

Wireless Communications and Mobile Computing

Energy Efficient Wireless Networks

Lead Guest Editor: Yujin Lim

Guest Editors: Gianluigi Ferrari, Hideyuki Takahashi, and Rossana M. C. Andrade






Energy Efficient Wireless Networks

Wireless Communications and Mobile Computing

Energy Efficient Wireless Networks

Lead Guest Editor: Yujin Lim

Guest Editors: Gianluigi Ferrari, Hideyuki Takahashi,
and Rossana M. C. Andrade




Copyright © 2019 Hindawi. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

- Javier Aguiar, Spain
Ghufran Ahmed, Pakistan
Wessam Ajib, Canada
Muhammad Alam, China
Eva Antonino-Daviu, Spain
Shlomi Arnon, Israel
Leyre Azpilicueta, Mexico
Paolo Barsocchi, Italy
Alessandro Bazzi, Italy
Zdenek Becvar, Czech Republic
Francesco Benedetto, Italy
Olivier Berder, France
Ana M. Bernardos, Spain
Mauro Biagi, Italy
Dario Bruneo, Italy
Jun Cai, Canada
Zhipeng Cai, USA
Claudia Campolo, Italy
Gerardo Canfora, Italy
Rolando Carrasco, UK
Vicente Casares-Giner, Spain
Luis Castedo, Spain
Ioannis Chatzigiannakis, Italy
Lin Chen, France
Yu Chen, USA
Hui Cheng, UK
Ernestina Cianca, Italy
Riccardo Colella, Italy
Mario Collotta, Italy
Massimo Condoluci, Sweden
Daniel G. Costa, Brazil
Bernard Cousin, France
Telmo Reis Cunha, Portugal
Laurie Cuthbert, Macau
Donatella Darsena, Italy
Pham Tien Dat, Japan
André de Almeida, Brazil
Antonio De Domenico, France
Antonio de la Oliva, Spain
Gianluca De Marco, Italy
Luca De Nardis, Italy
Liang Dong, USA
Mohammed El-Hajjar, UK
Oscar Esparza, Spain
- Maria Fazio, Italy
Mauro Femminella, Italy
Manuel Fernandez-Veiga, Spain
Gianluigi Ferrari, Italy
Ilario Filippini, Italy
Jesus Fontecha, Spain
Luca Foschini, Italy
A. G. Fragkiadakis, Greece
Sabrina Gaito, Italy
Óscar García, Spain
Manuel García Sánchez, Spain
L. J. García Villalba, Spain
José A. García-Naya, Spain
Miguel Garcia-Pineda, Spain
A.-J. García-Sánchez, Spain
Piedad Garrido, Spain
Vincent Gauthier, France
Carlo Giannelli, Italy
Carles Gomez, Spain
Juan A. Gómez-Pulido, Spain
Ke Guan, China
Antonio Guerrieri, Italy
Daojing He, China
Paul Honeine, France
Sergio Ilarri, Spain
Antonio Jara, Switzerland
Xiaohong Jiang, Japan
Minho Jo, Republic of Korea
Shigeru Kashiara, Japan
Dimitrios Katsaros, Greece
Minseok Kim, Japan
Mario Kolberg, UK
Nikoz Komninos, UK
Juan A. L. Riquelme, Spain
Pavlos I. Lazaridis, UK
Tuan Anh Le, UK
Xianfu Lei, China
Hoa Le-Minh, UK
Jaime Lloret, Spain
Miguel López-Benítez, UK
Martín López-Nores, Spain
Javier D. S. Lorente, Spain
Tony T. Luo, Singapore
Maode Ma, Singapore
- Imadeldin Mahgoub, USA
Pietro Manzoni, Spain
Álvaro Marco, Spain
Gustavo Marfia, Italy
Francisco J. Martinez, Spain
Davide Mattera, Italy
Michael McGuire, Canada
Nathalie Mitton, France
Klaus Moessner, UK
Antonella Molinaro, Italy
Simone Morosi, Italy
Kumudu S. Munasinghe, Australia
Enrico Natalizio, France
Keivan Navaie, UK
Thomas Newe, Ireland
Tuan M. Nguyen, Vietnam
Petros Nicopolitidis, Greece
Giovanni Pau, Italy
Rafael Pérez-Jiménez, Spain
Matteo Petracca, Italy
Nada Y. Philip, UK
Marco Picone, Italy
Daniele Pinchera, Italy
Giuseppe Piro, Italy
Vicent Pla, Spain
Javier Prieto, Spain
Rüdiger C. Prys, Germany
Sujan Rajbhandari, UK
Rajib Rana, Australia
Luca Reggiani, Italy
Daniel G. Reina, Spain
Jose Santa, Spain
Stefano Savazzi, Italy
Hans Schotten, Germany
Patrick Seeling, USA
Muhammad Z. Shakir, UK
Mohammad Shojafar, Italy
Giovanni Stea, Italy
Enrique Stevens-Navarro, Mexico
Zhou Su, Japan
Luis Suarez, Russia
Ville Syrjälä, Finland
Hwee Pink Tan, Singapore
Pierre-Martin Tardif, Canada






Mauro Tortonesi, Italy
Federico Tramarin, Italy
Reza Monir Vaghefi, USA
J. F. Valenzuela-Valdés, Spain

Aline C. Viana, France
Enrico M. Vitucci, Italy
Honggang Wang, USA
Jie Yang, USA

Sherali Zeadally, USA
Jie Zhang, UK
Meiling Zhu, UK

Contents

Energy Efficient Wireless Networks

Yujin Lim , Gianluigi Ferrari , Hideyuki Takahashi , and Rossana M. C. Andrade
Editorial (1 page), Article ID 1726458, Volume 2019 (2019)



A Heterogeneous Energy Wireless Sensor Network Clustering Protocol

Mengjia Zeng , Xu Huang , Bo Zheng , and Xiangxiang Fan 
Research Article (11 pages), Article ID 7367281, Volume 2019 (2019)

Enhancing Parasitic Interference Directional Antennas with Multiple Director Elements

Javier Schandy , Leonardo Steinfeld , Benigno Rodríguez , Juan P. González , and Fernando Silveira
Research Article (9 pages), Article ID 7546785, Volume 2019 (2019)

Online Supervisory Control and Resource Management for Energy Harvesting BS Sites Empowered with Computation Capabilities

Thembelihle Dlamini , Ángel Fernández Gambín , Daniele Munaretto, and Michele Rossi
Research Article (17 pages), Article ID 8593808, Volume 2019 (2019)

A Fuzzy Logic System for Vertical Handover and Maximizing Battery Lifetime in Heterogeneous Wireless Multimedia Networks

Thiago Coqueiro , José Jailton , Tássio Carvalho, and Renato Francês
Research Article (13 pages), Article ID 1213724, Volume 2019 (2019)

Autonomous Transmission Power Decision Strategy for Energy Efficient Operation of a Dense Small Cell Network

Jaesung Park , and Heejung Byun 
Research Article (8 pages), Article ID 6472982, Volume 2018 (2019)

Editorial

Energy Efficient Wireless Networks

Yujin Lim ¹, **Gianluigi Ferrari** ², **Hideyuki Takahashi** ³ and **Rossana M. C. Andrade**⁴

¹*Sookmyung Women's University, Seoul, Republic of Korea*

²*University of Parma, Parma, Italy*

³*Tohoku Gakuin University, Sendai, Japan*

⁴*Federal University of Ceará, Fortaleza, Brazil*

Correspondence should be addressed to Yujin Lim; yujin91@sookmyung.ac.kr

Received 30 May 2019; Accepted 30 May 2019; Published 12 June 2019

Copyright © 2019 Yujin Lim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As IoT services become vitalized, this leads to an explosion of sensor nodes more and more. As the energy consumption of a wireless network increases with the number of nodes, it becomes critical to reduce the energy consumption of a wireless network. The sensor nodes are required to function for from a few months to a few years without recharging once they are installed. In order to enlarge the lifetime of the nodes, an energy efficient scheme is needed to save the energy of the nodes. Now, the energy efficiency has become one of the key criteria for designing wireless networks, not only because of the environmental concerns, but also practical reasons of the nodes in wireless networks. Even though a significant number of studies taking evolutionary approaches are in progress to give advancements of the energy efficient wireless networks, much effort is still required to crystallize the energy efficient wireless networks.

The scope of this special issue is in line with recent contributions from academia and industry to tackle the technical challenges in order to concretize the energy efficient wireless networks. For the current issue, we are pleased to introduce a collection of papers covering a range of topics as follows:

- (i) Switching strategy for energy efficient wireless networks
- (ii) Clustering protocol for energy efficient wireless networks
- (iii) Transmission power control for energy efficient wireless networks

(iv) Vertical handover algorithm for energy efficient wireless networks

(v) Optimization of directional antennas for energy efficient wireless networks

As always, we appreciate the high-quality submissions from authors and the support of the community of reviewers.

Conflicts of Interest

The guest editors declare no conflicts of interest.

*Yujin Lim
Gianluigi Ferrari
Hideyuki Takahashi
Rossana M. C. Andrade*

Research Article

A Heterogeneous Energy Wireless Sensor Network Clustering Protocol

Mengjia Zeng ^{1,2}, Xu Huang ^{1,3}, Bo Zheng ¹ and Xiangxiang Fan ¹

¹School of Information Engineering, Huzhou University, Huzhou, Zhejiang Province 313000, China

²Qiuzhen School of Huzhou Teachers College, Huzhou, Zhejiang Province 313000, China

³College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang Province 310058, China

Correspondence should be addressed to Xu Huang; hx@zjhu.edu.cn

Received 10 October 2018; Revised 6 March 2019; Accepted 19 May 2019; Published 26 May 2019

Academic Editor: Gianluigi Ferrari

Copyright © 2019 Mengjia Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Low-Efficiency Adaptive Clustering Hierarchical (LEACH) protocol, a hierarchical routing protocol, has the advantage of simple implementation and can effectively balance network loads. However, to date there has been a lack of consideration for its use in heterogeneous energy network environments. To solve this problem, the Energy-Coverage Ratio Clustering Protocol (E-CRCP) is proposed, which is based on reducing the energy consumption of the system and utilizing the regional coverage ratio. First, the energy model is designed. The optimal number of clusters is determined based on the principle of “minimum energy consumption”, and the cluster head selection is based on the principle of “regional coverage maximization”. In order to balance the network load as much as possible, in the next iteration of cluster head selection, the cluster head with the lowest residual energy and the highest energy consumption is replaced to prolong the network's life. Our simulated results demonstrate that the proposed method has some advantages in terms of longer network life, load balancing, and overall energy consumption in the environment of a heterogeneous energy wireless sensor network.

1. Introduction

Wireless sensor networks are complex and changeable working environments, and they require a large number of sensor nodes to complete measurement tasks cooperatively [1]. The reasonable placement of nodes and the optimization of parameters according to different environments help improve the overall efficiency and reduce wireless sensor network (WSN) costs. In recent years, researchers have tried to solve the energy consumption and life optimization problem of wireless sensor networks from different angles, and they have put forward many effective methods. Data routing is a problem that must be considered in wireless sensor networks, and one of the most important goals of data routing is energy saving [2]. Therefore, many researchers have tried to explore this issue from the aspects of low computational complexity, energy balance, and efficient routing. Reference [3] proposes a data decoding and fusion scheme for wireless sensor networks, which achieves data fusion in resource-constrained scenarios with low computational complexity. This method

has explicit form of state estimation and residuals and is suitable for online computing. However, this method is aimed at the application of CEO (central evaluation officer) scenarios, which limits the migration of the algorithm. Document [4] studies the expansion state problem in data aggregation based on mobile agents, but the algorithm needs to calculate the dynamic migration path of mobile agents and deal with the fault and passive nodes, which increases the computational cost and hardware cost. Specific WSN software architecture design is important for maximizing network lifetime [5]. The token-based wireless sensor network cluster communication architecture in document [6] is to achieve energy-saving goals from this aspect, but the cost factor is introduced in the next hop node selection process, which increases the computing cost. Data volume in wireless sensor networks tends to grow continuously in both input and output [7]. Literature [8, 9] discussed the problem of reducing energy consumption from aspect of reducing the scale of data fusion, but literature [8] only considered the correlation of data and did not consider the correlation of adjacent sensor nodes.

Literature [9] used broadcasting mode to reduce the delay rate and prolong the life but did not discuss the flooding problem caused by broadcasting mechanism. The PED and PAD protocols proposed in literature [10] are event-driven and query-driven. The network load only becomes heavier when triggering data conditions, and most other time is lighter. Therefore, the energy consumption of network load is reduced to a certain extent. However, dynamic switching of network reporting schemes is needed in the implementation process, which increases the control overhead. Reference [11] proposes an energy-efficient cluster adaptive time division multiple access protocol EA-TDMA, which is a communication protocol between sensors in railway transportation system. This protocol improves energy efficiency by collecting information about future data packets rather than dispatching data packet exchanges in the competition stage [12], it is especially suitable for high-flow load characteristics of train operation [13], but its universality needs further verification.

The above-mentioned literature has actively explored and addressed the energy consumption of wireless sensor networks from different perspectives, while other researchers have studied this issue from the perspective of hierarchical mechanisms. The use of hierarchical mechanisms can optimize data delay to increase network scalability, reduce data redundancy and communication load, and optimize a network's lifespan [14]. Reference [15] studied the impact of uniform and nonuniform clustering on the performance of cluster sensor networks using numerical methods. It is concluded that uniform clustering has lower probability of decision-making errors than nonuniform clustering. Reference [16] implemented an efficient clustering protocol for wireless sensor networks from the perspective of fuzzy search to dynamically generate the optimal cluster number in each round using decentralization mechanism. However, the recalculation of the number of clusters per round increases the computational overhead. The EACA protocol proposed in [17] achieves better system lifetime prolongation effect, but it only considers the energy consumed by cluster head transmission and does not consider the condition that the base station is located at the far end of the network and the transmission energy consumption between cluster head and intracluster nodes. Document [18] introduces cluster sender for data transmission. Cluster heads are only responsible for allocating the time slots of TDMA within the cluster, which reduces the transmission burden of cluster heads and prolongs the network lifetime. But this method increases the election cost of cluster heads and cluster senders. From the above literatures, it can be seen that layering or clustering is an important technical means to reduce network energy consumption and prolong network life.

The typical representative of the hierarchical clustering routing protocol is the LEACH protocol [19] proposed by W. R. Heintzelman. The basic idea of this algorithm is that the cluster head nodes are randomly selected in a cyclic way, and the energy load of the whole network is equally distributed to each sensor node, which results in a reduction of the energy load and low network energy consumption that improves the overall network lifetime. The LEACH algorithm is divided into three parts: (1) cluster head election; (2) cluster members

join clusters; and (3) cluster routing. At the time of election, each node generates a random number between 0 and 1. If the number is less than the threshold $T(n)$, the node will become the cluster head. The formula to calculate $T(n)$ [19] is

$$T(n)_{LEACH} = \begin{cases} \frac{p}{1 - p(\theta \bmod (1/p))} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where p represents the cluster head ratio in the network; that is, the ratio of the number of cluster head nodes to the total number of nodes in the network. θ represents the current number of iterations. The set G indicates the set of nodes during the first $1/p$ iterations that are not cluster heads. From (1), we know that all nodes have the same probability of being selected as a cluster head. The energy consumption of all nodes in the system is balanced, which prolongs the life cycle of the system. Unfortunately, the LEACH algorithm has the following shortcomings: (1) the cluster head election is completely random, which may cause the cluster heads to be distributed unevenly in the monitored area. In turn, this will create an uneven global energy consumption distribution, especially for the node farthest away from the base station which may die early; (2) the scalability is poor, and communication between the cluster head and the base station is "single-hop", which is unsuitable for large-scale network applications; and (3) poor adaptability, due to diverse applications of wireless sensor networks, where the requirements of each may not be the same. As the LEACH algorithm adopts a unified, whole-network sampling and transmission period, it cannot be applied to heterogeneous networks [20–22].

In view of the shortcomings of the LEACH algorithm, considerable research has been done to address them. In [23], based on the LEACH protocol, energy-efficient and cooperative target tracking was regarded as a utility function of a cross-layer cluster optimization problem, which can obtain better simulation results, but does not involve any discussion of heterogeneous networks. A distributed algorithm proposed in [24] was used as an extension of the LEACH clustering algorithm. Although the lifetime of a network is longer than that of a LEACH network, the LEACH algorithm is only applicable to the problem of unit and nonunit circles, which has its limitations. Hence, the NEAP (Novel Energy Adaptive Protocol) energy-adaptive protocol was proposed [25]. Unlike the LEACH protocol, the threshold of NEAP is a function of the residual energy of the nodes, as shown in [25] (2):

$$T(n)_{NEAP} = \frac{p}{1 - p(\theta \bmod (1/p))} \times \left[\frac{E_{cur}}{E_{ini}} + \left(r_s \operatorname{div} \frac{1}{p} \right) \left(1 - \frac{E_{cur}}{E_{ini}} \right) \right] \quad (2)$$

where E_{cur} and E_{ini} represent the current and initial energy of the nodes, respectively, and r_s indicates the iterations where nodes have not been selected as cluster heads. Compared with the LEACH protocol, the NEAP protocol has better performance when selecting cluster heads, but to date there

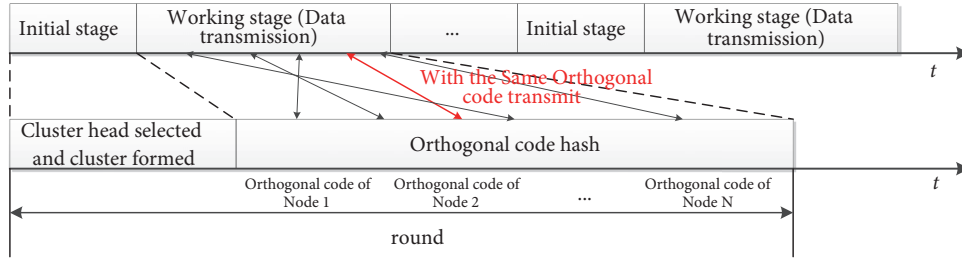


FIGURE 1: Graph of system timing.

has not been any discussion regarding energy optimization when selecting cluster heads. The authors of [26] proposed the Distributed Energy Efficient Cluster (DEEC) protocol, which adopts a heterogeneous, two-level energy structure network model, where each node chooses its cluster head based on its own residual energy. However, it does not consider the problem of system energy balance. The authors of [27] proposed an improved DEEC algorithm called the DDEEC (developed DEEC) protocol, which dynamically changes the standard for selecting cluster heads and then balances the energy consumed by the nodes. The authors of [28] also proposed an improved DEEC protocol called SEP (the Stable Election Protocol). The basic idea of SEP is based on the different initial energies of the nodes, which are divided into two categories: ordinary nodes and advanced nodes. Advanced nodes have a higher initial energy than ordinary nodes, and the probability of an advanced node being selected as a cluster head is higher. Reference [29] discussed the clustering protocol based on residual energy and distance information. Although [28, 29] carried out some useful explorations on the life cycle of wireless sensor networks with heterogeneous energy, the heterogeneous characteristics of node energy are not universal enough.

In view of the above problems, we propose a clustering protocol that can be applied to a heterogeneous energy wireless sensor network, the Energy-Coverage Ratio Clustering Protocol (E-CRCP). E-CRCP is an improvement of the LEACH protocol in terms of the selection of cluster heads, where we consider the lowest energy consumption in each communication iteration and the maximum cluster head coverage ratio. E-CRCP balances both the cluster head selection and the system energy load to extend the system's life cycle.

2. System Model

In order to simplify the problem, the following assumptions are made in our study: (1) the wireless sensor network is composed of a large number, N , of fixed sensor nodes; that is, once the sensor nodes are arranged in a monitoring area, the locations of the sensor nodes are no longer changed; (2) the nodes arranged in the monitoring area are subjected to a certain method to get their positional information (such as GPS); (3) all nodes are basically synchronized in second precision; (4) only one base station exists in the monitoring area, and its position is fixed in region A, which is at the

center; (5) the N sensor nodes have heterogeneous energy levels; i.e., they have different initial energies; (6) the system routing model is based on a hierarchical routing protocol cluster that consists a cluster head (CH) node and several noncluster head (non-CH) nodes, which are called normal nodes. First, the normal nodes transmit their sensing data to their respective CHs, where each CH node is responsible for fusing the data from the normal nodes and forwarding it to the base station (BS).

2.1. System Timing. In this paper, we divide the system timing into several rounds, where a cycle is called an iteration. The initial and working stages are set for each iteration, as shown in Figure 1. At the initial stage, CHs are selected, and clusters are formed. Data are transmitted at the working stage. Data transmission from non-CH to CH nodes follows the principle of Code-Division Multiple Access (CDMA); that is to say, all nodes share spectrum resources in the form of orthogonal address codes; only nodes with the same orthogonal code can transmit information between sender and receiver. Nodes transmit data do not interfere with each other.

2.2. Energy Model. Each sensor node in the system needs to receive and send information in the process of data transmission. From the point of view of energy consumption, the sensor node is simplified to consist of only a receiver and an emitter, in which the emitter consists of an emitting component and a power amplifier. During data transmission, the sensor node will switch between emitting and receiving states, which means that the node is in an emitting or receiving state at any given moment. When a sensor node emits or receives data, it consumes energy. The emitter consumes energy when it runs the emitter components and power amplifiers. Assuming that the receiving and sending ends are placed at a distance d away from each other, if d is small, the free space transmission model is adopted. When d is large, the multipath fading channel model is adopted. Figure 2 shows the system radio energy consumption model.

Based on [30], we use the energy consumed by transmitting a q bits message between the transmitter and receiver of d is

$$E_{tr}(q, d) = \begin{cases} q \times E_{el} + q \times E_{frs} d^2 & \text{if } d < d_0 \\ q \times E_{el} + q \times E_{tworay} d^4 & \text{if } d \geq d_0 \end{cases} \quad (3)$$

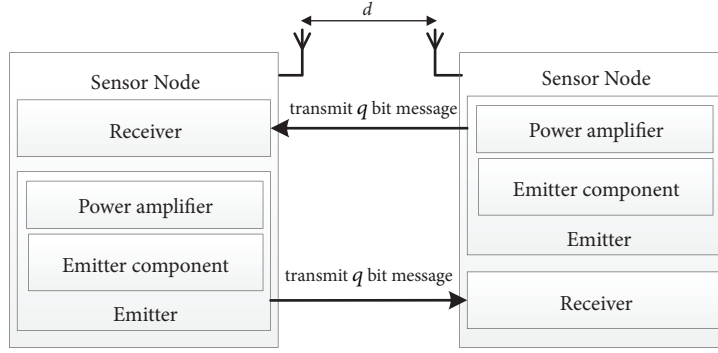


FIGURE 2: Radio energy consumption model.

where E_{el} represents the energy consumed per bit when the emitter components are running, E_{frs} and E_{tworay} represent the energy consumed by the unit power amplifier in the free space and the double path propagation model (two-ray ground model), respectively, and d_0 is as follows:

$$d_0 = \sqrt{\frac{(4\pi)^2 \times l \times h_t^2 \times h_r^2}{\lambda^2}} = \sqrt{\frac{E_{frs}}{E_{tworay}}} \quad (4)$$

In (4), h_t and h_r are the respective ground clearance of the sending and receiving ends and λ is the wavelength. Correspondingly, the energy consumed by receiving a q bits message is

$$E_{re}(q) = q \times E_{el} \quad (5)$$

2.3. Determination of the Optimal Number of Clusters. Cluster-based hierarchical routing protocols first divide nodes in the network into different clusters. How to assemble the cluster and select CH nodes is the problem that needs to be solved. The optimal probability p_{opt} of a node being a CH is an important embodiment of the clustering results. The authors of [31] proved that if the cluster number is nonoptimal, the energy consumption of the system will increase exponentially. Therefore, in this paper, we first calculate the optimal number of clusters from the perspective of the minimum energy consumption of the system. Routing protocols follow point 6 at the beginning of Section 2.

The number of nodes in the network and the initial energy of each node are different. We assume that, in iteration M th, C -number of CHs are generated, and there are $N/C - 1$ cluster member nodes in each cluster. Then, the member nodes send a q bits control message to the CH. The basic principle of a cluster is that the communication cost of all nodes in the cluster should be as low as possible. Generally, the CH is located at the center of a cluster, and the distance from other nodes to the CH is small. So, all member nodes that have a close distance to a CH are generally added to that cluster. It is assumed that the member nodes transmit the data to the CH based on the free space channel model. Therefore, we can get the energy consumption of each cluster member node that sends a q bits control message to the CHs, $E_{normal-CH}$, as

$$E_{normal-CH} = q \times E_{el} + q \times E_{frs} d_{CH}^2 \quad (6)$$

where d_{CH} is the average distance between the cluster members and the CH.

Next, let us consider a general situation where a CH is located far from the BS, and the message transmission model is a multipath fading channel model. The CH receives q bits control messages from the $N/C - 1$ cluster member nodes and performs data fusion. Then, the total energy consumed by the transmission of the message to the base station is E_{CH-BS} :

$$E_{CH-BS} = q \times E_{el} \times \left(\frac{N}{C} - 1\right) + q \times E_{data} \times \left(\frac{N}{C} - 1\right) + q \times E_{tworay} d_{BS}^4 \quad (7)$$

The first term in (7) is the energy consumed by the q bits control message from the $N/C - 1$ cluster member nodes. The second term is the energy consumed when fusing the data, and the last term is the energy used to transmit the data to the base station. Among them, E_{data} indicates the energy consumed when fusing a one-bit message, and d_{BS} indicates the average distance from the CH to the base station. Therefore, the total energy of a cluster message in a communication iteration is $E_{cluster}$:

$$E_{cluster} = E_{CH-BS} + \left(\frac{N}{C} - 1\right) \times E_{normal-CH} \quad (8)$$

Then, in the M th iteration, the total energy consumed by the network is E_{round} :

$$\begin{aligned} E_{round} &= C \times E_{cluster} \\ &= C \times E_{CH-BS} + (N - c) \times E_{normal-CH} \\ &= q \times N \times E_{el} + q \times N \times E_{data} + q \times C \\ &\quad \times E_{tworay} \times d_{BS}^4 + (N - C) \times q \times E_{el} \\ &\quad + (N - C) \times q \times E_{frs} \times d_{CH}^2 \\ &= 2 \times q \times N \times E_{el} + q \times N \times E_{data} + q \times C \\ &\quad \times (E_{tworay} \times d_{BS}^4 - E_{el} - E_{frs} \times d_{CH}^2) \end{aligned} \quad (9)$$

It is hoped that the energy consumed by the network is the lowest in every iteration. Therefore, we can obtain the

partial derivative C through (9) and set it to 0; this will give the optimal number of clusters C_{opt} . Accordingly, the optimal cluster head ratio p_{opt} is

$$p_{opt} = \frac{C_{opt}}{N} \quad (10)$$

2.4. Cluster Head Selection Based on the Maximum Coverage Ratio. Although implementation of the LEACH algorithm is simple, the random selection of CHs may result in a high density of them in one area, with other regions containing few or even no CHs. Hence, the CHs may be distributed unevenly throughout the system. In the process of CH selection, we should take account of the coverage ratio to prevent the uneven distribution of CHs. Coverage generally refers to the area coverage ratio [32]. Although all sensors may operate in a system, it is difficult to ensure that the coverage ratio of the target area is 100% [33]. In practical applications, small monitoring vulnerabilities have little impact on the system and are deemed acceptable. A coverage mechanism is used to ensure that nodes are kept active while meeting coverage expectations, and in this work we use the cluster coverage ratio. Based on previous work [34], the CH selection process is as follows.

2.4.1. Coverage Problem Description. Suppose the monitoring area is a rectangle with a length of h meters and a width of w meters, and the area is $h * w$ m². Taking h as the ordinate and w as the abscissa when establishing the two-dimensional coordinate system, we can get the coordinates of the N sensor nodes in the two-dimensional coordinate system. Let us further suppose that the sensing radius of each sensor is r , and the communication radius is R . In order to ensure network connectivity and wireless interference, $R = 2r$ [35]. Using $c_i = \{x_i, y_i, r\}$, it is shown that a circle with a radius of r is the center of the node coordinates $\{x_i, y_i\}$ ($i \in 1, \dots, N$). Assuming that the monitoring target coordinates are (x, y) , the distance between the target and sensor nodes is $d(c_i) = \sqrt{(x_i - x)^2 + (y_i - y)^2}$. The event that the monitoring target is covered by the sensor node is e_i , and the probability $P\{e_i\}$ of the event is the probability that the target (x, y) is covered by the sensor node c_i . Next, we consider the monitoring environment and noise interference, where the probability distribution [36] of the sensor node measurement model in the actual application is given as

$$P_{cov}(x, y, c_i) = \begin{cases} 1, & \text{if } d \leq r - r_e \\ e^{-\alpha_1 \varphi_1 \beta_1 / \varphi_2 \beta_2 + \alpha_2}, & \text{if } r - r_e < d < r + r_e \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$(i \in 1, \dots, N)$

where r_e ($0 < r_e < r$) is the measurement reliability parameter of the sensor nodes and $\alpha_1, \alpha_2, \beta_1,$ and β_2 are measured

parameters related to the sensing node characteristics. φ_1 and φ_2 are input parameters:

$$\varphi_1 = r_e - r + d(c_i), \quad (i \in 1, \dots, N) \quad (12)$$

$$\varphi_2 = r_e + r - d(c_i), \quad (i \in 1, \dots, N) \quad (13)$$

To improve the probability that a target is measured, multiple sensor nodes are used to measure targets simultaneously. The combined measurement probability is as follows [29]:

$$P_{cov}(Cov) = 1 - \prod_{c_i \in Cov} (1 - P_{cov}(x, y, c_i)), \quad (14)$$

$(i \in 1, \dots, N)$

The monitoring area is a rectangle of $h * w$ m² and is discretized into pixels. The pixel size is determined according to the actual application scenario. Whether each pixel is covered or not is measured by the joint measurement probability of node set $P_{cov}(Cov)$. In this paper, the area coverage $R_{area}(C)$ of node set C is defined as the ratio of the coverage area of node set C to the total monitoring area:

$$R_{area}(C) = \frac{\sum_{i=1}^N P_{cov}(Cov)}{h * w} \quad (15)$$

Assuming that the monitoring area is a square of 20 m * 20 m, it is divided into 100 pixels of equal size, and 20 sensor nodes are put into the area. A diagram of the monitoring area is shown in Figure 3, which shows the location of sensor nodes in the area. The coverage problem is described as follows: (1) use (11)-(13) to calculate the coverage of a sensor node to each pixel; (2) use (14) to calculate the joint coverage of the sensor nodes to each pixel; (3) repeat steps (1) to (2) to calculate the joint coverage rate from the sensor nodes to the pixel points; and (4) use (15) to calculate the area coverage and consider (15) as the optimization objective function of the coverage control algorithm.

2.4.2. Cluster Head Selection Algorithm. The N sensors in the monitoring area are numbered as 1- N , and we randomly select a node as the CH, assuming that the selected node is K . According to the optimal number of CHs calculated before, we need to select $C_{opt} - 1$ nodes from the remaining $N-1$ nodes as CHs. The selection principle is to compute the node coverage rate according to the steps mentioned above, followed by the maximum coverage rate, which is determined using (15). In the process of data communication, the energy consumption of all CHs is recorded as $E_{con_{i\vartheta}}$ ($i \in 1, \dots, C_{opt}, \vartheta \in 1, \dots, \theta$), which is calculated according to (3)-(7). The cumulative energy consumption of the first θ iterations is recorded as E_{con_i} ($i \in 1, \dots, C_{opt}$), and the residual energy of each CH is recorded as E_{rem_i} ($i \in 1, \dots, C_{opt}$), where

$$E_{con_i} = \sum_{\vartheta=1}^{\theta} E_{con_{i\vartheta}}, \quad (i \in 1, \dots, C_{opt}, \theta \in 1, \dots, R_n), \quad (16)$$

$$E_{rem_i} = E_{0_i} - E_{con_i}, \quad (i \in 1, \dots, C_{opt}), \quad (17)$$

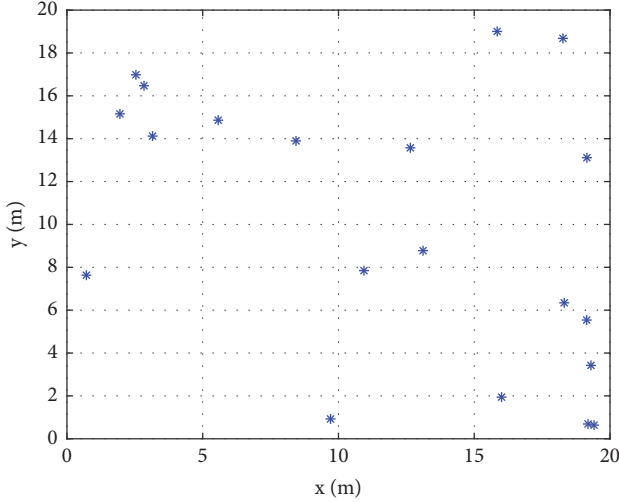


FIGURE 3: Random distribution of sensor nodes in 20 m * 20 m monitoring area.

where E_{0_i} is the initial energy of the sensor node, ϑ is the current number of iterations, and R_n is the number of iterations at the point where the energy of the system is exhausted or the maximum number of cycles assumed by the algorithm has been reached.

Next, the CH vitality parameter η_i is introduced, which is defined as follows:

$$\eta_i = \frac{E_{rem_i}}{E_{con_i}}, \quad (i \in 1, \dots, C_{opt}) \quad (18)$$

With the same energy consumption, the more the remaining energy is, the greater η_i is, and the higher the vitality of the CH is, the longer will be the life cycle. For the same residual energy, the more energy consumed by the current iteration is and the smaller the η_i is, the lower will be the vitality of the CH and the shorter will be the life cycle. After the end of a communication iteration, η_i of C_{opt} CHs are sorted from small to large. In order to extend the life cycle of the system as long as possible, we hope that the greater the vitality of CH nodes, the better.

The above algorithm is a one-iteration CH selection algorithm. After a data communication iteration is finished, the next iteration of CHs is selected. Because a CH needs to collect data from common nodes first and then fuse the data to send to the base station, the energy consumption of the CHs will be far greater than that of ordinary nodes. In order to balance the energy load of the whole system, we should try to let every node have the chance of becoming a CH. So, when the next iteration of CHs is selected, we need to replace the local CH and retain as many of the strongest CHs as possible. The replace proportion ρ ($\rho \in (0, 1]$) is a pure decimal in the $(0, 1]$ interval, where $\rho = 1$ indicates that the C_{opt} CHs of the current iteration are all replaced, and all the next iteration CHs are selected from the noncluster head set G . A value of $\rho = 0$ indicates that all CHs remain unchanged in the next iteration, which is not exist in this model, so $\rho \neq 0$. The number of replaced CHs is $C_{rep} = C_{opt} \times \rho$, so, we replace the

first C_{rep} CHs with η_i from small to large, selecting the C_{rep} CHs from the non-CH set G to maximize the region coverage in (15) and complete the next CH selection iteration. Because the data transmission of this model follows the CDMA, it can be seen from Section 2.1 that each communication iteration is divided into the initial stage of cluster head selection and clustering and the data transmission process in the working stage. The process of CH selection and clustering in the initial stage is actually a computing process. By selecting the control nodes with computing power to complete this process, the time required in the application process is fast and meets the needs. However, to ensure system efficiency, the time proportion of CH selection to the clustering process should not be too large in any communication iteration, or else system throughput will be affected. Therefore, in addition to the normal process of CH selection and clustering calculation, the system sets a time upper limit t_c for the initial stage, which is the average time for CH selection and clustering in the first θ communication iterations. If the initial stage is not completed when the t_c arrives, the system randomly generates the remaining CHs that should have been generated, but have not yet been generated. At this time, part of the algorithm degenerates to the LEACH algorithm. Figure 4 shows a flowchart of the CH selection algorithm.

In each CH selection iteration, the calculation process for each CH is mainly divided into area coverage calculation and low-energy CH replacement. In the process of calculating the regional coverage, the joint measurement probability $P_{cov}(Cov)$ is correlated with N^2 by (14), and the regional coverage $R_{area}(C)$ is also correlated with N^2 by (15), and therefore the computational time complexity of the regional coverage calculation is $O(N^2)$. In the process of CH selection and replacement, (18) shows that the calculation process is linearly related to N . Therefore, the time complexity of the whole algorithm is $O(N^2)$.

2.4.3. Clustering Process and Working Stage. After the C_{opt} CH nodes in one communication iteration have been elected, the CHs broadcast request messages to other normal nodes to join the cluster. After the non-CH nodes receive the message, they choose the nearest cluster to join until the cluster process terminates when the number of cluster nodes reaches N/C_{opt} or there are no remaining nodes. Because the value of N/C_{opt} may not be an integer, the number of nodes in the last cluster may be less than N/C_{opt} .

The working stage is also known as the data transmission stage. The CH broadcasts a CDMA data stream to notify its member nodes to start the data-acquisition process. The cluster member nodes send data to the CH according to the system timing in Figure 1, where the CH collects the node data and then transfers them to the base station. After the data transmission is completed, the algorithm will enter the next iteration of CH selection and form a new cluster. In the working stage, data acquisition begins with the CH sending a CDMA broadcast to its member nodes. The cluster members send the collected data to their respective CHs during the CDMA process. After receiving all the data, the CHs integrate them to reduce the noise in the signal and then send the

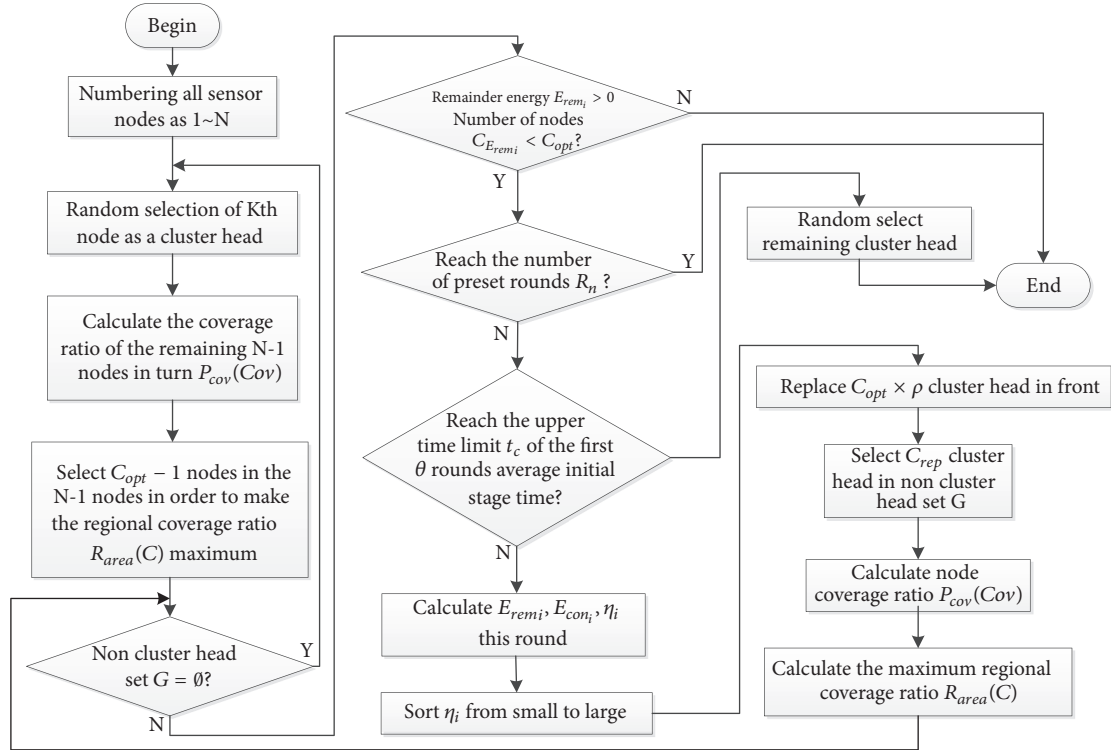


FIGURE 4: Cluster head selection algorithm flow chart.

TABLE 1: Simulation parameters.

Parameter	Value	Parameter	Value
Size of monitoring area $h \times w$	100 m \times 100m	Node energy coefficient α_i	Random number in (0,10), $i \in [1, N]$
Number of nodes N	100	E_{el}	50 nJ/bit
Node distribution	Random distribution	E_{frs}	10 nJ/bit/(m ²)
Sensing radius r	10 m	E_{tworay}	0.013 pJ/bit/(m ⁴)
Communication radius R	20 m	E_{data}	5 nJ/bit/signal
BS position	(50,50)	Size of data package	525 bytes
Node initial essential energy E_0	0.5 J	Size of CDMA package	25 bytes

integrated data to the base station in a single-hop or multihop manner. Then, the network begins to choose new CHs and form new clusters in the next iteration. When all nodes have become CHs, the next cycle will start.

3. Simulation Analysis

3.1. Simulation Parameter Hypothesis. In this section, we evaluate the performance of the proposed E-CRCP. The simulation was built using MATLAB R2016b and then compared with the LEACH, DDEEC, and SEP protocols. The specific simulation parameters are shown in Table 1.

In this experiment, the size of the monitoring area was fixed at 100m \times 100m, the coordinate axis range was [(0,0), (100,100)], the number of sensor nodes N was 100, and the position coordinates were obtained by coordinate axis if they were randomly distributed in the monitoring area. The base station was located in the center of the monitoring area (50,

50). If the basic initial energy $E_0 = 0.5J$ and the energy coefficient α_i ($i \in [1, N]$) is a random number in (0, 10), then the initial energy of each node is

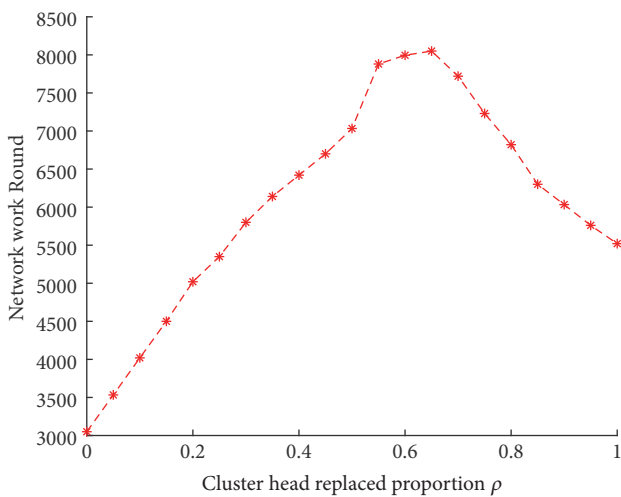
$$E_{0i} = E_0 \times \alpha_i \quad (19)$$

where E_{el} , E_{frs} , and E_{tworay} follow the parametric interpretation in Section 2.2 and E_{data} follows the parametric interpretation in Section 2.3.

3.2. Impact of the Cluster Head Replacement Ratio ρ on the Network Life Cycle. The purpose of the introduced CH replacement ratio ρ is to replace some low-vitality CHs in each cluster to prolong the network life cycle and balance the energy load of the whole system. The life cycle of the network, namely, the number of network lives, is expressed by the number of iterations. The value is equal to the iteration number when the last node in the network dies. The influence

TABLE 2: Comparison of several algorithms in terms of stability duration and network life cycle.

	Stability duration				Network life cycle			
	LEACH	DDEEC	SEP	E-CRCP	LEACH	DDEEC	SEP	E-CRCP
1	934	1255	1399	1689	5530	5860	8621	8650
2	955	1253	1396	1679	5532	5858	8625	8648
3	969	1258	1402	1682	5535	5861	8628	8651
4	936	1261	1392	1699	5540	5863	8629	8650
5	978	1251	1395	1692	5538	5864	8624	8649
6	933	1258	1403	1695	5534	5858	8626	8656
7	922	1247	1398	1698	5536	5869	8628	8655
8	955	1262	1400	1690	5538	5870	8629	8658
9	948	1263	1402	1688	5535	5865	8630	8652
10	972	1260	1398	1695	5539	5868	8625	8654

FIGURE 5: Impact of the cluster head replacement ratio ρ on the network life cycle.

of ρ on the life cycle of the system network is shown in Figure 5.

From Figure 5, we can see that ρ has a great influence on the life cycle of the network. A value of $\rho = 0$ indicates that all CHs in the next iteration are used in the local CH. In this case, the lifetime of the network is only maintained at about more than 3000 iterations, which is approximately half the highest point that is more than 8000 iterations. This shows that the life cycle of the CH nodes in the network determines the life cycle of the system. If the CHs are not replaced by other nodes, once all the CHs in the network are dead, the network will no longer work. It can be seen that the network without considering load energy consumption balance is not applicable. From the graph, the number of network iterations reaches the highest point near $\rho = 0.65$, which will prolong the life cycle of the network. A value of $\rho = 1$ indicates that all CHs in the current iteration are replaced, and the algorithm is identical to the LEACH algorithm. It is seen that the value of ρ first rises, reaches a maximum near $\rho = 0.65$, and then decreases, which shows that a too-large or too-small ρ has no

positive influence on the life cycle of the network. Only the appropriate ρ value can prolong the network life cycle.

3.3. Comparison of Several Algorithms in Terms of Network Life Cycle. In this section, we analyze the performance of the proposed E-CRCP algorithm in two aspects: stability time and network lifetime. The stability time and network lifetime are represented by the number of iterations. The stability time is equal to the number of iterations from the initial time to the iteration when the first node dies. The network lifetime is equal to the number of iterations from the initial time to the iteration when the last node dies. First, we analyze the stability time and network lifetime of four kinds of protocols: LEACH, DDEEC, SEP, and E-CRCP. Table 2 lists the experimental data of the 10 tests. From Table 2, we can see that the stability duration of LEACH, DDEEC, SEP, and E-CRCP is 950.2, 1256.8, 1398.5, and 1690.7, respectively. The average life spans of the networks are 5535.7, 5863.6, 8626.5, and 8652.3, respectively. These indicate that the proposed E-CRCP protocol can effectively extend the stability period and the network lifetime.

From Table 2, we can see that the E-CRCP proposed in this paper can extend the stability time and network lifetime better than LEACH, DDEEC, and SEP. The reason is that the LEACH protocol does not consider the residual energy of the node and simply gives the same opportunity to each node, whereas DDEEC only considers the residual energy of the node, and SEP only considers the node energy level. However, these considerations are not conducive to the selection of a good CH. The proposed E-CRCP dynamically adjusts the replacement ratio of the CHs, which promotes the E-CRCP to select the best CH, reduces energy consumption, and thus prolongs the stability time and network life. A prolonged network lifetime means that more nodes can collect data, which helps the base station receive more data packets.

3.4. Relationship between the Number of Active Nodes and the Network Lifetime. An active node is a working node. Once the energy of the node is exhausted, the node will no longer work (i.e., it dies). As time goes on, there will be fewer and fewer active nodes in the system. This section compares the lifetime of different protocols in the network

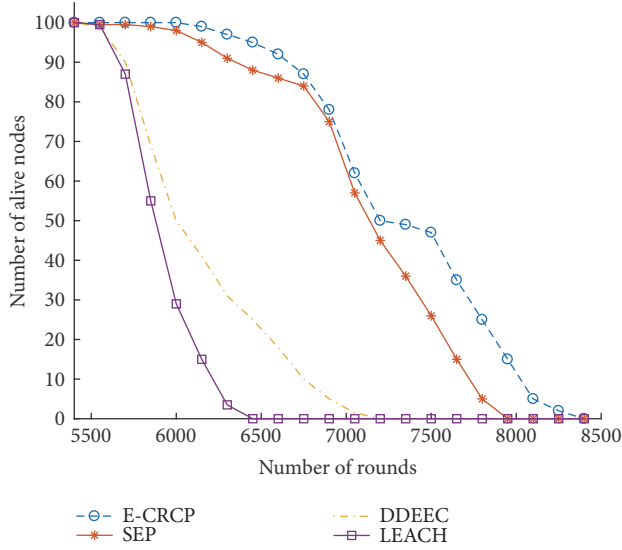


FIGURE 6: Relationship of the number of active nodes in the network and the network lifetime.

when the number of active nodes is different. Figure 6 reflects the influence of the number of active nodes on the network lifetime.

Figure 6 illustrates the relationship between the number of active nodes and the network lifetime. As we can see from the graph, with decrease of the number of alive nodes in the network, LEACH’s lifetime is about 6500 rounds, DDEEC’s lifetime is about 7200 rounds, and SEP’s lifetime is about 8000 rounds. That is to say, LEACH, DDEEC, and SEP to prolong the lifespan are failed. Conversely, the E-CRCP protocol can prolong the life span as the number of nodes decreases. This is because LEACH, DDEEC, and SEP do not take the cluster coverage mechanism into account. Each node needs to send all the collected environmental information to its CH nodes, including redundant information, thus increasing the energy consumption. The E-CRCP protocol not only considers the network energy load balance, but also takes the maximum coverage in the cluster into account, so that network life increases with the number of nodes decrease.

3.5. Influence of the Number of Sensor Nodes on Network Lifetime and Energy Consumption. To reveal the influence of the number of sensor nodes on network lifetime and energy consumption, the number of sensor nodes was chosen as $N = 100, 50,$ and $25,$ and the system was given the same average initial energy (assumed to be 0.5 J) under all three models, with the other parameters set as in Table 1. Network life was still measured by the number of iterations of network operation. Figure 7 shows the relationship between network lifetime and average residual energy for several protocols.

As can be seen from Figure 7, when the number of nodes in the system is $100, 50,$ and $25,$ the average residual energy of each protocol decreases as the number of iterations increases. When $N = 100$ and $50,$ the change trends of several protocols are very similar. The curve of each protocol shows a convex trend, but the curve shows a concave trend

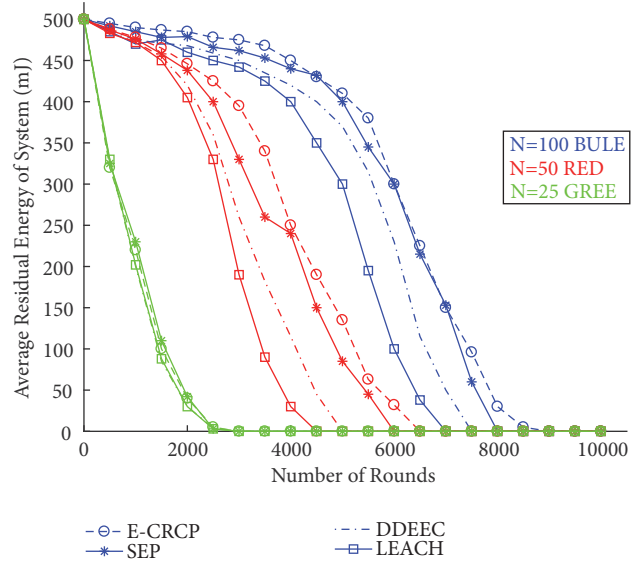


FIGURE 7: Influence of the number of sensor nodes on network lifetime and energy consumption.

when $N = 25.$ When $N = 100,$ the number of iterations to termination of the E-CRCP protocol was about 8500, whereas the number of iterations to termination when $N = 50$ was about 6500. In other words, the average residual energy of the system does not decrease synchronously as the number of nodes decreases exponentially but moves horizontally on the coordinate axis. This trend also exists for several other protocols. Moreover, from the graph, the performance of the proposed E-CRCP protocol is optimal when $N = 100$ or $N = 50.$ This shows that there should be an appropriate range to determine the number of nodes in the monitoring area. The appropriate number of sensor nodes in the monitoring area is conducive to extending the life of the network system. However, when N drops to $25,$ the performance of several protocols is similar, and the average residual energy of the system shows a rapid downward trend. The reason for this is that when the number of nodes in the system is small, the communication distance between nodes is generally longer. The energy consumption transmission model of the nodes therefore changes from a free space transmission model to a multipath fading transmission model, which makes the data transmission need more energy and makes the system energy rapidly decay to $0.$ In addition, when N is small, the E-CRCP algorithm proposed in this paper does not have any advantages in the calculation of regional coverage and optimal cluster number, which makes its performance not necessarily superior to other protocols. Hence, when N is appropriately chosen, the performance of the E-CRCP algorithm proposed in this paper is better than that of other algorithms.

3.6. Impact of Coverage Area on Network Life. To simplify the model, it was assumed that the monitoring area was equal to the coverage area, which was a square area with equal length and width, and that the base station was located in the center

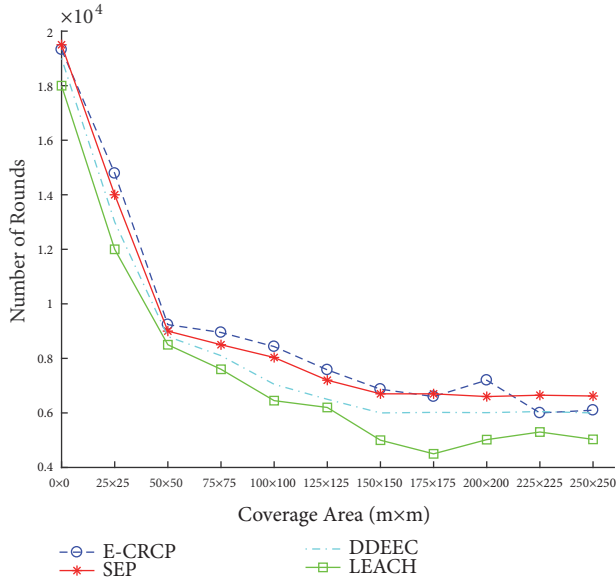


FIGURE 8: Impact of coverage area on network life.

of the area. The other parameters are listed in Table 1. Figure 8 shows the influence of changes in coverage area on network lifetime.

From Figure 8, it is clear that, in the range of $[(0,0), (50,50)]$, changes in coverage have a strong influence on network lifetime, but in the range of $[(50,50), (150,150)]$, this influence diminishes. In these two ranges, the influence of coverage on network lifetime shows a linear trend. As the coverage area continues to increase, there is no longer a stable linear trend between coverage area and network life. The reason for this is that the larger coverage area leads to a change in the data transmission model between nodes, making the state of the model unstable. It can be seen that the appropriate choice of coverage area has a positive impact on system stability.

4. Conclusions

In this paper, a CH selection protocol (E-CRCP), which is effectively applied to heterogeneous energy wireless sensor networks, is proposed as the solution to the CH selection problem in wireless sensor networks. First, a system-wide energy consumption model is established. The optimal number of system clusters is determined in the case of minimum energy consumption. Next, CH nodes are selected under the condition that the CH coverage is at a maximum, and CH nodes that consume a large amount of energy are replaced in the next communication iteration. The remaining members of each cluster join their nearest cluster and send their own data to the CH node. The CH node then sends the data to the base station after the data of each member node are fused, thus completing a single communication iteration. Our simulated results show that the algorithm proposed in this paper has obvious advantages over the LEACH, DDEEC, and SEP protocols in terms of the network

lifetime of heterogeneous energy network applications. In the process of CH selection, E-CRCP reduces the overall energy consumption of the network, balances the network load, and prolongs the network life.

Data Availability

The data used to support the findings of this study are included within Table 2 of the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partly supported by Zhejiang Province Soft Science Foundation of China (2019C35006), major project of ZJFHSSC (2015Z004), major project of Center for Research in Regional Economic Open and Development of Zhejiang (16JDGH011), Zhejiang Province Natural Science Foundation of China (LY18G020008; LQ18F020002), postdoctoral fund in China (2017M621489), National Natural Science Foundation of China (61202290), Huzhou University's Scientific Research Foundation in 2018 (2018XJKJ63), and Huzhou City Zhejiang Province Key Industry Project of China (2018GG29).

References

- [1] J. Yick, A. Bharathidasan, G. Pasternack, B. Mukherjee, and D. Ghosal, "Optimizing placement of beacons and data loggers in a sensor network - A case study," in *Proceedings of the 2004 IEEE Wireless Communications and Networking Conference, WCNC 2004*, pp. 2486–2491, USA, March 2004.
- [2] M. H. Anisi, A. H. Abdullah, S. A. Razak, and M. A. Ngadi, "An overview of data routing approaches for wireless sensor networks," *Sensors*, vol. 12, no. 4, pp. 3964–3996, 2012.
- [3] G. Ferrari, M. Martalò, and A. Abrardo, "Information fusion in wireless sensor networks with source correlation," *Information Fusion*, vol. 15, no. 1, pp. 80–89, 2014.
- [4] G. P. Gupta, M. Misra, and K. Garg, "Towards scalable and load-balanced mobile agents-based data aggregation for wireless sensor networks," *Computers and Electrical Engineering*, vol. 64, no. 10, pp. 262–276, 2017.
- [5] T. Cevik, "ECMTADR: energy conservative multitier architecture with data reduction for cluster-based wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, no. 1, pp. 1–11, 2015.
- [6] T. Çevik and A. H. Zaim, "EETBR: energy efficient token-based routing for wireless sensor networks," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 21, no. 2, pp. 513–526, 2013.
- [7] H. Teng, X. Liu, A. Liu et al., "Adaptive Transmission Power Control for Reliable Data Forwarding in Sensor Based Networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2068375, 22 pages, 2018.
- [8] H. Harb, A. Makhoul, D. Laiymani, and A. Jaber, "A distance-based data aggregation technique for periodic sensor networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 4, pp. 1–40, 2017.

- [9] S. Saginbekov and A. Jhumka, "Many-to-many data aggregation scheduling in wireless sensor networks with two sinks," *Computer Networks*, vol. 123, pp. 184–199, 2017.
- [10] B.-D. Lee and K.-H. Lim, "An energy-efficient hybrid data-gathering protocol based on the dynamic switching of reporting schemes in wireless sensor networks," *IEEE Systems Journal*, vol. 6, no. 3, pp. 378–387, 2012.
- [11] G. M. Shafiq, S. A. Azad, and A. B. M. S. Ali, "Energy-efficient wireless mac protocols for railway monitoring applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 649–659, 2013.
- [12] M. Molodova, Z. Li, A. Nunez, and R. Dollevoet, "Automatic detection of squats in railway infrastructure," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1980–1990, 2014.
- [13] Y. Chu, S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, "Application of reinforcement learning to medium access control for wireless sensor networks," *Engineering Applications of Artificial Intelligence*, vol. 46, pp. 23–32, 2015.
- [14] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000.
- [15] G. Ferrari, M. Martalò, and R. Pagliari, "Decentralized detection in clustered sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 959–973, 2011.
- [16] O. M. Alia, "A dynamic harmony search-based fuzzy clustering protocol for energy-efficient wireless sensor networks," *Annals of Telecommunications*, vol. 73, pp. 353–365, 2018.
- [17] O. Zytoune, Y. Fakhri, and D. Aboutajdine, "A fairly balanced clustering algorithm for routing in wireless sensor networks," *Sensor Review*, vol. 30, no. 3, pp. 242–249, 2010.
- [18] F. Bajaber and I. Awan, "An efficient cluster-based communication protocol for wireless sensor networks," *Telecommunication Systems*, vol. 55, no. 3, pp. 387–401, 2014.
- [19] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS '00)*, vol. 2, pp. 55–60, IEEE, Washington, DC, USA, January 2000.
- [20] Y. Li, N. Yu, W. Zhang et al., "Enhancing the performance of LEACH protocol in wireless sensor networks," in *Proceedings of the IEEE Conference on Computer Communications*, pp. 16–21, 2011.
- [21] Y. Gui, "Research on HEDSA data aggregation of wireless sensor network," *Computer Engineering*, vol. 37, no. 7, pp. 160–164, 2011.
- [22] K. S. Ahn, D. G. Kim, B. S. Sim et al., "Balanced chain-based routing protocol(BCBRP) for energy efficient wireless sensor networks," in *Proceedings of the 9th IEEE International Symposium on Parallel and Distributed Processing with Applications Workshops*, pp. 12–21, Republic of Korea, 2011.
- [23] L. Arienzo, "An information-theoretic approach for energy-efficient collaborative tracking in wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, pp. 1–14, 2010.
- [24] C.-F. Huang and Y.-C. Tseng, "The coverage problem in a wireless sensor network," in *Proceedings of the International Conference on ACM Wireless Sensor Networks and Applications*, pp. 115–121, ACM Press, New York, NY, USA, September 2003.
- [25] S.-S. Wang and Z.-P. Chen, "LCM: A link-aware clustering mechanism for energy-efficient routing in wireless sensor networks," *IEEE Sensors Journal*, vol. 13, no. 2, pp. 728–736, 2013.
- [26] L. Qing, Q. Zhu, and M. Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks," *Computer Communications*, vol. 29, no. 12, pp. 2230–2237, 2006.
- [27] S. H. Kang and T. Nguyen, "Distance based thresholds for cluster head selection in wireless sensor networks," *IEEE Communications Letters*, vol. 16, no. 9, pp. 1396–1399, 2012.
- [28] G. Smaragdakis, M. Ibrahim, and A. Bestavros, "SEP: a stable election protocol for clustered heterogeneous wireless sensor networks," in *Proceedings of the International Workshop on SANPA*, pp. 13–21, 2004.
- [29] X. Limou, L. Changyun, and M. Junfeng, "Residual energy and distance based clustering protocol for heterogeneous WSN," *Application Research of Computers*, vol. 33, pp. 2763–2766, 2016.
- [30] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application on specific protocol architecture for wireless micro sensor networks," *IEEE Trans on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.
- [31] C.-Y. Wan, A. T. Campbell, and L. Krishnamurthy, "PSFQ: A reliable transport protocol for wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pp. 1–11, ACM Press, New York, NY, USA, 2002.
- [32] L. Zhen, M. Yingchi, and C. Dong, "Effective data gathering scheme in heterogeneous energy WSN," *Application Research of Computers*, vol. 27, no. 6, pp. 2267–2270, 2010.
- [33] Z. Lin, Y. Feng, and L. Yu, "Research on the strategy of wireless sensor networks coverage by the particle optimization evolutionary," *Chinese Journal of Sensors and Actuators*, vol. 22, no. 6, pp. 873–877, 2009.
- [34] Y. Mao, Z. Wang, and Y. Liang, "Energy aware partial coverage protocol in wireless sensor networks," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing IEEE*, pp. 2535–2538, China, September 2007.
- [35] D. Tian and N. D. Georganas, "A coverage-preserving node scheduling scheme for large wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA '02)*, pp. 32–41, September 2002.
- [36] S. Li, C. Xu, W. Pan, and Y. Pan, "Sensor deployment optimization for detecting maneuvering targets," in *Proceedings of the 7th International Conference on Information Fusion*, pp. 1629–1635, IEEE Press, Piscataway, NJ, USA, 2005.

Research Article

Enhancing Parasitic Interference Directional Antennas with Multiple Director Elements

Javier Schandy , Leonardo Steinfeld ,
Benigno Rodríguez , Juan P. González , and Fernando Silveira

Department of Electrical Engineering, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay

Correspondence should be addressed to Javier Schandy; jschandy@fing.edu.uy

Received 21 September 2018; Accepted 6 February 2019; Published 21 February 2019

Academic Editor: Gianluigi Ferrari

Copyright © 2019 Javier Schandy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Swedish Institute of Computer Science Parasitic Interference Directional Antenna (SPIDA) is an electrically switched directional antenna that uses switched beamforming techniques to shape the antenna radiation pattern focusing the transmitted power in a given direction, increasing the maximum gain, and simultaneously reducing interference in other directions. This work extends the use of the SPIDA antenna, showing that using multiple director elements results in an improved performance in terms of maximum gain, narrower Half Power Beamwidth (HPBW), and a lower module of the S_{11} parameter. Measurements show that using three directors improves the maximum gain about 1.4 dB (6.8 dBi for the single director element antenna against 8.2 dBi for the antenna with three directors); the input impedance matching was also improved, obtaining a module of S_{11} parameter of -9.8 dB at the central frequency ($f_c = 2.4525$ GHz) against -7.5 dB for the antenna with a single director element. Finally, new intermediate directions of transmission can be achieved by using two successive director elements, where the power is focused in the bisectrix of the angle formed by the two directors. This converts a six-sector antenna like the SPIDA into a twelve-sector antenna without changing the hardware.

1. Introduction

In the last decades lots of new applications have emerged thanks to the availability of small devices capable of wireless communications, which allows sensing, processing, and communicating multiple physical variables or interacting with the physical world with a very low power consumption. These devices are expected to be of low cost and small size and to reach years of autonomy with small batteries, conforming large Wireless Sensor Networks (WSNs) with low operational and maintenance costs.

The IEEE 802.15.4 protocol is nowadays the most widely accepted standard in the 2.4 GHz ISM band for WSN. For the last decade most of the effort was in the development and optimization of wireless communication protocols. In comparison, the effort to improve the antenna of the WSN nodes has been very small, which could have helped the achievement of lower power consumptions and better efficiency. Many of the commercial nodes come with integrated omnidirectional antennas which radiate the energy in a

suboptimal way. Improving the antenna may provide better gain and SNR without increasing the overall irradiated power or may extend the battery lifetime, if the output power is reduced, keeping the distance range and the received signal strength. Another advantage of improving the antenna is that it may reduce the interference with other nodes by concentrating the radiated power in a certain direction, thus reducing the congestion that is known to be a common problem in multihop WSNs [1].

Also, directional antennas have been proposed as an alternative to increase the security in WSN [2].

One of the prominent ways to optimize an antenna is using dynamic beamforming. These techniques enable the increase of the antenna gain in some directions selected electronically on demand for each transmission [3, 4].

An example of this kind of antennas is the SPIDA antenna [5], which is an electrically switched directional antenna designed for WSN. This antenna has the advantage of having a low cost and an easy fabrication process, and also it has a

very small size that makes it very convenient to use in large scale deployments.

In this work, the simulation stage, the building, and characterization of a SPIDA antenna in the 2.4 GHz ISM band are described. During this process the use of multiple director elements to enable a complete new set of beam patterns for this antenna was explored. With the use of the professional electromagnetic simulation tool Computer Simulation Technology (CST) (<https://www.cst.com/>), the first steps of the performance analysis were made. Finally the SPIDA antenna with three directors was built and characterized, verifying in this way the predicted improvements with respect to the reference design antenna with only one director.

The main contributions of this work are (i) a complete characterization, simulation, and measurement of the reference design (single director element), including radiation pattern and S_{11} parameter, the last one missing in the bibliography; (ii) assessment of different configurations, by means of simulations, analyzing the resulting performance in terms of maximum gain in the main direction, Half Power Beamwidth (HPBW), and S_{11} parameter; and (iii) measurement of the three director elements configuration, identified as an equivalent to the reference but with improved performance, to confirm the simulation results.

The rest of this document is organized as follows. In Section 2 the main characteristics of the SPIDA antennas are introduced and the innovative idea of using multiple elements as directors is presented. In Section 3 the simulations and the results for different configurations are described. Section 4 describes the fabrication and characterization of the antenna and finally, in Section 5, the conclusions are summarized.

2. SPIDA Antenna

SPIDA antenna is a kind of antenna that allows us to perform switched beamforming [5–7]. Being able of controlling the beam direction dynamically is a very useful feature for wireless communication systems, also present in a similar kind of antennas based in this case in Electronically Steerable Passive Array Radiators (ESPAR) [8–10]. Comparing SPIDA antennas with ESPAR ones, the first are simpler and cheaper to fabricate, which represents an important advantage.

Phase-shifting antennas are also widely used in communications systems, but their use of heavy signal processing techniques makes them inadequate for WSN. In [11], a phase shifting directional antenna for WSN is proposed, but this antenna requires custom hardware to manage the signal processing and cannot be used with regular sensor nodes. In this aspect, the SPIDA antenna takes the lead as it can be attached to any sensor node with six output pins available.

These dynamic beamforming features are a promising alternative to optimize WSNs. This is the reason why several researchers have been developing this area [12–15], and it is the main motivation of this work.

2.1. SPIDA Baseline Design. The original antenna proposed by Nilsson [6] has six parasitic elements; thus the legs are separated 60° forming a hexagon. Figure 1 shows a sketch of

the constructed antenna. The overall size is such that it can be fitted in a cylinder with radius 52 mm and height 60 mm .

The antenna is composed of a planar structure in a symmetrical arrangement and a central vertical active element of 29.2 mm . Each of the six structures attached to the central hexagon is formed by a “leg” (that resembles the leg of a spider) with a vertical parasitic element. The length of this element is 27 mm and it can act either as a director, if it is left isolated (i.e., not connected), or as a reflector, if it is connected to ground. Each connection can be controlled electronically by a RF-switch that allows a microcontroller to manage dynamically the configuration of the antenna.

This antenna configuration has been characterized by other works, featuring approximately between 4 and 7 dBi in the principal direction with a 130° beamwidth [6, 13]. However, some important data, such as the S_{11} parameter, are missing in the corresponding reports. Thus, in this work a complete characterization is included, consisting in simulations and measurements.

2.2. SPIDA with Multiple Director Elements. Several works explored the use of SPIDA antennas for WSNs; initial efforts were in the antenna design itself [6], focusing, later, on solving problems not present in omnidirectional antennas, such as direction mismatch between main lobes of neighbor nodes during discovery phase. All these works were based on a six element SPIDA antenna with only one as a director. To the best of our knowledge, there is not any published report proposing the use of multiple director elements for this kind of antennas.

In this work different configurations were considered aiming to obtain higher gains in the main direction. Other aspects to study are how the adoption of multiple director elements affects the complete shape of the radiation pattern, and the input impedance matching.

Table 1 lists the eight different configurations considered, and Figure 2 depicts them graphically, where the direction of maximum gain is aligned with the horizontal axis (0°) for experiments 1, 2, 3, 4, and 5. For experiments 6 and 7, the direction of maximum gain is also aligned with the horizontal axis but equally high for 0° and 180° , and experiment 8 is omnidirectional. All these configurations were simulated in order to assess their performance. Section 3 describes the simulation results in detail. Among them, the most promising configuration (using three director elements) was measured and analyzed more deeply in Section 4.

3. Simulation

In order to assess the performance of the different configurations, the CST tool was used to simulate the antenna. This electromagnetic simulator was used to obtain the radiation pattern and the S_{11} parameter.

3.1. One Director (Conf. #1). This configuration, in which a single element acts as director, corresponds to the original configuration previously reported and is the reference design for comparison. From the simulated radiation pattern results, the maximum gain is 5.98 dBi with a HPBW of 129° and

TABLE 1: List of eight different configurations simulated with the location of the director element(s).

#	Description	Location
1	One director	0°
2	Two consecutive directors	30° and 330°
3	Three consecutive directors	0°, 60°, and 300°
4	Four consecutive directors	30°, 90°, 270°, and 330°
5	Five consecutive directors	0°, 60°, 120°, 240°, and 300°
6	Two opposed directors	0° and 180°
7	Four opposed directors	30°, 150°, 210°, and 330°
8	Six directors (all)	-

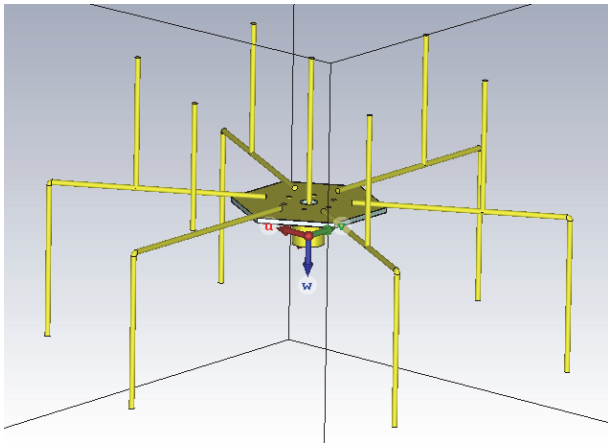


FIGURE 1: SPIDA antenna model in CST.

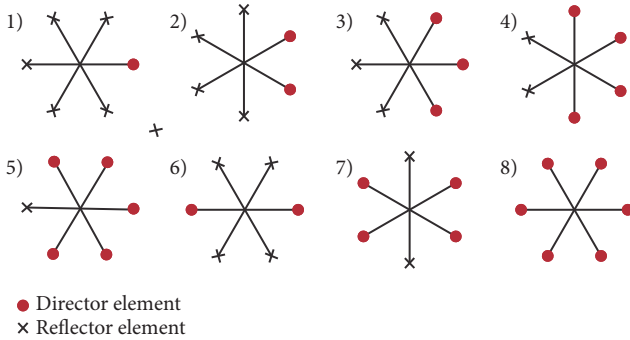


FIGURE 2: Eight different experiments, showing the location of the director and reflector element(s).

no side lobes, in agreement with previous reports (see the curve “Reference SPIDA” in Figure 3(a)). The front-to-back ratio (FTBR) is 21 dB and the S_{11} parameter varies from 2 GHz to 3 GHz as what the curve “Reference SPIDA” shows in Figure 3(b). These two curves are plotted together with the simulation results of other configurations for comparison purposes.

3.2. *Two Consecutive Directors (Conf. #2)*. As shown in Figure 3(a), the radiation pattern in H plane of this configuration presents a good directivity with a HPBW of 87°, significantly narrower than the 129° from the reference design. The maximum gain is 7.70 dBi, 1.72 dB higher than the configuration

with one director, but the FTBR is 9 dB, 12 dB lower than the reference design. Figure 3(b) shows the S_{11} parameter over the Smith Diagram (SD) when it varies from 2 GHz to 3 GHz. The S_{11} parameter in the central frequency of the IEEE 802.15.4 band 2.4525 GHz is -7.73 dB, about 4.00 dB lower than the reference design. According to these simulation results, this configuration outperforms the original one in maximum gain, presenting a narrow beamwidth which could be favorable in many scenarios, and having a better input impedance matching.

Another interesting characteristic of this configuration is that the main lobe direction is in the middle of the two directors, so with a six element antenna and using two consecutive directors it would be possible to direct the main beam in 12 different directions.

3.3. *Three Consecutive Directors (Conf. #3)*. The radiation pattern in H plane of this configuration is shown in Figure 4(a). It can be observed that this configuration has an even better directivity than that for one and two consecutive directors, with a HPBW of 76° compared to the 129° of the original design. The maximum gain is also better than the corresponding one for one and two directors, achieving 8.35 dBi, 2.37 dB higher than the reference design. This configuration presents a backlobe resulting in a FTBR of 13 dB, lower than that in the reference design but better than that in the configuration with two directors. Figure 4(b) shows the radiation pattern in E plane; for the sake of brevity E plane is shown for this configuration only, since the others have very similar characteristics.

The S_{11} parameter is shown in Figure 4(c), which achieves -11.25 dB at 2.4525 GHz, outperforming previous configurations.

The main lobe direction is aligned with the central director element (of the three used director elements), allowing the directional transmission in any of the original six directions, but with these improved characteristics.

These simulation results so far show that using two and three directors can focus the main beam to 12 different directions with better performance, in terms of gain, directivity, and impedance matching, than the original configuration using only one director. An aspect to consider in these configurations is that the FTBR is lower than the reference design, resulting in the radiation of more energy in the opposite direction.

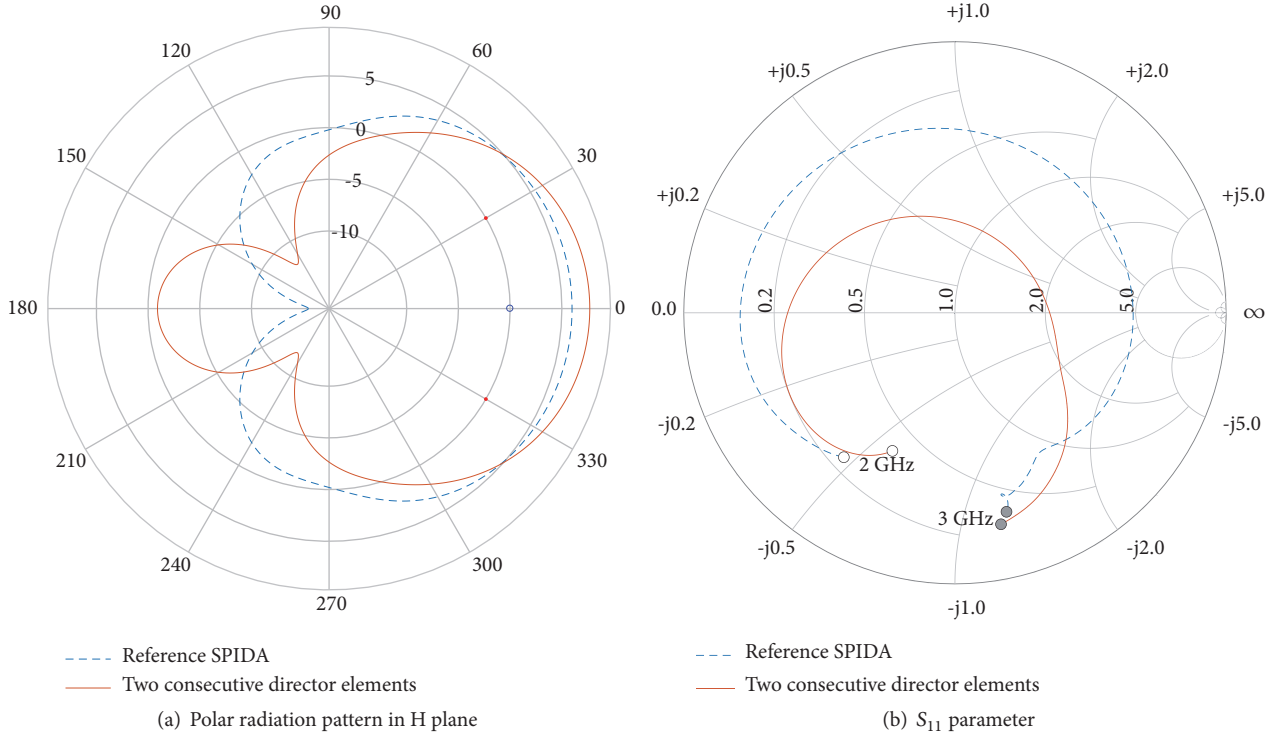


FIGURE 3: Two consecutive director elements (conf. #2).

TABLE 2: Simulation results.

#	Description	Half Power Beamwidth (3 dB)	Maximum Gain	$ S_{11} $ at 2.4525 GHz	FTBR
1	One director	129°	5.98 dBi	-3.73 dB	21dB
2	Two consecutive directors	87°	7.70 dBi	-7.73 dB	9 dB
3	Three consecutive directors	76°	8.35 dBi	-11.25 dB	13 dB
4	Four consecutive directors	137°	6.42 dBi	-11.94 dB	25 dB
5	Five consecutive directors	103°	5.65 dBi	-8.39 dB	-
6	Two opposed directors	70°	5.91 dBi	-8.70 dB	-
7	Four opposed directors	58°	7.40 dBi	-12.20 dB	-
8	Six directors (all)	Omni	3.90 dBi	-5.81 dB	-

3.4. Four and Five Consecutive Directors (Confs. #4 and #5). The remaining configurations using consecutive directors, that is four and five (configurations #4 and #5, respectively), do not show improvements, in terms of maximum gain, over the two previous analyzed configurations, so these results are not plotted for the sake of brevity.

3.5. Two and Four Opposed Directors (Confs. #6 and #7). Configurations using opposing directors could be interesting to be assessed, since these can be used to radiate simultaneously power in two opposite directions, in order to minimize the broadcast transmissions in linear deployments. The configuration #7 uses four opposed director elements. The resulting radiation pattern in H plane is shown in Figure 5(a). We observe that the power is actually radiated in two opposite directions, each one with a gain of 7.40 dBi (1.42 dB higher than the reference design). The S_{11} parameter

is shown in Figure 5(b) presenting a value of -12.20 dB at 2.4525 GHz, about 8.47 dB lower than the reference design. The HPBW is 58°, being much narrower than the reference design.

The results obtained for two opposed directors (conf. #6) can be observed in Table 2; these are not better than the corresponding ones for conf. #7.

3.6. Six Directors (Conf. #8). A special case is the configuration with six director elements; obtaining an omnidirectional pattern with a gain of 3.9 dBi, it is 1.8 dBd, which shows an improvement in the radiation characteristics with respect to the dipole, being in this way a very good option as an omnidirectional antenna.

The use of a combination of directional antennas together with omnidirectional antennas has been proposed in some communication protocols [16–18], where the omnidirectional

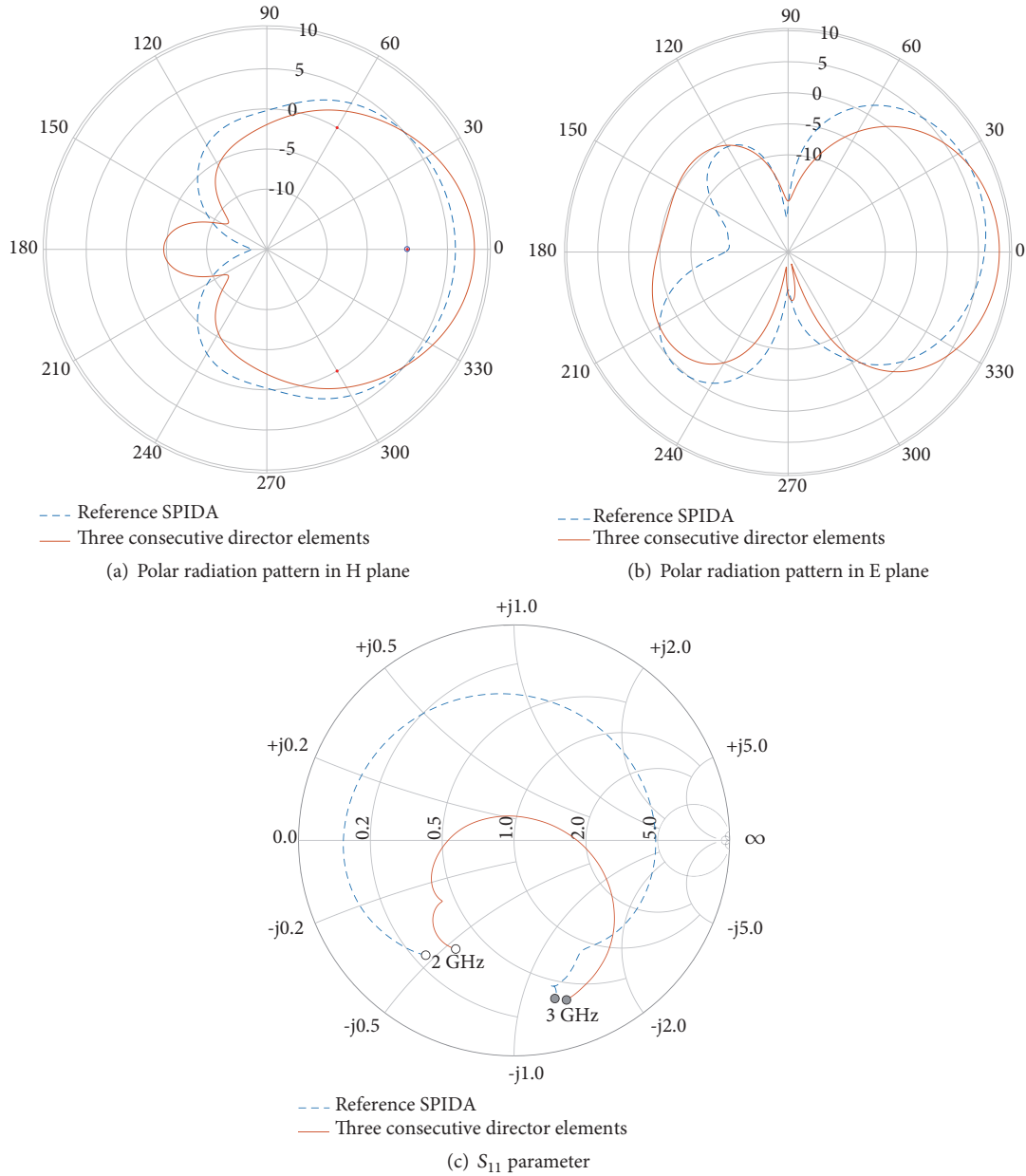


FIGURE 4: Three consecutive director elements (conf. #3).

antenna is adopted for broadcast messages. The configuration with six directors can be adopted for obtaining omnidirectionality.

Table 2 summarizes the main performance parameters of the simulation for the different configurations of the SPIDA. All configurations using consecutive directors outperform the reference design in terms of input impedance matching. The configurations with two, three, and five consecutive directors (confs. #2, #3 and #5) present narrower beamwidth, while the configuration with four directors (conf. #4) has a wider beamwidth. In terms of maximum gain confs. #2, #3, and #4 outperform the reference design. Considering the FTBR, the reference antenna performs similar to the one with four consecutive directors, but better than the

configurations with two and three consecutive director elements. Considering configurations with opposed directors, confs. #6 and #7 outperform the reference design in terms of input impedance matching and beam directivity in the desired directions, with conf. #7 having a higher maximum gain than the reference design and conf. #6 a lower one than it. These configurations have the potential to direct RF power to opposite directions with the benefit of enhanced performance. Also this antenna can be used as an omnidirectional one, with a better gain than a dipole. The final choice would depend on network design aspects, such as medium access protocols requirements and sensor nodes arrangement. Moreover, the increased RF power delivered to the air due to better impedance matching (achieved in all the

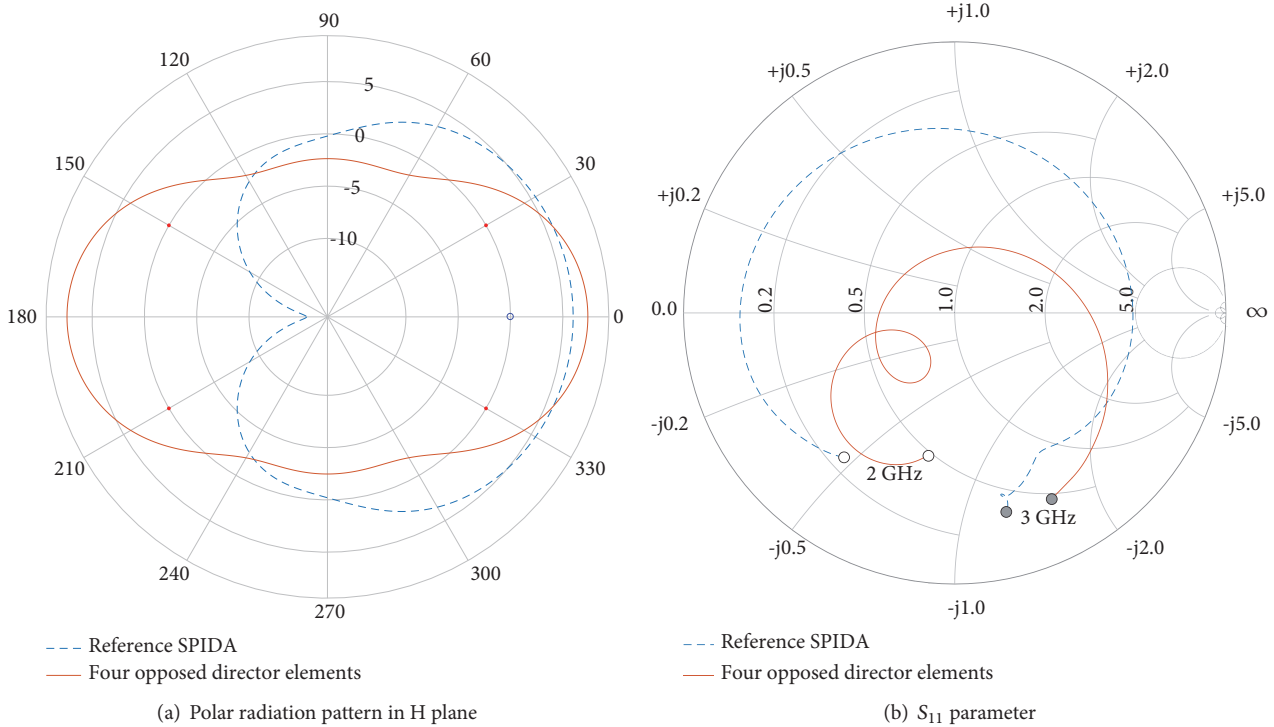


FIGURE 5: Four opposed director elements (conf. #7).

cases) represents an important improvement to the energy efficiency of the system.

Measured and simulated values of S_{11} for the more interesting cases are shown in Figures 3(b), 4(c), 5(b), and 7. If we take the criterion for the impedance bandwidth of considering a VSWR lower than two (module of S_{11} in dB lower than -9.54 dB), the configuration with three director elements presents an impedance bandwidth of 310 MHz, from 2180 MHz to 2490 MHz. For the band used in the IEEE 802.15.4 standard (from 2400 MHz to 2483.5 MHz), the module of S_{11} in dB for this antenna configuration is always lower than -9.7 dB, satisfying the selected criterion. During the development of this research, it was observed that the gain, directivity, and impedance bandwidth of the reference antenna could be modified (and improved) by changing its geometry, in particular changing the distance between the parasitic and the active elements. But it is important to have in mind that the number of director elements in use in this antenna is selected dynamically, and any geometrical optimization of this antenna has a different impact in each of the possible configurations.

4. Fabrication and Characterization

From the simulation results analysis, it turns out that one very promising configuration to improve the performance of the reference design is using three directors. This particular configuration is quite useful for our general research in the WSN area, so we decided to measure the performance of this configuration.

4.1. Fabrication. Two antennas were fabricated in a fixed configuration: one with one director and another with three directors. The first antenna is used as a reference for comparison with the second configuration in which the SPIDA antenna uses multiple director elements. Both antennas were built following the dimensions provided by [5, 6] and using six “legs” and six parasitic elements.

The elements were made using copper wire of 1 mm^2 of section (the dielectric shield was removed). A central PCB hexagon was used to fix the legs and connect them to ground. This hexagon was made of standard two-layer 1.6 mm FR4 PCB board of $35 \mu\text{m}$ of copper thickness. Both copper layers of the hexagon were connected using vias of 1 mm^2 of section, welded with tin (the via placement can be seen in Figure 6). A SMA connector was welded to the lower copper layer of the hexagon to feed the antenna through it. The active element of the antenna (the central element) was connected through the SMA connector to the central wire of the coaxial cable used to feed the antenna. A coaxial cable with SMA connectors was used to feed the antenna.

The hexagon was designed using CadSoft Eagle PCB Design Software and fabricated with a LKPF ProtoMat S63 circuit board plotter. The circuit board plotter features a resolution of $0.5 \mu\text{m}$ and an accuracy of $\pm 0.02 \text{ mm}$, allowing a very precise fabrication. This equipment enables the production of identical hexagons for the fabrication of these antennas, which facilitates the fabrication process repetitiveness.

The parasitic elements defined as directors were glued with silicone and the parasitic elements defined as reflectors were welded to its corresponding “legs” to ground (which

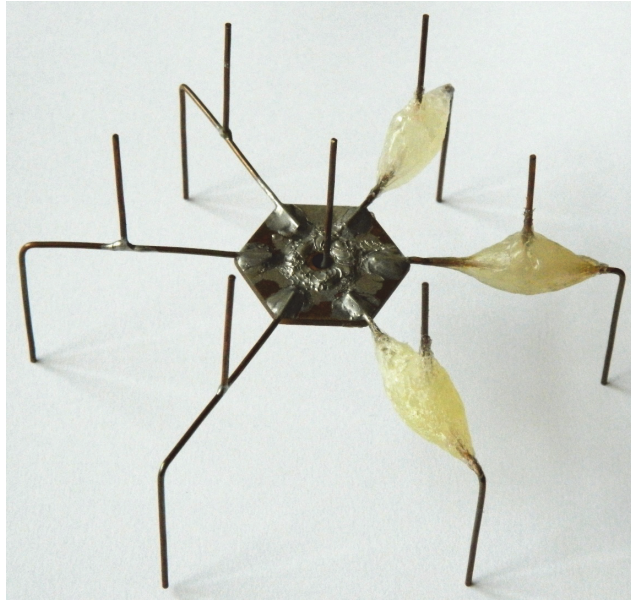


FIGURE 6: SPIDA antenna with three consecutive director elements.

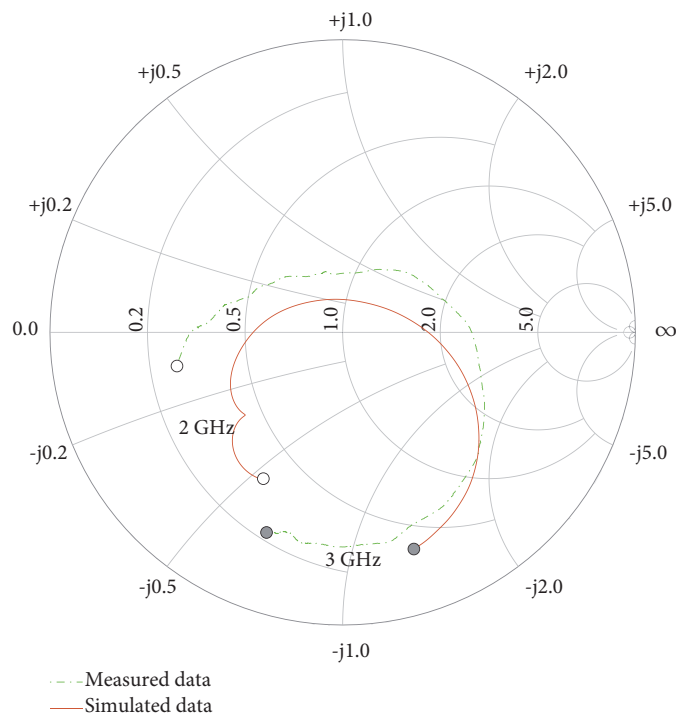


FIGURE 7: S_{11} parameter for the fabricated SPIDA antenna with three consecutive director elements.

are connected to the hexagon and through it to the cable shield). In a future work, switches will be inserted between the parasitic elements and its corresponding “leg” in order to have dynamic beamforming (controlled by their switches which are able to connect the parasitic element to the “leg” or not). According to chip manufacturer the typical switch attenuation is less than 1.6 dB.

Figure 6 shows a photography of the fabricated antenna with three directors.

4.2. Measurements and Results. For the characterization process a vectorial network analyzer (Rohde & Schwarz ZVB 8 Vector Network Analyzer, 300 kHz - 8 GHz), a RF generator (Agilent, E4438C, 250 kHz - 3 GHz, ESG Vector Signal Generator), and a spectrum analyzer (Agilent Technologies, EXA Signal Analyzer, N9010 A, 9 kHz - 7 GHz) were used.

During the antenna characterization the effort was concentrated on S_{11} parameter and the radiation pattern in the H and E planes for the case with three director elements. The

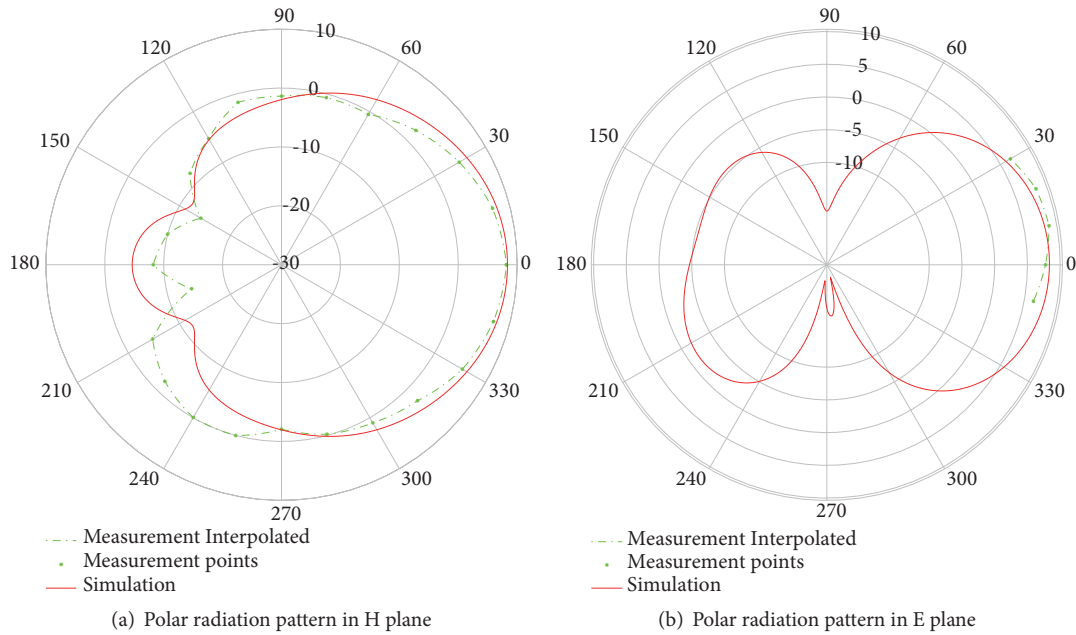


FIGURE 8: Fabricated SPIDA antenna with three consecutive director elements.

measurements obtained can be, respectively, seen in Figures 7 and 8. In these figures the measurements and the simulations are superimposed showing a good correspondence. The E plane was only measured in five points near 0° , as this is the area of interest for the WSN applications that we are considering, where all nodes are placed almost in the same horizontal plane.

For the SPIDA antenna with three director elements, a maximum gain of 8.2 dBi according to the measurements (8.35 dBi according to the simulations) was obtained, being 1.4 dB higher than the gain for the single director antenna taken as reference. A maximum gain of 8.2 dBi is a very good result, being a better gain than some previously reported results for similar antennas (e.g., in [6] (4.3 dBi), in [9] (5.1 dBi), and in [8] (8.08 dBi)).

According to the measurements, the HPBW for this antenna is 59° (76° according to the simulations) against 113° for the reference antenna (129° according to simulations).

The module of the S_{11} parameter according to the measurements for this antenna was -9.8 dB (-11.25 dB according to the simulations) against -7.5 dB for the reference antenna (-3.73 dB according to the simulations).

All these results show a very important improvement compared with the single director SPIDA antenna, which justify the use of three director elements instead of only one for this kind of antennas.

5. Conclusion

In this paper the advantages of using multiple director elements were discussed. An improved radiation pattern was obtained in this way, having an increase of the maximum gain of approximately 1.4 dB, 6.8 dBi for the single director element antenna against 8.2 dBi for the antenna with three

director elements. Also the input impedance matching was improved having a module of S_{11} parameter of -9.8 dB at the central frequency ($f_c = 2.4525$ GHz) for the three directors antenna against -7.5 dB for the antenna with a single director. By considering multiple director elements it was also shown that the flexibility in the beam orientation can be duplicated (having twelve beam directions instead of only six). Also it was shown that the use of multiple director elements can be very useful for specific situations as broadcasting where omnidirectional radiation patterns are generally better.

Once the director elements are controlled dynamically by using switches, then a very flexible beamforming scheme is obtained which can improve the performance of a wireless sensor network significantly.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Fondo María Viñas for supporting this project (FMV_1_2014_1_104872). The authors also thank Rodrigo Enjuu (from CST) for his cooperation with the simulation tasks and Professor Thiemo Voigt for discussions on the use of directional antennas for WSNs.

References

- [1] O. Chughtai, N. Badruddin, M. Rehan, and A. Khan, "Congestion detection and alleviation in multihop wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.
- [2] D. Curiac, "Wireless sensor network security enhancement using directional antennas: state of the art and research challenges," *Sensors*, vol. 16, no. 4, p. 488, 2016.
- [3] V. V. Khairnar, B. V. Kadam, C. K. Ramesha, and L. J. Gudino, "A reconfigurable parasitic antenna with continuous beam scanning capability in H-plane," *AEÜ - International Journal of Electronics and Communications*, vol. 88, pp. 78–86, 2018.
- [4] N. Tiwari and T. R. Rao, "A switched beam antenna array with butler matrix network using substrate integrated waveguide technology for 60 GHz wireless communications," *AEÜ - International Journal of Electronics and Communications*, vol. 70, no. 6, pp. 850–856, 2016.
- [5] E. Öström, L. Mottola, and T. Voigt, "Evaluation of an electronically switched directional antenna for real-world low-power wireless networks," *REALWSN 2010, LNCS 6511*, pp. 113–125, 2010.
- [6] M. Nilsson, "Directional antennas for wireless sensor networks," in *Proceedings of the 9th Scandinavian Workshop on Wireless Adhoc Networks (Adhoc'09)*, 2009.
- [7] B. Rodríguez, J. Schandy, J. P. González, L. Steinfeld, and F. Silveira, "Fabrication and characterization of a directional SPIDA antenna for wireless sensor networks," in *Proceedings of the 2017 IEEE URUCON, URUCON 2017*, pp. 1–4, Uruguay, October 2017.
- [8] R. Schlub, J. Lu, and T. Ohira, "Seven-element ground skirt monopole ESPAR antenna design from a genetic algorithm and the finite element method," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 11, pp. 3033–3039, 2003.
- [9] J. Lu, D. Ireland, and R. Schlub, "Dielectric embedded ESPAR (DE-ESPAR) antenna array for wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 8 I, pp. 2437–2443, 2005.
- [10] A. Kalis, A. G. Kanatas, and C. B. Papadias, *Parasitic Antenna Arrays for Wireless MIMO Systems*, Springer New York, New York, NY, USA, 2014.
- [11] L. Selavo and O. Chipara, "Directional antenna platform for low power wireless networks," in *Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks*, pp. 258–259, Junction Publishing, 2017.
- [12] A. Varshney, T. Voigt, and L. Mottola, "Using directional transmissions and receptions to reduce contention in wireless sensor networks," in *Real-World Wireless Sensor Networks*, vol. 281, pp. 205–213, Springer International Publishing, 2014.
- [13] L. Mottola, T. Voigt, and G. P. Picco, "Electronically-switched directional antennas for wireless sensor networks: A full-stack evaluation," in *Proceedings of the 2013 10th Annual IEEE Communications Society Conference on Sensing and Communication in Wireless Networks, SECON 2013*, pp. 176–184, USA, June 2013.
- [14] B. S. Geletu, L. Mottola, T. Voigt, and F. Österlind, "Modeling an electronically switchable directional antenna for low-power wireless networks," in *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN'11*, pp. 163–164, USA, April 2011.
- [15] Y. Jiang, H. Zhang, B. Zhao, and S. Rangarajan, "Optimizing multicast delay with switched beamforming in wireless networks," in *Proceedings of the 2011 IEEE International Conference on Communications*, pp. 1–6, Kyoto, Japan, June 2011.
- [16] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins, and S. Polit, "Ad hoc networking with directional antennas: a complete system solution," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 496–506, 2005.
- [17] R. Santosa, B. Lee, C. Yeo, and T. Lim, "Distributed neighbor discovery in ad hoc networks using directional antennas," in *Proceedings of the The Sixth IEEE International Conference on Computer and Information Technology (CIT'06)*, pp. 97–97, Seoul, September 2006.
- [18] S. Zhang and A. Datta, "A directional-antenna based MAC protocol for wireless sensor networks," in *Proceedings of the Computational Science and Its Applications, ICCSA*, pp. 551–586, 2005.

Research Article

Online Supervisory Control and Resource Management for Energy Harvesting BS Sites Empowered with Computation Capabilities

Thembelihle Dlamini ^{1,2}, Ángel Fernández Gambín ¹,
Daniele Munaretto,² and Michele Rossi¹

¹Department of Information Engineering, University of Padova, Padova, Italy

²Athonet, Bolzano Vicentino, Vicenza, Italy

Correspondence should be addressed to Thembelihle Dlamini; dlamini@dei.unipd.it

Received 24 October 2018; Accepted 3 February 2019; Published 19 February 2019

Academic Editor: Gianluigi Ferrari

Copyright © 2019 Thembelihle Dlamini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The convergence of communication and computing has led to the emergence of *multi-access edge computing (MEC)*, where computing resources (supported by virtual machines (VMs)) are distributed at the edge of the mobile network (MN), i.e., in base stations (BSs), with the aim of ensuring reliable and ultra-low latency services. Moreover, BSs equipped with *energy harvesting (EH)* systems can decrease the amount of energy drained from the power grid resulting into energetically self-sufficient MNs. The combination of these paradigms is considered here. Specifically, we propose an online optimization algorithm, called Energy Aware and Adaptive Management (ENAAM), based on foresighted control policies exploiting (short-term) traffic load and harvested energy forecasts, where BSs and VMs are dynamically switched on/off towards energy savings and Quality of Service (QoS) provisioning. Our numerical results reveal that ENAAM achieves energy savings with respect to the case where no energy management is applied, ranging from 57% to 69%. Moreover, the extension of ENAAM within a cluster of BSs provides a further gain ranging from 9% to 16% in energy savings with respect to the optimization performed in isolation for each BS.

1. Introduction

The full potential of 5G radio access technology can be realized through the use of distributed intelligence, whereby content, control, and computation are moved closer to mobile users, hereby referred to as the *network edge*. This evolution has led to the emergence of the multi-access edge computing (MEC) paradigm, which allows network functions to be virtualized and then deployed at the network edge to guarantee the low latency required by some applications. In this paper, we consider a hybrid edge computing architecture where computing servers are co-located with each base station (BS), and a centralized controller (a point within range to a set of BSs) is utilized to manage them, deciding upon the allocation of their computing and transmission resources. This type of architecture is in line with recent trends [1].

The convergence of communication and computing (MEC [2]) within the mobile space poses new challenges related to energy consumption, as BSs are densely deployed to maximize capacity and also empowered with computing capabilities to minimize latency. To cope with these challenges, previous studies have put forward BS sleep modes [3, 4], as BSs are dimensioned for the expected maximum capacity, yet traffic varies during the day. In addition, energy savings within the virtualized computing platform are of great importance, as virtualization can also lead to energy overheads. Therefore, a clear understanding and a precise modeling of the server energy usage can provide a fundamental basis for server operational optimizations. The experimental results in [5, 6] show that the locus of energy consumption for the Virtualized Network Function (VNF) components is the virtual machine (VM) instance where the VNF is instantiated and executed. Thus, for a given expected

traffic load, the energy consumption can be minimized by launching an optimal number of VMs, a technique referred to as VM *soft-scaling*, together with BS power saving methods, i.e., BS sleep modes.

Along these lines, we propose a controller-based network architecture for managing energy harvesting (EH) BSs empowered with computation capabilities where on/off switching strategies allow BSs and VMs to be dynamically switched on/off, depending on the traffic load and the harvested energy forecast, over a given look-ahead prediction horizon. To solve the energy consumption minimization problem in a distributed manner, the *controller* partitions the BSs into clusters based on their location; then, for each cluster, it minimizes a cost function capturing the individual communication site energy consumption and the users' Quality of Service (QoS). To manage the communication sites, the controller performs online supervisory control by forecasting the traffic load and the harvested energy using a Long Short-Term Memory (LSTM) neural network [7], which is utilized within a Limited Look-ahead Control (LLC) policy (a predictive control approach [8]) to obtain the system control actions that yield the desired trade-off between energy consumption and QoS. This work is an extension of [9], where we consider energy savings within a single off-grid BS scenario (i.e., BS powered by either wind or solar energy sources) taking into account the need for MEC in remote/rural areas. In this paper, however, a dense environment is considered, similar to an urban or semi-urban scenario, where each BS is powered by hybrid energy supplies (solar and power grid) and empowered with computation capabilities. Moreover, the optimization problem is extended for multiple BSs where energy management procedures are executed within a BS cluster in contrast with the single BS case of [9].

The rest of the paper is structured as follows. The related work is discussed in Section 2, and the system model is presented in Section 3. In Section 4, we detail the optimization problem and the proposed LLC-based online algorithm for a *single* communication site. The *multiple* BS communication site case is addressed in Section 5. Our contribution is evaluated in Section 6, and, lastly, concluding remarks are given in Section 7.

2. Related Work and Paper Contribution

Next, we first provide a literature review related to BS sleep modes techniques. Then, we review the mathematical tools that we use in this paper, followed by the literature review related to energy savings in virtualized computing platforms (i.e., works related to soft-scaling). Finally, we put forward our contributions and novelty of our work.

Sleep-Mode Strategies in Mobile Networks. Cellular networks are dimensioned to support traffic peaks; i.e., the number of BSs deployed in a given area should be able to provide the required QoS to the mobile subscribers during the highest load conditions. However, during off-peak periods the network may be underutilized, which leads to an inefficient use of network resources and to excessive

energy consumption. For these reasons, sleep modes have been proposed to dynamically turn off some of the BSs when the traffic load is low. This has been extensively studied in the literature; here we highlight the main applied techniques that are related to this work.

Clustering algorithms have been proposed as a way of switching off BSs to reduce the energy consumption. In [12], centralized and distributed algorithms group BSs exhibiting similar traffic profiles over time. In [13], a dynamic switching on/off mechanism locally groups BSs into clusters based on location and traffic load. The optimization problem is formulated as a non-cooperative game aimed at minimizing the BS energy consumption and the time required to serve their traffic load. Simulation results show energy costs and load reductions, while also providing insights of when and how the cluster-based coordination is beneficial.

Reducing the energy consumption involves some trade-offs in the optimization problem. QoS has been widely used as a trade-off metric [14, 15]. The Quality of Experience (QoE) is included in [16], where a dynamic programming switching algorithm is put forward. Other parameters that have been considered are the coverage probability and the BS state stability parameter, i.e., the number of on/sleep state transitions. For instance, a set of BSs switching patterns engineered to provide full network coverage at all times, while avoiding channel outage, is presented in [17]. According to the BS state stability concept, a two-objective optimization problem is formulated in [18] and solved with two algorithms: (i) near optimal but not scalable and (ii) low complexity, based on particle swarm optimization. The QoE is also affected by the UE position due to channel propagation phenomena. To this respect, in [19] the selection of the BSs to be switched off is taken so as to minimize the impact on the UEs' QoE, according to the distance from the handed off BSs.

To support sleep modes, neighboring cells must be capable of serving the traffic from the switched off cells. To achieve this, proper *user association* strategies are required. A framework to characterize the performance (outage probability and spectral efficiency) of cellular systems with sleeping techniques and user association rules is proposed in [20]. In that paper, the authors devise a user association scheme where a user selects its serving BS considering the maximum expected channel access probability. This strategy is compared against the traditional maximum SINR-based user association approach and is found superior in terms of spectral efficiency when the traffic load is inhomogeneous. User association mechanisms that maximize energy efficiency in the presence of sleep modes are addressed in [21]. There, a downlink HetNet scenario is considered, where the energy efficiency is defined as the ratio between the network throughput and the total energy consumption. Since this leads to a rather complex integer optimization problem, the authors propose a quantum particle swarm optimization algorithm to obtain a suboptimal solution.

A marketing approach to foster the opportunistic utilization of the unexploited small cell (SC) BS capacity in dense heterogeneous networks (HetNets) is presented in [22]. There, an offloading mechanism is introduced, where the

operators lease the capacity of a SC network owned by a third party in order to switch off their BSs (macro-BSs) and maximize their energy efficiency, when the traffic demand is low. The allocation of the SC resources among a set of competing operators is mathematically formulated as an auction problem.

A comprehensive power management model employing a BS switching on/off mechanism, within a BS system powered by green energy, is presented in [23]. The model considers weather conditions, user mobility, different green energy harvesting rates, energy storage with self-discharge effect, and switching on/off frequency. The authors propose two algorithms: the first decides which BSs are to be active based on the minimum energy cost, i.e., the energy price per time period, while the second one determines the active BSs by first prioritizing the minimum power consumption of the system and then the energy cost. The relationship between installing a solar harvesting system to power a BS and the energy management under varying demand is investigated in [24]. The authors present a solar installation planning model by explicitly modeling solar panels, batteries, inverters, and charge controllers, as well as the cellular network demand and energy management. They found that the solar installation and the energy management of the base stations are so coupled that even the order in which these technologies are introduced can have a major impact on the network cost and performance.

The survey paper [25] presents taxonomy of existing energy sustainable paradigms and methods to address energy savings in network elements (i.e., BSs) equipped with EH capabilities. Here, the authors discuss the shortcomings of previous studies related to efficient energy management procedures, the lack of relevant discussion related to the integration of EH into future networks, and, lastly, energy self-sustainability in future networks. The current work is a technical contribution where we address some of the shortcomings that were identified in [25], also proposing the use of machine learning (ML) tools for pattern forecasting and adaptive control schemes for decision-making. In addition, this work is in line with the research topics which can be found in our review paper [26].

The majority of the works on BS switching off mechanism considered clusters of BSs from a *single* mobile operator perspective, where some functions of the BS can be switched off and then the remaining active BSs handle the upcoming traffic. A new approach is presented in [27] which exploits the coexistence of multiple BSs from different mobile operators in the same area. An intracell roaming-based infrastructure-sharing strategy is proposed, followed by a distributed game-theoretic switching off scheme that takes into account the conflicts and interaction among the different operators. Moreover, in [28], the authors investigate the energy and cost efficiency of multiple HetNets (i.e., each HetNet is composed of eNodeBs (eNBs) and SC BSs from one operator) that share their infrastructure and also are able to switch off part of it. Here, a form of roaming-based sharing is also adopted, whereby the operator can roam its traffic to a rival operator during a predefined period of time and area. An energy efficient optimization problem is formulated and solved using

a cooperative greedy heuristic algorithm. Regarding the cost efficiency, the cooperation and cost sharing decisions among the operators are modeled using a Shapley Value based bankruptcy game.

Pattern Forecasting along with Foresighted Optimization. Control-theoretic and machine learning (ML) methods for resource management have been successfully applied to various problems, e.g., task scheduling, bandwidth allocation, and network management policies. In the paradigm of supervisory control for managing mobile networks (MNs), online forecasting using ML techniques and the LLC method can yield the desired system behavior when taking into account the environmental expectations, i.e., traffic load and energy to be harvested. Next, we briefly review the mathematical tools that we use in this paper, namely, the LLC method and LSTM neural network [7].

Control-theoretic algorithms and the LLC method have been used to obtain control actions that optimize the system behavior, by employing a forecasting mathematical model, over a limited look-ahead prediction horizon. LLC is conceptually similar to model predictive control (MPC) [29]. In [30], an online supervisory control scheme based on LLC policies is proposed. Here, after the occurrence of an event, the next control action is determined by estimating the system behavior a few steps into the future, using the currently available information as inputs. The control action exploration is performed using a search tree assuming that the controller knows all future possible states of the process over the prediction horizon. Moreover, in [8], an online control framework for resource management in switching hybrid systems is proposed, where the system's control inputs are finite. The relevant parameters of the operating environment, e.g., workload arrival, are estimated and then used by the system to forecast future behavior over a look-ahead horizon. From this, the controller optimizes the predicted system behavior following the specified QoS through the selection of the system controls.

To model time series datasets, the LSTM network is used as it is able to handle the long-term dependencies due to its inherent capability of storing past information and then recalling it. In [31], a distributed LSTM online method based on the particle filtering algorithm is presented with an aim of investigating the performance of online training of LSTM architectures in a distributed network of nodes. An LSTM based model for variable length data regression is proposed and then put into a nonlinear state-space form to train the model in an online fashion. Then, financial and real life datasets are used for performance evaluation, and it is observed that the distributed online approach yields the same results that are obtained in the centralized case, when considering the mean square errors as the performance measure. Moreover, an LSTM forecasting method is utilized in [9] within an LLC-based algorithm to obtain the system control actions yielding the desired trade-off between energy consumption and QoS, for a remote site powered by only green energy.

Energy Savings in Virtualized Platforms through Soft-Scaling. With the advent of virtualization, it is expected that the Network Function Virtualization (NFV) framework can

exploit the benefits of virtualization technologies to significantly reduce the energy consumption of large scale network infrastructures. In virtualized computing environments, the locus of energy consumption for components is due to the VMs running in the server(s). Thus, energy saving studies within the virtualized computing environment have involved the scaling down of the number of computing nodes/servers (autoscaling [32]), VM migration [33] (movement of a VM from one host to another), and soft resource scaling [34] (shortening of the access time to physical resources), all hereby referred to as VM *soft-scaling*, i.e., the reduction of computing resources per time instance.

Algorithms for the dynamic on/off switching of servers have been proposed as a way of minimizing energy consumption in computing platforms. In [32], at the beginning of each time slot computing resources are provisioned depending on the expected server workloads via a reinforcement learning-based resource management algorithm, which learns on-the-fly the optimal policy for dynamic workload offloading and the autoscaling of servers. Then in [9], computing resources (VMs) are provisioned based on a LLC policy after forecasting the future workloads and harvested energy. In [33], the Central Processing Unit (CPU) utilization thresholds are used to identify overutilized servers. Hence, migration policies, enabled by the live VM migration method [35], are applied for moving the VMs between physical nodes (servers). The VMs are only moved to hosts that will accept them without incurring high energy cost, i.e., without any increase in the CPU utilization. Subsequently, the idle servers are switched off.

Power management is also of interest in virtualized computing platforms, i.e., data centers using virtualization technologies. In [34], a power management approach called *VirtualPower* is presented. The algorithm exploits hardware power scaling, i.e., the dynamic power management strategies using Dynamic Voltage and Frequency Scaling (DVFS) [36, 37], and software-based methods, i.e., scaling the allocation of physical resources to VMs using the hypervisor scheduler, for controlling the power consumption of underlying platforms. Due to the low power management benefits obtained from hardware scaling, a *soft resource scaling* mechanism is proposed whereby the scheduler shortens the maximum resource usage time for each VM, i.e., the time slice allocated for using the underlying physical resources.

Novelty of this Work. Here, we consider the aforementioned scenario, where each BS is equipped with EH hardware (a solar panel for EH and an Energy Buffer (EB) for energy storage) and a MEC server co-located with the BS for computation purposes, under the management enabled by the *controller*.

Motivated by the potential capabilities of EH and MEC and the presence of the controller,

- (1) we introduce the use of virtualization with the aim of investigating how VMs can be soft-scaled based on the forecasted server workloads, as VMs are the source of energy consumption in computing environments;

- (2) we put forward the edge controller-based architecture for small cell BSs management, as one of the future trends for small cells [1] in 5G MNs;
- (3) we reconsider the BS sleeping control mechanism under the new MEC paradigm, which has not been sufficiently covered in the literature. In addition, we use a clustering method for enabling energy savings within the MN;
- (4) we estimate the short-term future traffic load and harvested energy in BSs, by using LSTM neural network [38];
- (5) we develop an online supervisory control algorithm for the radio access (edge) network management based on a predictive method, specifically the LLC method, along with clustering and energy management procedures. The main goal is to enable energy savings (ES) strategies within the access network, BS sleep modes, and VM soft-scaling, following the energy efficiency requirements of a virtualized infrastructure from [39]. The proposed management algorithm is called Energy Aware and Adaptive Management (ENAAM) and is hosted in the edge controller. The ENAAM algorithm considers the future BS traffic load, onsite green energy in the EB, and then provisions access network resources, per communication site, based on the learned information; i.e., energy saving decisions are made in a forward-looking fashion.

The proposed optimization strategy leads to a considerable reduction in the energy consumed by the edge computing and communication facilities, promoting self-sustainability within the mobile network through the use of green energy. This is achieved under the controller guidance, which makes use of forecasting, clustering, control theory, and heuristics.

3. System Model

As a major deployment of MEC and in line with current trends for future mobile networks as suggested by prominent network operators (e.g., Huawei Technologies [1]), the considered scenario is illustrated in Figure 1. It consists of a densely deployed MN featuring N BSs and colocated cache-enabled MEC servers. Each MEC server hosts M VMs. Each communication site, i.e., the BS and the colocated MEC server, is empowered with EH capabilities through a solar panel and an EB that enables energy storage. Energy supply from the power grid is also available. Moreover, the Energy Manager (EM) is an entity responsible for selecting the appropriate energy source and for monitoring the energy level of the EB. All BSs communicate with a centralized entity called the *edge controller*, which is responsible for managing the access network apparatuses. The energy level information is reported periodically to the edge controller through the pull file transfer mode procedure (e.g., File Transfer Protocol [40]). Moreover, we consider a discrete-time model, whereby time is discretized as $t = 1, 2, \dots$, and each time slot t has a

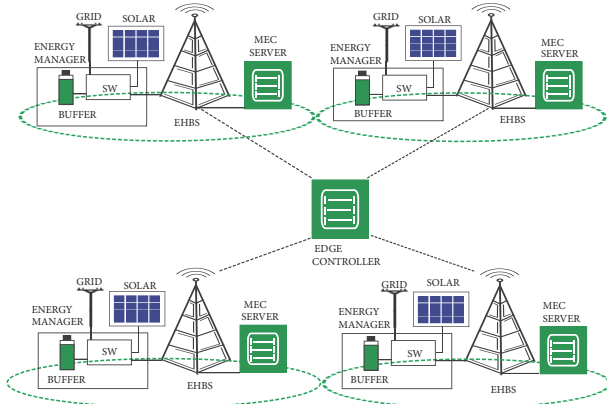


FIGURE 1: Edge network topology. The electromechanical switch (SW) selects the appropriate source of energy.

fixed duration τ . The list of symbols that are used in the paper is reported in Table 1.

3.1. Traffic Load and Energy Consumption. Mobile traffic volume exhibits temporal and spatial diversity and also follows a diurnal behavior [41]. Therefore, traffic volume at individual BSs can be estimated using historical mobile traffic datasets. In this paper, real MN traffic load traces obtained from the Big Data Challenge organized by Telecom Italia Mobile (TIM) [10] are used to emulate the computational load (in fact, the dataset is not a true representative of future applications that require processing at the edge but contains data that is exchanged with the purpose of communication. We nevertheless use it due to the difficulties in finding open datasets containing computing requests). Specifically, the used data was collected in the city of Milan during the month of November 2013, and it is the result of users interaction within the TIM MN, based on Call Detail Record (CDR) files for a day considering four BS sites representing the traffic load profiles. A CDR file consists of SMS, Calls, and Internet records with timestamps. To understand the behavior of the mobile data, we have applied the X-means clustering algorithm [42] to classify the load profiles into several categories. In our numerical results, each BS $n = 1, 2, \dots, N$ is assigned a load profile $L_n(t)$, which is picked at random as one of the four clusters (each cluster represents a typical BS load profile) in Figure 2. $L_n(t)$ consists of computation workloads $\Gamma_n(t)$ ([MB]) and standard workloads $\Gamma'_n(t)$ ([MB]). According to [43], we assume that 80% of $L_n(t)$ is *delay sensitive* and, as such, requires processing at the edge, i.e., $\Gamma_n(t) = 0.8L_n(t)$, whereas the remaining 20% pertains to standard flows, *delay tolerant* traffic, i.e., $\Gamma'_n(t) = L_n(t) - \Gamma_n(t)$.

The total energy consumption ([J]) for the communication site n at time slot t is formulated as follows, inspired by [9, 44–47]

$$\theta_{\text{tot},n}(t) = \theta_{\text{BS},n}(t) + \theta_{\text{MEC},n}(t) + \theta_{\text{TX},n}(t), \quad (1)$$

where $\theta_{\text{BS},n}(t)$ is the BS energy consumption term, $\theta_{\text{MEC},n}(t)$ is the MEC server consumption term due to computation

TABLE 1: Notation: list of symbols used in the analysis.

Symbol	Description
Input Parameters	
N	number of BSs, indexed by n
M	maximum number of VMs hosted by each MEC server
τ	time slot duration
$L_n(t)$	BS n traffic load profile in time slot t , n is the BS index
$\Gamma_n(t)$	workload handled by the MEC server at BS n in time slot t
$\Gamma'_n(t)$	standard (non MEC) traffic at time t
θ_0	BS load independent energy consumption or operation energy
f_{max}	maximum processing rate for VM m
\mathcal{F}	a finite set of available processing rates for VM m
$\theta_m^{\text{ov}}(t)$	energy overheads incurred when turning on/off VMs
$\theta_{\text{idle},m}(t)$	static energy consumed by VM m in the idle state
$\theta_{\text{max},m}(t)$	maximum energy consumed by VM m at maximum processing rate
$\gamma_m(t)$	workload fraction to be computed by the m -th VM
γ^{max}	maximum computation load per-VM
Δ	maximum per-slot and per-VM allowed processing time
$\theta_{\text{idle}}(t)$	energy consumption of network interfaces in idle mode
$\theta_{\text{data}}(t)$	energy cost of exchanging one unit of data between the server and the BS
β_{max}	maximum energy buffer capacity
$\beta_{\text{up}}, \beta_{\text{low}}$	upper and lower energy buffer thresholds
Variables	
$\theta_{\text{tot},n}(t)$	total energy consumption for the communication site n
$\theta_{\text{BS},n}(t)$	BS n energy cost at t
$\theta_{\text{MEC},n}(t)$	server consumption due to computation activities
$\theta_{\text{TX},n}(t)$	data transmission energy consumption between the BS and the MEC server
$\zeta_n(t)$	BS n switching status indicator at t
$M(t)$	number of VMs to be active in time slot t
$\theta_{\text{load}}(t)$	total wireless transmission power
$f_m(t)$	instantaneous processing rate
$\theta_m^{\text{op}}(t)$	energy consumption of VM m operation
$\alpha_m(t)$	load dependent factor
$\mu_m(t)$	the expected processing time
$B_n(t)$	the total amount of load that is served by the BS site
$\beta_n(t)$	energy buffer level in slot t
$H_n(t)$	harvested energy profile in slot t
$Q_n(t)$	purchased grid energy in slot t

activities, and $\theta_{\text{TX},n}(t)$ represents the data transmission energy consumption between the BS and the MEC server.

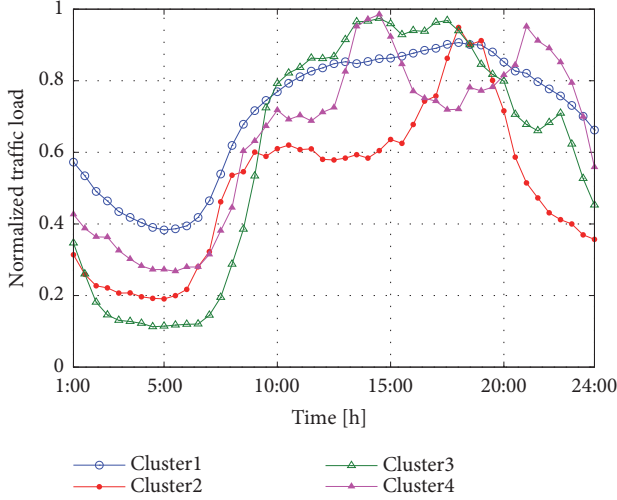


FIGURE 2: Example traces for normalized BS traffic loads. The data from [10] has been split into four representative clusters.

BS Energy Consumption. $\theta_{BS,n}(t) = \zeta_n(t)\theta_0 + \theta_{load}(t)$, where $\zeta_n(t) \in \{\varepsilon, 1\}$ is the BS switching status indicator (1 for *active mode* and ε for *power saving mode*) and θ_0 is a constant value (load independent), representing the operation energy which includes baseband processing, radio frequency power expenditures, etc. The constant $\varepsilon \in (0, 1)$ accounts for the fact that the baseband energy consumption can be scaled down as well whenever there is no or little channel activity, into a power saving mode. $\theta_{load}(t)$ represents the total wireless transmission (load dependent) power to meet the target transmission rate from the BS to the served user(s) and to guarantee low latency at the edge. Since we assume a noise-limited channel and the guarantee of low latency requirements at the edge, $\theta_{load}(t)$ is obtained by using the transmission model in [44] (see (5) in this reference). Here, we neglect the imbalance of traffic volumes in uplink and downlink, and also we do not account for the switching energy cost for the BS mode transition [46] due to the fact that future BS functions will be virtualized [48].

MEC Server Energy Consumption. It depends on the number of VMs running in time slot t , named, $M(t) \leq M$, and on the CPU frequency that is allotted to each virtual machine. Specifically, VMs are instantiated on top of the physical CPU cores, and each VM is given a share of the host server CPU, memory, and network input/output interfaces. The CPU is the main consumer of energy in the server [33] due to the VM-to-CPU share mapping. Hence, in this work we focus on the CPU utilization only. With $f_m(t) \in [0, f_{max}]$ we mean the instantaneous processing rate [49], expressed in bits per second that are computed, and f_{max} is the maximum processing rate for VM m . In this paper, $f_m(t)$ is set within a finite set $\mathcal{F} = \{f_0, f_1, \dots, f_{max}\}$ where $f_0 = 0$ represents zero speed of the VM (e.g., deep sleep or shutdown). At any given time t , the total energy consumption of a virtualized server, with $M(t)$ running VMs, is

$$\theta_{MEC,n}(t) = \sum_{m=1}^{M(t)} (\theta_m^{op}(t) + \theta_m^{ov}(t)), \quad (2)$$

where $\theta_m^{op}(t)$ is the energy consumption of VM m operation and $\theta_m^{ov}(t) \geq 0$ is the energy cost incurred through the turning on/off the VM; i.e., $\theta_m^{ov}(t) > 0$ only when VM m is switched on/off and it is zero otherwise. $\theta_m^{op}(t)$ is obtained using the linear relationship between the CPU utilization contributed by VM m and the energy consumption, from [49, 50] (see (4) in the second reference):

$$\theta_m^{op}(t) = \theta_{idle,m}(t) + \alpha_m(t) (\theta_{max,m}(t) - \theta_{idle,m}(t)), \quad (3)$$

where $\theta_{idle,m}(t)$ represents the *static* energy drained by VM m in the idle state, and $\theta_{max,m}(t)$ is the *maximum* energy it drains. The quantity, $\alpha_m(t)(\theta_{max,m}(t) - \theta_{idle,m}(t))$, represents the *dynamic* energy component, where $\alpha_m(t) = (f_m(t)/f_{max})^2$ [8] is a load dependent factor. Note that $\alpha_m(t)$ and $f_m(t)$ are deterministically related as f_{max} is a constant. $\theta_m^{ov}(t)$ is obtained from [50] (see (5) in this reference) as a constant and is typically limited to a few hundreds of mJ per MHz².

Conventionally, for each BS site, the hypervisor, i.e., the software that provides the environment in which the VMs operate, is in charge of allocating $f_m(t)$ and the workload fraction to be computed by the m -th VM, named $\gamma_m(t)$. In our setup, we have $\sum_{m=1}^{M(t)} \gamma_m(t) \leq \Gamma_n(t)$, where equality is achieved when the workload is fully served by $M(t)$ VMs. We also note that, in practical application scenarios, the maximum per-VM computation load to be computed is generally limited up to an assigned value, named γ^{max} . Motivated by the energy efficient requirements from [39], i.e., the hypervisor's ability to accept and implement policies from a management entity, in this paper, the *edge controller* usage is pursued. Here, the edge controller determines the $f_m(t)$ value that will yield the desired or expected processing time, $\mu_m(t) = \gamma_m(t)/f_m(t)$, considering the workload $\gamma_m(t)$ allotted to VM m . $\mu_m(t)$ must be less than or equal to the maximum per-slot and per-VM processing time (in seconds), named, Δ ; i.e., $\mu_m(t) \leq \Delta$. Note that Δ is also the server's response time, i.e., the maximum time allowed for processing the total computation load.

We remark that, as a result of the allocation procedure that is developed in this paper, for any BS site n , the processing rates $f_m(t)$ shall be found, similar to [50] (see Remark 1 from this reference). Then, the total amount of load that is served by the BS site may be set as follows: $B_n(t) = \sum_{m=1}^{M(t)} \gamma_m(t) \leq \Gamma_n(t)$. The objective of the considered optimization is to find the operating mode for the BS (either "on" or "power saving"), the number of VMs $M(t)$ that are to be allocated and, for each of them, the processing rate $f_m(t)$. In doing so, (1) the amount of delay sensitive load that is not served at the edge, $\Gamma_n(t) - \sum_{m=1}^{M(t)} \gamma_m(t)$, shall be minimized, while exploiting as much as possible the energy harvested from the solar panels, so that the mobile network will be energetically self-sufficient and (2) the load is computed in a time shorter than or equal to Δ . The details of the proposed optimization algorithm are provided in Section 4.

Data Transmission Energy Consumption. We assume that the intercommunication between the BS and the MEC server is bidirectional and symmetric. Hence, under steady-state operating conditions, for the communication site n , $\theta_{TX,n}(t)$ is obtained as $\theta_{TX,n}(t) = \theta_{idle}(t) + \theta_{data}(t)B_n(t)$ by using the

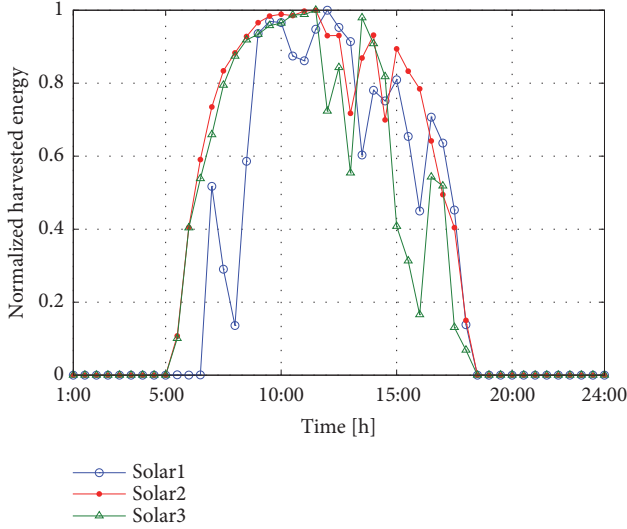


FIGURE 3: Example traces for harvested solar energy from [11].

VM migration hint from [51], where $\theta_{\text{idle}}(t)$ (fixed value in J) is the energy drained by the network interfaces in idle mode over a time slot t , θ_{data} (fixed value in J/byte) is the cost of exchanging one byte of data between the MEC server and the BS per time slot t , and $B_n(t)$ is the amount of data exchanged. These parameters, $\theta_{\text{idle}}(t)$ and $\theta_{\text{data}}(t)$, are obtained from [51]. Note that $B_n(t)$ also corresponds to the amount of data to be processed at the MEC server in bytes.

3.2. Energy Patterns and Storage. The energy buffer is characterized by its maximum energy storage capacity β_{max} . At the beginning of each time slot t , the EM provides the energy level report to the edge controller through the local MEC server; thus the EB level $\beta_n(t)$ is known, enabling the provision of the required computation resources, i.e., the VMs. The energy level report/file from the EM to the MEC server is transferred using the pull mode procedure (e.g., File Transfer Protocol) [40].

In this work, the amount of harvested energy $H_n(t)$ in time slot t in the communication site n is obtained from open source solar traces [11] (see Figure 3). The dataset is the result of daily environmental records. In our numerical results, $H_n(t)$ represents a daily solar radiation record for three different areas. From the three solar profiles, each communication site energy profile is picked at random to represent the daily energy harvested and then scaled to fit the EB capacity β_{max} of 490 kJ. Thus, the available EB level $\beta_n(t+1)$ at the beginning of time slot $t+1$ is calculated as follows:

$$\beta_n(t+1) = \beta_n(t) + H_n(t) - \theta_{\text{tot},n}(t) + Q_n(t), \quad (4)$$

where $\beta_n(t)$ is the energy level in the battery at the beginning of time slot t , $\theta_{\text{tot},n}(t)$ is the energy consumption of the communication site over time slot t (see (1)), and $Q_n(t) \geq 0$ is the amount of energy purchased from the power grid. We remark that $\beta_n(t)$ is updated at the beginning of time slot t whereas $H_n(t)$ and $\theta_{\text{tot},n}(t)$ are only known at the end of it.

For decision-making in the edge controller, the received EB level reports are compared with the following thresholds: β_{low} and β_{up} , respectively termed the lower and the upper energy threshold with $0 < \beta_{\text{low}} < \beta_{\text{up}} < \beta_{\text{max}}$. β_{up} corresponds to the desired energy buffer level at the BS and β_{low} is the lowest EB level that any BS should ever reach. If $\beta_n(t) < \beta_{\text{low}}$, then BS n is said to be *energy deficient* and our optimization in the following section makes sure that $\beta_n(t)$ never falls below β_{low} due to its transmission and computing activities within a time slot. Instead, if for any time slot we have $\beta_n(t) < \beta_{\text{up}}$, then the following amount of energy $Q_n(t) = \beta_{\text{up}} - \beta_n(t)$ is purchased from the energy grid to compensate for the deviation from the desired EB level (due to previous BS activity).

4. Optimization for a Single Communication Site

In this section, we formulate an optimization problem to obtain *energy savings* through short-term traffic load and harvested energy predictions, along with energy management procedures for a *single* communication site. The optimization problem is defined in Section 4.1, and the communication site management procedures are presented in Section 4.2.

4.1. Problem Formulation. At the beginning of each time slot t , the edge controller receives the energy level report $\beta_n(t)$ from each EM (via the MEC application responsible for energy profiles in the MEC server), using the pull mode file transfer. Here, we aim to minimize the overall energy consumption in the communication site over time, i.e., the consumption related to the BS transmission activity and the MEC server, by applying BS power saving modes and VM soft-scaling, i.e., tuning the number of active virtual machines. To achieve this, we first consider the optimization for a single communication site. We define two cost functions as follows:

(F1) $\theta_{\text{tot},n}(t)$, which weighs the energy consumption due to transmission (BS) and computation (MEC server);

(F2) a quadratic term $(\Gamma_n(t) - B_n(t))^2$, which accounts for the QoS cost.

In fact, (F1) tends to push the system towards self-sustainability solutions; i.e., $\zeta_n(t) \rightarrow \varepsilon$. Instead, (F2) favors solutions where the delay sensitive load is entirely processed by the local MEC server; i.e., $B_n(t) \rightarrow \Gamma_n(t)$. A weight $\eta \in [0, 1]$ is utilized to balance the two objectives (F1) and (F2). The corresponding (weighted) cost function is defined as

$$J(\zeta, \alpha, t) \triangleq \bar{\eta} \theta_{\text{tot},n}(\zeta_n(t), \{\alpha_m(t)\}, t) + \eta (\Gamma_n(t) - B_n(t))^2, \quad (5)$$

where $\bar{\eta} \triangleq 1 - \eta$; with $\{\alpha_m(t)\}$ we mean the sequence of factors $\alpha_1(1), \alpha_2(1), \dots, \alpha_{M(t)}(1)$. Hence, letting l be the current time

slot and T be the time horizon, the following optimization problem is formulated over time slots $1, \dots, T$:

$$\begin{aligned}
 \mathbf{P1}: \min_{\zeta, \alpha} \quad & \sum_{t=1}^T J(\zeta, \alpha, t) \\
 \text{subject to:} \quad & \text{C1: } \zeta_n(t) \in \{\varepsilon, 1\}, \\
 & \text{C2: } b \leq M(t) \leq M, \\
 & \text{C3: } \beta_n(t) \geq \beta_{\text{low}}, \\
 & \text{C4: } 0 \leq f_m(t) \leq f_{\text{max}}, \\
 & \text{C5: } 0 \leq \gamma_m(t) \leq \gamma^{\text{max}}, \\
 & \text{C6: } \mu_m(t) \leq \Delta, \quad t = 1, \dots, T,
 \end{aligned} \tag{6}$$

where $m = 1, \dots, M(t)$ (VM index) and vectors ζ (BS switching status in time slots $1, \dots, T$) and α (load dependent factor) contain the *control actions* for the considered time horizon, per communication site; i.e., $\zeta = [\zeta(1), \zeta(2), \dots, \zeta(T)]$ and $\alpha = \{[\alpha_m(1)], [\alpha_m(2)], \dots, [\alpha_m(T)]\}$. Constraint C1 specifies the BS operation status (either *power saving* or *active*), C2 forces the required number of VMs, $M(t)$, to be always greater than or equal to a minimum number $b \geq 1$: the purpose of this is to be always able to handle mission critical communications. C3 makes sure that the EB level is always above or equal to a preset threshold β_{low} , to guarantee *energy self-sustainability* over time. Note that this constraint may imply that in certain time slots the BS is to be switched off, although the workload may be nonnegligible. When managing a single BS site (the formulation in this section), this implies that the load will not be served, but this fact may be compensated for when multiple communication sites are jointly managed, e.g., handing off the workload to another, energy richer, and BS. This is dealt with in Section 5. Furthermore, C4 and C5 bound the maximum processing rate and workloads of each running VM m , with $m = 1, \dots, M(t)$, respectively. Constraint C6 represents a hard-limit on the corresponding per-slot and per-VM processing time.

To solve P1 in (6), we leverage the use of LLC [8, 30] and heuristics, obtaining the controls $\zeta(t) \triangleq (\zeta(t), \{\alpha(t)\})$ for $t = 1, \dots, T$. Note that (6) can iteratively be solved at any time slot $t \geq 1$, by just redefining the time horizon as $t' = t, t+1, \dots, t+T-1$.

4.2. Communication Site Management. In this subsection, a traffic load and energy harvesting prediction method and an online management algorithm are proposed to solve the previously stated problem P1. In Section 4.2.1, we discuss the prediction of the future (short-term) traffic load and harvested energy processes, and then in Section 4.2.2, we solve P1 by first constructing the state-space behavior of the control system, where online control key concepts are introduced. Finally, the algorithm for managing the single communication site is presented in Section 4.2.3.

Modeling steps

- Step 1: load and normalize the dataset
- Step 2: split dataset into training and testing
- Step 3: reshape input to be [samples, time steps, features]
- Step 4: create and fit the LSTM network
- Step 5: make predictions
- Step 6: calculate performance measure

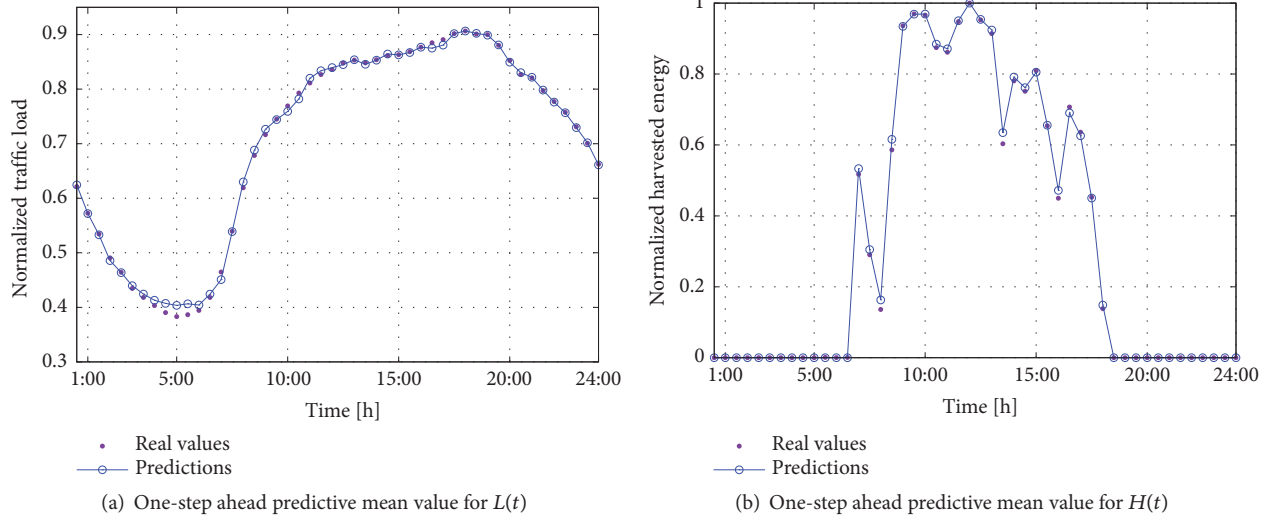
Box 1: LSTM prediction model steps.

4.2.1. Traffic Load and Energy Forecasting. ML techniques constitute a promising solution for network management and energy savings in cellular networks [52, 53]. In this work, given a time slot duration of $\tau = 30$ min, we perform time series prediction; i.e., we obtain the $T = 3$ estimates of $\hat{L}_n(t)$ and $\hat{H}_n(t)$, by using an LSTM network developed in Python using Keras deep learning libraries (Sequential, Dense, LSTM) where the network has a visible layer with one input, one hidden layer of four LSTM blocks or neurons, and an output layer that makes a single value prediction. This type of recurrent neural network uses backpropagation through time for learning and memory blocks for regression [7]. The dataset is split as 67% for training and 33% for testing. The network is trained using 100 epochs (2,600 individual training trials) with batch size of one. As for the performance measure of the model, we use the Root Mean Square Error (RMSE). The prediction steps are outlined in Box 1. Figures 4(a) and 4(b) show the prediction results that will be discussed in Section 6.

4.2.2. Edge System Dynamics. We denote the system state vector at time t by $\mathbf{x}(t) = (M(t), \beta_n(t))$, which contains the number of active VMs, $M(t)$, and the EB level, $\beta_n(t)$, for the BS site n . $\boldsymbol{\zeta}(t) = (\zeta(t), \{\alpha_m(t)\})$ is the input vector, i.e., the control action that drives the system behavior at time t . The system evolution is described through a discrete-time state-space equation, adopting the LLC principles [8, 30]:

$$\mathbf{x}(t+1) = \Phi(\mathbf{x}(t), \boldsymbol{\zeta}(t)), \tag{7}$$

where $\Phi(\cdot)$ is a behavior model that captures the relationship between $(\mathbf{x}(t), \boldsymbol{\zeta}(t))$, and the next state $\mathbf{x}(t+1)$. Note that this relationship accounts for (1) the amount of energy drained $\theta_{\text{tot},n}(t)$ that harvested $H_n(t)$ and that purchased from the power grid $Q_n(t)$, which together lead to the next buffer level $\beta_n(t+1)$ through (4) and (2) to the traffic load $L_n(t)$, from which we compute the server workloads $\Gamma_n(t)$ that leads to $M(t)$ and to the control $\boldsymbol{\zeta}(t)$. The network management algorithm in the edge controller, the ENAAM algorithm, finds the best control action vector for the communication site, following a *model predictive control approach*. Specifically, for each time slot t , problem (6) is solved, obtaining control actions for the whole time horizon $t, t+1, \dots, t+T-1$. The control action that is applied at time t is $\boldsymbol{\zeta}^*(t)$, which is the first one in the retrieved control sequence. This control amounts to setting the BS radio mode according to $\zeta^*(t)$, i.e., either active or power saving, and the number of instantiated VMs,

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

$M^*(t)$, along with their obtained $\{\alpha_m^*(t)\}$ values (see Remarks 1 and 2 below). This is repeated for the following time slots $t+1, t+2, \dots$

Remark 1 (role of prediction). State $\mathbf{x}(t)$ and control $\boldsymbol{\zeta}(t)$ are, respectively, measured and applied at the beginning of time slot t , whereas the offered load $L_n(t)$ and the harvested energy $H_n(t)$ are accumulated during the time slot and their value becomes known only by 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 we formally write as

$$\hat{\mathbf{x}}(t+1) = \Phi(\mathbf{x}(t), \boldsymbol{\zeta}(t)), \quad (8)$$

and the same applies to the subsequent time slots in the optimization horizon $t+2, t+3, \dots, t+T-1$. For these estimations we use the forecast values of load $\hat{L}_n(t)$ and harvested energy $\hat{H}_n(t)$, from the LSTM forecasting module.

Remark 2 (VM number and workload allocation). A remark on the provisioned VMs per time slot per-MEC server, $M(t)$, is in order. Specifically, the number of active VMs (i.e., the VM computing cluster) depends on the predicted load, $\hat{L}_n(t+1)$, where the expected server workload is $\hat{\Gamma}_n(t+1) = 0.8\hat{L}_n(t+1)$. Each VM can compute an amount of up to γ^{\max} . Then, an estimate of the number of virtual machines that shall be active in time slot t to serve the predicted server workloads is here obtained as follows: $M(t) = \lceil (\hat{\Gamma}_n(t+1)/\gamma^{\max}) \rceil$, where $\lceil \cdot \rceil$ returns the nearest upper integer. We heuristically split the workload among virtual machines by allocating a workload $\gamma_m(t) = \gamma^{\max}$ to the first $M(t) - 1$ VMs, $m = 1, \dots, M(t) - 1$, and the remaining workload $\gamma_m(t) = \hat{L}_n(t+1) - (M(t) - 1)\gamma^{\max}$ to the last one $m = M(t)$.

Controller Decision-Making. The controller is obtained by estimating the relevant parameters of the operating environment, i.e., the BS load $\hat{L}_n(t)$ and the harvested energy $\hat{H}_n(t)$, and subsequently using them to forecast the future system behavior through (8) over a look-ahead time horizon of T time slots. The control actions are picked by minimizing $J(\boldsymbol{\zeta}, \boldsymbol{\alpha}, t)$ (see (5)). At the beginning of each time slot t the following process is iterated:

- (1) Future system states, $\hat{\mathbf{x}}(t+k)$, for a prediction horizon of $k = 1, \dots, T$ steps are estimated using (8). These predictions depend on past inputs and outputs up to time t , on the estimated load $\hat{L}_n(\cdot)$ and energy harvesting $\hat{H}_n(\cdot)$ processes, and on the control $\boldsymbol{\zeta}(t+k)$, with $k = 0, \dots, T-1$.
- (2) The sequence of controls $\{\boldsymbol{\zeta}(t+k)\}_{k=0}^{T-1}$ is obtained for each step of the prediction horizon by optimizing the weighted cost function $J(\cdot)$ (see (5)).
- (3) The control $\boldsymbol{\zeta}^*(t)$ corresponding to the first control action in the sequence with the minimum total cost is the applied control for time t and the other controls $\boldsymbol{\zeta}^*(t+k)$ with $k = 1, \dots, T-1$ are discarded.
- (4) At the beginning of the next time slot $t+1$, the system state $\mathbf{x}(t+1)$ becomes known and the previous steps are repeated.

4.2.3. The ENAAM Algorithm. Let t be the current time. $\hat{L}_n(t+k-1)$ is the forecast load in slot $t+k-1$, with $k = 1, \dots, T$, i.e., over the prediction horizon. For the control to be feasible, we need $\underline{\Gamma}_n(t) \leq B_n(t) \leq \hat{\Gamma}_n(t+k-1)$, where $\underline{\Gamma}_n(t)$ is the smallest Γ such that $\text{round}(\hat{\Gamma}_n(t+1)/\gamma^{\max}) = b$. For the buffer state, we heuristically set $\zeta(t+k-1) = \varepsilon$ if either $\beta_n(t+k-1) < \beta_{\text{low}}$ or $L_n(t+k-1) < L_{\text{low}}$, and $\zeta(t+k-1) = 1$; otherwise β_{low} and L_{low} are preset low thresholds for the EB and the BS load, respectively. For slot $t+k-1$, the feasibility

```

Input:  $\mathbf{x}(t)$  (current state)
Output:  $\boldsymbol{\zeta}^*(t) = (\zeta^*(t), \{\alpha_m^*(t)\})$ 
01: Initialization of variables
 $\mathcal{S}(t) = \{\mathbf{x}(t)\}, \text{Cost}(\mathbf{x}(t)) = 0$ 
02: for  $k = 1, \dots, T$  do
    (i) forecast the load  $\widehat{L}_n(t+k-1)$ 
    (ii) forecast the harvested energy  $\widehat{H}_n(t+k-1)$ 
    (iii)  $\mathcal{S}(t+k) = \emptyset$ 
03: for all  $\mathbf{x} \in \mathcal{S}(t+k-1)$  do
04:   for all  $\boldsymbol{\zeta} = (\zeta, \{\alpha_m(t)\}) \in \mathcal{A}(t+k-1)$  do
05:      $\widehat{\mathbf{x}}(t+k) = \Phi(\mathbf{x}(t+k-1), \boldsymbol{\zeta})$ 
06:      $\text{Cost}(\widehat{\mathbf{x}}(t+k)) = J(\zeta, \alpha, t+k-1) + \text{Cost}(\mathbf{x}(t+k-1), \boldsymbol{\zeta})$ 
07:      $\mathcal{S}(t+k) = \mathcal{S}(t+k) \cup \{\widehat{\mathbf{x}}(t+k)\}$ 
   end for
end for
08: Find  $\widehat{\mathbf{x}}_{\min} = \text{argmin}_{\widehat{\mathbf{x}} \in \mathcal{S}(t+T)} \text{Cost}(\widehat{\mathbf{x}})$ 
09:  $\boldsymbol{\zeta}^*(t) :=$  control leading from  $\mathbf{x}(t)$  to  $\widehat{\mathbf{x}}_{\min}$ 
10: Return  $\boldsymbol{\zeta}^*(t)$ 

```

ALGORITHM 1: ENAAM.

set $\mathcal{A}(t+k-1)$ contains the control pairs $(\zeta(t), \{\alpha_m(t)\})$ that obey these relations.

The algorithm is specified in Algorithm 1 as it uses the technique in [8]: the search starts (line 01) from the system state at time t , $\mathbf{x}(t)$, and continues in a breadth-first fashion, building a tree of all possible future states up to the prediction depth T . A cost is initialized to zero (line 01) and is accumulated as the algorithm travels through the tree (line 06), accounting for predictions, past outputs, and controls. The set of states reached at every prediction depth $t+k$ is referred to as $\mathcal{S}(t+k)$. For every prediction depth $t+k$, the search continues from the set of states $\mathcal{S}(t+k-1)$ reached at the previous step $t+k-1$ (line 03), exploring all feasible controls (line 04), obtaining the next system state from (8) (line 05), updating the accumulated cost as the result of the previous accumulated cost, plus the cost associated with the current step (line 06), and updating the set of states reached at step $t+k$ (line 07). When the exploration finishes, the initial action (at time t) that leads to the best final accumulated cost, at time $t+T-1$, is selected as the optimal control $\boldsymbol{\zeta}^*(t)$ (lines 08, 09, 10). Finally, for line 04, we note that Γ_n belongs to the continuous set $[\underline{\Gamma}_n, \widehat{L}_n(t+k-1)]$. To implement this search, we quantized this interval into a number of equally spaced points, obtaining a search over a finite set of controls.

ENAAM Complexity. The computation complexity of the algorithm is $O(N_x N_\zeta T)$, where $N_x \triangleq |\mathbf{x}(t)|$ and $N_\zeta \triangleq |\boldsymbol{\zeta}(t)|$, respectively, represent the number of system states and the number of feasible actions at time t . Note that state and action space are, respectively, quantized into $N_x = M \times N_\beta$ and $N_\zeta = 2 \times M \times N_\alpha$ levels, where M is the number of virtual machines, N_β is the number of quantization levels for the energy buffer, and N_α is the number of quantization levels for the load variable $\alpha_m(t)$. Such quantization facilitates the

search in Algorithm 1. Note that exhaustive search would entail a complexity of $O((N_x N_\zeta)^T)$.

5. Multiple Communication Sites

In this section, we extend the work from Section 4 by considering the energy savings for *multiple* communication sites. We formulate an optimization problem to obtain energy savings through short-term traffic load and harvested energy predictions and clustering, along with energy management procedures for the clustered BS sites. The problem formulation for multiple communication sites is described in Section 5.1; then cluster formation is discussed in Section 5.2, and the edge management procedure for each cluster, enabled by the edge controller, is presented in Section 5.3.

5.1. Problem Formulation. Our objective is to improve the overall energy savings of the network by clustering BSs based on their location (or distance measures) similarity and then optimizing the energy savings within each cluster by employing the single optimization case described in Section 4. From an energy efficiency perspective, in a cluster of BS nodes, one BS (or more) might have a preference of switching off, by first offloading its (their) traffic load to its (their) neighboring BS that have enough spare capacity for handling extra traffic load and then switching off. The whole offloaded traffic load from the BS, denoted by BS n , is allocated to the neighboring cluster member (active BS) in which orthogonal resource allocation helps mitigate intracluster interference, such that the selected neighboring BS, denoted by BS n' , is allocated the incremental load, denoted by $L_{nn'}(t) \triangleq L_n(t)$. Whenever a BS is switched off, it should maintain service to its users via a reassociation process in order to offload the users to the neighboring active BS having extra resources for handling upcoming extra traffic load. The reassociation process involves notifying the connected users to try and connect to neighboring BSs with extra resources.

In the view of the above, we consider that all BSs are grouped into sets of clusters $\mathcal{O} = \{O_1, \dots, O_{|\mathcal{O}|}\}$. Here, a given cluster $O_i \in \mathcal{O}$, with $i = 1, \dots, |\mathcal{O}|$, consists of a set of BSs that coordinate with the controller. The clustering mechanism is discussed in Section 5.2. For each cluster $O_i \in \mathcal{O}$, we aim to minimize the energy consumption, i.e., the consumption due to BS transmission and the running VMs in the servers, using BS power saving modes and VM soft-scaling per active cluster member. To do so, we define a cost function which captures the individual communication site energy consumption and its QoS. The (weighted) cost for each cluster member, BS $n \in O_i$, is redefined as follows:

$$J_n(\zeta, \alpha, t) \triangleq \overline{\eta} \theta_{\text{tot},n}(\zeta_n(t), \{\alpha_m(t)\}_n, t) + \eta (\Lambda_n(t) - B_n(t))^2, \quad (9)$$

where $\zeta_n(t)$ is the activity status of BS n (either *power saving* or *active*) and $\{\alpha_m(t)\}_n$ is the set of factors for the allocated VMs at BS n . Moreover, $\Lambda_n(t) \leftarrow L_n(t)$ if BS n only handles its own traffic, whereas $\Lambda_n(t) \leftarrow L_n(t) + \Delta L_n(t)$, in case one

(or multiple) BSs are switched off in time slot t and its (their) traffic is redirected (handed off) to BS n . The computation of $\Delta L_n(t)$ is addressed in Section 5.3. The per cluster cost $Y_{O_i}(\zeta_i, \alpha_i, t)$ is the aggregated cost of all cluster members, $Y_{O_i}(\zeta_i, \alpha_i, t) = \sum_{n \in O_i} J_n(\zeta, \alpha, t)$. Hence, over time horizon, $t = 1, \dots, T$, the following optimization problem is defined:

$$\begin{aligned} \mathbf{P2}: \min_{\mathcal{E}} \quad & \sum_{O_i \in \mathcal{O}} Y_{O_i}(\zeta_i, \alpha_i, t) \\ \text{subject to:} \quad & \text{C1 - C6: from Eq. (6),} \\ & \text{C7: } |O_i| \geq 1, \quad \forall O_i \in \mathcal{O}, \\ & \text{C8: } O_i \cap O_j = \emptyset, \\ & \quad \forall O_i, O_j \in \mathcal{O}, O_i \neq O_j, \end{aligned} \quad (10)$$

where $\mathcal{E} \triangleq \{\zeta_i, \alpha_i\}$ is the collection of variables to be reconfigured for all the BS clusters (the whole MN), for all time slots $t = 1, \dots, T$. As for the constraints, C7 and C8 ensure that each BS is part of only one cluster. Solving **P2** in (10) involves BS clustering, the forecasting method from Section 4.2.1, a heuristic rule for the selection of which BSs have to be switched off, and the ENAAM algorithm from Section 4.2.3. Once **P2** is solved, the control action to be applied at time t , per cluster O_i , corresponds to the elements in $\{\zeta_i, \alpha_i\}$ that are associated with the first time slot 1 in the optimization horizon. As above, (10) can iteratively be solved at any time slot $t \geq 1$, by just redefining the time horizon as $t' = t, t + 1, \dots, t + T - 1$.

5.2. Cluster Formation. Clustering algorithms have been proposed as a way of enabling energy saving mechanisms in BSs, where groups of inactive BSs or BSs with low loads are switched off. With the advent of EH BSs, the BSs with $\beta_n(t) < \beta_{\text{low}}$ can be switched off, while still guaranteeing the QoS through the other active BSs. That is, within each formed cluster, the controller tries to minimize the cost function, which captures the tradeoff between the energy efficiency and the QoS of each cluster member. The key step in clustering is to identify similarities or distance measures between BSs in order to group BSs with similar characteristics. In this paper, we use the location of the BSs as it defines the relative neighborhood (the distance measures) with the other BSs. Using the location of the BSs and the distance between the BSs, we obtain a distance-based similarity matrix \mathbf{W}^d . In addition, we assume that the network topology is static during the clustering algorithm execution.

In Section 5.2.1 we detail the clustering measure that we use to obtain the similarities between BSs based on location, followed by the distance-based clustering algorithm in Section 5.2.2.

5.2.1. Relative Neighborhood Based on BS Adjacency and Gaussian Similarity. Similar to [13], we model the MN as a graph $G = (\mathcal{N}, E)$, where \mathcal{N} represents the set of BSs, while the set E contains the edges between any two BSs. There is an edge $(n, n') \in E$ if and only if n and n' can mutually receive

each other's transmission. In this case, we say that n and n' are neighbors. We use a parameter $r_{nn'}$ to characterize the presence of a link between nodes, where $r_{nn'} \in \{0, 1\}$. Let y_n be the coordinates of BS $n \in \mathcal{N}$ in the Euclidean space. The relative neighborhood of BS n is defined by the nearness of the BSs in its e_d -radio propagation space (or neighborhood):

$$\mathcal{X}_n = \{n' \text{ s.t. } \|y_n - y_{n'}\| \leq e_d\}. \quad (11)$$

If $n' \in \mathcal{X}_n$ we say that BSs n and n' are neighbors, and we set $r_{nn'} = 1$; otherwise $r_{nn'} = 0$. The links between the vertices in \mathcal{N} are weighted based on their similarities. Based on the distance between BS n and n' , we can classify the BSs based on their location using the Gaussian similarity measure [13] (a classification kernel function used in machine learning), which is defined as

$$w_{nn'}^d = \begin{cases} \exp\left(\frac{-\|y_n - y_{n'}\|^2}{2\sigma_d^2}\right) & \text{if } \|y_n - y_{n'}\| \leq e_d, \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

where $2\sigma_d^2$ adjusts the impact of the neighborhood size. In (12), we assume that the BSs located far from each other have low similarities, compared to those that are close to each other, as those that are close are more likely to cooperate with each other. The distance-based similarity matrix \mathbf{W}^d is formed using $w_{nn'}^d$ as the (n, n') -th entry.

5.2.2. Distance-Based Clustering. The BS clustering is performed after obtaining the similarity matrix \mathbf{W}^d of the MN graph $G = (\mathcal{N}, E)$. Given the matrix \mathbf{W}^d , we employ a centralized clustering method, specifically the K-means [54], as the matrix provides the full location knowledge. K-means partitions the set of nodes into clusters in which each node belongs to the cluster with the nearest mean distance. In addition, the value of K , i.e., the number of clusters ($|O_i|$), is known prior and is a design parameter. This algorithm requires knowledge of all the BS locations; thus, it is categorized as a centralized method. In our case, this process does not incur any computation delay as the edge controller is assumed to have high computation capabilities.

5.3. Edge Network Management. Our aim is to implement and validate an LLC framework for dynamic resource provisioning in multiple communication sites with the goal of achieving energy savings within the access network through BS sleep modes and VM soft-scaling. Given the formation of clusters, load, and energy forecasting, our next goal is to develop a mechanism for solving **P2** (see (10)) where each cluster of BSs adjusts its transmission parameters and its computing cluster entities based on the forecast information. In order to minimize the per cluster cost function, we introduce the notion of *network impact* in Section 5.3.1, whereas we describe the edge management procedure in Section 5.3.2.

5.3.1. Network Impact. The dynamic BS switching off strategies may have an impact on the network due to the traffic

load that is offloaded to the neighboring BSs. To avoid this, the BS to be switched off must be carefully identified within a BS cluster. To determine whether a particular BS can be switched off or not, we follow the work done in [55]. As an example, we consider one cluster O_i , together with its cluster members $n \in O_i$, then from it we choose one BS, BS n , where BS n neighbors set is denoted by \mathcal{N}_n . Note that the BS $n' \in \mathcal{N}_n$ is the BS to which the traffic load will be offloaded to after turning off BS n . Also, BS n can only be switched off if there exists a neighboring BS n' that satisfies the following feasibility constraint [55]:

$$L_{n'}(t) + L_{nn'}(t) \leq 1, \quad n' \in \mathcal{N}_n, \quad (13)$$

where $L_{n'}(t)$ is the original BS n' traffic load and $L_{nn'}(t)$ is the incremental traffic load from BS n (the switched off BS) to BS n' (the neighboring BS). We recall that the load $L_{n'}(t)$ is normalized with respect to the maximum load that a BS can sustain, so the inequality in (13) means that it is feasible for BS n' to take the extra load from BS n . To quantify how the incremental system load affects the overall network load due to the switching off process, we introduce the notion of *network impact*. For every BS n within cluster O_i , $i = 1, \dots, K$, its *network impact* due to the offloaded system load onto one of the neighboring BSs is defined as follows:

$$I_n(t) = \max_{n' \in \mathcal{N}_n} [L_{n'}(t) + L_{nn'}(t)], \quad \forall n \in O_i. \quad (14)$$

Here, the maximum network impact value $I_n(t)$ over the neighboring BSs is considered as a measure for each BS towards switching off and generating extra traffic loads for its neighboring BSs. In this work, considering cluster O_i , we switch off the BS n^* that has the least network impact; i.e.,

$$n^* = \arg \min_{n \in O_i} I_n(t). \quad (15)$$

The BS that takes the load from n^* is selected as the BS n' that minimizes $L_{n'}(t) + L_{n^*n'}(t)$ over the set of active BSs that are on within the cluster O_i . For BS n' , we then set $L_{n'}(t) \leftarrow L_{n'}(t) + L_{n^*n'}(t)$. This procedure is sequentially repeated for all the cluster members until there is no active BS whose neighbors satisfy the feasibility condition of (13). Note that here, we focus only on which BS to switch off, as for the BS turning on state, we assume that the *commitment time* (time configured so that the BS automatically wakes up without external triggers) is a system parameter that is preconfigured when the BS is switched off.

5.3.2. Edge Management Procedure. Here, we propose a distributed edge network management procedure that makes use of the ENAAM algorithm (see Section 4.2.3). The decision-making criterion only depends on the BS information and on its neighboring BSs; thus, the BS switching off decision can be localized within each cluster. To decide which BSs shall be switched off, we follow a sequential decision process. While this is heuristic, it allows coping with the high complexity associated with an optimal (all BSs are jointly assessed) allocation approach. The edge management procedure is as follows.

For each BS cluster O_i , with $i = 1, \dots, K$, we have the following:

- (1) Initialize an allocation variable $\Delta L_n(t) = 0$ for all BSs $n \in O_i$. Compute $I_n(t)$, using (14), for all BSs n and obtain the BS with the least *network impact* $n^*(t)$, using (15). Switch off BS $n^*(t)$ and assign its load to the neighboring BS $n' \in O_i$ that minimizes $L_{n'}(t) + \Delta L_{n'}(t) + L_{n^*n'}(t)$. Update the extra allocation for BS n' as $\Delta L_{n'}(t) \leftarrow \Delta L_{n'}(t) + L_{n^*n'}(t)$. Recompute $I_n(t)$ for all the BSs that are still on and identify the next BS that can be switched off, i.e., the one with the *least network impact*. This procedure is repeated until none of the BSs in the cluster verifies Eq. (13). At this point, we have identified all the BSs n^* that shall be switched off in O_i .
- (2) For each active BS $n' \in O_i$, the ENAAM algorithm is executed using $L_{n'}(t) + \Delta L_{n'}(t)$, where $\Delta L_{n'}(t) = 0$ if BS n' does not take extra load, whereas it is greater than zero otherwise. Note that, $\Delta L_{n'}(t)$ corresponds to the total traffic that is handed over to BS n' , possibly from multiple nearby BSs.

Edge Network Management Complexity. The algorithm is independently executed for each cluster and the corresponding time complexity is obtained as follows. Considering the action *Step (1)*, from above, the time complexity associated with the computation of the BS having the least network impact is linear with the size of the cluster $|O_i|$. Once that is computed, the complexity associated with updating the load allocation for the active BSs is $|O_i| - 1$, which leads to a total complexity of $|O_i|(|O_i| - 1) = O(|O_i|^2)$. Moreover, such process is iterated for each BS that is switched off. In the worst case, where all the BSs but one are switched off, the final complexity of step 1 is $O(|O_i|^3)$. As for *Step (2)*, from above, the computation complexity depends on the ENAAM algorithm, which is independently executed by each *active* BS. Thus, in the worst case (no BSs are switched off), the total aggregated complexity is as follows: $O(|O_i|N_xN_\zeta T)$, which is linear in all variables, namely, number of cluster members, number of BS states, number of actions, and time horizon T .

6. Performance Evaluation

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

6.1. Simulation Setup. We consider multiple BSs, each one collocated with a MEC server and a coverage radius of 40 m. In addition, we use a virtualized server with specifications from [56] for a VMware ESXi 5.1-ProLiant DL380 Gen8. Our time slot duration τ is set to 30 min and the time horizon is set to $T = 3$ time slots. The simulations are carried out by exploiting the Python programming language.

6.2. Numerical Results. Pattern Forecasting. We show real and predicted values for the traffic load and harvested energy over time in Figures 4(a) and 4(b), where we track the

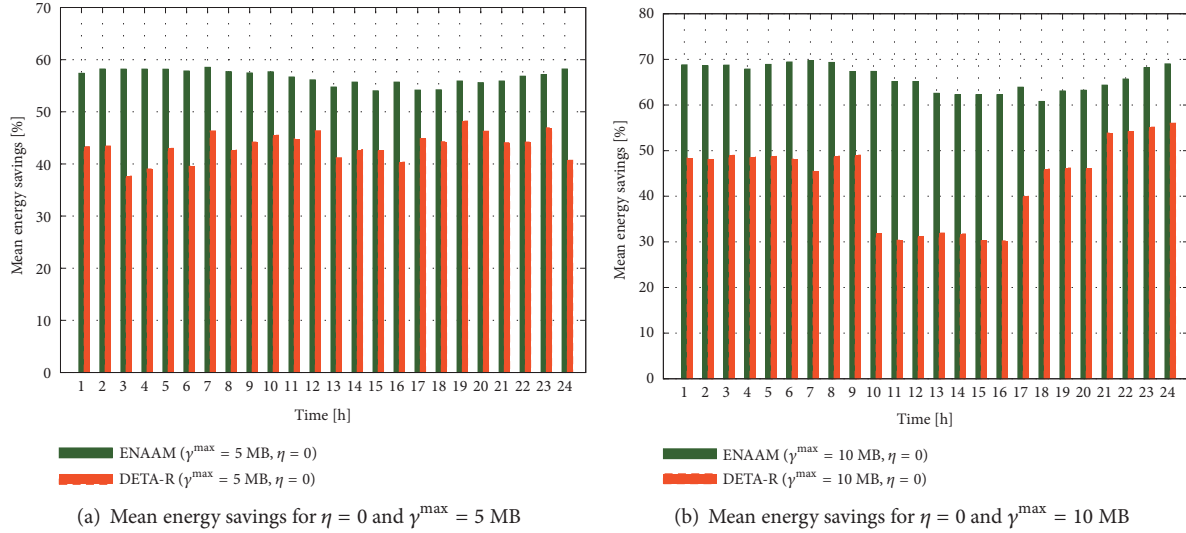


FIGURE 5: Mean energy savings for the single BS case.

TABLE 2: System parameters.

Parameter	Value
Total BSs, N	24
Max. number of VMs, M	27
Min. number of VMs, b	1
Time slot duration, τ	30 min
Operating power, θ_0	10.6 W
Energy overheads for switching VM, $\theta_m^{ov}(t)$	0.05 J/MHz ²
Max. computation workload per VM, γ^{\max}	{5, 10} MB
Max. allowed processing time, Δ	0.8 s
Energy cons. of network interfaces, $\theta_{idle}(t)$	3 J
Cost of exchanging one unit of data, $\theta_{data}(t)$	6 J/byte
Processing rate set, \mathcal{F}	{0, 4, 8, 12, 16, 20}
Static energy consumed by VM, $\theta_{idle,m}(t)$	4 J
Max. energy cons. by VM at f_{\max} , $\theta_{\max,m}(t)$	10 J
Energy storage capacity, β_{\max}	490 kJ
Lower energy threshold, β_{low}	30% of β_{\max}
Upper energy threshold, β_{up}	70% of β_{\max}
Low traffic threshold, L_{low}	4 MB

one-step predictive mean value at each step of the online forecasting routine. Then, Table 3 shows the average RMSE of the normalized harvested energy and traffic load processes, for different time horizon values, $T \in \{1, 2, 3\}$. Note that the predictions for $H(t)$ are more accurate than those of $L(t)$ (confirmed by comparing the average RMSE), due to differences in the used dataset granularity. However, the measured accuracy is deemed good enough for the proposed optimization.

Single Communication Site. Figures 5(a) and 5(b) are computed with $\eta = 0$ using Cluster 1 and Solar 1 as traffic load and harvested energy profiles for each BS (see Figures 2 and 3). Moreover, $\gamma^{\max} = 5$ MB and 10 MB, respectively.

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

	$T = 1$	$T = 2$	$T = 3$
$L(t)$	0.037	0.042	0.048
$H(t)$	0.011	0.016	0.021

They show the mean energy savings achieved over time when on-demand and energy aware edge resource provisioning are enabled (i.e., BS sleep modes and VM soft-scaling), in comparison with the case where they are not applied. Our edge network management algorithm (ENAAM) is benchmarked with another one that heuristically selects the amount of traffic that is to be processed locally, $B_n(t) \leq \Gamma_n(t)$, depending on the expected load behavior. It is named Dynamic and Energy-Traffic-Aware algorithm with Random behavior (DETA-R). Both ENAAM and DETA-R are aware of the predictions in future time slots (see Section 4.2.1); however, DETA-R provisions edge resources using a heuristic scheme. DETA-R heuristic works as follows: if the expected load difference is $\hat{L}(t+1) - \hat{L}(t) > 0$, then the normalized workload to be processed by BS n in the current time slot t , $B_n(t)$, is randomly selected in the range $[0.6, 1]$; otherwise, it is picked evenly at random in the range $(0, 0.6)$.

Average results for the ENAAM scheme show energy savings of 69% ($\gamma^{\max} = 10$ MB) and 57% ($\gamma^{\max} = 5$ MB), while DETA-R achieves 49% ($\gamma^{\max} = 10$ MB) and 43% ($\gamma^{\max} = 5$ MB) on average, where these savings are with respect to the case where *no energy management* is performed; i.e., the network is dimensioned for maximum expected capacity (maximum value of $\theta_{\text{tot},n}(t)$, with $M = 27$ VMs, $\forall t$). The results show that the maximum load allocated to each VM, γ^{\max} , has an impact towards energy savings. An increase in energy savings is observed when $\gamma^{\max} = 10$ MB due to the fact that the number of VMs demanded per time slot is reduced, when compared to the allocation of $\gamma^{\max} = 5$ MB.

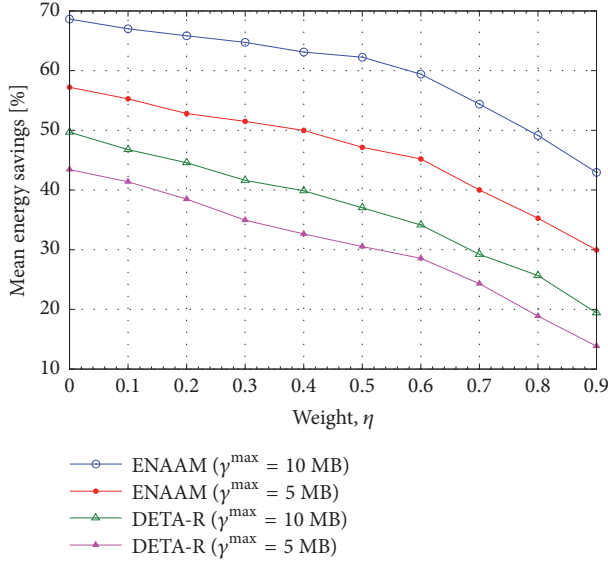


FIGURE 6: Energy savings versus weight η (single BS case).

The ESs evolution with respect to η is presented in Figure 6, taking into account the load allocated to each VM, γ^{\max} . The results were obtained using Cluster 1 and Solar 1 as traffic load and harvested energy profiles (see Figures 2 and 3). As expected, a drop in energy savings is observed when QoS is prioritized, i.e., $\eta \rightarrow 1$, as in this case the BS energy consumption is no longer considered. It can be observed that ENAAM achieves a 50% (or above) from $\eta = [0, 0.4]$ when $\gamma^{\max} = 5$ MB and from $\eta = [0, 0.7]$ when $\gamma^{\max} = 10$ MB. This shows that the higher the load allocated to each VM, the lesser the energy that is drained, as few VMs are running. DETA-R operates at below 50% for all η and γ^{\max} values.

Multiple Communication Sites. Figures 7(a) and 7(b) present the mean energy savings achieved with respect to the cluster size and the weight η , using all the traffic load and harvested energy profiles from Figures 2 and 3. Each BS randomly picks its own traffic load and harvested energy profile at the beginning of the optimization process. Here, to select the BS to be switched off, we use the management procedure of Section 5.3. As for DETA-R, a BS is randomly selected to evolve its operating mode to power saving mode and offload its load to a nearby BS (in this case, the least loaded neighboring BS is selected), without taking into account its network impact measure.

Figure 7(a) shows the average energy savings obtained when clustering is adopted, i.e., here, the cluster size is increased from $|O_i| = 1$ to 10 and $\eta = 0$. The obtained energy savings are with respect to the case where all BSs are dimensioned for maximum expected capacity (maximum value of $\theta_{\text{tot},n}(t)$, with $M = 27$ VMs, $\forall t, \forall n \in O_i$). It should be noted that the energy savings increase as the size of the cluster grows, thanks to the load balancing among active BSs, which cannot be implemented in the single communication site scenario (i.e., when BSs are independently managed).

Then, Figure 7(b) shows the average energy savings with respect to η , when the cluster size is set to an intermediate

case ($|O_i| = 6$). Again, here the energy savings are obtained with respect to the case where all the BSs are dimensioned for maximum capacity. As expected, there is a drop in the energy savings achieved as the value of η increases, as QoS is prioritized. It can be observed that ENAAM achieves a value of 50% or above when $\eta = [0, 0.8]$ (at $\gamma^{\max} = 10$ MB) and when $\eta = [0, 0.6]$ (at $\gamma^{\max} = 5$ MB). DETA-R achieves value above 50% or above when $\eta = [0, 0.4]$ (at $\gamma^{\max} = 10$) and $\eta = [0, 0.1]$ (at $\gamma^{\max} = 5$ MB).

Comparing Figures 6 and 7(b), an average gain of 9% on the energy savings is observed when clustering is applied, by considering the mean energy savings with respect to η achieved with ENAAM for both cases. From Figure 7(a) we see that this gain can be as high as 16% for ENAAM with $\gamma^{\max} = 5$ MB (red curve) and bigger for the DETA-R approach. These results support the notion that performing a clustering-based optimization is beneficial thanks to the additional cooperation within each neighborhood of BSs. This cooperation allows switching off more BSs through load balancing, increasing the energy savings while still controlling the users' QoS.

7. Conclusions

In this paper, we have envisioned an edge network where a group of BSs are managed by a controller, for ease of BS organization and management, and also a mobile network where the edge apparatuses are powered by hybrid supplies, i.e., using green energy in order to promote energy self-sustainability and the power grid as a backup. Within the edge, each BS is endowed with computation capabilities to guarantee low latency to mobile users, offloading their workloads locally. The combination of energy saving methods, namely, BS sleep modes and VM soft-scaling, for single and multiple BS sites helps to reduce the mobile network's energy consumption. An edge energy management algorithm based on forecasting, clustering, control theory and heuristics, is proposed with the objective of saving energy within the access network, possibly making the BS system self-sustainable. Numerical results, obtained with real-world energy and traffic load traces, demonstrate that the proposed algorithm achieves energy savings between 57% and 69%, on average, for the single communication site case, and a gain ranging from 9% to 16% on energy savings is observed when clustering is applied, with respect to the allocated maximum per-VM loads of 5 MB and 10 MB. The energy saving results are obtained with respect to the case where no energy management techniques are applied, either in one BS or single cluster.

Data Availability

In this paper, we have used open source datasets for the mobile network (MN) traffic load and the harvested solar energy. 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

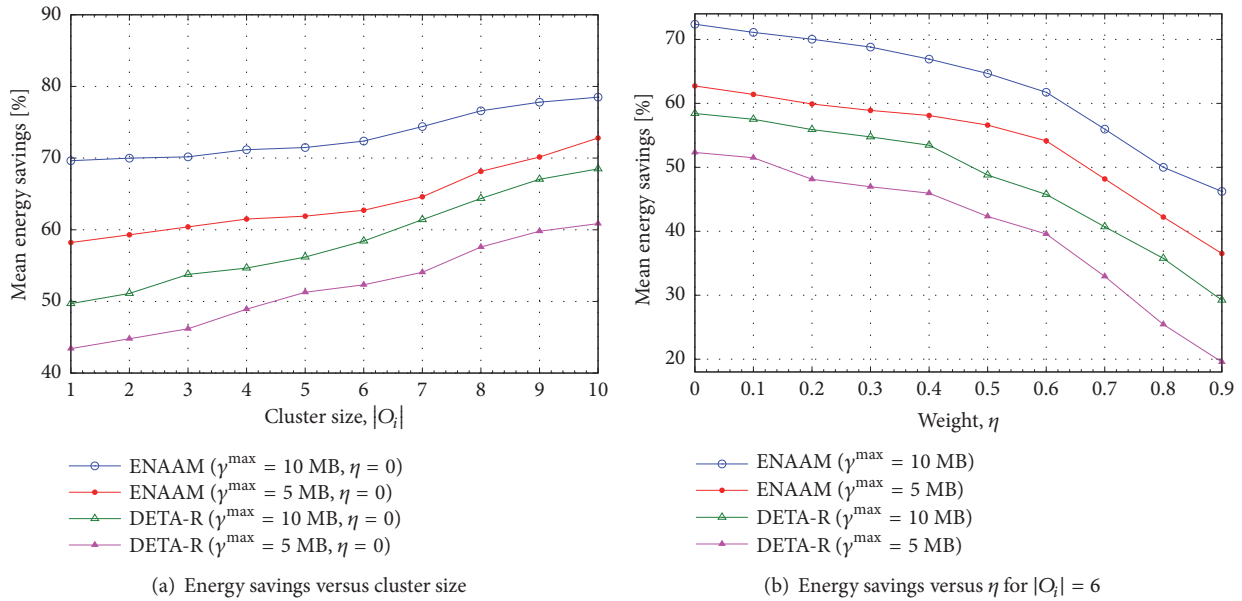


FIGURE 7: Energy savings for the multiple BSs case.

article. (2) The real solar energy 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.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement no. 675891 (SCAVENGE).

References

- [1] "Five Trends to Small Cells 2020," Tech. Rep., Huawei Technologies, Helsinki, Finland, 2016.
- [2] M. Patel, Y. Hu, P. Hédé et al., "Mobile edge computing introductory technical white paper," Tech. Rep., ETSI, Sophia-Antipolis, France, 2014.
- [3] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.
- [4] J. Erman and K. Ramakrishnan, "Understanding the super-sized traffic of the super bowl," in *Proceedings of the 2013 conference on Internet measurement conference*, Barcelona, Spain, October 2013.
- [5] R. Morabito, "Power consumption of virtualization technologies: An empirical investigation," in *Proceedings of the in IEEE International Conference on Utility and Cloud Computing (UCC)*, Limassol, Cyprus, Dec 2015.
- [6] Y. Jin, Y. Wen, and Q. Chen, "Energy efficiency and server virtualization in data centers: An empirical investigation," in *proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM Workshops)*, Orlando, FL, USA, Mar 2012.
- [7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Mass, USA, 2016.
- [8] S. Abdelwahed, N. Kandasamy, and S. Neema, "Online control for self-management in computing systems," in *Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, Ontario, Canada, 2004.
- [9] T. Dlamini, A. F. Gambin, D. Munaretto, and M. Rossi, "Online resource management in energy harvesting bs sites through prediction and soft-scaling of computing resources," in *Proceedings of the 2018 IEEE 29th annual international symposium on personal, indoor and mobile radio communications (PIMRC)*, Bologna, Italy, September 2018.
- [10] "Open Big Data Challenge," <https://dandelion.eu/datamine/open-big-data/>.
- [11] "Solar Radiation Measurement Data," <https://energydata.info/dataset/armenia-solar-radiation-measurement-data-2017>.
- [12] H. Zhang, J. Cai, and X. Li, "Energy-efficient base station control with dynamic clustering in cellular network," in *Proceedings of the IEEE International Conference on Communications and Networking (CHINACOM)*, Guilin, China, August 2013.
- [13] S. Samarakoon, M. Bennis, W. Saad, and M. Latva-Aho, "Dynamic clustering and on/off strategies for wireless small cell networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2164–2178, 2016.
- [14] S. Cai, L. Xiao, H. Yang, J. Wang, and S. Zhou, "A cross-layer optimization of the joint macro-and picocell deployment with sleep mode for green communications," in *Proceedings of the 22nd Wireless and Optical Communications Conference, WOCC 2013*, Chongqing, China, May 2013.
- [15] Y. Zhu, Z. Zeng, T. Zhang, and D. Liu, "A QoS-aware adaptive access point sleeping in relay cellular networks for energy efficiency," in *Proceedings of the IEEE Vehicular Technology Conference (VTC Spring)*, Seoul, Korea, May 2014.

- [16] Y. Yuan and P. Gong, "A QoE-orientated base station sleeping strategy for multi-services in cellular networks," in *Proceedings of the International Conference on Wireless Communications and Signal Processing (WCSP 2015)*, Nanjing, China, October 2015.
- [17] F. Han, Z. Safar, and K. J. R. Liu, "Energy-efficient base-station cooperative operation with guaranteed QoS," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3505–3517, 2013.
- [18] C. Liu, Y. Wan, L. Tian, Y. Zhou, and J. Shi, "Base station sleeping control with energy-stability tradeoff in centralized radio access networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, USA, December 2015.
- [19] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, "'Green' distance-aware base station sleeping algorithm in LTE-Advanced," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 1347–1351, Ottawa, Canada, June 2012.
- [20] H. Tabassum, U. Siddique, E. Hossain, and M. J. Hossain, "Downlink performance of cellular systems with base station sleeping, user association, and scheduling," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5752–5767, 2014.
- [21] Y. Zhu, Z. Zeng, T. Zhang, L. An, and L. Xiao, "An energy efficient user association scheme based on cell sleeping in LTE heterogeneous networks," in *Proceedings of the 2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 75–79, Sydney, Australia, September 2014.
- [22] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multiobjective auction-based switching-off scheme in heterogeneous networks: to bid or not to bid?" *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9168–9180, 2016.
- [23] B. Z. Dongsheng Han and Z. Chen, "Sleep mechanism of base station based on minimum energy cost," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 4202748, 13 pages, 2018.
- [24] M. D'Amours, A. Girard, and B. Sanso, "Planning solar in energy-managed cellular networks," *IEEE Access*, vol. 6, pp. 65212–65226, 2018.
- [25] N. Piovesan, A. Fernandez Gambin, M. Miozzo, M. Rossi, and P. Dini, "Energy sustainable paradigms and methods for future mobile networks: A survey," *Elsevier - Computer Communications*, vol. 119, pp. 101–117, 2018.
- [26] 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*, vol. 2017, Article ID 8618364, 21 pages, 2017.
- [27] A. Antonopoulos, E. Kartsakli, A. Bousia, L. Alonso, and C. Verikoukis, "Energy-efficient infrastructure sharing in multi-operator mobile networks," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 242–249, 2015.
- [28] M. Oikonomakou, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Evaluating cost allocation imposed by cooperative switching off in multi-operator shared HetNets," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11352–11365, 2017.
- [29] J. Maciejowski, *Predictive Control with Constraints*, Prentice Hall, 2002.
- [30] S.-L. Chung, S. Lafortune, and F. Lin, "Limited lookahead policies in supervisory control of discrete event systems," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 37, no. 12, pp. 1921–1935, 1992.
- [31] T. Ergen and S. S. Kozat, "Online training of LSTM networks in distributed systems for variable length data sequences," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 5159–5165, 2017.
- [32] J. Xu and S. Ren, "Online learning for offloading and autoscaling in renewable-powered mobile edge computing," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, Washington, DC, USA, December 2016.
- [33] A. Beloglazov, J. Abawayj, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing," *Future Generation Computer Systems*, vol. 28, no. 5, pp. 755–768, 2012.
- [34] R. Nathuji and K. Schwan, "VirtualPower: coordinated power management in virtualized enterprise systems," in *Proceedings of the 21st ACM SIGOPS symposium on operating systems principles, SOSP'07*, pp. 265–278, Washington, DC, USA, October 2007.
- [35] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast transparent migration for virtual machines," in *Proceedings of the annual conference on usenix annual technical conference*, Berkeley, Calif, USA, Apr 2005.
- [36] C. Jeffrey, A. Darrell, T. Prachi, V. Amin, and D. Ronald, "Managing energy and server resources in hosting centers," in *Proceedings of the 18th ACM Symposium on Operating Systems Principles*, Alberta, Canada, Oct 2001.
- [37] J. Lorch and A. J. Smith, "PACE: A new approach to dynamic voltage scaling," *IEEE Transactions on Computers*, vol. 53, no. 7, pp. 856–869, 2004.
- [38] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, Melbourne, Australia, 2013.
- [39] Network., Network Functions Virtualisation (NFV): Hypervisor Domain, ETSI, Sophia-Antipolis, France, Jan 2015.
- [40] "3GPP TS 32.2.297, charging data record (CDR) file format and transfer," Tech. Rep., ETSI, Sophia-Antipolis, France, 2016.
- [41] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3G cellular networks," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking (MobiCom '11)*, pp. 121–132, Las Vegas, Nev, USA, September 2011.
- [42] D. Pelleg and A. W. Moore, "X-means: Extending K-means with efficient estimation of the number of clusters," in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*, San Francisco, Calif, USA, Jun 2000.
- [43] 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*, pp. 1–9, Sydney, Australia, June 2014.
- [44] L. Chen, S. Zhou, and J. Xu, "Energy efficient mobile edge computing in dense cellular networks," in *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.
- [45] P.-S. Yu, J. Lee, T. Q. S. Quek, and Y.-W. P. Hong, "Traffic offloading in heterogeneous networks with energy harvesting personal cells-network throughput and energy efficiency," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1146–1161, 2016.
- [46] J. Wu, Y. Bao, G. Miao, S. Zhou, and Z. Niu, "Base-station sleeping control and power matching for energy-delay tradeoffs with bursty traffic," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3657–3675, 2016.

- [47] L. Haikun, X. Cheng-Zhong, J. Hai, G. Jiayu, and L. Xiaofei, "Performance and energy modeling for live migration of virtual machines," in *Proceedings of the 20th international symposium on high performance distributed computing*, California, Calif, USA, Jun 2011.
- [48] "Virtualization for small cells: Overview," Tech. Rep., Small Cell Forum, Draycott, England, 2015.
- [49] K. Li, "Performance analysis of power-aware task scheduling algorithms on multiprocessor computers with dynamic voltage and speed," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 11, pp. 1484–1497, 2008.
- [50] M. Shojafar, N. Cordeschi, and E. Baccarelli, "Energy-efficient adaptive resource management for real-time vehicular cloud services," *IEEE Transactions on Cloud Computing*, 2016.
- [51] C. Canali, L. Chiaraviglio, R. Lancellotti, and M. Shojafar, "Joint minimization of the energy costs from computing, data transmission, and migrations in cloud data centers," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 580–595, 2018.
- [52] M. Chen, W. Saad, and C. Yin, "Machine learning for wireless networks with artificial intelligence: a tutorial on neural networks," *IEEE Wireless Communications*, 2017, <https://arxiv.org/abs/1710.02913>.
- [53] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, 2017.
- [54] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Springer, 2010.
- [55] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2126–2136, 2013.
- [56] "Standard Performance Evaluation Corporation," Tech. Rep., SPEC, Virginia, USA, https://www.spec.org/virt_sc2013/results/res2013q2/.

Research Article

A Fuzzy Logic System for Vertical Handover and Maximizing Battery Lifetime in Heterogeneous Wireless Multimedia Networks

Thiago Coqueiro , José Jailton , Tássio Carvalho, and Renato Francês

Institute of Technology, Federal University of Pará (UFPA), PA, Brazil

Correspondence should be addressed to Thiago Coqueiro; thiago.sidonio@gmail.com

Received 21 September 2018; Revised 13 November 2018; Accepted 17 December 2018; Published 10 January 2019

Academic Editor: Gianluigi Ferrari

Copyright © 2019 Thiago Coqueiro et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bandwidth and power hungry applications are proliferating in mobile networks at a rapid pace. However, mobile devices have been suffering from a lack of sufficient battery capacity for the intensive/continuous use of these applications. In addition, the mobile ecosystem is currently heterogeneous and comprises a plethora of networks with different technologies such as LTE, Wi-Fi, and WiMaX. Hence, an issue must be addressed to ensure that quality of experience (QoE) is provided for the users in this scenario: an energy-efficient strategy that is designed to extend the battery lifetime of mobile devices. This paper proposes an architecture which provides an intelligent decision-making support system based on Fuzzy Logic for saving the energy of mobile devices within an integrated LTE and Wi-Fi network. The simulated experiments show the benefits of the solution this architecture can provide by using QoE metrics.

1. Introduction

The increasing demand for new services, technologies, and content is changing the way users obtain access to the Internet. According to Cisco, by 2021, 74% of the mobile devices will generate 98% of the traffic data, and 78% of this will originate from video traffic [1]. The popularization of the use of multimedia applications, together with the increase in the number of mobile users, makes it essential to supply services with a high transmission rate and improved quality.

Both in the current climate and in the future, the wireless network environment will be based on the coexistence of multiple networks that provide access to a wide range of technologies such as Bluetooth, Wi-Fi, WiMAX, and LTE. This will be a place where mobile users, equipped with devices supporting multiple network interfaces, will be able to obtain access to multimedia services through different access networks by means of the radio. In other words, the heterogeneity of a wireless environment provides the opportunity to assess and select the best network from a range of others, on the basis of the required conditions of a multimedia service.

In light of this, handover is a procedure that allows a mobile device to be disconnected from a network so that it can be connected to another. Its goal is to allow mobile users to be always connected (ABC, Always Best Connected [2]) to a network, so that their application can keep operating while they are relocating between different places. When the decoupling and connection involve the same network technologies, the phenomenon is called horizontal handover, whereas vertical handover involves the use of different technologies.

Since users want a better multimedia experience in their mobile devices, the delivery of video of a high quality is a more challenging task in wired networks. This is also owing to restrictions and mobility behavior and the environment of heterogeneous wireless networks itself; however, it mainly lies in the challenge of meeting the required conditions for multimedia applications with regard to the transfer of data and ensuring low latency and an insignificant loss.

The decision of when and where to carry out the handover will depend on several factors or attributes such as the following: QoS (Quality of Service), RSS (Received Signal Strength), bandwidth, the battery consumption rate, and

mobile user speed. The concept of Quality of Experience (QoE) is becoming a key factor because it can measure the degree of quality of a multimedia service through the perception of the user. In other words, the satisfaction of the user can be measured through required conditions based on social psychology, cognitive science, and engineering science [3].

Expectations about satisfaction for different services and application vary among different users. The traditional concept of QoS fails to take account of the fact that the satisfactions of the user should be used as an indicator, or, rather, it is only concerned with the network properties through metrics designed for the delivery of content [14–16]. This means that QoE should be an important attribute to take into account in the handover decision-making process.

Another key factor that needs to be considered is the level of power consumption of the mobile device, since video applications consume a large amount of energy. Today, energy saving is a concern that must be addressed, particularly in heterogeneous wireless networks which provide a wide range of opportunities to choose networks that allow improved energy saving through vertical handover. However, it is a real challenge to select an ideal network that takes account of the users' preferences and is, at the same time, energy efficient.

The different types of technology have different bandwidth as well as different power consumption. Therefore, there is a need for a balance between bandwidth and energy consumption, because there will be times when the user opts for a network with more bandwidth, reducing the battery life, as there will be times when the user will choose to migrate to a network with less bandwidth but with a longer battery life [17, 18]. For this reason, the Fuzzy System will have as one of its inputs the battery consumption.

This paper creates an algorithm for vertical handover decision-making based on fuzzy logic and which takes account of the following when selecting the best network: QoE criteria, energy consumption and the mobility of the user.

The paper is structured in the following way: The Section 2 provides a brief overview of studies related to vertical handover; The Section 3 describes the architecture of the fuzzy logic system for heterogeneous wireless networks; The Section 4 provides the analysis of the results and the Section 5 summarizes the conclusion of the study and makes suggestions for further research in the field.

2. Related Works

This section provides an overview of several related works on heterogeneous architectures that provide QoE support, together with seamless mobility for mobile users of different technologies or energy-saving strategies for battery devices.

There are several researches in the literature such as [4] that proposes a handover mechanism based on the coordination between MIH (Media-Independent Handover) and PMIPv6 (Proxy Mobile IPv6) to support user mobility.

The main focus of the research was the reduction of failed handovers, packet loss, and QoS requirements. However, the proposal does not consider support for energy saving and QoE. The work [5] proposes an algorithm that uses the parameters Time-to-Trigger (TTT) and Hysteresis Margin. Such parameters offered improvements in energy efficiency and Ping Pong Handover Ratio system. However, the proposal itself does not take into consideration the quality of user experience in its mechanism.

The Proposals [6, 7] Decision mechanisms use strategies based on Multiple Attributes Decision-Making (MADM). The proposal [6] developed a handover decision algorithm with energy efficiency support based on the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), which has the function of offering a rank of alternatives on the best networks for the selection. This algorithm uses as parameters power consumption, traffic class, and battery level of each network interface of the mobile device. The work of [7] presents a comparison of different MADM methods, considering the battery level standard. The main methods evaluated were AHP (The Analytic Hierarchy Process), ANP (Analytical Network Process), Fuzzy AHP, and Fuzzy ANP, where such methods were combined in 120 combinations for evaluation, in which the research results conclude that the best methods of combinations were Euclidian-normalization-TOPSIS-FANP and Sum normalization-GRANFANP. However, the proposals [6, 7], even considering the energy consumption, could not enable a relation with the standard of QoE.

The paper [8] proposes the use of the fuzzy system to handover decision, and the strategy was the combination of QoS parameters and QoE (Mean Opinion Score) indicators. The proposal [9] uses the QoE to select the best connection; the QoE was evaluated using Mean Opinion Score (MOS) in real time, through the PSQA (Pseudosubjective Quality Assessment) technique based on statistical learning through RNN (Random Neural Network). In [10] a Software-Defined Networking (SDN) architecture was proposed to implement a handover decision strategy based on fuzzy system considering QoS and QoE requirements; the fuzzy system is able to monitor a set of APs (Access Points) for the selection of the best AP to the user. In [11], the proposal is a multicriteria algorithm combining fuzzy system and utility function as a decision strategy. The proposal uses the fuzzy system to support the input imprecision information, while the utility function is responsible for reducing the number of handovers. The papers [8, 10], though proposing improvements to QoE, do not provide support for energy consumption. Although the proposal [11] considers multiple criteria such as delay, available bandwidth, and received signal strength, however the QoE and power consumption were not considered.

In [12], we proposed the vertical handover decision algorithm, Artificial Neural Networks (ANN), which uses a learning method based on neural network. The algorithm considers the parameters of QoE and QoS in its decision mechanism but does not have support in relation to the user satisfaction. Despite the collected results, the ANN obtained a reduced number of handoffs performed, as well as the reduction of the delays. The proposal [13] uses a QoE evaluation

TABLE 1: Related Works.

Proposal	QoE	Energy Efficiency	Decision Strategy	Proposal Focus
[4]	No	No	Coordination between MIH and PMIPv6	Reduce number of Ping Pong events and handover failures
[5]	No	Yes	Time-to-Trigger (TTT) and Hysteresis Margin	Improve energy efficiency and Reduce Ping Pong Handover Ratio
[6]	No	Yes	MADM - TOPSIS	Reduce power consumption
[7]	No	Yes	MADM - Euclidian-normalization-TOPSIS-FANP and Sum normalization-GRA-FANP	Support for reducing energy consumption
[8]	Yes	No	Fuzzy Logic System	Selection of the best network using QoE
[9]	Yes	No	RNN (Random Neural Network)	Selection of the best network using QoE
[10]	Yes	No	SDN and Fuzzy Logic System	QoS and QoE support
[11]	No	No	Fuzzy Logic System and Utility Function	Reduction in the number of handoffs
[12]	Yes	No	Artificial Neural Networks (ANN)	Reduction in the number of handoffs and delay
[13]	Yes	Yes	RNN (Random Neural Network)	Correlation between QoE and QoS in heterogeneous networks
Current Proposal	Yes	Yes	Fuzzy Logic System	Selecting the best network considering trade-off between QoE and energy efficiency

mechanism based on RNN in order to search the mapping relationship between QoS values and MOS (Mean Opinion Score) values. In addition, a QoE-Q was proposed as a vertical handover algorithm, considering Q-learning theory, to maximize user experience quality. Simulation results point to an increase in QoE performance as well as improvements in mobile device energy consumption. However, the proposal has, as main focus, a vertical handover mechanism based on the correlation between QoE and QoS in heterogeneous networks.

Therefore, none of them provides a joint approach that involves a solution for both energy-saving and vertical handover with QoE support. The proposal of this article is to propose a handover decision mechanism which provides the choice of the network taking into consideration the trade-off between energy consumption and QoE. Table 1 compares the related works in relation to the current proposal.

3. A Fuzzy Logic System for a Heterogeneous Wireless Network Architecture with an Energy-Efficiency Support System

This section presents a proposal with a Fuzzy System for seamless handover and energy-efficient support for mobile multimedia communication. The objective of the proposed architecture is to save battery in heterogeneous environment

networks with multiple devices as expected for Future Internet environments (see Figure 1).

3.1. Problem Statement and Major Contributions. One of the main contributions made by this article is to examine a dynamic heterogeneous architecture composed of an LTE technology (which is currently being adopted as the standard for 4G networks) and also formed of the IEEE 802.11n technology. This dynamic heterogeneous architecture is mobile-oriented to Future Internet environments since it will offer multiple connectivity options and allow the user to choose the best network for a particular period of time. Even if the user moves away from his starting place to a new destination point, the architecture will continue to offer multiple connectivity options.

When the mobile device is within the area of coverage of two or more networks, it has to select the network to which it will be connected, although the devices do not have any kind of mobile support or intelligence that can assist in its decision-making. Not even the technology standards can provide any kind of support for making decisions with regard to the selection and connection of the network.

At first, the only adopted criterion is the RSSI, or, in other words, the mobile device will only be connected to the nearest access point or base station. As a result, the lack of support for decision-making when selecting the network can cause the user problems since a mobile device with low battery

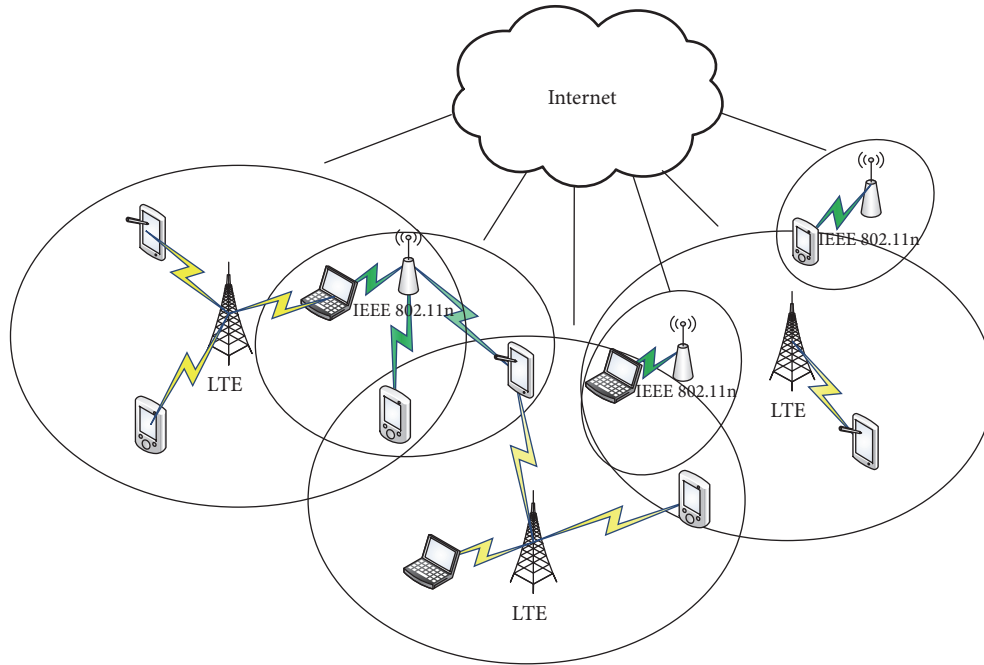


FIGURE 1: A Heterogeneous Wireless System Network.

power might be connected to a network which increases its energy consumption and hence reduces its battery life. Similarly, a user with a battery that has a prolonged life might be connected to a network that is saturated and offers a low bandwidth, and this can damage the user's application.

User mobility can also impair the quality of experience since users with high mobility can carry out excessive and unnecessary handovers, as in the case of the ping-pong handover. This means a mobile device will be connected to a new network even though it will only remain within the area of coverage of this new network for a short time, on account of its high mobility; this means that it will have to reestablish connection with the point of contact or previous base station.

Owing to the lack of mobile support or intelligence in the devices, another contribution made by this article is the Fuzzy Logic System which is oriented to heterogeneous wireless environments to assist decision-making with regard to the selection and connection of the network. However, the technologies must have different features before they can become a heterogeneous environment and, for this reason, this article puts forward a Fuzzy Logic System with two sets of rules. If the newly detected network is a Wi-Fi network, the Fuzzy system will be governed by a set of rules designed for this technology, whereas if the newly detected network is an LTE network, the Fuzzy system will also be governed by a set of rules designed for this technology.

The Fuzzy Logic System that is guided by two sets of rules that are applicable to each technology will allow the rational use of the mobile device. This system will prevent a mobile device with low energy from being connected to a network with high energy consumption and thus reduce the battery life. Similarly, the Fuzzy Logic System will prevent a device with a high battery capacity from being connected

to a network that is saturated and hence has a low bandwidth; in this way, the Fuzzy System can also carry out load balancing by distributing the mobile devices among the networks. In addition, the Fuzzy System will be able to prevent high-mobility devices from being either rapidly connected to the network or disconnected and hence avoid the ping pong handover. For this reason, Fuzzy is unlike traditional systems, since it was implemented with particular technological features in mind and incorporates three input variables: mobility, battery level and Quality of Experience.

The technologies have different battery consumption and this needs to be taken into account; for this reason, this article makes another contribution which is the use of a realistic battery model that has specific battery discharge parameters for each technology. The Wi-Fi and LTE technologies have different battery levels and as well as this, each mobile device status also has a different level of battery consumption. This is because the battery discharge rate of a mobile device which is transmitting messages is different from when it is receiving them and in addition, is different when the mobile device is idle. This explains why the battery model used in the Simulator was Rakhmatov-Vrudhula, since the model traditionally used in the simulators is the linear model which has a single battery discharge rate regardless of its technology or status.

Another contribution made by this article is the use of Quality of Experience as an input metric of the Fuzzy Logic System. A measurement algorithm that could be used for Quality of Experience was implemented in the base stations and access points. The mobile device provides information (by means of network signaling messages) to the base station and access point to which it is connected, together with the level of quality of experience being offered. This measurement

is derived from the number of video packets received by the mobile device, and this number allows the level of QoE being offered, to be classified.

3.2. The Fuzzy Rules Architecture. Fuzzy logic is an extension of conventional logic (binary). Unlike traditional logic that works with exact values, fuzzy logic allows a degree of uncertainty, by seeking to manipulate imprecise terms that are normal in human language. This enables a digital system to be designed, which only works with binary logic, can mimic human thought and is able to process information in a subjective way.

In a fuzzy system, the result appears in a range from 0 to 1. The value 0 means complete exclusion and a value 1 means full membership. The other values represent intermediate degrees of relevance. A certain element may also belong to two or more fuzzy sets that are defined in the same universe, where the values of the membership functions for each fuzzy set may be different. Thus, an element can belong more to one fuzzy set and less to others.

The Fuzzy sets are usually defined in linguistic terms. Where the values represent the height of an individual, there may be three fuzzy sets that are defined as follows: low, medium, and high. If, for example, there are five fuzzy sets in the universe, one possibility would be as follows: very low, low, medium, high, and very high.

In this paper, three metrics were chosen as input: mobile speed, the remaining battery power and QoE (Quality of Experience). These metrics are used for the dynamic operations during the simulations. Mobile speed is also a very important metric, since a user with a high speed can be degraded if handover is performed within a small coverage area of the network. The QoE value indicates the video quality of a link network with a current Base Station/Access Point and it is an important metric to trigger the handover. The battery consumption varies depending on which network is selected, and there will be situations when the selected network should not only take into account the QoS/QoE requirements, but also the remaining degree of energy of the device.

It was decided to divide the speed (m/s), into three sets: low speed (interval [0 to 6]), medium speed (interval [4 to 14]), and high speed (when the speed is higher than 12m/s). Three sets were also defined for the QoE (dB): Low (interval [0 25], medium (interval [22 31] and high (when the QoE is higher than 31dB). The battery power (in percentage terms) is defined in three sets: Low [0%–25%], medium [20%-80%], high [more than 70%]. These three fuzzy sets generate four outputs: NOT to make handover, PROBABLY NOT to make handover, PROBABLY YES to make handover and YES to make handover (see Figure 2).

The architecture consists of two technologies: Wi-Fi (the IEEE 802.11n) and 4G (the LTE); the Fuzzy System will receive the necessary information to decide what is the best connection for that moment. Precisely because the architecture is behaved by two technologies, the Fuzzy System has two sets of specific rules for each one that will be explained below.

The Fuzzy Rules determine when the mobile user will change the network (handover). When the mobile user

TABLE 2: Wi-Fi Rules.

Speed	QoE	Energy	Handover
Low	Medium	Medium	ProbablyYes
Low	Medium	High	ProbablyYes
Medium	Medium	Medium	ProbablyYes
Medium	Medium	High	ProbablyYes
Low	Low	High	Yes
Low	High	Low	Yes
Medium	Low	Low	Yes
Medium	Low	Medium	Yes
Medium	Low	High	Yes
Medium	Medium	Low	Yes
Medium	High	Low	Yes
High	Low	Low	Yes
High	Low	Medium	Yes
High	Low	High	Yes
High	Medium	Low	Yes

TABLE 3: LTE Rules.

Speed	QoE	Energy	Handover
Low	Medium	Medium	ProbablyYes
Low	Medium	High	ProbablyYes
Medium	Medium	Medium	ProbablyYes
Medium	Medium	High	ProbablyYes
Low	Low	Low	Yes
Low	Low	Medium	Yes
Low	Low	High	Yes
Low	Medium	Low	Yes
Low	High	Low	Yes
Medium	Low	Low	Yes
Medium	Low	Medium	Yes
Medium	Low	High	Yes
Medium	Medium	Low	Yes
Medium	High	Low	Yes
High	Low	Low	Yes
High	Low	Medium	Yes
High	Low	High	Yes
High	Medium	Low	Yes

is connected to a Wi-Fi network and detects a new LTE network, the Fuzzy System will trigger the handover in the situations (Table 2):

In general, a mobile user connected to the Wi-Fi network will change to an LTE network when in Wi-Fi network the QoE is Low, or the energy level is medium/high and it has a high speed. This situation prioritizes the quality of experience for the mobile users.

When the mobile user is connected to an LTE network and detects a new Wi-Fi network, the Fuzzy System will trigger the handover in the situations (Table 3).

In general, a mobile user connected to the LTE network will change to a Wi-Fi network when in LTE network the QoE

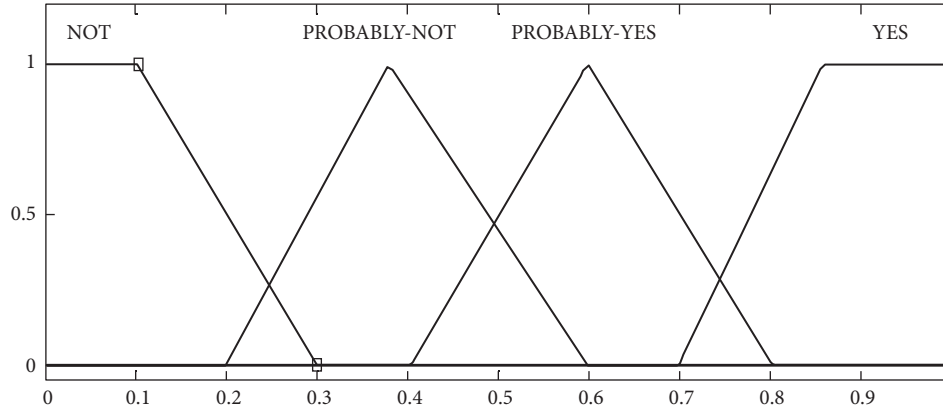


FIGURE 2: Output Fuzzy System.

is Low, or energy level is low/medium and it has a low speed. This situation prioritizes the energy level for the mobile users.

3.3. The Rakhmatov–Vrudhula Model. The purpose of a mathematical model for batteries is to study/predict their performance or behavior. A battery is an assembly of electrochemical cells, each of which consists of two electrodes: one with negative polarity nodes and a cathode which has positive polarity. The electrochemical reactions generate electrons that are released to provide electric power to the devices causing the battery drain [19], the battery model used in this paper was Rakhmatov–Vrudhula model.

The Rakhmatov–Vrudhula battery model [19] is more realistic than the linear model because it takes account of the transmission states, and for each state there is a different type of electric discharge (see (1)). This model includes different battery capacity and rates of recovery (in idle mode it is possible to increase the lifetime of the battery) for different types of batteries (alkaline, lithium ions).

$$\alpha = \sum_{k=1}^n 2I_{k-1}A(L, t_k, t_{k-1}, \beta) \quad (1)$$

where, I_{k-1} = the current discharge during the period $k - 1$, A = calculates the discharge rate of the non-linear battery model, L = battery lifetime, t_k = duration time of k period, and t_{k-1} = duration time of $k - 1$ period.

3.4. The Distributed Mobility Management (DMM) Scheme. In the application scenario, the handover decision-making is shared between the mobile station and point of attachments (POA) (e node B or access point). The Figure 3 illustrates the vertical handover process between the current and the target network in the proposed network architecture.

(1) When the MS detects a targeted neighbor network, it sends a *Link_Detected* message to the target network, and this message allows the targeted network to recognize the mobile station. (2) The targeted network replies with a *Link_Parameters_Report* message, which contains network information, such as RSSI and the level of Quality of Experience being offered by the new network. (3) The MS passes

the energy, the QoE and mobility information to the Fuzzy System (Wi-Fi or LTE rules on the network detected). The fuzzy output states whether the handover is necessary or not, the decision is made based on the value of output inference, the network that presents the highest value of inference will be the one chosen by the mobile device. (4) After the Fuzzy decision, the MS sends a *Handover_Initiate* message to the targeted network to trigger the vertical handover. (5) The MS sends the current network a *Link_Down* message about the need to change to a new network.

4. Evaluation of the Architecture

This section evaluates the architecture designed to provide QoE with maximized battery power. The performance assessment was carried out through Network Simulator 2 (NS2) and Evalvid Tool (to transmit the video in the simulation) [20]. The objective is to demonstrate the benefits of the proposed architecture in relation to the original architecture, as well as to compare the performance of the proposal with other papers researched in the literature. The proposal has been compared in relation to papers [8–11]. The proposals [8–11] were adapted for the NS-2 simulator to the performance comparison. The mobility model used was Random Way Point, so the movements and speed of the mobile users were random in the simulation.

The video used in this simulation was “Sintel”, which consists of 1253 frames with the YUV format, sampling 4:2:0 and dimensions of 1080x720, which was compressed through a MPEG-4 CODEC and sent at a 30 frame/s rate [21]. The video was chosen because the “Sintel” video is high definition.

The simulated parameters that were configured for all the experiments, are shown in Table 4 and represent the normal values of IEEE 802.11n and LTE networks.

The simulation parameters for the evaluation of energy consumption are described in Table 5. The initial battery voltage was 1000 joules and different power transmission and reception charges were adopted for each technological system. In the case of Wi-Fi technology, the same values were adopted as in [21] and the values used in [22] for LTE.

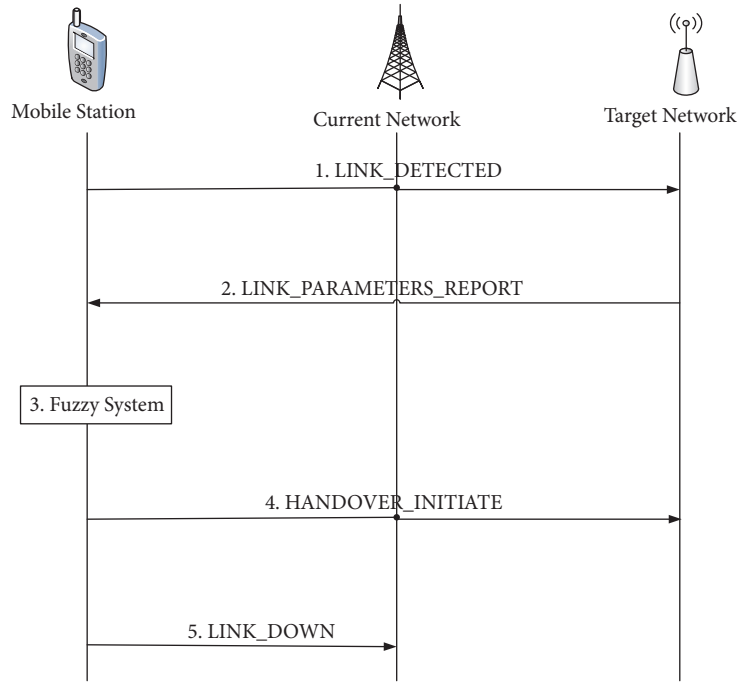


FIGURE 3: Handover signaling.

TABLE 4: Simulated Parameters.

	IEEE 802.11n	LTE
Rate transmission	108 Mbps	150 Mbps
Coverage area	100 m	1000 m
Videos	Resolution: 176 x 144 CIF Resolution: 352 x 288 CIF Resolution: 1080 x 720 CIF Frames rate: 30 frames/s Colour Mode: Y, U, V	
Queue	Drop Tail (40 ms delay)	
Packet size	1052 bytes	
Maximum Fragmentation packet	1024 bytes	
Number of simulations	50	
Confidence interval	95 %	
Number of videos	3	

TABLE 5: Energy Simulated Parameters.

	Transmission (W)	Reception (W)
WiFi	1.3	0.9
LTE	2.5	1.7

4.1. The Power Consumption Results. Initially, the results were used to compare the battery life when the mobile user is connected in the architecture without Fuzzy System (the architecture with original protocols) with the mobile user connected in the architecture with Fuzzy System, during the simulations the mobile user performs handovers.

The mobile user in the architecture without Fuzzy System (Original Architecture) was connected for approximately 47 minutes, in the paper [11] was 48,5 minutes, in the paper [10] was 50,3 minutes, in the paper [9] was 51,2 minutes, in the paper [8] was 52 minutes and while the mobile user with an energy-saving policy was connected for 58 minutes (Figure 4). There is a gain in energy because with the Fuzzy System the handover in the network takes place when the battery level of the device is low. In this case, there will be a loss in quality for the user but an increase in the lifetime of the battery.

4.2. QoS Results. This paper assesses the benefits of the scheme in terms of the network throughput rate. In the first situation, the mobile device was within the intersection of areas of coverage, while in the simulation in the original architecture without the Fuzzy Logic System, in the papers [8–11], the mobile device chose the network that was nearest even though it was saturated and offered a low quality of service. However, in the simulation in the architecture with the fuzzy system, the mobile device chose the network that offered the best quality of service. In Graph (Figure 5), it can be seen that the network throughput rate in the architecture with the fuzzy system was superior. The average throughput rate in the original architecture was 0,51 Mbps, in the paper [11] was 0,65Mbps, in the paper [10] was 0,76Mbps, in the paper [9] was 1,6 Mbps, in the paper [8] was 1,8 Mbps and while in the architecture with the fuzzy system the average throughput rate was 2,23 Mbps.

In the second situation, the simulations had mobile devices with high mobility within the coverage area of the networks. In a scenario where the original architecture, the

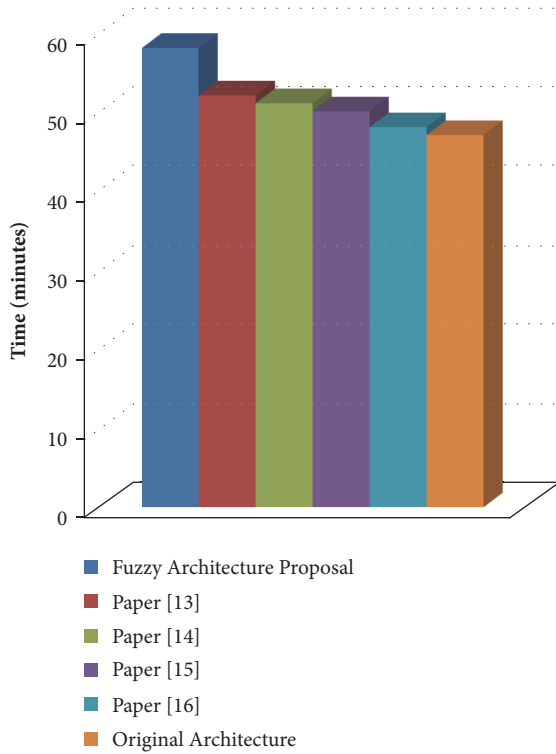


FIGURE 4: Energy-Saving over time.

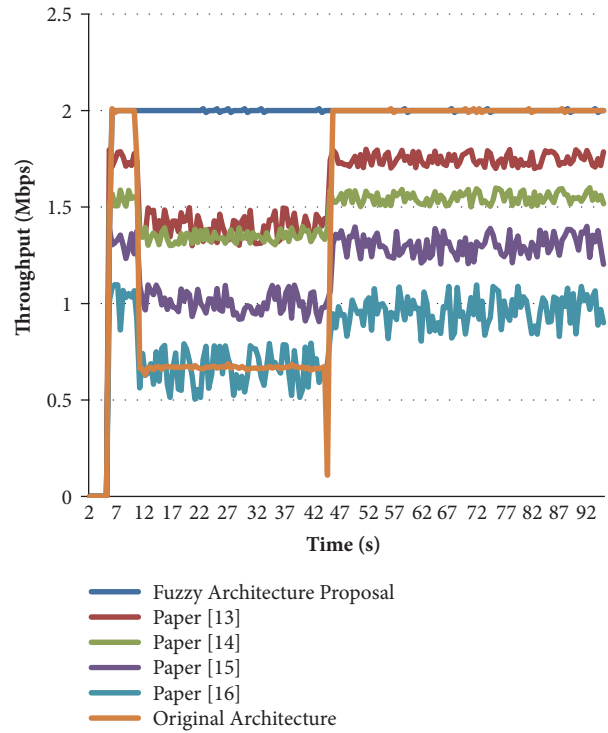


FIGURE 6: Throughput in ping-pong handover situation.

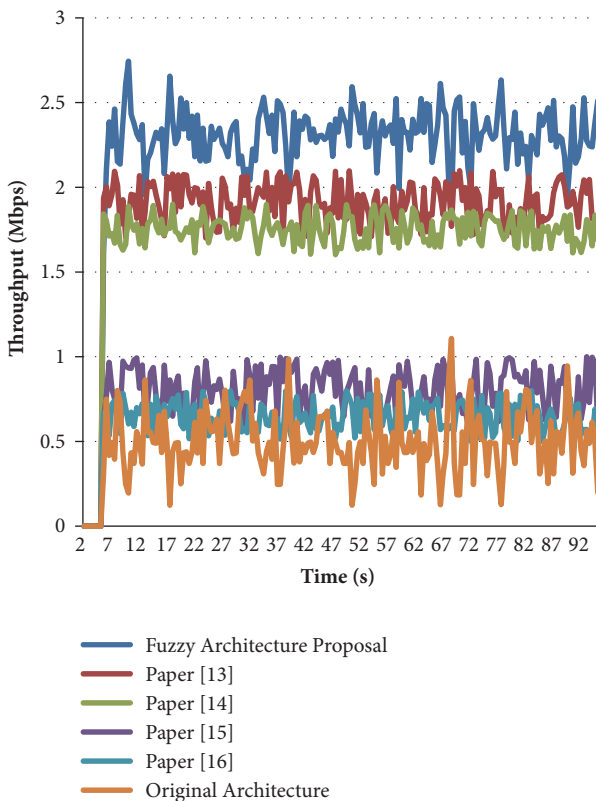


FIGURE 5: Throughput over Time.

papers [8–11] had no mobile support for the selection of a network, the high-mobility devices carried out the ping-pong handover. In other words, they were connected to a new network, (which was offering less bandwidth) and were later connected again to their previous network. This unnecessary exchange of network impaired the customer flow rate and for this reason, in a situation where the architecture has a fuzzy logic system, the high-mobility mobile devices failed to carry out the ping-pong handover or maintain its quality of service. In Graph (Figure 6), it can be noted that there is a loss of flow in the high-mobility mobile devices which carried out the ping-pong handover.

4.3. QoE Results. This paper also analyses the simulations with a video application. The video results are evaluated by means of objective QoE metrics: (i) Peak Signal to Noise Ratio (PSNR), (ii) Structural Similarity Metric (SSIM) and (iii) Video Quality Metric (VQM). The data are collected by using the MSU Video Quality Measurement Tool (VQMT). The PSNR is a traditional objective method to estimate the standards of multimedia services based on the opinions of the user. The SSIM index is a decimal value between 0 and 1, where 0 means there is a zero correlation with the original image, and 1 means exactly the same image. The VQM determines the level of multimedia quality based on human eye perception and subjective factors, including blurring, global noise, block distortion and colour distortion. The results of the VQM estimates range from values of 0 to 5, where 0 is the best possible score. In accordance with the network parameters, the Fuzzy system will keep the user longer in the network that offers the best quality.

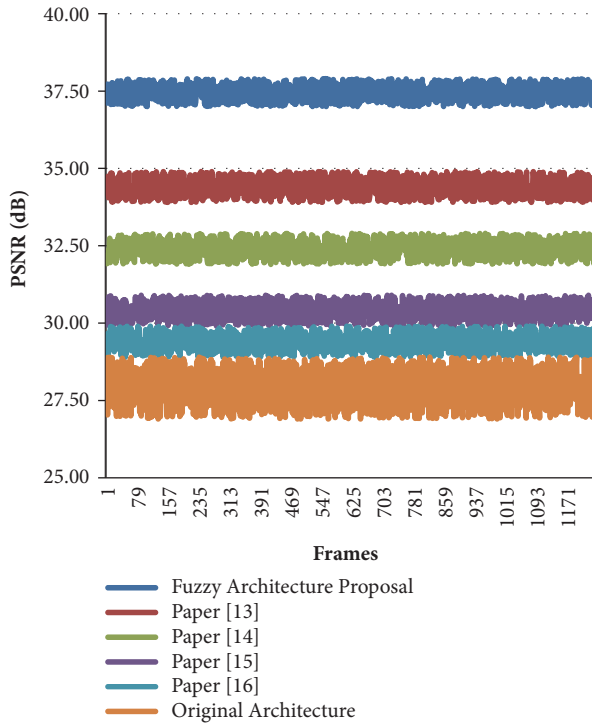


FIGURE 7: PSNR over Frames.

The videos transmitted with the Fuzzy System were superior to those transmitted without the Fuzzy System, to the papers [8–11]. The PSNR value for a video without the Fuzzy System was 28dB which can be rated as a fair video, the PSNR value for video in paper [11] was 29dB and rated as fair video, the PSNR value for video in paper [10] was 30dB which can be rated as a fair video, the PSNR value for video in paper [9] was 32dB and rated as good video, the PSNR value for video in paper [8] was 34dB which can be rated as a good video, whereas the PSNR value for the video with the Fuzzy System was 37dB and rated as an excellent video. The video transmitted with the fuzzy system had a better performance during the transmission and kept a balance between energy and QoE (see Figure 7).

The SSIM value for a video without the Fuzzy System was 0.71, the SSIM value for video in paper [11] was 0.73, the SSIM value for video in paper [10] was 0.79, the SSIM value for video in paper [9] was 0.81, the SSIM value for video in paper [11] was 0.85 and while the SSIM value for the video with the Fuzzy System was 0.93 (see Figure 8).

The VQM value for a video without the Fuzzy System was 3.95, the VQM value for video in paper [11] was 3.02, the VQM value for video in paper [10] was 2.8, the VQM value for video in paper [9] was 2.41, the VQM value for video in paper [8] was 2.1 and the VQM value for the video with Fuzzy System was 1.44 (see Figure 9).

Figure 10 shows that the video transmitted with the Fuzzy System had a better performance, since the delay was less than the video transmitted without the Fuzzy System, in the papers [8–11]. The lower the delay, the faster and more efficient the delivery of the frames, which confirms the superiority of the

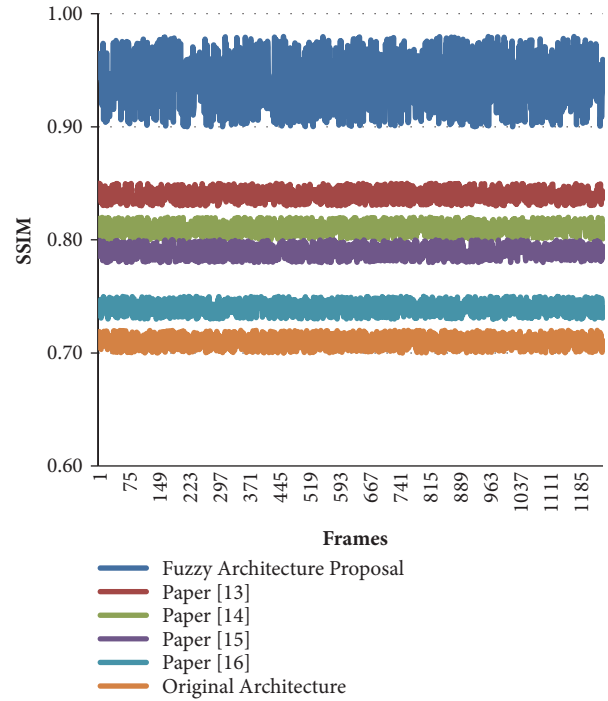


FIGURE 8: SSIM over Frames.

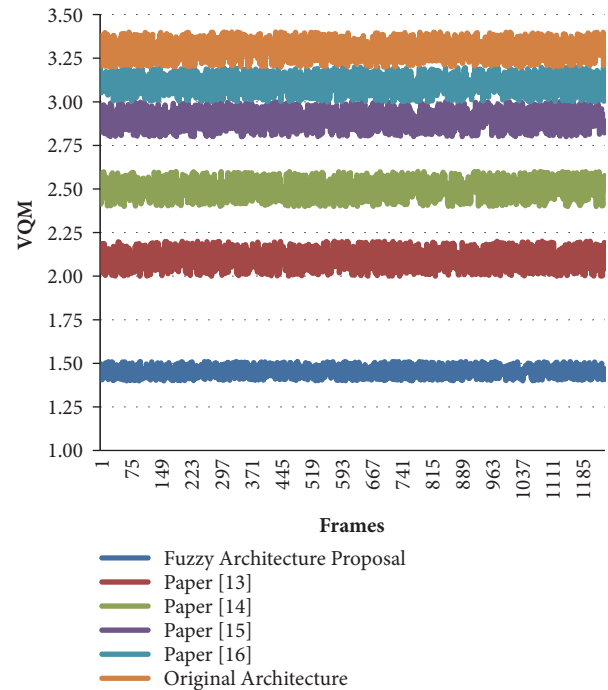


FIGURE 9: VQM over Frames.

video transmitted with the fuzzy scheme by means of the QoE metrics (Figure 10).

The superiority of the video transmitted with the Fuzzy System can also be determined by making a visual comparison of the frame transmitted with the mechanism (Figure 16), and with a frame of the video transmitted without the

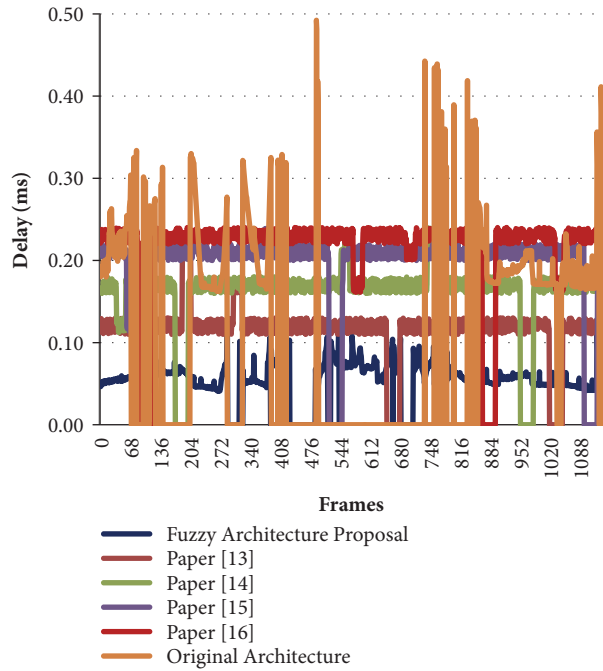


FIGURE 10: Delay over Frames.



FIGURE 11: Without the Fuzzy System.

mechanism (Figure 11), in the paper [11] (Figure 12), in the paper [10] (Figure 13), in the paper [9] (Figure 14), in the paper [8] (Figure 15).

5. Conclusion and Suggestions for Future Work

The mobile devices for the new heterogeneous wireless communication architecture need to be connected to networks that are able to provide the best quality of experience. However, there are occasions when it is important is to save energy and extend the battery life, even if this reduces the quality, to ensure continuity and avoid an abrupt termination of the services. This paper proposes a Heterogeneous Wireless

System formed of IEEE 802.11n and LTE networks that makes use of Fuzzy Rules to support an energy-efficient approach for saving battery power, while keeping QoE at satisfactory levels. Simulation evaluations show the benefits of this intelligent-based middleware solution for energy-efficient seamless vertical handover.

In future studies, the architecture will include new technologies, inputs for the Fuzzy Systems, and battery models, as well as dynamic scenarios with mobile users competing in IEEE 802.11n and LTE networks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.



FIGURE 12: In the Paper [11].



FIGURE 13: In the Paper [10].



FIGURE 14: In the Paper [9].



FIGURE 15: In the Paper [8].

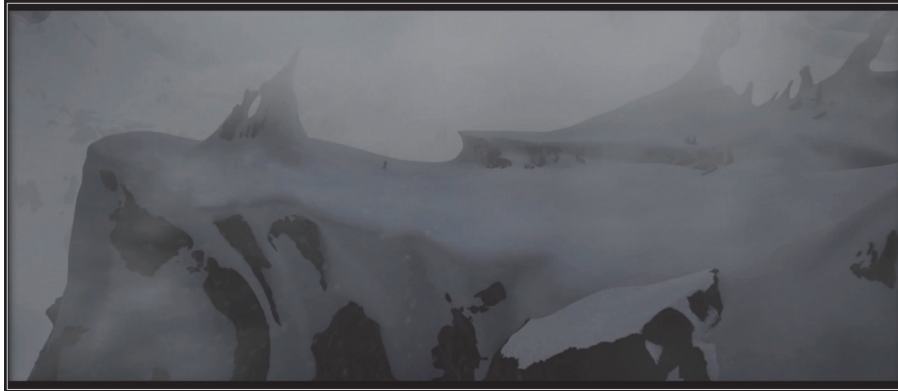


FIGURE 16: With the Fuzzy System.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by CNPQ, CAPES and UFPA.

References

- [1] CISCO, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021," white paper, available online: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf> (accessed 28.03.2018).
- [2] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications Magazine*, vol. 10, no. 1, pp. 49–55, 2003.
- [3] L. Zou, R. Trestian, and G.-M. Muntean, "E3DOAS: Balancing QoE and Energy-Saving for Multi-Device Adaptation in Future Mobile Wireless Video Delivery," *IEEE Transactions on Broadcasting*, vol. 64, no. 1, pp. 26–40, 2018.
- [4] A. A. Mansour, N. Enneya, and M. Ouadou, "A velocity-aware handover trigger in two-tier heterogeneous networks," *Future Internet*, vol. 10, no. 1, 2018.
- [5] B. Zhang, W. Qi, and J. Zhang, "An energy efficiency and ping-pong handover ratio optimization in two-tier heterogeneous networks," in *Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 532–536, Las Vegas, Nev, USA, January 2018.
- [6] R. Priyanka, S. Baghla, and M. Himanshu, "Energy Efficient Network Selection in Heterogeneous Network using TOPSIS".
- [7] M. Mansouri and C. Leghris, "A battery level aware MADM combination for the vertical handover decision making," in *Proceedings of the 13th IEEE International Wireless Communications and Mobile Computing Conference, IWCMC 2017*, pp. 1448–1452, Spain, June 2017.
- [8] A. B. Zineb, M. Ayadi, and S. Tabbane, "QoE-fuzzy VHO approach for heterogeneous wireless networks (HWNs)," in *Proceedings of the 30th IEEE International Conference on Advanced Information Networking and Applications, AINA 2016*, pp. 949–956, Switzerland, March 2016.
- [9] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin, "QoE-aware vertical handover in wireless heterogeneous networks," in *Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC '11)*, pp. 95–100, Istanbul, Turkey, July 2011.
- [10] O. Aldhaibani, F. Bouhaf, M. Makay, and A. Raschella, "An SDN-Based Architecture for Smart Handover to Improve QoE in IEEE 802.11 WLANs," in *Proceedings of the 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 287–292, Krakow, Poland, May 2018.
- [11] A. Kammoun and N. Tabbane, "Fuzzy utility decisional vertical handover algorithm for enhancing network performances," in *Proceedings of the 5th International Conference on Multimedia Computing and Systems, ICMCS 2016*, pp. 337–343, Marrakech, Morocco, October 2016.
- [12] A. B. Zineb, M. Ayadi, and S. Tabbane, "QoE-based vertical handover decision management for cognitive networks using ANN," in *Proceedings of the 2017 Sixth International Conference on Communications and Networking (ComNet)*, pp. 1–7, Hammamet, March 2017.
- [13] J. Chen, Y. Wang, Y. Li, and E. Wang, "QoE-Aware Intelligent Vertical Handoff Scheme Over Heterogeneous Wireless Access Networks," *IEEE Access*, vol. 6, pp. 38285–38293, 2018.
- [14] Y. Xu, R. Q. Hu, L. Wei, and G. Wu, "QoE-aware mobile association and resource allocation over wireless heterogeneous networks," in *Proceedings of the 2014 IEEE Global Communications Conference, GLOBECOM 2014*, pp. 4695–4701, USA, December 2014.
- [15] I. V. S. Brito and G. B. Figueiredo, "Improving QoS and QoE Through Seamless Handoff in Software-Defined IEEE 802.11 Mesh Networks," *IEEE Communications Letters*, vol. 21, no. 11, pp. 2484–2487, 2017.
- [16] L. Yala, P. A. Frangoudis, and A. Ksentini, "QoE-aware computing resource allocation for CDN-as-a-service provision," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, USA, December 2016.
- [17] M. F. Tuysuz, "An energy-efficient QoS-based network selection scheme over heterogeneous WLAN - 3G networks," *Computer Networks*, vol. 75, pp. 113–133, 2014.
- [18] M. F. Tuysuz and R. Trestian, "Energy-Efficient Vertical Handover Parameters, Classification and Solutions over Wireless Heterogeneous Networks: A Comprehensive Survey," *Wireless Personal Communications*, vol. 97, no. 1, pp. 1155–1184, 2017.
- [19] D. N. Rakhmatov and S. B. K. Vrudhula, "An analytical high-level battery model for use in energy management of portable

- electronic systems,” in *Proceedings of the International Conference on Computer-Aided Design 2001*, pp. 488–493, USA, November 2001.
- [20] “Evalvid,” <http://www.tkn.tu-berlin.de/research/evalvid/>, 2018.
- [21] L. Marie and M. Nilsson, “Investigating the energy consumption of a wireless network interface in an ad hoc networking environment,” in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 3, pp. 1548–1557, Anchorage, Alaska, USA, April 2001.
- [22] M. Deruyck, W. Vereecken, E. Tanghe et al., “Comparison of power consumption of mobile WiMAX, HSPA and LTE access networks,” in *Proceedings of the 9th Conference of Telecommunication, Media and Internet (CTTE '10)*, pp. 1–7, June 2010.

Research Article

Autonomous Transmission Power Decision Strategy for Energy Efficient Operation of a Dense Small Cell Network

Jaesung Park  and Heejung Byun 

School of Information and Communication, The University of Suwon, 17 Wauan-gil, Bongdam-eup, Hwaseong-si, Gyeonggi-do 18323, Republic of Korea

Correspondence should be addressed to Heejung Byun; heejungbyun@suwon.ac.kr

Received 27 September 2018; Accepted 13 November 2018; Published 25 November 2018

Guest Editor: Hideyuki Takahashi

Copyright © 2018 Jaesung Park and Heejung Byun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Smart interference management methods are required to enhance the throughput, coverage, and energy efficiency of a dense small cell network. In this paper, we propose a transmit power control for energy efficient operation of a dense small cell network. We cast the power control problem as a noncooperative game to satisfy the design requirement that small cells do not need any information exchange among them. We analyze the sufficient condition for the existence of a Nash equilibrium (NE) state of the proposed game. We also analyze that the NE state is unique by transforming the original nonlinear fractional programming problem into a nonlinear parametric programming problem. Through simulation studies, we verify our analysis results. In addition, we show that the proposed method achieves higher energy efficiency of a network and balances the energy efficiency among cells more evenly than the methods based on the AIMD (additive increase and multiplicative decrease) algorithm.

1. Introduction

The amount of data traffic transferred through wireless networks has been increased exponentially and the growth rate is expected to increase in the future [1]. Various efforts have been made to accommodate the ever increasing traffic demand [2]. Massive MIMO (Multiple Input and Multiple Output) and CoMP (Coordinated Multipoint/Reception) increase the spectral efficiency of a wireless link [3, 4]. New spectrums for wireless networks and new radio access technologies have been sought [5, 6]. Small cells are deployed to increase the coverage and capacity of a network through spatial reuse [7]. Among those, densifying small cells are regarded as a promising way to enhance the spectral efficiency of a network in a cost-effective way. However, the interference among cells also increases as the cell density increases. Thus, efficient interference management methods have been devised to fully exploit the advantage of a dense small cell network (DSCN) [8, 9]. In the meantime, not only the interference but also the energy efficiency (EE) of a DSCN

becomes one of the key design requirements to reduce the power consumption of a DSCN [10, 11].

Energy efficient DSCN operation methods can be classified into two categories. The methods belonging to the first category utilize that the density of small cells is high. Since a user equipment (UE) can be served by many cells, a DSCN can accommodate the traffic demands in an energy efficient way if some cells are running in a sleep mode [12–15]. Thus, the schemes in this category attempt to find an optimal set of active cells that maximizes the EE of a DSCN for a given spatiotemporal traffic distribution. The methods in the second category consider a set of active cells and try to use network resources in an energy efficient manner. Radio resource scheduling [16], association control [17], cell clustering [18], and transmit power control belong to this category. Usually, scheduling, clustering and association management operate in a large timescale while the transmit power control works in a small timescale. Thus, following the timescale separation approach [18, 19], we focus on the power control problem for energy efficient operation of a DSCN by

assuming that the clusters of active cells and the service order of UEs in each cell are determined.

For a single cell network, iterative algorithms are proposed in [20, 21] that determine an optimal transmit power maximizing EE. In [22, 23], a Markovian approach is taken for energy efficient power control. The authors in [24] use the parameter free fractional programming to derive an energy efficient power allocation to maximize the system energy efficiency and propose a power adaptation algorithm based on the analysis result. In [25], authors formulate a nonconvex optimization problem to maximize the energy efficiency of a cell. By reforming the problem into a convex optimization problem with the property of the parameter-free fractional programming and the concept of perspective function, they devise a distributed power control algorithm that requires the minimum amount of information to be exchanged among cells. A Nash bargaining cooperative game-theoretic framework is proposed in [26]. After showing the relationship between the energy efficiency and the spectral efficiency, they formulate the energy efficiency maximization problem. Then, they propose a distributed algorithm that gives a suboptimal solution and guarantees the efficiency and fairness. The authors in [27] propose an energy-efficient power allocation and wireless backhaul bandwidth allocation for a small cell network using OFDMA. They devise algorithms for the original non-convex problem so that each small cell can jointly determine the transmit power for serving UEs and the bandwidth for backhauling. In [28], a Stackelberg game model is adopted to increase the energy efficiency of small cell networks. The authors propose a pricing scheme between a macrocell and small cells and devise a power control method by transforming the original nonlinear fractional programming problem into a subtractive form.

However, we note that the nature of a DSCN is amorphous [29] because small cells are deployed in an unplanned manner by different entities. Thus, it is difficult to expect that small cells exchange information for cooperative control. In addition, each small cell is selfish and rational in that it attempts to increase its own EE in response to the environments. Therefore, in this paper, we design a power control method by which each cell determines its transmit power autonomously without any message exchange among cells. To achieve the goal, we model the power control problem for enhancing EE of a DSCN as a noncooperative game and propose a power control algorithm based on the best response function. Furthermore, we provide a sufficient condition for the existence of Nash equilibrium (NE) of the game. We also prove that the NE of the proposed power control game is unique.

The rest of the paper is organized as follows. In Section 2, we describe the system model and formally define the energy efficiency of a DSCN. In Section 3, we present the power control game and analyze its property. We especially prove that the proposed power control game has a unique NE. After we verify that the proposed method is superior to the other distributed method based on the additive increase and multiplicative decrease algorithm in Section 4, we conclude the paper in Section 5.

2. System Model

We consider the downlinks of a small cell network where the spectrum reuse factor is one. We assume that radio resources are divided into resource blocks which is the smallest resource unit that can be allocated to an UE. We assume the time is divided into a time slot with equal size.

We denote by \mathcal{C} the set of cells interfering with each other. We also denote the set of UEs served by a cell i by \mathcal{U}_i . Let us denote by p_i the transmit power of a cell i . If we denote by \mathcal{P}_i the set of transmit power that each cell can choose from (i.e., $p_i \in \mathcal{P}_i$), the SINR at an UE $u \in \mathcal{U}_i$ can be given as

$$\mathfrak{s}_{i,u} = \frac{p_i h_{i,u}}{\sum_{j \neq i, j \in \mathcal{C}} p_j h_{j,u} + N_0}, \quad (1)$$

where $h_{i,u}$ is the channel gain between a cell i and a UE u and N_0 is the noise power. Let us denote by $b_{i,u}$ the number of resource blocks that i allocates to u . We also denote by M_R the maximum number of resource blocks that a small cell has. If we assume that small cells use a round-robin scheduler [30], $b_{i,u}$ becomes $M_R/|\mathcal{U}_i|$, where $|X|$ is the cardinality of a set X . If the bandwidth of a resource block is W_R and we denote by $B_{i,u} = b_{i,u}W_R$, from the Shannon's capacity formula, the downlink data rate provided to a UE u by a cell i is expressed as

$$\mathbf{r}_{i,u} = B_{i,u} \log_2(1 + \mathfrak{s}_{i,u}). \quad (2)$$

Therefore, the throughput of a cell i can be given as

$$\mathfrak{R}_i = \sum_{u \in \mathcal{U}_i} \mathbf{r}_{i,u}. \quad (3)$$

From the measurement studies, the authors in [31] proposes a power consumption model for various cell types which is composed of a load-independent part and a load-dependent part. The resulting power consumption model of a cell is expressed as follows.

$$\mathfrak{P}_i = N_A P_0 + \Delta_p p_i, \quad (4)$$

where N_A is the number of transmit antenna, P_0 is the power consumption at the minimum non-zero output power, and Δ_p is the slope of the load-dependent power consumption.

Then, the energy efficiency of a cell i is defined as

$$\mathfrak{E}_i = \frac{\mathfrak{R}_i}{\mathfrak{P}_i} = \frac{\sum_{u \in \mathcal{U}_i} B_{i,u} \log_2(1 + \mathfrak{s}_{i,u})}{N_A P_0 + \Delta_p p_i}. \quad (5)$$

An optimization problem that attempts to find a transmit power vector maximizing the total energy efficiency of a network (i.e., $\sum_{i \in \mathcal{C}} \mathfrak{E}_i$) can be formulated and a central controller can find a globally optimal solution. In terms of implementation, all the channel conditions between all UEs and cells are needed to solve the optimization problem. However, the signaling overhead increases exponentially with the number of cells. Thus, the signaling overhead for a central controller to obtain the required channel information will be prohibitive in a DSCN where the number of cells is very large. In addition, since a DSCN is assumed to be composed of autonomous cells, it is unlikely that there is a central

controller gathering necessary information, solving the optimization problem, and distributing the optimal transmit power for each cell. Therefore, a distributed algorithm that enables each cell to determine an optimal transmit power using its local information is needed. To tackle the issue, in this paper, we propose a noncooperative power control game model for energy efficient operation of a DSCN even if each cell behaves in a self-interested way without any message exchange with other cells.

3. Power Control Game

The distributed power control game is defined as follows.

$$\Gamma = \{\mathcal{C}, \mathcal{S}, (g_i)_{i \in \mathcal{C}}\}, \quad (6)$$

where \mathcal{C} is the set of game players, \mathcal{S} is the set of power allocation profiles, and g_i is the utility function of a cell i . Since the power control game is played among small cells, the set of players is the set of small cells. For each cell i , \mathcal{P}_i is the set of strategies (i.e., transmit powers) that i can choose. Thus, a power allocation profile is a combination of strategies of all players $\mathcal{S} = \times_{i \in \mathcal{C}} \mathcal{P}_i$. A utility function of a player maps each strategy profile $\mathbf{p} = (p_1, \dots, p_{|\mathcal{C}|}) \in \mathcal{S}$ to a real number. Since the purpose of the game is to determine the transmit power of a cell to optimize the energy efficiency, we set the utility function of a cell as \mathcal{G}_i . We also denote the strategy profile except p_i by $\mathbf{p}_{-i} = (p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{|\mathcal{C}|}) \in \mathcal{S}_{-i} = \times_{j \in \mathcal{C}-\{i\}} \mathcal{P}_j$.

Then, in the noncooperative power control game Γ , each cell i attempting to maximize its utility faces the problem of determining its best response when other cells commit to play \mathbf{p}_{-i} . The best response of a cell i to the strategy profile \mathbf{p}_{-i} is a strategy $p_i^* \in \mathcal{P}_i$ such that $\mathcal{G}_i(p_i^*, \mathbf{p}_{-i}) \geq \mathcal{G}_i(p_i, \mathbf{p}_{-i})$ for all $p_i \in \mathcal{P}_i$. Since a cell i measures the influence of \mathbf{p}_{-i} on \mathcal{G}_i

during each time slot, it determines its transmit power for the t -th time slot $p_i^*(t)$ at the end of the $(t-1)$ -th time slot as

$$p_i^*(t) = \arg \max_{p_i \in \mathcal{P}_i} \mathcal{G}_i(p_i, \mathbf{p}_{-i}). \quad (7)$$

For the noncooperative game, we analyze the existence and the uniqueness of the Nash equilibrium (NE) as Propositions 1 and 2.

Proposition 1. *There exists an NE in the game Γ if $p_i < (2^{\Omega_{i,u}} - 1)/z_{i,u} \forall i \in \mathcal{C}, u \in \mathcal{U}_i$, where*

$$\Omega_{i,u} = \frac{1}{\log 2} \frac{x + \Delta_p p_i z_{i,u} (2\Delta_p + x z_{i,u} + 3\Delta_p z_{i,u} p_i)}{2\Delta_p^2 (1 + z_{i,u} p_i)^2}, \quad (8)$$

$$x = N_A P_0,$$

$$z_{i,u} = \frac{h_{i,u}}{\sum_{j \neq i, j \in \mathcal{C}} p_j h_{j,u} + N_0}.$$

Proof. We prove the existence of a NE in the game Γ by showing that \mathcal{P}_i is nonempty, convex, and compact subset in an Euclidean space and \mathcal{G}_i is continuous and quasi-concave in \mathcal{P}_i [32]. \mathcal{P}_i is the set of transmit power that a cell i can choose. Since the transmit power of a cell i is bounded by the minimum transmit power p_m and the maximum power p_M , \mathcal{P}_i is not empty, convex, and compact subset in an Euclidean space E^1 . In addition, \mathcal{G}_i is a continuous function of p_i .

To show the concavity of \mathcal{G}_i , we consider a UE $u \in \mathcal{U}_i$. Then,

$$\mathcal{G}_{i,u} = \frac{B_{i,u} \log_2(1 + z_{i,u} p_i)}{x + \Delta_p p_i}. \quad (9)$$

Therefore,

$$\begin{aligned} \frac{\partial \mathcal{G}_{i,u}}{\partial p_i} &= \frac{(B_{i,u}/\log 2)(z_{i,u}/(1 + z_{i,u} p_i))(x + \Delta_p p_i) - B_{i,u} \Delta_p \log_2(1 + z_{i,u} p_i)}{(x + \Delta_p p_i)^2} \\ &= \frac{B_{i,u} z_{i,u}}{\log 2 (1 + z_{i,u} p_i) (x + \Delta_p p_i)} - \frac{B_{i,u} \Delta_p \log_2(1 + z_{i,u} p_i)}{(x + \Delta_p p_i)^2}. \end{aligned} \quad (10)$$

We let

$$\begin{aligned} X_{i,u} &= \frac{z_{i,u}}{(1 + z_{i,u} p_i) (x + \Delta_p p_i)} \\ Y_{i,u} &= \frac{\log_2(1 + z_{i,u} p_i)}{(x + \Delta_p p_i)^2} \end{aligned} \quad (11)$$

Then,

$$\begin{aligned} \frac{\partial^2 \mathcal{G}_{i,u}}{(\partial p_i)^2} &= \frac{B_{i,u}}{\log 2} \frac{\partial X_{i,u}}{\partial p_i} - B_{i,u} \Delta_p \frac{\partial Y_{i,u}}{\partial p_i} \\ &= B_{i,u} \left(\frac{1}{\log 2} \frac{\partial X_{i,u}}{\partial p_i} - \Delta_p \frac{\partial Y_{i,u}}{\partial p_i} \right). \end{aligned} \quad (12)$$

Since

$$\frac{\partial X_{i,u}}{\partial p_i} = - \frac{z_{i,u} (\Delta_p + x z_{i,u} + 2\Delta_p z_{i,u} p_i)}{(1 + z_{i,u} p_i)^2 (x + \Delta_p p_i)^2} \quad (13)$$

and

$$\begin{aligned} \frac{\partial Y_{i,u}}{\partial p_i} &= \frac{1}{(x + \Delta_p p_i)^4} \left(\frac{1}{\log 2} \frac{z_{i,u} (x + \Delta_p p_i)^2}{1 + z_{i,u} p_i} \right. \\ &\quad \left. - \log_2(1 + z_{i,u} p_i) 2(x + \Delta_p p_i) \Delta_p \right) \end{aligned}$$

$$= \frac{1}{(x + \Delta_p p_i)^2} \left(\frac{1}{\log 2} \frac{z_{i,u}}{1 + z_{i,u} p_i} - \frac{2\Delta_p \log_2(1 + z_{i,u} p_i)}{x + \Delta_p p_i} \right) \quad (14)$$

thus (12) becomes

$$\begin{aligned} \frac{\partial^2 \mathfrak{G}_{i,u}}{(\partial p_i)^2} &= B_{i,u} \left(-\frac{z_{i,u}}{\log 2} \frac{\Delta_p + xz_{i,u} + 2\Delta_p z_{i,u} p_i}{(1 + z_{i,u} p_i)^2 (x + \Delta_p p_i)^2} \right. \\ &\quad - \frac{\Delta_p}{\log 2} \frac{z_{i,u}}{1 + z_{i,u} p_i} \frac{1}{(x + \Delta_p p_i)^2} \\ &\quad \left. + \frac{2\Delta_p^2 \log_2(1 + z_{i,u} p_i)}{(x + \Delta_p p_i)^3} \right) \\ &= \frac{B_{i,u}}{(x + \Delta_p p_i)^2} \left(-\frac{z_{i,u}}{\log 2} \frac{\Delta_p + xz_{i,u} + 2\Delta_p z_{i,u} p_i}{(1 + z_{i,u} p_i)^2} \right. \\ &\quad \left. - \frac{\Delta_p}{\log 2} \frac{z_{i,u}}{1 + z_{i,u} p_i} + \frac{2\Delta_p^2 \log_2(1 + z_{i,u} p_i)}{x + \Delta_p p_i} \right). \end{aligned} \quad (15)$$

Therefore, the condition that makes $\partial^2 \mathfrak{G}_{i,u}/(\partial p_i)^2 < 0$ becomes

$$\begin{aligned} \frac{2\Delta_p^2 \log_2(1 + z_{i,u} p_i)}{x + \Delta_p p_i} &< \frac{z_{i,u}}{\log 2} \frac{\Delta_p + xz_{i,u} + 2\Delta_p z_{i,u} p_i}{(1 + z_{i,u} p_i)^2} \\ + \frac{\Delta_p}{\log 2} \frac{z_{i,u}}{1 + z_{i,u} p_i} &= \frac{1}{\log 2} \frac{z_{i,u}}{1 + z_{i,u} p_i} \\ \cdot \frac{\Delta_p + xz_{i,u} + 2\Delta_p z_{i,u} p_i + \Delta_p (1 + z_{i,u} p_i)}{1 + z_{i,u} p_i} &= \frac{1}{\log 2} \\ \cdot \frac{z_{i,u} (2\Delta_p + xz_{i,u} + 3\Delta_p z_{i,u} p_i)}{(1 + z_{i,u} p_i)^2}. \end{aligned} \quad (16)$$

By rearranging (16), we obtain

$$\begin{aligned} \log_2(1 + z_{i,u} p_i) \\ < \frac{1}{\log 2} \frac{x + \Delta_p p_i}{2\Delta_p^2} \frac{z_{i,u} (2\Delta_p + xz_{i,u} + 3\Delta_p z_{i,u} p_i)}{(1 + z_{i,u} p_i)^2}. \end{aligned} \quad (17)$$

```

1: at the end of each time slot, each cell  $i \in \mathcal{C}$ :
2: Init:  $\epsilon > 0, \delta > \epsilon, n = 1$ 
3: while  $\delta > \epsilon$  do
4:    $\mathbf{p}_i(n+1) = \arg \max_{p_i \in \mathcal{P}_i} \mathfrak{G}_i(p_i(n), \mathbf{p}_{-i}(n))$ 
5:    $\delta = |\mathfrak{G}_i(p_i(n+1), \mathbf{p}_{-i}(n+1)) - \mathfrak{G}_i(p_i(n), \mathbf{p}_{-i}(n))|$ 
6:    $n = n + 1$ 
7: if  $p_i(n) \in \mathcal{P}_i$  then
8:    $p_i^* = p_i(n)$ 
9: else
10:  if  $\partial \mathfrak{G}_i / \partial p_i > 0$  then
11:    $p_i^* = p_M$ 
12:  else
13:    $p_i^* = p_m$ 

```

ALGORITHM 1: Transmit power decision algorithm based on the best response function.

If we let

$$\Omega_{i,u} = \frac{1}{\log 2} \frac{x + \Delta_p p_i}{2\Delta_p^2} \frac{z_{i,u} (2\Delta_p + xz_{i,u} + 3\Delta_p z_{i,u} p_i)}{(1 + z_{i,u} p_i)^2}, \quad (18)$$

$\partial^2 \mathfrak{G}_{i,u}/(\partial p_i)^2$ becomes negative if

$$p_i < \frac{2^{\Omega_{i,u}} - 1}{z_{i,u}}. \quad (19)$$

Since $\mathfrak{G}_i = \sum_{u \in \mathcal{U}_i} \mathfrak{G}_{i,u}$, $\partial^2 \mathfrak{G}_i/(\partial p_i)^2 < 0$ if $\partial^2 \mathfrak{G}_{i,u}/(\partial p_i)^2 < 0$ for all $i \in \mathcal{C}$ and $u \in \mathcal{U}_i$. Since \mathfrak{G}_i is a strict concave function in p_i if $p_i < (2^{\Omega_{i,u}} - 1)/z_{i,u}$, there exists an NE in the game Γ . \square

Proposition 2. An NE in the game Γ is unique.

Proof. The transmit power of a cell i is in \mathcal{P}_i which is a compact and connected subset of an Euclidean space E^1 . Both \mathfrak{R}_i and \mathfrak{P}_i are a continuous function of p_i and produce real numbers. In addition, since $\mathfrak{P}_i > 0, \forall p_i \in \mathcal{P}_i$, the nonlinear fractional programming problem that maximizes \mathfrak{G}_i (i.e., the problem in (7)) can be transformed into the following nonlinear parametric programming problem.

$$\begin{aligned} F(q) &= \max \{ \mathfrak{R}_i(p_i) - q \mathfrak{P}_i(p_i) \mid p_i \in \mathcal{P}_i \}, \\ &\quad \forall q \in E^1. \end{aligned} \quad (20)$$

Then, from the Lemma 4 in [33], $F(q) = 0$ has a unique solution q_0 . In addition, from the theorem in [33], q_0 is $\mathfrak{R}_i(p_i^*)/\mathfrak{P}_i(p_i^*) = \max \mathfrak{G}_i$ if and only if $F(q) = 0$. Therefore, the NE of the game Γ is unique. \square

The transmission power control algorithm based on the proposed noncooperative game model can be summarized as in Algorithm 1.

4. Performance Evaluation

In this section, we compare the performance of the proposed transmit power control method with the other method that do not require a cell to exchange messages with other cells when it determines its transmit power. One of the most popular methods that determines a control variable in a distributed manner is the AIMD (additive increase and multiplicative decrease) adopted in TCP for the congestion control. The basic idea of the AIMD is to increase the current control variable linearly if a node does not detect performance degradation. When a node detects performance degradation, it considers that the competition for the shared resources in a network is severe. Thus, the node decreases its control variable multiplicatively to reduce the competition level fast. Therefore, AIMD increases the performance of each node opportunistically while stabilizing a network by inducing an implicit cooperation among the competing nodes without explicit message exchanges among them.

In case of the transmit power control, the control variable of a node is the transmit power of a cell and the performance metric is the energy efficiency of a cell. Thus, when AIMD is used, each cell i determines the transmit power for the t -th time slot denoted by $p_i(t)$ based on $p_i(t-1)$, $\mathfrak{E}_i(t-1)$ and $\mathfrak{E}_i(t)$, where $\mathfrak{E}_i(t)$ is the energy efficiency of a cell i at the beginning of the time slot t when i uses $p_i(t-1)$. Specifically, if $\mathfrak{d}_i(t) = \mathfrak{E}_i(t) - \mathfrak{E}_i(t-1) < 0$, $p_i(t) = p_i(t-1)/2$. If $\mathfrak{d}_i(t) > 0$, $p_i(t) = p_i(t-1) + \mathfrak{a}$. We consider two types of \mathfrak{a} . One is $\mathfrak{a} = p_M/10$, and the other is $\mathfrak{a} = p_M - p_i(k-1)$. Henceforth, the AIMD using the former \mathfrak{a} will be referred as a static AIMD and the AIMD using the latter \mathfrak{a} will be called a dynamic AIMD.

We deployed $N_c = 40$ cells in $100m \times 100m$ area according to the homogeneous Poisson point process (HPPP) to reflect the uncoordinated cell deployments. In the same area, we positioned N_m UEs following the HPPP and associate each of them to its nearest cell. According to the 3GPP Pico base station system specification [30], we configure the system parameters of each cell as follows. The system bandwidth is 10MHz, the bandwidth of a resource block is 180kHz, an antenna height is set to be 10m, and the antenna gain is configured as 2dBi. The maximum and the minimum transmit power are set to $p_M = 24dBm$ and $p_m = -20dBm$, respectively. The path loss model of $38 + 30 \log_{10}(d)dB$ is used, where d is the distance between a sender and a receiver in meters. The parameters related to the cell power consumption are configured according to the Pico cell parameters in [31]. Specifically, we set $N_a = 2$, $P_0 = 6.8W$, and $\Delta_p = 4$. The initial transmit power of each cell is randomly selected from \mathcal{P}_i according to the uniform distribution.

In Figure 1, we compare the energy efficiency of a network over time with different N_m . The proposed method represented as NonCoGame in the figure makes \mathfrak{E} to the steady state fast. In case of AIMD, \mathfrak{E} also converges to a steady point which is smaller than that of NonCoGame case. However, \mathfrak{E} s obtained by AIMD increases in the

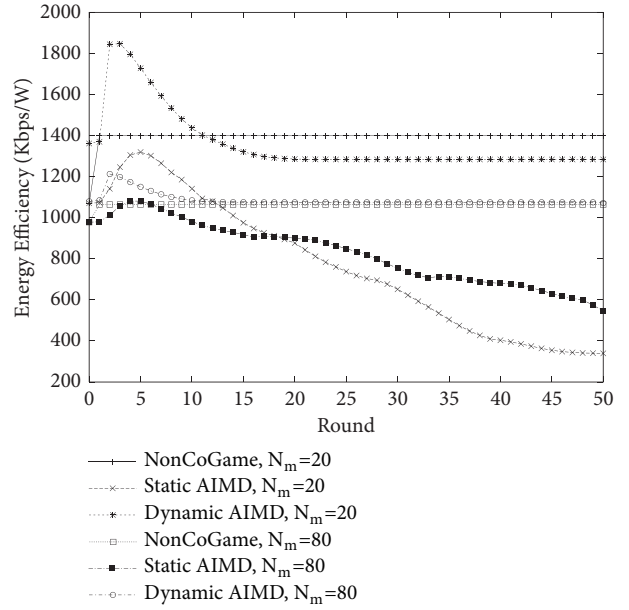


FIGURE 1: Comparison of energy efficiency.

beginning of the simulation and decreases to the steady state.

To scrutinize the phenomenon, we analyze the dynamics of the total throughput of a network (i.e., $\mathfrak{R} = \sum_{i \in \mathcal{C}} \mathfrak{R}_i$) and the total power consumption of a network (i.e., $\mathfrak{P} = \sum_{i \in \mathcal{C}} \mathfrak{P}_i$) in Figures 2 and 3, respectively. We choose $p_i(0)$ randomly at the start of the simulation. \mathfrak{R}_i is determined not only by the received signal power at the UEs in \mathcal{U}_i but also by the total interference imposed on these UEs. If $p_i(0)$ and $p_j(0) (\forall j \in \mathcal{C} - \{i\})$ are not large enough, both the received signal power at the UEs in \mathcal{U}_i and the total interference on these UEs are small. Thus, $\mathfrak{E}_i(0)$ is relatively smaller than its maximum value achievable. In case of AIMD, each cell gradually increases its power whenever $\mathfrak{E}_i(t+1) - \mathfrak{E}_i(t) > 0$. By increasing $p_i (\forall i \in \mathcal{C})$, \mathfrak{P} increases and \mathfrak{R} increases if $p_i h_{i,u}$ is bigger than $\sum_{j \neq i \in \mathcal{C}} p_j h_{j,u} + N_0$. If the amount of the increment in \mathfrak{R} is larger than that in \mathfrak{P} for a given $p_i(t)$ and $p_i(t-1)$, \mathfrak{E} increases. If a cell detects $\mathfrak{E}_i(t+1) - \mathfrak{E}_i(t) < 0$, it decreases p_i by half, which decreases both \mathfrak{R}_i and \mathfrak{P}_i . However, \mathfrak{P}_i is a linear function of p_i while \mathfrak{R}_i involves not only p_i but also $p_j (\forall j \neq i \in \mathcal{C})$. We observe in Figure 3 that \mathfrak{P} does not change much while \mathfrak{R} decreases sharply during the round 1 to 15. The result indicates that the amount of increments in \mathfrak{P} by the cells increasing transmit powers is close to the amount of decrements in \mathfrak{P} by the cells who reduce their transmit powers. Since AIMD decreases a transmit power multiplicatively, the received signal powers at the UEs in the cell that decreases transmit power also reduces while the total interference experienced by these UEs increases because of the other cells that increase transmit powers. Thus, \mathfrak{R}_i reduces abruptly which leads to sharp decrease in \mathfrak{R} during this period.

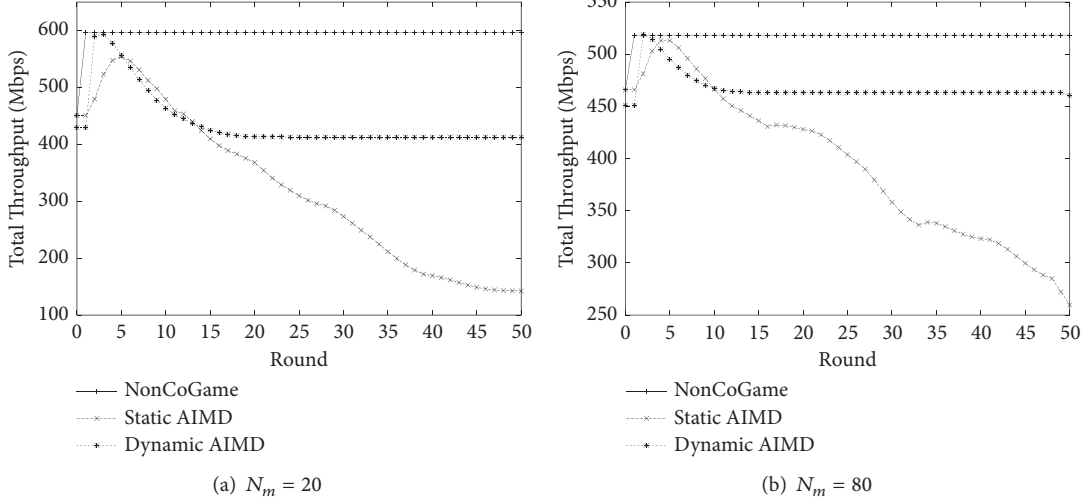


FIGURE 2: Comparison of total throughput.

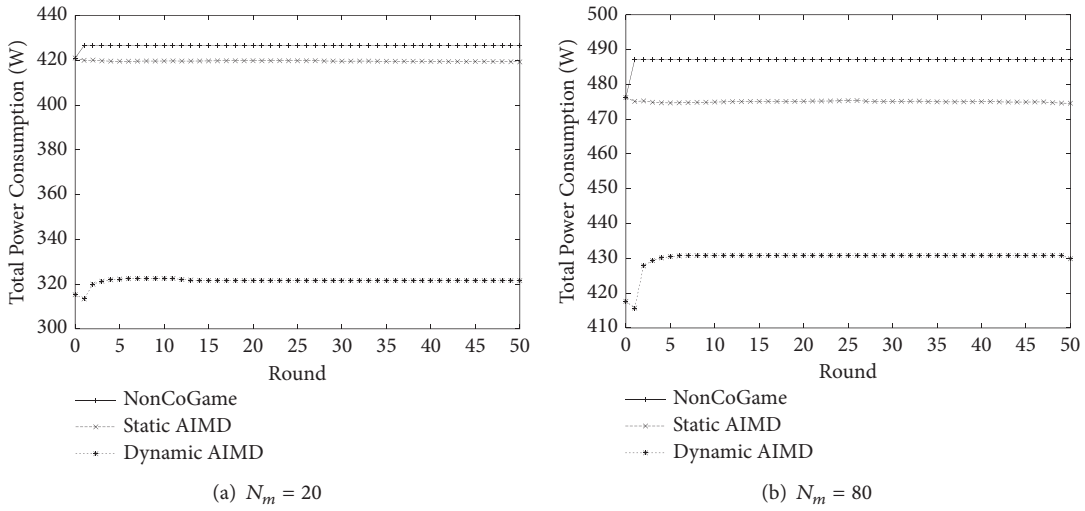


FIGURE 3: Comparison of total power consumption.

To inspect the effects of the fairness in terms of \mathcal{G}_i , we observe the Jain's fairness index \mathfrak{F}_{fi} in Figure 4, where \mathfrak{F}_{fi} is given as

$$\mathfrak{F}_{fi} = \frac{(\sum_{i \in \mathcal{E}} \mathcal{G}_i)^2}{|\mathcal{E}| \sum_{i \in \mathcal{E}} \mathcal{G}_i^2} \quad (21)$$

As \mathfrak{F}_{fi} is closer to 1, \mathcal{G}_i s are more similar to each other. We can see that the proposed method can increase \mathfrak{F}_{fi} . Since cells and UEs are uniformly distributed, each cell can obtain similar energy efficiency when the proposed method is used.

5. Conclusions

In this paper, we propose a noncooperative game model for a cell to determine its transmit power autonomously

to optimize its energy efficiency. We provide a sufficient condition for the existence of Nash equilibrium of the proposed game. We also prove that the game has unique Nash equilibrium by transforming the nonlinear fractional programming problem into a parametric programming problem. Through simulation studies that compare the performance of the proposed method with those of the AIMD method, we show that the proposed power control method can stabilize a system at higher total energy efficiency and balance cell energy efficiency more evenly than the AIMD methods.

Data Availability

The data used to support the findings of this study are included within the article.

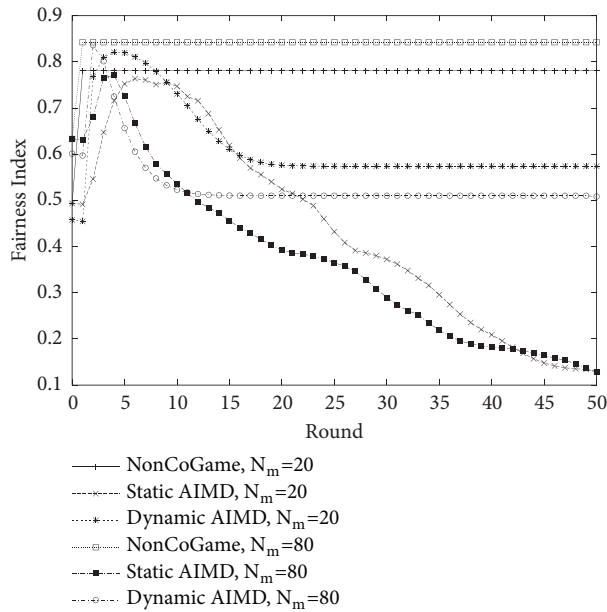


FIGURE 4: Comparison of fairness index in terms of the energy efficiency.

Conflicts of Interest

The authors declare that there are no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government under Grant NRF-2018RID1A1B07050893.

References

- [1] Ericsson, "Ericsson Mobility Report," Nov. 2017.
- [2] J. G. Andrews, S. Buzzi, and W. Choi, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [3] D. C. Araujo, T. Maksymuk, A. L. F. de Almeida, T. Maciel, J. C. M. Mota, and M. Jo, "Massive MIMO: Survey and future research topics," *IET Communications*, vol. 10, no. 15, pp. 1938–1946, 2016.
- [4] S. Bassoy, H. Farooq, M. A. Imran, and A. Imran, "Coordinated Multi-Point Clustering Schemes: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 743–764, 2017.
- [5] B. Soret, A. De Domenico, S. Bazzi, N. H. Mahmood, and K. I. Pedersen, "Interference Coordination for 5G New Radio," *IEEE Wireless Communications Magazine*, pp. 1–7, 2017.
- [6] M. Xiao, S. Mumtaz, Y. Huang et al., "Millimeter Wave Communications for Future Mobile Networks (Guest Editorial), Part I," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1425–1431, 2017.
- [7] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-Dense Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.
- [8] A. Y. Al-Zahrani, F. R. Yu, and M. Huang, "A Joint Cross-Layer and Colayer Interference Management Scheme in Hyperdense Heterogeneous Networks Using Mean-Field Game Theory," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1522–1535, 2016.
- [9] J. Liu, M. Sheng, L. Liu, and J. Li, "Interference Management in Ultra-Dense Networks: Challenges and Approaches," *IEEE Network*, vol. 31, no. 6, pp. 70–77, 2017.
- [10] B. Yang, G. Mao, X. Ge, M. Ding, and X. Yang, "On the Energy-Efficient Deployment for Ultra-Dense Heterogeneous Networks With NLoS and LoS Transmissions," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 369–384, 2018.
- [11] L. D. Nguyen, H. D. Tuan, T. Q. Duong, O. A. Dobre, and H. V. Poor, "Downlink Beamforming for Energy-Efficient Heterogeneous Networks With Massive MIMO and Small Cells," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3386–3400, May 2018.
- [12] Z. Li, D. Grace, and P. Mitchell, "Traffic-Aware Cell Management for Green Ultradense Small-Cell Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2600–2614, 2017.
- [13] J. Zheng, Y. Cai, X. Chen, R. Li, and H. Zhang, "Optimal base station sleeping in green cellular networks: a distributed cooperative framework based on game theory," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4391–4406, 2015.
- [14] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, 2016.
- [15] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multiobjective Auction-Based Switching-Off Scheme in Heterogeneous Networks: To Bid or Not to Bid?" *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9168–9180, 2015.
- [16] F. Héliot, T. Yang, and C. H. Foh, "Low-complexity green scheduling for the coordinated downlink of HetNet system," in *Proceedings of the 20th IEEE International Workshop on Computer Aided Modelling and Design of Communication Links and Networks, CAMAD 2015*, pp. 216–220, UK, September 2015.
- [17] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. M. Leung, and H. V. Poor, "Energy Efficient User Association and Power Allocation in Millimeter-Wave-Based Ultra Dense Networks with Energy Harvesting Base Stations," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1936–1947, 2017.
- [18] L. Liang, W. Wang, Y. Jia, and S. Fu, "A cluster-based energy-efficient resource management scheme for ultra-dense networks," *IEEE Access*, vol. 4, pp. 6823–6832, 2016.
- [19] S. Samarakoon, M. Bennis, W. Saad, M. Debbah, and M. Latva-Aho, "Ultra dense small cell networks: Turning density into energy efficiency," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1267–1280, 2016.
- [20] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3292–3304, 2012.
- [21] L. Sboui, Z. Rezk, and M.-S. Alouini, "Energy-efficient power control for OFDMA cellular networks," in *Proceedings of the 27th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2016*, Spain, September 2016.
- [22] G. Ozcan, M. Ozmen, and M. C. Gursoy, "QoS-driven energy-efficient power control with random arrivals and arbitrary input

- distributions,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 376–388, 2017.
- [23] M. Ozmen and M. C. Gursoy, “Energy-Efficient Power Control in Fading Channels with Markovian Sources and QoS Constraints,” *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5349–5364, 2016.
- [24] A. Helmy, L. Musavian, and T. Le-Ngoc, “Energy-efficient power adaptation over a frequency-selective fading channel with delay and power constraints,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4529–4541, 2013.
- [25] Y. Jiang, N. Lu, Y. Chen et al., “Energy-Efficient Noncooperative Power Control in Small-Cell Networks,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7540–7547, 2017.
- [26] C. Yang, J. Li, Q. Ni, A. Anpalagan, and M. Guizani, “Interference-aware energy efficiency maximization in 5G ultra-dense networks,” *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 728–739, 2017.
- [27] H. Zhang, H. Liu, J. Cheng, and V. C. M. Leung, “Downlink energy efficiency of power allocation and wireless backhaul bandwidth allocation in heterogeneous small cell networks,” *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1716–1705, 2017.
- [28] M. Lashgari, B. Maham, and H. Kebriaei, “Energy Efficient Price Based Power Allocation in a Small Cell Network by Using a Stackelberg Game,” in *Proceedings of the 2018 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 1–5, Batumi, Georgia, June 2018.
- [29] X. Guozhen, W. Sen, and C.-L. I, “On Amorphous Nature of Ultra Dense Networks,” in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2016.
- [30] *3GPP Radio Access (E-UTRA) Radio Frequency (RF) requirements for LTE Pico Node B*, 3GPP TR 36.931 V9.0.0, May 2011.
- [31] G. Auer, V. Giannini, C. Desset et al., “How much energy is needed to run a wireless network?” *IEEE Wireless Communications Magazine*, vol. 18, no. 5, pp. 40–49, 2011.
- [32] R. B. Myerson, *Game Theory: Analysis of Conflict*, Harvard University Press, 1991.
- [33] W. Dinkelbach, “On nonlinear fractional programming,” *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.