the network during low-traffic periods. In their work, a game-theoretic framework was proposed in order to enable the MN operators to individually estimate the switching-off probabilities that reduce their expected financial cost. Apart from the energy efficiency benefits, the proposed scheme allows the participating MN operators to minimize their spending costs independently of the strategies of the coexisting MN operators. Despite of the presented benefits, it is worth noting that infrastructure sharing should be considered for both low- and high-traffic periods, which is the focus of this paper. However, due to the existence of competition between the different MNs, collaboration in this infrastructure sharing is a primary requisite. In order to enforce such a collaboration between competitors, the authors in [14] proposed a strategic network infrastructure sharing framework for contractual backup reservation between a small/local network operator of limited resources and uncertain demands, and one resourceful operator with potentially redundant capacity. Here, one MN operator pays for network resources reserved for use by its subscribers in another MN operator, while in turn, the payee guarantees the availability of the resources. Then, in [15], the problem of infrastructure sharing among MN operators is presented as a multiple-seller single-buyer business. In their contribution, each BS is utilized by subscribers from other operators and the owner of the BS is considered as a seller of the BS infrastructure while the owners of the subscribers utilizing the BS are considered as buyers. In the presence of multiple seller MN operators, it is assumed that they compete with each other to sell their network infrastructure resources to potential buyers.

The aforementioned works consider BS infrastructure sharing towards lowering operational cost, either by switching on/off the BSs, while maintaining the network control. In addition, infrastructure sharing is treated as a business case instead of a cooperative effort towards boosting connectivity in remote/rural areas. If one MN operator is treated as a seller while the other one as a buyer if it uses its network resources, this becomes a business venture. For example, one MN operator might be using the resource reservation technique, whereby it reserves resources for other small operators. Again, here, the other party has to pay in order to use those facilities. However, it is worth mentioning that the works done in [12–15] do not consider infrastructure sharing with the MEC paradigm and the consideration of green energy has been overlooked. Those that are within

the MEC paradigm share their *own* network resources, among themselves in order to handle spatially uneven computation workloads in the network. Their objective being to avoid large computation latency at overloaded small BSs as well as to provide high quality of service (QoS) to end users. The details of how internal infrastructure sharing is conducted cannot be covered in this article; interested readers are referred to [16]. Table 1 summarizes the differences of the infrastructure sharing strategy from existing works.

3. System Model

In this paper, we consider a remote/rural site network scenario as illustrated in Figure 1. Each network apparatus (BS, MEC server) in the figure is mainly powered by renewable energy harvested from wind and solar radiation, and it is equipped with an EB for energy storage. The stored energy is shared by the edge server and the BS system. The EM is an entity responsible for selecting the appropriate energy source to fulfill the EB and also for monitoring the energy level of the EB. Then, the intelligent electromechanical switch (I-SW) aggregates the energy sources to fulfill the EB level. The proposed model in Figure 1 is cache-enabled, TCP/IP offload capable (i.e., enables *partial* offloading in the server's NIC such as checksum computation [17]). The virtualized MEC server, which is colocated with the BS, is assumed to be hosting C containers (see C1 and C2 in Figure 1). Also, it has an input and output buffer for holding the workloads. It is assumed that some of the BS functions are virtualized as pointed in [18] as the MEC node is composed of a virtualized access control router which acts as an access gateway for admission control. The virtualized access control router (ACR) is responsible for local and remote routing, and it is locally hosted as an application. Here, it is assumed that the remote/rural site infrastructure is shared between *two* MN operators through a preexisting agreement, where a common microwave backhaul or a multihop wireless backhaul relaying is used for accessing remote clouds or the Internet. Moreover, a discrete-time model is considered, whereby the time is discretized as $t = 1, 2, \cdots$ time slots of a fixed duration τ.

3.1. Input Traffic and Queue Model. In the communication site, the BS is the connection point anchor and the computing platform processes the currently assigned delay-sensitive tasks by self-managing its own local virtualized storage/computing resources. Also shown in Figure 1 is an input buffer of size L_{in}, a reconfigurable computing platform and the related switched virtual LAN; an output queue of size L_{out}; and a controller that reconfigures the computing-plus-communication resources and also performs the control of input/output traffic flows. Since the workload demand exhibits a diurnal behavior in remote/rural areas, forecasting the mobile operator's workload can help towards network infrastructure sharing. Thus, in order to emulate the remote site traffic load $L(t)$ (from $|\nu(t)|$ users), real MN traffic load traces from [19] are used. It is assumed that *only* operators A and B share the remote/rural BS site, and their traffic load profiles are denoted by $L_{\text{A}}(t)$ and $L_{\text{B}}(t)$ (bits), respectively. It is also assumed that $L_{\text{A}}(t)$ (or $L_{\text{B}}(t)$) consists of 0.8 delay-sensitive workloads $\gamma_{\text{A}}(t)$ (or $\gamma_{\text{B}}(t)$) and the remainder is delay-tolerant. The total admitted workload is denoted by $\gamma^*(t) = \gamma_{\text{A}}(t) + \gamma_{\text{B}}(t)$, i.e., $\gamma^*(t) \le L_{\text{in}}$). The input/output (I/O) queue of the system is assumed to be loss-free such that the time evolution of the backlog queues follows Lindley's equations. The normalized BS traffic load behavior representation of the two mobile operators is illustrated in Figure 2.

3.2. Communication and Computing Energy Cost Model. For the BS system deployed in the remote/rural area, the total energy consumption $\theta_{\text{SITE}}(t)$ (measured in J) at time slot t consists of the BS communications, denoted by $\theta_{\text{COMM}}(t)$, and computing platform processes, related to computing, caching, and communication, which is denoted by $\theta_{\text{COMP}}(t)$. Thus, the energy consumption model at time slot t is formulated as follows, inspired by [20]:

$$\theta_{\text{SITE}}(t) = \theta_{\text{COMM}}(t) + \theta_{\text{COMP}}(t). \tag{1}$$

The BS energy consumption processes $\theta_{\text{COMM}}(t)$ constitutes of the sum of the following:

$$\theta_{\text{COMM}}(t) = \sigma(t)\theta_0 + \theta_{\text{load}}(t) + \theta_{\text{bk}} + \theta_{\text{data}}(t)\gamma^*(t), \tag{2}$$

where $\sigma(t) \in \{0, 1\}$ is the BS switching status indicator, with 1 representing the active mode while 0 indicates the power saving mode. θ_0 is the load-independent constant value representing the operation energy, $\theta_{\text{load}}(t) = L(t)(2^{r_0/\zeta(t)W} - 1)N_0(K)^{\alpha}\beta^{-1}$ the load-dependent transmission power to the served subscribers that guarantees low latency services at a target rate r_0. The term W is the channel bandwidth, and $\zeta(t)$ is the fraction of the bandwidth used by the mobile users from operators A and B, while α and β are the path loss exponent and the path loss constant, respectively. The term K denotes the average distance between two BSs within the same region, and N_0 is the noise power. The parameter θ_{bk} represents the constant microwave backhaul transmission energy cost, and $\theta_{\text{data}}(t)$ (fixed value in J/byte) is the intercommunication cost incurred by exchanging data between the BS and MEC interfaces.

Next, we discuss the MEC server processes that make up $\theta_{\text{COMP}}(t)$. With $\gamma^*(t)$ being the currently admitted workload to be processed, let $\gamma_c(t) \le \gamma_{\max}$, $c = 1, \cdots, C(t)$, denote the size of the task that the scheduler allocates, per container, bounded by the set maximum amount $\gamma_{\max}$. This is such that the following constraint: $\sum_{c=1}^{C(t)} \gamma_c(t) = \gamma^*(t)$, guarantees that the overall workload is partitioned into $|C(t)|$ parallel tasks. This load distribution is motivated by the share feature [21] that is inherent in virtualization technologies. This enables the resource scheduler to efficiently distribute resources among contending containers, thus guaranteeing the completion of the computation process within the expected time. Thus, the set of attributes which characterize each container are $\{\psi_c(t), \theta_{\text{idle},c}(t), \theta_{\max,c}(t), \Delta, f_c(t)\}$, where $\psi_c(t) = (f_c(t)/f_{\max})^2$ is the container utilization function, and $f_{\max}$ is the maximum available processing rate for the container. Here, $f_c(t) \in [f_0, f_{\max}]$ denote the processing rates of container c, whereby the term f_0 is the zero speed of the container, e.g., deep sleep or shutdown. The term

Table 1: Comparison with existing works.

Feature	Edge computing	Method used	Forecasting	Objective
RAN sharing [12]	No	Linear programming	No	Max. QoS
Traffic load exploitation [13]	No	Game theory	No	Min. spending cost
Contractual backup [14]	No	Contract design under symmetric information	No	Max. resource utilization and profits
Multiple-seller single-buyer [15]	No	Stochastic geometry	No	Cost minimization Guarantee of QoS
Communication and computation (proposed)	Yes	LSTM LLC	Yes	Min. energy consumption Guarantee of QoS

Yes: considered; No: not considered.

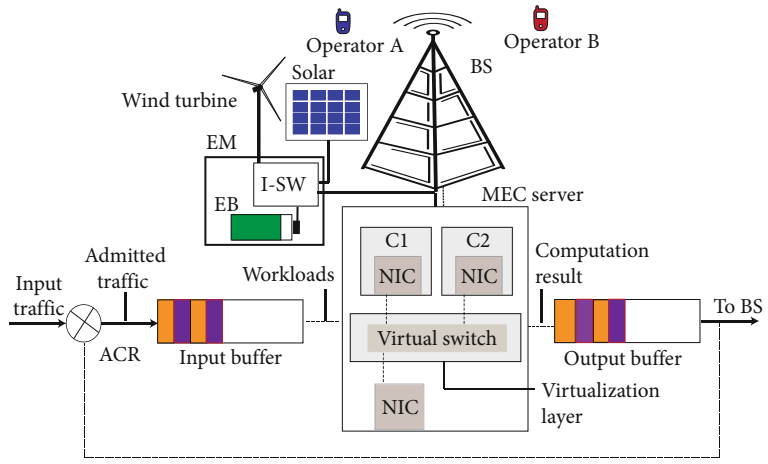


Figure 1: The remote/rural BS site infrastructure consisting of the BS colocated with the MEC server both powered by green energy obtained from solar radiation and wind turbine.

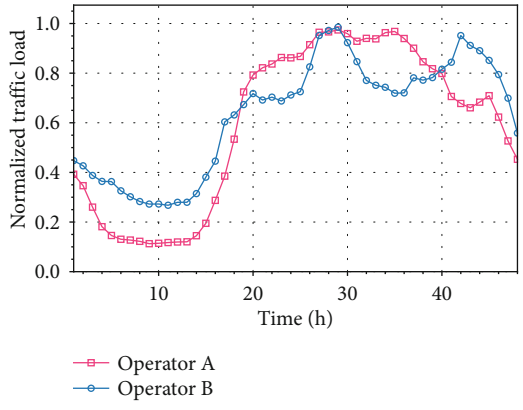


Figure 2: Normalized BS traffic load behavior representing two MN operators represented as operators A and B.

$\theta_{\text{idle},c}(t)$ represents the static energy drained by the container c in its idle state, $\theta_{\max,c}(t)$ is the maximum energy that container c can consume, and Δ is the maximum per-slot and per-container processing time (s).

Within the computing platform, the energy drained due to the active containers, denoted by $\theta_{\text{CP}}(t)$, is induced by the CPU share that is allocated for the workload, and it is given by

$$\theta_{\text{CP}}(t) = \sum_{c=1}^{C(t)} \theta_{\text{idle},c}(t) + \psi_c(t)(\theta_{\max,c}(t) - \theta_{\text{idle},c}(t)). \quad (3)$$

It should be noted that within the edge server, there is the virtualization layer with switching capabilities (see Figure 1). Thus, the processing rates are switched from the processing rates of the previous time instance $(t-1)$, denoted by $f_c(t-1)$, to the present instance (t), denoted by $f_c(t)$. This entails an energy cost, denoted by $\theta_{\text{SW}}(t)$, which is defined as

$$\theta_{\text{SW}}(t) = \sum_{c=1}^{C(t)} k_e(f_c(t) - f_c(t-1))^2, \quad (4)$$

where k_e represents the per-container reconfiguration cost caused by a unit-size frequency switching which is limited to a few hundreds of per $(\text{MHz})^2$.

The MEC server can perform TCP/IP computation processing in the network adapter in order to minimize the CPU utilization. Such process incurs an energy that is drained, denoted by $\theta_{\mathrm{OF}}(t)$, which is obtained as

$$\theta_{\mathrm{OF}}(t) = \delta(t)\theta^{\mathrm{nic}}_{\mathrm{idle}}(t) + \theta^{\mathrm{nic}}_{\mathrm{max}}(t), \tag{5}$$

where $\theta^{\mathrm{nic}}_{\mathrm{idle}}(t)$ (a nonzero value) is the energy drained by the adapter when powered but with no data transfer processes. This avails an opportunity to reduce the nonzero value to zero energy. For this, $\delta(t) = (0, 1)$ is the switching status indicator, with 1 indicating the active state and 0 representing the idle state. Then, $\theta^{\mathrm{nic}}_{\mathrm{max}}(t)$ is the maximum energy drained by the network adapter process and it is obtained in a similar way as in [20].

In order to keep the intracommunication delays at a minimum, it is assumed that each container c communicates with the resource scheduler through a dedicated reliable link that operates at the transmission rate of $r_c(t)$ (bits/s). Thus, the power drained by the c^{th} end-to-end connection is given by

$$P^{\mathrm{net}}_c(t) = \Psi_c(r\bar{t}t_c r_c(t))^2, \tag{6}$$

where $c = 1, \cdots, C(t)$, $r\bar{t}t_c$ is the average round-trip-time of the c^{th} intraconnection, and Ψ_c (measured in W) is the power consumption of the connection when the product, i.e., the round-trip-time, is by communication-rate-unit-valued. Therefore, after $\gamma_c(t)$ has been allocated to container c, the corresponding communication energy consumed by the c^{th} links, denoted by $\theta_{\mathrm{LK}}(t)$, is obtained as

$$\theta_{\mathrm{LK}}(t) = P^{\mathrm{net}}_c(t)\left(\frac{\gamma_c(t)}{r_c(t)}\right) \equiv \left(\frac{2\Psi_c}{(\tau - \Delta)}\right)(r\bar{t}t_c\gamma_c(t))^2. \tag{7}$$

In practical application scenarios, the maximum per-slot communication rate within the intracommunications is generally limited by a preassigned value r_{max}; thus, the following hard constraint must hold: $\sum_{c=1}^{C(t)} r_c(t) = \sum_{c=1}^{C(t)} (2\gamma_c(t)/(\tau - \Delta)) \le r_{\mathrm{max}}$. We also note that there exists a two-way per task execution delay where each link delay is denoted by $\varrho_c(t) = \gamma_c(t)/r_c(t)$. In this work, we assume that the overall delay equates to $2\varrho_c(t) + \Delta$.

To dequeue the computational results from the output buffer, denoted by L_{out}, the optical tunable drivers are used for the data transfer processes. A trade-off between the transmission speed and the number of active drivers per time instance is required to reduce the energy consumption. For data transfers, $|D(t)| \le D$ drivers are required for transferring $l_d(t) \in L_{\mathrm{out}}$. The energy drained by the data transfer process, denoted by $\theta_{\mathrm{LS}}(t)$, consists of the energy for utilizing each fast tunable driver, denoted by $m_d(t)[(J/s)]$ (a constant value), the target transmission rate r_0, and L_{out}. Thus, the energy is obtained as follows:

$$\theta_{\mathrm{LS}}(t) = \sum_{d=1}^{D(t)} \frac{(m_d(t)l_d(t))}{r_0}, \tag{8}$$

where the parameters are obtained similar to [20].

To minimize the network traffic from the remote/rural site to the remote clouds, some of the frequently requested internet contents are cached locally, more especially viral contents. The caching process contribute to the energy consumption within the site, denoted by $\theta_{\mathrm{CH}}(t)$, and it is obtained as [20]:

$$\theta_{\mathrm{CH}}(t) = \bar{\lambda}(t)(\theta_{\mathrm{TR}}(t) + \theta_{\mathrm{CACHE}}(t)), \tag{9}$$

where $\theta_{\mathrm{TR}}(t)$ represents the power consumption due to transmission processes, $\theta_{\mathrm{CACHE}}(t)$ is the power consumption contributed by the caching process with its intracommunication, and $\bar{\lambda}(t)$ is the response time function for viral content [22].

Overall, the resulting communication-plus-computing processes incurs an energy cost $\theta_{\mathrm{COMP}}(t)$, per slot t, which is given by Equations (3), (4), (5), (7), (8), and (9), as follows:

$$\theta_{\mathrm{COMP}}(t) = \theta_{\mathrm{CP}}(t) + \theta_{\mathrm{SW}}(t) + \theta_{\mathrm{OF}}(t) + \theta_{\mathrm{LK}}(t) + \theta_{\mathrm{LS}}(t) + \theta_{\mathrm{CH}}(t). \tag{10}$$

3.3. Energy Harvesting and Demand Profiles. The rechargeable energy storage device is characterized by its finite energy storage capacity E_{max}, and the energy level reports are periodically pushed to the DRC-RS application in the MEC server. In this case, the EB level $B(t)$ is known, which enables for the provisioning of the required communication and computing resources in the form of the required containers, the transmission drivers, and the transmission power in the BS. To emulate the profiles, the amount of harvested energy $H(t)$ in time slot t is obtained from open-source solar and wind traces from a farm located in Belgium [23], and they are as shown in Figure 3. The data in the dataset matches the time slot duration of 30 min used in this work, and it is the result of daily environmental records. In this work, the wind energy is selected as a power source during the solar energy off-peak periods. The available EB level $B(t+1)$ located at the offgrid site evolves according to the following dynamics:

$$E(t+1) = \min\{E(t) + H(t) - \theta_{\mathrm{SITE}}(t) - a(t), E_{\mathrm{max}}\}, \tag{11}$$

where $E(t)$ is the energy level in the battery at the beginning of time slot t, $\theta_{\mathrm{SITE}}(t)$ represents the site energy consumption, *see* Equation (1), and $a(t)$ is the leakage energy. However, it is worth noting that the energy level $E(t)$ is updated at the beginning of time slot t, whereas $H(t)$ and $\theta_{\mathrm{SITE}}(t)$ are only known at the end of t. Thus, the energy constraint at the off-grid site must be satisfied for every time slot: $\theta_{SITE}(t) \le E(t)$. Therefore, for decision making, the online controller simply compares the received EB level

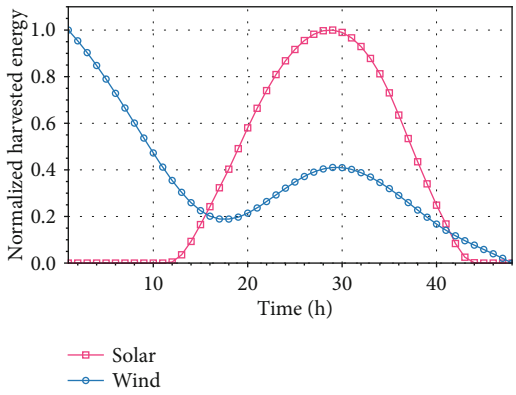


FIGURE 3: Example traces for harvested solar traces and wind traces from [23].

reports with two set points $(0 < E_{\text{low}} < E_{\text{up}} < E_{\max})$, the lower E_{low} and upper E_{up} energy thresholds. Here, E_{low} is the lowest EB level that the off-grid site should reach and E_{up} corresponds to the desired energy buffer level at the site. If $E(t) < E_{\text{low}}$, then the site is said to be energy deficient, and a suitable energy source at each time slot t is selected on the forecast expectations, i.e., the expected harvested energy $\widehat{H}(t)$.

4. Problem Formulation

In this section, the optimization problem is formulated to obtain an energy-efficient infrastructure sharing and resource management procedures through short-term traffic load and harvested energy forecasting. The overall goal is to enable energy-efficient infrastructure sharing and resource management, within remote and rural communication sites, and in turn guaranteeing a comparable QoS to that of urban areas, with reduced energy consumption in remote/rural sites.

4.1. Optimization Problem. Within the BS, the allocated bandwidth W is shared between mobile subscribers from operators A and B, and within the computing platform, the containers (i.e., as the computing resources) and the underlying physical resources (e.g., CPU) are shared among the users who offloaded their delay-sensitive workloads. To address the aforementioned problem, two cost functions are defined, namely, F1 and F2, where (F1) is defined as $\theta_{\text{SITE}}(t)$ (F1), weighs the energy drained in the BS site due to transmission and computing processes; and (F2) which accounts for the comparable QoS is defined as $(\gamma^*(t) - L_{\text{in}})^2$. Regarding this formulation, it is worth noting that F1 tends to push the system towards self-sustainability solutions and F2 favors solutions where the delay-sensitive load is entirely admitted in the computing platform by the router application, taking into account the expected energy to be harvested. The corresponding (weighted) cost function is defined as

$$J(\zeta, \sigma, C, D, t) \triangleq^{\Delta} Y\theta_{\text{SITE}}(\zeta(t), \sigma(t), C(t), D(t), t) + \bar{Y}(\gamma^*(t) - L_{\text{in}}(t))^2, \tag{12}$$

where $Y = [0, 1]$ is the weight used to balance the two functions, and $\bar{Y} = ^{\Delta} 1 - Y$. Hence, starting from the current time slot $t = 1$ to the finite horizon T, the time is discretized as follows: $t = 1, 2, \cdots, T$); thus, the optimization problem is formulated as follows:

$$\mathbf{P1}: \min_{\mathcal{N}} \sum_{t=1}^{T} J(\zeta, \sigma, C, D, t) \quad \text{subject to :} \tag{13}$$

$$\begin{aligned}
&\text{A1}: \sigma(t) \in \{0, 1\},\\
&\text{A2}: \beta \le C(t) \le C,\\
&\text{A3}: E(t) \ge E_{\text{low}},\\
&\text{A4}: 0 \le \gamma_c(t) \le \gamma_{\max},\\
&\text{A5}: 0 \le f_c(t) \le f_{\max},\\
&\text{A6}: r_{\min} \le r_c(t) \le r_{\max},\\
&\text{A7}: \theta_{\text{SITE}}(t) \le E(t),\\
&\text{A8}: \max\{2\rho_c(t)\} + \Delta = \tau_{\max}, \quad t = 1, \cdots, T,
\end{aligned} \tag{14}$$

where the set of objective variables to be configured at slot t in the BS system and MEC server is defined as $\mathcal{N} = ^{\Delta}\{\zeta(t), \sigma(t), C(t), \{\psi_c(t)\}, \{P_c^{\text{net}}(t)\}, \{\gamma_c(t)\}, \delta(t), D(t)\}$. These settings handle the transmission and computing activities using the following constraints. Here, constraint A1 specifies the BS operation status (either *power saving* or *active*), and A2 forces the required number of containers, $C(t)$, to be always greater than or equal to a minimum number $\beta \ge 1$. The purpose of this is to be always able to handle mission-critical communications. The constraint A3 ensures that the EB level is always above or equal to a preset threshold E_{low}, to guarantee *energy self-sustainability* over time. Furthermore, A4 bound the maximum workloads of each running container c, with $c = 1, \cdots, C(t)$, and A5 represents a hard limit on the corresponding per-slot and per-VM processing time. A6 forces $r_c(t)$ to fall in a desired range: $[r_{\min}, r_{\max}]$ of transmission rates, and A7 ensures that the energy consumption at the site is bounded by the available energy in the EB. A8 offers the hard QoS guarantees within the computing platform. From **P1**, it is noted that there exists a nonconvex component $P_c^{\text{net}}(t)$, from $\theta_{\text{LK}}(t)$. In this case, the geometric programming (GP) concept can be used to convert $\theta_{\text{LK}}(t)$ into a convex function similar to [20]. Thus, in order to solve **P1** in (13), the LLC approach [24], GP technique, and heuristics are used towards obtaining the feasible system control inputs $\eta(t) = (\zeta(t), \sigma(t), C(t), \{\psi_c(t)\}, \{P_c^{\text{net}}(t)\}, \{\gamma_c(t)\}, \delta(t), D(t))$ for $t = 1, \cdots, T$. Well, it should be noted that (13) can iteratively be solved at any time slot $t \ge 1$, by just redefining the time horizon as $t' = t, t+1, \cdots, t+T-1$.

4.1.1. Feasibility and QoS Guarantees. Regarding the feasibility of the problem, the following formal results hold.

Input: $s(t)$ (current state)
Output: $\eta^*(t)$ (control input vector)
01: Parameter initialization
$\mathcal{G}(t) = \{s(t)\}$
02: **for** (k within the prediction horizon of depth T) **do**
- $\widehat{L}(t+k)$:= forecast the workload
- $\widehat{H}(t+k)$:= forecast the energy
- $\mathcal{G}(t+k) = \varnothing$
03: **for** (each $s(t)$ in $\mathcal{G}(t+k)$) **do**
- generate all reachable states $\widehat{s}(t+k)$
- $\mathcal{G}(t+k) = \mathcal{G}(t+k) \cup \{\widehat{s}(t+k)\}$
04: **for** (each $\widehat{s}(t+k)$ in $\mathcal{G}(t+k)$) **do**
- calculate the corresponding $\theta_{SITE}(\widehat{s}(t+k))$ taking into account of $\zeta(t)$, and $l_d(t)$ from $L_{out}(t)$
end for
end for
end for
05: - obtain a sequence of reachable states yielding the best system input
06: $\eta^*(t) \coloneqq$ control leading from $s(t)$ to $\widehat{s}_{\min}$
07: **Return** $\boldsymbol{\eta}^*(\boldsymbol{t})$

Algorithm 1: DRC-RS algorithm pseudocode.

Proposition 1. *Feasibility conditions.*
The following two inequalities:

$$\left(\frac{r_{\max}}{2}\right)(\tau - \Delta) \geq L_{in},$$
$$\sum_{c=1}^{C(t)} f_c(t)\Delta \geq r_{\min}, \tag{15}$$

guarantee that the infrastructure sharing and resource reconfiguration problem is feasible.

Since the reported conditions assure that P1 admits the solution, we then consider the corresponding QoS properties. In this regard, it is safe to point out that A6 and A8 lead to the following hard bounds on the resulting communication-plus-computing delay.

Proposition 2. *Hard QoS guarantees.*
Firstly, the feasibility conditions of Proposition 1 must be met. Next, we let random variables measure the following: the random queue delay of the input queue τ_{IQ}, the service time of the input queue τ_{SI}, the queue delay of the output queue τ_{OQ}, and the service time of the output queue τ_{SO}. Thus, the following QoS guarantees hold: the random total delay ($\tau_{tot} =^{\Delta} \tau_{IQ} + \tau_{SI} + \tau_{OQ} + \tau_{SO}$) induced by the computing platform is limited (in a hard way) up to

$$\tau_{tot} \leq \left(\frac{(L_{in} + L_{out})}{r_{\min}}\right) + 2. \tag{16}$$

Thus, the reported QoS guarantee leads to the conclusion that the remote/rural site can handle delay-sensitive workloads while meeting the bound in A8.

4.2. Infrastructure Sharing and Resource Allocation. In this subsection, the predictions for the BS traffic load and energy consumption, the description of the remote/rural site system dynamics, and the proposed online controller-based algorithm are presented.

4.2.1. Prediction of Exogenous Processes. Two exogenous processes are considered in this work: the harvested energy $H(t)$ and the BS traffic loads $L(t)$. In order to generate the predictions $(\widehat{H}(t), \widehat{L}(t))$, the LSTM neural networks [11] were adopted. Thus, the LSTM-based predictor has been trained to give an output of the forecasts for the required number of future time slots T. The trained LSTM network consists of an input layer, a single hidden layer consisting of 40 neurons, for 80 epochs, for a batch size of 4; and an output layer. For training and testing purposes, the dataset was split as 70% for training and 30% for testing. As for the performance measure of the model, the RMSE is used.

4.2.2. Remote/Rural Site System Dynamics. In order to effectively manage the remote/rural site, an adaptive implementation of the controller is developed. Its purpose is to compute the solutions of both the infrastructure sharing and resource configurations on-the-fly. For this purpose, an online controller-based algorithm is proposed and is outlined in Algorithm 1.

At this point, it should be noted that at time slot t, the system state vector is $s(t) = (\zeta(t), \sigma(t), C(t), D(t), E(t))$ and the applied input vector that drives the system towards the desired behavior. These drivers perform bandwidth sharing, adaptive BS power transmission, autoscaling and reconfiguration of

containers, and tuning of the optical drivers and are denoted by $\eta^*(t) = \{\zeta(t), \sigma(t), C(t), \{\psi_c(t)\}, \{P_c^{\text{net}}(t)\}, \{\gamma_c(t)\}, \delta(t), D(t)\}$. The system behavior is described by the discrete-time state-space equation, adopting the LLC principles [24]:

$$s(t+1) = \Phi(s(t), \eta(t)), \tag{17}$$

where $\Phi(\cdot)$ is a behavioral model that captures the relationship between $(s(t), \eta(t))$ and the next state $s(t+1)$. This relationship accounts for the amount of energy drained $\theta_{\text{SITE}}(t)$, that harvested $H(t)$, which together lead to the next buffer level $E(t+1)$ through Equation (11). The DRC-RS algorithm finds the best control action vector $\eta^*(t)$ that yields the desired system behavior within the remote/rural site. Note that $P_c^{\text{net}}(t)$ is obtained using the CVXOPT toolbox and $\gamma_c(t), C(t)$ is obtained following the procedure outlined in remark 1 in [20]. The entire process is repeated every time slot t when the controller can adjust the behavior given the new state information. The state values of $s(t)$ and $\eta(t)$ are measured and applied at the beginning of the time slot t, whereas the offered load $L(t)$ and the harvested energy $H(t)$ are accumulated during the time slot and their value becomes known only at the end of it. This means that, being at the beginning of time slot t, the system state at the next time slot $t+1$ can only be estimated, which is formally written as

$$\hat{s}(t+1) = \Phi(s(t), \eta(t)). \tag{18}$$

At this regard, it is worth noting that the control actions are taken after exploring only a limited prediction horizon, yielding a limited number of possible operating states. In order to ensure system stability, we rely on the notion that a system is said to be stable under control, if for any state, it is always possible to find a control input that forces it closer to the desired state or within a specified neighborhood of it [25].

4.2.3. Dynamic Resource Controller for Remote/Rural Sites. The edge network management algorithm pseudocode is outlined in Algorithm 1, and it is based on the LLC principles, where the controller obtains the best control action $\eta^*(t)$. Starting from the *initial state*, the controller constructs, in a breadth-first fashion, a tree comprising all possible future states up to the prediction depth T. The algorithm proceeds as follows:

A search set $\mathcal{G}$ consisting of the current system state is initialized (line 01), and it is accumulated as the algorithm traverse through the tree (line 03), accounting for predictions, accumulated workloads at the output buffer, past outputs, and controls, operating intervals. The set of states reached at every prediction depth $t+k$ is referred to as $\mathcal{G}(t+k)$ (line 02). Given $s(t)$, the traffic load $\hat{L}(t+k)$ and harvested energy $\hat{H}(t+k)$ are estimated first (line 02), and generate the next set of reachable control actions by applying the accepted workload $\gamma^*(t+k)$, energy harvested and shared bandwidth fraction $\zeta(t+k)$. The cost function corresponding to each generated state $\hat{s}(t+k)$ is then computed (line 04), where $\hat{s}(t+k)$ take into account of l_d as observed from $L_{\text{out}}(t)$. Once the prediction horizon is explored, a sequence of reachable states yielding minimum energy consumption is obtained (line 05). The control action $\eta^*(t)$ corresponding to $\hat{s}(t+k)$ (the first state in this sequence) is provided as input to the system while the rest are discarded (line 06). The process is repeated at the beginning of each time slot t.

Table 2: System parameters.

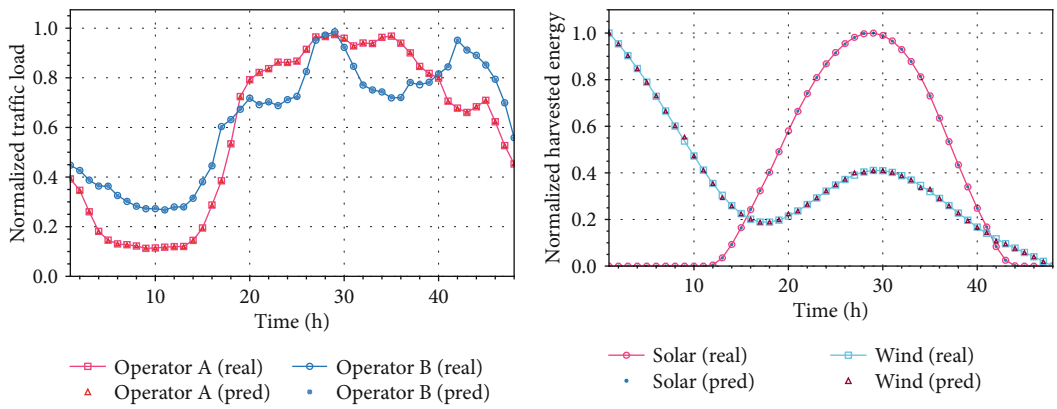
Parameter	Value
Microwave backhaul power, θ_{bk}	50 W
BS operating power, θ_0	10.6 W
Max. number of containers, C	20
Min. number of containers, β	1
Time slot duration, τ	30 min
Container c (idle state), $\theta_{\text{idle}_c}(t)$	4 J
Container c (max), $\theta_{\text{max},c}(t)$	10 J
Reconfiguration cost, k_e	0.005 J/(MHz)^2
NIC in idle state, $\theta_{\text{idle}}^{\text{nic}}(t)$	13.1 J
Max. allowed processing time, Δ	0.8 s
Processing rate set, $\{f_c(t)\}$	{0,50,70,90,105}
Bandwidth, W	1 MHz
Max. allocated c workload γ_{max}	10 MB
Max. number of drivers, D	6
Noise spectral density, N_0	-174 dBm/Hz
Max. container c load, γ_{max}	10 MB
Driver energy, $m_d(t)$	1 J/s
Target transmission rate, r_0	1 Mbps
Leakage energy, $a(t)$	2 μJ
Energy storage capacity, E_{max}	490 kJ
Lower energy threshold, E_{low}	30% of E_{max}
Upper energy threshold, E_{up}	70% of E_{max}

5. Performance Evaluation

In this section, some selected numerical results for the scenario of Section 3 are shown. The parameters that were used in the simulations are listed in Table 2.

5.1. Simulation Setup. A BS empowered with computation capabilities deployed in a rural/remote area is considered in this setup. Our time slot duration τ is set to 30 min, and the time horizon is set to $T = 3$ time slots. For simulation, Python is used as the programming language.

5.2. Numerical Results. Data preparation: the information from the used mobile and energy traces is aggregated to the set time slot duration. The mobile traces are aggregated from 10 observation time to τ. As for the wind and solar traces, they were aggregated from 15 min observation time



(a) One-step ahead predictive mean value for $L(t)$ (b) One-step ahead predictive mean value for $H(t)$

FIGURE 4: One-step online forecasting for both $L(t)$ and $H(t)$ patterns.

to τ. The used datasets are readily available in a public repository (*see* [26]).

In Figure 4, the real and predicted values for traffic load from operators A and B, harvested energy is shown. Here, the forecasting routing tracks each value and predicts it over one step. The shown selected prediction results are for operators A and B, solar, and wind. Then, Table 3 shows the average RMSE of the normalized harvested energy and traffic load processes (L_A, L_B), for different time horizon values, $T \in \{1, 2, 3\}$. In the table, the term $H_{wind}(t)$ represents the forecasted values for energy harvested from wind turbines and $H_{solar}(t)$ is for the harvested energy from solar panels. From the obtained results, the prediction variations are observed between $H(t)$ and $L(t)$ when comparing the average RMSE. The measured accuracy is deemed good enough for the proposed optimization.

The DRC-RS algorithm is benchmarked with another one, named Resource Reservation Manager (RRM), which is inspired by the backup reservation agreement from [14]. In the RRM, the network resources are reserved per time slot based on a set-point threshold percentage. Both algorithms make use of the learned information.

Figure 5 shows the average energy savings obtained within the off-grid system. Here, the number of users connected to the remote site is increased from $|\nu(t)| = 5$ to 50 , using an incremental step size of 5. The obtained energy savings are with respect to the case where the BS site is dimensioned for maximum expected capacity (maximum value of $\theta_{COMM}(t), \theta_{COMP}(t)$). From the results, as expected, it is observed that the energy savings decrease as the number of mobile users connected to the remote site increases. The DRC-RS outperforms the RRM algorithm. At this regard, we note that the communication site will accept users as long as energy harvesting projections are positive.

Then, Figure 6 shows the average energy savings for the edge system. Here, the BS group size is set to $|\nu(t)| = 20$ and the obtained energy saving results are with respect to the case where no energy management procedures are applied, i.e., the BS is dimensioned for maximum expected capacity (maximum value of $\theta_{SITE}(t)$, $\forall t$) and the MEC server provisions the computing resources for maximum expected computation workload (maximum value of $\theta_{MEC}(t)$, with $C = 20$ containers, $\forall t$). The average results of DRC-RS ($k_e = 0.05$, $\gamma_{max} = 10$ MB) show energy savings of 51%, while RRM achieves 43% on average. The effectiveness of the BS management procedure, autoscaling and reconfiguration of the computing resources, and on/off switching of the fast tunable laser drivers, coupled with foresighted optimization, is observed in the obtained numerical results.

TABLE 3: Average prediction error (RMSE) for harvested energy and traffic load processes, both normalized in [0,1].

	$T = 1$	$T = 2$	$T = 3$
$L_A(t)$	0.070	0.090	0.011
$L_B(t)$	0.050	0.070	0.010
$H_{wind}(t)$	0.011	0.013	0.016
$H_{solar}(t)$	0.050	0.070	0.090

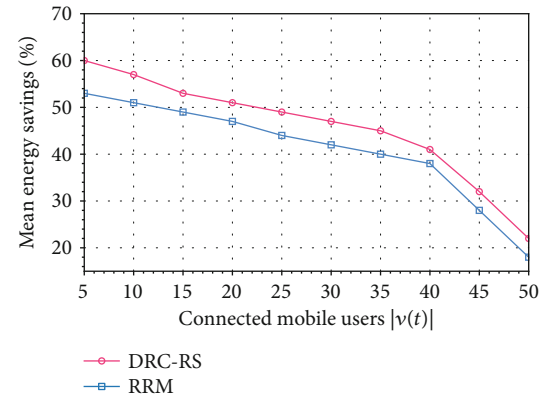


FIGURE 5: Energy savings versus number of users connected to the BS.

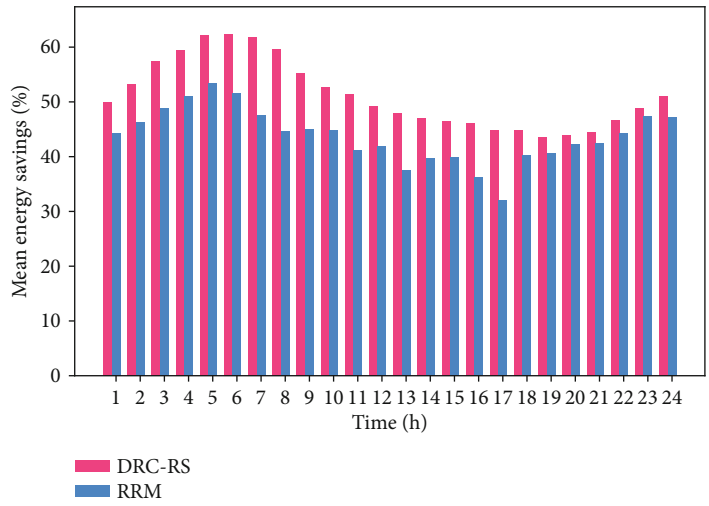


Figure 6: Mean energy savings for the remote/rural site system.

6. Conclusions

The challenge of providing connectivity to remote/rural areas will be one of the pillars for future mobile networks. To address this issue, in this paper, we present an infrastructure sharing and resource management mechanism for handling delay-sensitive workloads within a remote/rural site. Numerical results, obtained with real-world energy and traffic load traces, demonstrate that the proposed algorithm achieves mean energy savings of 51% when compared with the 43% obtained by our benchmark algorithm. Also, the energy that can be saved decreases as the number of users connected to the BS increases, with a guarantee of serving more users as long the green energy is available. The energy-saving results are obtained with respect to the case where no energy management techniques are applied in the remote site.

Data Availability

In this paper, open-source datasets for the mobile network (MN) traffic load, solar, and wind energy have been used. The details are as follows: (1) the real MN traffic load traces used to support the findings of this study were obtained from the Big Data Challenge organized by Telecom Italia Mobile (TIM) and the data repository has been cited in this article. (2) The real solar and wind traces used to support the findings of this study have also been cited in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] "Ericsson mobility report," Tech. Rep., Ericsson, Stockholm, Sweden, 2020, https://uk5g.org/media/uploads/resource_files/november-2020-ericsson-mobility-report.pdf.

[2] D. Thembelihle, M. Rossi, and D. Munaretto, "Softwarization of mobile network functions towards agile and energy efficient 5G architectures: a survey," Wireless Communications and Mobile Computing, 2017.

[3] H. Saarnisaari, S. Dixit, M.-S. Alouini et al., "6G white paper on connectivity for remote areas," Tech. Rep., University of Oulu, Oulu, Finland, 2020.

[4] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: state-of-the-art and the road ahead," *Computer Networks*, vol. 182, 2020.

[5] T. Taheri, R. Nilsson, and J. van de Beek, "The potential of massive-MIMO on TV towers for cellular coverage extension," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 8164367, 14 pages, 2021.

[6] J. Lun, P. Frenger, A. Furuskar, and E. Trojer, "5G new radio for rural broadband: how to achieve long-range coverage on the 3.5 GHz band," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pp. 1–6, Honolulu, USA, 2019.

[7] L. Feltrin, N. Jaldén, E. Trojer, and G. Wikström, "Potential for deep rural broadband coverage with terrestrial and non-terrestrial radio networks," *Frontiers in Communications and Networks*, vol. 2, 2021.

[8] T. Dlamini, Á. F. Gambín, D. Munaretto, and M. Rossi, "Online resource management in energy harvesting BS sites through prediction and soft-scaling of computing resources," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1820–1826, 2018.

[9] "Software-defined and cloud-native foundations for 5G networks," Tech. Rep., InterDigital, Denver, USA, 2019, https://www.interdigital.com/download/5cf010bb566d07680800072d.

[10] "Mobile infrastructure sharing: trends in Latin America," https://www.itu.int/en/ITU-D/Regulatory-Market/Documents/CostaRica/Presentations/Session8Daniel%20Leza%20-%20Mobile%20Infrastructure%20Sharing%20-%2012%20March%202014.pdf.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Rogers Street in Cambridge, Massachusetts MA 02142, 2016.

[12] L. Cano, A. Capone, G. Carello, M. Cesana, and M. Passacantando, "On optimal infrastructure sharing strategies in mobile radio networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3003–3016, 2017.

[13] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Game-theoretic infrastructure sharing in multioperator cellular networks," *IEEE transactions on vehicular communications*, vol. 65, no. 5, pp. 3326–3341, 2016.

[14] J. Hou, L. Sun, T. Shu, Y. Xiao, and M. Krunz, "Strategic network infrastructure sharing through backup reservation in a competitive environment," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1–9, Boston, MA, USA, 2019.

[15] S. Tachporn, G. Sudarshan, H. Ekram, R. Nandana, and L.-a. Matti, "Infrastructure sharing for mobile network operators: analysis of trade-offs and market," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2804–2817, 2018.

[16] L. Chen, S. Zhou, and J. Xu, "Computation peer offloading for energy-constrained mobile edge computing in small-cell networks," *IEEE/ACMTransactions on Networking*, vol. 26, no. 4, pp. 1619–1632, 2018.

[17] S. Ripduman, R. Andrew, A. W. Moore, and M. Kieran, "Characterizing 10 Gbps network interface energy consumption," in *IEEE Local Computer Network Conference*, pp. 268–271, Colorado, USA, 2010.

[18] "Virtualization for small cells: overview," Tech. Rep., Small Cell Forum, Draycott, England, 2015, https://scf.io/en/documents/106__Virtualization_for_small_cells_Overview.php.

[19] "Open big data challenge," https://dandelion.eu/datamine/open-big-data/.

[20] T. Dlamini and S. Vilakati, "LSTM-based traffic load balancing and resource allocation for an edge system," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8825396, 15 pages, 2020.

[21] M. Cardosa, M. R. Korupolu, and A. Singh, "Shares and utilities based power consolidation in virtualized server environments," in *2009 IFIP/IEEE International Symposium on Integrated Network Management*, pp. 327–334, New York, NT, USA, 2009.

[22] F. B. Abdesslem and A. Lindgren, "Large scale characterisation of YouTube requests in a cellular network," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, pp. 1–9, Sydney, Australia, 2014.

[23] "Total generation," https://www.elia.be/en/grid-data/power-generation.

[24] J. P. Hayes, "Self-optimization in computer systems via on-line control: application to power management," in *International Conference on Autonomic Computing, 2004. Proceedings.*, pp. 54–61, Washington, DC, USA, 2004.

[25] S. Abdelwahed, N. Kandasamy, and S. Neema, "Online control for selfmanagement in computing systems," in *Proceedings. RTAS 2004. 10th IEEE Real-Time and Embedded Technology and Applications Symposium, 2004*, pp. 368–375, Ontario, Canada, 2004.

[26] "Mobile and energy datasets," https://github.com/lihles/mobile-datasets.