Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Electrical, Electronics and Communications Engineering

(34^{th}cycle)

Green Mobile Networks: from 4G to 5G and Beyond

Greta Vallero

******

Supervisor:

Professor Michela Meo

Doctoral Examination Committee:

Marco Ajmone Marsan - IMDEA Networks Institute, Leganes (Madrid), Spain

Cicek Cavdar - KTH Royal Institute of Technology, Stockholm, Sweden

Emilio Leonardi - Politecnico di Torino, Torino, Italy

Vincenzo Mancuso - IMDEA Networks Institute, Leganes (Madrid), Spain

Loutfi Nuaymi - IMT Atlantique, Nantes, France

Politecnico di Torino

2022

To Gin

Acknowledgements

Before presenting the work, I would like to thank everyone who helped me during this PhD program. I would like to express my gratitude to my supervisor, professor Michela Meo, for her constant guidance, for being supportive and encouraging, offering me valuable opportunities to grow in my work as a researcher and as a woman. Furthermore, I am grateful for the remarkable opportunity I have been given to carry out part of my PhD research activity working with professor Marco Ajmone Marsan and Dr. Daniela Renga, who have always been keen to share their deep and distinguished expertise in the field of Telecommunication Networks. In addition, my PhD experience has been enriched by the collaborations with professor Margot Deruyck, professor Wout Joseph and German Castellanos, from Ghent University, in Belgium.

A huge thanks to Linda and Sor'Anna for being my light and lovely refuge. A special thanks to Alessandro and Noemi, for being the craziest and, also for this reason, my favourite couple. I would like to thank my awesome colleagues, who started and shared with me this journey Giuseppe, Elena, Marco, Andrea and Michele. Thanks also to my "new" colleagues Antonio, Tommaso, Federico, Franco, Giuseppe Jr, Tuni, Barbara and Alessandro.

A special thanks to the Maigret & Magritte, for being the community I have been looking for. In particular thanks to the *Condominio Giuseppi*. Thanks guys for the amazing moments spent together, for the Capogiallo night and for the countless Saturday nights on jit.si during the pandemic months.

Thanks to Maurizio, Sam and my teammates, the most amazing team, which a captain could have. Thank you girls!

Finally, thanks to my best friends, my lovely, sometimes hateful, friends, who make me feel like a lucky person: Elena, Martina and Mommo.

Last but not least, thanks to my lifelong friend Lorenzo and all his family. Someone would ask: "After all this time?", I will reply: "Always!"

Abstract

In order to comply with the Paris Agreement and the European Green Deal, the communication community has recognised the network energy efficiency as a fundamental and urgent aspect, to make the communication network sustainable.
In line with this, the contribution of this thesis consists in designing, analysing and evaluating high energy efficient RANs, in the 4G environment and investigating various critical issues raised by the introduction of the MEC technology and the UAV-BSs, pillar technologies for 5G and beyond RANs.
The Base Stations (BSs) account for 80% of the total consumption of the Radio Access Network (RAN) and their energy needs are expected to further grow because of the rise of mobile data traffic in the next few years. The BS switching is one of the most studied approaches for the reduction of the energy consumption of RANs. We propose different RAN managements, which reach up to 40% energy saving and good Quality of Service (QoS). The decision to switch a BS to sleep mode is driven by the future traffic demand and/or the future generation of a Renewable Energy Source (RES), typically a Photovoltaic (PV) panel system, which is locally installed. In order to access these predictions, Artificial Intelligence (AI) and Machine Learning (ML) based approaches are used in this work. The network performances reveal that ML approaches are necessary to achieve significant energy saving and good QoS. Nevertheless, a careful processing of the traffic predictions and the understanding of the overall traffic pattern is fundamental and result more impacting on the achieved RAN performances than the careful selection of the traffic predictor. The impact of the BS switching on the BS failure rate is also analysed in this work.
Besides the RAN energy efficiency, the future generation of RANs is envisioned to expand the existing mobile networks, achieving ultra-low delays, extensive coverage, as well as ultra-high reliability. In order to achieve this, the Multi-access Edge Computing (MEC) is considered a promising solution. In RANs, it consists of the placement of computing and storage servers, directly at each BS of these

networks. In this work, the simultaneous employment of the MEC technology and BSs switching is considered, providing an overview of their mutual effects. The employment of the MEC technology increases the RAN energy consumption, since the MEC platforms have to be powered to provide the service, while the BS switching dynamically activates and deactivates resources, and consequently the MEC servers, impacting its performance. New user association policies are proposed, in order to totally exploit the MEC technology and reduce the network energy consumption. Simulation results reveal that, thanks to the MEC and the proposed methodologies, the experienced delay and the energy consumption drop, respectively, up to 60% and 40%.

Moreover, Unmanned Aerial Vehicles equipped with Base Stations (UAV-BSs) are considered an effective support for 5G RAN, to dynamically provide additional capacity, in case of network congestion or emergency situations. Our evaluation reveals that UAV-BSs need frequent replacement because of the scarce on-board energy availability on UAV-BSs, provided by on-board batteries. In addition, simulations show that the Backhaul (BH) network often saturates, due to its low available bandwidth, deteriorating the network QoS. To cope with the latter, we use the MEC paradigm, to cache popular contents on each UAV-BS and decrease the occupancy of the BH network. It results effectively, reducing the lost traffic by 33%. In order to address the former issue, we consider a solar-powered UAV-BSs. The model of a Long Term Evolution (LTE) Multi User (MU)-Multiple Input Multiple Output (MIMO) UAV-BS, powered by a PV panel system, is formalised, to derive the energy production levels that are needed to satisfy the traffic demand, the probability to waste energy and the proper PV panel capacity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The supply of the Information and Communication Technology (ICT) sector was responsible for 3% of the global carbon emissions in 2018 and, according to forecasts, following its current trend, will be responsible for up to 14% in 2040, making this field a direct contributor to the climate change [5]. For this reason, the European Commission, in [6], under the need for actions to improve the energy efficiency in communications, in line with the Paris Agreement and the European Green Deal, which aim at achieving net zero greenhouse gas emissions by 2050, has formalised a policy for the regulation of the energy consumption and the carbon emissions of the Broadband Communication Equipment. In particular, according to this document, known as Code of Conduct, the total European consumption of at least 50 TWh per year was estimated for the year 2015 for broadband equipment. This document discusses the policy for slightly reducing or keeping constant the broadband equipment electricity consumption compared to a business as usual scenario with the current growing usage and penetration of broadband equipment in the EU, through the smart stand-by equipment and the energy aware management.

Meanwhile, the communication community has recognised the network energy efficiency as a fundamental and urgent aspect, to make more sustainable the communication network. As well known, the BSs have been identified as the most energy consuming components of mobile networks [7], accounting for 80% of the total energy consumption of the RANs. The BSs energy consumption is expected to further grow because of the rise of the mobile IP traffic, which will reach 77.5 exabyte (EB) per month by 2022 and 5 016 EB per month in 2030 [7–9], more than,

respectively, 6 and 400 times larger than 11.5 EB per month occurred in 2017.
Because of this, a lot of effort has been put in the transition towards more sustainable and energy efficient RAN and results can be classified in two domains. The first refers to the improvements of the energy characteristics of the telecommunication equipment themselves, aiming at reducing the RAN power consumption through the use of low-power components, the adoption of more efficient power amplifiers, the increase of the thermal operation range to reduce the need for cooling, etc [10]. This has been largely adopted by Mobile Network Operators (MNOs). For example, BSs of 3G systems have a typical peak power consumption of about 3.5 kW, while 4G BSs consume less than 1 kW.
The second domain refers to network-level resource management algorithms, that leverage the temporal and spatial variations of traffic loads, using sleep (or Low Power Idle) modes, to achieve energy efficiency [11–14]. In particular, Resource on Demand (RoD) strategies are RAN management approaches that dynamically adapt the active resources to the traffic demand. The traffic demand in a RAN varies over time, presenting peaks for relatively short periods, and long periods of resource under-utilisation that induce waste of energy. This waste of energy is because the energy consumption of the 4G BSs is very little proportional to the traffic demand and when it is low, unneeded resources are kept active, consuming energy. For this reason, while during traffic peaks all the available RAN resources must be activated, when the traffic demand is low, unnecessary BSs are switched in sleep mode. In this way, the amount of active networking equipment is reduced to the minimum necessary, to provide the desired capacity to serve the instantaneous traffic demand while keeping the desired QoS level. Nevertheless, activating and deactivating BSs impacts the BS failure rate: on the one side, switching is harmful to BS failure rate; on the other side, the time spent in sleep mode saves the BS from deterioration. The balance between these two phenomena depends on the hardware (HW) components of the BS, and on the RAN management strategy. The deterioration of the failure rate of the BS directly impacts its maintenance and operation cost for the operator. According to [15], this cost accounts for 3 k€ and 1 k€, per year, for each macro and micro cell BS, respectively. Up to 4 billions of BSs were counted worldwide back in 2012, as described in [16], and this number is bound to remarkably increase because of the RAN densification, planned with the 5G RAN deployment [17]. Thus, in addition to energy saving and QoS, these RAN management strategies have to be designed taking into consideration the impact on the BS failure rate, in order to

avoid the explosion of the RAN Operational (OPEX) cost.
Besides the reduction of the absolute value of the energy consumption of telecommunication networks through this dynamic resource allocation, a trend in networking considers the availability of the energy produced by RES [18, 19]. In particular, RES are integrated in the mobile networks and are installed in proximity to the BSs networks. This means that the RAN is powered by both the "green" energy, produced by RES and the energy taken from the power grid, which can be considered "brown", since it is largely produced by burning fossil fuels. Besides the improvement of the RAN sustainability, this solution is effective to reduce the network electricity bill, by reducing the amount of energy that has to be purchased from the power grid, which is the key contributor for the increase of the OPEX. Because of this, according to [20] between 2014 and 2020 the number of renewable powered BSs increases by 45%, accounting for 70 000 towers powered by solar powered worldwide in 2020, and their deployment is expected to significantly advance in 6G RANs [9]. Also in literature, many works consider a PV panel system to reduce the CO_2 emissions used to generate energy by burning fossil fuels, when powering a RAN [21, 22]. While the solution is promising, various issues need to be addressed, among which the solar panel dimensioning and the possible lack of energy generation due to its intermittent nature. Indeed, the solar energy harvesting presents, as other RES, randomness, dependence on the weather conditions and daily and seasonal variability, making these BSs self-survival unstable. To address this issue, the green energy generation has to be combined with resource management strategies, in order to optimise its usage with respect to the brown energy. If the available green energy is properly managed, the network becomes more independent from the power grid, making this solution very suitable for remote areas and regions where the electricity system is unreliable, because of frequent black-outs. To this end, AI and ML based solutions offer great support. The growth of computational power, the availability of data, the improvement of learning algorithms are the boosts behind the pervasiveness of new AI- and ML-based mechanisms to respond to the new challenges of today's networks, which are often too complex to be properly understood, modelled, and managed with traditional approaches, as the case of decision making for network management for RAN energy efficiency. In addition, these approaches are also fundamental to catch the effect of complex interactions among multitudes of heterogeneous users, network elements (such as macro and small cells in heterogeneous networks) and different energy supplies, as well as to understand the hidden correlations among systems.

These advances respond to the energy efficiency requirements of the new generations of networks. Indeed, 5G systems aim at consuming a fraction of the energy consumption of 4G mobile networks, even if, as mentioned above, the amount of traffic which 5G networks are supposed to manage is much larger than in 4G ones [23]. In addition, discussions about 6G networks have identified 1 Tbit/W as the energy efficiency constraint for that generation of RANs. Nevertheless, besides the RAN energy efficiency, the next generation of networks is envisioned to expand the existing mobile networks, achieving ultra-low delays, extensive coverage, as well as ultra-high reliability. This allows to deploy plenty of services and applications, which require strict constraints in terms of latency, throughput and reliability, from various sectors such as automotive, entertainment, e-Health and e-Industry. In order to reply to these requirements, the MEC paradigm has been introduced [24–26]. It pushes computing and storage resources in physical proximity of end users, placing servers on the edges of the network [27]. In this way, the execution of applications, the pre-processing of data and the caching of popular contents are performed in proximity of end users. In RANs, these servers are co-located on BSs. Therefore, these infrastructures provide storage and computation services, in addition to access services [28]. Several benefits are derived from the introduction of this technology [29, 30]. First, the backhaul traffic load is reduced, since the access to the cloud is unneeded and handled at the network edge. Second, as proved in [31–33], the ultra-low delays are ensured, since the service and content location are very close to the end users. On the other hand, the employment of the MEC technology worsens the RAN energy efficiency. Indeed, in addition to the energy needed for the supply of the communication unit, the computing and/or caching platforms has to be powered to provide the service, generating the growth of the RAN energy consumption [34, 35]. Meanwhile, the BSs switching used by the RoD approaches dynamically activates and deactivates resources, and consequently the MEC servers, impacting its performance.

Moreover, in order to satisfy the new generation RAN requirements, the Third Generation Partnership Project (3GPP), in [36], identified the aerial networks as the complementary infrastructure to the terrestrial 5G RAN [37]. This aerial network relies mainly on UAVs. UAVs are typically drones and can operate between 80 m and 400 m from the ground. Mounting a BS on UAVs has been proposed as a promising solution to dynamically deploy fast and flexible communication facilities, where traditional ground infrastructures are not feasible or cost-effective [38]. Through

the additional capacity provided by UAV-BSs, connectivity is brought to the users that are suffering from low-quality service. In particular, UAV-BSs adapt their aerial position where needed, based on position and traffic requirements of users, who are connected to the UAV-BSs through access links, while UAV-BSs are connected to the Core Network (CN), establishing BH links between them and an Access Point (AP). The BH network is the most challenging part of this network. Indeed, its links represent the bottleneck of the system. According to [4], if the LTE traditional frequency is used, not enough spectrum is available, while higher frequency employment suffers from high path loss. Another important challenge that needs to be addressed is related to the scarce on-board energy availability that is provided to UAV-BSs by on-board batteries. This is an issue which actually characterises UAVs in general but, when they are used as communication infrastructure, the situation worsens. Indeed, besides the energy needed for the flight of the UAV, also the communication unit has to be powered to provide the service [39], resulting in a higher energy consumption and reduction of the UAV-BS lifetime. To cope with this, solar-powered UAV-BSs are a viable solution, which result in a longer UAV-BS survival, without adding significant mass to the device [40, 41]. This is possible because of the advance of the UAV structure, which makes most recent drones able to carry up heavy payloads. Nevertheless, the solar panel dimensioning and the understanding of the effect of intermittent energy production on the communication service provided by UAV-BSs network are fundamental steps in the design of this network infrastructure.

1.1 Main Contributions

The contribution of this thesis consists in designing, analysing and evaluating high energy efficient RANs, investigating various critical issues raised by the introduction of the MEC technology and the UAV-BSs, pillar technologies for 5G and beyond RANs. Firstly, we design and investigate the suitability of ML algorithms as support for the implementation of BS switching mechanisms, which aim at the network energy consumption reduction. Then, we design, implement and assess the performances of the mutual effect of the MEC paradigm and the BS switching mechanisms. Finally, we design and validate an UAV-BS-aided RAN, used in case of crowded and emergency situations. Further details about the main contributions and the topics covered in this work are presented below.

Suitability, adaptation and analysis of ML-based solutions for RAN energy efficiency

Dynamic resource allocation is one of the most effective approaches to reduce the heterogeneous RAN energy consumption. It is based on the sleep mode of BSs. In particular, when RoD based approaches are used, a BS is switched to sleep mode, in case its traffic demand is low. Alternatively, in case the RAN is supplied by a RES system, the BS switching decisions are based on the amount of available renewable energy, which is locally produced. In particular, a BS is put in sleep mode, when that quantity is not sufficient for the network supply, to minimise the energy which has to be purchased from the grid. We propose different approaches for this scenario and results highlight the fundamental role of the macro cell BSs in hierarchical RANs, in order to provide the adequate QoS, while reducing the network energy consumption. The decisions for the BS switching are driven by the future traffic demand and/or the future renewable energy production. Assuming them perfectly known is an optimistic and unrealistic assumption. In order to overcome it, we use different ML algorithms for their prediction. First, a clear understanding of the effects on the network operation of the introduction of ML algorithms for the forecast of future traffic and/or the PV panel production is provided. Results, obtained through simulations, reveal that they are suitable for this application and allow to achieve significant energy saving. Nevertheless, the achieved energy saving strongly depends on the traffic pattern. For this reason, ML approaches are necessary, since they are a versatile framework, which adapts the network operation to the traffic characteristics typical of each area and to its evolution, which cannot be performed autonomously and that needs continuous updates to follow traffic pattern variations. Nevertheless, QoS may be compromised because of incorrect BS deactivation, with a limited sensitivity to the type of ML algorithm, which is used. This is because critical BS (de)activation decisions are taken in correspondence of specific traffic values, and high accuracy in the estimations is not required in general, but only close to the values which trigger a BS switching. For this reason, traffic predictions performed over a shorter time scale, combined with a careful processing of the predictions are proposed, resulting in very effective performance improvement, measured in RAN energy saving and QoS. To do this, different Artificial Neural Network (ANN)-based approaches are proposed for the traffic forecast, whose input features are analysed and discussed. The processing of the predictions aims at detecting the overall trend

of the traffic profile and combining predictions at different time lags. They result more impacting on the achieved RAN energy efficiency and QoS than the careful selection of the traffic predictor.
The impact of the BS switching on the BS failure rate is also analysed in this work. To do this a failure rate model is employed, which accounts for the time a BS spends in sleep mode, which saves from deterioration, as well as for the BS switching frequency, which is harmful for its HW components. Energy saving strategies based on conservative processing of the traffic demand forecasts reach significant energy consumption reduction, preserving QoS, as well as the BSs failure rate, also in case the BS HW has not been designed for dynamic switching. When the processing of the traffic demand predictions results in a dynamic BSs activation and deactivation, energy saving is further slightly improved, at the expense of a small loss in QoS and in the BS failure rate, suggesting that these approaches are suitable only in case the BS HW design is optimised for BS switching.
Within the ML-based solution for RAN energy efficiency, our contributions can be found in:

- Donevski, I., Vallero, G., & Marsan, M. A. (2019, April). Neural networks for cellular base station switching. In I*EEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 738-743). IEEE.

- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2019). Greener RAN operation through machine learning. *IEEE Transactions on Network and Service Management, 16*(3), 896-908.

- Vallero, G., Renga, D., Meo, M., & Ajmone Marsan, M. (2020, November). Processing ANN Traffic Predictions for RAN Energy Efficiency. In *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (pp. 235-244).

- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2021). RAN energy efficiency and failure rate through ANN traffic predictions processing. *Computer Communications*.

Analysis and optimisation of the simultaneous employment of MEC technology and RAN dynamic resource allocation

The MEC paradigm has been introduced in 5G networks, in order to reply to the requirements of services and applications, which require strict constraints in terms of delay, from different sectors (automotive, e-Health, etc). Analysing the simultaneous employment of the MEC technology and BSs switching for the RAN energy efficiency is fundamental. Indeed, because of the MEC servers' supply, the RAN energy consumption increases, while BS switching deactivates RAN resources and, as a consequence, also these MEC platforms. We evaluate these aspects through a simulation-based approach, which simulates the RAN located in the city centre of Ghent, in Belgium. The impact of the MEC server capacity, as well as of the traffic characteristics, on the RAN performances is analysed. Results show that caching at the edge and the dynamic activation of the BSs is promising in reducing latency and the network energy consumption, respectively, without deteriorating their performances because of their coexistence. In addition, the spread of the cache capacity among the BSs in heterogeneous RAN is investigated. Caching on the macro BSs is always needed to significantly reduce delays, while caching also on the micro cells relieves the effort on the macro cell.
Different association policies are proposed, which minimise the RAN energy consumption and/or the experienced delay, in order to maximise the benefits provided by the MEC technology usage, ensuring also the achievement of the network energy efficiency. Because of the high complexity of the problem, we propose a greedy approach for its resolution. This is strictly necessary, since the association procedure is performed while the system is operating, which means that a solution is needed on the fly.
Within the the simultaneous employment of the MEC technology and BSs switching, our contributions can be found in:

- Vallero, G., Deruyck, M., Joseph, W., & Meo, M. (2020, June). Caching at the edge in high energy-efficient wireless access networks. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.
- Vallero, G., Deruyck, M., Meo, M., & Joseph, W. (2021). Base Station switching and edge caching optimisation in high energy-efficiency wireless access network. *Computer Networks*, 192, 108100.

Design, modelling and evaluation of UAV-BS networks

UAV-BSs-aided networks are a suitable solution, to dynamically deploy fast and flexible communication facilities, to cope with the terrestrial network failures, due to network overloads or physical unavailability. They are also envisioned as a support to reply to the large traffic demand of the 5G and 6G RAN. In order to evaluate UAV-BSs networks in crowded scenarios, we implement an ad-hoc simulator, which accounts for the real 3D environment of the considered area and traffic demand collected in crowded environments. From our evaluation, the BH links emerge as the most challenging part of the network because of its frequent saturation, due to the scarce available bandwidth. We propose the employment of the MEC paradigm, in order to cache popular contents on each UAV-BS and decrease the occupancy of the BH network. Results reveal that this is a very effective solution, even if the achieved performances are strongly dependent on the traffic characteristics.
In addition, another important problem which is revealed by our simulations is related to the scarce on-board energy availability, provided by on-board batteries. As previously mentioned, the UAV lifetime is constrained by the time for the depletion of the on-board energy battery. When UAVs are used as BS, the UAV-BS lifetime further decreases because, besides the energy needed for the flight, also the communication HW needs to be powered. In order to address this issue, we consider a solar-powered UAV-BSs. In particular, in order to properly design and quantify the effects of the changes of its components to satisfy the needed requirements, the model of a LTE MU-MIMO UAV-BS, powered by a PV panel system, is formalised. In order to do this, the queuing theory and a discretized representation of the data and energy flows, as Data Packets (DPs) and Energy Packets (EPs), respectively, are used, to investigate the interplay between the traffic demand and the energy generation. Through this model, the energy production levels that are needed to satisfy the traffic demand, the probability that an EP is unused and the proper PV panel dimensioning are analytically derived, also considering the time dependence of both the traffic demand and the PV panel production.
Within the UAV-BS topic, our contributions can be found in:

- Castellanos, G., Vallero, G., Deruyck, M., Martens, L., Meo, M., & Joseph, W. (2021). Evaluation of flying caching servers in UAV-BS based realistic environment. *Vehicular Communications*, 32, 100390.

- Vallero, G., & Meo, M. (2021, June). Modelling Solar Powered UAV-BS for 5G and Beyond. In *2021 19th Mediterranean Communication and Computer Networking Conference (MedComNet)* (pp. 1-8). IEEE.

1.2 Outline of the Thesis

The rest of the thesis is organised as follows:

- **Chapter 2** contains the related work information, along with the related open issues and further details about the objectives and the various topics addressed in our study.
- **Chapter 3** considers a portion of an heterogeneous RAN and, in order to dynamically manage the RAN resources, uses AI/ML-based approaches, which receive as input real data traffic. In particular, in this part of the study, the RoD switching approach is used: micro cell BSs are put in sleep mode in case the forecast traffic demand is low, to reduce the network energy consumption. The effects of the usage of ML algorithms for the forecast of future traffic demand on BS switching are evaluated.
- **Chapter 4** discusses the impact of the time granularity of the traffic predictions, which drive the BS switching decisions. Different ANN-based methodologies for the traffic prediction are proposed and evaluated. Approaches which aim at carefully processing the traffic forecasts and understanding the overall traffic pattern are proposed to improve the network performance. Finally, the impact of the BS activation/deactivation on the BS lifetime is quantified.
- **Chapter 5** envisions the employment of a RES system for the RAN supply. In particular, in this part of the work, the BS switching operates according to the available renewable energy, which is produced by a PV panel system, locally installed. Different BS switching strategies are proposed and evaluated. Afterwards, ML algorithms are used for the prediction of the RES production, as well as of the traffic demand, and decisions for the BS deactivation are taken according to them. The effects of the accuracy of the ML algorithms are discussed and evaluated.

- **Chapter 6** investigates the MEC paradigm and, in particular, analyzes the scenario which uses the BS switching and the MEC paradigm simultaneously. In order to do this, the portion of RAN located in the city centre of Ghent, in Belgium, is simulated. The growth of the energy consumption due to the supply of the MEC servers installed on each BS is discussed, as well as the effects of the dynamic activation/deactivation of the BSs, and consequently of the MEC servers, on the experienced delay. Moreover, the impact of the different spread of the cache capacity among the BSs of the network is analysed. New association policies are introduced, in order to minimise the MEC-enabled RAN energy consumption and/or the experienced delay, in order to maximise the benefits provided by the MEC technology usage, ensuring also the achievement of the RAN energy efficiency.

- **Chapter 7** is devoted to investigate and design UAV-BSs-aided networks. The MEC paradigm is used to address the issue related to the BH network congestion, analysing the derived benefits, as well as the growth of the network energy consumption and the drop of the time of life of each drone.

- **Chapter 8** focuses on the modelling of a UAV-BS, equipped with a PV-panel, for the communication unit supply, in order to address the issue related to the scarce on-board energy availability. Using queuing theory and data and energy flows discretization, the PV panel is dimensioned and the energy wasting is formalised, considering the time variation of the traffic demand and of the PV panel production.

- **Chapter 9** summarises the work and the contributions presented in this manuscript.

Chapter 2

Contribution and Related Works

The related works in relation to the main topics of this study are detailed in this chapter, highlighting the open issues which are addressed and the specific contribution which are provided by this study. Section 2.1 revises the literature about the dynamic resource allocation in RAN. In section 2.2 an overview about the MEC paradigm is provided, while section 2.3 describes related works reviewing the usage of UAVs in wireless communications.

2.1 Dynamic Resource Allocation

In the green networking literature, many RAN management solutions have been proposed. The RoD-based approaches exploit the daily variability of the traffic demand and adapt the RAN capacity to it. An overview of these RoD strategies is presented in [11, 42, 43].
The RoD is typically employed for the reduction of the RAN energy consumption. This is the case of the work presented in [13]: it formalises the optimisation problem, whose objective is the minimisation of the total power consumption of the RAN through the regulation of the transmission power and the BS switching to/from sleep mode, while satisfying the minimum data rate for each served user. With the same objective, authors in [44] adapt the energy consumption to the actual traffic load. In

Table 2.1 Summary of references for RoD topic.

[11, 42, 43]	Overviews of RoD strategies
[13, 44, 14]	RoD for RAN energy consumption minimisation
[45, 46]	RoD for the improvement of the interaction with the smart grid, reducing the electricity bill and providing ancillary services
[47]	Prediction of mobile data traffic through ARIMA
[48]	Prediction of mobile data traffic through SA
[49]	Markovian models for traffic prediciton
[48, 50]	Hourly TCP/IP traffic prediction through ANNs
[51, 52]	Mobile traffic prediction through RNNs
[53]	Mobile traffic prediction through an ANN combined with a RNN
[54]	Mobile traffic prediction through Support Vector Machines
[55]	Mobile traffic prediction through Linear Regression
[56]	Effects of sleep mode on the BS lifetime
[57, 58]	Formalisation of the optimal BS switching to maximise the RAN lifetime
[59, 60]	Effects of network device switching in optical backbone networks
[61–63]	RES employment for RAN energy efficient improvement and electricity bill reduction
[64], [65], [66], [67]	PV panels dimensioning problem
[19, 21]	Energy-aware managements of RAN BSs, supplied by a RES system

order to do this, they analytically derive the optimum value which triggers the BS deactivation, in order to reduce the network energy consumption. Similarly, the work presented in [14] dynamically allocates the RAN resources, through a time varied probabilistic on/off switching approach, whose switching decisions are based on a risk level, which provides a measure of the reliability of the network. With a different goal, RoD strategies are applied in [45, 46], in order to improve the interaction with the smart grid in a demand-response scenario, thus reducing the electricity bill and providing ancillary services.

Many of these works, which focus on BSs switching, aim at dynamically allocating resources, under the assumption that the future traffic demand is exactly known. This means that predictions of the amount of traffic demand is necessary in order to make the proposed approaches viable. This aspect is very critical, since errors in these forecasts can significantly affect the performance of these strategies. If the traffic demand

is overestimated, waste of energy occurs; in case of traffic underestimation, incorrect BS deactivations may deteriorate the QoS. In literature, many works focus on traffic estimation. In [47], an Auto-Regressive Integrated Moving Average (ARIMA) is used for the prediction of mobile data traffic and a Seasonal ARIMA (SA) model is used in [48]. These works demonstrate that these two methods provide high accuracy, but require slow training and forecasting, which make them impractical for on-line forecasting. The work presented in [49], uses Markovian models, while [48], [50], [51], [52], [53], [54], [55] employ ML approaches. According to [48] and [50], ANNs provide promising results in forecasting the hourly amount of traffic in TCP/IP networks. In [51], very good performance is reached in the forecast of the mobile traffic of an LTE BS, using a Recurrent Neural Network (RNN) and 1 ms resolution data. High accuracy in traffic predictions is achieved with the same approach, in [52]. A hybrid scheme, structured in an ANN and a RNN is discussed in [53]. Moreover, [54] and [55] predict traffic demand with Least Squares Support Vector Machines and a a Linear Regression based approach, respectively.
In accordance with these studies available in literature, different ML algorithms for the traffic prediction are used in our work, and these predictions are used to drive the RoD decisions. Similarly to what done in [68], a clear understanding of the effects, measured in energy consumption reduction and QoS loss, on network operation of the introduction of ML algorithms for the forecast of future traffic is provided in chapter 3. Differently from the previous literature, in chapter 4, traffic predictions are performed over a shorter time scale, combined with processing techniques on the traffic forecasts, which aim at the understanding of the overall traffic trend. The resulting energy efficiency and QoS is more impacted by these processing techniques than by the careful selection of the traffic predictor.
Recently, the effects of the BS switching on the lifetime of the BSs have been investigated. The authors of [56] showed that putting a BS in sleep mode, besides reducing its energy consumption, increases its lifetime, since the BS operating temperature is reduced. This reduction depends on its HW components, i.e., on the materials used to build the device, and on the time spent in sleep mode. However, the same paper highlights also that power states transitions, which imply transition in the HW operating temperature, negatively affect the BS lifetime. For this reason, the works presented in [57, 58], formalise the optimal BS switching to maximise the RAN lifetime; the problem is solved through an heuristic, which allows to save up to 40% of energy during night, without decreasing the network survival duration.

The effects of the network device switching are also analysed in optical backbone networks, in [59, 60]. Chapter 4 introduces the BS failure rate as a new variable in the design space of the RAN management. For networks not designed for highly dynamic resource allocation, conservative approaches better prevent BSs from HW failure; with the deployment of new devices suited for strongly dynamic networks, less conservative approaches can be used, which achieve higher energy saving.
Many works consider RES power supply for the BSs of RANs. Indeed, RES are widely adopted in real implementations to make communication networks more energy efficient and to reduce the electricity bill [61–63]. Various papers address the critical issue of properly dimensioning these RES systems. The sizing process entails trading off self-sustainability, cost and feasibility constraints due to the installation of a RES system. In [64], the problem of a proper dimensioning of the PV panels is investigated via simulation, whereas authors in [65], [66] and [67] deploy models to derive the optimal RES system dimensioning for powering a BS. In [65] the predefined constraint on the worst month outage probability is considered, whereas in [66] costs are minimised taking into account the limit on the maximum allowed battery depletion probability. The PV panel system capacity model in [67] is derived as the balancing between the need for satisfying the network energy demand and the target degree of network independence from the power grid. Works presented in [19, 21] investigate the integration of energy-aware managements of RAN BSs, supplied by a RES system. These management strategies switch BSs to/from sleep mode, according to the renewable energy availability, for both reducing the RAN energy consumption and complying with the typically intermittent availability of RES production, due to the varying weather conditions. With these objectives, in chapter 5, different RAN management strategies are proposed. These strategies switch BSs to/from sleep mode according to the amount of renewable energy, which is produced by a PV panel system, locally installed, aiming at coping with the intermittent nature of the solar energy. As for the RoD approach, in literature, the future traffic demand and the future RES production, on which BS switching decisions are based, are unrealistically and optimistically assumed perfectly known. For this reason, in section 5.4, ML approaches are employed in order to forecast the future RES production, as well as the future traffic demand, using real historical data. These predictions are used for the RAN management, and their effects on the RAN performance are evaluated.
A summary of the cited works is given in table 2.1.

Table 2.2 Summary of references for MEC topic.

[69, 70]	Overview of the MEC technology utilisation
[31–33]	MEC paradigm for content caching
[2, 25, 26, 29, 35, 71–73]	Formulation of the optimisation problem to select the contents to cache in MEC servers used for content caching.
[2]	Formulation of the optimisation problem to select the contents to cache in MEC servers used for content caching, to minimise the experienced delay.
[71, 29]	Formulation of the optimisation problem to select the contents to cache in MEC servers used for content caching, to maximise the local hit.
[72, 35]	Formulation of the optimisation problem to select the contents to cache in MEC servers used for content caching, to maximise the local hit and minimise the power consumption.
[73]	Allocation of femto cells and WiFi off-loading, to optimise the experienced delay.
[26]	Formulation of the optimisation problem to select the contents to cache in MEC servers used for content caching, to minimise the experienced delay and network cost.
[25]	Content caching replacement algorithm, through predictions of the future content request.

2.2 Multi Access Edge Computing

The MEC technology that uses computing and caching power at the edges of the network, has received much attention in the literature. It provides several benefits, such as the reduction of the experienced latency and of the load in the core network [29, 30]. In [69, 70], surveys of the MEC technology utilisation are provided. The overview presented in [69] distinguishes the MEC mechanisms into mechanisms for computation offloading and mechanisms for caching, while [70] mainly focuses on caching features, discussing in details the caching optimisation and the content insertion/expulsion policies. The reduction of the latency given by the adoption of the MEC paradigm, used for content caching, in wireless networks is demonstrated in [31–33]. In [31], different use cases show the drop of the delay and backhaul links utilisation, with the employment of the MEC paradigm, whereas the work

presented in [32] highlights that the spectral and energy efficiency are improved with caching at the wireless edges. Authors in [33] propose the Dynamic Programming based Adaptive Caching Algorithm, which, besides the minimisation of the experienced delay, optimise also the provided video quality. Many works formulate an optimisation problem to select the contents to cache in order to improve the network performance, when the MEC technology is used, [2, 25, 26, 29, 35, 71–73]. Authors in [2] aim at minimising the experienced delay in an heterogeneous RAN, where caching servers are placed on each BS. In [71], the optimisation problem maximises the local hit, proposing a reduction of the problem in order to treat it analytically. With the same objective, in [29], the optimal content placement problem is given. The work discussed in [72] formulates through an integer programming problem the most appropriate content placement in order to optimise the system towards the hit occurrences and the power consumption, considering users mobility. Also in [35], the experienced delay and power consumption are jointly optimised and results are obtained through simulations. In [73], the allocation of femto cells and WiFi off-loading, used as helper where some contents are cached, is optimised in order to minimise the time needed for each download. [25] proposes a content caching replacement algorithm, which uses predictions of the future request references, derived from a polynomial fit algorithm. Authors in [26] formulate an optimisation problem, which minimise the network cost and content delivery delay, in order to determine which chunks of a content should be cached and how they should be transcoded. The problem is solved applying the relaxation to the constraints and proposing an heuristic.

The energy efficiency in RAN and the employment of the MEC technology in these networks have been largely investigated in the literature as detailed above, but the impact of the MEC technology employment on the network energy efficiency is usually neglected. Meanwhile, the effect of the BSs switching on the MEC technology performance is ignored, as well. Indeed, these two topics are typically considered separately and their coexistence has not been investigated yet. For this reason, chapter 6 considers their simultaneous employment and provides an evaluation of their mutual effects. Using different traffic characteristics and different MEC server capacity, the growth of the energy consumption due to the supply of the MEC servers installed on each BS is analysed in chapter 6, as well as the effects of the BSs switching, and consequently of the MEC server deactivation, on the experienced delay. Association procedures, which aim at further improving the network energy

Table 2.3 Summary of references for UAV-BS networks topic.

[74–81, 39]	Overviews that cover different aspects of the UAV-aided networks.
[80, 81]	UAV-BS network design and planning in 3D aerial networks.
[82, 82, 4, 83, 84]	Study, evaluaiton and design of BH network in the UAV-BS-aided networks.
[85]	Prototype of a mm-Wave antenna steering for the BH link to an aerial BS.
[86]	Simulations of realistic access and BH links.
[87–90]	Optimisation for solving the BH constraints.
[9, 91–94]	Simulations which highlight the scarce on-board energy availability.
[40, 41]	Prototype of a solar-powered drone.
[95]	Simulations of a solar-powered drone.
[96]	UAV-BSs network, powered by PV panel on each drone, and maximisation of the system throughput.
[97]	Optimisation of the trajectory and resource allocation in a solar powered UAV-BSs network.
[98, 99]	Discretization of data flow, in DPs, and energy flow, in terms of EPs.
[100–102]	Markov model of a BS powered by renewable energy sources.
[103]	Formalization of the outage and the overflow probabilities due to lack of energy and the throughput.

efficiency, as well as the latency reduction, are proposed. The MEC switching is also introduced to guarantee load balancing among the active BSs.
A summary of the cited works is given in table 2.2.

2.3 UAV-BS networks

Recently, several studies about the concept of BSs mounted on UAVs, have been published [74–81]. Mozaffari et al. in [74] introduce an extensive tutorial about wireless networks aided by UAVs. In this work, the classification of a wireless network with UAVs, the open issues related to the channel modelling, the optimal trajectory and deployment, the network planning, and the resource management are presented. Authors in [75–77] provide diverse overviews that cover different aspects of the

UAV-aided networks, starting from the fundamentals of aerial channel modelling to specific performance indicators to evaluate aerial wireless networks. The survey in [78] focuses on the challenges of the physical aspects of multi-UAV networks such as the flight control and trajectory planning and cross-layer network design for mission control applications. The work in [79] presents the characterisation of UAV-BSs based networks, their architectures and the problematic of data routing, handovers and the impact of the energy efficiency, highlighting its impact on the UAV-BS network lifetime. In [80] and [81], the concept of 3D aerial networks is presented and discussed, providing a framework to properly UAV-BS network design and planning.

The study about the design of the BH network in the UAV-BS-aided networks is relatively recent. In [82] a multi-hop BH network is proposed to maximise the bit rate of the UAV-BS nodes through a network formation game algorithm. The authors in [82] provide a BH network for UAV-BS based on mm-Wave 3GPP standards, evaluating the impact of moving 3D objects on the propagation links of mm-waves. Works presented in [4] and [83] propose similar evaluations for LTE based BH links for UAV-BS. The authors agree that the increment of the users density leads to a deterioration of the BH network quality, due to the saturation of the Resource Blocks (RBs) of this network. Similarly, in [84], an optimal solution for an LTE in-band integrated access and BH network is evaluated for a single drone where the usage of UAV-BS enhances the throughput. Pokorny et al., in [85], present a prototype of a mm-Wave antenna steering for the BH link that provides connectivity to an aerial BS. Likewise, in [86] simulations and measured camping results are provided to explore realistic access and BH links. This work highlights that the flight time, the BH links and the mm-Wave alignments are the most challenging aspects of the UAV-BS. The BH constraints are solved through optimisation methods in [87–90].

In this work (chapter 7), we first investigate the performance of the UAV-BS-aided RAN in a crowded scenario. In order to evaluate it, we implement a simulation tool, which considers the real 3D environment located in the city centre of Ghent, in Belgium, using real traffic data to shape the traffic demand. While the solution is promising, two big challenges emerge from our results. First, the QoS is significantly deteriorated, during the traffic demand peaks, because of the BH links saturation. To address this issue, the MEC paradigm is considered (section 7.5). In particular, each UAV-BS is equipped with a MEC server, which is used to store the most popular contents. In case a user requires a content which is cached on the MEC server of

the UAV-BS, with which he/she is associated, that content is directly transmitted to the user, making the access to the cloud and to the BH network unneeded. The employment of MEC servers significantly improves the QoS, but its impact strictly depends on the traffic characteristics. In addition, this installation does not significantly impact the UAV-BS energy consumption, and, consequently, its lifetime.
The second issue which is highlighted by our simulation results is the scarce energy availability, provided on-board batteries. Because of this, each UAV-BSs needs an hourly replacement, because of the energy battery depletion. In literature, this issue is highlighted in [9, 91–94]. To overcome this, solar-powered UAV-BSs have been proposed in literature. Besides the studies discussed in section 2.1, about the usage of the solar energy for the RAN BS supply for the network self-sustainability, energy efficiency and reduction of the electricity bill [104, 105, 63–66], many studies discuss UAV-BSs, equipped with PV panels to produce electricity, without adding significant mass or size [39]. Authors in [40] and [41] prototype a solar-powered drone, which can fly for 28 hours, while in [95], authors simulate an UAV powered by a PV panel, via MATLAB/Simulink software, analysing also the effects of weather conditions. In [96], Sun et al. maximise the system throughput of a UAV-BSs network, powered by PV panels, which are installed on each drone, maximising the system throughput. In [97], the optimisation problem for the optimal trajectory and resource allocation in a solar powered UAV-BSs network is formulated and results highlight that the flight altitude is a trade-off in order to harvest more energy, without deteriorating the quality of the signal.
Also the modelling community focuses on the dynamics of the renewable energy harvesting in communication systems, to derive its formal representation, which is fundamental for its proper design, dimensioning and quantification of the effects of the changes of its components to satisfy the needed requirements. These works typically use a discretized representation of data flow, digitalized in DPs, as well as of energy, in terms of EPs, which correspond to the amount of energy which is needed to process a DP [98, 99]. Authors in [100] provides the Markov model of a BS totally powered by renewable energy sources. The model consists of a three-queue system: the first acts as an energy storage, the second is the data queue and the third is used as a reserve energy queue. An analytical model for a generic energy harvesting transmitter, equipped with an energy battery, where incoming EPs are stored, is presented in [103], which formulates the outage and the overflow probabilities due to lack of energy and the throughput. In [101, 102], in order to

study the performance of an energy-harvesting sensor node, a Markovian queuing system, composed by two queues, one used as energy battery and the other as data buffer, is proposed and analysed, to evaluate the impact of the energy harvesting process and the optimal data collection process, respectively, on the communication performances. Nevertheless, the survey in [39] highlights that the solar panel dimensioning and the understanding of the effect of intermittent energy production on the communication service provided by UAV-BSs network is usually neglected in literature. The proper size of the PV panel system is a challenging aspect. On one hand, it has to be large enough to provide the needed energy and extend the UAV-BS lifetime as long as possible. On the other hand, it can not be too large because of the UAV-BS payload constraint, i.e. the area and the weight which a UAV-BS is able to carry. For this reason, in this work (chapter 8), the queuing theory and a discretized representation of data and energy flows, as DPs and EPs are used to model an LTE MU-MIMO UAV-BS equipped with a PV panel, used as energy source. The PV panel dimensioning that is needed to satisfy the traffic demand and the probability that an EP is unused are analytically derived, also considering the time dependence of the traffic demand and the PV panel production.
A summary of the cited works is given in table 2.2.

Chapter 3

Greener RAN operation through Machine Learning

Part of the work presented in this chapter has already been published in:

- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2019). Greener RAN operation through machine learning. *IEEE Transactions on Network and Service Management, 16*(3), 896-908.

As discussed in chapter 1, the mobile access equipment, i.e. the BSs, has been identified as the most energy consumer actor in the communication network. The RoD mechanism is one of the most studied approaches, to reduce the BS energy consumption. It dynamically switches the BSs, according to the user traffic demand. It exploits the variability of the traffic demand in RAN, which presents peaks for relatively short periods, and long periods of resource under-utilisation that leads to waste of energy, since useless resources are kept active, consuming energy. The RoD is hence needed to dynamically adapt the available resources to the current traffic demand. While during traffic peaks all the available resources must be activated, when the traffic demand is low, unnecessary BSs are switched to sleep mode and some energy is saved. The RoD strategy has already proved to be very effective in optimising radio resource usage, reducing the RAN energy consumption and minimising the operational costs [12, 13]. Nevertheless, these works make the optimistic and unrealistic assumption that both the future traffic demand is exactly known.

In this chapter, we overcome this unrealistic assumption and consider a scenario in which the network operation is decided based on traffic predictions. These predictions are derived through ML algorithms trained with past traffic patterns. The estimation of traffic demand is clearly extremely critical: if traffic is underestimated, BSs deactivation may lead to QoS deterioration; conversely, if traffic is overestimated, lower energy saving than possible is achieved. Our key contribution is to provide a clear understanding of the effects on network operation of the introduction of ML algorithms for the forecast of future traffic demand. The general conclusions that come out of our investigation are that a large error in the traffic forecast does not always imply an increase in the energy consumption of the network or a deterioration of QoS. Many of the considered ML algorithms succeed in achieving a good trade-off between energy consumption and QoS. Results also highlight that energy savings strongly depend on traffic patterns that are typical of the considered area. This implies that a widespread implementation of these energy saving strategies without the support of ML would require a careful tuning that cannot be performed autonomously and that needs continuous updates to follow traffic pattern variations. On the contrary, ML approaches provide a versatile framework for the implementation of the desired trade-off that naturally adapts the network operation to the traffic characteristics typical of each area and to its evolution.
Section 3.1 describes our scenario and section 3.2 presents the employed traffic predictors. Details of the RoD strategy are given in section 3.3. Our methodologies are evaluated using the Key Performance Indicators (KPIs) defined in section 3.4. Results and comparisons among the used traffic forecast algorithms are discussed in section 3.5 and our findings are drawn in section 3.6.

3.1 RAN Management for Energy Saving

In this part of the work, we look at portions of a LTE-A RAN, structured as depicted in Fig. 3.1. As can be noticed from the figure, each RAN portion offers services over a specific service area, and consists of one cluster comprising one macro BS that defines a cell over the whole service area, and a few micro BSs that define small cells providing additional capacity in hot spots within the service area during peak traffic demand. Thus, the small cells area coverage overlaps with the coverage of the macro cell. This is a typical scenario considered for 5G and beyond RAN architectures that

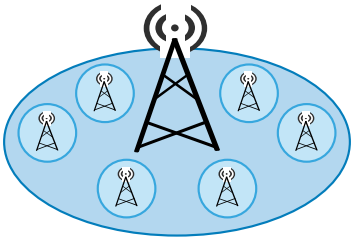


Fig. 3.1 A cluster composed by one macro BS and a few micro BSs.

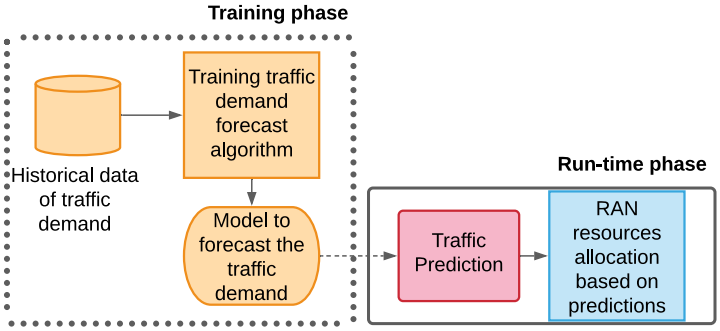


Fig. 3.2 Flowchart of the two-step network operation.

leverage small cell BSs exploiting high frequency bands (typically millimetre wave). In order to reduce the RAN energy consumption, so as to obtain both a reduction in energy cost, and a more energy parsimonious network operation, the centralised Management and Orchestration System (MANO) predicts the future traffic demand in the cluster and, according to these predictions, it applies some resource management strategy.

As shown in Fig. 3.2, the MANO operation consists of 2 steps:

1. **Training phase**. The algorithms used to predict the the traffic demand (orange shapes in Fig. 3.2) are trained using historical data. This phase is performed off-line.

2. **Run-time phase**. At the beginning of each time slot (that we assume to last 1h), the traffic demand is forecast, using the predictor previously trained (red rectangle in Fig. 3.2). Given this prediction, the network is operated according to the implemented power saving strategy (blue shape in Fig. 3.2).

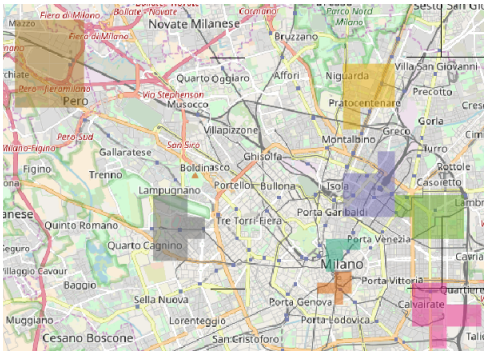


Fig. 3.3 Considered traffic areas: the Duomo di Milano (red), a business (dark green), a residential (yellow), the train station (purple), the San Siro (grey), the Politecnico di Milano (light green), an industrial (magenta) and the Rho (brown) areas.

3.2 Traffic Prediction

In this section, we present the traffic prediction tools used during the BS run-time as a preliminary step to the BS management decision. For the prediction of the traffic load of each BS, we consider and compare a number of machine learning techniques, which are trained in the training phase (see orange shapes in Fig. 3.2), using historical data.

3.2.1 Traffic Input Data

Data provided by a large Italian mobile network operator is used in this study. They report the traffic demand volume, in bits, of 1420 BSs, located in the city of Milan (Italy) and in a wide area around it, for two months in 2015, with granularity of 15 minutes. This time period includes typical weeks, as well as Easter week, when a brief vacation occurs. During typical weeks, people follow their usual working and activities routine. As a result, the considered period provides a good representation of the traffic demand volume dynamics. The traffic traces are normalised; hence, the peak of each traffic pattern is equal to the maximum capacity of each BS. Note that this is a pessimistic assumption with respect to energy saving possibilities, since the capacity of the network is usually overdimensioned. Then, data is aggregated, in order to obtain the hourly granularity.
For our work, eight portions of the city are selected, which are shown in Fig. 3.3. The traffic pattern of a BS taken in each of these areas during two weeks is reported

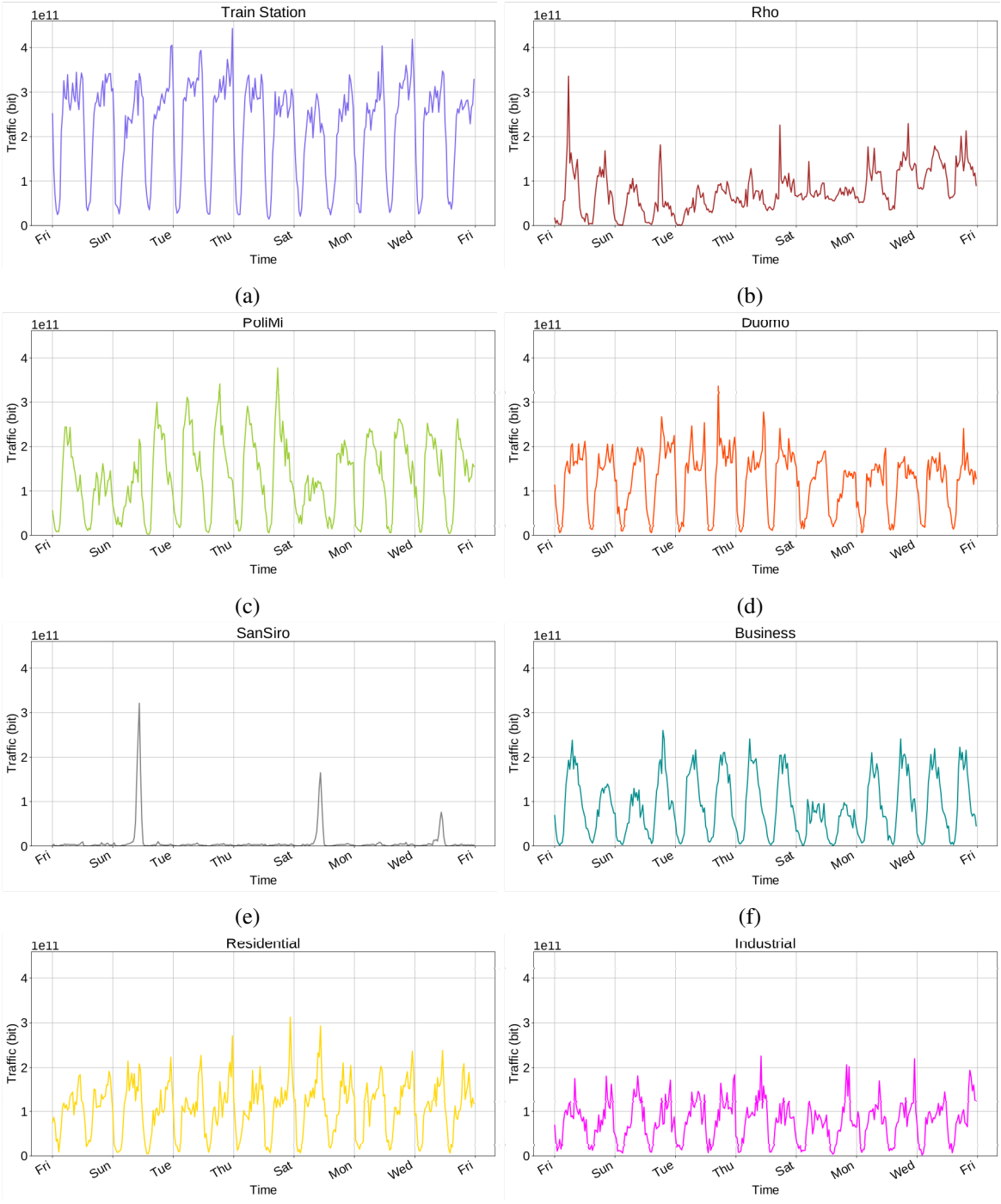


Fig. 3.4 Traffic pattern in the considered traffic areas: (a) Tran Station, (b) Rho, (c) PoliMi, (d) Duomo, (e) San Siro, (f) Business (dark green), (g) Residential, (h) Industrial areas.

in Fig. 3.4. These areas were selected as samples of quite different scenarios, and, hence, traffic patterns. All together, the selected areas are representative of the various zones that coexist in an urban environment.
The train station area (purple square in Fig. 3.3) is characterised by intense activity levels, especially at the end of the working hours, specifically at 8 p.m., as shown by the purple curve in Fig. 3.4a. The Rho district (brown in Fig. 3.3) is an area that hosts big events, fairs and exhibitions that last for a few days which determine the growth of the traffic demand, see brown curve in Fig. 3.4b. The Duomo di Milano area (red square in Fig. 3.3) is a touristic area, with high activity during several hours of the day. As indicated by the curve in Fig. 3.4d, it remains close to its peak value from 8 a.m. to 10 p.m. with no significant difference between weekends and working days. The Politecnico di Milano area (light green in Fig. 3.3) hosts a large campus with many students. Its peak during the weekends is about half than in working days, as shown by the light green curve in Fig. 3.4c. The San Siro neighbourhood includes a large soccer stadium (grey in Fig. 3.3), and the activity here is quite low, but with very high peaks according to the scheduled matches and concerts, see grey curve in Fig. 3.4e. A part of a business neighbourhood and some residential streets (dark green and yellow squares in Fig. 3.3) are also considered: the traffic in these areas follows the typical behaviour of people in their daily life. The peak of the business area on weekends is almost half that of the working days, whose traffic demand reaches the peak during the central hours of the day and sharply drops from 5 p.m. on, see dark green curve in Fig. 3.4f. In the residential area a traffic rise is observed in the evening, as highlighted by the yellow curve in Fig. 3.4g and during the week ends. Finally, the industrial zone (magenta in Fig. 3.3 and in Fig. 3.4h) is a particular case of a business area.
In each of these portions of the RAN, we assume that one macro BS and 6 small cell BSs are present, so that the service area is covered by one macro cell which overlaps with 6 small cells. To do this, for each area, we selected 7 traffic patterns recorded in that area. The trace which presents the highest traffic demand is chosen as the macro cell BS, while the remaining six as micro cell BSs.
After having investigated the performance of the considered ML approaches in support of network management decisions for these eight areas, we will look also into other areas, to make our conclusions more robust.

3.2.2 ML Approach for Traffic Prediction

In this work, different ML techniques are employed, to evaluate and compare their performances. The data of 47 days (out of the 61 for which we have data) are used for the training phase, i.e. for their training.

Block Linear Regression

Block Linear Regression (BLR) is proposed in [55] to forecast the traffic demand in wireless communication networks. This predictor reflects the daily and weekly periodicity of mobile traffic, and is formulated using linear regression. A single model is constructed to forecast the traffic of all the considered BSs, starting from the past traffic data of all BSs. As in [55], to predict the traffic demand of BS b at hour t of day d, the model receives as input 24 traffic samples, which are the traffic demand of the previous 24 hours.

Artificial Neural Network

The ANN model proposed in [106] is used, considering separate ANNs for each BS. Each ANN is composed of 4 layers: the input layer which has 5 nodes, 2 hidden layers with 4 nodes each, and the output layer with one node.
In order to predict the traffic demand, the ANN must be fed with carefully selected input features. As in [106], the input features are chosen according to the daily periodicity of the traffic and to the high correlation between consecutive samples. In particular, the selected input features are chosen according to the daily periodicity, as well as the high correlation between consecutive samples. In more detail, to predict the traffic demand of BS b at hour t, the following values are given as input to the network:

- $T^{(b,t-1)}$: traffic on BS b at hour t-1;
- $T^{(b,t-24)}$: traffic one day before the current time slot t on BS b;
- $T^{(b,t-1-24)}$: traffic one day before the time slot just passed, on BS b;
- $T^{(b,t-48)}$: traffic on BS b two days before the current time slot t;

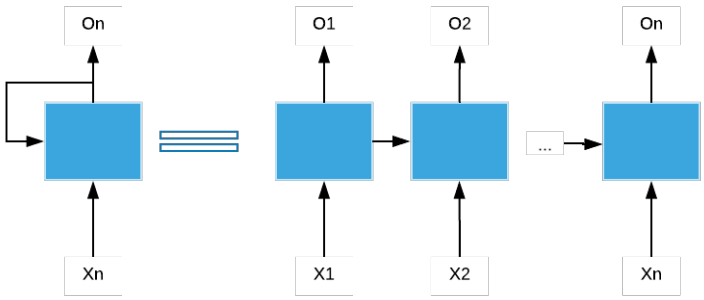


Fig. 3.5 Scheme of the LSTMC.

- $T^{(b,t-48-1)}$: traffic two days before the time slot just passed, at BS *b*.

Four different prediction methods are considered; for each BS:

- 1 ANN is used, without distinction between week-day and week-end traffic patterns;
- 24 ANNs are separately trained, one per each hour of the day;
- 2 ANNs are separately trained, one for the week-day traffic pattern and the other for the week-end pattern;
- 48 ANNs are separately trained, 1 for each of the 24 hours of the week-day traffic pattern and the remaining 24 for each of the 24 hours of the week-end pattern.

When the distinction between week-day and week-end patterns is considered, the inputs of the ANN, which belong to past days, are taken from the corresponding pattern. Therefore, the samples of past days of a week-end day (or a week-day) are taken in the previous week-end days (or a week-day).

Long Short Term Memory Cell

When Long Short Term Memory Cell (LSTMC) is used, RNN is trained for each BS. This kind of neural network is called recurrent since the output of the hidden layers is fed back into the network itself, as shown in Fig. 3.5, in which on the right

part we show the unrolled version of what the left part shows [107]. As in [107], the RNN is obtained with a single LSTMC layer. An LSTMC has an internal memory state which is added to the processed input. Therefore, the cell is responsible for remembering and computing the value of the state.

Baseline

In order to assess the effectiveness of the prediction approaches presented above, we need to compare their performance against some simple baseline cases. The simplest baseline estimation consists in computing for each BS the average traffic demand for each hour, and in using such an average as the predictor. In particular, for each BS *b*, an hourly predictor for week-days and an hourly predictor for week-ends are computed with the traffic samples belonging to the training data-set:

$$\overline{T}^{(b,t^*)} = \frac{1}{D} \sum_{\substack{d=0 \\ t\%24==t^*}}^{D} T^{(b,t)}, \qquad t^* = 1, ..., 24 \tag{3.1}$$

where $T^{(b,t)}$ is the traffic demand at the BS *b* at time interval *t* and D is the number of days available for the training phase, in the pattern considered (week-day or week-end). The traffic demand on BS *b* at hour *t* is forecast using the t^*-th sample of the baseline of *b*, given by t%24:

$$\hat{T}^{(b,t)} = \overline{T}^{(b,t\%24)} + D^{(b,t\%24)} \tag{3.2}$$

where $D^{(b,t)}$ is the difference between the baseline and the actual traffic demand of base station *b* at time *t*. The estimation of $D^{(b,t)}$ is given by:

$$D^{(b,t)} = K \cdot std^{(b,t^*)} \tag{3.3}$$

where K is a scalar parameter set to achieve always at least 90% of carried user data in a slot, and $std^{(b,t^*)}$ is the standard deviation at time t^*, computed as t%24, on BS *b*.

Table 3.1 Values of the parameters of the consumption model for macro and small cell BSs.

BS type	N_{trx}	P_{max} (W)	P_0 (W)	Δ_p
Macro	6	20	84	2.8
Micro	2	6.3	56	2.6

Baseline with ANN

This prediction algorithm is similar to the previous one, but differs in the way in which the estimation of $D_{b,t}$ is computed. In this case, $D_{b,t}$ is forecast using an ANN. To forecast the traffic demand of BS b at hour *t*, the ANN inputs are the following:

- $\overline{T}^{(b,t)} - T^{(b,t)}$: difference on BS *b* between the baseline and the actual traffic demand at the time slot just past;
- $\overline{T}^{(b,t)} - T^{(b,t-24)}$: difference on BS *b* between the baseline and the traffic one day the current time slot;
- $\overline{T}^{(b,t)} - T^{(b,t-1-24)}$: difference on BS *b* between the baseline and the actual traffic demand one day before the time slot just past;
- $\overline{T}^{(b,t)} - T^{(b,t-48)}$: difference on BS *b* between the baseline and the actual traffic demand two days before the current time slot;
- $\overline{T}^{(b,t)} - T^{(b,t-1-48)}$: difference on BS *b* between the baseline and the actual traffic demand two days before the time slot just past.

3.3 Energy Reducing Strategy

In the run time phase, at the beginning of each time slot, the future traffic demand is forecast as described in the previous sections. The power saving strategy is employed to allocate the needed network resources, making decisions based on the predictions. We consider the **RoD** strategy, proposed in [44]. It aims at minimising the network energy consumption by adapting the network capacity to the traffic demand. When the traffic demand of a micro cell BS is low, that micro cell BS is switched in sleep mode and its corresponding traffic is carried by the macro cell. In this thesis the traditional sleep mode is considered. It completely deactivates the BS, i.e. it puts

it in a deep sleep state where the BS is out of operation, with negligible power consumption This sleep mode differs from the Advanced Sleep Modes, typical of 5G BSs, see [108], since they consist of different levels of sleep mode, which gradually deactivate the BS components and differ for energy consumption, time of duration, activation and deactivation to return to the active state for the BS.
In order to apply the RoD, at the beginning of each time slot, the traffic demand of each BS is predicted. Then, for each micro cell BS, if both the following conditions are verified, that micro BS is put in sleep mode:

- The forecast traffic load of the considered micro cell BS is lower than the threshold ρ^*. As demonstrated in [44], the optimal value of ρ^* is 37% of the maximum load of the BS. This threshold depends on the energy consumption per carried bit: when the traffic is below ρ^*, the energy needed to carry a unit of traffic in the micro cell is larger than in the macro, so that it is more convenient to switch off the small cell BS, if this is possible in terms of total capacity. In particular, $P(t,\rho(t))$ is the input power required for the operation of a BS at time t, when its traffic load is $\rho(t)$, defined as the ratio between the traffic carried by the BS, in bit/s, and the BS capacity, in bit/s. It is derived according to the EARTH model proposed in [109]:

$$P(t,\rho(t)) = N_{trx} \cdot [P_0 + \Delta_p P_{max} \rho(t)], \qquad 0 \leq \rho \leq 1 \tag{3.4}$$

  where N_{trx} is the number of transceivers, P_0 represents the power consumption, in watt, when the radio frequency output power is null, Δ_p is the slope of the load dependent power consumption. P_{max} is the maximum radio frequency output power, in watt, at maximum load. Table 3.1 summarises the value of the parameters for macro and small cell BSs [109].
  Given $\rho_m(t)$ and $\rho_M(t)$, the traffic load on a micro BS and on a macro cell BS, respectively, switching that micro cell BS off is convenient if:

$$P(t,\rho_m(t)+\rho_M(t)) < P(t,\rho_m(t)) + P(t,\rho_M(t)) \tag{3.5}$$

  It follows that the deactivation of a micro cell BS saves energy if:

$$\rho_m(t) < \frac{N_{trx,m} \cdot P_{0,m}}{(N_{trx,M} \cdot \Delta_{p,M} \cdot P_{max,M}) - (N_{trx,m} \cdot \Delta_{p,m} \cdot P_{max,m})} \tag{3.6}$$

where $N_{trx,m}$, $P_{max,m}$, $P_{0,m}$, $\Delta_{p,m}$ and $N_{trx,M}$, $P_{max,M}$, $P_{0,M}$, $\Delta_{p,M}$ are, for the micro and macro cell BS respectively, N_{trx}, P_{max}, P_0, Δ_p, respectively. Adopting the parameter values reported in Table 3.1, the optimal value for ρ^* is 0.37. This means that when the traffic load on the micro is larger than 0.37, the additional cost to carry the traffic of the micro BS through the macro BS does not compensate for the savings that can be achieved by switching off the micro BS.

- In addition to the condition discussed above, a micro cell BS can be deactivated if the available capacity of the macro cell BS is enough to carry the traffic of the considered micro cell BS.

Each of the considered micro cell BSs is analysed for its possible deactivation as described above, starting from the least loaded to the most loaded, in the following hour. Given that the load of a micro BS is lower than ρ^*, its energy consumption per bit is larger, if its load is smaller. Thus, giving larger priority to micro BSs in the deactivation procedure leads to minimum network energy consumption [44].

3.4 Key Performance Indicators

Our scenario and methodologies are evaluated in terms of the following Key Performance Indicators (KPIs).

Average Relative Error

In order to evaluate the accuracy of the prediction algorithms, we use ARE (Average Relative Error), which measures the average ratio between the real and predicted traffic patterns. It is computed as:

$$\mathrm{ARE} = \frac{1}{N_{BS}} \sum_{b=1}^{N_{BS}} \mathrm{RE}^{(b)} \tag{3.7}$$

where N_{BS} is the number of the considered BSs and $RE^{(b)}$ is the RE (Relative Error) on BS b, derived as:

$$RE^{(b)} = \frac{1}{H}\sum_{t=1}^{H}\frac{|T^{(b,t)} - \hat{T}^{(b,t)}|}{T^{(b,t)}} \tag{3.8}$$

where $T^{(b,t)}$ is the real traffic demand at time t on BS b, $\hat{T}^{(b,t)}$ is the forecast traffic demand at time t on BS b, H is the duration of the testing period, in number of time slots.

Average Mean Error

In addition to ARE, the AME (Average Mean Error) is employed as an error metric defined as:

$$AME = \frac{1}{B}\sum_{b=1}^{N_{BS}} ME^{(b)} \tag{3.9}$$

where N_{BS} is the number of the considered BSs and $ME^{(b)}$ (Mean Error) on BS b is derived as:

$$ME_b = \frac{1}{H}\sum_{t=1}^{H}(T^{(b,t)} - \hat{T}^{(b,t)}) \tag{3.10}$$

where $T_{b,t}$ is the real traffic demand at time t on BS b, $\hat{T}_{b,d,t}$ is the forecast traffic demand at time t on BS b, and H is the duration of the testing phase. In the AME, negative and positive values are possible; indeed, AME indicates whether the considered algorithm systematically overestimates or underestimates the traffic. When the AME is positive, the algorithm tends to underestimate the traffic demand, while in case of negative values of the AME, the traffic is usually overestimated. AME is complementary to ARE which is a measure of the prediction error in absolute terms.

Average capacity

The average capacity of the BS cluster is computed again considering that in each time slot some BSs are active and provide capacity, while some others are sleeping and do not provide capacity.

Energy Consumption

The energy consumption of the BS cluster, E_{TOT}, in watt, is computed considering that in each time slot some BSs are active and consume energy, while some others are in sleep mode and consume no energy. It is given by:

$$E_{TOT} = \sum_{t=1}^{H} \sum_{b=1}^{N_{BS}} E_{TOT}^{(b,t)} \tag{3.11}$$

where $E_{TOT}^{(b,t)}$ is the energy consumption of BS *b*, at time *t*; it is computed as in Eq. (3.4). N_{BS} is the number of the considered BSs, H is the number of time slots of the testing phase. The energy consumption of a BS in sleep mode is considered negligible.

Energy Consumption Reduction

When the resource allocation strategies presented in section 3.3 are used, in each time slot some BSs are active and consume energy, while some others may be in sleep mode and thus consume no (or very little) energy. The energy consumption is computed as described above. In order to measure the effectiveness of the energy reducing strategy, the energy saving is computed. It is calculated with respect to the *Always ON* scenario: this is the case in which all BSs are always active regardless of the amount of traffic demand. It is computed as follows:

$$E_{RED} = 100 \cdot \frac{E_{ON} - E_{TOT}}{E_{ON}} \tag{3.12}$$

where E_{ON} is the energy consumption in the Always ON scenario; E_{TOT} is the energy consumption with the considered strategy, computed as detailed above.

Lost Traffic

In the Always ON scenario, each BS is always active and able to carry its traffic demand. In case resources are dynamically allocated according to the strategies described above, the situation is different. The Lost Traffic, L_T, is defined as the percentage of the traffic demand that cannot be carried by the network, accounting for the fact that in each time slot some BSs are active and can handle their traffic

demand, while some others may be off and thus cannot provide any service. Let us define the traffic that overflows from the micro cell BS b to the macro cell as:

$$O^{(b,t)} = \begin{cases} T^{(b,t)} & \text{if b is in sleep mode} \\ 0 & \text{if b is active} \end{cases} \tag{3.13}$$

the lost traffic is given by:

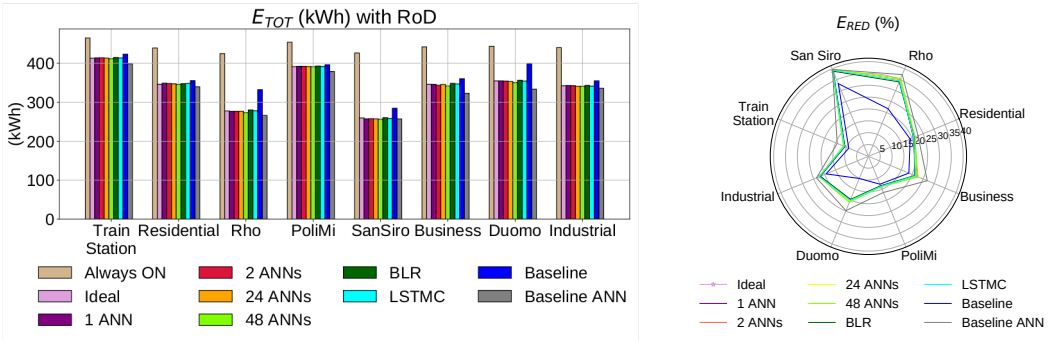
$$L_T = \frac{\sum_{t=1}^{H} \max(0, T^{(B,t)} + \sum_{b=1}^{N_{BS}} O^{(b,t)} - C)}{\sum_{t=1}^{H} \left(T^{(B,t)} + \sum_{b=1}^{N_{BS}} T^{(b,t)} \right)} \cdot 100 \tag{3.14}$$

where C is the capacity of a macro BS and H the duration of the testing phase. The lost traffic is the percentage of traffic that cannot be carried by the macro cell BS when traffic overflows from deactivated small cell BSs.

The performance indicators obtained using the traffic and energy predictions provided by the proposed algorithms are compared to each other. In addition, they are compared with the performance of the *Always ON* scenario, where all the BSs are always active regardless of the amount of traffic demand and produced energy. Moreover, the achieved results are compared with the *Ideal Case* in which the RoD energy saving strategy is applied, according to the perfect knowledge of the future traffic demand.

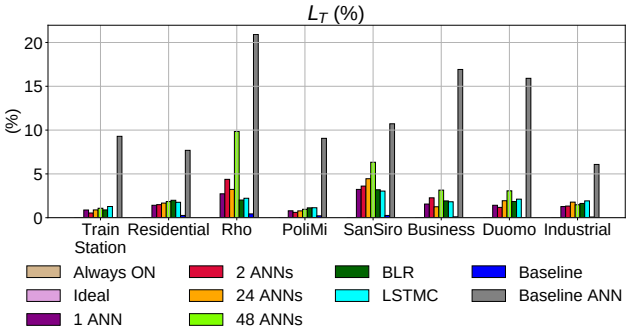
3.5 Performance Analysis

In this section, we compare the accuracy of the traffic prediction algorithms and we investigate their effectiveness when introduced in the RoD strategy of the considered RAN portions. As previously mentioned, we use the data of 47 days for the training phase, while the remaining 14 days for the run-time phase.



(a) (b)

Fig. 3.6 Comparison of the effectiveness of prediction techniques under RoD: (a) Energy consumption E_{TOT}, (b) Energy consumption reduction.



(a)

Fig. 3.7 Comparison of the effectiveness of prediction techniques under RoD: Percentage of lost traffic L_T.

3.5.1 Comparison among ML algorithms

Figs. 3.6 and 3.7 report the results of the application of the RoD strategy coupled with the various traffic forecast algorithms, in each one of the 8 considered RAN portions. The plot of Fig. 3.6a shows the energy consumption values, while the radar plot of Fig. 3.6b reports the corresponding energy consumption reduction, computed with respect to the Always ON scenario, in which no action is taken to reduce the energy consumption. Fig. 3.7 shows the percentage of lost traffic.

The figures indicate that, in all the considered areas, the benefit of the proposed approach is significant, with energy saving up to 40% and never below 10%. Comparing the effectiveness of the different algorithms, results indicate that, with all the considered traffic forecast algorithms, except for the Baseline with and without ANN, energy consumption is very close to the ideal case (where we assume perfect

knowledge of future traffic demand). The Baseline with and without ANN show the highest and the lowest energy consumption drops, respectively. From Fig. 3.7 we see, however, that energy reduction is achieved at the cost of QoS deterioration. With the exception of the Baseline with ANN, the other ML algorithms have equivalent performance within each zone, with QoS deterioration that depends on the area but is usually below 5%. When BLR, LSTMC and one or more ANNs are used, differences among the network performance indicators are limited, so that these approaches can be considered equivalent. Even if one of them may provide a more energy consuming configuration during some hours, it will provide a less energy consuming one during others.

Observe now in Fig. 3.8 the ARE (i.e., the error of the ML algorithms in predicting the traffic). ARE is, in general, quite small. Interestingly, the Baseline, despite showing good QoS (low lost traffic), reaches 3.17 in ARE, which is the largest error among the considered prediction techniques. This happens because for our combined energy saving and QoS goals, the correct estimation of traffic is important only around the values that are taken as thresholds for the decision to switch on or off some small cell BSs. Large errors in traffic estimations in periods of low or high traffic are irrelevant, since no network operation action is taken. This is revealed also in Fig. 3.9, where the hourly lost traffic and the hourly ARE, for the Industrial zone, obtained with Baseline and 1 ANN are plotted. The figure highlights that peaks in the ARE do not correspond to peaks of lost traffic. Indeed, at 2 a.m., the Baseline presents the maximum ARE. Nevertheless, no traffic is lost. Moreover, even if the Baseline usually presents higher ARE than 1 ANN, it does not provide larger lost traffic. For example, at 10 p.m., even though 1 ANN has lower ARE than the Baseline, it causes larger lost traffic. Therefore, a larger error could not imply the deterioration of the quality of service.

Now, we analyse the pros and cons of each ML algorithm. They are reported in Table 3.2. The table contains the number of trained models. As mentioned in Section 3.2, a single model is trained when BLR is used. When LSTMC or the Baseline are used, a model is employed for each BS. The number of models with ANN varies: 1, 2, 24, or 48 models are needed, when 1 ANN, 2 ANNs, 24 ANNs, or 48 ANNs are used, respectively. In the case of Baseline with ANN, 2 models are trained for each BS. As a result, when BLR is used, increasing the number of BSs does not increase the number of trained models, but the dimension of the training set. When the other ML approaches are employed, the number of trained models linearly grows with the BSs.

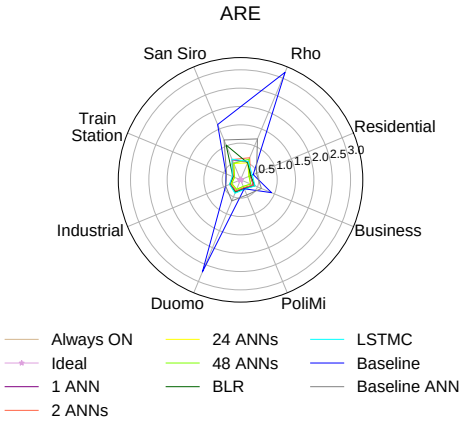


Fig. 3.8 ARE of the traffic forecast algorithms.

Increasing the number of trained models, the computational power and time increase. Nevertheless, the training of prediction models is computed off-line and for this reason the time for the training phase does not represent a strict constraint for this application. Moreover, the table includes the adaptability to variation of each traffic pattern. It means that when a traffic pattern modifies for example its mean value, the ML algorithm is able to detect this variation. In Fig. 3.10a, a traffic trace which changes its mean value in time is plotted, as well as the predicted samples obtained with each considered forecast algorithm. The figure shows that the Baseline and the Baseline with ANN are not able to follow the actual trace. This is because the Baseline does not use recent samples for the forecast of the current one. The table also indicates if hyperparameters are needed for the considered algorithm. When ANNs or the LSTMC are used, the number of hidden layers, as well as the number of neurons for each layer should be decided. If the Baseline and the BLR are chosen, no hyperparameters are needed. Therefore in these two cases, the design of the model is easier.

3.5.2 Impact of traffic patterns

We now investigate the impact of traffic patterns. As already visible in the previous figures, there is quite some difference among the performance indicators in the various geographical areas. To better catch this, Fig. 3.11 combines energy consumption and lost traffic by representing each prediction algorithm in each area with a marker

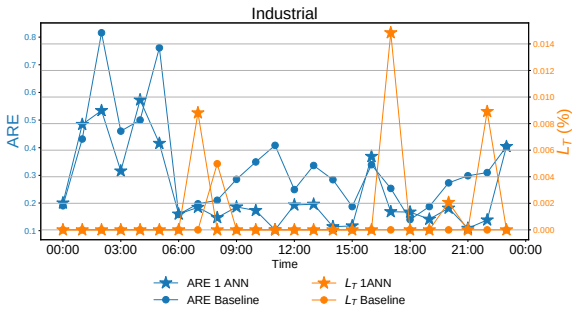


Fig. 3.9 Hourly ARE and hourly lost traffic in the Industrial zone

Table 3.2 Pros and cons of each ML algorithm.

ML algorithm	Number trained models	Adaptability to variation	Hyper parameters
1 ANN	1 for each BS	Yes	Yes
2 ANNs	2 for each BS	Yes	Yes
24 ANNs	24 for each BS	Yes	Yes
48 ANNs	48 for each BS	Yes	Yes
BLR	1	No	No
LSTMC	1 for each BS	Yes	Yes
Baseline	1 for each BS	No	No
Baseline with ANN	2 for each BS	No	Yes

positioned so that the x-axis value corresponds to the energy consumption and the y-axis value corresponds to the percentage of lost traffic. We clearly see that results are clustered according to the geographical area. The area-dependent behaviour derives from the traffic patterns that characterise the areas. The Train Station and PoliMi areas have a typical traffic demand which is larger than the threshold for many hours (around 40% of the testing phase). Because of this, the energy consumption can not be reduced significantly (Fig. 3.6b), since each BS is needed to carry the traffic and cannot enter sleep mode. In the Business, Duomo, Residential and Industrial zones the traffic pattern is larger than the threshold for shorter periods than in the previous case. For this reason, the RoD strategy results in more effective energy consumption reduction, as can be seen in Fig. 3.6b. The average (among all ML algorithms) ARE in the Industrial area increases by 23% with respect to the one given by the Train Station area (Fig. 3.8). Nevertheless, Fig. 3.7 reveals that the lost traffic is almost the same: 1.9%, on average. This results from different average distances from the threshold. Indeed, in the Train Station area the average distance of the hourly amount

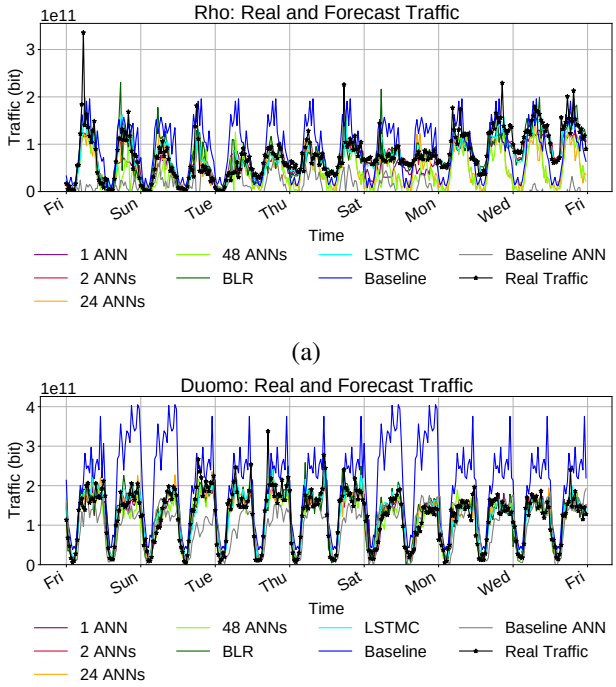


Fig. 3.10 Forecast of the traffic demand: Rho (a) and Duomo (b)

of traffic is approximately $7.8 \cdot 10^{10}$ bit, while in the Industrial zone it is larger than $10 \cdot 10^{10}$ bit. Thus, even presenting larger ARE, an incorrect deactivation is less frequent thanks to the large distance from the threshold, preventing the deterioration of the QoS. Finally, the RAN portions that we called San Siro and Rho exhibit the most favourable patterns, with long periods of low traffic demand. This leads to the largest reductions of energy consumption (see Fig. 3.6b), since, in these areas, during the day, the traffic demand is usually lower than the threshold. Nevertheless, these areas present sometimes unpredictable and random very high peaks which result in the largest ARE and the worst QoS (Fig. 3.6b, 3.8). Because of the shortness of the duration of the data trace, the prediction models can not learn these peculiar and irregular behaviours.

We now consider two quite different areas: Rho and Duomo (see Figs. 3.4b, 3.4d, respectively). Figs. 3.12a and 3.12b highlight the inter-relations among the different performance indicators obtained with RoD, using different traffic demand forecast algorithms in each area. As expected, the performance in the two areas is significantly different for all indicators. This means that a widespread implementation of energy efficient strategies without the support of ML algorithms would require a complex

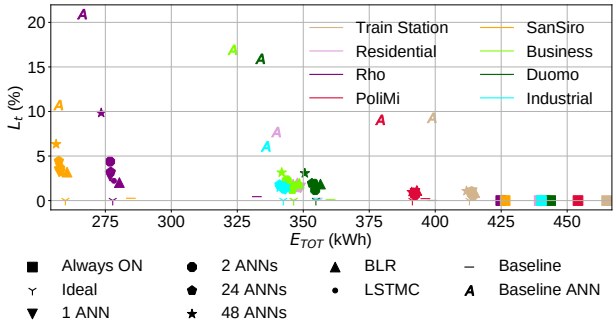
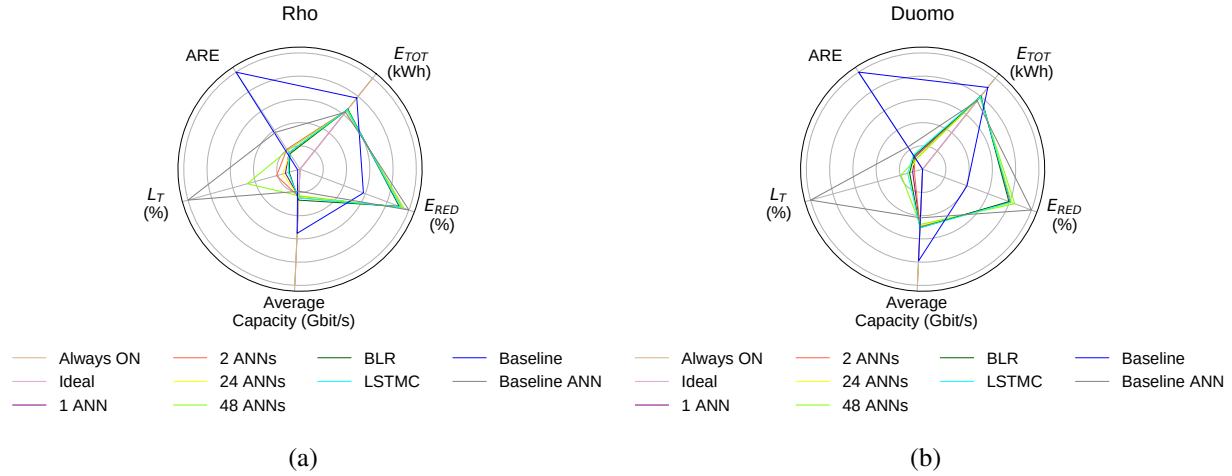


Fig. 3.11 Energy consumption and lost traffic, with different prediction techniques under RoD.



(a) (b)

Fig. 3.12 Comparison among performance indicators under RoD for two quite different zones: Rho (a) and Duomo (b).

tuning of the network management algorithms. In addition, since traffic patterns change with time, a careful adaptation to these variations would also be necessary. An automatic learning process of the traffic patterns, and a consequent automatic tuning of the energy efficiency strategies, are clearly necessary.

We can again observe from Figs. 3.12a and 3.12b, that higher values of ARE do not always imply higher energy consumption or worse QoS. Indeed, energy consumption and amount of lost traffic depend on the configuration of the network, which can be correct (i.e., the same as in the ideal case), even with a large error in the traffic forecast.

The average capacity of the considered RAN portion, which depends on the average number of active BSs, and the AME, which depends on the prediction trend, provide information about the behavior of the traffic forecast algorithms. Large average

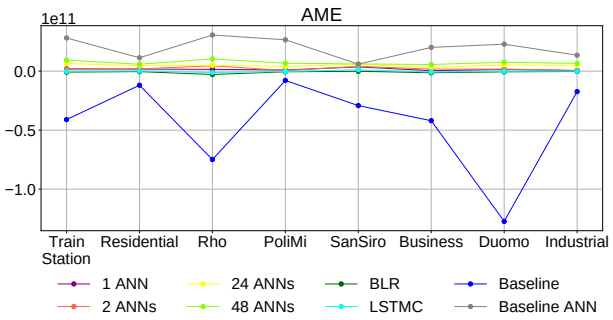


Fig. 3.13 AME of the traffic forecast algorithm.

capacity means that the conditions to put in sleep mode the micro BSs are verified infrequently. The same happens when negative AME is achieved. In such a case, the forecast traffic may be overestimated. On the contrary, low average capacity and positive AME indicate traffic underestimation. These trends affect QoS and energy consumption. A detailed comparison of the traffic forecasts for the areas Rho and Duomo is shown in Figs. 3.10a and 3.10b, respectively. When the Baseline algorithm is used, traffic is overestimated and, hence, the average capacity is always larger than in all the other cases (blue curve in the figures) and the AME is always more negative than in all the other cases (blue curve in Fig. 3.13). The energy consumption drops by 10% and 22%, respectively in the two areas, without any lost traffic. In the cases of BLR, LSTMC and one or more ANNs, the average capacity decreases, since the traffic is better estimated. In such a case, the AME (Fig. 3.13) is closer to zero. This means that the overestimated samples compensate for the underestimated ones. The average capacity further decreases when the Baseline is combined with the ANN. The AME reaches large positive values, since this model underestimates traffic.

3.5.3 Performances evaluation over many additional areas

The last part of this work consists in the evaluation of the KPIs obtained with the proposed methodology over 8 additional areas to test the robustness of our considerations. We use 8 new portions of the RAN that are indicated in Fig. 3.14. As before, each area is covered by one macro cell BS which overlaps with 6 micro cell BSs. The results are presented in Fig. 3.15, where the power consumption and the percentage of lost traffic are plotted, using RoD. Each area is identified by an integer. Integers from 1 to 8 correspond to the previously considered areas numbered

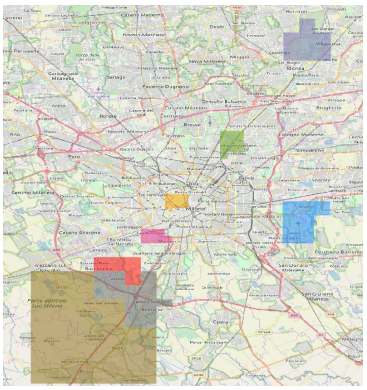


Fig. 3.14 Considered traffic areas: a touristic (orange), a theatre (green), a residential (magenta), a portion of Monza city (purple), the Mediolanum Forum sports facility in Assago (red), the Milano's airport (light blue), an highway (grey) and a park (brown) areas.

according to the following order: 1 - Train Station; 2 - Rho; 3 - Duomo; 4 - PoliMi; 5 - San Siro; 6 - Business; 7 - Residential; 8 - Industrial areas. Numbers from 9 to 16 are related respectively to: 9 - an area rich in tourist attractions (orange in Fig. 3.14); 10 - a urban zone, including a theatre (green); 11 - a portion of the Assago village, where a famous sports facility is located (red); 12 - a residential neighborhood (magenta); 13 - an area containing a section of the highway (grey); 14 - the zone of the Milano's airport (light blue); 15 - a district in Monza city (purple); 16 - a park (brown).

Results are similar to the ones presented in the previous sections. When RoD is used, the power consumption is reduced by up to 40% and never below 9% (Fig. 3.15a). The reduction in power consumption strictly depends on the traffic pattern. The ML algorithms do not impact the energy consumption: each of them provides a reduction very close to the ideal case. As shown in Fig. 3.15b, this reduction is achieved at the expense of a deterioration of QoS: some traffic is lost, but usually not more than 5%, except for the Baseline with ANN. As before, the Baseline combined with the ANN provides the highest power consumption reduction, but the worst QoS: up to 20% of traffic is lost (Figs. 3.15a, 3.15b).

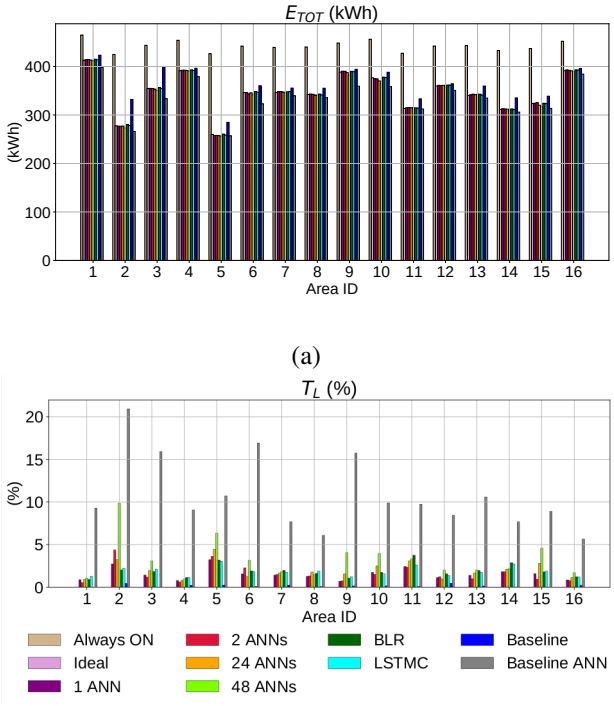


Fig. 3.15 Comparison of the energy consumption and the percentage of lost traffic: (a) Energy consumption and (b) Percentage of lost traffic with RoD.

3.6 Final remarks

In this part of the work, we considered different portions of a RAN providing services in the city of Milan, each one corresponding to a cluster comprising one macro cell and 6 micro cell BSs.
We investigated the effectiveness of ML algorithms to forecast the future traffic demand. The predictions are used by the MANO, to allocate the RAN resources according to the RoD, used as dynamic resource allocation, to reduce the RAN energy consumption. In particular, the RoD approach activates the minimum number of small cell BSs necessary to satisfy the forecast traffic demand. Our results show that ML algorithms are necessary to achieve a good trade-off between energy efficiency and QoS. Moreover, results point out that large errors in the forecast do not always imply bad network performance. Indeed, the correct estimation of traffic is important only around the values that are taken as thresholds for the decision to activate or deactivate some micro cell BSs. This makes the RAN performance slightly sensible to the used ML algorithm.
Obviously, forecast algorithms that tend to overestimate traffic, yield lower energy saving without deteriorating QoS. On the contrary, forecast algorithms that tend to underestimate traffic yield losses of traffic, but higher energy saving: up to 5 percentage points more than the ideal case.
Results also reveal that energy savings strongly depend on traffic patterns that are typical of the considered area. This implies that a widespread implementation of these energy saving strategies without the support of ML would require a careful tuning that cannot be performed autonomously and that needs continuous updates to follow traffic pattern variations. On the contrary, ML approaches provide a versatile framework for the implementation of the desired trade-off that naturally adapts the network operation to the traffic characteristics typical of each area and to its evolution.

Chapter 4

ANN Traffic Predictions Processing for RAN Energy Efficiency

Part of the work presented in this chapter has already been published in:

- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2019). Greener RAN operation through machine learning. *IEEE Transactions on Network and Service Management, 16*(3), 896-908.
- Vallero, G., Renga, D., Meo, M., & Ajmone Marsan, M. (2020, November). Processing ANN Traffic Predictions for RAN Energy Efficiency. In *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (pp. 235-244).
- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2021). RAN energy efficiency and failure rate through ANN traffic predictions processing. *Computer Communications*.

In the previous chapter, we show that the ML approaches are a necessary support to solve the problem of RAN management for network energy efficiency. Indeed, they provide traffic predictions, which allow the activation of the minimum number of micro cell BSs, necessary to satisfy the traffic demand. This means that when the

traffic is predicted to be small enough, some micro cell BSs can be deactivated, and their traffic is carried by their macro BSs that remain active. Conversely, when the forecast traffic grows and additional capacity is needed, some BSs in sleep mode are re-activated. Results pointed out that a good trade-off between energy efficiency and QoS is achieved when ML is used for the traffic prediction. Nevertheless, the results showed a limited sensitivity to the type of employed ML techniques. Indeed, critical BS (de)activation decisions are taken in correspondence of specific traffic values, and high accuracy in the estimations is not required in general, but only close to the values where decisions are taken. For this reason, in this chapter, to significantly improve the performance, traffic predictions are carefully processed in order to understand the overall pattern. With this objective, we use traffic predictions over a shorter time scale and their processing and results show that this is fundamental in order to significantly improve the QoS.
In addition, we also introduce the BS failure rate as a new variable in the design space of RAN management. Activating and deactivating BSs has an impact on BS failure rate. On the one side, switching is harmful to the BS failure rate; on the other side, the time spent in sleep mode saves the BS from deterioration. The balance between these two phenomena is discussed in this chapter. It depends on the HW components of the BS, as well as on the RAN management strategy.
Section 4.1 presents the scenario of our study. We model each BS as defined in section 4.2 and the employed RoD strategies use traffic predictions, obtained with the tools that are presented in section 4.3, and on the prediction processing algorithms that are reported in section 4.4. After presenting performance indicators in section 4.5, results are discussed in section 4.6. Section 4.7 summarises our findings.

4.1 Scenario

As in chapter 3, a portion of an heterogeneous LTE RAN is considered, comprising one macro cell BS, and a few small cell BSs, whose coverage overlaps with the macro cell, see Fig. 3.1. Small cell BSs are deployed to provide additional capacity during high traffic demand periods. As in the previous chapter, the MANO decides the activation of resources (i.e., micro cell BSs), according to predictions of the future traffic demand. In this part of the work, these predictions are performed on a temporal horizon of 15 minutes (which is the time granularity chosen by the operator

whose data we used in this work, and is thus taken as the time slot). According to the RoD energy saving strategy, micro cell BSs can be switched on and off to reduce the RAN energy consumption, with attention to the QoS. This means that, when not all the capacity is needed to satisfy the predicted traffic demand, some small cell BSs are put in sleep mode. On the contrary, all BSs are activated in those periods in which all the capacity is required for the traffic demand satisfaction. The activation/deactivation of a BS cannot occur at intervals shorter than one hour, to avoid too frequent switches.
In this part of the work, only ANNs are employed for traffic predictions. Nevertheless, these predictions are also processed before reaching a decision about BS activation and deactivation. Similar to what depicted in Fig. 3.2 and presented in chapter 3, the BS management works in two phases. First, **Training phase**, performed only once, as a preliminary step of our online management system. In this step, the ML algorithm used to predict the traffic demand is trained using historical data. Then, the **Run-time phase**, which uses the previously trained ANN for the BS activations or deactivations. During this phase, at every time slot, i.e., every 15 minutes, the traffic demand is forecast for the following 4 time slots through the predictor trained in the training phase. In addition to what has benn done in chapter 3, these four predictions are processed by the MANO, in order to understand the traffic pattern, detect the overall trend and correctly decide which small cell BS must be active in the next hour.

4.2 Modelling the BS

The input power, in watt, required for the operation of a BS at time slot t, denoted as $P_{in}(t)$, is derived according to the EARTH linear model proposed in [109] and described in section 3.3. According to this model, the BS energy consumption is given by a fixed and a traffic load dependant component.
We adopt the model of the BS failure rate depicted in [58] and [110]. The model treats the BS as a whole, i.e. as a single entity, rather than using a model for each single component of the BS, even if this assumption results in a less detailed model, which does not specify the dependencies among the BS components. The failure rate of a BS, and in general, of an arbitrary device, is given by the Arrhenius law [111],

in failure/hour:

$$\gamma = \gamma_0 e^{-\frac{E_A}{KT}} \tag{4.1}$$

where γ_0 is the failure rate, in failure/hour, assuming an infinite operating temperature, E_A is the minimum activation energy, in joule, K is the Boltzmann constant, in joule/kelvin, and T is the operating temperature, in kelvin. From the model, we notice that putting a BS in sleep mode positively impacts the failure rate, since the operating temperature of a device in sleep mode is usually lower than in normal operating condition.

Nevertheless, the power state change negatively affects the failure rate, tending to increase it [58, 110, 57]. This is because, as explained in [112], the metal struggles with temperature variations and, in particular, to state cycling. For a BS, which is dynamically activated/deactivated, a state cycle is the variation of operating temperature between the active and sleep mode and is modelled with the Coffin-Manson equation, in failure/hour:

$$\gamma_{TR} = \frac{f_{TR}}{N_F} \tag{4.2}$$

where f_{TR} is the frequency of the thermal cycling, in failure/hour. This means that it refers to the periodicity between two consecutive events of the same type (switch on-switch on, i.e. the cycle is from the time in which the BS is activated to the next time in which it is reactivated after a deactivation). N_F is the number of cycles to failure, computed as

$$N_F = C_0 \left(\frac{\Delta_T - \Delta_{T_0}}{F}\right)^{-q} \tag{4.3}$$

where C_0 is a constant which depends on the material and q is the Coffin-Manson exponent. Δ_T is the temperature variation of the cycle, Δ_{T_0} is maximum variation of temperature which avoids variation of the failure rate and F is a factor of temperature normalisation.

In order to derive the total failure rate of a BS, Eqs. (4.1), (4.2) and (4.3) are used, assuming that the failure rates due to the different effects, i.e. to the time spent in sleep mode and the switching, are statistically independent from each other and act in an additive manner. As a result, the resulting failure rate of a aBS b, $\gamma^{(b)}$ is computed by:

$$\gamma^{(b)} = (1 - \tau_S)\gamma_{ON} + \tau_S \gamma_S + \frac{f_{TR}}{N_F} \tag{4.4}$$

where τ_S is the fraction of time the device spends in sleep mode, γ_{ON} and γ_S, in failure/h, are the failure rates when the BS is active and in sleep mode, respectively,

computed with the Arrhenius law, see (4.1). The parameter f_{TR}, in cycle/h, is the frequency of the sleep mode cycle and N_F, in cycle/failure, is the number of cycles supported by the device before a failure occurs, derived as in (4.3). Usually, to measure the impact of the device switching on its lifetime, the Accelerator Factor (AF) is estimated. This indicator provides the mean lifetime increase/decrease with respect to the always on condition, as the ratio between the resulting failure rate and the failure rate of the always on scenario. A value of AF larger than 1 means that the failure rate increases, while a value smaller than 1 indicates that the failure rate decreases. Similar to (4.4), the resulting AF is given by two contributions: the time spent in sleep mode, which decreases the BS failure rate (and AF), and the frequency of the operating state changes, which deteriorates the BS failure rate (and increases AF). For each BS b, the AF can be computed over a period of duration ϑ, as follows:

$$AF^{(b,\vartheta)} = \frac{\gamma^{(b)}}{\gamma^{(b)}_{ON}} = 1 - \underbrace{(1 - AF_S)\tau_S}_{\text{Lifetime Increase}} + \underbrace{\chi f_{TR}}_{\text{Lifetime Decrease}} \tag{4.5}$$

where AF_S is the AF, computed assuming that the device is always kept in sleep mode. According to [111], it is always lower than 1, otherwise putting the device in sleep mode would mean increasing the failure rate of the device. Then, τ_S is the fraction of time the BS has spent in sleep mode in the period of duration t. The parameter f_{TR}, in cycle/h, is the frequency of the switching cycle which is measured over t and χ, in h/cycle, is defined as $\frac{1}{\gamma^{(b)}_{ON} N_F}$ and acts as weight of the frequency f_{TR}. As a result, the drop of the BS failure rate is achieved when $\chi f_{TR} < (1 - AF_S)\tau_S$. Notice that the parameters χ and AF_S depend on the HW component used to build the BS, while τ_S and f_{TR} depend on the switching strategy.

4.3 Traffic predictions

In this section, we present the traffic prediction tools used during the BS run-time as a preliminary step to the BS management decision. An ANN-based approach is used for this purpose. This is because, besides the potentiality of ANN has been widely demonstrated in [48] [50], chapter 3 shows that the performance of the BS activation/deactivation are not significantly affected by the ML approach used for

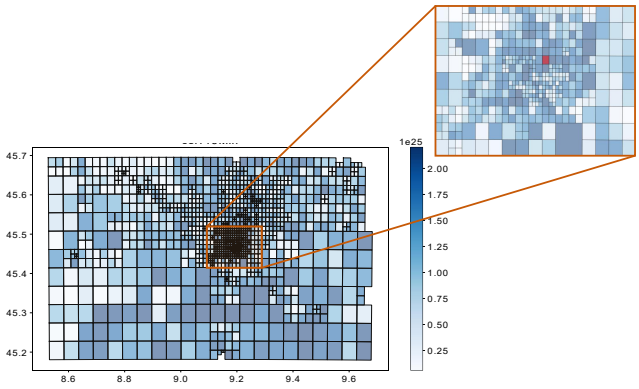


Fig. 4.1 Spatial cross-correlation among cells with lag= 15 minutes.

the traffic predictions. Thus, the usage of a simple ANN represents a good trade-off between performance and complexity.

4.3.1 Input Data

We use data provided by a large Italian mobile network operator, presented in section 3.2. They report the traffic demand volume, in bit, of 1420 BSs located in the city of Milan (Italy) and in a wide area around it, for two months in 2015, with granularity of 15 minutes. As in section 3.2, the traffic traces are normalised and eight portions of the city are selected as selected areas are representative of the various zones that coexist in an urban environment. As in chapter 3, in each of these portions of the RAN, we assume that one macro BS and 6 small cell BSs are present. For each area, we selected 7 traffic patterns recorded in that area and the trace which presents the highest traffic demand is chosen as the macro cell BS, while the remaining six as micro cell BSs.

4.3.2 Selection of the ANN input features

In order to predict traffic demand, the ANN must be fed with carefully selected input features. In this part of the work, because of the lower employed granularity than the on used in chapter 3, the investigation of the best choice for the ANN input features was made accounting for the temporal, as well as the spatial correlations of traffic. In particular, we exploit the traffic temporal periodicity (which we observed to be

present in most traffic patterns) due to the periodicity of human activities, and we investigate the possibility of also using the spatial correlation which is expected to be present among adjacent cells. In Fig. 4.1, the cross-correlation obtained between the traffic at one BS in the city centre, indicated in red, and all others is plotted, choosing as time lag one time slot, i.e., 15 minutes. We can see that correlation only mildly depends on the spatial closeness to the considered BS (darker colours correspond to higher correlation values). Indeed, high correlation values are present even among cells that are very far from each other. For this reason, we focus only on input features based on the temporal periodicity of traffic patterns.
Let us define by $T_{b,i}$ the traffic demand at BS b and time slot i. For simplicity of notation, in what follows we drop the index b if there is no ambiguity. At the beginning of each time slot *t*, the traffic demand at time slot *t* (the time slot that is just beginning), *t+1* (the following time slot), *t+2* and *t+3* are predicted; predictions are denoted by $\hat{T}^{(t)}$, $\hat{T}^{(t+1)}$, $\hat{T}^{(t+2)}$, $\hat{T}^{(t+3)}$, respectively.

Similar to chapter 3, the prediction tool receives as inputs:

- $T^{(t-1)}$: the traffic at the time slot just past, i.e., $t-1$;
- $T^{(t-(24\cdot4))}$: the traffic one day before the current time slot (the factor 4 comes from our time slots being 15 minutes long);
- $T^{(t-1-(24\cdot4))}$: the traffic one day before the time slot just past;
- $T^{(t-(48\cdot4))}$: the traffic two days before the current time slot;
- $T^{(t-1-(48\cdot4))}$: the traffic two days before the time slot just past.

4.3.3 Traffic Forecast Approach

Different ANN-based prediction approaches are tested.

1 ANN-4 outputs

One ANN for each BS is used. At time *t*, the ANN outputs the traffic demand samples at time slots *t*, *t+1*, *t+2* and *t+3* (see Fig. 4.2a).

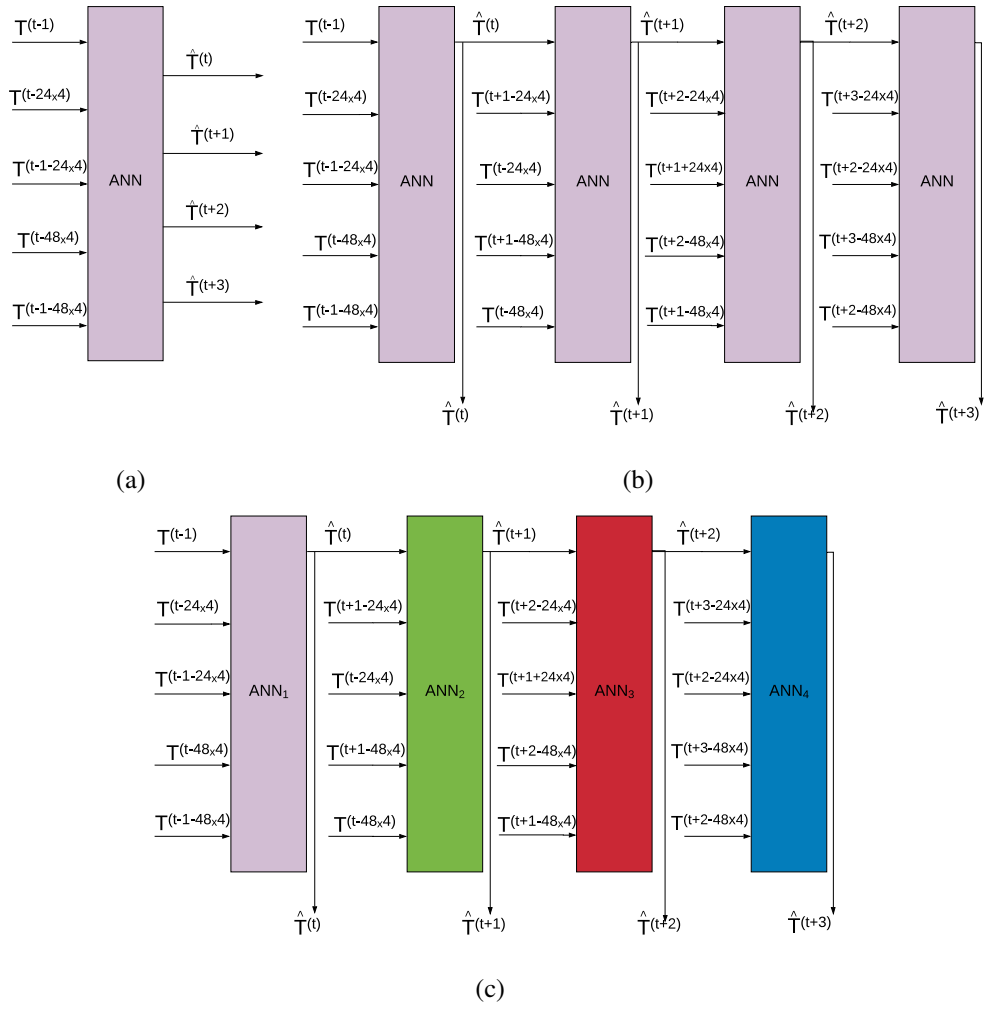


Fig. 4.2 Scheme of the three proposed prediction techniques: (a) 1 ANN-4 outputs, (b) 1 ANN-1 output, (c) 4 ANNs-1 output

1 ANN-1 output

One ANN for each BS is used. The ANN is trained to predict the traffic demand at the current time slot, e.g. at time t, and it is used in cascade to predict also the three future traffic samples, e.g. at time $t+1$, $t+2$, $t+3$. This means that the ANN produces the prediction of the traffic demand at time *t*, namely $\hat{T}^{(t)}$, using in input the traffic at previous time slot, $T^{(t-1)}$, as well as the traffic of previous days. Once $\hat{T}^{(t)}$ is computed, for predicting the traffic at time $t+1$, the same ANN is used but it receives as input the predicted traffic $\hat{T}^{(t)}$ instead of $T^{(t)}$ that is unknown. Similarly, for the prediction of traffic at times $t+2$ and $t+3$, predictions are used instead of

traffic samples for the unknown values of the input. The logical schema is reported in Fig. 4.2b.

4 ANNs-1 output

Four ANNs are used for each BS. Each ANN is dedicated to the prediction of the traffic demand at a given time lag. This means that the 4 future traffic samples are separately predicted, using 4 different ANNs, but the inputs are as in the previous case: predictions are used instead of missing samples whenever needed. The schema is reported in Fig. 4.2c.

As in chapter 3, each ANN mentioned above is structured in 3 layers: the input layer which has 8 nodes, one hidden layer with 17 nodes, and the output layer with one node, if *1 ANN-1 output* and *4 ANNs-1 output* are employed, or 4 nodes in case of *1 ANN-4 outputs* usage. The number of layers, as well the number of nodes for each layer are among the hyper-parameters which need to be selected. These have been chosen in order to achieve a good trade off between the accuracy and the time needed to train the network. Each ANN is trained minimising the Mean Squared Error (MSE) over the data of the first 47 days of the considered time period.

4.4 Processing traffic predictions

After the ANN has generated traffic predictions, they must be processed by the MANO to decide about micro cell BS (de)activation, with the objective to save energy, without compromising QoS. In this section, we propose strategies for processing the predictions and deciding micro cell activation and deactivation.

4.4.1 Resource Allocation

Different algorithms can be used to combine traffic predictions in a BS management strategy, based on the approach in chapter 3, which states that a micro cell BS is switched off if its traffic demand is lower than a threshold ρ^*, provided that such amount of traffic can be carried by the macro cell BS. As in chapter 3, the threshold depends on the energy consumption per carried bit: when the traffic is below ρ^*, the

energy needed to carry a unit of traffic in the micro cell is larger than in the macro, so that it is more convenient to switch off the small cell BS, if this is possible in terms of total capacity. As discussed in chapter 3, the optimal value of the threshold is 37% of the maximum load of the BS.

Max2Max

In this case, resources can be allocated only at the beginning of each hour: at 00:00, 01:00, etc. At the beginning of each hour, $\hat{T}^{(b,t)}$, $\hat{T}^{(b,t+1)}$, $\hat{T}^{(b,t+2)}$, $\hat{T}^{(b,t+3)}$, the 4 traffic demands corresponding to that hour are predicted for each micro cell BS *b*, as well as for the macro cell *B*; predictions in the macro are denoted by $\hat{T}^{(B,t)}$, $\hat{T}^{(B,t+1)}$, $\hat{T}^{(B,t+2)}$, $\hat{T}^{(B,t+3)}$. Among these 4 samples, the maximum, $\hat{M}^{(b)}$, for each micro cell BS *b* and the maximum, $\hat{M}^{(B)}$, for the macro cell are computed:

$$\hat{M}^{(b)} = \max\left(\hat{T}^{(b,t)}, \hat{T}^{(b,t+1)}, \hat{T}^{(b,t+2)}, \hat{T}^{(b,t+3)}\right) \tag{4.6}$$

$$\hat{M}^{(B)} = \max\left(\hat{T}^{(B,t)}, \hat{T}^{(B,t+1)}, \hat{T}^{(B,t+2)}, \hat{T}^{(B,t+3)}\right) \tag{4.7}$$

A micro cell BS *b* is switched off if $\hat{M}^{(b)}$ is lower than the threshold, and its traffic can be carried by the macro, given that the macro is expected to be carrying an amount of traffic $\hat{M}^{(B)}$:

$$\text{if } \left(\hat{M}^{(b)} < \rho^*\right) \wedge \left(\hat{M}^{(B)} + \hat{M}^{(b)} < C\right) \rightarrow \text{switch off b} \tag{4.8}$$

where C is the capacity of the macro cell B. Basically, the decision is taken based on the maximum of the predicted traffic samples. As the decision is taken, $\hat{M}_B$ is updated accordingly, to account for the traffic load that will be transferred from the considered BS.

Max2Max Continuous

This strategy is very similar to *Max2Max*, but, in this case, it is applied at the beginning of each 15 minute time slot and not only at the beginning of an hour, as in the previous case. The decision to switch off a cell for 4 consecutive time slots (1 hour) can be taken in any time slot.

I2I

When this strategy is used, the switch on/off is possible only at the beginning of each hour. Given the four predicted traffic demands belonging to the considered hour, for each micro cell BS b and for the macro cell BS B, a micro cell BS b is switched off when, for every slot $t+i$ with $i = 0, 1, 2, 3$, the estimated traffic $\hat{T}^{(b,t+i)}$ is lower than the threshold ρ^* and there is enough available capacity on the macro BS:

$$\begin{aligned} \text{if } \forall i = 0, \cdots, 3 \quad & (\hat{T}^{(b,t+i)} < \rho^*) \wedge (\hat{T}^{(b,t+i)} + \hat{T}^{(B,t+i)} < C) \\ & \rightarrow \text{switch off b} \end{aligned} \tag{4.9}$$

In this case, the decision to switch off is taken if the requested conditions are verified slot by slot.

I2I Continuous

When this strategy is used, *I2I* is applied at the beginning of each time slot. As in *Max2Max Continuous*, each micro cell BS remains active or in sleep mode for at least 1 hour (4 consecutive time slots), but a change of state can happen in any time slot.

I2I Flexible

This is a further variation of *I2I Continuous*. As before, at the beginning of each time slot, *I2I* is applied. Nevertheless, when a micro cell BS has been put in sleep mode for at least one hour, it remains sleeping if the necessary conditions are verified for one more time slot. This means that when we are at time t, given that the micro cell BS has been deactivated since at least *t-4*, that micro cell BS remains in sleep mode, if $\hat{T}^{(b,t)}$ is lower than ρ^*, provided that $\hat{T}^{(b,t)}$ can be carried by the macro BS during the t-th time interval.

As in chapter 3, the conditions to switch to sleep mode a micro cell BS are verified, starting from the least loaded to the most loaded, in the following hour. Indeed, in case of traffic load lower than ρ^*, the energy consumption per bit is larger. As explained in chapter 3, in order to minimise the network energy consumption, larger

priority is given to micro BSs which carry the minimum traffic load. The load during the following hour on each micro BS is given by summing the traffic demand during the 4 time slots belonging to that hour.

4.4.2 Descending front detection

The presence of noise in traffic patterns may result in incorrect deactivation of the small cell BSs, thus deteriorating QoS. For this reason, the concept of Descending Front Detection (DFD) is introduced in the processing of traffic predictions. In particular, the switching from active to sleep mode of a micro cell BS is permitted only if a descending front is detected: if an active micro cell BS is detected to be in a descending phase, the necessary conditions for the micro cell switch off are checked. Because of the noise inherent in traffic patterns, a negative first derivative is not a sufficiently good indicator of a descending front. Therefore, a moving average filter is used for this purpose. It smooths data by replacing each traffic sample with the average of the neighbouring samples. This operation practically acts as a low-pass filter on traffic patterns. In our case, a triangular smoothing is applied twice. In particular, at time *t*, the following expression is computed, for z = t-4, t-5, t-6:

$$S'^{(b,z)} = \frac{1}{81} \sum_{j=-2}^{2} (3-|j|) \sum_{i=-2}^{2} (3-|i|) T^{(b,z+j+i)} \tag{4.10}$$

where $T^{(b,z+j+i)}$ is the real traffic demand on BS *b* at time $z+j+i$. However, notice that for $z = t-4$ and for $j = 2$ and $i = 2$, $T^{(b,z+j+i)}$, is $T^{(b,t)}$, which is not known. Thus, its prediction, $\hat{T}^{(b,t)}$ is used in this case. The maximum z is chosen equal to *t-4*, in order to avoid using other predicted samples.
If $S'^{(b,t-4)} < S'^{(b,t-5)} < S'^{(b,t-6)}$, we conclude that a descending front is detected. If this is the case, the necessary micro cell BS switch off conditions are checked. If they are verified, as described in section 4.4.1, the considered micro cell BS can be deactivated.

4.5 Key Performance Indicators

In this part of the study, the ARE, i.e. the average relative error, the Energy Consumption Reduction, E_{RED}, i.e. the percentage of energy saving with respect to the Always on, and the percentage of Lost Traffic, L_T, detailed in section 4.5 are employed. The following additional KPIs are also evaluated.

Accelerator Factor for each micro cell BS

For each BS *b*, we measure $AF^{(t,b)}$ and $AF^{(b)}$ which are, respectively, the AF of that BS *b*, measured at time t and at the end of the considered operating period, i.e. in steady state, given by (4.5).

Accelerator Factor

For each considered portion of network, we measure AF which is the average AF of that area at the end of the considered operating period, i.e. in steady state:

$$AF = \frac{1}{N_{BS}} \sum_{b=1}^{N_{BS}} AF^{(b)} \tag{4.11}$$

where N_{BS} is the number of micro cell BS in the considered portion of RAN and $AF^{(b)}$ is the AF, for each BS *b* in the considered area, computed as in (4.5).

4.6 Performance evaluation

In this section, we discuss numerical results obtained by experimenting with the different prediction, processing and decision algorithms presented in the previous sections on the considered RAN portions. Out of the 61 days for which we have real traffic data, the first 47 are used for the ANN training phase, while the remaining 14 days are used for the run-time phase.

Table 4.1 Average relative error, ARE, with the different approaches at different time lags.

	1 ANN-4 outputs	**1 ANN-1 output**	**4 ANNs-1 output**	**4 ANNs-1 output (spatial)**
ARE_t	0.33	0.33	0.33	0.37
ARE_{t+1}	0.43	0.44	0.43	0.47
ARE_{t+2}	0.52	0.52	0.48	0.52
ARE_{t+3}	0.61	0.57	0.52	0.54

4.6.1 Choice of the ANN

As a first step, we analyse the effectiveness of the different ANN configurations for traffic predictions, using the previously defined ARE (Average Relative Error) as a performance metric. The results provided by the considered ANN configurations, namely *1 ANN-4 outputs*, *1 ANN-1 output* and *4 ANNs-1 output*, for each time lag, are reported in Table 4.1, averaged over the eight considered geographical areas. Observe that numerical results confirm what is intuitively expected, and was quantitatively shown in [51]: the error increases with the time horizon of the predictions. Moreover, typically, the *1 ANN-4 outputs* provides the largest ARE. This is because, when the other 2 approaches are used, the sample corresponding to the most recent traffic demand, even if only predicted, is provided as an input feature. In Fig. 4.3, the percentage of the reduction of ARE gained with *1 ANN-1 output* and *4 ANNs-1 output*, with respect to *1 ANN-4 outputs*, are shown in blue and orange, respectively. The reduction of the estimation error is the largest for *4 ANNs-1 output*, especially when the time horizon of the predictions is longer. This is because this ANN configuration uses 4 ANNs: during the training phase, each ANN learns how to forecast the desired output, managing the error which affects the input traffic sample derived from a prediction. For these reasons, in the rest of this study we will use *4 ANNs-1 output* for traffic forecast, unless otherwise specified.
In order to confirm the mild correlation among adjacent cells, we also report the ARE which is obtained when we provide to *1 ANN-4 outputs* an additional input feature. In order to select this additional input feature, the cross-correlation between the traffic demand of the current BS and each of its adjacent ones, is performed. Then, the argmax function is computed, in order to select the BS bs_{MAX} and the time lag l_{MAX} which provide the largest value of cross-correlation. Thus, when we

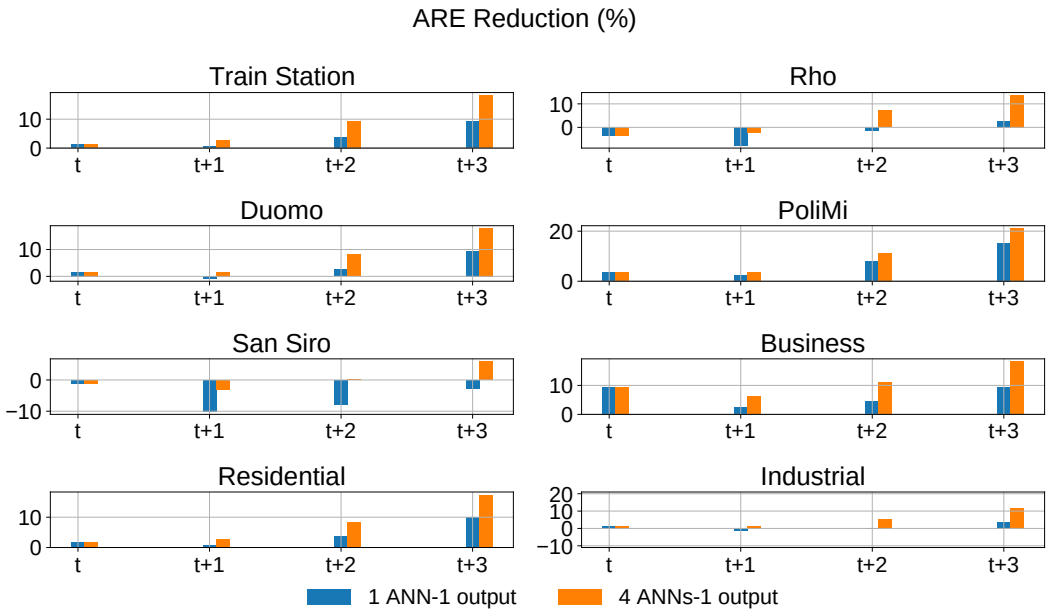
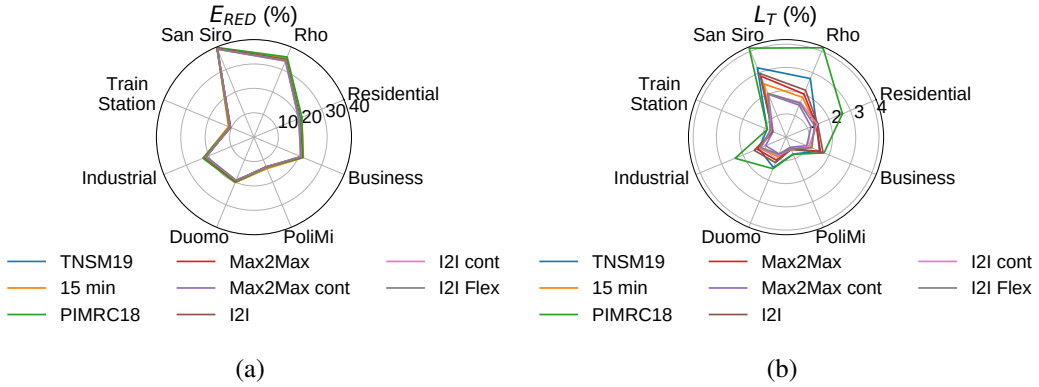


Fig. 4.3 Percentage of the reduction of absolute relative error, ARE, obtained by *1 ANN-1 output* or *4 ANNs-1 output* with respect to the *1 ANN-4 outputs*, for different time lags.

are predicting the traffic demand at time t, the additional input feature is the traffic demand on bs_{MAX} at time t-l_{MAX}. Similarly, for the prediction of the traffic demand at t+*1*, t+*2* and t+*3*, the traffic demand on bs_{MAX} at time t+*1*-l_{MAX}, t+*2*-l_{MAX}, t+*3*-l_{MAX}, respectively, are given as additional input feature. From Table 4.1, it is possible to notice that the presence of this feature deteriorates the precision of the forecast.

4.6.2 Dynamic resource allocation performance

We now investigate the performance of the resource allocation strategies presented in Section 4.4.1. Our solutions are compared against 3 benchmarks: (*i*) the *TNSM19* approach presented in chapter 3, which allocates the resources of a RAN according to the hourly traffic predictions obtained using an ANN, with no processing of the ANN outputs; (*ii*) the *PIMRC18* approach: in this case the traffic is predicted using the LSTM network proposed in [51] and resources are allocated based on *I2I*; (*iii*) the *15 min* approach, similar to the *TNSM19* case, but operated over 15 minutes time slots. With this approach, each small cell BS can be switched to/from sleep mode as soon as needed, with no constraint on the frequency of switching.



(a) (b)

Fig. 4.4 Comparison of dynamic resources allocation strategy in the various areas: (a) E_{RED} and (b) L_T.

Effect of traffic pattern shape and load distribution

For each strategy and zone, Fig. 4.4a reports the energy consumption reduction computed with respect to the *always ON* scenario, and Fig. 4.4b reports the percentage of lost traffic. First, it is possible to confirm that, as discussed in chapter 3, the percentage energy saving directly depends on the shape of the traffic pattern (peak/off-peak ratio, duration of peaks, ...), which is characteristic of the considered area. If the traffic demand is low for many hours, the BS management approach can be very effective: up to 40% of the energy consumed with respect to the always ON approach can be saved, as we see in the San Siro and Rho areas. When the traffic demand is larger than ρ^* for longer periods, the small cell BSs can be switched off for shorter periods, and a lower amount of energy is saved. This is the case of the PoliMi and Train Station areas, where the energy saving is lower than 15%. In Fig. 4.5a, the traffic demand during 5 days of the simulation of a micro cell BS in the Train Station and San Siro areas is plotted, in dark and light grey, respectively. The former presents a traffic volume usually larger than the threshold, indicated by the black horizontal line, while the latter almost always lower than ρ^*. As a consequence, their sleeping time ratio τ_S is very different, as can be noticed in Fig. 4.5b, where τ_S is plotted, for each dynamic resource allocation approach. In particular, τ_S is never larger than 0.25 and lower than 0.9 for the micro cell BS located in the Train Station and San Siro areas, respectively. Nevertheless, the frequency switching f_{TR}, reported in Fig. 4.5c, assumes very close values for both cases, since the micro cell BS located in the Train Station zone is usually ON and is deactivated only during the night, while the one placed in the San Siro area is usually in sleep mode and

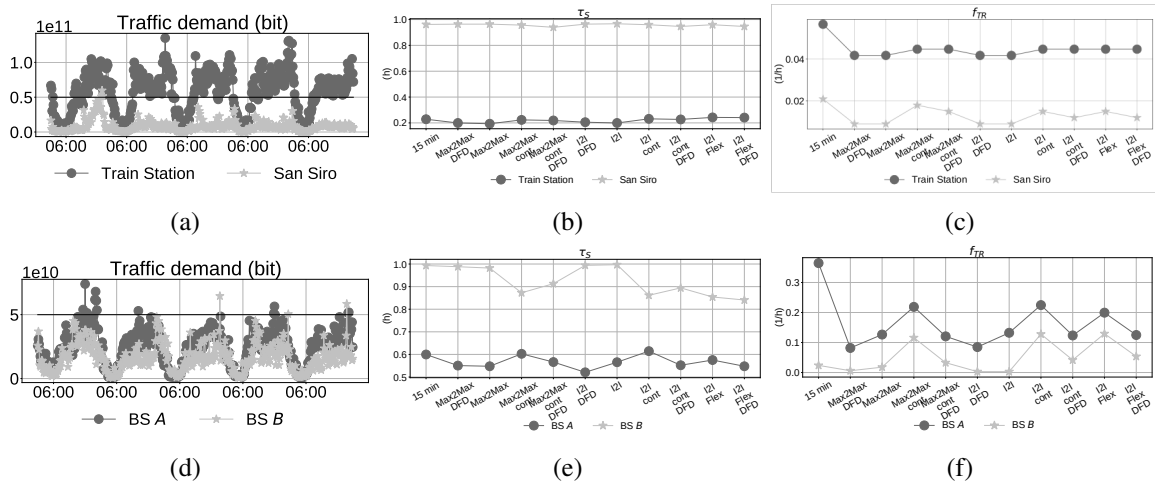
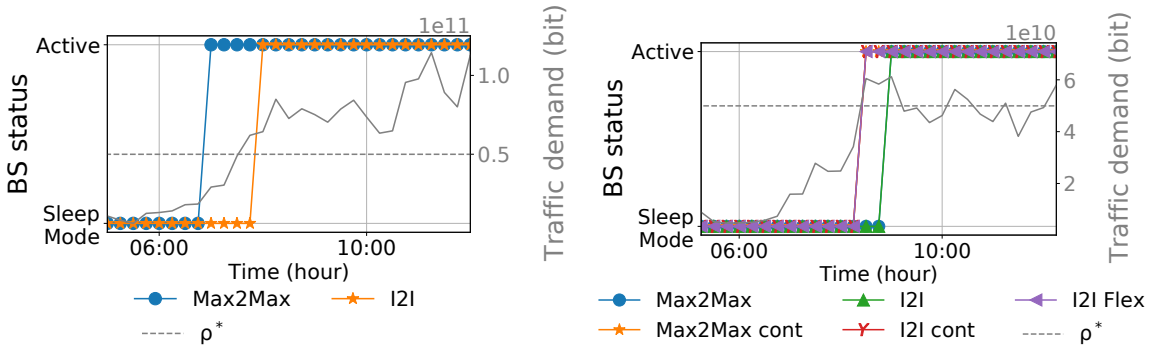


Fig. 4.5 Comparison of dynamic resources allocation strategies with different traffic demand pattern: (a) traffic demand, (b) τ_S and (c) f_{TR} of a BS in San Siro and Train Station areas; (a) traffic demand, (b) τ_S and (c) f_{TR} of two different micro cell BSs in the Residential areas.

is active only during public events, when additional capacity is needed in order to satisfy the traffic demand. Fig. 4.5d reports the traffic demand during 5 days of simulation of two micro cell BSs, BS *A* and BS *B*, of the residential district. These two patterns, reported in dark and light grey in the figure, are very similar in shape and volume and are lower than the threshold for most of the time. However, their corresponding τ_S assume very different values. This is because micro cell BS *A* presents a lower traffic demand than micro cell BS *B*. Therefore, it has priority to be put in sleep mode, and is in sleep mode for more than 80% of the time, for each dynamic resource allocation. Because of this, the macro cell BS carries also the amount of traffic demand of micro cell BS *A*, making not possible the deactivation of micro cell BS *B*, to avoid macro cell BS overloading, even if its traffic demand is lower than the threshold. This suggests that the switching of a single micro cell BS and, consequently, its energy consumption, depends on its traffic demand, as well as on the traffic demand of the other micro cell BSs, which belong to the same hierarchical RAN cluster.

Comparing resource allocation strategies

Focusing again on Fig. 4.4a, it is possible to notice that the reduction of the energy consumption obtained with our proposals is slightly lower than with the chosen benchmarks. Indeed, with our proposals the energy consumption increases, at most,



(a) (b)

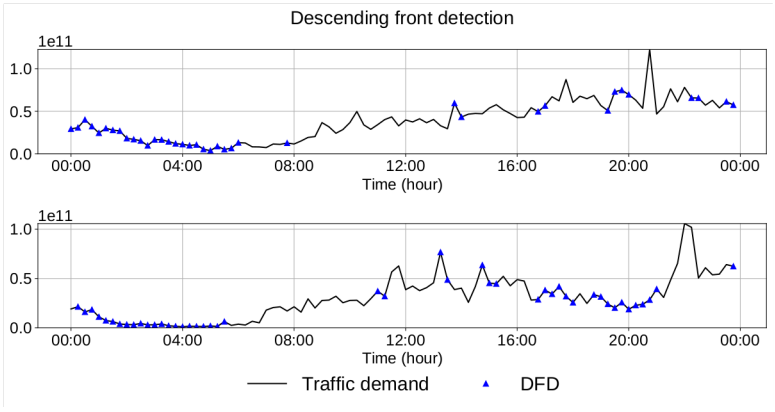
Fig. 4.6 Comparison of dynamic resources allocation strategies in the various areas: (a) *Max2Max* and *I2I* and (b) *Cont* version.

by 1.7%, 1.9% and 2.9%, with respect to the *TNSM19*, *15 min* and *PIMRC18* benchmarks. This is because our approaches are slightly more conservative in energy saving, but better preserve QoS, measured as the percentage of lost traffic. When *TNSM19* is used, resources are allocated under the unrealistic assumption that the traffic demand is uniformly distributed within a whole hour. For this reason, the lost traffic is higher than with the other approaches. With *PIMRC18*, up to 4% of traffic is lost. Even if it provides traffic predictions affected by lower *ARE* (0.29, 0.37, 0.42, 0.47, for the forecast at time t, $t+1$, $t+2$ and $t+3$, respectively), it usually generates more underestimated traffic samples that contribute to QoS deterioration.

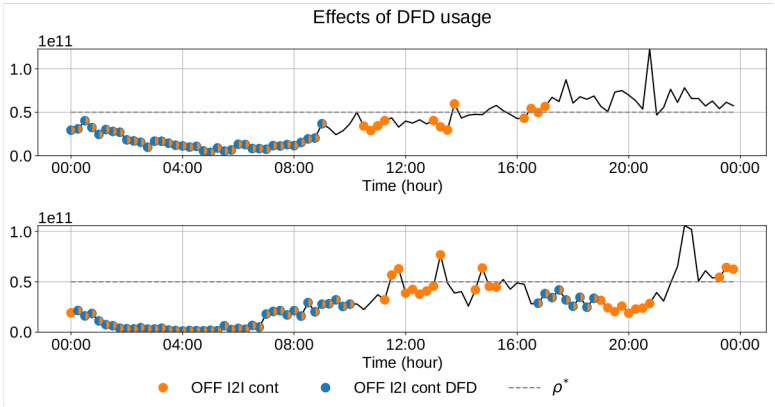
The comparison with the *15 min* case is also interesting. The *15 min* case is based only on traffic predictions performed over a time horizon of 15 minutes for which the error is lowest (see table 4.1). Nonetheless, in this case no processing of the ANN outputs is performed; hence, despite the small error in predictions, the lost traffic is quite large. This is a clear indication of the importance of the processing of ANN outputs.

Let us now focus on the proposed approaches. The lost traffic is lower in the *Max2Max* case than in the *I2I* one, since its switching condition is stricter. Fig. 4.6a shows the status of a micro cell BS of the Train Station area in orange and blue, when *I2I* and *Max2Max* are used, respectively. The micro cell BS traffic demand is reported in grey and ρ^* with the dashed grey curve. Even if these two approaches use the same prediction samples for resource allocation, *Max2Max* makes the BS active sooner than *I2I*. At 7.00 a.m., predictions of the traffic demand during the following hour are erroneously smaller than the threshold. Nevertheless, *Max2Max*

switches BSs if the maximum, among the traffic demand samples belonging to that hour, is smaller than ρ^* and can be carried by the macro cell, supposing that it is managing an amount of traffic which is the maximum traffic demand among the 4 traffic demand samples of that hour. Thus, the micro BS is activated, since its traffic demand cannot be carried by the macro BS because of the capacity constraint.
The use of the *cont* variation provides benefits in terms of QoS in both strategies, *Max2Max* and *I2I*. When *cont* is used, the effect of higher errors, which characterises traffic predictions over longer time lags, is further mitigated, so that a more accurate resource allocation can be performed. As a result, the lost traffic is always less than 2%. Specifically, in the areas where the resource allocation is more difficult due to the unpredictability of traffic demand, i.e., Rho and San Siro, 1.6% and 2% of the traffic is lost. In areas where patterns are more regular, values are always lower than 1.4%. Similar results are given by *I2I Flex*: the lost traffic is lower than the chosen benchmarks because the BS switching can react to the traffic demand every 15 minutes, provided that the last switching has occurred for at least 1 hour. This can be observed in Fig. 4.6b, where each curve corresponds to the status of a micro cell BS of the Duomo area, obtained with each of the considered allocation approaches. Also its traffic demand (in grey) and ρ^* (grey dashed curve) are reported. The *cont* variation and *I2I Flex* react as soon as the traffic demand increases (at 8.30). When *I2I* and *Max2Max* are used, resources are allocated at 8.00 a.m., and the prediction with lag equal to 3 is used for that time slot. Because of the large error which affects this forecast, this sample results lower than the threshold, and the micro cell BS is not activated. Thus, from Figs. 4.6a and 4.6b, it is possible to notice that strategies behave similarly if the traffic demand is far from the threshold. Indeed, in this case, the large error, which affects typically deeper forecasts used by *I2I* and *Max2Max*, does not impact resource allocation, correctly detecting the value of the traffic demand with respect to the threshold. As soon as the traffic demand moves closer to the threshold, even if based on the same predictions, the resources allocation is different. In case of *max*, conditions for the deactivation are stricter; with *cont* based approaches, more accurate predictions can be used. This results in more likely activation of micro BSs and, consequently, in lower lost traffic.



(a)



(b)

Fig. 4.7 Impact of descending front detection for two areas: (a) detection of fronts and (b) switch off decisions with and without DFD.

4.6.3 Impact of descending front detection

We now investigate the impact of the use of DFD in resource allocation. When DFD is used, the deactivation of a micro cell BS is possible only if a descending front is detected, according to the conditions described in Sec. 4.4.1. In Fig. 4.7a, blue triangles mark the detection of a descending front during one day of the run-time phase of 2 micro cell BSs, belonging to the PoliMi and Rho Fiere areas. As can be observed, descending fronts are mostly correctly identified. Since the current predicted traffic demand has lower impact on DFD than past samples, see equation (4.10), it is possible that DFD is activated after a local minimum.

Fig. 4.8 reports the energy consumption reduction, in bars, and the lost traffic, indicated by the blue and red lines with circle markers. Blue markers refer to no DFD, while red markers refer to DFD. The results of the chosen benchmarks are

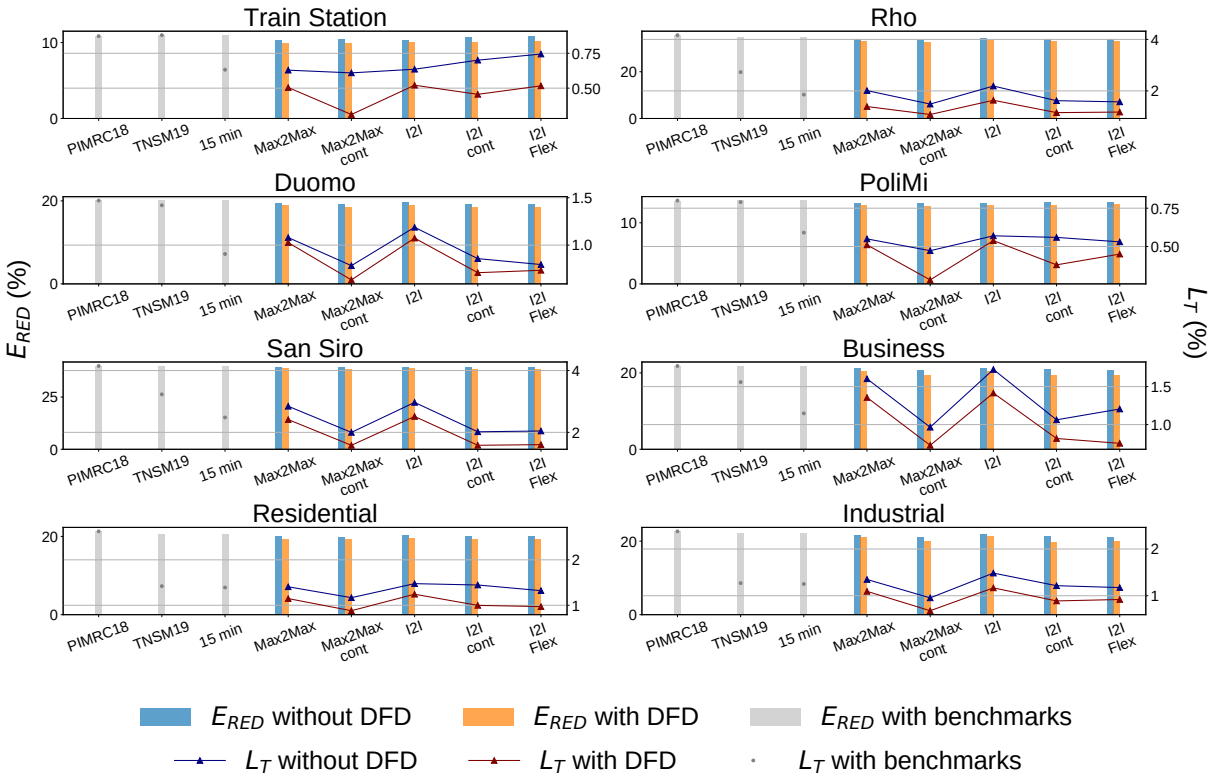


Fig. 4.8 Energy consumption reduction and lost traffic in each area, with each dynamic resource allocation with and without descending front detection, DFD.

reported in grey. The figure reveals that the usage of DFD generates a systematic drop in both energy efficiency, for a small amount, and lost traffic, for more significant values. The energy consumption reduction remains between 10% and 39%, similar to the case of no DFD, when *TNSM19*, *15 min* and *PIMRC18* are used, but QoS improves significantly: lost traffic is usually below 1%. In the San Siro and Rho Fiere areas, because of the critical characteristics of traffic patterns, this value is between 1% and 1.5%. The reductions of lost traffic are due to the stricter conditions to switch off the micro cell BSs. This can be seen in Fig. 4.7b, which illustrates an example of the traffic demand, in black, of the 2 small cells BSs of Fig. 4.7a. In Fig. 4.7b, the orange and blue points indicate the time slot during which the considered micro cell BS is in sleep mode when *I2I cont* and *I2I cont with DFD* are used, respectively. During periods of almost constant but noisy traffic demand, if traffic values are close to the threshold ϱ^*, incorrect small cell BSs deactivations may occur. Indeed, for those traffic values, even a small error in the traffic predictions can determine a wrong allocation of resources. This is the case reported in the

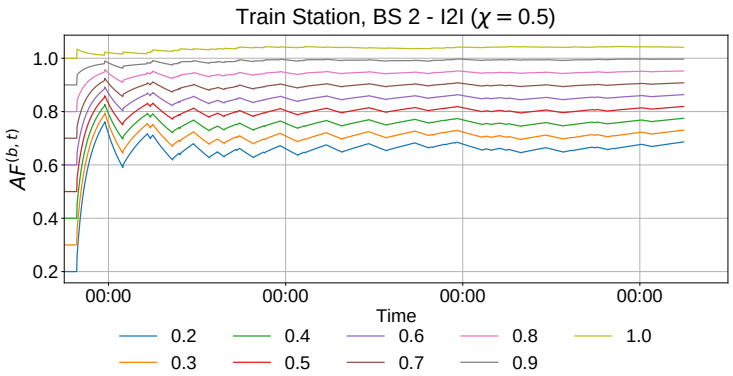
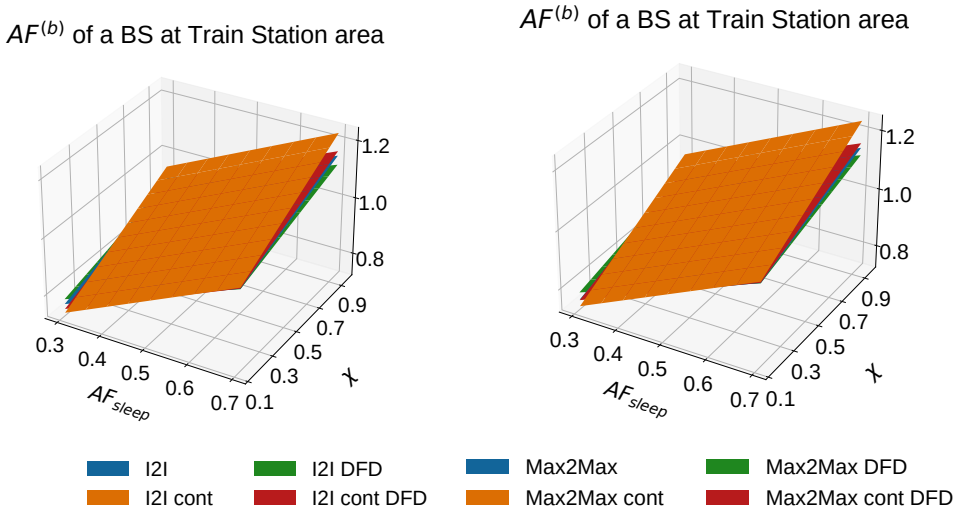


Fig. 4.9 $AF^{(t,b)}$ for a micro cell BS of the Train Station area, with χ=0.5, varying AF_{sleep}

figure: with DFD, incorrect deactivation of the considered small cell BSs is avoided since a descending front is not detected. Without DFD, with the *I2I cont* alone, the estimation error (even if small) makes the predictions lower than ρ^*, and a wrong switch off decision is taken. With DFD, the small cell BS is not switched off because the descending front is not detected. This behaviour explains the slight increase of energy consumption when DFD is applied. However, in spite of a very limited raise in energy consumption, the traffic loss can be reduced by up to 74% with respect to the benchmarks.

4.6.4 Impact on the BS failure rate

The impact of the proposed dynamic resource allocation schemes on the BS failure rate is now discussed. Each curve in Fig. 4.9 represents the behaviour of $AF^{(t,b)}$ versus time for a micro cell BS in the Train Station area, obtained with a different value of AF_{sleep}, when *I2I* is employed, with χ equal to 0.5, that is the value measured in an LTE BS in [57]. At the beginning of each simulation, $AF^{(t,b)}$ is lower than 1, since the simulation starts at midnight and the micro BS can be put in sleep mode, making $AF^{(t,b)}$ small. Then, at 7 a.m. it starts growing, since the BS is activated due to increasing daily traffic demand. After some fluctuations, due to BS activation and deactivation that follow the daily traffic demand variations, $AF^{(t,b)}$ stabilises, since τ_S and f_{TR} stabilise as well. When large values of the parameter AF_{sleep} are considered, $AF^{(t,b)}$ is large due to more significant BS deterioration in sleep mode.
Fig. 4.10 reports the value of $AF^{(b)}$, on the z-axis for a BS in the Train Station area; different values of AF_{sleep} in the interval $[0.1, 0.9]$ are considered on the x-axis, and



(a) (b)

Fig. 4.10 $AF^{(b)}$ for a micro cell BS of the Train Station area: (a) with I2I-based approaches and (b) with Max2Max-based approaches.

different values of χ in $[0.1, 2.0]$, on the y-axis. Each plotted plane corresponds to a different dynamic resource allocation approach. In particular, the *I2I* and *Max2Max* strategies are considered, with and without the *cont* variant and the use of *DFD*. From these figures, we first notice that the growth of AF_{sleep} and χ implies a growth of $AF^{(b)}$. If AF_{sleep} is large, the time in sleep mode is less beneficial to the BS failure rate; while large values of χ corresponds to the growth of the cost of the BS switching, see (4.5). When AF_{sleep} and χ are large enough, more conservative dynamic resource allocations provide lower values of $AF^{(b)}$, than more dynamic ones. Indeed, the *DFD* variant, which uses the strictest switching conditions, provides the lowest $AF^{(b)}$, because of the reduction of the switching frequency f_{TR}, without significant reduction of the sleeping time ratio τ_S, see Figs. 4.5b, 4.5c, 4.5e and 4.5f. The largest values of $AF^{(b)}$ are obtained when the *cont* variation is used, since, as mentioned in the previous section, it promptly reacts to the low traffic demand. This increases f_{TR} and, as a consequence, $AF^{(b)}$. For small values of AF_{sleep} and χ (bottom left part of the plots in Fig. 4.10), the situation is different. Indeed, with small values of χ and AF_{sleep} the cost of a BS switching does not significantly impact the BS failure rate, and spending time in sleep mode largely decreases it. Therefore, in this interval of values, the approaches which put the micro cell BS in sleep mode for a longer time, provide lower values of $AF^{(b)}$, as in the case of *cont* variants.
Fig. 4.11 combines energy consumption and AF by representing each dynamic

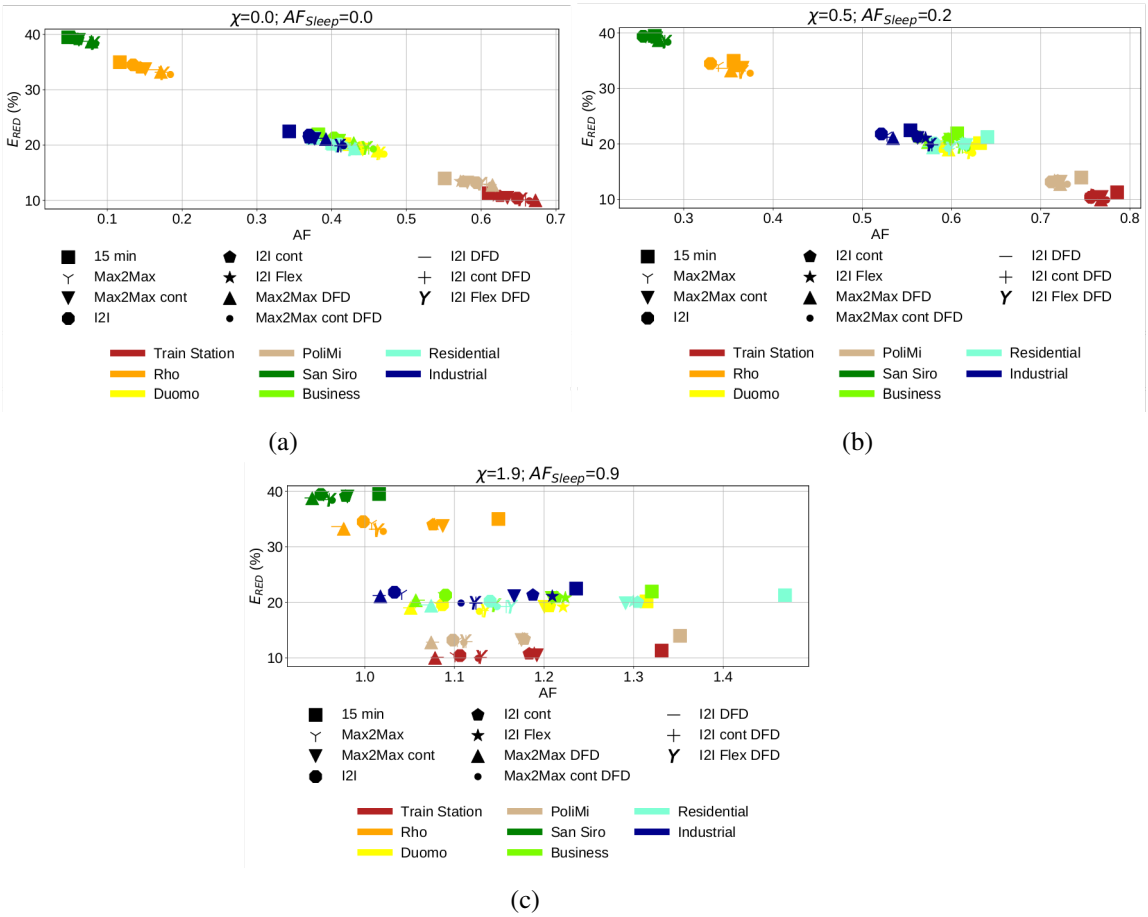


Fig. 4.11 Energy Consumption Reduction and AF, obtained using different dynamic resource allocation, in each area, with χ and AF_{sleep} equal to (a) 0, (b) 0.2, 0.5 and (c) 1.9, 0.9.

resource allocation algorithm in each area with a marker positioned so that the y coordinate corresponds to the energy consumption reduction and the x coordinate corresponds to the value of AF. Fig. 4.11a reports the AF values, with AF_{sleep} and χ equal to 0, which corresponds to the ideal case in which in sleep mode the BS failure rate goes to 0, meaning that its lifetime goes to infinity, and the BS switching does not affect it. Results in Fig. 4.11b are provided for $AF_{sleep} = 0.2$ and $\chi = 0.5$, as measured in [57] for an LTE BS. Finally, in Fig. 4.11c, the parameters are set pessimistically to 0.9 and 1.9, meaning that the sleep mode only slightly reduces the BS failure rate and the BS switching is highly costly, significantly affecting the BS deterioration. As expected and discussed in section 4.6, results are clustered according to the geographical area, because of the different achieved energy consumption reduction, which strictly depends on the traffic pattern that is characteristic of each zone. When

AF_{sleep} and χ are 0, AF is always lower than 0.7, meaning that the average failure rate of the BS is decreased by 30%. In addition, we notice that AF is directly proportional to the energy consumption reduction, since AF is only affected by the time spent in sleep and active mode, not by the BS switching. As a result, the *15 min* approach always provides the lowest AF, since it rapidly reacts to the low traffic demand, immediately turning the micro cell BSs into sleep mode. When AF_{sleep} and χ increase, the dynamism of this resource allocation approach negatively impacts the AF, which results the largest among the ones provided by our strategies, see Figs. 4.11b and 4.11c. Indeed, when the parameter χ grows, each BS switching is very costly. Thus, when AF_{sleep} and χ are 0.2 and 0.5, *I2I* and *Max2Max* approaches provide the lowest value of AF, since the most suitable balance between τ_S and f_{TR} is achieved. This does not occur with the *cont* variant: the large values of τ_S are not sufficient to compensate for the large values of f_{TR}. Similarly, with the adoption of *DFD*, the small values of τ_S, because of the strict conditions for the BS deactivation, generate larger values of AF than with *I2I* and *Max2Max*. If AF_{sleep} and χ are 0.9 and 1.9, the situation further worsens and AF is usually larger than 1, meaning that dynamic resource allocation increases the BS failure rate, because of the high cost of switching and the low benefit of being in sleep mode. Only for the San Siro area, values of AF lower than 1 are obtained because of the very long time the micro cell BSs spend in sleep mode. In this scenario, *I2I DFD* and *Max2Max DFD* are needed to reach the minimum AF values, since their strict deactivation requirements prevent frequent highly costly switching. Furthermore, under high values of AF_{sleep} and χ, the variable traffic patterns observed in the different zones can make even more critical the selection of the proper resource allocation scheme, whose impact on AF may result more significant. Indeed, whereas in the Train Station area the worst performing prediction algorithm increases AF by about 10% with respect to the lowest values obtained under *I2I DFD* and *Max2Max DFD*, in other traffic zones, like the Residential area, the worst performing scheme provide AF values that result up to almost 40% higher than the AF value under the best approach, that is anyway larger than 1.

4.6.5 Impact of the traffic prediction technique

Finally, let us consider the impact of the traffic prediction technique. Fig. 4.12 reports with the blue and the orange bars, the energy consumption reduction achieved with

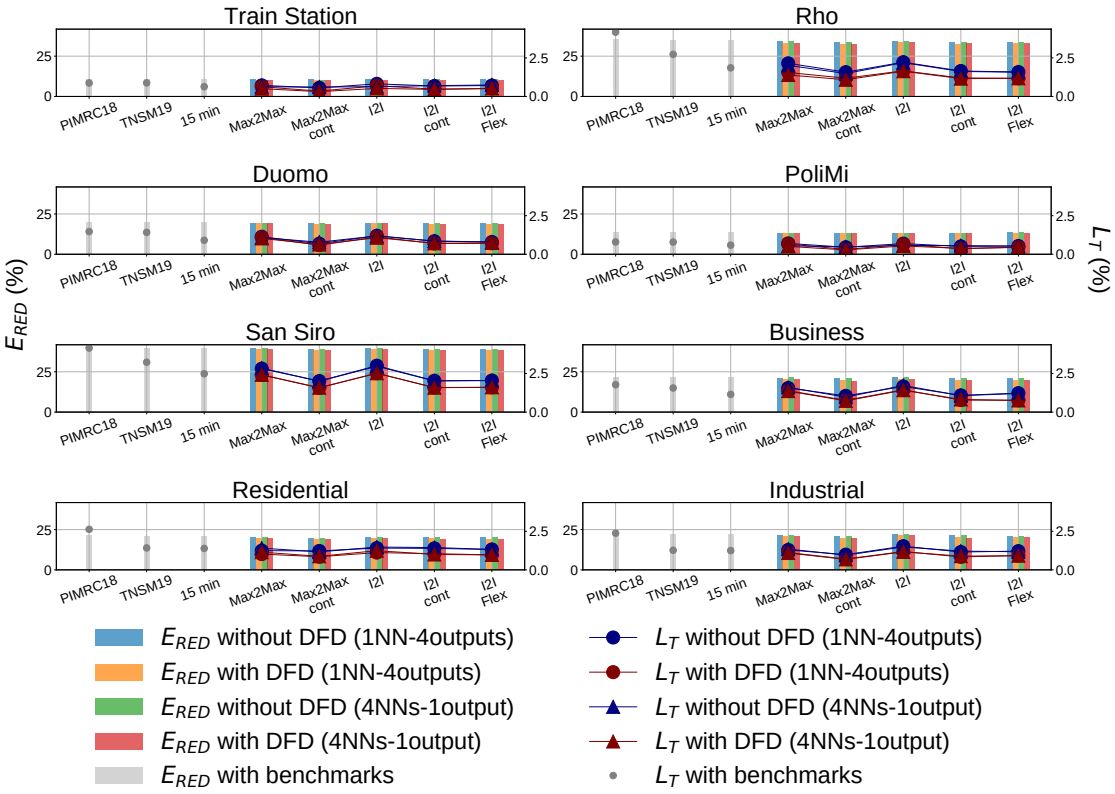


Fig. 4.12 Energy consumption reduction and lost traffic in each area, with each dynamic resource allocation with and without descending front detection, DFD, in each area, using *1 ANN-4 outputs* and *4 ANNs-1 output*.

and without DFD, if the traffic demand is forecast with *4 ANNs-1 output*, in each area. The resulting lost traffic is shown with the red and blue lines with triangle markers, if the DFD is used or not, respectively. Similarly, the green and red bars in Fig. 4.12 indicate the energy consumption reduction obtained with and without DFD, when the traffic demand is forecast with *1 ANN-4 outputs*, which is the ANN that we identified as the one performing worst in predicting traffic. The obtained lost traffic is reported, respectively, with the red and blue lines with circle markers. In spite of the larger estimation error with respect to *4 ANNs-1 output* (see Table 4.1), performance is very similar: the values of lost traffic and energy consumption are almost equal to the previous case. Indeed, lost traffic drops up to 1%, while energy consumption is reduced between 9% and 40%. Similar results are achieved in the other areas. This means that the choice of an effective processing algorithm can have more impact on performance than the choice of the ANN. Only with careful

processing, the ANN prediction errors are mitigated, and a good trade-off between energy consumption reduction and QoS is achieved.

4.7 Final Remarks

This part of the work confirms that the allocation of heterogeneous hierarchical RAN resources according to traffic demand is promising, but the provided energy saving of each BS is strictly related to its traffic demand pattern, as well as to the traffic patterns over the whole considered area. Dynamic resource allocation requires the knowledge of the actual traffic demand and, hence, ML approaches are needed to accurately predict it so as to enable network management mechanisms that adapt to traffic variability. This is interesting in perspective, for the promising possibilities offered toward the deployment of new networks that are easily and automatically reconfigurable. However, ML approaches become particularly effective only if their outputs are integrated into decision processes that are driven by a deep domain knowledge. If the traffic predictions are carefully processed, QoS deterioration is avoided, while significant energy saving can be achieved. Prediction processing requires both the understanding of traffic patterns over long time scales, so as to detect the overall trend of increasing or decreasing traffic, as well as strategies to combine predictions at different time lags.
Finally, prediction processing and the consequent dynamic resource allocation affect the BS failure rate in different ways. Switching a BS is harmful to its failure rate while the time spent in sleep mode prevents its deterioration. The actual impact of the combination of these two phenomena depends on the HW components of the BS, as well as on the RAN management strategy. In case the switching of a BS is not costly, less strict switching conditions can be applied: the BS failure rate is not affected while larger energy saving is achieved. Conversely, when the BS is sensitive to switching, more conservative resource allocations should be employed. For existing networks, not designed for highly dynamic resource allocation, conservative approaches better prevent BSs from HW failures; however, in perspective, with the deployment of new devices suited for strongly dynamic networks, less conservative approaches, which frequently activate and deactivate BSs, can be used, and higher energy saving is expected.

Chapter 5

Renewable Energy Sources for RAN Energy Efficiency

Part of the work presented in this chapter has already been published in:

- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2019). Greener RAN operation through machine learning. *IEEE Transactions on Network and Service Management, 16*(3), 896-908.
- Vallero, G., Deruyck, M., Joseph, W., & Meo, M. (2020, June). Caching at the edge in high energy-efficient wireless access networks. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.
- Vallero, G., Deruyck, M., Meo, M., & Joseph, W. (2021). Base Station switching and edge caching optimisation in high energy-efficiency wireless access network. *Computer Networks*, 192, 108100.

As discussed in chapters 1 and 2, a recent trend in networking is to consider the availability of energy produced by Renewable Energy Source (RES), which are integrated in the mobile networks and installed in proximity of the BSs networks. In this way, the RAN is supplied by the "green" energy produced by RES and by the energy taken from the grid, which is considered "brown", since it is largely produced by burning fossil fuels. As a results, the RAN sustainability is improved and the electricity bill is reduced, by reducing the amount of energy that has to be purchased

from the power grid, which is the key contributor for the increase of the OPEX.
While the solution is promising, various issues need to be addressed, among which the possible lack of energy generation due to its randomness, dependence on the weather conditions and daily and seasonal variability. To address these issues, the green energy generation has to be combined with resource management strategies, so as to optimise its usage with respect to the brown energy. If the available green energy is properly managed, the network results more independent from the power grid, making this solution very suitable for remote areas and regions where the electricity system is unreliable, because of frequent black-outs.
In this chapter, we consider a portion of a RAN, powered by a PV panel system, energy batteries and the power grid. Details of the considered scenario are given in section 5.1. In case the green available energy, given by the amount of energy produced by the PV panel system and stored in the batteries, is not sufficient for the RAN supply, an energy reducing strategy is applied. As detailed in section 5.2, the proposed strategies switch micro cell BSs in order to reduce the network energy consumption and, consequently, the amount of energy which has to be bought from the grid; the performances are discussed in section 5.3. As discussed in chapter 2, the BS switching is performed under the unrealistic and optimistic assumption that the future traffic demand and the PV panel generation are perfectly know. For this reason, in section 5.4, ML algorithms are used to predict the future traffic demand and PV panel production and these predictions drive the BS switching.

5.1 Scenario

For our evaluation, the portion of the heterogeneous RAN in the orange rectangle of Fig. 5.1 is considered, which provides service to an area of 0.3 km^2. It is the real portion of the RAN located in the city centre of Ghent, in Belgium. It is composed of 8 macro cell BSs, marked by the blue points in Fig. 5.1, each supported by 4 micro cell BSs, indicated with the brown points in the figure, whose radio coverage overlaps with the macro cell. Thus, as in scenarios used in chapters 3 and 4, the micro cell BSs are deployed to provide additional capacity during high traffic demand periods. As in the previous chapters, the LTE-A is the wireless technology considered with a frequency of 2.6 GHz and a channel bandwidth of 5 MHz with a single transmitting and receiving antenna, i.e., Single Input Single Output (SISO) for both the micro

Fig. 5.1 Considered portion of RAN of the city centre of Ghent (Belgium), composed by 8 macro cell BSs, each supported by 4 micro cell BSs.

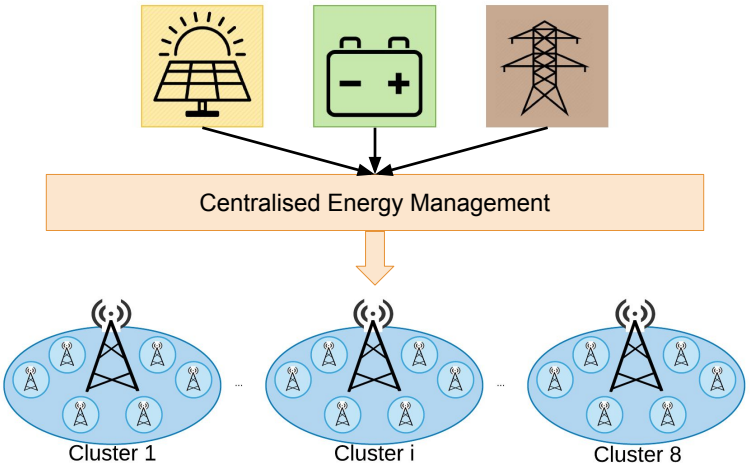


Fig. 5.2 Scheme of the the energy supply system in the considered portion of RAN.

and macro cell BSs. The link budget assumed in this work is the same of [113] and they are reported in Table 5.1.

With respect to the energy supply of the cluster, as depicted in Fig. 5.2, a centralised PV panel system, an energy battery, and the power grid are considered. As in [21], the capacity of the PV panel is 100 kWp, while the effective battery size is 50 kWh, with actually 71 kWh, since we consider a maximum Depth of Discharge (DOD) of 70%, which allows the battery to operate for more than 500-600 cycles before being replaced [114], [115]. Losses of 25% in energy efficiency due to the charging and discharging processes are considered [116]. The energy generated by the PV panel is used to power the BSs and in case of additional production, is conserved in the battery. If no renewable energy is available, the BSs take the required energy from the power grid. The data of the produced energy are taken by PVWATT [117].

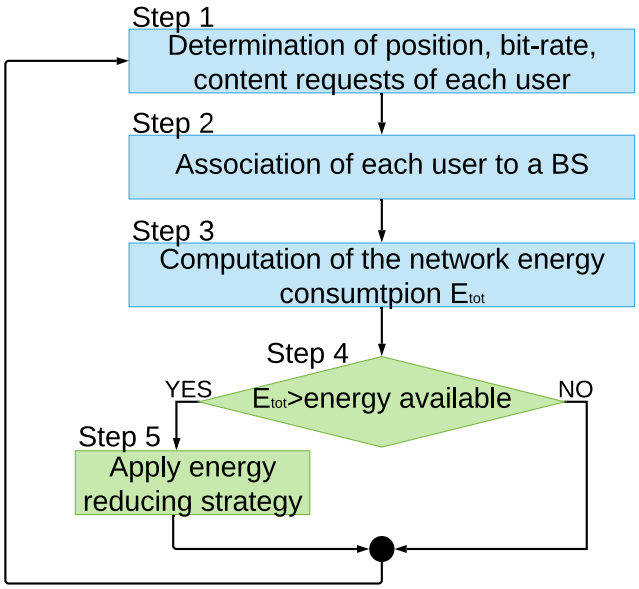


Fig. 5.3 Different steps of each time interval of each simulation.

The data estimates the electricity production of a typical poly- or mono-crystalline silicon PV-panel system, taking into consideration realistic solar irradiation patterns, corresponding to the typical meteorological year in the considered area, with a granularity of one hour. The main typical losses occurring in a real PV panel system during the process of solar radiation conversion into electricity are accounted for 14% and the efficiency is 20%. The data collected during the week from 3 January to 9 January, in Turin (Italy) are used in our simulations. The use of winter data leads to a worst case scenario in terms of produced energy.

5.2 Simulations

In order to evaluate the scenario described above, see section 5.1, we simulate it, assuming that it operates for 1 week and accounts for the 3D city environment, located in the considered portion of the city of Ghent. Each time interval of each simulation is structured in different steps. First, the traffic during each hour is generated: the number of active users, as well as their position and required bit rate (Step 1 in Fig. 5.3). Once, each user has been associated with a BS (Step 2 in Fig. 5.3), if possible, the total RAN energy consumption is computed (Step 3 in Fig. 5.3).

Table 5.1 Link budget parameters for the LTE-A macro cell and micro cell BS.

Parameters	Macro cell BS	Micro cell BS
Frequency	2.6 GHz	
Maximum input power antenna	43 dBm	33 dBm
Antenna gain base station	18 dBi	4 dBi
Antenna gain mobile station	2 dBi	
Soft hand over gain	0 dB	
Feeder loss base station	0 dB	
Feeder loss mobile station	0 dB	
Fade margin	10 dB	
Yearly availability	99.995%	
Cell interference margin	2 dB	
Bandwidth	5 MHz	
Receiver SNR	1/3 QPSK = –1.5 dB, 1/2 QPSK = 3 dB, 2/3 QPSK = 10.5 dB, 1/2 16-QAM = 14 dB, 2/3 16-QAM = 19 dB, 4/5 16-QAM= 23 dB, 2/3 64-QAM = 29.4 dB	
Used sub-carriers	301	
Total sub-carriers	512	
Noise figure mobile station	8 dB	
Implementation loss mobile station	0 dB	
Height mobile station	1.5 m	
Coverage requirements	90%	
Shadowing margin	13.2 dB	
Building penetration loss	8.1 dB	

If it is larger than the available green energy, i.e. the amount of energy stored in the battery and produced by the PV panel system, an energy reducing strategy is applied (Step 5 in Fig. 5.3). Details of each step are given in the following sections.

5.2.1 Generation of the traffic

In this stage (see Step 1 in Fig. 5.3), the number of active users is determined. As in [18, 21, 118], in order to do this, a user distribution is used. This distribution varies according to the hour of the day, to reflect the typical behaviour of the daily

traffic demand. Then, the position of each user u is determined, as 3D coordinates $(x^{(u)}, y^{(u)}, z^{(u)})$, located inside the orange area depicted in Fig. 5.1. The coordinates are generated according to a uniform distribution. The choice of the uniform distribution is because in the considered area there are no hot-spots such as tourist attractions or parks, meaning that each location has the same chance to be chosen as possible location for each user. In addition to the position, the requested bit rate is determined (Step 1 in Fig. 5.3). We assume that the bit rate required by each user is 1 Mbps [18, 119].

5.2.2 Creation of the network

The RAN adapts its capacity to the instantaneous bit rate requested by the active users. This means that, as in [18], the capacity of the network is not always totally used, but it dynamically responds to the instantaneous traffic demand. In order to do this, the input power of each BS is reduced as much as possible, while guaranteeing the user coverage [18]. In particular, in each time slot, once the position and the required bit rate have been determined for each user, the association process starts (Step 2 in Fig. 5.3). In order to associate each user with a BS, a list of possible BSs is created, to which the user can be connected. A BS is inserted in the list, if (i) it is active, (ii) it can provide the requested bit-rate and (iii) the experienced path loss is lower than an allowable maximum. To determine the experienced path loss, the direct line between the user and each considered BS is determined. In order to do this, the presence of existing buildings is taken into account, whose 3D data are provided by a shape file of the city of Ghent. As a result, it is possible to determine if the user is in a Line-of-Sight (LoS) or Non-Line-of-Sight (NLoS). According to this, the appropriate Walfish Ikegami (WI) propagation model is used [120]. This path loss has to be lower than the maximum allowable, in order to guarantee to still have sufficient quality at the receiver side. If the experienced path loss is larger than the allowable maximum, the input power of the BS is increased until the path loss becomes acceptable. In case the input power reaches the maximum allowable input power, but the path loss is still larger than the maximum, that BS is not inserted in the list. The user is associated with the BS of the list from which he/she experiences the lowest path loss (Step 2 in Fig. 5.3). In case the list results empty, the same procedure is performed using only the inactive BSs, i.e. removing the first condition for the insertion of a BS in the list. In case also this procedure provides an empty

list, the user remains uncovered. When an off BS is activated, the simulator checks if it is possible to transfer users already covered by other BSs to this "new" BS, if the available bit rate is sufficient and the experienced path loss is lower than the one experienced from the "old" BS, eventually increasing the transmitting power of the "new" BS, if possible and necessary. If all users of a certain BS are moved to this "new" BS, that BS is switched off.

5.2.3 Energy reduction strategy

In each time slot, once each user has been associated to a BS, if possible, the RAN energy consumption for the current time slot t is computed (Steps 3 in Fig. 5.3). Details of this computation are given in section 5.2.4. If the RAN energy consumption is larger than the renewable energy, which is available during that time interval, given by the energy produced by the PV panel system and stored in the battery, an energy reduction strategy is applied (Steps 4-5 in Fig. 5.3). The following energy reducing strategies are used in this part of the work:

1. **Always ON**: in this case, no action is taken during that time slot.
2. **Resource on Produced Energy (RoPE)**: micro cell BSs are put in sleep mode in case of energy shortage, i.e. in case the available renewable energy is lower than the energy consumption of the network for the considered hour, but are switched off gradually. To do this, the network energy consumption is computed, when switching off 1, 2, 3 or 4 micro cell BSs per macro cell. As soon as the network consumption becomes smaller than the amount of available renewable energy, that number of micro cell BS per macro cell BS is deactivated. The order in which the micro cell BSs are turned off follows the ascending number of users served by each micro cell BS. The users who have been connected to each deactivated micro cell BS are reconnected to an active BS, if possible, e.g., if there is an active BS that has enough available capacity and if the experienced path loss is lower than the maximum possible.
3. **Strict RoPE**: all micro cell BSs are deactivated during that hour, in case the generated and the stored renewable energy are not enough to power the network. Similar to RoPE strategy, the users who were connected to a micro

Table 5.2 Values of the parameters of the consumption model for the access and BH network.

	Macro BS	**Micro BS**
N_{sec}	3	1
n_{TX}	1	1
P_{rect} [W]	100	100
P_{mwl} [W]	80	0
P_{airco} [W]	225	60
P_{amp} [W]	12.8	12.8
P_{trans} [W]	100	100
P_{dsp} [W]	100	100

cell BS, which has been switched to sleep mode, are reconnected to an active BS, if possible.

4. **Super Strict RoPE**: once the network energy consumption is computed, in case the available renewable energy is smaller than the energy consumption of the network for the considered hour, macro cell BSs and their corresponding micro cell BSs are gradually switched in sleep mode. We do this by computing the energy consumption of the network, when switching off 1, 2,.., 8 macro cell BSs and the corresponding micro cell BSs. As soon as the available renewable energy becomes sufficient for the network supply, that number of BSs is deactivated. The order in which the macro BSs are selected to be turned off, follows the number of users served by each macro BS. After the deactivation, the users who were connected to a sleeping BSs are reconnected to an active BS, if possible.

5.2.4 Key Performance Indicators

The detailed scenario is evaluated with the following KPIs.

Energy consumption

The total energy consumption of the network E_{TOT}, in watt hour, during the simulation is computed as:

$$E_{TOT} = \sum_{t=1}^{H} E_C^{(t)} \tag{5.1}$$

where H is the duration of the simulation, i.e. the number of time slots, 168 in our simulations, given by 7 days in a week times 24 hours in a day, and $E_C^{(t)}$ is the energy consumption of the network at time *t*, in watt hour. It is computed according to the models for the macro cell and micro cell BS proposed and discussed in [113], which was validated by temporal measurements in the considered portion of RAN. Similar to the EARTH model, see [109] and section 3.3, these models are given by a constant and a load dependent components. In particular, for each BS, the energy consumption is modelled as:

$$E_C^{(b,t)} = \left[N_{sec} P_{rect} + P_{mwl} + P_{airco} + N_{sec} \left[n_{TX} \cdot (P_{amp} + P_{trans}) + P_{dsp} \right] \rho^{(t)} \right] \tau \quad (5.2)$$

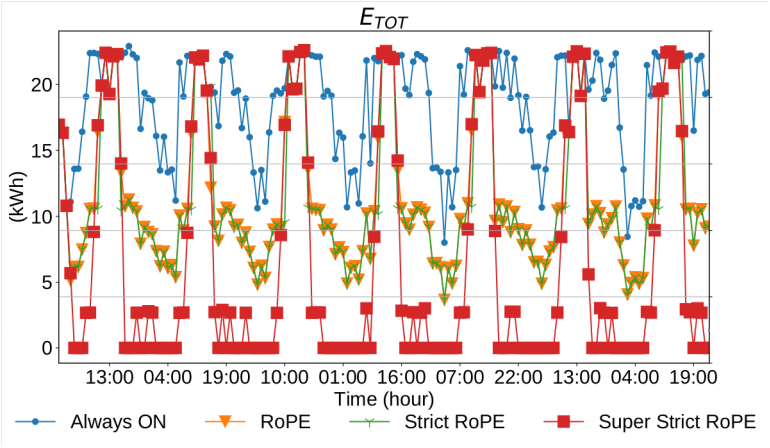
where N_{sec} is the number of sector of the BS, P_{rect}, P_{mwl} and P_{airco} are the power consumption, in watt, respectively, of the rectifier, the microwave link (if present) and the air conditioning system. The parameter n_{TX} represents the number of transmitting antennas, P_{amp}, P_{trans} and P_{dsp} are the power consumption of the power amplifier, the transceiver and the Digital Signal Processor, respectively, $\rho^{(t)}$ is the traffic load at time t and τ is the time duration of a time slot (in hour, 1 in our case). The values, for micro and macro BSs, are reported in Table 5.2.
If a BS is in sleep mode, its energy consumption is assumed negligible.

Lost Traffic

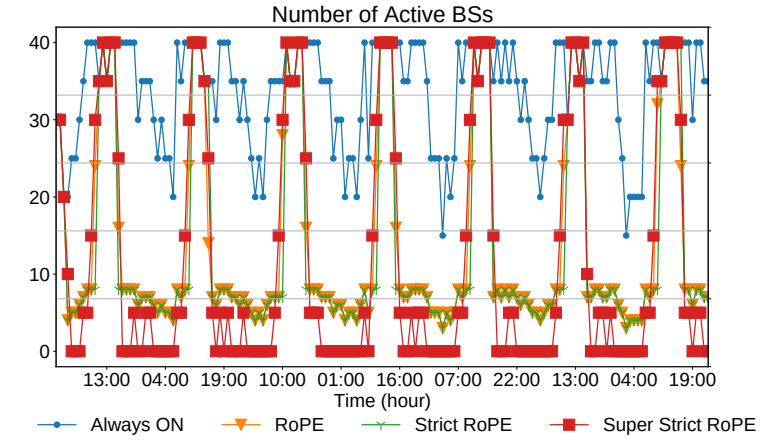
The lost traffic, L_T, is the percentage of unserved traffic, considering that a user can be associated with a BS only if he/she experiences a path loss lower than a given threshold and if that BS has enough capacity to provide the required bit rate.

5.3 Performance Evaluation

In this section we discuss the impact of the proposed energy reducing strategies. In Figs. 5.4a and 5.4b, the network energy consumption and the number of active BSs, are plotted, respectively, for each time slot of the simulation. The curves in blue and orange in the figures correspond to, respectively, Always ON and RoPE, which gradually deactivates micro cell BSs in case the renewable energy is not sufficient for the network supply. The green and red lines in Figs. 5.4a and 5.4b are the cases



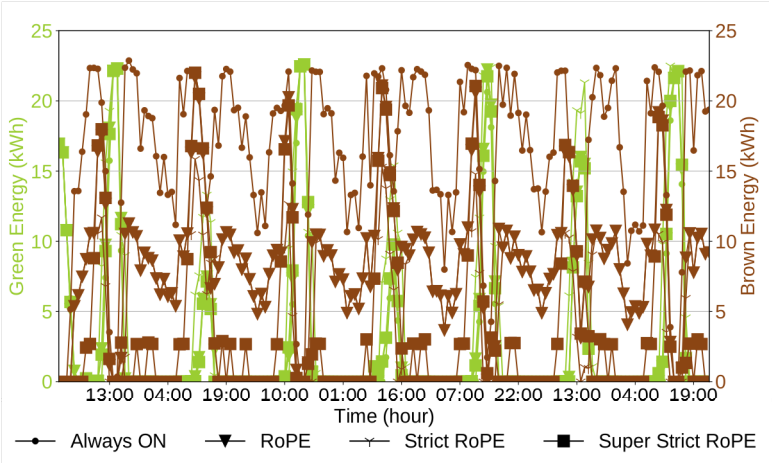
(a)



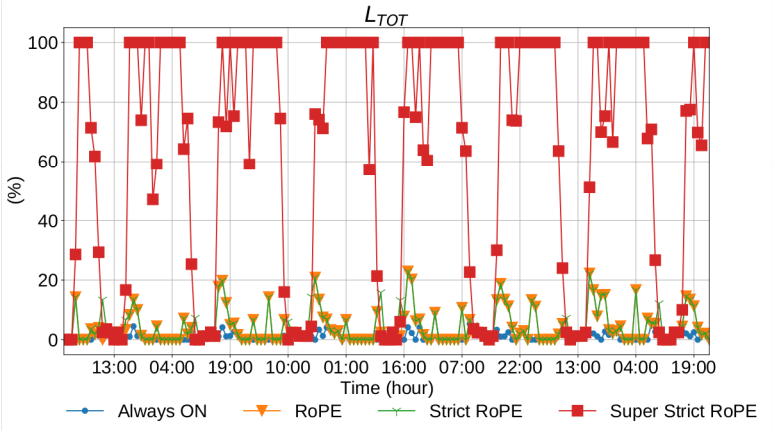
(b)

Fig. 5.4 Usage of the energy reducing strategies: (a) Energy Consumption and (b) Number of active BSs.

in which Strict RoPE and Super Strict RoPE are employed. In case the renewable energy is smaller than the network consumption, Strict RoPE deactivates each micro BSs, RoPE gradually deactivates each macro BSs and its micro BSs, to make the available green energy sufficient for the RAN supply. From these figures, we notice that, while the energy consumption and the number of active BSs are almost the same during the light hours for each used strategy, see Figs. 5.4a and 5.4b, the situation is different from 14:00 to 8:00. Indeed, in this time interval, since the renewable energy results insufficient to power the network because of lack of solar production, when Strict RoPE, RoPE and Super Strict RoPE are used, BSs are deactivated, reducing the RAN energy consumption, see orange, green and red curves in Figs. 5.4a and 5.4b. Strict RoPE and RoPE behave very similarly in terms of number of switched



(a)



(b)

Fig. 5.5 Usage of the energy reducing strategies: (a) Amount of Used Green and Brown Energy and (b) User Coverage.

micro cell BSs and energy consumption. This is because, even if the latter puts in sleep mode only the number of micro cell BSs which makes the renewable energy sufficient for the network supply, actually it often needs to deactivate all micro BSs (see Fig. 5.4b). As a consequence, it switches only a slightly lower number of BSs than the former strategy. This results in a very similar energy consumption, which is, respectively for Strict RoPE and RoPE, 1900 and 1940 kWh, 40% lower than Always ON case, whose consumption accounts for 3148 kWh. Besides the reduction of the network energy consumption, Strict RoPE and RoPE allow to buy from the grid a lower amount of brown energy. Fig. 5.5a reports, for each hour of the simulation, in green the amount of used renewable energy and in brown the amount of brown energy, which is bought from the grid, for each energy reducing

strategy, indicated with a different marker. The amount of green energy is almost the same for each strategy, but with Strict RoPE and RoPE, between 46% and 55% less energy is bought from the grid than when Always ON is used, which needs to buy 86% of its energy consumption from the grid. This shows that making the RAN more sustainable through the usage of renewable energy sources has to be coupled with the micro cell BS switching, in order to be effective. Indeed, if this is the case, the energy consumption and the bought energy are strongly decreased, reducing the OPEX expenditure and improving the RAN self-sufficiency. The inverse trend of the energy consumption and of active BSs presented by Strict RoPE and RoPE strategies occurs for the lost traffic, which results larger than when Always ON is employed, because of micro cell BSs deactivation, but rarely larger than 5%. This growth of the lost traffic underlines the key impact of the micro cell BSs on the QoS: they are necessary in order to always provide an optimal user coverage, i.e. lost traffic lower than 1%, which is provided when the micro cell BSs are not deactivated, as when Always ON strategy is used.
The situation is different for Super Strict RoPE. Indeed, it often deactivates all BSs during the night, generating interruption of the service, as shown by the curve marked by triangles of Fig. 5.4b. As a consequence, even if more than 64% of energy is saved, its nightly lost traffic goes to 100%, resulting in unacceptable QoS (see green curve in Fig. 5.5b). This highlights the fundamental role of the macro cell BSs in hierarchical RANs, in order to provide an adequate QoS, revealing that the deactivation of the macro cell BSs to save energy is not a feasible solution for our scenario. For this reason, this strategy is not considered in the following discussions.

5.4 RES production management through Machine Learning

As seen in the previous sections, micro cell BS switching according to the green energy availability allows to significantly reduce the usage of the brown energy, as well as to reduce the RAN energy consumption. Nevertheless, decisions are made assuming a perfect knowledge of the future green energy generation and RAN energy consumption, which depends on the traffic demand. In this part of the work, similar to what done in chapter 3, in order to overcome this unrealistic and optimistic assumption, we use RES production and traffic demand predictions.

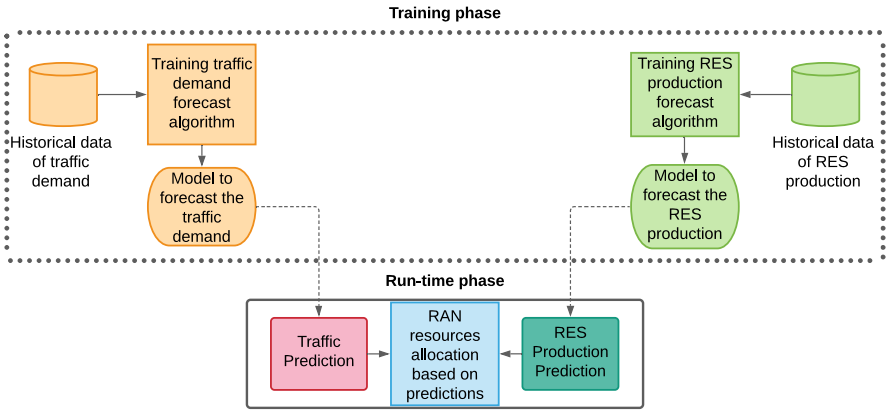


Fig. 5.6 Flowchart of the two-step network operation.

To do this, we look at the portions of a LTE-A RAN of chapter 3. In each of these RAN portions, there is a macro BS that defines a cell over the whole service area, and a few micro BSs that define small cells, whose coverage overlaps with the one of the macro, thus providing additional capacity in hot spots during peak traffic demand.
In this case, we assume that each cluster operates by using green energy provided by a PV panel system, and an energy storage (one or more batteries). The BSs use green energy generated by this PV panel; if extra energy is produced, it is stored in the battery, while if additional energy is needed to power the BSs, it is drained from the battery. Additionally, a connection to the power grid is available, to survive periods of insufficient or no RES production and empty battery; this is the brown energy component of the cluster. Hence, brown energy is used if the energy generated by the PV panel and the one stored in the battery are not enough to power the BSs.
In order to reduce the RAN energy consumption, so as to obtain both a reduction in energy cost, and a more energy parsimonious network operation, the MANO predicts the future PV panel production and the future traffic demand, to derive the RAN energy consumption, and, according to these predictions, it applies some resource management strategy.
Similar to chapter 3, as shown in Fig. 5.6, the MANO operation consists of 2 steps:

1. **Training phase**. The algorithms used to predict the RES production (green shapes in Fig. 5.6) and the traffic demand (orange shapes in Fig. 5.6) are trained using historical data. This phase is performed off-line.

2. **Run-time phase**. This is the operating part of the MANO operation. At the beginning of each time slot (that we assume to last 1h), the RES production and the traffic demand is forecast, using the predictor previously trained (red and dark green rectangles in Fig. 5.6). Given these predictions, the RAN energy consumption is computed and in case it is larger than the forecast available green energy, given by the amount of forecast energy generated by the PV panel and the stored energy, the network resources are allocated according to the implemented energy saving strategy (blue shape in Fig. 5.6).

5.4.1 Traffic Prediction

For the prediction of the traffic load of each BS, we use the data about many BSs located in Milan and in a wide area around it, discussed in 3.2, which report the traffic demand volume for two months in 2015, with granularity of 15 minutes. Data are aggregated to derive an hourly granularity and split in training set and data set. The former accounts for 75% of the data and is used in the training phase, see Fig. 5.6, and the latter corresponds to the remaining 25% of the data and is employed for the run time phase, see Fig. 5.6. The ML approaches presented and discussed in section 3.2 are employed for the forecast: 1/2/24/48 ANNs, LSTMC, BLR, Baseline and Baseline with ANN.

5.4.2 PV Panel Production Prediction

In this section, the ML approach used for the prediction of the energy generated by the PV panel is described. As for the traffic prediction, see section 5.4.1, it is trained in the run time phase, with historical data and is employed to the resource allocation in the run time phase.
Data about RES production in Milan are estimated using the data set used in section 5.1, provided by PVWATT [117]. As depicted in section 5.1, the PVWATT data are derived based on realistic solar irradiation patterns, corresponding to the typical meteorological year in the considered area, with a granularity of one hour. The main typical losses occurring in a real PV system during the process of solar radiation conversion are taken into account. In our work, the PV panel system has a capacity of 10 kWp. Lead-acid batteries are considered for energy storage. Each element has capacity 200 Ah and voltage 12 V. The storage capacity, defined as the number of

battery units, is equal to 5. As in section 5.1, the maximum DoD is 70% and the losses in energy efficiency in the charging and discharging processes account for 25%.
The amount of the energy produced at time t is predicted using linear regression, giving as input the following features:

- month, day, current hour,
- average of the beam irradiance at time *t-1* (W/m^2),
- average of the diffuse irradiance at time *t-1* (W/m^2),
- average of the ambient temperature at time *t-1* (°C),
- average of the wind speed at time *t-1* (m/s) ,
- average of the plane of array irradiance at time *t-1* (W/m^2),
- average of the PV cell temperature at time *t-1* (°C)

5.4.3 Energy Reducing Strategy

As energy reducing strategy, we use RoPE and Super Strict RoPE, implemented similarly to the ones used in section 5.2. Super Strict RoPE, discussed in section 5.2, is not considered because of the unacceptable QoS which is obtained when it is used. The micro cell BSs are switched off if the local available green energy is not enough to power the BSs and the macro cell can carry the traffic of the micro cells. In particular, in each time slot, the traffic demand of each BS and the PV panel production are forecast. Then, the RAN energy consumption is computed, using the traffic predictions. If **RoPE** is employed, when the predicted amount of available green energy (produced and stored) is lower than the predicted energy necessary to operate the cluster in the next time slot, the micro cell BSs are switched off gradually, provided that their forecast traffic can be managed by the macro cell, until the predicted available green energy becomes sufficient to operate the cluster in the next time slot.
When **Strict RoPE** is used, in case the predicted available green energy (produced and stored) is lower than the predicted energy required to power the cluster in the next time slot, the micro cell BSs are switched off, if their forecast traffic can be carried by the macro cell BS.

5.4.4 Key Performance Indicators

As in chapter 3, we evaluate the scenario in terms of total energy consumption E_{TOT}, in Wh, computed as in section 3.4, where the energy consumption of each BS is derived through the EARTH model. Also the percentage of lost traffic, L_T, is measured, as discussed in section 3.4. The predictions of the traffic demand have been already discussed in section 3.5, in terms of AME and ARE. In this chapter, in order to evaluate the RES prediction accuracy, they are computed for the forecast of the green energy generation. In particular, AME refers to the difference between the real and predicted energy generation by RES, it is computed as:

$$ARE = \frac{1}{H}\sum_{t=1}^{H}\frac{|E^{(t)} - \hat{E}^{(t)}|}{E^{(t)}} \tag{5.3}$$

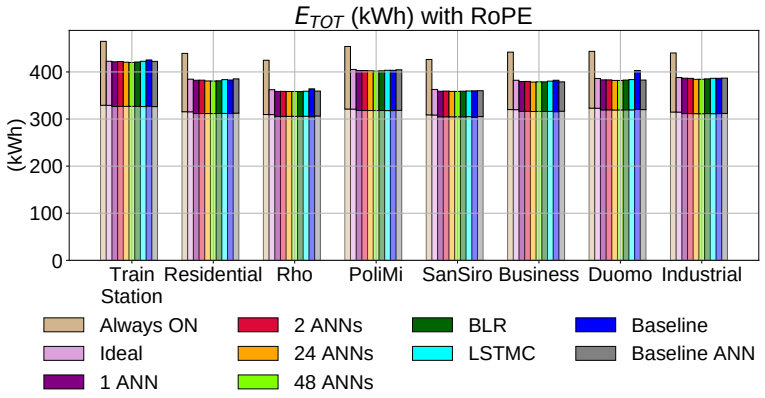
where $E^{(t)}$ is the real renewable energy production at time *t*, $\hat{E}^{(t)}$ is the forecast renewable energy production at time *t*, H is the duration of the testing period, in number of time slots.
The AME indicator is also evaluated. It is the difference between the real and predicted energy generation by RES:

$$AME = \frac{1}{H}\sum_{t=1}^{H}(E^{(t)} - \hat{E}^{(t)}) \tag{5.4}$$

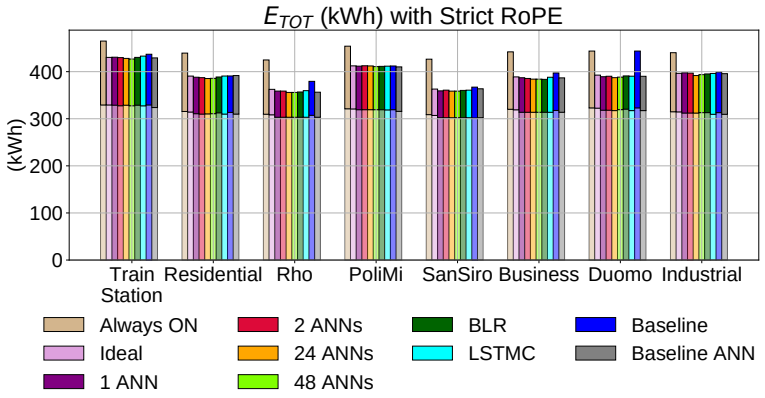
where $E^{(t)}$ is the real renewable energy production at time *t*, $\hat{E}^{(t)}$ is the forecast renewable energy production at time *t* and H is the number of time slots of each simulation.

5.4.5 Performance Evaluation

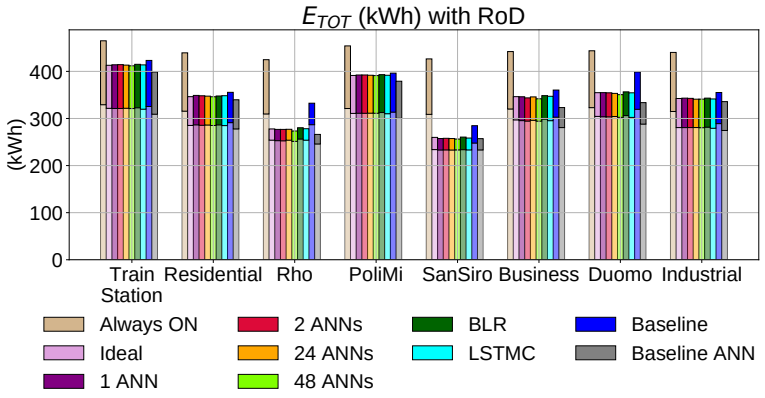
We now evaluate the performance of the RoPE and Strict RoPE energy consumption reduction strategies, which deactivate micro BSs, gradually and totally, respectively, when the available energy is lower than the consumed one, provided that their traffic can be carried by the macro BS. Super Strict RoPE, discussed in section 5.2, is not considered because of the unacceptable QoS which is obtained when it is used. We compare the results with the ones presented in section 3.5, obtained with RoD as energy consumption reduction strategy, which turns to sleep mode a micro cell BS, if



(a)



(b)



(c)

Fig. 5.7 Total Energy Consumption E_{TOT}, in kWh, using RoPE (a), Strict RoPE (b) and RoD (c).

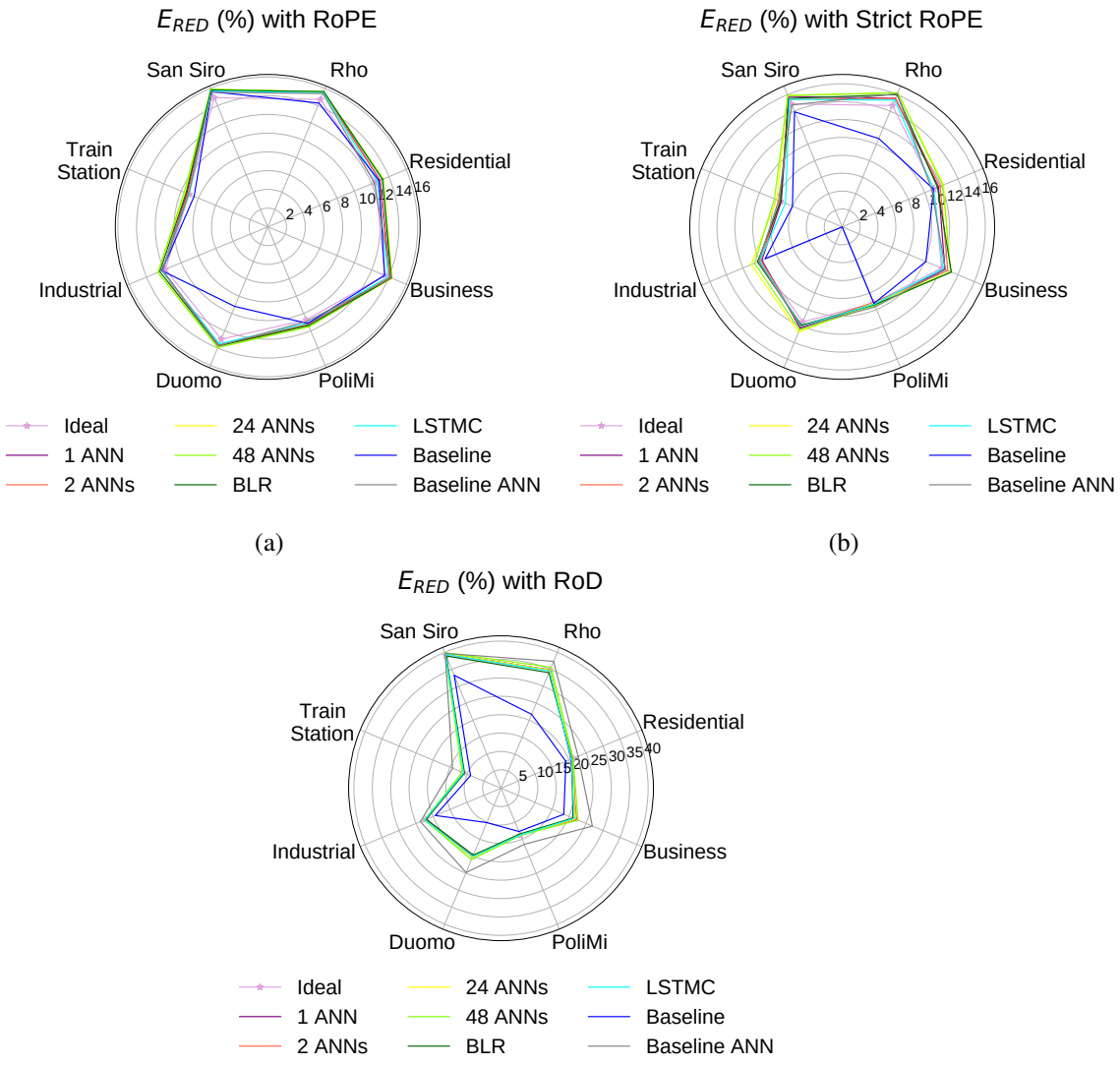
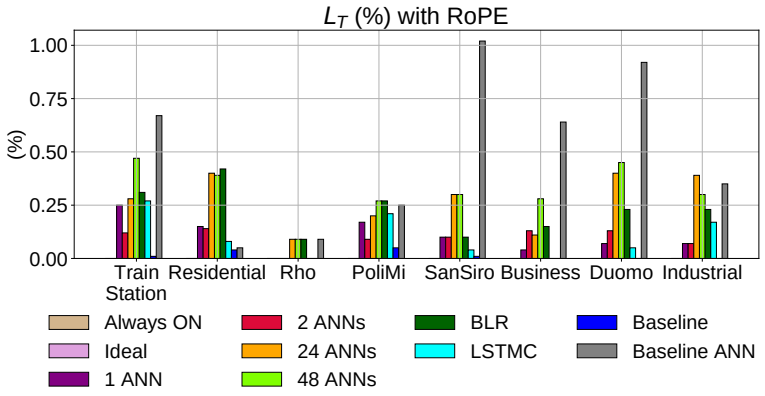
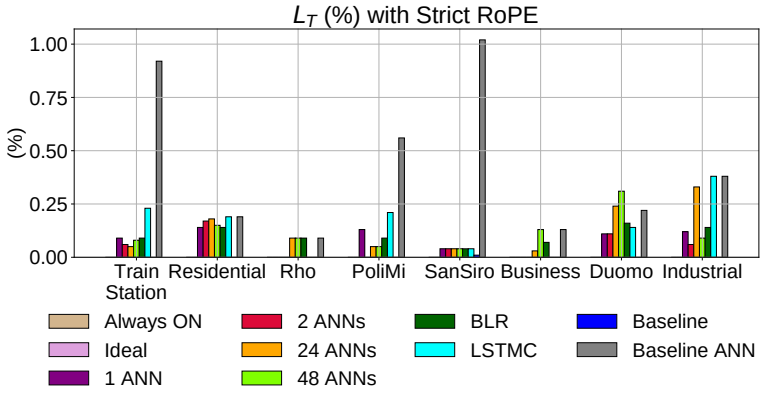


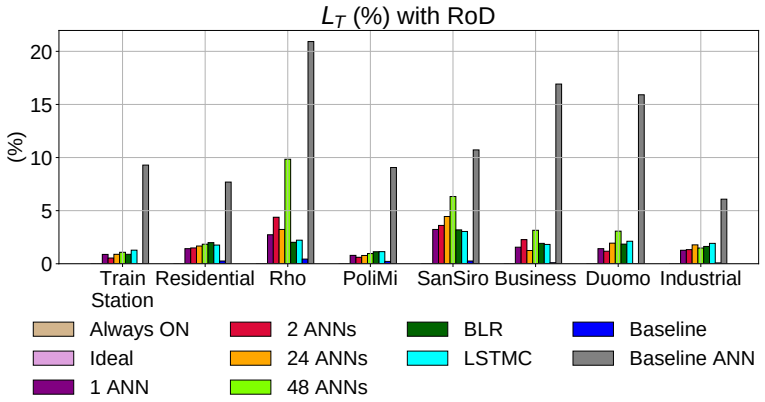
Fig. 5.8 Total Energy Consumption Reduction E_{RED}, using (a) RoPE, (b) Strict RoPE and (b) RoD.



(a)



(b)



(c)

Fig. 5.9 Lost Traffic L_T, using (a) RoPE, (a) Strict RoPE and (c) RoD.

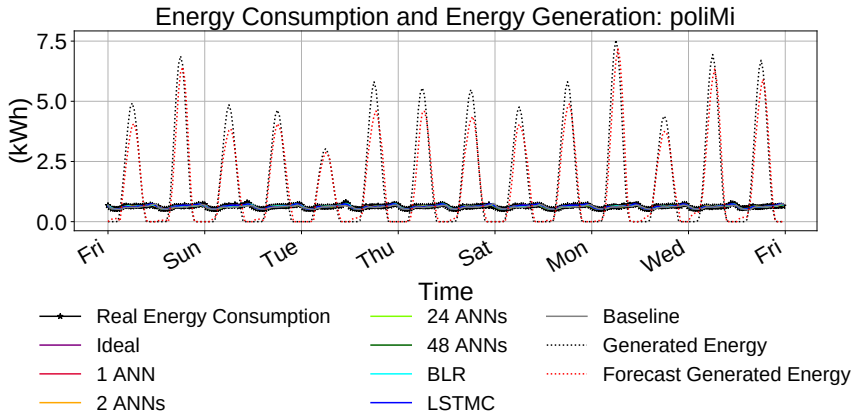


Fig. 5.10 Real and forecast energy generation by RES and real and forecast energy consumption.

its traffic demand is lower than ρ^*, provided that that traffic demand can be handled by the macro BS, see section 3.1.

The bar plots of Figs. 5.7a, 5.7b and 5.7c show the total energy consumption, E_{TOT}, distinguishing the green (lighter) and the brown used energy, when RoPE, Strict RoPE and RoD are used, respectively. The energy consumption reduction, E_{RED}, computed with respect to the Always ON scenario, in which no action is taken to reduce the energy consumption, is reported in the radar plots of Figs. 5.8a, 5.8b and 5.8c, when these three strategies are employed, respectively. Figs. 5.9a, 5.9b and 5.9c show the percentage of lost traffic with RoPE, Strict RoPE and RoD, respectively.

Figs. 5.7a, 5.7b and 5.7c show that the energy consumption achieved when ML algorithms are used is almost identical to the ideal case, which assumes a perfect knowledge of the future traffic demand, as well as of the PV panel production. In addition, we notice that most of the consumed energy is produced locally by PV panels, and only a small amount, between 14% and 22% with RoPE, between 14% and 23% with Strict RoPE and between 7% and 22% with RoD, is purchased from the power grid, typically during the night, when energy is cheaper.

The figure indicates that, in all the considered areas, the energy saving with RoPE and Strict RoPE is between 8% and 16%, lower than from 9% to 40%, which is achieved with RoD. This is because the condition to switch micro cell BSs off with RoPE and Strict RoPE occurs rarely, if compared with the conditions for the micro BS deactivation in RoD. To show this, Fig. 5.10 reports the real and the forecast renewable energy generation (black and red dashed lines, respectively) and the real and forecast energy consumption in the PoliMi area, computed with the traffic

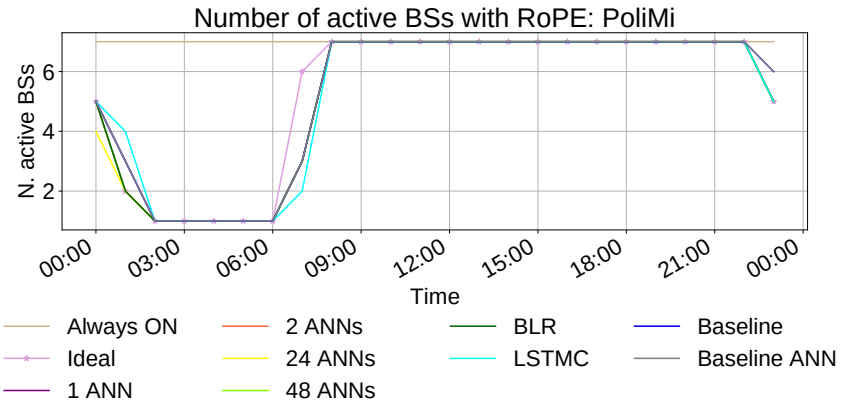
predictions given by the different ML algorithms. The ARE and the AME of the production of RES is 15 and $137 \cdot 10^3$, respectively. These values and Fig.5.10 reveal that the estimation algorithm fails mainly in the forecast of the peaks. The predicted amount of RES production is lower than the actual value, but greater than the forecast and real power consumption. For this reason, the RAN is usually configured properly. In addition, the figure reveals also that the amount of available renewable energy is often sufficient to power the cluster, keeping all the micro BSs active (see Fig. 5.10). During these periods, the obtained network configurations are identical, even using different traffic prediction algorithms (from 8 a.m. to 10 p.m. in Fig. 5.11a). This does not occur between 11 p.m. and 7 a.m., when the green energy produced by the PV panels, collected and stored in the battery during the day, is not sufficient, see Fig. 5.10. Part of this time interval coincides with low traffic demand periods. In this case, even with different errors in the forecast, the configurations obtained with any traffic forecast model are minimal: typically, all the micro BSs are put in sleep mode. As a consequence, the energy consumption reduction, E_{RED}, achieved with RoPE and Strict RoPE is less dependent on the traffic pattern, than RoD, see Figs. 5.8a and 5.8c.

For the same reason, under the RoPE and Strict RoPE strategies, the behaviour is almost the same whatever ML algorithm is used, as can be seen in Figs. 5.11a and 5.11b, which reports the number of active BSs during one day in the PoliMi area. Nevertheless, when Strict RoPE is used and Baseline is employed as the traffic forecast algorithm, the energy saving is no larger than 14% and always smaller than the reduction achieved with Strict RoPE, when the other ML algorithms are used, and with RoPE. This is because the Baseline algorithm typically overestimates the traffic demand, as highlighted in section 3.5. Since Strict RoPE considers the deactivation of the six micro cell BSs together and not gradually, as in RoPE, the traffic overestimation makes the available capacity of the macro insufficient to carry the traffic demand of the micro BSs and, consequently, the micro cell BSs are not turned off.

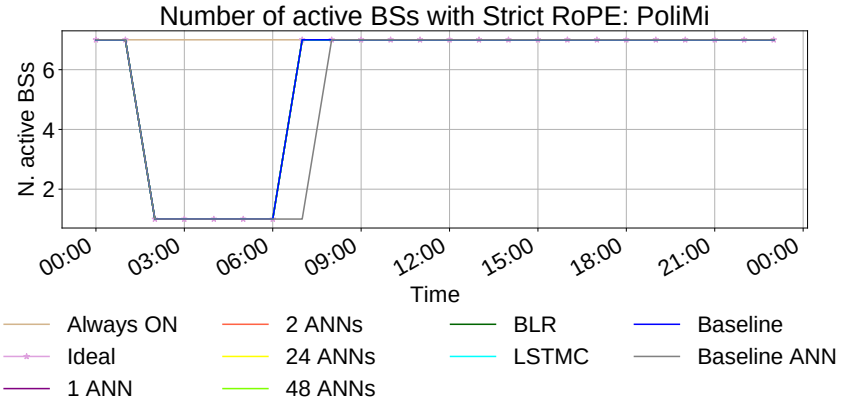
Because of these reasons, with RoPE and Strict RoPE, QoS is less compromised than with RoD: less than 1% of traffic is lost with RoPE and Strict RoPE, see Figs. 5.9a and 5.9b, while up to 20% with RoD, as can be noticed in see Fig. 5.9c.

5.5 Final Remarks

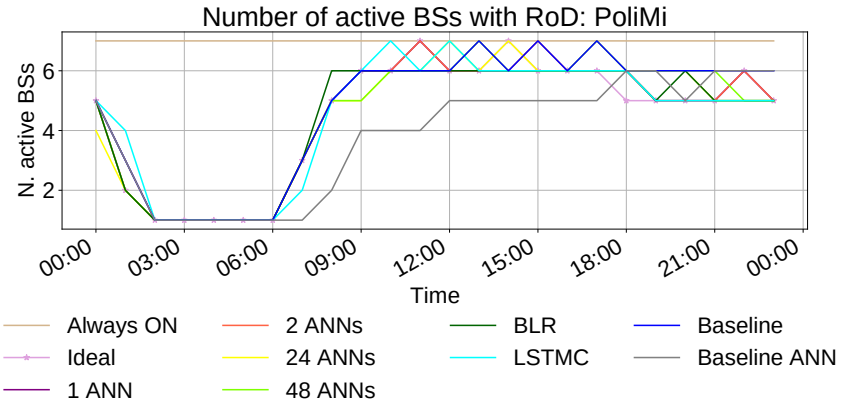
In this part of the work, RES is considered for the RAN supply. In particular, we considered a portion of the network, which is powered by a PV panel system and energy batteries and, if needed, by the power grid. The energy generated by the PV panel is used to power the BSs and in case of additional production, is stored in the battery. If the renewable energy is not sufficient for the network supply, the BSs take the missing energy from the power grid. In order to reduce the RAN energy consumption and the brown energy bought from the grid, energy reducing strategies are applied in case of lack of green available energy. Results highlight that making the RAN more sustainable through the usage of renewable energy sources has to be coupled with the micro cell BS switching, in order to be effective. Indeed, if this is the case, the energy consumption and the bought energy are strongly decreased, reducing the OPEX expenditure and improving the RAN self-sufficiency. Nevertheless, our results highlight the fundamental role of the macro cell BSs in hierarchical RANs, in order to provide an adequate QoS, revealing that the deactivation of the macro cell BSs to save energy is not a feasible solution for our scenario.
In addition, the usage of ML approaches is effective for the predictions of RES production and traffic demand. These forecasts drive the decision for the energy reducing strategies and results show that the network performance does not depend much on the used ML algorithms. This suggests that the use of RoPE is preferable to RoD, and that a simple ML algorithm like BLR, which constructs a single model for all the BSs of all the city areas, or 1 or 2 ANNs combined with RoPE can produce a very effective and flexible approach for energy efficiency and QoS.



(a)



(b)



(c)

Fig. 5.11 Number of the active BSs during a day in the PoliMi area, using (a) RoPE, (b) Strict RoPE and (c) RoD.

Chapter 6

Energy Efficiency and Edge Caching in RAN

Part of the work presented in this chapter has already been published in:

- Vallero, G., Deruyck, M., Joseph, W., & Meo, M. (2020, June). Caching at the edge in high energy-efficient wireless access networks. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.
- Vallero, G., Deruyck, M., Meo, M., & Joseph, W. (2021). Base Station switching and edge caching optimisation in high energy-efficiency wireless access network. *Computer Networks*, 192, 108100.

In chapters 3, 4 and 5, strategies based on the BS switching are used, to reduce the network energy consumption, without deteriorating the QoS, through the support of ML approaches. In the new generation of network, in order to cope with the network traffic demand growth and the strict latency requirements of many services and applications, the MEC technology has been introduced. It uses caching and computing platforms, located at the edge of the network and specifically at the BSs, in the RAN. Nevertheless, the impact of the MEC technology employment on the network energy efficiency is usually neglected and the effect of the BSs switching on the MEC technology performance is ignored, as well. For this reason, the simultaneous employment of the MEC technology and the BSs switching is considered in this chapter, providing an overview of their mutual effects. In particular, using a

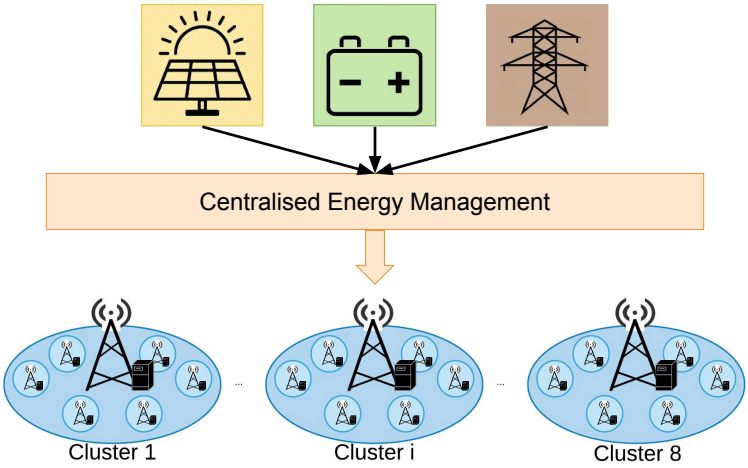


Fig. 6.1 Scheme of the the considered portion of RAN.

simulation-based approach, we quantify the gain as well as the cost, which derive from the usage of caching servers, placed at each BS. The gain is measured in experienced delay drop, and the cost is expressed in growth of the energy consumption of the network. The effects of the traffic characteristic, as well as of the distribution of the cache among BSs are analysed. These quantities are also evaluated when an energy reducing strategy is used, which deactivates micro cell BSs in case of local renewable energy shortage, similar to chapters 3 and 5. In addition, in section 6.5, we use different association policies which we have designed, in order to minimise the RAN energy consumption and/or the experienced delay, i.e. to maximise the benefits provided by the MEC technology usage, ensuring also the achievement of the network energy efficiency.

6.1 Scenario

The portion of the heterogeneous RAN, illustrated in Fig. 6.1, covering an area of 0.3 km^2, used in chapter 5, in the orange rectangle of Fig. 5.1, is considered. As explained in chapter 5, it includes the 8 macro cell BSs (blue points in Fig. 5.1), located in the city centre of Ghent, in Belgium. As depicted in Fig. 6.1, each macro BS is supported by 4 micro cell BSs (brown points in Fig. 5.1), which are deployed to provide additional capacity during high traffic demand periods. As in the

previous chapter, the LTE-A is the wireless technology considered with a frequency of 2.6 GHz and a channel bandwidth of 5 MHz with SISO for both the micro and macro cell BSs. The link budget assumed in this work is the same as [113] and they are reported in Table 5.1.
We assume that each BS of the cluster, both macro and micro BS, is equipped with a caching server, see Fig. 6.1, to push contents closer to the users. Similarly to [1], the HW technology of each cache is Dynamic Random Access Memory (DRAM). As in [35] and [73], these servers update their contents according to the Least Frequently Used (LFU) cache algorithm, to store the most popular contents of a file library composed of 1000 files, of 100 Mbit size each. Each file has the same dimension, but this assumption can be easily removed, since in real systems, files can be split into blocks of the same length [121, 122].
As in chapter 5 and illustrated in Fig. 6.1, a centralised PV panel system, whose capacity is 100 kWp, is used as primary RAN energy supply. In addition, an energy battery is used. Its effective battery size is 50 kWh, with actually 71 kWh, since a maximum Depth of Discharge of 70%, respectively, is assumed, to ensure the maximum battery life. The energy generated by the PV panel is used to power the BSs and in case of extra production, is conserved in the battery. If no renewable energy is available, the required energy is withdrawn from the power grid. As in previous chapters, the data of the produced energy are taken by PVWATT [117]. The data estimates the hourly electricity production of a typical PV panel system, considering realistic solar irradiation patterns of the meteorological year in the considered area, accounting for 14% and 20% of losses occurring in a real PV panel system during the process of solar radiation conversion into electricity and efficiency, respectively. Also for this part of the work, the data production of the week from 3 January to 9 January, in Turin (Italy) are used in our simulations, to lead to a worst case scenario in terms of produced energy.

6.2 Simulations

In order to evaluate the scenario described above, we simulate it, using an extended version of the simulator described in section 5.2. We assume an operating time of 1 week, considering the 3D city environment, located in the considered portion of the city of Ghent. In the simulations, the traffic during each hour is generated: the

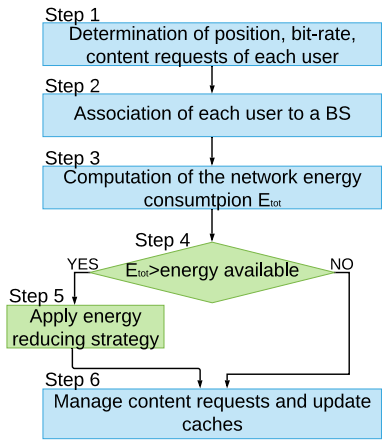


Fig. 6.2 Different steps of the simulations.

users, as well as their position, required bit rate and number of requested contents are determined (Step 1 in Fig. 6.2). Then, each user is associated with a BS (Step 2 in Fig. 6.2), if possible, and the content requested by each user is determined (Step 3 in Fig. 6.2). Then, the energy consumption of the RAN is computed (step 3 in Fig. 6.2) and if it is lower than the available energy, given by the amount of energy which is produced by the PV panel system and stored in the battery, an energy reducing strategy is applied (steps 4 and 5 in Fig. 6.2). Finally, the requested contents are delivered and caches are updated (Step 7 in Fig. 6.2). Details of each step are given in the following sections.

6.2.1 Generation of the traffic

As described in section 5.2, in order to determine the number of active users in a time slot, a user distribution is used, which varies according to the hour of the day, to shape the hourly variation of the daily traffic demand. Then, for each user of each time slot, the 3D coordinates of each user u, $(x^{(u)}, y^{(u)}, z^{(u)})$, are determined through a uniform distribution, and we assume that each user requires 1 Mbit/s as bit rate (Step 1 in Fig. 6.2). Afterwards, the number of generated content requests for each user is determined. To do this, as in [31], a Poisson distribution is used, whose parameter λ is 1 request/minute.

6.2.2 Creation of the network

In each time slot, once the position, the required bit rate and the number of content requested have been determined for each user (Step 1 in Fig. 6.2), the association process begins (Step 2 in Fig. 6.2). Similar to the procedure described in section 5.2, a user is associated to BS, from which he/she experiences the lowest path loss, computed as in section 5.2.2, provided that (i) that BS is active, (ii) it has enough capacity to serve that user and (iii) the experienced path loss is lower than the maximum allowable path loss to receive the signal with acceptable quality. In order to verify this last condition, the input power of the BS can be increased until the path loss becomes acceptable, up to the maximum allowable input power. If there is no active BS which satisfy these requirements, conditions (ii) and (iii) are verified on the inactive BSs and the BS from which the user experiences the lowest path loss is picked. In case an off BS is activated, the simulator checks if it is possible to transfer users already covered by other BSs to this "new" BS, if the available bit rate is sufficient and the experienced path loss is lower than the one experienced from the "old" BS, eventually increasing the transmitting power of the "new" BS, if possible and necessary. If all users of a certain BS are moved to this "new" BS, that BS is switched off.
Once a user is associated to a BS, the requested files are determined, according to the popularity distribution of the BS from which he/she experiences the lowest path loss, derived as in section 5.2.2 (Step 3 in Fig. 6.2). According to [35, 73, 123, 72, 2, 1], this popularity distribution is a Zipf distribution, characterised by the parameter α. This parameter impacts the difference among contents in terms of popularity. A large α means that the most popular contents are significantly more popular than the other contents, and decreasing α, the popularity of content behaves more similarly to the uniform distribution. The level of popularity of each content on each macro cell BSs is determined starting from a reference popularity and performing random shuffles on it. In particular, sorting the files of the library from the most popular to the least popular, according to this reference popularity, the popularity of 30% of content is randomly swapped to generate the popularity on each macro cell BS, so that slight differences among the files popularity at different locations (i.e., at different BSs) are introduced [31]. A similar procedure is performed to determine the popularity at each micro cell BS. In this case, starting from the popularity of the corresponding macro cell BS, the popularity of 15% of the contents is shuffled.

6.2.3 Energy reduction strategy

In each time slot, once each user has been associated to a BS, if possible, we compute the RAN energy consumption for the current time slot t (Step 3 in Fig. 6.2), as detailed in section 6.2.6. In case the renewable energy, which is available during that time interval, given by the energy produced by the PV panel system and stored in the battery is not sufficient for the RAN energy supply, an energy reduction strategy is applied (Steps 5-6 in Fig. 6.2):

1. **Always ON**: in this case, no action is taken during that time slot.
2. **RoPE**: as in section 5.4.3, micro cell BSs are gradually put in sleep mode. To do this, the network energy consumption is computed, when switching off 1, 2, 3 or 4 micro cell BSs per macro cell BSs. As soon as the network consumption becomes smaller than the amount of available renewable energy, that number of micro cell BS per macro cell BS is put in sleep mode. The order in which the micro cell BSs are turned off follows the number of users served by each micro cell BS. The users who have been connected to each deactivated micro cell BS are reconnected to a macro cell BS, if possible, e.g., if there is a macro cell BS that has enough available capacity and if the experienced path loss is lower than the maximum possible.
3. **Strict RoPE**: all micro cell BSs are deactivated during that hour and the users who were connected to a micro cell BS, which has been switched to sleep mode, are reconnected to an active BS, if possible.

6.2.4 Content delivery

In each time interval, once the energy reduction strategy is applied, if needed, requested contents are delivered (Step 7 in Fig. 6.2). When the requested content is cached in the server of the serving BS, the content is transmitted directly to the user. Notice that the content transfer to the user, from the BS, which he/she is associated with, always incurs [124]. We assume that this access latency $T_{bs,u}$ for the cells is identical and equal to 30 ms, as in [2]. If the requested content is not cached by the serving micro cell BS but by the macro cell BS, the macro cell BS transmits the content to that micro. The link between a micro cell BS and its corresponding

macro is wired, through optic fibres. Nevertheless, due to the insufficient capacity of the BH links, which causes a bottleneck in this segment of the network, its latency contribution, $T_{BS,bs}$, is significant [125], but lower than the one in the wireless link, between the user and the micro BS. We assume that it is equal to 20 ms [2]. The resulting latency is given by $T_{BS,bs} + T_{bs,u}$ [2]. If the content is not present not even on the macro cell BS, the request is forwarded to the content provider. In this case, the experienced latency is given by $T_{cp,BS} + T_{BS,bs} + T_{bs,u}$, with $T_{cp,BS}$ equal to 50 ms, giving a total latency equal to 100 ms, as in [2, 124]. In case a user is associated with a macro cell BS, which is caching the requested content, that content is received with latency $T_{bs,u}$. If that content is not stored in the server of that macro cell BS, it is retrieved on the content provider and the user receives it with delay $T_{cp,BS} + T_{bs,u}$. The values of the latency are summarised in Table 6.1.
After each content delivery, the cache is updated (Step 7 in Fig. 6.2), according to the LFU cache algorithm, so as to always cache the most popular contents.

6.2.5 Initial state of caches

We determine which contents are stored in each cache at the beginning of each simulation, through a preliminary phase. For each value of α, the scenario is simulated, assuming that the considered RAN operates for 100 weeks. Each simulation begins with empty caches and the occurrences of a content request are updated, as well as the stored files in each cache, at each time slot. When the number of variations in each cache stabilises, the transient phase for each cache filling is assumed to be over. As a result, the files in each cache at that time interval are the initial state of the caches in simulations.

6.2.6 Key Performance Indicators

The detailed scenario is evaluated with the following KPIs.

Energy consumption

As in chapter 5, the total energy consumption of the network E_{TOT}, in watt hour, is given by $E_{TOT} = \sum_{t=1}^{T} E_{TOT}^{(t)}$, where T is the number of time slots, 168 in our

simulations, i.e. the number of hours in a week, and $E_{TOT}^{(t)}$ is the energy consumption of the network at time *t*, in watt hour, which takes into account both the energy needed for the communication unit supply and the MEC server supply:

$$E_{TOT}^{(t)} = \sum_{b=1}^{N_{BS}} E_{C}^{(b,t)} + \sum_{b=1}^{N_{BS}} E_{MEC}^{(b,t)} \tag{6.1}$$

where N_{BS} is the number of the active BSs, $E_{C}^{(b,t)}$ is the energy consumption of the BS *b* at time *t*, due to the communication features and $E_{MEC}^{(b,t)}$ is the energy consumption of the BS *b* at time *t*, due to the supply of the cache located on that BS. According to [1] and [35], $E_{MEC}^{(b,t)}$, in watt hour, is given by:

$$E_{MEC}^{(b,t)} = \omega_{MEC} C_{MEC} \tau \tag{6.2}$$

where ω_{MEC} is in W/bit, C_{MEC} is the capacity of the server and τ the time duration of a time slot (in hour, 1 in our case).
The $E_{C}^{(b,t)}$ is the energy needed for the communication unit and is computed with the model for the macro cell and micro cell BS of (5.2), with parameter values reported in Table 5.2.

Lost Traffic

As in chapter 5, the lost traffic, L_T, accounts for the percentage of unserved users, considering that a user can be associated to a BS only if he/she experiences a path loss lower than a given threshold and if that BS has enough capacity to provide the required bit rate.

Average Delay

This is the average delay experienced by users and it is given by:

$$D_{AVG} = \frac{1}{\sum_{t=1}^{T} U_t \sum_{u=1}^{U_t} R_u} \sum_{t=1}^{T} \sum_{u=1}^{U_t} \sum_{r=1}^{R_u} d^{(u,r)} \tag{6.3}$$

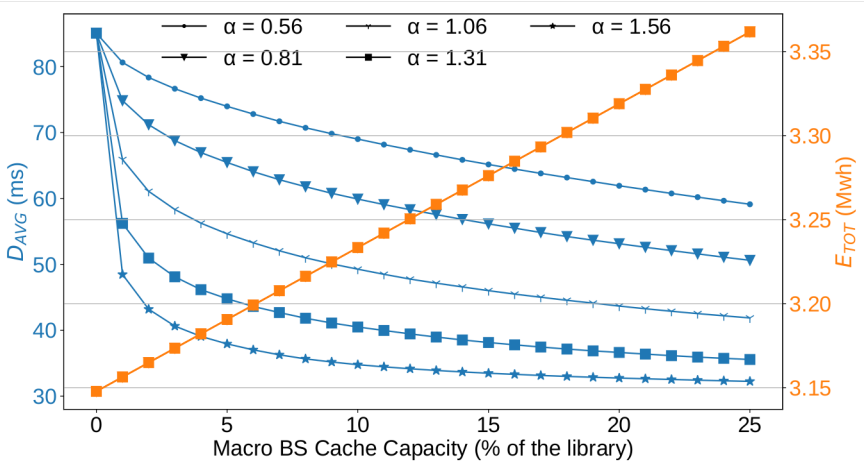


Fig. 6.3 Average delay (in blue) and energy consumption (in orange) varying the dimension of each cache, for different values of the parameter α.

where $d^{(u,r)}$ is the delay which is experienced by the user u, for the content request r. R_u, U_t and T correspond to the number of requests required by the user u, the number of served users at time t and the duration of the simulation, respectively.

Hit - 1 hop probability

This is the probability that the requested content is stored locally on the BS to which the considered user is connected.

Hit - 2 hops probability

This is the probability that the content requested by a user associated with a micro cell BS is not cached on that micro cell BS but on the corresponding macro cell.

Miss probability

This is the probability that the requested content is taken from the content provider, since it is not stored in the cache of the BS, to which the user is associated nor in the one of the corresponding macro cell BS, if the considered user is associated to a micro cell BS.

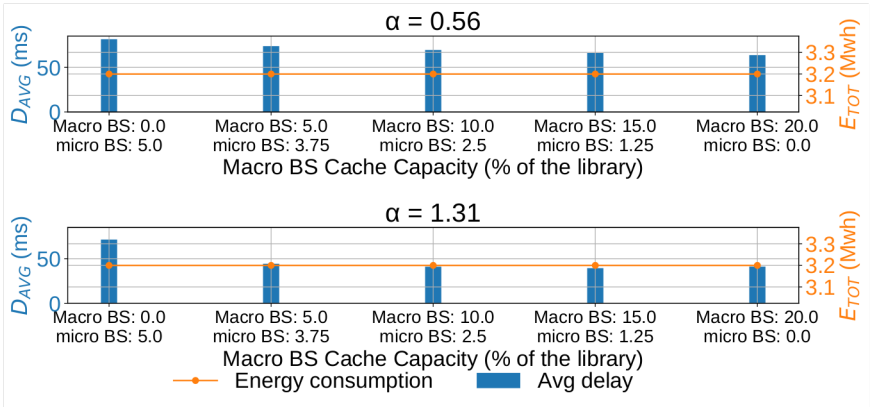
Table 6.1 Values of parameter used in simulations [1, 2].

Parameter	Value
$T_{bs,u}$	30 ms
$T_{BS,bs}$	20 ms
$T_{cp,BS}$	50 ms
ω_{MEC}	$2.5 \cdot 10^{-9}$ W/bit

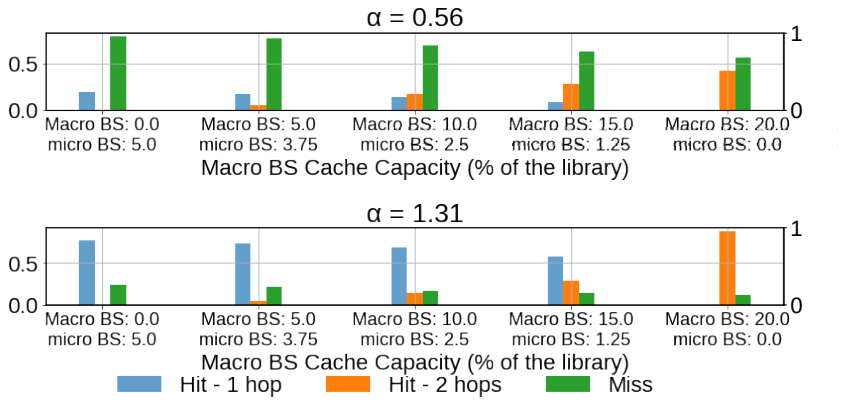
6.3 Energy Efficiency Effects of MEC

6.3.1 Impact of the size of the cache

In this section, we analyse the benefits provided by the caching on the BSs, as well as the effects of the parameters which affect the performance of the local caching. To do this, we simulate the scenario described in section 6.1, using Always ON as the energy reduction strategy. Fig. 6.3 shows on the left y-axis the average experienced latency, in blue, and on the right y-axis the energy consumption, in orange, varying the size of the cache on each micro and on each macro (the cache on the macro is double the one on the micro BSs), given in percentage of stored library. Each curve of the figure corresponds to different values of the parameter α, which characterises the Zipf's distribution. When the percentage of stored library is zero, we are in the case in which no local caching is performed. The growth of the size of the cache generates a reduction of the experienced delay, since more content can be stored locally. This reduction strictly depends on the characteristics of the popularity, i.e., on the parameter α. Indeed, as already mentioned, a large value of α means that there is a small part of the library which is very popular. If this is the case, even a small cache drastically reduces the experienced delay. When α is larger than 1, the experienced delay is reduced up to 50%, if 1% of the library is locally stored. Conversely, a small value of α indicates that the files have similar popularity. In this scenario, larger caches are needed to achieve significant delay reduction: if the popularity distribution is described by the Zipf's function with parameter α equal to 0.56, 10% of the library should be stored to reduce the experienced delay by 30%. The energy consumption increases linearly with the cache size, see Eq. (6.2). Nevertheless, this growth is limited to 13%, when the cache stores 25% of the library and lower than 3% if 5% of the library is cached.



(a)



(b)

Fig. 6.4 Given a fixed caching capacity equal to 20% of the total library, change of its distribution among BSs: (a)Avg delay and energy consumption, (b) hit/miss occurrences probability on micro cell BS.

6.3.2 Impact of the distribution of the cache

Besides the impact of the size of the cache on the user experience, we also investigate the impact of its distribution. In particular, for each macro cell BS and its 4 micro cell BSs, a total capacity equal to 20% of the library is considered, and we vary its distribution among the BSs. We consider the case in which the cache on the macro cell BSs stores 0%, 5%, 10%, 15% and 20% of the library and, correspondingly, each micro cell BSs stores 5%, 3.75%, 2.5%, 1.25% and 0%. In Fig. 6.4a, the average delay (blue bars) and the energy consumption (orange line) are reported, for these values of cache capacity, for α equal to 0.56 and 1.31, when Always ON strategy is used. Fig. 6.4b shows the probability of the possible events that a user might experience when served by a micro cell BS. Blue bins indicate the probability to

experience a *hit - 1 hop*, i.e., is the content is cached on that micro cell BS, the orange bins show the *hit - 2 hops* probability and the green bins report the *miss* probability. As shown by the orange lines in Fig. 6.4a, the energy consumption is constant since the total capacity does not change. Moreover, the plot reveals again that the delay reduction strictly depends on the popularity, i.e., on the parameter α and marginally depends on the cache distribution among BSs. Indeed, as reported in Fig. 6.4b, if α is 0.56 and the micro cell BSs can store up to 5% of the library, no more than 24% of the requests on a micro cell BSs can be satisfied locally (on that BS), while this number grows to 83% if α is 1.31. Similarly, when all the considered caching capacity is put on the macro cell BS, that BS satisfies 43% and 83% of the requests, respectively. Furthermore, even if the hit with a single hop is less frequent due to the reduction of the cache capacity installed on each micro cell BS (see Fig. 6.4b), from Fig. 6.4a, it is evident that putting more cache on each macro cell BS generates the drop of the experienced delay. This is because the cache on each macro is reachable by the users connected to it, as well as by users connected to each corresponding micro cell BS. Therefore, the growth of the capacity on macro cell BSs corresponds to the growth of the cache capacity, which is reachable by all the users. For the same reason, the miss probability decreases, when the capacity on each macro cell BS increases (Fig. 6.4b). Nevertheless, the resources on the macrocell are precious and it is convenient to install some capacity on micro cell BSs too: this relieves the effort on the macrocells and allows to achieve some local (1 hop from users) hits, especially if α is large. Indeed, when α is equal to 1.31 and all the cache capacity is located on the macro cell BSs, the average delay is larger than the case where 15% and 1.25% of the library are stored on the macro and on the micro cell BSs, respectively.

6.4 Impact of Energy Reduction Strategy

We now analyse the effect of the energy reduction strategy on the caching paradigm and vice versa, when Strict RoPE and RoPE are used. As detailed in section 6.2, Strict RoPE deactivates each micro cell BS in case the RAN energy consumption is lower than the produced and stored renewable energy. When this occurs, RoPE gradually switches off micro BSs, until the available renewable energy becomes enough to power the RAN.

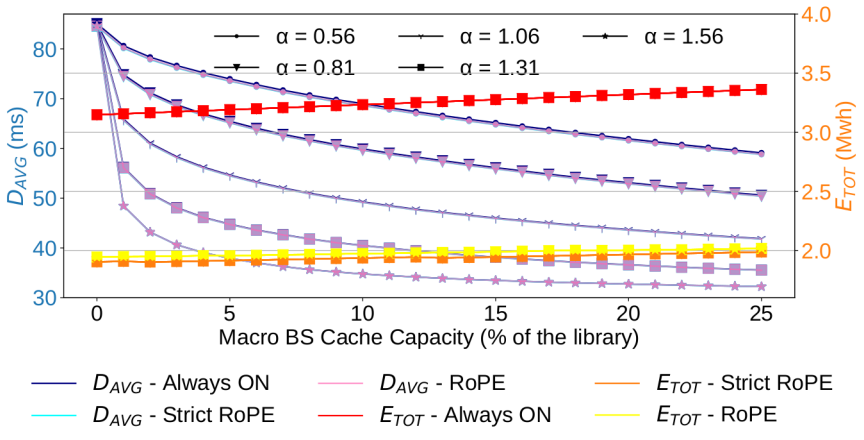


Fig. 6.5 Delay and energy consumption varying the dimension of each cache, for different values of the parameter α, when Always ON, RoPE and Strict RoPE are used.

In Fig. 6.5, the impact of the variation of the size of each cache is shown, in terms of experienced delay (left y-axis) and of energy consumption (right y-axis) in blue, light blue, pink and in red, orange, yellow, when Always ON, Strict RoPE and RoPE strategies are used, respectively. As already discussed, the usage of Strict RoPE and RoPE significantly reduces the energy consumption of the network: when they are employed, the system drops its consumption by up to 41% and 40%, respectively. With Strict RoPE and RoPE, the energy consumption does not grow as much as in the case of Always ON, with the increase of the cache capacity. Indeed, when each macro cell BS stores up to 25% of the library, the network consumes 4% more than the case with no caching servers. This is because with a larger cache, it is more likely that the available renewable energy is insufficient to power the RAN. As a result the switching of micro cell BSs occurs more often and above a given storage size (up to when 5% of the library is stored), the system stabilises: micro cell BSs are deactivated in the same period, since the energy is not sufficient for the network supply in the same instant. Moreover, with Strict RoPE and RoPE, the experienced delay is slightly reduced. This is due to the fact that when the micro cell BSs are deactivated, the users are closer to the content provider, since they are always at 2 hops distance. This is more evident for low values of α, since in these cases the content needs to be taken from the content provider more often. Therefore, the impact of this reduction of distance is higher. The employment of Strict RoPE and RoPE strategies reduce the user coverage, as noticed above. As can be noticed in Fig. 6.6, by varying the size of each cache, the lost traffic grows from 1% given with

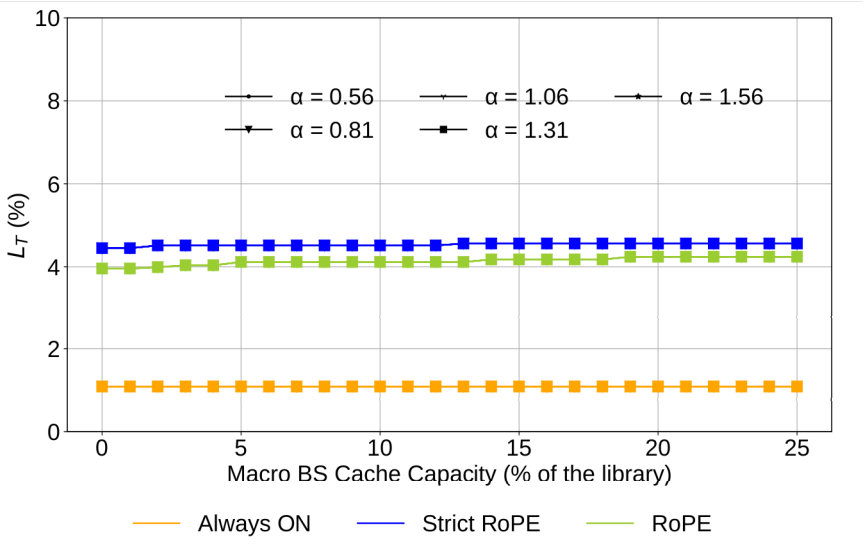


Fig. 6.6 User Coverage varying the size of each cache, for different values of the parameter α, when Always ON, Strict RoPE and RoPE are used.

Always ON, in orange, to, respectively, 4.5% and 4%, in blue and green, which is acceptable. When Strict RoPE and RoPE are used, the growth of the cache capacity does not significantly impact the user coverage, since the number of time slots during which micro cell BSs are deactivated slightly grows by 4%, if 25% of the library is cached on macro BSs, with respect to the scenario that does not employ MEC server.

6.5 RAN Optimisation through User Association

In order to maximise the benefits provided by the MEC technology usage, ensuring also the achievement of the energy efficiency, we design different association policies, which aim at minimising the RAN energy consumption and/or the experienced delay. Once the traffic has been generated, see section 6.2, for each user u, the list of possible BSs to which the user can be associated is created. As depicted in section 6.2, in order to be inserted in the list, a BS b must be active, have enough available capacity and the path loss which the user experiences from that BS b, PL_{b-u}, derived as in section 5.2.2, has to be acceptable, i.e. lower than PL_{MAX}. Our association policies determine the BS of the list the user should be associated with, through a fitness function f. This fitness function f is computed for each BS of the list and

Algorithm 1 Given a user u, select the BS the user is associated with.

Input: User u, BaseStation [] BSs
Output: Association of a user (for the given user, the BS which is associated with)

```
1:  function SELECTION OF THE BS(User, BaseStation[ ])
2:      MAX_f ← -1
3:      BS_MAX_f ← null
4:      for BaseStation b in BSs do                          ▷ Iterate over the BSs
5:          if b is ON and b has enough capacity and PL_b–u ≤ PL_MAX then
6:              f_b ← w_1 · (1 − Ê_C^(t) / E_MAX) + w_2 · (1 − D̂_AVG^(t) / D_MAX)
7:              if MAX_f ≤ f_b then                          ▷ BS b improves the solution
8:                  MAX_f ← f_b
9:                  BS_MAX_f ← b
10:             end if
11:         end if
12:     end for
13:     return BS_MAX_f
14: end function
```

gives a measure of the energy consumption and of the delay experienced by users, assuming that the considered user is associated with that BS. The fitness function is defined as follows:

$$f = w_1 \cdot (1 - \frac{\hat{E}_C^{(t)}}{E_{MAX}}) + w_2 \cdot (1 - \frac{\hat{D}_{AVG}^{(t)}}{D_{MAX}}) \tag{6.4}$$

where w_1 and w_2 are weight factors between 0.0 and 1.0. The value of $\hat{E}_C^{(t)}$, in watt, is the energy consumption of the current solution, E_{MAX}, in watt, is the energy power, consumed by the network when all BSs are active and consuming the maximum energy. $\hat{D}_{AVG}^{(t)}$ is the average experienced delay, in milliseconds, if the current network is used and D_{MAX} is the average experienced delay if each requested content is retrieved in the cloud. Higher values of f mean that the network performs better in terms of power consumption and/or experienced delay. Once this function has been computed for each BS, the one that maximises it is picked. The pseudo code of the selection of the BS is reported in Alg. 1. In case this procedure does not provide any BS for the association of the given user, then the same procedure is repeated using the inactive BS, i.e. removing the first condition in line 5 of Alg. 1. As for the simulations presented in section 6.2, if a "new" BS is activated, we check

Table 6.2 Summary of the different user association strategies.

Strategy	Objective	Macro BS MEC Switching
PL Opt	Min. path loss	No
D_{AVG} Opt	Min. D_{AVG}	No
E_{TOT} Opt	Min. E_{TOT}	No
D_{AVG}-E_{TOT} Opt	Min. D_{AVG} and E_{TOT}	No
PL Opt WS	Min. path loss	Yes
D_{AVG} Opt WS	Min. D_{AVG}	Yes
E_{TOT} Opt WS	Min. E_{TOT}	Yes
D_{AVG}-E_{TOT} Opt WS	Min. D_{AVG} and E_{TOT}	Yes

if it is possible to move users already covered by already active BSs to this "new" one, if the available bit rate is enough, the experienced path loss is acceptable and the value of the fitness function increases, increasing the transmitting power of the BS, if needed and possible. If all users of a certain BS are moved to this "new" BS, that BS is switched off.
The way the weights are set, determines which aspect of the network is optimised while the association procedure is performed:

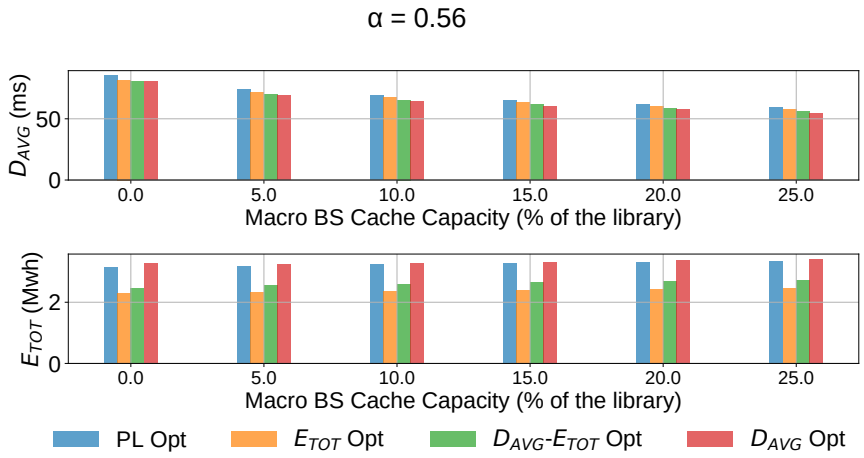
- E_{TOT} *Opt*: the network energy consumption is optimised; to realise this, w_1 is 1 and w_2 is 0;
- D_{AVG} *Opt*: the network is optimised towards the experienced delay, thus w_1 and w_2 are 0 and 1, respectively;
- D_{AVG}-E_{TOT} *Opt*: in this case the optimisation is performed with respect to both the network energy consumption and the experienced delay, w_1 and w_2 are 0.5.
- *PL Opt*: the goal is to select the solution that makes the users suffer the lowest path loss, and w_1 and w_2 are equal to 0. This is the typical association procedure and we use it as a benchmark. This is the case used in simulations discussed in section 6.3.

The usage of the fitness function for the association procedure to minimise the network energy consumption and/or the experienced delay, provides a greedy solution. Indeed, it takes optimal local decisions for each user and the association of a “new” user does not consider the association of the already associated ones, unless

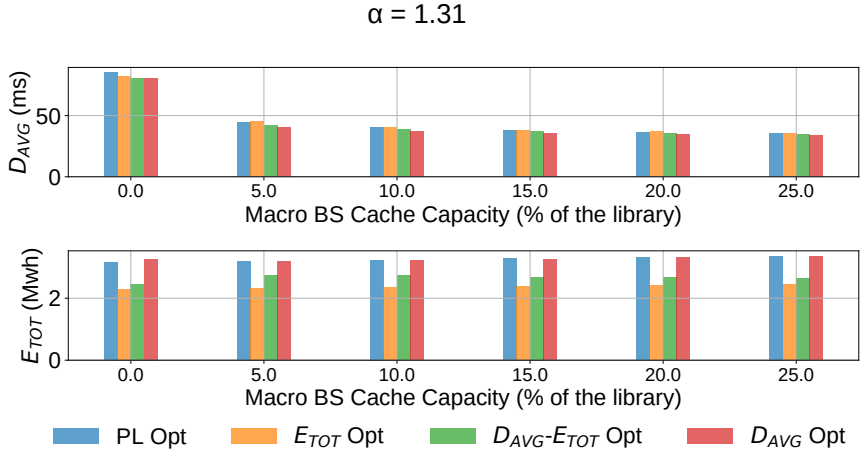
a new BS is activated. This reflects the actual temporal succession of user arrival and the possible handover when a BS is activated, making our approach realistic for the real RAN environment. Moreover, the greedy approach is necessary since the optimisation of the network energy consumption and/or the experienced delay, controlling the BS emitted power and the user association, is an NP-Hard problem, as illustrated in [126]. Nevertheless, the association procedure is performed while the system is operating that means that a solution is needed on the fly. This makes the optimisation approach a not feasible solution.
Finally, we introduce the variant which uses the MEC switching. In this case, the MEC servers, which are installed on each macro BS, are deactivated during the day, from 5 a.m. to 11.00 p.m. and the cache capacity of each macro BS is distributed equally to its 4 micro cell BSs. The described association policies are named as E_{TOT} *Opt WS*, D_{AVG} *Opt WS*, D_{AVG}-E_{TOT} *Opt WS* and *PL Opt WS*. A summary of the used association strategies is given in Table 6.2.

6.5.1 Performance Evaluation of User Association Policies

Now we analyse the effects of these user association policies. Figs. 6.7a and 6.7b show the delay and the energy consumption with the different user association policies, with α equal to 0.56 and 1.31, respectively, increasing the cache capacity from 0% to 25% of the library. First, since the characteristic of the traffic demand, i.e. of the parameter α, has no impact on the network energy consumption, when E_{TOT} Opt is used, variations on the characteristic of the traffic do not impact the associations and, as a consequence, the network energy consumption, as can be noticed by the orange bars in the two figures. Moreover, when compared to the case in which the experienced path loss is minimised, between 22% and 27% of network energy consumption reduction is achieved. Using this users association procedure, users are typically associated with macro BSs, meaning that having a few but heavily loaded macro BSs is more efficient than having many active BSs under utilised. This also shortens the time needed to reach the cloud and because of this, when retrieving contents in the cloud occurs often, i.e. α is lower than 1 and/or the server size is small, a slight delay reduction, between 2.5% and 4%, with respect to PL Opt policy, is achieved.
When users are associated while minimising the experienced delay, i.e., D_{AVG} Opt is used, the delay drops up to 10% (see red bars in Fig. 6.7) with respect to the case with



(a)



(b)

Fig. 6.7 Energy consumption and delay achieved with the proposed users association policies, with (a) α=0.56 and (b) α=1.31.

the same cache capacity, but where PL Opt is used as the association approach. This drop is more significant when misses are more likely, i.e. for decreasing cache size and α parameter, because performance can be improved significantly. Moreover, in order to make the access to the cloud faster, this policy forces users to be associated with the macro BS, which maximises the local hit. This is also because if a user is associated with a macro BS, the cloud is reached with 2 hops, while if associated with a micro BS, 3 hops are necessary to reach it. Thus, being associated with a macro makes the access to the cloud faster. Because of this trend, to achieve a path loss low enough to receive an acceptable quality of the signal, the output power of a BS results higher than when PL Opt is used. This determines the growth of

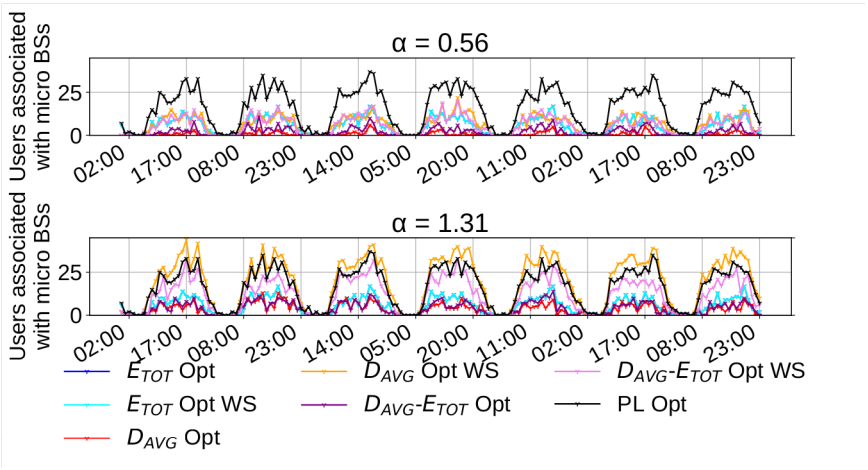


Fig. 6.8 Number of users who are associated to micro BSs, the cache capacity on each macro is 10% of the library.

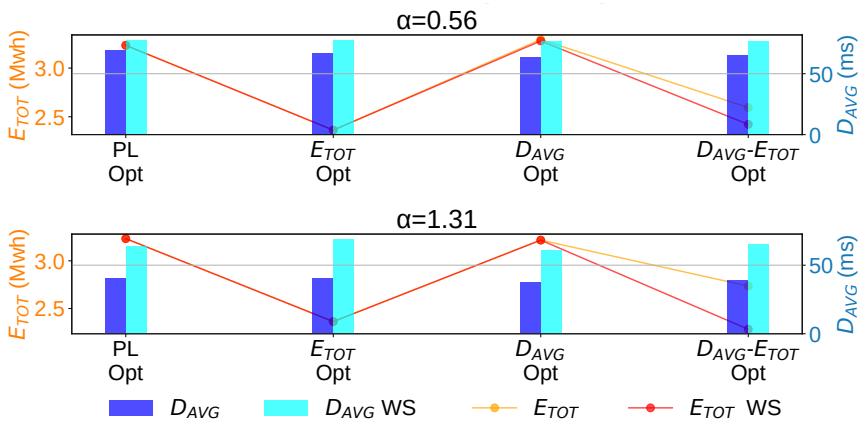


Fig. 6.9 Energy Consumption and Delay with different user association policies, if the cache capacity on each macro is 10% of the library.

the network energy consumption by 4%. When α is larger than 1, this policy is not particularly effective, as already mentioned, making the energy consumption and the delay almost unchanged. In case the association procedure aims at the joint minimisation of delay and energy consumption, both of them are reduced by 5% and 20%, respectively.

6.5.2 Effects of the MEC switching

We have observed that the proposed association procedures tend to associate users to macro BSs, but this violates the load balancing. Fig. 6.8 shows the number of users, that are associated with a micro cell BS, in each time slot of the simulation. This value is shown for α equal to 0.56, on the left and to 1.31, on the right, assuming that each cache, which is installed on each macro cell BS, stores up to 10% of the library. We consider all the association policies, with and without variant which switches the MEC servers installed on the macro BSs. As can be observed from the figure, if users are associated according to D_{AVG} Opt, in blue, D_{AVG}-E_{TOT} Opt, in purple, and E_{TOT} Opt, in red, users associated with micro cell BSs are 0, 1 and 6, respectively, against 19, obtained when PL Opt is used. As shown in [118], this increases the electromagnetic exposure of human beings, besides worsening the QoS, as claimed in [127, 128]. To solve this issue, we introduce the variant, which deactivates the MEC server of each macro BSs, between 5 a.m. and 11 p.m., distributing equally that amount of capacity from each macro cell BSs to its 4 micro cell BSs). Therefore, in this time interval, the MEC capacity on each micro cell is a quarter of the MEC capacity of its macro cell BS, in addition to the capacity of its MEC server, while no cache capacity operates on macro BSs. Fig. 6.9 shows the delay, for each association policies and for α equal to 0.56, on the left and to 1.31, on the right, assuming that each cache on each macro stores up to 10% of the library, with the light blue and blue bins, when MEC switching variant is used or not, respectively. The network energy consumption is given by the red and orange lines, respectively.
First, we focus our attention on the cases that associate users according to D_{AVG} Opt or D_{AVG}-E_{TOT} Opt policies. In these cases, users tend to be associated with micro cell BSs during the day (from 5.00 a.m. to 11.00 p.m.), to minimise the delay. As a consequence, when α is 0.56, the WS variant slightly increases the number of users served by the micro BSs from 0 with D_{AVG} Opt to 11 with D_{AVG} Opt WS and from 1 with D_{AVG}-E_{TOT} Opt to 7 with D_{AVG}-E_{TOT} Opt WS. With α equal to 1.31, this growth is more evident: from 4 to 22 and from 3 to 9, if D_{AVG} Opt, D_{AVG} Opt WS and D_{AVG}-E_{TOT} Opt and D_{AVG}-E_{TOT} Opt WS are used, respectively. With low values of α this improvement is less evident. This is because in these cases access to the cloud is frequently needed. For this reason, users continue to be associated with macro BSs to reduce the time to retrieve contents, despite the WS variant employment. With large values of α, several local hits occur. Hence, the WS variant induces the association of users with micro BSs and this increases the network energy consumption, which grows up to 16% (see Fig. 6.9). As regards the experienced

delay, it grows by up to 68%. This is because, since the cache capacity is moved from each macro BS towards its micro cell BSs, less cache is reachable by users. In addition, since users tend to be associated with micro cell BSs, they are further from the cloud (3 hops) than if they are associated with the macro BS (2 hops), making the access to the cloud slower. When E_{TOT} Opt is employed as association policy, the usage of the WS variant does not improve the load-balancing issue, since it is independent of the MEC server presence and capacity. As a consequence, the WS does not impact the energy consumption of the PL Opt, as well as of the E_{TOT} Opt. Nevertheless, the delay grows by up to 70%, in case the WS variant is used. As already mentioned, removing cache capacity on the macro and spreading it among its micro cell BSs, deteriorates the delay, as less capacity is reachable from users.

6.6 Final Remarks

In this chapter, the delay reduction and the network energy efficiency are revised and discussed. A portion of an heterogeneous hierarchical RAN is considered, where the MEC technology is employed, to push the most popular contents closer to users so as to reduce latency. The considered RAN is powered by a PV panel system and an energy battery and is connected to the power grid. Different users association policies are proposed in order to further improve the experienced delay and the energy consumption of the network. We notice that, even if strictly dependent on the characteristics of the traffic and on the server capacity, the MEC technology reduces the experienced delay up to 60%, without generating significant growth of the network energy consumption, limited to 7%. In addition, the employment of an energy reduction strategy, applied in case of renewable energy shortage, reduces the energy consumption but does not impact the experienced delay. The proposed users association policies effectively provide reductions of delay and power consumption. Our results lead to three conclusions. First, caching at the edge and dynamic activation of the BSs, can be very effective in reducing latency and reducing the network power consumption, without deteriorating their performances because of their coexistence. Second, caching on the macro BSs is always needed to significantly reduce delays, while caching also on the micro cells relieves the effort on the macro cell. Finally, association procedures which minimise the delay and/or the energy consumption tend to associate users with macro BSs. In this way, when the network

energy consumption is optimised, also the experienced delay is slightly reduced. Meanwhile, in case the association procedure is based on the minimisation of the delay, the energy consumption increases, since users can be associated to BS which are far, and the emitted power has to be increased so as to receive the signal with the adequate quality.

Chapter 7

Evaluation of flying caching servers in UAV-BS based realistic environment

Part of the work presented in this chapter has already been published in:

- Castellanos, G., Vallero, G., Deruyck, M., Martens, L., Meo, M., & Joseph, W. (2021). Evaluation of flying caching servers in UAV-BS based realistic environment. *Vehicular Communications*, 32, 100390.

UAV-BSs are drones on which BS equipment is mounted. This is possible because of the advance of the structure of the most recent drones, which makes them able to carry up heavy payloads. Up to some years ago, no more than 600 g payload could be brought, while the most recent drones can transport up to 8 kg, much more than the 2 kg weight needed to bring the network equipment [129]. This has contributed to consider the UAV-BSs one of the possible key actors to dynamically and rapidly deploy communication facilities, where traditional ground network infrastructures are not feasible or cost-effective. These UAV-BSs fly where users, who are suffering from low-quality of service, are located, to bring them connectivity. Users are connected to the UAV-BSs through access links, while UAV-BSs are connected to the CN, establishing BH links between them and an AP. In 5G and beyond 5G systems, UAV-BSs are envisioned as support for the ground RAN, in case of terrestrial network

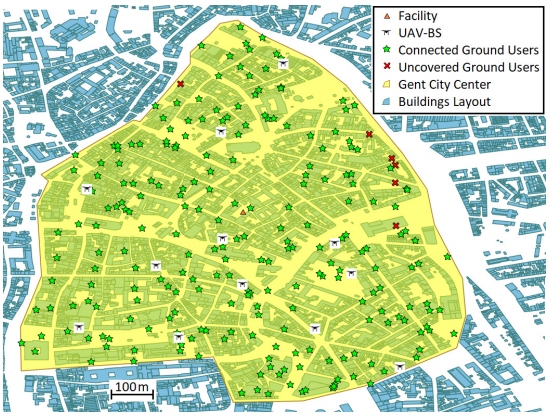


Fig. 7.1 Scenario for the city centre of Ghent with 250 users

failures due to network overloads or physical unavailability, while in 6G networks, UAV-BSs, beside being employed to provide additional radio coverage, are also thought as content providers and computing servers, in order to bring these platforms in proximity to users and meet the strict delay requirement of some applications.
In this chapter, the potential of mounting BSs on drones to offer a solution for crowded scenarios where the ground RAN is overwhelmed is considered. This scenario is detailed and modelled in sections 7.1 and 7.2, respectively. Then, the performance of the support of UAV-BSs is evaluated, through an ad-hoc simulator, presented in section 7.3. Results, shown in section 7.4, reveal that, because of the scarce on-board energy availability on UAV-BSs, which is provided by on-board batteries, UAV-BSs need frequent replacement because of the lack of energy. In addition, simulations show that the BH network often saturates, due to its low available bandwidth, which limits its capacity and significantly deteriorates the network QoS. To cope with this issue, we propose to resort to the MEC paradigm. In particular, MEC servers which provide caching are installed on each UAV-BS, in order to store popular contents and reduce the usage of the BH network. Section 7.5 shows the improvement achieved when MEC servers are used; in section 7.6, the final remarks of the chapter are drawn.

7.1 Scenario

In this part of the work, we consider a crowded environment, located in an area of

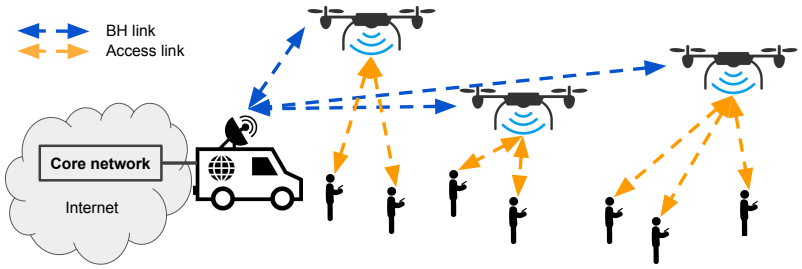


Fig. 7.2 Network structure

1 km^2, delimited by the orange shape in Fig.7.1, situated in the city centre of Ghent, in Belgium. We assume that the typical yearly festival is taking place. The total event duration is 6 hours, from 7 p.m. to 1 a.m.. We assume that the users in this scenario are mostly static since typically in situations such as festivals and sport events, they either maintain their position or move at a very slow pace.
We assume that, in order to support the user communication requirements in a crowded scenario, like the festival we are considering, an UAV-BS aided network is employed. Similarly to [3, 4, 91], the network is structured as depicted in Fig.7.2. The ground users, green stars in Fig. 7.1 and human icons in Fig.7.2, access the communication service through the access network (orange links in Fig.7.2), provided by UAV-BSs, acting as traditional BSs, indicated by the drone icon in Fig. 7.1. These UAV-BSs reach the CN, by BH links (blue links in Fig.7.2). Each BS equipment is mounted on a quadcopter MD4-1000, which can carry wireless communication equipment payload, which accounts for less than 2 kg, according to [129]. These drones have a maximum speed of 45 km/h powered by a 22.2 V battery with a capacity of 17.33 Ah [130]. As described in [4], the connection between the UAV-BSs and the CN for a fast deployable network is given by an antenna on a truck or on a crane to increase its height, as shown in Fig.7.2, indicated by the orange triangle in Fig. 7.1. This facility can provide up to 25 simultaneous drones, and up to 500 battery packages to rapidly replace drones that are running out of energy, allowing the system to maintain up to 25 simultaneous UAV-BSs with nearly 20 takeovers per drone. In the considered network, the height of this antenna is 60 m to avoid the majority of the building heights in the centre of Ghent [4].

Table 7.1 Link budget parameters for simulation [3, 4]

Parameters	Access	BH
Frequency	2.6 GHz	3.5 GHz
Bandwidth	5 MHz	20 MHz
Max Tx Power	33 dBm	43 dBm
Antenna Gain	4 dBi	5 dBi
Antenna Height	1.5 m	60 m
Noise Figure	8 dB	5 dB
Fade Margin	10 dB	1 dB0
Shadowing Margin	8.2 dB	8.2 dB
Interference Margin	2 dB	0 dB
MCS and SNR	1/3 QPSK = -1.5 dB 1/2 QPSK = 3 dB 2/3 QPSK = 10.5 dB 1/2 16-QAM = 14 dB 2/3 16-QAM = 19 dB 1/2 64-QAM = 23 dB 2/3 64-QAM = 29.4 dB	

LTE release 14 working on the 3.5 GHz band is used for the BH links with a bandwidth of 20 MHz consisting of 100 RBs [131, 132]. The access network uses a 2.6 GHz LTE femtocell based technology. Its details are discussed in chapter 6. The link budget for the access and BH network is reported in Table 7.1.

7.2 System Model

In this section we provide details about the model of our scenario. We assume that time is slotted in T time intervals, 5 minutes long. The used notation is reported in Table 7.2.

7.2.1 Traffic Demand Modelling

In order to model the traffic demand during public events, such as football matches, concerts, we use data presented in chapters 3 and 4. These data report the traffic demand volume, in bits, of many BSs located in the city of Milan (Italy) and in a wide area around it, for two months in 2015, with granularity of 15 minutes. In this

Table 7.2 Summary of the system model notation.

Notation	Definition
N	Number of UAV-BS in the fleet
U	Number of users
T	Number of time slots
C	Maximum number of cached contents
B_a	Maximum access capacity of each UAV-BS
B_{BH}	Maximum BH network capacity
$b_{a,n}^{(t)}$	Traffic volume carried by UAV-BS n in the access network
$b_{BH,n}^{(t)}$	Traffic volume carried by UAV-BS n in the BH network
$\mathscr{F}$	Set of contents in the library
F	Number of contents in the library
S	Size of each content
$P_{\mathscr{F},n}(f)$	PDF of the content popularity of UAV-BS n
α	Parameter of the Zipf's distribution
p_{hit}	Hit probability
p_{miss}	Miss probability
$l_u^{(t)}$	Traffic volume requested by user u at time t
$P(t)$	The network power consumption at time t
$P_{a,n}^{(t)}$	The access power consumption of UAV-BS n, at time t
$P_{BH,n}^{(t)}$	The BH power consumption of UAV-BS n, at time t
$P_{c,n}^{(t)}$	The caching power consumption of UAV-BS n, at time t
D	Average transmission delay
D_{hit}	Time needed to send a content from an UAV-BS to a user
D_{miss}	Time needed to send a content from the cloud to a user

part of the work, we update the value of the traffic based on the average increment, equal to 7.5 times, of the mobile market in Italy from 2015 to nowadays, as reported in [133]. Then, we transform the traffic volume in bit rate, normalising the values by the granularity of samples, i.e., 15 minutes. Then, we select the trace corresponding to the BS which covers the soccer stadium of San Siro, which typically hosts public events, and we investigate its traffic profile. The black line in Fig.7.3a is the mean bit rate in each 15-minutes long time interval, computed using the whole traffic trace, while each coloured curve shows the bit rate, in Mbps, of each match, that took place there during the Spring of 2015. From the figure, it is visible that the traffic demand significantly increases during matches with respect to the average traffic demand (black line in Fig.7.3a). Then, we also notice that the bit rate during each event

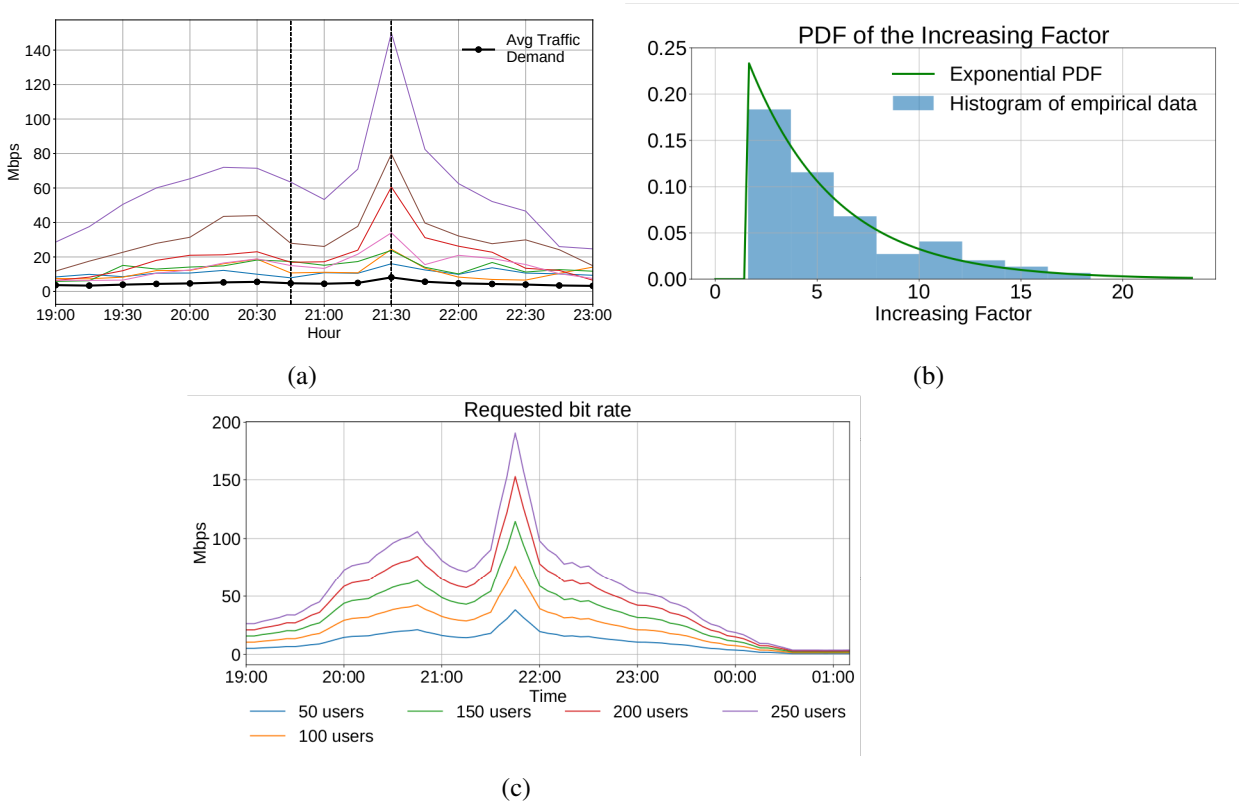


Fig. 7.3 Traffic Analysis, Modelling and Generation: (a) Traffic demand, in bits, during public events at the San Siro soccer stadium, (b) PDF of the Increasing Factor, (c) Requested capacity for an average value of 0.24 for different users scenarios. U:Users

typically starts growing prior to the beginning of a match, indicated by the vertical black line in the figure. Shortly before the kickoff, which starts in correspondence to the second vertical black line, it slightly decreases and when the intermission starts, the traffic demand always presents the peak, decreasing again as soon as the match resumes.

To shape this behaviour, we assume that the bit rate $\beta^{(t)}$ which is requested during these special events, at time t, is given by:

$$\beta^{(t)} = B_{AVG} \cdot U^{(t)} \cdot K^{(t)} \tag{7.1}$$

where B_{AVG} is the average required bit rate, in Mbps, given by the average of each point of the black circle marked line in Fig.7.3a; $U^{(t)}$ is a shape function, used to model the trend of the bit rate during each event and $K^{(t)}$ is a scalar, which scales the average traffic demand. In order to build the $U^{(t)}$ function, we normalise the

bit rate trace corresponding to each match, reported in Fig.7.3a, by its peak. Then, we average these normalised bit rates, time interval by time interval, to obtain its average shape during a public event. In the end, in order to determine the value of $K^{(t)}$, we collect the increasing factors for each bit rate sample, with respect to the average. These increasing factors are computed as $\beta^{(t)}/\beta^{(t)}_{AVG}$, where $\beta^{(t)}_{AVG}$ is the the th–samples of the black curve in Fig. 7.3a. We build the Probability Density Function (PDF) of these increasing factors which is reported in Fig.7.3b, where in blue there is the histogram plot of these data, while the green line is the PDF of an exponential distribution that best fits it, whose parameter is λ=0.24. As a result, assuming that the number of users remains constant during the event, the bit rate of each user u, $\lambda^{(t)}_u$, from which it is possible to derive the traffic demand volume $l^{(t)}_u$ discussed in section 7.2, is as follows:

$$\lambda^{(t)}_u = \Lambda_{Avg} \cdot U^{(t)} \cdot K^{(t)} \tag{7.2}$$

where $U^{(t)}$ is the shape function, computed as described above, and $K^{(t)}$ is a sample extracted from an exponential distribution. Λ_{Avg} is the average bit rate, assumed equal to 0.225 Mbps, which is acquired empirically from the data provided by the MANO, assuming that there are 250 users. As a result, the total required bit rate is shown in Fig.7.3c, where each curve corresponds to a different number of users U. Observe that the shape of each curve well fits the typical behaviour which characterises the public events (see Fig.7.3a). The requested traffic presents a peak of 235 Mbps, in the scenario where there are 250 users, as can be noticed in Fig.7.3c.

7.2.2 Network Modelling

In order to model the architecture depicted in Fig.7.2, a fleet of N_{UAV-BS} UAV-BSs, acting as traditional BSs is used. The UAV-BSs provide the communication service to the *U* ground users. Each ground user u access this service through the access network (orange links in Fig.7.2), provided by the UAV-BSs. The UAV-BSs reach the CN through BH links (blue links in Fig.7.2). Each UAV-BS $n \in [1, N_{UAV-BS}]$ has a certain access capacity, i.e., in the access network, it can deliver up to $B_A \geq 0$ data volume, in bits, within each time interval t. As regards the BH network, $B_{BH} \geq 0$ is the maximum traffic volume which can be carried in the BH network, within each time slot t. This means that, given $T^{(b,t)}_A$ and $T^{(b,t)}_{BH}$ the traffic volume carried by

Table 7.3 Values of the parameters of the consumption model for the access and BH network.

Network	N_{sec}	P_{rect}	P_{mwl}	P_{airco}	n_{TX}	P_{amp}	P_{trans}	P_{dsp}
Access/BH	1	0	0	0	1	12.8	1.8	7.9

Table 7.4 Summary of the channel model notation.

Notation	Definition
$\mathcal{N}$	Normal Distribution for the excess path loss for the access network
μ	Mean of $\mathcal{N}$
σ	Standard Deviation of $\mathcal{N}$
ϑ	Elevation angle between users and UAV-BS
a	Frequency parameter for access PL model
b	Environmental parameter for access PL model
d_{3D}	3D distance between TBS and UAV-BS
f_c	Central frequency of the BH PL model
h_{TBS}	Height of the Terrestrial BS

each UAV-BS b in the access and BH network at time t, respectively, the following conditions have to be satisfied:

$$T_A^{(b,t)} \leq B_A \qquad \forall b \in [1, N_{UAV-BS}], B_A \geq 0, T_A^{(b,t)} \geq 0 \tag{7.3}$$

$$\sum_{n=1}^{N} T_{BH}^{(b,t)} \leq B_{BH} \qquad B_{BH} \geq 0, T_{BH}^{(b,t)} \geq 0 \tag{7.4}$$

Since the traffic demand of each user is transmitted in the access, as well as in the BH networks, we have $T_A^{(b,t)}$=$T_{BH}^{(b,t)}$.

7.2.3 Path Loss and Channel Modelling

Several path loss models are used to evaluate the aerial to ground communications as described in [134]. For the access network link, we consider the aerial to ground model from Al-Hourani et al. [135], which offers a statistical path loss model for sub-6GHz bands. It is based on the Free Space Path Loss (FSPL) model using ray tracing for urban and suburban environments accounting for the effects of buildings

into the model. The Al-Hourani path loss model is modelled as :

$$PL_A = FSPL + \mathcal{N}(\mu, \sigma^2) \tag{7.5}$$

where FSPL is the Free Space Path Loss model from the Friss equation. $\mathcal{N}$ is the normal distribution for the excess path loss with a mean of μ, and standard deviation σ as:

$$\sigma = a \cdot e^{(-b \cdot \vartheta)} \tag{7.6}$$

where a and b are frequency and environment-dependent variables and ϑ is the elevation angle between the ground and the aerial nodes. This model is suitable for urban environments with building representation. To account for that, we use the model in the 2.6 GHz band setting the μ, a and b values to 1.67, 8.59 and 0.04 for the LoS model, while the values for the NLoS are 18, 26.53 and 0.003 respectively. For the BH network we consider the TR 36.777 path loss model since it is more suitable for the communication between the terrestrial BSs and aerial nodes in the 3.5GHz band [136]. This model is suitable for antennas higher than 22 m and distances up to 4 km. The LoS path loss model is computed as:

$$PL_{BH_{LoS}} = 28 + 22\log_{10}(d_{3D}) + 20\log_{10}(f_c) \tag{7.7}$$

where d_{3D} is the three-dimensional distance between the terrestrial BS and the UABS and f_c is the central transmission frequency.
The NLoS path loss model is described as:

$$PL_{BH_{NLoS}} = \left[46 - 7\log_{10}(h_{TBS})\right]\log_{10}(d_{3D}) + 20\log_{10}\left(\frac{40\pi f_c}{3}\right) - 17.5 \tag{7.8}$$

where h_{TBS} is the height of the terrestrial BS. Detailed parameters of the link budget that influence the channel model are described in Table 7.1.

7.2.4 Power Consumption Modelling

The network power consumption at time t is the total power consumption of the wireless network, considering the power needed to provide the access and BH

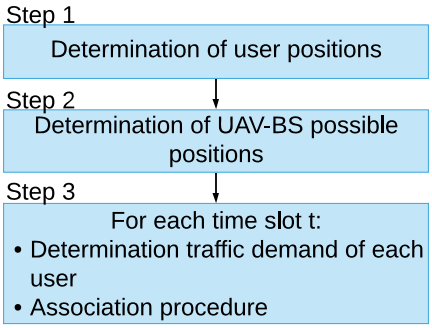


Fig. 7.4 Steps of each simulation.

features in the network, at time t:

$$E_{TOT}^{(t)} = \sum_{b=1}^{N_{UAV-BS}} \left(E_{A}^{(b,t)} + E_{BH}^{(b,t)} \right) \tag{7.9}$$

where $E_{A}^{(b,t)}$ and $E_{BH}^{(b,t)}$ are the access and BH energy consumption of each UAV-BS *b* during time slot *t*. If UAV-BS n is not active, $E_{A}^{(b,t)}$ and $E_{BH}^{(b,t)}$ are zero. The LTE energy consumption models, in watt hour, for the access and BH networks, for each UAV-BS b are as in Eq. (5.2), with parameters reported in Table 7.3. In the computation, the power needed by an UAV-BS for flying is not considered.

7.3 Simulation

We evaluate the detailed scenario through simulations. The employed tool is based on the one implemented and discussed in chapter 6. We assume that each simulation covers an operating time of 6 hours, from 7:00 p.m. to 01:00 a.m.. Time is slotted in 5 minutes long time intervals.
In Fig.7.4 the structure of the simulations is depicted. Before the actual simulation begins, we first determine the coordinate of the location of each user u in the yellow area depicted in Fig.7.1. As mentioned above, differently to the scenario depicted in chapter 6, the position of each user is assumed static, meaning that it remains the same for the whole simulation, i.e. for the whole duration of the public event. This is a realistic assumption for the considered overcrowded scenario, where users maintain their position or move at a very slow pace and for a very small space. Because of this

static behaviour of the ground users, handovers and frequency allocation analysis are not considered. As in chapter 6, these coordinates are assumed to be uniformly distributed over the considered area, since in this area, attractions are uniformly distributed, resulting in an uniform distribution of users, meaning that each location has the same chance to be chosen as a possible location for the user. In order to place the UAV-BSs, a set of possible locations for the N_{UAV-BS} available drones is created. This set contains a list of possible aerial positions, where UAV-BS could hover. There are as many UAV-BS positions as ground users. Each grid point uses the same $(x^{(b)}, y^{(b)})$ coordinates of users while $(z^{(b)})$ is set to 80 m or above, if a building is present.
For each time slot t, the bit rate required by each user is determined, using the model described in section 7.2.1. Once the position of each user has been generated, he/she is associated with the UAV-BS from which he/she experiences the lowest path loss, provided that that UAV-BS has enough available capacity in both the access and BH network, to serve him/her. Details of each step are given below.

7.3.1 Dynamic traffic generation

At the beginning of each time slot t, $\lambda^{(u,t)}$, the bit rate of each user u, is determined. To do this, we use the model of Eq. (7.2). From the bit rate $\lambda^{(u,t)}$, the traffic demand volume $l^{(u,t)}$ is derived.

7.3.2 Network Generation

In this step of the simulation, each user is associated with an UAV-BS, if possible. As in chapter 6, the list of UAV-BS is generated, to which the user u can be connected. This list contains the active UAV-BSs, which provide the requested traffic demand in both the access and BH network and from which the experienced path loss is lower than an allowable maximum, in order to guarantee to the user to receive the signal with a sufficient quality. The model of (7.5) is employed for the path loss computation and if this path loss is greater than the allowable maximum, the input power of the UAV-BS is increased until it becomes acceptable. In case the input power reaches the maximum allowable input power, but the path loss is still larger than the maximum, that UAV-BS is not inserted in the list.
A user can be associated with a UAV-BS if that UAV-BS provides the required bit

rate in the access as well as in the BH network. The user traffic demand in the access and in the BH network is $l^{(u,t)}$, or $\lambda^{(u,t)}$ if expressed in bit rate. This means that an UAV-BS b is able to provide the required traffic demand in the access network, if the following condition is verified:

$$l^{(u,t)} + \sum_{v \in \mathscr{U}^{(b,t)}} l^{(v,t)} \leq B_A \tag{7.10}$$

where $\mathscr{U}^{(b,t)}$ is the set of user already associated with UAV-BS b, $l^{(v,t)}$ is their traffic demand, B_A is the maximum access volume, equal to $5.07 \cdot 10^3$ Mb, assuming that its maximum bit rate is 16.9 Mbps and each time slot 5 minutes long. As regards the BH network, an UAV-BS b is able to carry the needed traffic demand at time t in the BH network, if following condition is verified:

$$l^{(u,t)} + \sum_{\nu=1}^{N_{UAV-BS}} b_{BH}^{(\nu,t)} \leq B_{BH} \tag{7.11}$$

where N_{UAV-BS} is the number of UAV-BSs, B_{BH} is the maximum BH traffic volume in a time slot 5 minutes long, set equal to $21.6 \cdot 10^3$ Mb, assuming that its maximum bit rate is 72 Mbps. Then, $b_{BH}^{(\nu,t)}$ is the sum of the traffic demand of the set of users $\mathscr{U}^{(\nu,t)}$ associated with UAV-BS ν during the time slot t:

$$b_{BH}^{(\nu,t)} = \sum_{v \in \mathscr{U}^{(\nu,t)}} l^{(v,t)} \tag{7.12}$$

where $l^{(v,t)}$ is the traffic demand at time t of each user v, as above.
Summarising, if an UAV-BS transmits and the user receives the signal with acceptable quality, i.e. with acceptable path loss, and UAV-BS satisfies (7.10) and (7.11), that UAV-BS is inserted in the list. The user is associated with the UAV-BS from which he/she experiences the lowest path loss among the UAV-BSs in the list.
In case the list results empty, the same procedure is performed using only the inactive UAV-BSs. In this case, the path loss experienced by the considered user in the access network has to be acceptable, as well as the path loss in the BH network, experienced by the considered drone, computed using the models reported in Eq. (7.7) and (7.8). In case also this procedure provides an empty list, the user remains uncovered.

7.3.3 Key Performance Indicators

The performances of the proposed methodologies are evaluated using the KPIs described below.

Capacity

The capacity of the network is defined as follows:

- Access Capacity: This is the total used access capacity of the network in a time slot, measured in Mbps, indicated with $b_A^{(t)}$.
- BH Capacity: This provides the total used BH capacity, $b_{BH}^{(t)}$, during a time slot, in the network, in Mbps.
- Number of Used Resource Blocks: This is the total number of used RBs in the BH network, in each time slot. Its notation is $RBs^{(t)}$.

We also measure the average access and the BH capacity, as well as the number of RBs which are used on average in each simulation. In order to distinguish from the previous KPIs, we name them *Average Access Capacity*, b_A, *Average BH Capacity*, b_{BH}, computed as $\frac{1}{H}\sum_{t=0}^{H} b_A^{(t)}$ and $\frac{1}{H}\sum_{t=0}^{H} b_A^{(t)}$, respectively, where H is the duration of the simulation. The *Average Number of RBs*, RBs, is $\frac{1}{H}\sum_{t=0}^{H} RBs^{(t)}$.

Lost Traffic

As in previous chapters, the Lost Traffic measures the percentage of traffic which can not be handled by the network, since there is not enough available capacity in the access and/or in the BH network and/or the experienced path loss is unacceptable. It is denoted as $L_T^{(t)}$, if it is measured in a single time slot, while as L_T, if it is computed as $L_T=\frac{1}{H}\sum_{t=0}^{H} L_T^{(t)}$, where H is the number of time slots in the simulation.

Energy Consumption

As in previous chapters, the Energy Consumption, E_{TOT}, is the energy consumption of the wireless network. In this part of the work, it accounts for the energy needed

for the access and BH network supply:

$$E_{TOT} = \sum_{t=1}^{H} E_{TOT}^{(t)} \tag{7.13}$$

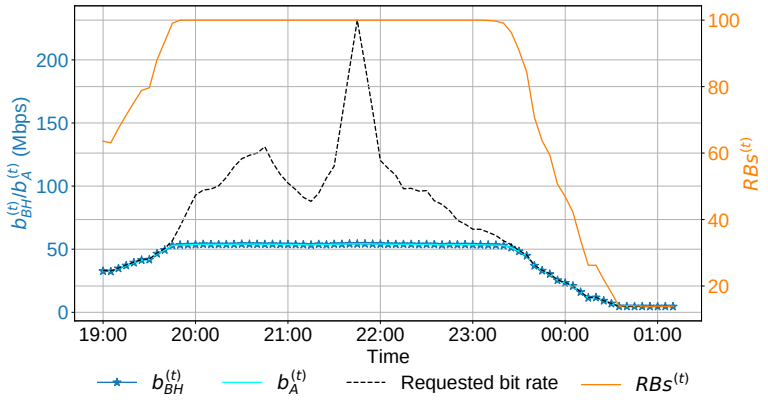
where $E_{TOT}^{(t)}$ is the power consumption in time slot t, computed as in (7.9), which sums the power needed by each UAV-BS to provide the access and BH network supply. H is the duration of the simulation.
We also measure the average energy consumption E_{AVG}, which is the average energy consumption of each active UAV-BS, given by $\sum_{t=1}^{H} \frac{E_{TOT}^{(t)}}{N_{UAV-BS}^{(t)}}$, where $N_{UAV-BS}^{(t)}$ is the number of active UAV-BSs at time slot t.

UAV-BS

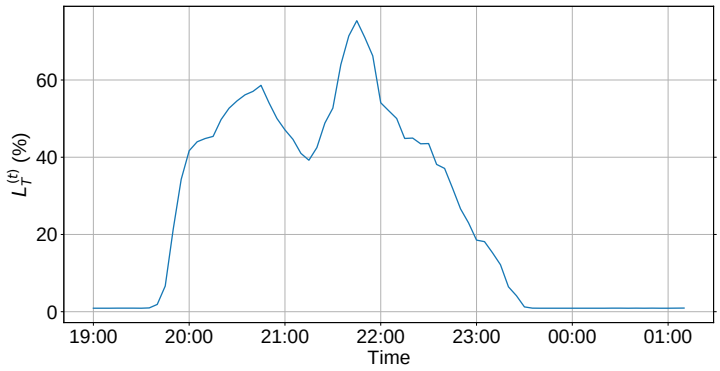
The UAV-BS provides the number of average UAV-BS locations needed during each simulation, given by $\frac{1}{H}\sum_{t=1}^{H} N_{UAV-BS}^{(t)}$.

7.4 Performance Evaluation

In this section the scenario previously presented is evaluated, assuming a user density equal to 250 user/km^2. In Fig. 7.5a, on the left y-axis, the requested bit rate, the provided access and BH capacity are plotted in black, blue and cyan curves, in each time interval of the considered operating time. On the right y-axis, the number of used RBs in the BH network is provided. From this figure, it is possible to notice that, until 19:50, the used access and BH capacity, see the blue and cyan curves, respond instantaneously to the required bit rate. Then, at 19:50, the used access and BH capacity become flat and are unable to provide the required bit rate (black curve in Fig.7.5a). This is because the available RBs are totally used (orange curve in Fig.7.5a), which means that the BH network is saturated. After the peak, the traffic demand decreases. After 23:10, traffic demand is so low that not all the BH RBs' are needed. This means that the network is not saturated anymore and can provide the required bit rate. This impacts the lost traffic, plotted in Fig. 7.5b, which is almost zero until 19:50. Then, because of the BH network saturation, it starts growing, up to 75%, at 21:45, when the traffic demand peak occurs. After this, since the traffic



(a)



(b)

Fig. 7.5 (a) Used capacity and (b) Lost Traffic $L_T^{(t)}$.

demand drops and the BH network is not saturated anymore, it decreases, as well.

Table 7.5 summarises the simulation results. It reports the total access and BH capacity, b_A and b_{BH}, respectively, as the average energy consumption E_{AVG} and the average number of UAV-BS locations. The table indicates that, in order to cover the considered area, 10.6 UAV-BS locations are needed, each consuming on average 8.8 Wh, as indicated by the E_{AVG} column, resulting in a total consumption of 90.8 Wh. This means that the flight duration of each drone is 69 minutes, meaning that 5 drones per each locations are needed in our scenario.

These results highlight two issues which need to be addressed, in order to make this solution effective to support RANs. First, the BH network saturation significantly deteriorates the QoS, since it limits the user coverage. Then, the scarce on-board energy availability that is provided to UAV-BSs by on-board batteries has to be faced. Indeed, each of them needs to be replaced 6 times, assuming a simulation time

Table 7.5 Simulation results.

b_{BH}/b_A	RBs	L_T	E_{AVG}	E_{TOT}	UAV-BS
41.85 Mbps	79	41.9%	0.7 kWh	6.8 kWh	10.6

equal to 6 hours. In the next session, we employ the MEC technology to improve the network capacity and consequently the QoS, while in chapter 8, solar-powered UAV-BSs are discussed, which result in a longer UAV-BS survival.

7.5 QoS improvement through MEC technology

From the previous analysis, it is evident that the BH network is the most challenging part of our scenario. As discussed, the average lost traffic is more than 25% because of the capacity constraints of the BH links. In this part of the work, to address this issue, the MEC technology is considered. As described in chapter 6, it pushes storage and computing platforms at the edge of the network. In this way, in case a user requires a content which is stored in the MEC server of its UAV-BS, that UAV-BS directly transmits that content to that user. If this is the case, the access to the cloud is unneeded and the BH network is not used to satisfy that request. In the following section the model and the simulator used for the evaluation of the proposed scenario are detailed.

7.5.1 MEC-enabled System Model

We assume that each UAV-BS b is equipped with a MEC server, that provides the storage capability. Similarly to chapter 6, the server of each UAB-BS b updates its cached contents according to the LFU cache algorithm, to store the C most popular contents. A library $\mathscr{F}$={1,2,..,F}, composed by F content items is considered and each file has size S, in bits. Moreover, each file has its popularity, which varies geographically, i.e., each UAB-BS b is characterised by a specific order of popularity of contents, described with a PDF $P^{(b)}_{\mathscr{F}}(f)$. As consequence, it is possible that the probability that the content f is required on UAV-BS b_0 is different than on UAV-BS b_1. Nevertheless, for each UAV-BS b, $b \in [1, N_{UAV-BS}]$, $\sum_{f=1}^{F} P^{(b)}_{\mathscr{F}}(f) = 1$. As in

chapter 6, the popularity is described by a Zipf's distribution, which is expressed as:

$$P_{\mathscr{F}}^{(b)}(f) = \frac{\Omega}{f^{\alpha}} \tag{7.14}$$

where $\Omega = (\sum_{i=1}^{F} \frac{1}{i^{\alpha}})^{-1}$ [35]. The parameter α impacts the difference among contents in terms of popularity and defines the steepness of the Zipf's distribution. A large value of α (i.e., $\alpha > 1$) means that the most popular contents are significantly more popular than the other contents. By decreasing α, the popularity of content behaves more similarly to a uniform distribution. As in [35], since the local most popular contents are stored in the server of each UAV-BS b, the hit probability is as follows:

$$p_{HIT} = 1 - \frac{\sum_{f=C}^{F} \frac{1}{f^{\alpha}}}{\sum_{f=1}^{F} \frac{1}{f^{\alpha}}} \tag{7.15}$$

where, as previously mentioned, C is the number of contents stored in each cache, F is the number of items in the library. As can be noticed in Eq. (7.15), p_{HIT} is monotonic increasing with C. The miss probability p_{MISS} is computed as $p_{MISS} = 1 - p_{HIT}$, resulting in monotonic decrease with C.

As in section 7.2, $l^{(u,t)}$ is the traffic volume which is associated with the user u, $u \in [1, U]$, at time t. The number of contents, which each user u requires is modelled as $\lfloor \frac{l^{(u,t)}}{S} \rfloor$, where the $\lfloor \cdot \rfloor$ operator is needed to represent an integer number of requests for content per user. Given the set of users who are associated with UAV-BS b at time t, $\mathscr{U}^{(b,t)}$, the access traffic demand on that UAV-BS b, is given by:

$$b_A^{(b,t)} = \sum_{u \in \mathscr{U}^{(b,t)}} l^{(u,t)} \qquad l^{(u,t)} \geq 0, \forall n \in [1, N] \tag{7.16}$$

Since the MEC technology is now employed, $b_A^{(b,t)} \neq b_{BH}^{(b,t)}$. This is because, in case a user u is associated with an UAV-BS b and he/she requires a content which is stored on that UAV-BS, that content is not transmitted in the BH network, but is directly sent to that user. As a result, the traffic demand in the BH network is:

$$b_{BH}^{(b,t)} = p_{MISS} \sum_{u \in \mathscr{U}^{(b,t)}} (\lfloor \frac{l^{(u,t)}}{S} \rfloor S) + \sum_{u \in \mathscr{U}^{(b,t)}} (l^{(u,t)} \mod S) \tag{7.17}$$

As derived from (7.15), the miss probability p_{MISS} is 1 if the MEC technology is not employed and it decreases with the growth of the server capacity. Knowing this,

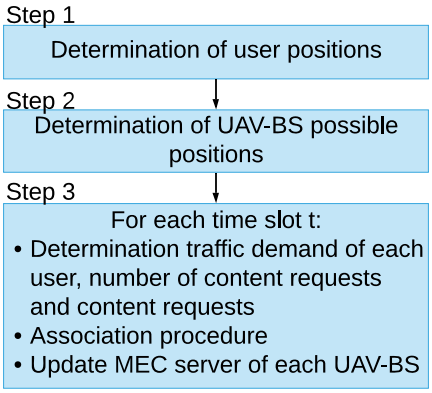


Fig. 7.6 Network allocation algorithm of the tool

(7.17) shows that the MEC technology and the rise of its size prevent the BH network from congestion, since $b_{BH}^{(b,t)}$ linearly drops with p_{MISS}.
The network energy consumption has to account also for the supply of the MEC server, which is installed on each UAV-BS. As a result, Eq. (7.9) has an additional contributor:

$$E^{(t)} = \sum_{b=1}^{N_{UAV-BS}} \left(E_A^{(b,t)} + E_{BH}^{(b,t)} + E_{MEC}^{(b,t)} \right) \tag{7.18}$$

where $E_A^{(b,t)}$ and $E_{BH}^{(b,t)}$, in watt hour, are the energy consumption for the access and BH network, respectively, and are computed as in Eq. (5.2), with the parameters reported in Table 7.3. As regards $E_{MEC}^{(b,t)}$, it is the energy consumption of the MEC server, in watt hour, and is computed as in Eq. (6.2).
As mentioned above, in case a user u needs a content which is stored locally on the UAV-BS b with which he/she is associated, that content is directly sent to him/her. If this is not the case, the content is retrieved in the cloud, reached through the BH network. The average transmission delay is:

$$D = p_{HIT} \cdot D_{HIT} + (1 - p_{HIT}) \cdot D_{MISS} \tag{7.19}$$

where p_{HIT} is given by (7.15), D_{HIT} is the time needed to send a content from an UAV-BS to a user and D_{MISS} is the time needed to bring a content from the cloud to a user. This results in $D_{HIT} < D_{HIT}$ and, hence, from (7.19), the growth of p_{HIT} makes D drop. As already mentioned and highlighted in (7.15), p_{HIT} grows with the growth of the number of stored contents C, i.e. with the cache capacity.

7.5.2 MEC-enabled Simulations

In order to evaluate our scenario, we extend the simulator described in section 7.3. Also for these simulations, the considered operating time starts at 7:00 p.m., ending at 01:00 a.m. and the time is discretized in 5 minutes time slots. Each UAV-BS of the fleet is equipped with a MEC server, which stores popular contents, whose HW technology is DRAM. The servers update their contents according to the LFU cache algorithm, to store the most popular contents of the file library composed of 1000 files, of 50 Mbit size each, as in chapter 6. In case a user requires a content which is stored in the MEC server of its UAV-BS, that UAV-BS directly transmits that content to that user, without accessing the cloud nor the BH network. The position of each user is determined through the uniform distribution previously described, which corresponds to the set of possible location for the UAV-BSs. Each of these UAV-BS locations is characterised by a specific order of popularity, which is described by a Zipf's distribution, as discussed in chapter 6. The level of popularity of each content on each UAV-BS is determined starting from a reference popularity and performing random shuffles from it, to introduce some slight differences among the file's popularity at different locations [31].
At the beginning of each time interval of the simulation, the traffic demand volume $l^{(u,t)}$ of each user is determined as described in section 7.2.1. From this, the number of content requests is derived as $\lfloor \frac{l^{(u,t)}}{S} \rfloor$, where S is the size of each content, which is 50 Mbit, as in [35].
Then, each user is associated with an UAV-BS, if possible. Before starting this procedure, for each user, which contents are requested is determined. To do this, for each user, given the set of possible UAV-BSs locations, we picked the one from which that user experiences the lowest path loss, computed as in section 7.2.3. The Zipf's distribution associated with that UAV-BS location is used to determine which contents are requested by that user. At this point, the association procedure actually begins. The list of active UAV-BS is generated, inserting an UAV-BS if (i) the experienced path loss is lower than an allowable maximum, increasing the transmitting power, if needed and (ii) it can provide the requested traffic demand in the access and BH network. An UAV-BS n is able to provide the required traffic demand in the access network, if Eq. (7.10) is verified. The traffic demand in the BH network depends on the contents, which are requested by the user and which are stored on the considered UAV-BS b. Indeed if a requested content is stored on

that UAV-BS b, in case that user is associated with b, that content is directly sent to the user and the BH network is not used for that delivery. As a result, in order to verify if an UAV-BS b is able to carry the needed traffic demand at time t in the BH network, the following condition is verified:

$$(l^{(u,t)} - Hit^{(u)} \cdot S) + \sum_{\nu=1}^{N_{UAV-BS}} b_{BH}^{(\nu,t)} \leq B_{BH} \tag{7.20}$$

where $Hit^{(u)}$ is the number of contents which are requested by user u and stored in the UAV-BS b, S is the size, in bits, of each content, assumed 50 Mbit, N_{UAV-BS} is the number of UAV-BSs, B_{BH} is the maximum BH traffic volume in a time slot 5 minutes long, set equal to $21.6 \cdot 10^3$ Mb, assuming that its maximum bit rate is 72 Mbps. Notice that in order to compute $Hit^{(u)}$, we are optimistically assuming to perfectly know which contents are requested by user u. To overcome this assumption, the prediction of the content requests is needed, which typically uses Machine Learning-based approaches or Lagrange interpolation, as in [25, 137]. Nevertheless, without loss of generality, this is out of the scope of this thesis, and we leave it as future work.
Then, $b_{BH}^{(\nu,t)}$ is the sum of the traffic demand of the set of users $\mathscr{U}^{(\nu,t)}$ associated with UAV-BS ν during the time slot t:

$$b_{BH}^{(\nu,t)} = \sum_{v \in \mathscr{U}^{(\nu,t)}} (l^{(v,t)} - Hit^{(v)} \cdot S) \tag{7.21}$$

where $l^{(v,t)}$ is the traffic demand at time t of each user v, as above, $Hit^{(v)}$ is the number of contents which are requested by user u and stored in the UAV-BS ν and S is 50 Mbit.
Summarising, if an UAV-BS transmits and the user receives the signal with acceptable quality, i.e. with acceptable path loss, and UAV-BS satisfies (7.10) and (7.11), that UAV-BS is inserted in the list. The user is associated with the UAV-BS from which he/she experiences the lowest path loss among the UAV-BSs in the list. In case the list remains empty, the inactive UAV-BSs are evaluated with the same procedure. If at its end the list is still empty, the user remains uncovered.
In each time slot *t*, once each user has been associated with an UAV-BS b, if possible, the requested contents are delivered. As already mentioned, if a content f, requested by the user u, is cached in the serving UAV-BS b, a hit occurs and the content is

directly transmitted to the user. In case of miss, i.e. the required content is not cached at the UAV-BS b, that content is retrieved from the content provider, reached by the CN, accessed through the BH network. We assume that each requested content is transmitted at the minimum bit rate which allows to receive it within a time slot, i.e. 0.17 Mbps, independently on the required bit rate $\lambda^{(u,t)}$. As a result, the transmission time in case of hit, similar to [35], is:

$$D_{HIT} = \frac{S}{\lambda_S} \tag{7.22}$$

where S is 50 Mbit and λ_S is 0.17 Mbps. In case of miss, the transmission time is the transmission time in the access network, in the BH network and in the CN:

$$D_{MISS} = D_A + D_{BH} + D_{CN} \tag{7.23}$$

where D_A, the time needed to transmit the content from the UAV-BS to the user is as in (7.22) and D_{BH} is the transmission time in the BH network. We assume that the bit rate for the download in the BH network is λ_S equal to 0.17 Mbps, but this is increased when the allocated bit rate in the BH network, for the UAV-BS b, given by the bit rate in each allocated RB, is larger than the needed:

$$D_{BH} = \frac{S}{(\lambda_S + \frac{1}{miss^{(b,t)}}(RB^{(b,t)} \cdot \beta_{RB} - b_{BH}^{(b,t)})} \tag{7.24}$$

where S is 50 Mbit, $miss^{(b,t)}$ is the number of miss at time t, on UAV-BS b, i.e. the number of requests which need the BH network to retrieve the content, λ_S equal to 0.17 Mbps, β_{RB} is the bit rate per RB, equal to 0.72 Mbps, $RB^{(b,t)}$ is the number of RB which are used by UAV-BS b at time t. D_{CN} is the time needed in the CN and, as in [2, 34], is 50 ms. Once each requested content is delivered, each cache is updated, according to the LFU cache algorithm, so as to always cache the most popular contents.

7.5.3 MEC-enabled Performance Evaluation

The results are evaluated varying the capacity of each MEC server and for different values of the parameter α, characterising the contents' popularity distribution. The effects of the growth of the user density, from 50 to 250 users/km^2 are also investi-

gated. First, we analyse the capacity of the access network, the capacity of the BH network and by the covered users. Next, the delay performance of the cache system is described after presenting the evaluation of the network power consumption. Finally, we discuss the impact of the user density on our system proposal.

Access and BH Capacity

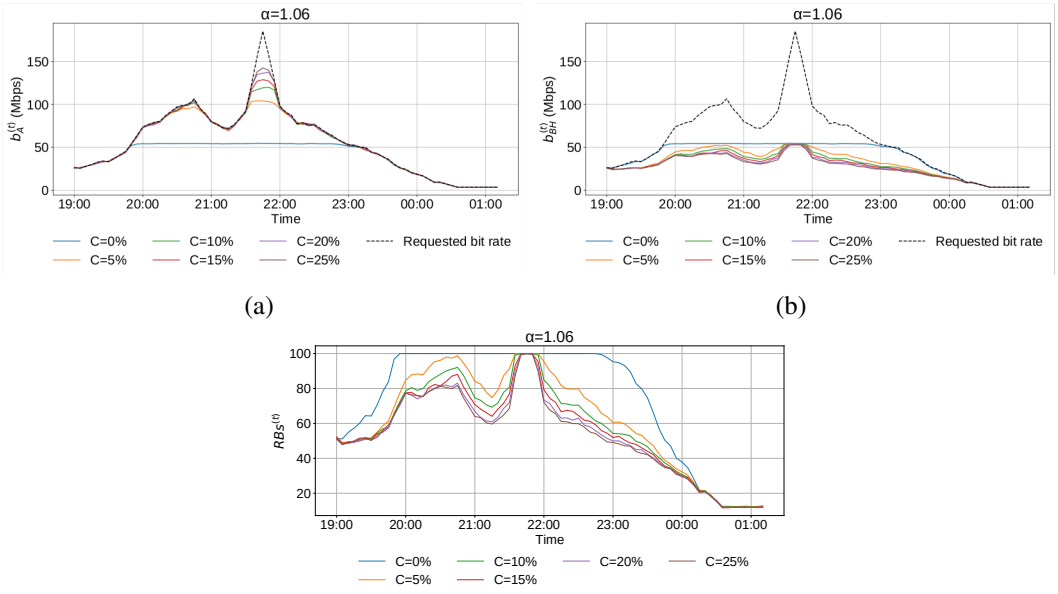


Fig. 7.7 Used capacity with α equal to 1.06: (a) Used Access Capacity (Mbps), (b) Used BH Capacity (Mbps), (c) Used BH RBs.

Figs. 7.7a, 7.7b and 7.7c show the behaviour of the used access, BH capacity, and the number of BH RBs, respectively, with 250 users/km^2, considering the α parameter to describe the popularity of contents. In these figures, α equals 1.06, an average value for the Zipfs's distribution. The dashed line is the total required bit rate and each of the other curves corresponds to the used access and BH capacity, with a different capacity of each MEC server, measured in percentage of the library that the server can store. If it is equal to 0%, we are considering the scenario that does not use the MEC technology. This is the case presented and discussed in sections 7.1 and 7.4, where each content request needs to access the CN, through the BH link in order to retrieve the file. This means that the BH bit rate needed by each user corresponds to the total bit rate required by that user, see section 7.1.

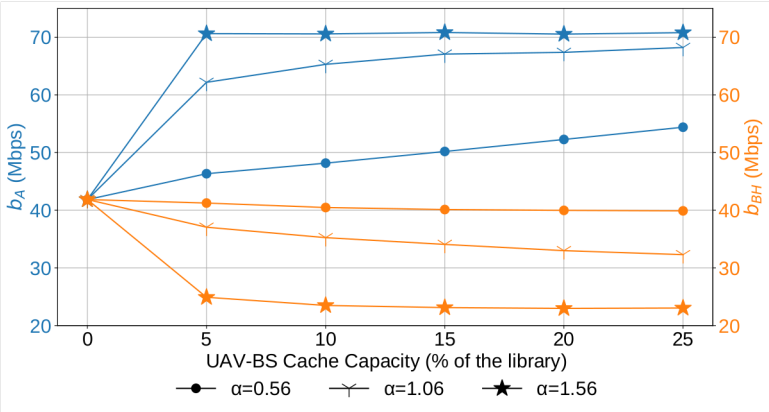


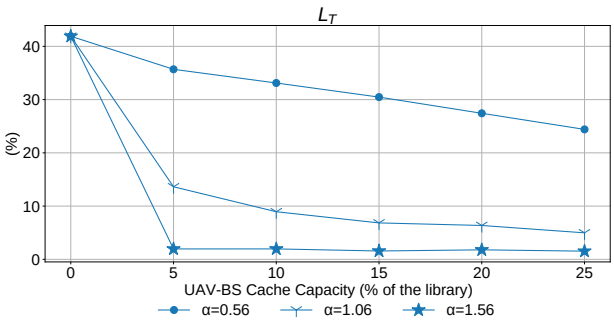
Fig. 7.8 Used capacity varying the capacity of the MEC Server, for different values of the parameter α.

From Fig.7.7a, it is possible to notice that, as discussed in section 7.4, until 19:50, the used access and BH capacity, given by the blue curve, respond instantaneously to the required bit rate. Then, at 19:50, the used access and BH capacity, are unable to provide the required bit rate (dashed curve in Fig.7.7b), because of the BH network saturation, see Fig.7.7c. After the peak, the traffic demand decreases and the network is not saturated anymore and can provide the required bit rate.
When the MEC technology is used, i.e. the capacity of each server is larger than 0%, the used BH capacity, as well as the number of used RBs, is lower than the case without the MEC technology. This is because, part of the contents that are requested by the users are stored locally, in the MEC server. Thus, in case of a hit, retrieving the content at the content provider, reachable through the BH network, is not needed. This occurs more often with larger cache size, since finding the requested content locally stored is more likely. For this reason, the BH used capacity, as well as the number of used BH RBs, is lower if larger caches are considered. The used access capacity is larger, since the BH network is prevented from saturation: when the required bit rate becomes significantly large, from 19:50 on, not all the RBs are used. Nevertheless, when the required bit rate reaches the peak, at 21:45, the BH network saturates, i.e., the needed RBs are larger than the available ones, and for this reason the used access capacity is between 50% and 80% of the required bit rate, depending on the size of the MEC server installed on each UAV-BS.
In Fig. 7.8 the average used access and BH capacity during the whole intervention are shown varying the capacity of each MEC server, on the left y-axis and right y-axis, respectively. The blue, orange and green curves provide the used access and

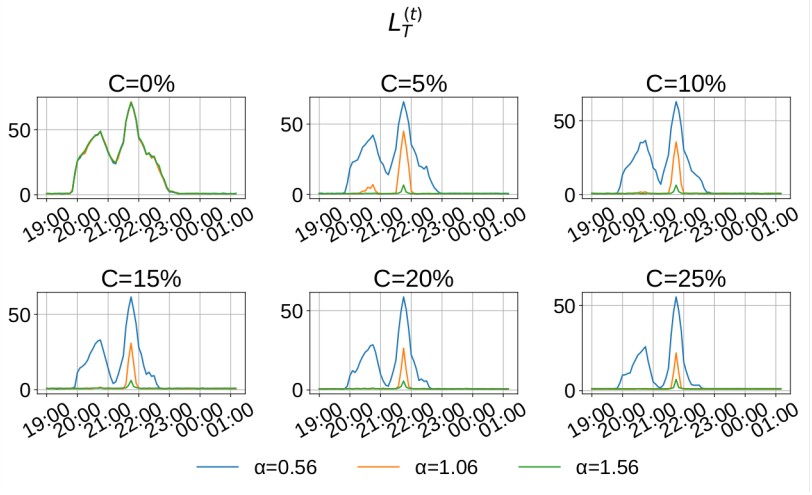
the BH capacity, with α equal to 0.56, 1.06 and 1.56, respectively. From the figure it is clear that, as already seen in Fig.7.7, increasing the size the MEC server installed on each UAV-BS determines the growth of the used access capacity and the drop of the used BH capacity. This is because if more contents are locally stored, the need to retrieve the content in the cloud is less likely, and, as consequence, less capacity is needed in the BH network, avoiding its saturation and permitting to use more capacity in the access one. Nevertheless, as can be observed in Fig.7.8, the trend of the capacities strictly depends on the parameter α, characterised by the Zipf's distribution, used to describe the popularity of files. Large values of α means that there is a small part of the library which is very popular. If this is the case, even a small cache significantly increases the used access capacity, while decreasing the needed BH capacity. This occurs when α is larger than 1: the used access capacity is larger than 60 Mbps and the BH capacity is not larger than 24 Mbps if only 5% of the library is locally stored. In case of a small value of α indicates that the files have similar popularity. In this case, larger caches are needed to significantly increase the used access capacity and decrease the used BH one: if the popularity distribution parameter α is equal to 0.56, 15% of the library should be stored to use 50 Mbps as the capacity of the access network. With respect to the BH network capacity, it is not significantly reduced, even if 25% of the library is locally stored, i.e. from 41 Mbps, when the MEC technology is not used, to 40 Mbps. when the cache capacity is 25% of the library. As an example, when the caching server is deactivated, the access and BH capacity are the same (41.8 MBps), but will represent 58% of the BH and only an average of 25% of the access network, and will vary depending on the number of active UAV-BS used.

Lost Traffic

In this subsection, we discuss the effects of the MEC technology on the lost traffic L_T, which provides a measure of the network QoS. Fig.7.9a shows the percentage of lost traffic L_T versus the capacity of each MEC server, for three different values of α in the scenario with 250 users/km^2. From the figure, it is visible that the growth of the cache capacity installed on each UAV-BS determines the drop of the lost traffic, meaning that this improves the network QoS. If larger caches are used, storing locally a requested content is more likely. This means that, more likely, the access to the cloud to retrieve that content is unneeded and less BH capacity is employed. As a



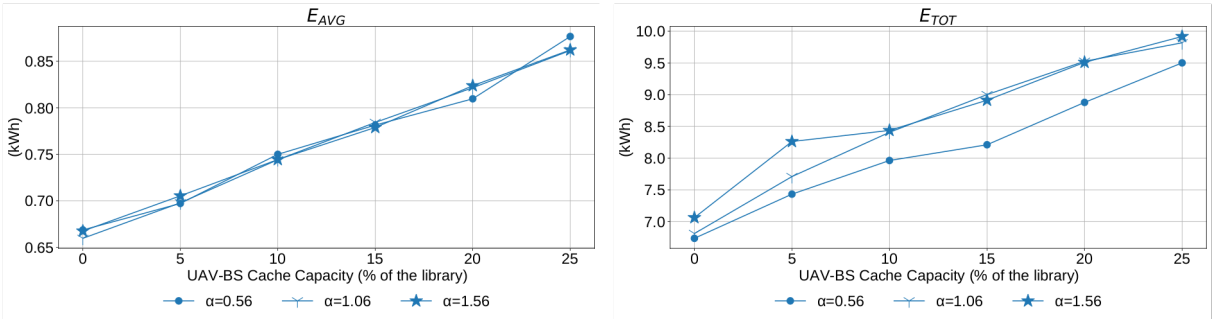
(a)



(b)

Fig. 7.9 Lost Traffic L_T (%), with different values of α and MEC server capacity C (a) Averaged for whole event, (b) Plot in time.

consequence, the whole network does not saturate and more users can be served, losing a low amount of traffic. When the MEC technology is not employed, i.e. each MEC server stores 0% of the file library, all the bit rate required by each user has to be provided in both access and BH network. This makes the BH network more exposed to saturation, resulting in more than 40% of lost traffic. In this case, the BH network saturates as soon as the traffic demand reaches nearly 54 Mbps (see Fig.7.7a). The MEC technology prevents the BH network from this, decreasing the lost traffic. This improvement highly depends on the distribution of content popularity. When there is a small group of very popular contents, ($\alpha > 1$) between 13% and less than 2% of the traffic is lost, when only 5% of the library is locally stored, since it is more likely that the required contents are cached. In case all contents have similar popularity, since requesting for a cached content is less frequent, larger capacity of



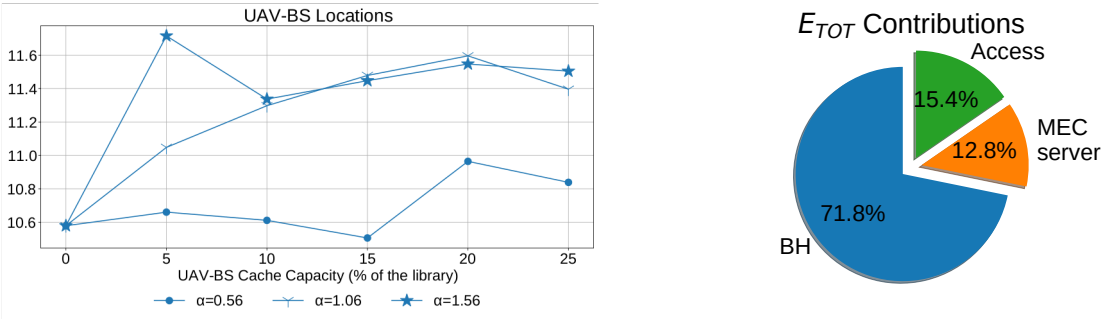
(a) (b)

Fig. 7.10 Energy consumption analysis: (a) E_{AVG} and (b) E_{AVG}, varying the capacity of each MEC server.

the MEC servers is needed in order to significantly reduce the lost traffic: when 10% of the library is locally stored, 33% of the traffic is lost.
A more detailed analysis of the peak hour and the provided users is found in Fig.7.9b, where the percentage of lost traffic versus time, $L_T^{(t)}$, is shown for different values of the MEC servers' capacity. Typically, at the beginning of each simulation no traffic is lost, but as soon as the required bit rate increases, the benefits provided by the employment of the MEC technology are more evident. When MEC is not used, i.e. the MEC server capacity is equal to 0%, the lost traffic significantly grows, see section 7.4. When the traffic demand reaches its peak, at 21:45, without MEC technology, only 76.5% of traffic is handled. This is between 71.2% and 62.7%, if MEC servers are installed on each UAV-BS and contents have similar popularity, i.e. if α is 0.56. Up to 16.3% and 10.8% of traffic is unserved during the peak if α is larger than 1 for MEC capacities of 5% and 25%, respectively. However, for $\alpha = 0.56$ the enhancement was not so good; as shown by the blue lines in Fig.7.9b several users remain unconnected at the beginning of the match and in the half time break. At this peak period, the served traffic ranges between 71.2% to 62.7% for the different sizes of the MEC server.

Energy Consumption

From Fig.7.10a, the energy consumption of each UAV-BS presents a linear behaviour only dependent on the size of the MEC server. It is given by the average energy needed by each UAV-BS to provide the access and BH links, as well as the power required for the MEC server during each simulation. The results highlight that, as



(a) (b)

Fig. 7.11 Power consumption analysis: (a) Number of employed UAV-BS locations, varying the capacity of each MEC server, (b) Power consumption ratio of different UAV-BS sections.

expected, the energy consumption of each UAV-BS is not affected by the characteristics of the popularity of files distribution and it linearly increases with the MEC server capacity. This linear trend is due to the model of the MEC server energy consumption which is used in our simulations, see (6.2). The energy consumption increases by 12%, if 10% of the library is locally stored and it grows up to 31% if each server stores 25% of the library, from 0.9 kWh when no caching is used, to 140 kWh.

Now, we investigate the impact of the usage of the MEC technology on the network energy consumption, reported in Fig.7.10b, where each line is the average network energy consumption for different values of the α parameter, varying the cache capacity of each MEC server. We notice that the network energy consumption is slightly higher when the α parameter and/or the capacity of each cache server grows. Indeed, when each MEC server stored 5% of the library, it increases by 10%, 13% and 17%, when the α is 0.56, 1.06 and 1.56, respectively, with respect to the case with no MEC technology employment. This is because the number of UAV-BSs locations slightly increases with the cache capacity, as can be seen in Fig.7.11a, where the rise of the number of UAV-BS locations grows for larger values of the cache size and the α parameter. In particular, the average number of UAV-BSs locations increases from an average of 10.1 BSs for α and cache capacity equal to 0.56 and 0%, respectively to 11.4 BSs if they are 1.56 and 25%, respectively. By enlarging the cache size, a requested content is more likely to be already cached, so that less BH network capacity is needed, avoiding its saturation and permitting to use more access capacity. As soon as the access capacity of an UAV-BS is totally used, an unused UAV-BS is activated, until the BH network saturates, generating a growth in the number of

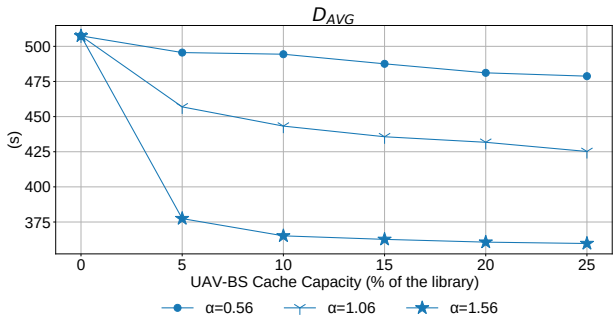


Fig. 7.12 Average Transmission Time, varying the capacity of each MEC server. U:Users. A:Altitude.

used UAV-BS locations, as well as of network power consumption. As a result, large values of α give higher network energy than low ones, because more drones are employed, providing more access capacity and, consequently, dropping the lost traffic.

Fig.7.11b shows each energy consumption contribution, distinguishing among the energy needed to provide communication (access and BH in green and yellow, respectively) and caching, in blue. Given the total energy consumption of an UAV-BS, 15.4% and 71.8% is used for the access and BH networks, respectively, while 12.8% for the MEC server supply. Even if the usage of the MEC technology impacts the energy consumption of an UAV-BS, this does not determine a relevant UAV-BS flight duration reduction. It decreases by 10%: from 77 min, when no MEC servers are used, to 69 min, if the MEC server stores 25% of the library. Since the duration of our event is 375 min, the drop of the flight duration of an UAV-BS determines the increase of the needed number of drones per location during the simulation from 5 to 6.

Average Transmission Time

Pushing caching resources closer to the users provides significant benefits in terms of reduction of the latency. Indeed, if a user requests a content which is locally stored, the BH and CN networks are not accessed and the transmission time is only the time needed to download the content from the UAV-BS. On the contrary, if that content is not locally cached, it is retrieved in the cloud. In this case, the content is downloaded from the cloud through the UAV-BS. The behaviour of the transmission time, while increasing the capacity of each caching server, installed on each UAV-BS

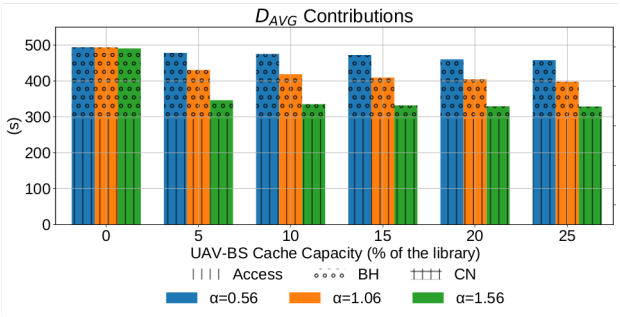


Fig. 7.13 D_{AVG} distinguishing between Access, BH and CN contributions, varying the capacity of each MEC server and the α parameter.

is given in Fig.7.12, where each curve corresponds to a different value of α. The growth of the size of the cache generates reduction of the transmission time, from 500 s down to 458 s, 398 s and 329 s depending on the value of α in a 25% server size. Notice that, even if these values appear high, they are as expected, according to the computation of the experienced latency, see (7.22), (7.23). This reduction of the transmission time is because more contents can be stored locally and the access to the cloud is more likely unneeded reducing the transmission time to only the access network. Also in this case, this reduction strictly depends on the characteristics of the popularity, used to model the popularity of contents. For larger values α, even a small cache size drastically reduces the transmission time compared with the no MEC scenario. When α is 1.56, the transmission time is reduced by 25%, from 500 s to 375 s, with 5% of the library locally stored. A small value of α indicates that the files have similar popularity and the probability to take a content from the CN is higher. As a result, for α equal to 0.56, at least 20% of the library should be stored to reduce the experienced delay by 5% with respect to the no MEC technology usage. In Fig.7.13, the transmission time is reported, for different values of α, increasing the cache capacity. The average time needed in the access network is reported by the bars marked by dark colours, while the time needed in the BH network is indicated with the bars with light colours. Finally, the time needed in the CN is given by the bars with the clearest pattern. As in Fig.7.12, the average transmission delay drops if the cache capacity and/or the α parameter rises. Moreover, from Fig.7.13, we notice that the CN transmission time is negligible with respect to the access and BH network transmission delay. In addition, the access transmission delay remains constant, despite the growth of the cache capacity, as well as of the parameter α, which makes the hit occurrences more likely. Indeed, the drop of the average transmission delay is

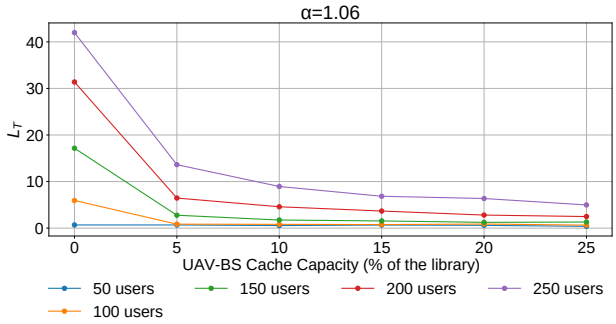


Fig. 7.14 L_T (%) varying the number of users.

due to the drop in the average BH network transmission time. This is because, in case of a hit, which occurs more often with large cache capacity and α parameter, the resulting BH transmission delay is zero. This reveals that the MEC technology effectively reduces the transmission time through the reduction of the transmission time in the BH network, but the access transmission time bounds this reduction.

Variation of the number of users

In this part of the work, we provide an overview of the effectiveness of our methodology, varying the user density.
Each curve in Fig.7.14 corresponds to the percentage of provisioned users for different user density, from 50 to 250 users/km^2, increasing the capacity of each MEC server installed on each UAV-BS and using 1.06 as value of the α parameter. Fig.7.14 shows that, if no local caching is performed, (MEC server = 0%), the percentage of lost traffic L_T, is drastically reduced because of the network saturation, with more than 100 users. Indeed in these cases, more than 15% of the traffic is lost. In case popular contents are locally stored, the percentage of lost traffic is never larger than 10%, if 10% of the library is locally stored and it is lower than 5%, in case 25% of the library is cached on each UAV-BS.

7.6 Final Remarks

The usage of fast deployable networks aided by UAV-BS is considered the future of dynamic network infrastructures. Indeed they result very useful to cope with the

growth of mobile traffic and suitable for overcrowded scenarios, to rapidly bring connectivity where needed. In the first part of this chapter, we show that the BH network is the most challenging part of this network, representing its bottleneck and, for this reason, limiting the user coverage of these networks.
In order to address this issue, the MEC paradigm is considered in this thesis. In particular, MEC servers which provide caching capability are installed on each UAV-BS. In these scenarios, where the attendants share a related interest, the chances that similar users download the same content is quite high and, for this reason, MEC caching servers result very suitable for these situations. Using a realistic traffic demand model and network architecture, composed of a set of UAV-BSs, which provide communication and caching services, we show that this is a very promising solution to avoid the BH network saturation. Nevertheless, the achieved performances are strictly dependent on the characteristics of the traffic, e.g. the popularity distribution of contents, and on the capacity of each MEC server. We investigate the effects of the capacity of the MEC server and of the content popularity, which is modelled as a Zipf's distribution. To this end, we enhance a capacity simulation tool that is applied to a realistic scenario in the city centre of Ghent, Belgium, with realistic traffic acquired from an Italian mobile operator in the San Siro Stadium (Milan). The results of our simulations show that the proposed architecture increases the user coverage by 33% and access capacity by 70% while reducing the BH bandwidth usage up to 55%, as well as the experienced delay by 33%. Also, we prove that for highly popular content, a small server capacity provides more than 93% of requested traffic in the peak hour.

Chapter 8

Modelling Solar Powered UAV-BS for 5G and Beyond

Part of the work presented in this chapter has already been published in:

- Vallero, G., & Meo, M. (2021, June). Modelling Solar Powered UAV-BS for 5G and Beyond. In *2021 19th Mediterranean Communication and Computer Networking Conference (MedComNet)* (pp. 1-8). IEEE.

As highlighted in chapter 7, mounting a BS on UAVs is a promising solution to dynamically deploy fast and flexible communication facilities, to cope with the terrestrial network failures, due to network overloads or physical unavailability. The previous chapter highlights that the BH network is the most challenging part of the UAV-BS-aided network, because of its low available bandwidth, which limits its capacity. To address this issue, we employ the MEC paradigm, to store popular contents and reduce the usage of the BH network, resulting in a QoS improvement. The results also reveal that another important challenge that needs to be addressed is related to the scarce on-board energy availability that is provided to UAV-BSs by on-board batteries. This is an issue which actually characterises UAVs in general but, when they are used as communication infrastructure, the situation worsens. Indeed, besides the energy needed for the flight of the UAV, also the communication unit has to be powered to provide the service, resulting in a higher energy consumption and reduction of the UAV-BS lifetime [39].
To avoid that the energy consumption of the communication unit negatively affects

the lifetime of the UAV, we assume that the amount of energy needed by the communication unit is provided by a PV panel, while the power for the UAV-BS flight is taken from the on-board battery. It results in a longer UAV-BS survival, without adding significant mass to the device [40, 41]. As discussed in chapter 2, even if the solution is promising, several issues need to be tackled. In [96, 138, 97], the optimal UAV-BSs location is optimised, revealing that a higher UAV-BSs altitude increases the PV panel production, while negatively affecting the QoS. The survey in [39] highlights that the solar panel dimensioning and the understanding of the effect of intermittent energy production on the communication service provided by UAV-BSs network is usually neglected in literature.
In this chapter, the dimensioning of the PV power supply system of the LTE MU-MIMO UAV-BS is investigated, focusing on the interplay between traffic demand and energy generation. To do this, we model the communication unit supply as a queuing system using a discretized representation of data and energy flows, as DPs and EPs. The transmission of a DP is triggered by the arrival of an EP. When an EP reaches the system, but no DPs are available or the system is already processing a DP, that EP is unused and, if enough battery capacity is available, moved to the battery and employed to power the UAV-BS flight. Details of the model of this system are given in section 8.1. In section 8.2, through the queue stability conditions, we investigate the energy production levels that are needed to satisfy the traffic demand. The probability that an EP is unused is derived in section 8.3, and the trade-off between the system stability and energy losses is discussed. Finally, in section 8.4, the time variation of both the produced energy and the traffic demand is considered to make the system more realistic and to derive the proper dimension of the PV panel which should be installed on the UAV.

8.1 System Description

The considered scenario is depicted in Fig. 8.1. The UAV-BS provides additional capacity to users, who access the CN using the access links, between users and the UAV-BS, and the BH links, which are established between the UAV and an AP. As in chapter 7, we consider a femto cell BS, equipped with multiple antennas, using the MU-MIMO. The access and BH networks are as in chapter 7. The access network uses a 2.6 GHz LTE femto cell based technology, whose link budget is set as in

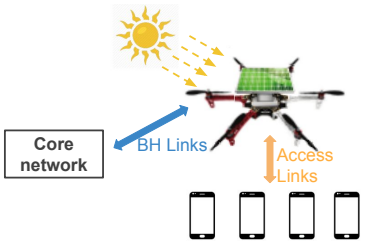


Fig. 8.1 Solar-powered UAV communication system, where the UAV-BS is equipped with solar panels that harvest energy from solar source.

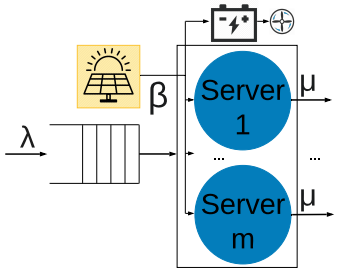


Fig. 8.2 Model of the considered system.

Table 7.1, according to which the radio coverage is 830 m, in Line-of-Sight. The BH links work on the 3.5 GHz frequency and bandwidth of 20 MHz, consisting of 100 RBs [131, 139, 132]. The UAV-BS is equipped with a PV panel that harvests energy from solar and converts it to electrical energy. The PV panel is used to supply the transmitting unit of the drone, while the battery is used to make the drone fly. Without loss of generality, we consider these two energy consuming entities as independent and we focus on the PV panel which produces energy for data transmission. The MU-MIMO femtocell BS is modelled as a queuing system with m servers, as depicted in Fig. 8.2, where each server represents a single MIMO antenna. As in [98, 99], data and energy flows are discretized in DPs and EPs, respectively, and the waiting line in which the DPs are waiting to be transmitted is assumed to be infinite. To transmit a DP, a server needs an EP. This means that as soon as the server ends the transmission of a DP, it switches to sleep mode until an EP arrives. When this occurs, if there is at least a DP in the DP waiting line and an available server, the server is activated and a new transmission starts. In case there are no DPs to be transmitted, that EP is considered lost and can not be used for the transmission of a DP. An EP that reaches the system while all the servers are busy, sending DPs, is lost, as well. As already mentioned and illustrated in Fig. 8.2, actually the lost EPs are not

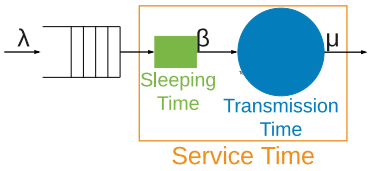


Fig. 8.3 Model for the single server DP queue.

wasted, but moved and stored in the energy battery, if enough capacity is available, and employed to supply the UAV-BS mobility. As in [101, 140], DPs and EPs arrive at the system according to Poisson processes with rate λ and β, respectively; the DP transmission time has an exponential distribution with parameter μ.

8.2 Model Description and Analysis

8.2.1 Analysis of the single server DP queue

We start by analysing the LTE BS equipped with a single MIMO antenna. This is modelled as a single server queuing system, as reported in Fig. 8.2. As described before, when the server completes a service, it switches to sleep mode, until an EP arrives; when an EP arrives, if there is a DP in the data queue, a new transmission starts. The system is equivalent to the one drawn in Fig. 8.3: when the server finishes a DP transmission, it enters into a vacation time, during which it is inactive and waiting for the next EP arrival. This is modelled as a *sleeping box* (represented in green in Fig. 8.3), acting as a delayer, before the server can start transmitting the next DP in the queue, if any. The time spent in the sleeping box is exponentially distributed, with parameter β and we call it *sleeping time*. Once this time expires, i.e., an EP arrives, the server is reactivated and, if there is a DP in the waiting line, the actual transmission starts (see blue part in Fig. 8.3) and lasts for the *transmission time*. The *service time* is defined as the time a DP spends in the orange rectangle in Fig. 8.3. It is given by the sum of the sleeping time (that is the time waiting for the EP) and the transmission time. In our case, this is given by the sum of two exponential random variables with parameters β and μ, respectively. The system can then be modelled as an M/G/1 queue, where the probability density function of the service time G, as discussed in [141], is $f_G(x) = (\beta\mu)/(\beta-\mu)(e^{-\mu x} - e^{-\beta x})$.
The mean value of the service time, E[G], and the variance Var[G] are, respectively,

$\frac{1}{\beta}+\frac{1}{\mu}$ and $\frac{\beta^2+\mu^2}{\beta^2\mu^2}$.
The M/G/1 queue is stable if $\lambda E[G] < 1$, that is $\lambda < \frac{1}{E[G]}$; hence, the system is stable if:

$$\lambda < \frac{\beta\mu}{\beta+\mu} \tag{8.1}$$

Probability that an EP is lost

The probability that an EP is lost, L_{EP}, can be written as :

$$L_{EP} = L_{EP,empty} + L_{EP,busy} \tag{8.2}$$

where $L_{EP,empty}$ is the probability that an EP is lost because when it arrives no DPs are present in the system, the server remains in sleep mode and no transmission starts; it is given by the probability to find the queuing system empty:

$$L_{EP,empty} = 1 - \lambda E[G] \tag{8.3}$$

$L_{EP,busy}$ is the probability that the EP is discarded since, when it reaches the system, the server is active and transmitting a DP:

$$L_{EP,busy} = \lambda E[G]\frac{1/\mu}{1/\mu+1/\beta} = \frac{\lambda}{\mu} \tag{8.4}$$

which is the probability that the queue is not empty ($\lambda E[G]$) multiplied by the probability that the server is processing instead of sleeping.
Out of the stability condition, which means that (8.1) is not satisfied, $L_{EP,empty}$ is zero, since the DP queue diverges and there are always DPs waiting to be transmitted. Similarly, since there is always some DPs waiting in the queue, transmission times alternate with times waiting for an EP to arrive, and an EP is lost if it arrives during a transmission time. The probability that an EP is lost is $L_{EP,busy} = \frac{1/\mu}{1/\mu+1/\beta} = \frac{\beta}{\mu+\beta}$.

8.2.2 Analysis of the DP queue with multiple servers

We now discuss the LTE BS, equipped with multiple MIMO antennas, using MU-MIMO. This is modelled as a multiple server system, each server representing a single MIMO antenna. The system can be modelled as a bidimensional Continuous-

Time Markov Chain (CTMC) $X = \{X(t)\}$ in which the state is given by $\bar{s} = (d, p)$, where d is the number of DPs in the system (both waiting or being transmitted) and p is the number of servers that are transmitting a DP, i.e., the number of transmitting MIMO antennas. Transitions are reported in Table 8.1. Arrivals occur with rate λ and increase by one the number of DPs in the system, d. A departure of a DP occurs with rate $p\mu$, where μ is the service rate and p is the number of servers that are transmitting a DP. Finally, a transmission starts if there are available servers ($p < m$), some DPs that are waiting ($d > p$) and an EP arrives with rate β.

To study the ergodicity of the system let us focus on the *service capacity* of the system with m servers, or the *maximum service rate* that the system can reach. The maximum service rate is reached when the load is high and there are always DPs to transmit. To investigate this situation, we introduce a new model that we call the *maximum service rate model*. The model is an M/M/m/m queue in which services represent DPs transmission, and the customers represent the EPs that enable a service to start: as soon as an EP arrives, a transmission can start if some server is available. Since the EPs cannot be stored, the capacity of the waiting line is zero. The probability to lose an EP in these conditions is denoted by $L^*_{EP,busy}$ and is given by the probability that an EP arrives and finds all the m servers already busy. This probability is the loss probability of the M/M/m/m queue, the very well-known Erlang-B formula, $E_B(m, \beta/\mu)$:

$$L^*_{EP,busy} = E_B(m, \beta/\mu) = \frac{1/m!(\beta/\mu)^m}{\sum_{i=0}^{m} 1/i!(\beta/\mu)^i} \tag{8.5}$$

The throughput of the service capacity model, which represents the maximum service rate of the DP queue with multiple servers, is equal to:

$$S_M = \beta \left(1 - L^*_{EP,busy}\right) \tag{8.6}$$

Hence, the DP queue with multiple servers is stable and the CTMC X is ergodic and reaches a steady-state distribution if the DP arrival rate is smaller than the maximum service rate, i.e., if the following condition holds:

$$\lambda < \beta \left(1 - L^*_{EP,busy}\right) \tag{8.7}$$

Table 8.1 Transition rate out of state $\bar{s} = (d, p)$

Destination State	Rate
$(d+1, p)$	λ
$(d-1, p-1)$ for $p > 0$	$p\mu$
$(d, p+1)$ for $p < m \wedge d > p$	β

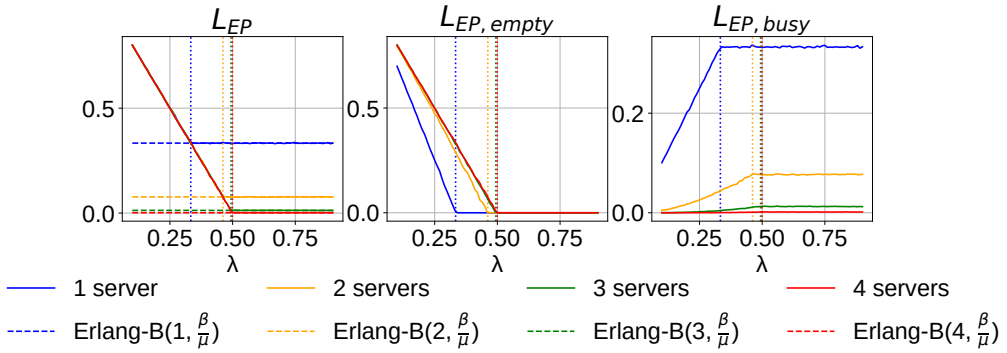


Fig. 8.4 L_{EP} (on the left), $L_{EP,Busy}$ (in the middle) and $L_{EP,Empty}$ (on the right), versus λ and with different number of servers. Vertical lines represent stability conditions, dashed horizontal lines report the EP loss probability given by the Erlang-B formula in (8.5).

From the solution of the CTMC X under the ergodicity condition, it is possible to derive the performance indicators when the system is stable. Let $\pi(d, p)$ be the steady-state probability to have d DPs in the system and p servers that are processing DPs. We can derive the probability that an EP is lost because it finds no DP that needs to be processed:

$$L_{EP,empty} = \sum_{p=0}^{m-1} \pi(p, p)$$

and the probability that an EP is lost because when it arrives at the system all the servers are already busy processing DPs:

$$L_{EP,busy} = \sum_{d=m}^{\infty} \pi(d, m)$$

8.3 Loss Probability

In this section, the models previously discussed are verified by simulation experiments. We simulate our scenario with $m = 1,2,3,4$ servers, β and μ equal to 0.5 and 1.0, respectively, and λ varying between 0.1 and 0.89, with granularity 0.01.
We measure the EP loss probability, which is the probability to lose an EP. Remember that lost EPs are actually not wasted, but moved to the battery and used for the supply of the UAV-BS flight. Fig. 8.4 reports the total probability of losing an EP, L_{EP} (on the left), the probability to lose an EP because there are no DPs to serve, $L_{EP,empty}$ (in the middle), and the probability to lose it because the servers are busy, $L_{EP,busy}$ (on the right). Loss probabilities are shown versus λ and the cases with $m = 1,2,3,4$ servers are plotted, respectively, in blue, orange, green and red. Vertical lines show the ergodicity condition on λ as in (8.7), while the horizontal dashed lines in the plot on the left indicate $L^*_{EP,busy}$ as in (8.5).
Observe the two different regions which correspond to the queue being stable or not and that differ for the behaviour of the loss probabilities. For low values of λ, the DP queue is stable. As λ grows the probability to lose an EP because it finds the queue empty ($L_{EP,empty}$) decreases while it becomes the more and more likely that the EP finds the server(s) busy ($L_{EP,busy}$ grows). Interestingly, when the system is stable, L_{EP} assumes the same values, independently on the number of servers. When the number of servers grows, the probability to lose EPs because the servers are all busy is smaller but this is compensated by the probability to lose and EP because the queue is empty.
When the queue becomes unstable, i.e, when λ exceeds the condition in (8.7), the probability $L_{EP,empty}$ is zero, while the probability $L_{EP,busy}$ is given by $L^*_{EP,busy}$, as in (8.5).
Let us now focus on the ergodicity condition. While increasing the number of servers from 1 to 2 has an important beneficial effect on the maximum service rate, any additional server produces only marginal improvements, which tend to vanish. Indeed, by increasing the number of servers, the probability $L^*_{EP,busy}$ reduces and reaches zero, but the maximum service rate in (8.6) converges to the EP arrival rate β. Hence, we can conclude that by increasing the number of servers, it is possible to increase the maximum service rate (less tight ergodicity condition) but no benefits are achieved in terms of EP loss probability. However, the benefits of additional servers can be observed only as far as their aggregate transmission capacity is small;

when the transmission capacity is large enough, the system is constrained by the energy production system. In order to further increase the system service capacity, a more powerful energy production system is needed.

8.4 Dimensioning UAV-BSs

8.4.1 EP and DP Rates Models

In order to properly model the size of the PV panel which is installed on each UAV-BS, we consider the hourly fluctuation of the energy production and of the data traffic. Hence, the EP and DP rates are not constant but time dependent and denoted by $\beta(t)$ and $\lambda(t)$, respectively. To model the hourly variation of the rates, we start from real PV panel energy production data. To do this, we use data employed in chapter 6, provided by PV-WATT [117], which reports the hourly estimation of the electricity production of a typical poly- or mono-crystalline silicon PV-panel system, accounting for realistic solar irradiation patterns, corresponding to the typical meteorological year in the considered area. It considers 14% for the main typical losses occurring during the process of solar radiation conversion into electricity in the PV panel system. The efficiency is 20%, which means that about 5 m^2 can be assumed per kWp of PV panel capacity [142], while its weight is 2.55 kg per m^2, for off-the-shelf portable PV panels. The elevation at which the PV panel operates, if it is installed on an UAV-BS, makes it more efficient, since it is working at lower temperature and this increases the energy production between 7% and 12%, as reported in [143]. For this reason, the data of energy production provided by PV-WATT are increased by 7%. We normalise the hourly energy generation during the year, so that the generated energy in each hour is given as a fraction of the yearly peak hour production. Then, the meteorological winter and summer periods are selected, using the normalised energy production from 1st December to 28th February and from 1st June to 31st August, respectively. The average hourly energy production during the day is computed, which provides the shape of the hourly EP arrival rate during the typical day in the winter and summer seasons, which are reported in Fig. 8.5, in blue and orange, respectively. Thus, the hourly EP arrival rate, in EP/s, is derived as:

$$\beta(t) = B \cdot f_E^{(t)} \tag{8.8}$$

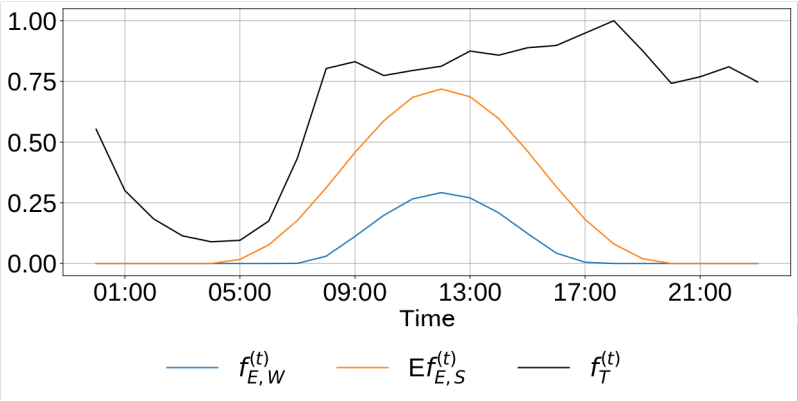


Fig. 8.5 Normalised pattern for the hourly traffic demand (in black) and for the hourly energy production in winter (in blue) and in summer (in orange).

where $f_E^{(t)}$ is the energy production shape as previously derived and B, in EP/s, is the factor which defines the arrival rate of EPs.
To formalise the variation of the traffic demand, we use the traffic data presented in chapter 6. As mentioned in the previous chapters, they report the traffic demand volume, in bits, in a wide area around Milan, in Italy, for a duration of two months in 2015, with granularity of 15 minutes. As in chapter 3, data are aggregated to have an hourly granularity and the average hourly traffic demand during the day is computed. Then, data are normalised and the daily pattern $f_T^{(t)}$ is obtained; it is reported in black in Fig. 8.5. The parameter $\lambda(t)$, in DP/s, is computed as follows:

$$\lambda(t) = L \cdot f_T^{(t)} \tag{8.9}$$

where L, in DP/s, is a factor, which scales the amplitude traffic rate.

8.4.2 Model of the PV Panel

We now derive the size of the PV panel which has to be installed on each UAV-BS to make the system stable; i.e., to serve the traffic. The parameter B, in EP/s, is defined as:

$$B = \frac{C_{PV}}{EP_J} R_{PEAK} \tag{8.10}$$

where C_{PV} is the capacity of the PV-panel, in watt, EP_J is the energy carried by an EP, in J/EP, and R_{PEAK} is the ratio between the yearly maximum power production per hour of the considered PV-panel and its nominal capacity and it varies according

Table 8.2 Parameters setting

Elevation [m]	**P_A [W]**	**P_{BH} [W]**	**EP_J [mJ]**
70	17	79.96	6
120	15.89	93.04	6.33

Table 8.3 Summary of PV panel capacity, surface and weight

B	**C_{PV} [W]**		**Surface [m^2]**		**Weight [kg]**	
	70 m	120 m	70 m	120 m	70 m	120 m
0.1	0.7	0.83	0.04	0.04	0.1	0.1
3	23.68	24.98	0.1	0.1	0.25	0.25
5	39.47	41.64	0.19	0.2	0.48	0.51
7	55.26	58.29	0.28	0.29	0.71	0.74
10	78.95	83.27	0.39	0.41	0.99	1.04

to the location of the PV-panel. Substituting Eqs. (8.8) and (8.10) in Eq. (8.1), the system results stable if the following condition holds:

$$\frac{\lambda(t)\cdot\mu}{\mu-\lambda(t)}\frac{EP_J}{R_{PEAK}f_E^{(t)}} < C_{PV} \tag{8.11}$$

8.4.3 Parameters Setting

In this part of the work, we consider a LTE MU-MIMO UAV-BS, equipped with a single MIMO antenna, which supports the ground RAN providing additional capacity to users. We assume that its bit rate BR is the weighted average of the access and BH bit rate, BR_A and BR_{BH}, computed as $(BR_A \cdot T_A + BR_{BH} \cdot T_{BH})/(T_A + T_{BH})$, where T_A and T_{BH} are the time needed to transmit a DP in the access and BH networks, respectively. BR_A and BR_{BH}, as in chapter 7, are 16.9 Mbit/s and 72 Mbit/s, respectively, while T_A and T_{BH} are, respectively, equal to $3.35 \cdot 10^{-4}$ s and $0.79 \cdot 10^{-4}$ s, assuming that the size of each DP is 709 bytes, as in [144]. From these parameters, we derive that BR is 27.38 Mbit/s and μ is 4.28 DP/ms. The size of each EP, which is the necessary amount of energy needed to transmit a DP, is the average between the energy needed for the transmission of a DP in the access and in the BH network. To compute these values, the models employed in [3, 4, 145] are used, which depend on the HW components of the antenna, as well as on the transmitted power, which varies with the elevation of the drone. Here, we assume

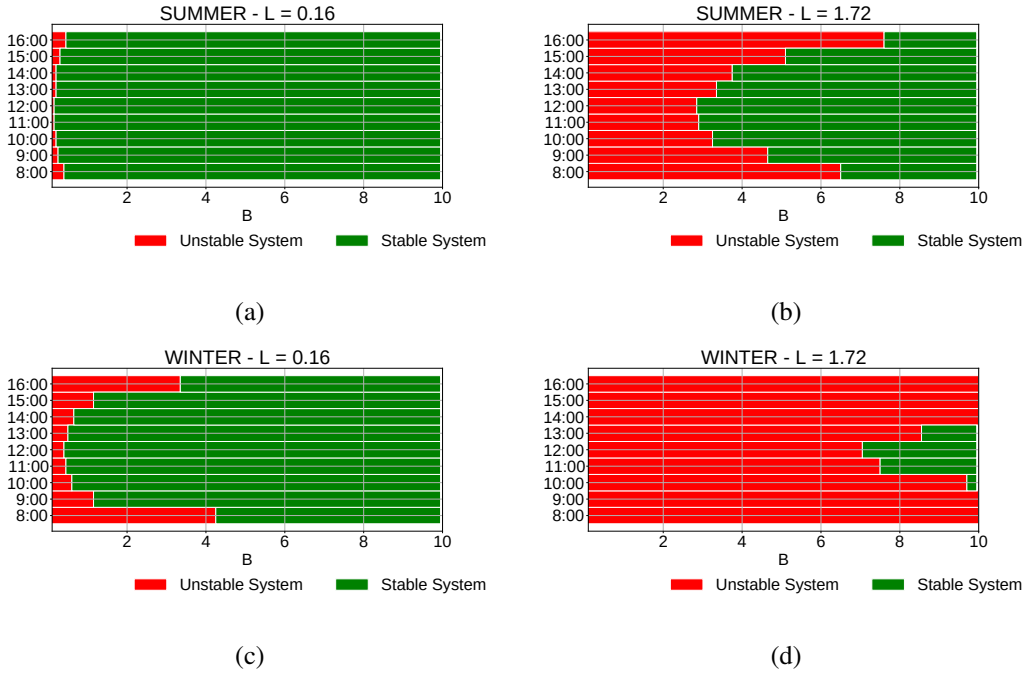


Fig. 8.6 Values of B which make the system stable (enough energy for traffic demand) during each hour under high traffic (L = 0.16) and low traffic (L = 1.72), in summer (a) and (b) and in winter (c) and (d).

two different elevations, set equal to 70 m and 120 m, since the maximum allowed drone elevation until 31st December 2020, in Italy, is 70 m but from 1st January 2021 considered drones are allowed to fly up to 120 m, according to the European regulation reported in [146]. Given the power consumption of the access and the BH interfaces, P_A and P_{BH}, respectively, derived with the simulator discussed in chapter 7 and reported in Table 8.2, at the two considered elevations, the size of each EP, EP_j, in joule, is $EP_j = 0.5 \cdot (P_A \cdot T_A + P_{BH} \cdot T_{BH})$, where T_A and T_{BH} denote the time needed to transmit a DP in the access and BH network, respectively. If the drone is at 120 m, it transmits more energy to reach the users than when it operates at 70 m. As a consequence, the size of each EP depends on the elevation of the drone (see Table 8.2). Finally, from solar production data provided by [117], we know that the yearly maximum power production of a PV panel installed in Turin, is equal to 76% of its nominal capacity and R_{PEAK} is set accordingly.

8.4.4 Performance evaluation

Through the stability of the system, we investigate the capability of the UAV-BS to serve the traffic demand and the relation between the traffic and the energy production. We focus on daylight time, i.e. from 8:00 to 17:00. On x-axis of Fig. 8.6, the parameter B, which determines the energy production as in (8.8), varies from 0.1 to 10.0. As previously mentioned, the nominal capacity of the PV panel corresponding to a value of B depends on the elevation at which the UAV-BS is operating. When B is 0.1, the nominal capacity of the PV panel is 0.79 W and 0.83 W, if it is 70 m and 120 m, respectively, corresponding to a surface of $0.4 \cdot 10^{-2}$ m^2 and a weight no larger than 0.1 kg; with B equal to 10.0, the nominal capacities are 78.95 W and 83.27 W, which means a PV panel surface of 0.4 m^2, with weight around 1 kg. For the considered elevations, the capacity, surface and weight for some values of B in the considered interval, are reported in Table 8.3. The table indicates that the considered PV panels are small enough, in terms of both surface and weight, to make their installation feasible on an UAV-BS.

The green and red bars in Fig. 8.6 indicate the values of B, which make the system stable and unstable during each hour reported on the y-axis: when the system is unstable, the energy is not enough to satisfy the traffic demand. Results are given for two different values of traffic, represented by L equal to 0.16 and 1.72, corresponding to 0.89 and 9.75 Mbps, respectively, in Figs. 8.6b and 8.6d, in summer and winter. The figure shows that, as expected, low traffic demand can always be satisfied even in winter with limited energy production; while high traffic demand is satisfied only in summer, when the energy production is high and large PV panels are considered. Larger values of B are needed to serve the traffic during low energy production hours, from 8:00 to 10:00 and from 14:00 to 16:00, than during the peak energy production period, from 11:00 to 13:00.

The probability to lose EPs, L_{EP}, is plotted in Figs. 8.7, where each curve corresponds to a given hour of the day; different values of B are considered, from 0.1, in blue, to 10.0, in red. L_{EP} grows when lower energy production levels are needed: that is when traffic is low and production is high. Indeed, the peak of L_{EP} always occurs at 12:00.

In particular, at 12:00, in summer, B equal to 0.15 and 2.85 stabilises the system, when L is 0.16 and 1.72, see Fig. 8.6b, respectively, maintaining L_{EP} no larger than 0.02 and 0.32. At 16:00 values of B larger than 0.45 and 7.6 are needed. This means

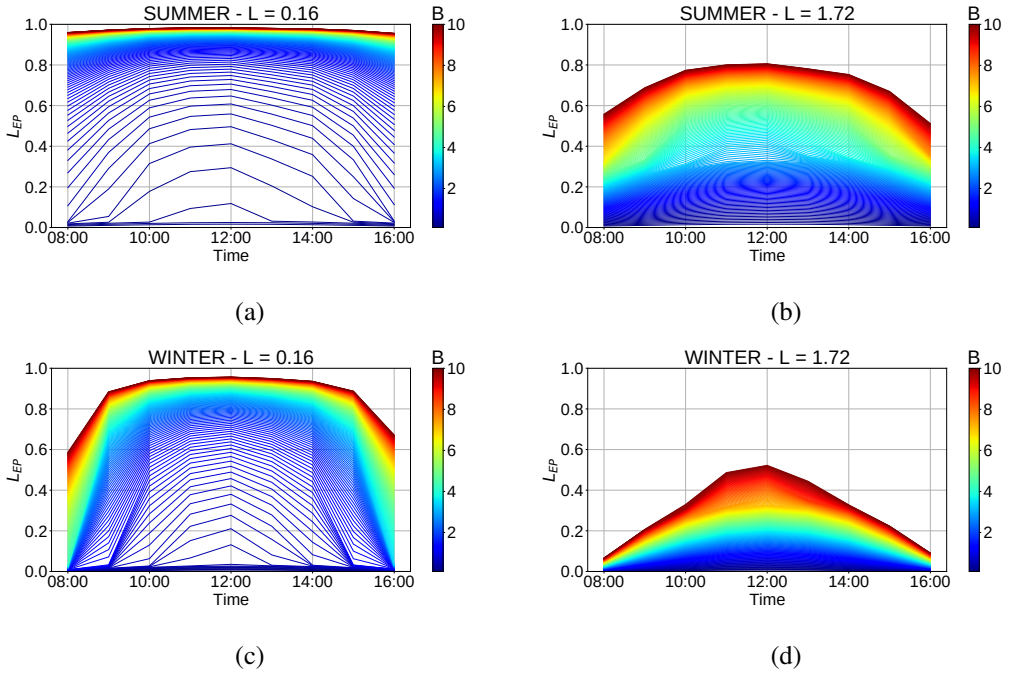


Fig. 8.7 Hourly L_{EP} in summer with L = 0.16 (a) and L = 1.72 (b) and winter with L = 0.16 (c) and L = 1.72 (d).

that in summer during peak hours production, if the UAV-BS operates at 120 m of altitude, it needs a PV panel with capacity of 1.2 W and 24 W, while, if it is at 70 m, the capacity slightly decreases to 1.1 W and 22.5 W. For both the elevations, these nominal capacities correspond to a panel surface of $0.6 \cdot 10^{-2}$ m^2 and 0.12 m^2, if L is 0.16 and 1.72, respectively. In case the UAV-BS provides the service out of the peak of the energy production, the needed PV panel capacity at 120 m increases up to 3.7 W and 63 W, corresponding to PV panel areas no larger than 0.02 m^2 and 0.3 m^2. In case the UAV-BS operates at 70 m, the needed nominal capacities are 3.55 W and 60 W, without significant reductions of PV panel surface.

In winter, the situation is similar as illustrated in Figs. 8.6c, 8.6d, 8.7c, 8.7d, even if larger values of B than in summer are needed and lower values of L_{EP} are reached, because of the low energy production, which characterises this season, as shown in Fig. 8.5. At 8:00 and 16:00, when L is 0.16, B greater than 4.2 and 3.35, respectively, is needed in order to stabilise the system, keeping L_{EP} lower than 0.03. This means that with this traffic intensity, a PV panel whose nominal capacity is between 28 W and 35 W is needed if the UAV-BS is located at 120 m, which has a surface between 0.17 m^2 and 0.14 m^2. This capacity rises between 26.4 W and 33 W, with

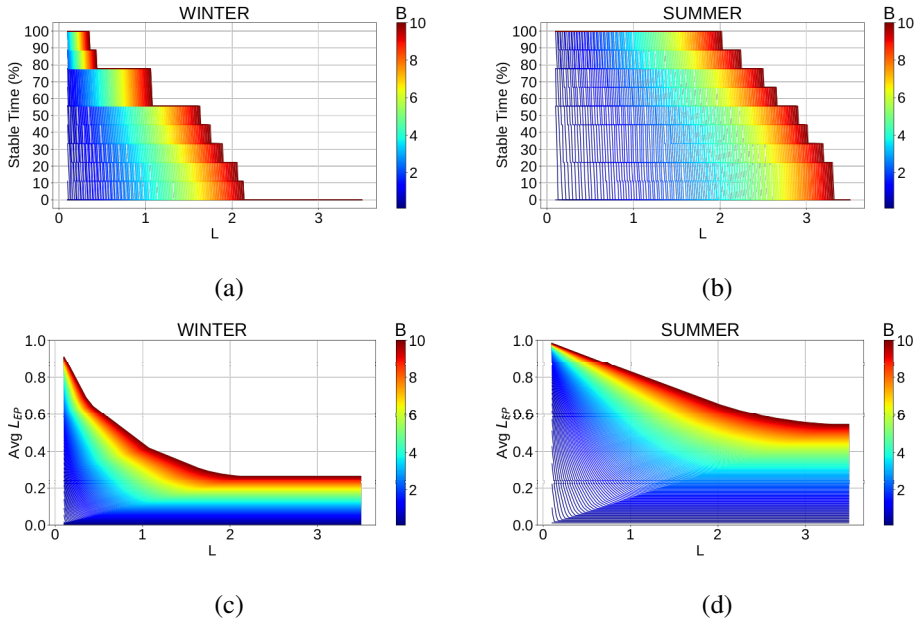


Fig. 8.8 Percentage of time during which the system is stable in winter (a) and summer (b); Average L_{EP} in winter (c) and summer (d).

no significant variation of the needed PV-panel area, if the UAV-BS elevation is 70 m. At 12:00 B equal to 0.4 is sufficient to make the system stable, provided by a PV panel with nominal capacity of 3.3 W and surface equal to $0.17 \cdot 10^{-1}$ m^2. In these cases, L_{EP} is 0.03. Increasing L up to 1.72, i.e. increasing the traffic intensity, only during the peak hour production, from 10:00 to 14:00, B larger than 7.05 makes the system stable, with L_{EP} no larger than 0.32, which is achieved when the installed PV panel capacity is 71 W, which has a surface of 0.3 m^2.

Separately considering winter and summer, Figs. 8.8a and 8.8b provide the percentage of time during a day for which the system is stable, increasing the value of L (and, hence, the traffic) from 0.1 to 3.5. Each curve in the figures corresponds to a different value of B, from 0.1, in blue, to 10.0, in red. These figures confirm that large values of L require large values of B to achieve the stability of the system. In winter, the system is stable 100% of the time (meaning it can serve all the traffic at each time of the day) only if traffic is low (L is between 0.1 and 0.22), and the production system is large (B larger than 7.8). The nominal capacity of the PV-panel that is needed is at least 61.58 W, if the UAB-BS is located at 70 m, while it is at least 64.95 W, when it works 120 m above the ground. If lower values of B are used, the system reaches stability less than 80% of the considered time and with B

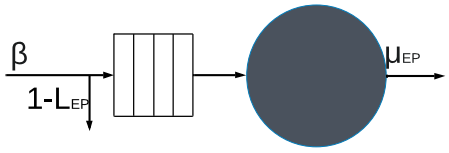


Fig. 8.9 Model of the UAV-BS flight supply system.

lower than 1, which corresponds to a nominal capacity smaller than 9 W, for both the considered elevations, instability occurs for more than half of time. In summer, the system is stable for the whole considered time even under high traffic (L up to 1.29), if B is larger than 9.95. The system results stable for more than 80% of the time when L is 0.43, if B is between 2.0 and 10, i.e. if the nominal capacity of the PV panel is 15.8 W, if the drone height is 70 m, and 16.7 W, if the UAV-BS is located at 120 m above the ground, but for less than 60%, if B is lower than 1.45, which is the scenario where a PV-panel with nominal capacity of 12 W is employed.
As discussed above, large values of B and small values of L, beside making the system stable, generate large values of energy losses L_{EP} as shown in Figs. 8.8c and 8.8d. L_{EP} decreases with B and L, until the region where the system is unstable, where L_{EP} assumes a constant value, which is up to 0.54, in summer and up to 0.27, in winter, according to the B and L setting.
These results show how the models proposed in the previous sections can be used to understand the relation between traffic demand and power supply and to dimension the UAV-BS power supply system. Results also highlight the importance of the evaluation of the period of the day and of the year in order to verify the feasibility of this solution and its proper sizing.

8.4.5 Impact of the PV panel supply on the UAV-BS flight time

In this section, we investigate the impact of the PV panel usage on the flight time of the LTE MU-MIMO UAV-BS, equipped with a single MIMO antenna. As mentioned above, the PV panel is used to supply the transmitting unit of the drone, while the battery is used to make the drone fly. When an EP, produced by the PV panel, reaches the transmitting unit, if there is at least a DP in the DP waiting line and an available

server, a new transmission starts, see section 8.1. As depicted in Fig. 8.2, in case an EP arrives at the transmitting unit but there are no DPs to be transmitted or all the servers are busy, sending DPs, that EP is considered lost and it is moved and stored in the energy battery, if enough capacity is available, and employed to supply the UAV-BS mobility.

In this section, we quantify the impact on the UAV-BS flight time of the EPs which are considered lost at the transmitting unit but reach the battery used for the supply of the drone flight. The part which makes the drone fly is modelled as a queuing system with a single server and a finite waiting line, where the customers, which are the EPs, are waiting to be consumed for the UAV-BS flight supply, as depicted in Fig. 8.9. Since the UAV-BS has full battery when it starts to operate, we assume that when the system starts, the waiting line is full and as soon as it is empty, the drone operation ends, since no energy is available to supply the flight of the drone. The EPs which reach the system are the ones which are lost at the transmitting unit. This means that the EPs are generated according to a Poisson process with rate β, but they actually reach the battery of the UAV-BS flight part with probability L_{EP}, see Fig. 8.9. The probability L_{EP} is computed as in (8.2). The processing time of the customers, i.e. EPs, corresponds to the time which is required to consume an EP for the supply of the flight of the UAV-BS. We assume that it is constant, with rate μ_{EP}, in EP/ms. As in [3, 147], the battery voltage and the current flight usage of the UAV-BS which we are considering are 14.3 V and 5 A, resulting in a power flight usage equal to 71.5 W. This means that the energy needed to fly in 1 ms is $71.5 \cdot 10^{-3}$ J, which corresponds to 11.29 EP/ms, assuming that each EP contains $6.33 \cdot 10^{-3}$ J, as in the case the UAV-BS operates at 120 m from the ground. As a result, since the UAV-BS has a time flight equal to 24 minutes, when no EPs reach the battery used for the supply of the drone flight, the waiting line can store up to 16265402 EPs.

In Fig. 8.10 the growth of the lifetime of the UAV-BS, because of the EPs which are lost in the transmitting unit and used for the supply of the drone flight, is reported. In particular, the duration of the UAV-BS flight is shown, in the summer scenario, in Figs. 8.10a, 8.10b and, in the winter one, in Figs. 8.10c, 8.10d, for L equal to 0.16 and 1.72, respectively, which correspond to 0.89 and 9.75 Mbps. Each curve in these figures is the flight time duration, in minutes, for a given hour of the day, from 8:00 to 16:00, increasing the B parameter, which determines the energy production as in (8.8) and varies from 0.1 to 10.0. As previously discussed, B equal to 0.1 corresponds to a nominal capacity of the PV panel equal to 0.83 W, while with B

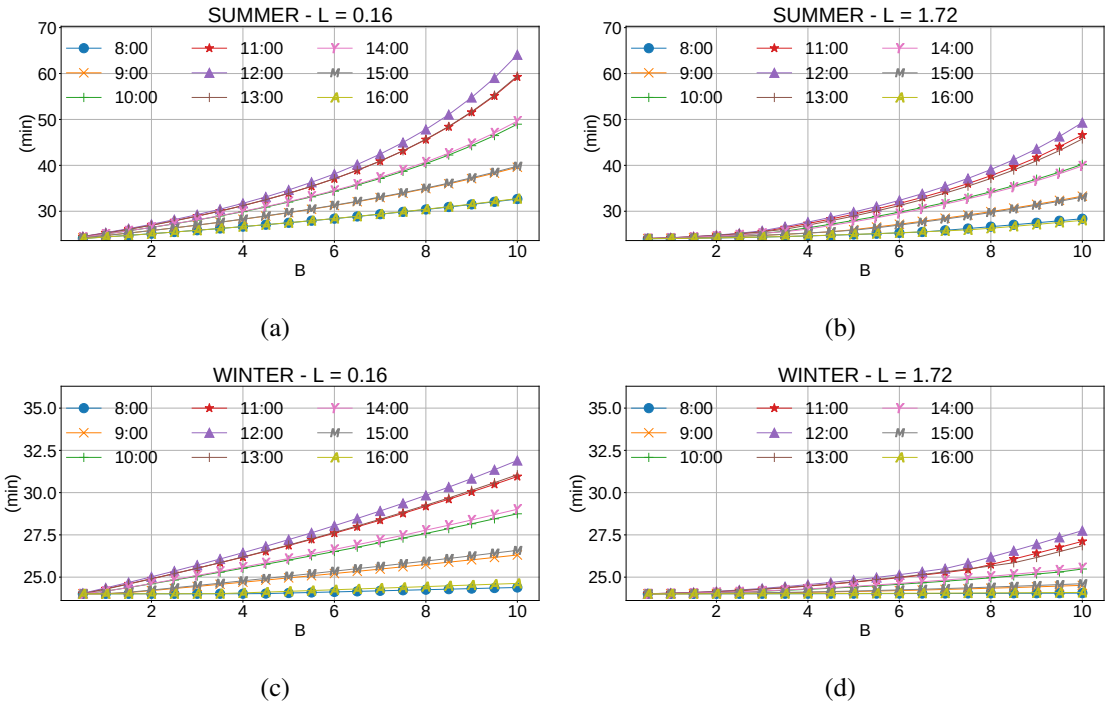


Fig. 8.10 UAV-BS flight time duration in summer with L=0.16 (a) and L=1.72 (b) and in winter with L=0.16 (c) and L=1.72 (d).

equal to 10.0, the PV panel nominal capacity is 83.28 W. Similar to the previous discussion, the growth of the flight time duration strictly depends on the period of the day and of the year and high traffic demand needs large values of B to significantly increase the drone flight duration. Indeed, in summer, with B equal to 5, the UAV-BS is able to fly for up to 45 and 30 minutes, meaning that it grows by up 44% and 24%, with L equal to 0.16 and 1.72, respectively. When B is equal to 10, the drone is able to operate for up to one hour. In winter large values of B are needed to achieve significant rise of flight of the UAV-BS. When B is 10, the lifetime of the drone lasts up 32 and 28 minutes, with L equal to 0.16 and 1.72, respectively.

8.5 Final Remarks

The use of UAVs, on which BS equipment is mounted, is a promising solution to dynamically provide additional capacity in RANs. Usually, UAV-BSs are powered by on-board batteries, which makes their survival short, because of the scarceness of the energy availability. For this reason, solar-powered UAV-BSs are an interesting

alternative which, however, raises a number of challenges related to its dimensioning and the intermittent nature of the energy generation.
In this part of the work, we model a PV-panel powered LTE MU-MIMO UAV-BS and investigate the different system operation regions, as a function of the traffic demand and the energy production. Our results reveal that the usage of the PV panel as a unique energy source for the communication unit of an UAV-BS is an effective solution but it has to be properly sized in order to operate in stability conditions. Keeping into consideration the period of the day, of the year and the traffic demand intensity is fundamental in order to properly design this solution.

Chapter 9

Conclusion

In line with the Paris Agreement and the European Green Deal, the communication community has recognised the network energy efficiency as a fundamental and urgent aspect, to make the communication network sustainable. The BS have been identified as the most energy consuming components of mobile networks, accounting for 80% of the total consumption of the RAN and their energy need is expected to further grow because of the rise of the mobile data traffic, expected in the next years. The contribution of this thesis consists in designing, analysing and evaluating high energy efficient RANs, investigating various critical issues raised by the introduction of the MEC technology and the UAV-BSs, pillar technologies for 5G and beyond RANs. One of the most studied approaches for the reduction of the energy consumption of RANs is the RAN management, based on the BS switching. It is often combined with the RAN renewable energy supply, which uses locally installed RES, typically composed of a PV panel system, to provide the energy needed for the network supply. The Resource on Demand approaches leverage the temporal variability of the traffic and it deactivates BSs in case of low traffic demand. Alternatively, in case the RAN is supplied by a RES system, the Resource on Produced Energy approaches can be employed. If this is the case, the BS switching decisions are based on the amount of available renewable energy, which is locally produced. In particular, a BS is put in sleep mode, when that quantity is not sufficient for the network supply, to minimise the energy which has to be purchased from the grid. We propose different approaches for this scenario, which provide up to 40% of energy saving and good QoS, in the considered scenarios. Results also highlight the fundamental role of

the macro cell BSs in hierarchical RANs, in order to provide the adequate QoS. With the employed strategies, the decision to switch a BS to sleep mode is driven by the future traffic demand and/or the future generation of RES. In order to access these predictions, AI and ML based approaches are used in this work. We use different ML algorithms and we also propose and discuss ANN-based solutions to derive the needed predictions. Our results show that ML algorithms are necessary to achieve a good trade-off between energy efficiency and QoS. Moreover, results point out that large errors in the forecast do not always imply bad network performance. Indeed, the correct estimation of traffic is important only around the values that trigger the decision to activate or deactivate some micro cell BSs. This makes the RAN performance slightly sensible to the used ML algorithm. Forecast algorithms that tend to overestimate traffic, yield lower energy saving without deteriorating QoS. On the contrary, forecast algorithms that tend to underestimate traffic yield losses of traffic, despite higher energy saving, increasing it up to 5 percentage points (which does not compromise QoS). Results also reveal that energy savings strongly depend on traffic patterns that are typical of the considered area. This implies that a widespread implementation of these energy saving strategies without the support of ML would require a careful tuning that cannot be performed autonomously and that needs continuous updates to follow traffic pattern variations. On the contrary, ML approaches provide a versatile framework for the implementation of the desired trade-off that naturally adapts the network operation to the traffic characteristics typical of each area and to its evolution. However, ML approaches become particularly effective only if their outputs are integrated into decision processes that are driven by a deep domain knowledge, which cannot be eliminated if the desired objectives are to be achieved. Indeed, this thesis shows that if the traffic predictions are carefully processed, QoS deterioration is avoided, while significant energy saving can be achieved. Prediction processing requires both the understanding of traffic patterns over long time scales, so as to detect the overall trend of increasing or decreasing traffic, as well as strategies to combine predictions at different time lags. Prediction processing and the consequent dynamic resource allocation affect the BS failure rate in different ways. In addition, this work investigates the impact of the BS switching on the BS HW lifetime. Switching a BS is harmful to its failure rate while the time spent in sleep mode prevents its deterioration. The actual impact of the combination of these two phenomena depends on the HW components of the BS, as well as on the RAN management strategy. In case the switching of a BS is not costly, less strict

switching conditions can be applied: the BS failure rate is not affected while larger energy saving is achieved. Conversely, when the BS is sensitive to switching, more conservative resource allocations should be employed. For existing networks, not designed for highly dynamic resource allocation, conservative approaches better prevent BSs from HW failure. However, in perspective, with the deployment of new devices suited for strongly dynamic networks, less conservative approaches, which frequently activate and deactivate BSs, can be used, and higher energy saving is expected.

Our work helps understand the simultaneous employment of the BSs switching and the MEC paradigm. In RANs, the MEC paradigm consists of the placement of computing and storage servers, directly at each BS of these networks. It is considered a promising solution, in order to expand the existing mobile networks, achieving ultra-low delays, extensive coverage and ultra-high reliability. In this work, RANs which use both the BS switching for the energy efficiency and the MEC for the delay reduction are investigated, providing an overview of their mutual effects. The employment of the MEC technology increases the RAN energy consumption, since the MEC platforms have to be powered to provide the service. Meanwhile, the BSs switching dynamically activates and deactivates resources, and consequently the MEC servers, impacting its performance. We notice that caching at the edge and dynamic activation of the BSs, can be very effective in reducing latency and reducing the network power consumption, respectively, without deteriorating their performances because of their coexistence. Indeed, even if strictly dependent on the characteristics of the traffic and on the server capacity, the MEC technology reduces the experienced delay up to 60%, without generating significant growth of the network energy consumption, limited to 7%. In addition, the employment of an energy reduction strategy, applied in case of renewable energy shortage, reduces the energy consumption but does not impact the experienced delay. Nevertheless, caching on the macro BSs is always needed to significantly reduce delays, while caching also on the micro cells relieves the effort on the macro cell. Finally, we propose association policies, whose objective is the minimisation of the RAN energy consumption and/or the experienced delay, in order to maximise the benefits provided by the MEC technology usage, ensuring also the achievement of the network energy efficiency. Association procedures which minimise the delay and/or the energy consumption tend to associate users with macro BSs. In this way, when the network energy consumption is optimised, it decreases by up to 27% and also the experienced

delay is slightly reduced by 4%. Meanwhile, in case the association procedure is based on the minimisation of the delay, it drops by 10%, but the energy consumption increases by 4%, since users can be associated with BS which are far, and the emitted power has to be increased so as to receive the signal with the adequate quality.
Moreover, this work investigates the usage of UAV-BS-aided RANs. UAV-BS networks are fast deployable networks, which use as network infrastructure the UAV-BSs. Indeed they cope with the growth of mobile traffic and are suitable for overcrowded and emergency scenarios, to rapidly bring connectivity where needed. Our simulations point out that UAV-BSs need hourly replacement because of the scarce on-board energy availability on these drones, provided by on-board batteries. In addition, simulations show that the BH network often saturates, due to its low available bandwidth, which limits its capacity and significantly deteriorates the network QoS. To cope with the BH network saturation, which significantly deteriorates the QoS, we use the MEC paradigm, to cache popular contents on each UAV-BS and decrease the occupancy of the BH network. Simulations, performed for different traffic characteristics and different MEC server capacity, reveal that this is a very promising solution to avoid the BH network saturation. In particular, the QoS is significantly improved, by reducing the lost traffic by 33%, thanks to a halving of the BH network capacity needs. As a consequence, the access capacity grows by 70% while the delay drops by 33%.
In order to address the lack of on-board available energy, we consider a solar-powered UAV-BSs. The model of a LTE MU-MIMO UAV-BS, powered by a PV panel system, is formalised. In order to do this, queuing theory is employed, coupled with a discretized representation of the data flow and energy flow, as DPs and EPs. The transmission of a DP is triggered by the arrival of an EP. When an EP reaches the system, but no DPs are available or the system is already processing a DP, that EP is unused and, if enough battery capacity is available, moved to the battery and employed to power the UAV-BS flight. Through this model, we investigate the interplay between the traffic demand and the energy generation, in particular the energy production levels that are needed to satisfy the traffic demand, the probability that an EP is unused and the proper PV panel dimensioning are analytically derived, also considering the time dependence of both the traffic demand and the PV panel production. Our results reveal that the usage of the PV panel as a unique energy source for the communication unit of an UAV-BS is an effective solution, but it has to be properly sized in order to operate in stability conditions. Keeping into

consideration the period of the day, of the year and the traffic demand intensity is fundamental in order to properly design this solution.

Concluding, this work shows that RAN resource management is an effective solution to respond to the sustainability and energy efficiency constraints, established by European policies, as well as the requirements which are emerging from the 6G discussions in the communication community. The urgency of the RAN energy efficiency is also due to the need for cooperation in facing the climate change problem, which will characterise the era which we are entering in.
This RAN aspect steers further research efforts in the next year to come. Indeed, in the future, the role of renewable energy, the optimisation of RAN resource management are bound to become even more relevant, in view of the deployment of 5G RANs, as well as the design of the 6G mobile network standard. This brings new directions to be taken. First, the employment of other RES, such as wind, geothermal, etc., is envisioned to become even more relevant because of sustainability requirements, also motivated by the imminent deployment and marketing of more efficient RES infrastructures. In addition, the application of Resource on Demand and Resource on Produced Energy approaches will have to be adapted to the new 6G scenarios. Indeed, the future network will be a large decentralised system, where intelligent decisions are made at different granular levels. To accelerate the learning and improve the reliability, distributed AI will leverage increasingly powerful distributed computation, communication, caching and control resources, meaning that BSs will also provide these features. Hence, the BS activation/deactivation decisions will no longer be based only on traffic demand and/or RES production, posing new challenges to this extent. In this scenario, as already investigated, the RAN can profitably interact with the Smart Grid and become an active actor in the energy market, acting as a prosumer, withdrawing and injecting energy from/in the electric grid. To this extent, dynamic and energy-aware service availability could be suitable, adapting the provided services or their prices, to meet the energy availability in the grid, as well as the locally produced. On the other hand, the deployment of shared distributed resources for computation, communication, caching and control among service providers leads to the employment of new economic models such as shared, or circular, economic, whose usage in this field has to be investigated.
Nevertheless, in this scenario, some challenges arise. While the application of AI in 5G networks is limited to the optimisation of the network architecture, since the

5G network did not take it into account at the beginning of its design, the situation is different for the 6G network. As a consequence, the high penetration of AI is envisioned as the key element, to make the network as smart, agile, and able to learn and adapt itself according to the changing network dynamics as possible. Because of the long training running on a huge number of Graphical Processing Units, using large data-sets, it brings to an ever-growing energy needs [148]. For this reason, in order to achieve the energy efficiency requirement, it will be necessary to include this aspect as a new variable in the design space of the optimisation of the network energy efficiency. Furthermore, with the high penetration of the services provided by RAN, the need for reliable power supply for this network becomes more and more urgent, mainly in emerging countries, whose RAN traffic demand is expected to significantly grow, but which can not rely on a continuous power supply service. Finally, the new quantum computing system has been emerging as the future of high performance computing platforms, able to solve complex problems faster than classical computers, representing a great support for the network. Its study is dawning, even if the investment in this field has increased in the public and private sectors, but its power consumption is one of its many open issues which has to be investigated.

References

[1] Nakjung Choi, Kyle Guan, Daniel C Kilper, and Gary Atkinson. In-network caching effect on optimal energy consumption in content-centric networking. In *2012 IEEE international conference on communications (ICC)*, pages 2889–2894. IEEE, 2012.

[2] Min Chen, Yongfeng Qian, Yixue Hao, Yong Li, and Jeungeun Song. Data-driven computing and caching in 5g networks: Architecture and delay analysis. *IEEE Wireless Communications*, 25(1):70–75, 2018.

[3] Margot Deruyck, Jorg Wyckmans, Wout Joseph, and Luc Martens. Designing uav-aided emergency networks for large-scale disaster scenarios. *EURASIP Journal on Wireless Communications and Networking*, 2018(1):1–12, 2018.

[4] German Castellanos, Margot Deruyck, Luc Martens, and Wout Joseph. Performance evaluation of direct-link backhaul for uav-aided emergency networks. *Sensors*, 19(15):3342, 2019.

[5] Lotfi Belkhir and Ahmed Elmeligi. Assessing ict global emissions footprint: Trends to 2040 & recommendations. *Journal of Cleaner Production*, 177:448–463, 2018.

[6] Bertoldi P and Lejeune A. Code of conduct on energy consumption of broadband equipment. (KJ-NA-30789-EN-N (online)), 2021.

[7] Azeddine Gati, Fatma Ezzahra Salem, Ana Maria Galindo Serrano, Didier Marquet, Stephane Le Masson, Thomas Rivera, Dinh-Thuy Phan-Huy, Zwi Altman, Jean-Baptiste Landre, Olivier Simon, et al. Key technologies to accelerate the ict green evolution–an operator's point of view. *arXiv preprint arXiv:1903.09627*, 2019.

[8] Cisco VNI Forecast. Cisco visual networking index: Forecast and trends, 2017–2022. *White paper, Cisco Public Information*, 2019.

[9] Faisal Tariq, Muhammad RA Khandaker, Kai-Kit Wong, Muhammad A Imran, Mehdi Bennis, and Merouane Debbah. A speculative study on 6g. *IEEE Wireless Communications*, 27(4):118–125, 2020.

[10] Luis M Correia, Dietrich Zeller, Oliver Blume, Dieter Ferling, Ylva Jading, István Gódor, Gunther Auer, and Liesbet Van Der Perre. Challenges and enabling technologies for energy aware mobile radio networks. *IEEE Communications Magazine*, 48(11):66–72, 2010.

[11] Łukasz Budzisz, Fatemeh Ganji, Gianluca Rizzo, Marco Ajmone Marsan, Michela Meo, Yi Zhang, George Koutitas, Leandros Tassiulas, Sofie Lambert, Bart Lannoo, et al. Dynamic resource provisioning for energy efficiency in wireless access networks: A survey and an outlook. *IEEE Communications Surveys & Tutorials*, 16(4):2259–2285, 2014.

[12] Mattia Dalmasso, Michela Meo, and Daniela Renga. Radio resource management for improving energy self-sufficiency of green mobile networks. *ACM SIGMETRICS Performance Evaluation Review*, 44(2):82–87, 2016.

[13] Hakim Ghazzai, Muhammad Junaid Farooq, Ahmad Alsharoa, Elias Yaacoub, Abdullah Kadri, and Mohamed-Slim Alouini. Green networking in cellular hetnets: A unified radio resource management framework with base station on/off switching. *IEEE Transactions on Vehicular Technology*, 66(7):5879–5893, 2017.

[14] Nadhir Ben Rached, Hakim Ghazzai, Abdullah Kadri, and Mohamed-Slim Alouini. A time-varied probabilistic on/off switching algorithm for cellular networks. *IEEE Communications Letters*, 22(3):634–637, 2018.

[15] Klas Johansson, Anders Furuskar, Peter Karlsson, and Jens Zander. Relation between base station characteristics and cost structure in cellular systems. In *2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No. 04TH8754)*, volume 4, pages 2627–2631. IEEE, 2004.

[16] Off-grid power for mobile base stations—renewable and alternative energy sources for remote mobile telecommunications: Global market analysis and forecasts. http://www.pikeresearch.com. Accessed: 2021-10-26.

[17] Dario Pompili, Abolfazl Hajisami, and Tuyen X Tran. Elastic resource utilization framework for high capacity and energy efficiency in cloud ran. *IEEE Communications Magazine*, 54(1):26–32, 2016.

[18] Margot Deruyck, Wout Joseph, Emmeric Tanghe, and Luc Martens. Reducing the power consumption in lte-advanced wireless access networks by a capacity based deployment tool. *Radio Science*, 49(9):777–787, 2014.

[19] Margot Deruyck, Daniela Renga, Michela Meo, Luc Martens, and Wout Joseph. Reducing the impact of solar energy shortages on the wireless access network powered by a pv panel system and the power grid. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–6. IEEE, 2016.

[20] Renewable energy for mobile towers: Opportunities for low- and middle-income countries - september 2020. https://www.gsma.com/mobilefordevelopment/resources/renewable-energy-for-mobile-towers-\opportunities-for-low-and-middle-income-countries/. Accessed: 2021-10-26.

[21] Margot Deruyck, Daniela Renga, Michela Meo, Luc Martens, and Wout Joseph. Accounting for the varying supply of solar energy when designing wireless access networks. *IEEE Transactions on Green Communications and Networking*, 2(1):275–290, 2017.

[22] Mst Rubina Aktar, Abu Jahid, Md Farhad Hossain, and Md Al-Hasan. Energy sustainable traffic aware hybrid powered off-grid cloud radio access network. In *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, pages 121–125. IEEE, 2018.

[23] Afif Osseiran, Federico Boccardi, Volker Braun, Katsutoshi Kusume, Patrick Marsch, Michal Maternia, Olav Queseth, Malte Schellmann, Hans Schotten, Hidekazu Taoka, et al. Scenarios for 5g mobile and wireless communications: the vision of the metis project. *IEEE communications magazine*, 52(5):26–35, 2014.

[24] Yaping Sun, Zhiyong Chen, and Hui Liu. Delay analysis and optimization in cache-enabled multi-cell cooperative networks. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2016.

[25] Emeka E Ugwuanyi, Saptarshi Ghosh, Muddesar Iqbal, Tasos Dagiuklas, Shahid Mumtaz, and Anwer Al-Dulaimi. Co-operative and hybrid replacement caching for multi-access mobile edge computing. In *2019 European Conference on Networks and Communications (EuCNC)*, pages 394–399. IEEE, 2019.

[26] Emna Baccour, Aiman Erbad, Kashif Bilal, Amr Mohamed, and Mohsen Guizani. Pccp: Proactive video chunks caching and processing in edge networks. *Future Generation Computer Systems*, 105:44–60, 2020.

[27] Xiaotong Yang, Xueyong Yu, Hao Huang, and Hongbo Zhu. Energy efficiency based joint computation offloading and resource allocation in multi-access mec systems. *IEEE Access*, 7:117054–117062, 2019.

[28] Mohammad Asif Habibi, Meysam Nasimi, Bin Han, and Hans D Schotten. A comprehensive survey of ran architectures toward 5g mobile communication system. *IEEE Access*, 7:70371–70421, 2019.

[29] Bartlomiej Blaszczyszyn and Anastasios Giovanidis. Optimal geographic caching in cellular networks. In *2015 IEEE international conference on communications (ICC)*, pages 3358–3363. IEEE, 2015.

[30] Konstantinos Poularakis, George Iosifidis, Antonios Argyriou, and Leandros Tassiulas. Video delivery over heterogeneous cellular networks: Optimizing cost and performance. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 1078–1086. IEEE, 2014.

[31] Tuyen X Tran, Abolfazl Hajisami, Parul Pandey, and Dario Pompili. Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges. *IEEE Communications Magazine*, 55(4):54–61, 2017.

[32] Dong Liu, Binqiang Chen, Chenyang Yang, and Andreas F Molisch. Caching at the wireless edge: design aspects, challenges, and future directions. *IEEE Communications Magazine*, 54(9):22–28, 2016.

[33] Rui Deng. Dash based video caching in mec-assisted heterogeneous networks. *MULTIMEDIA TOOLS AND APPLICATIONS*, 2020.

[34] Greta Vallero, Margot Deruyck, Wout Joseph, and Michela Meo. Caching at the edge in high energy-efficient wireless access networks. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2020.

[35] Zhaohui Luo, Minghui LiWang, Zhijian Lin, Lianfen Huang, Xiaojiang Du, and Mohsen Guizani. Energy-efficient caching for mobile edge computing in 5g networks. *Applied sciences*, 7(6):557, 2017.

[36] 3GPP. Study on new radio (nr) to support non-terrestrial networks. 2019.

[37] Gunes Karabulut Kurt, Mohammad G Khoshkholgh, Safwan Alfattani, Ahmed Ibrahim, Tasneem SJ Darwish, Md Sahabul Alam, Halim Yanikomeroglu, and Abbas Yongacoglu. A vision and framework for the high altitude platform station (haps) networks of the future. *IEEE Communications Surveys & Tutorials*, 23(2):729–779, 2021.

[38] Qingqing Wu, Yong Zeng, and Rui Zhang. Joint trajectory and communication design for multi-uav enabled wireless networks. *IEEE Transactions on Wireless Communications*, 17(3):2109–2121, 2018.

[39] Bin Li, Zesong Fei, and Yan Zhang. Uav communications for 5g and beyond: Recent advances and future trends. *IEEE Internet of Things Journal*, 6(2):2241–2263, 2018.

[40] Philipp Oettershagen, Amir Melzer, Thomas Mantel, Konrad Rudin, Thomas Stastny, Bartosz Wawrzacz, Timo Hinzmann, Kostas Alexis, and Roland Siegwart. Perpetual flight with a small solar-powered uav: Flight results, performance analysis and model validation. In *2016 IEEE Aerospace Conference*, pages 1–8. IEEE, 2016.

[41] Scott Morton, Ruben D'Sa, and Nikolaos Papanikolopoulos. Solar powered uav: Design and experiments. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 2460–2466. IEEE, 2015.

[42] T Shankar et al. A survey on techniques related to base station sleeping in green communication and comp analysis. In *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, pages 1059–1067. IEEE, 2016.

[43] Stefano Buzzi, I Chih-Lin, Thierry E Klein, H Vincent Poor, Chenyang Yang, and Alessio Zappone. A survey of energy-efficient techniques for 5g networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(4):697–709, 2016.

[44] Mattia Dalmasso, Michela Meo, and Daniela Renga. Radio resource management for improving energy self-sufficiency of green mobile networks. *ACM SIGMETRICS Performance Evaluation Review*, 44(2):82–87, 2016.

[45] Daniela Renga, Hussein Al Haj Hassan, Michela Meo, and Loutfi Nuaymi. Energy management and base station on/off switching in green mobile networks for offering ancillary services. *IEEE Transactions on Green Communications and Networking*, 2(3):868–880, 2018.

[46] Muhammad Ali, Michela Meo, and Daniela Renga. Cost saving and ancillary service provisioning in green mobile networks. In *The Internet of Things for Smart Urban Ecosystems*, pages 201–224. Springer, 2019.

[47] Jia Guo, Yu Peng, Xiyuan Peng, Qiang Chen, Jiang Yu, and Yufeng Dai. Traffic forecasting for mobile networks with multiplicative seasonal arima models. In *2009 9th International Conference on Electronic Measurement & Instruments*, pages 3–377. IEEE, 2009.

[48] Paulo Cortez, Miguel Rio, Miguel Rocha, and Pedro Sousa. Multi-scale internet traffic forecasting using neural networks and time series methods. *Expert Systems*, 29(2):143–155, 2012.

[49] M Zubair Shafiq, Lusheng Ji, Alex X Liu, and Jia Wang. Characterizing and modeling internet traffic dynamics of cellular devices. *ACM SIGMETRICS Performance Evaluation Review*, 39(1):265–276, 2011.

[50] Paulo Cortez, Miguel Rio, Miguel Rocha, and Pedro Sousa. Internet traffic forecasting using neural networks. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 2635–2642. IEEE, 2006.

[51] Hoang Duy Trinh, Lorenza Giupponi, and Paolo Dini. Mobile traffic prediction from raw data using lstm networks. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1827–1832. IEEE, 2018.

[52] Sebastian Troia, Rodolfo Alvizu, Youduo Zhou, Guido Maier, and Achille Pattavina. Deep learning-based traffic prediction for network optimization. In *2018 20th International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2018.

[53] Leether Yao and Teng-Shih Tsai. Novel hybrid scheme of solar energy forecasting for home energy management system. In *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social*

Computing (CPSCom) and IEEE Smart Data (SmartData), pages 150–155. IEEE, 2016.

[54] Shaojun Wang, Jia Guo, Qi Liu, and Xiyuan Peng. On-line traffic forecasting of mobile communication system. In *2010 First International Conference on Pervasive Computing, Signal Processing and Applications*, pages 97–100. IEEE, 2010.

[55] Huimin Pan, Jingchu Liu, Sheng Zhou, and Zhisheng Niu. A block regression model for short-term mobile traffic forecasting. In *2015 IEEE/CIC International Conference on Communications in China (ICCC)*, pages 1–5. IEEE, 2015.

[56] Luca Chiaraviglio, Marco Listanti, and Edoardo Manzia. Life is short: The impact of power states on base station lifetime. *Energies*, 8(12):14407–14426, 2015.

[57] Luca Chiaraviglio, Francesca Cuomo, Marco Listanti, Edoardo Manzia, and Martina Santucci. Fatigue-aware management of cellular networks infrastructure with sleep modes. *IEEE Transactions on Mobile Computing*, 16(11):3028–3041, 2017.

[58] Luca Chiaraviglio, Francesca Cuomo, Marco Listanti, Edoardo Manzia, and Martina Santucci. Sleep to stay healthy: Managing the lifetime of energy-efficient cellular networks. In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2015.

[59] Carlos Natalino, Luca Chiaraviglio, Filip Idzikowski, Carlos RL Francês, Lena Wosinska, and Paolo Monti. Optimal lifetime-aware operation of green optical backbone networks. *IEEE Journal on Selected Areas in Communications*, 34(12):3915–3926, 2016.

[60] Pawel Wiatr, Jiajia Chen, Paolo Monti, and Lena Wosinska. Energy efficiency versus reliability performance in optical backbone networks. *Journal of Optical Communications and Networking*, 7(3):A482–A491, 2015.

[61] Vinay Chamola and Biplab Sikdar. Solar powered cellular base stations: current scenario, issues and proposed solutions. *IEEE Communications magazine*, 54(5):108–114, 2016.

[62] Hussein Al Haj Hassan, Loutfi Nuaymi, and Alexander Pelov. Renewable energy in cellular networks: A survey. In *2013 IEEE online conference on green communications (OnlineGreenComm)*, pages 1–7. IEEE, 2013.

[63] Tao Han and Nirwan Ansari. Powering mobile networks with green energy. *IEEE Wireless Communications*, 21(1):90–96, 2014.

[64] Michela Meo, Yi Zhang, Raffaella Gerboni, and Marco Ajmone Marsan. Dimensioning the power supply of a lte macro bs connected to a pv panel and the power grid. In *2015 IEEE International Conference on Communications (ICC)*, pages 178–184. IEEE, 2015.

[65] Vinay Chamola and Biplab Sikdar. Resource provisioning and dimensioning for solar powered cellular base stations. In *2014 IEEE Global Communications Conference*, pages 2498–2503. IEEE, 2014.

[66] Vinay Chamola and Biplab Sikdar. A multistate markov model for dimensioning solar powered cellular base stations. *IEEE Transactions on Sustainable Energy*, 6(4):1650–1652, 2015.

[67] Daniela Renga and Michela Meo. Dimensioning renewable energy systems to power mobile networks. *IEEE Transactions on Green Communications and Networking*, 3(2):366–380, 2019.

[68] Meysam Masoudi, Mohammad Galal Khafagy, Ebrahim Soroush, Daniele Giacomelli, Simone Morosi, and Cicek Cavdar. Reinforcement learning for traffic-adaptive sleep mode management in 5g networks. In *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–6. IEEE, 2020.

[69] Mahshid Mehrabi, Dongho You, Vincent Latzko, Hani Salah, Martin Reisslein, and Frank HP Fitzek. Device-enhanced mec: Multi-access edge computing (mec) aided by end device computation and caching: A survey. *IEEE Access*, 7:166079–166108, 2019.

[70] Sunitha Safavat, Naveen Naik Sapavath, and Danda B Rawat. Recent advances in mobile edge computing and content caching. *Digital Communications and Networks*, 6(2):189–194, 2020.

[71] Konstantinos Poularakis, George Iosifidis, and Leandros Tassiulas. Approximation algorithms for mobile data caching in small cell networks. *IEEE Transactions on Communications*, 62(10):3665–3677, 2014.

[72] Min Chen, Yixue Hao, Long Hu, Kaibin Huang, and Vincent KN Lau. Green and mobility-aware caching in 5g networks. *IEEE Transactions on Wireless Communications*, 16(12):8347–8361, 2017.

[73] Karthikeyan Shanmugam, Negin Golrezaei, Alexandros G Dimakis, Andreas F Molisch, and Giuseppe Caire. Femtocaching: Wireless content delivery through distributed caching helpers. *IEEE Transactions on Information Theory*, 59(12):8402–8413, 2013.

[74] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, Young-Han Nam, and Mérouane Debbah. A tutorial on uavs for wireless networks: Applications, challenges, and open problems. *IEEE communications surveys & tutorials*, 21(3):2334–2360, 2019.

[75] Evgenii Vinogradov, Hazem Sallouha, Sibren De Bast, Mohammad Mahdi Azari, and Sofie Pollin. Tutorial on uav: A blue sky view on wireless communication. *arXiv preprint arXiv:1901.02306*, 2019.

[76] Jingjing Wang, Chunxiao Jiang, Zhu Han, Yong Ren, Robert G Maunder, and Lajos Hanzo. Taking drones to the next level: Cooperative distributed unmanned-aerial-vehicular networks for small and mini drones. *Ieee vehIcular technology magazIne*, 12(3):73–82, 2017.

[77] Aziz Altaf Khuwaja, Yunfei Chen, Nan Zhao, Mohamed-Slim Alouini, and Paul Dobbins. A survey of channel modeling for uav communications. *IEEE Communications Surveys & Tutorials*, 20(4):2804–2821, 2018.

[78] Reza Shakeri, Mohammed Ali Al-Garadi, Ahmed Badawy, Amr Mohamed, Tamer Khattab, Abdulla Khalid Al-Ali, Khaled A Harras, and Mohsen Guizani. Design challenges of multi-uav systems in cyber-physical applications: A comprehensive survey and future directions. *IEEE Communications Surveys & Tutorials*, 21(4):3340–3385, 2019.

[79] Lav Gupta, Raj Jain, and Gabor Vaszkun. Survey of important issues in uav communication networks. *IEEE Communications Surveys & Tutorials*, 18(2):1123–1152, 2015.

[80] Mohammad Mozaffari, Ali Taleb Zadeh Kasgari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Beyond 5g with uavs: Foundations of a 3d wireless cellular network. *IEEE Transactions on Wireless Communications*, 18(1):357–372, 2018.

[81] Bander Alzahrani, Omar Sami Oubbati, Ahmed Barnawi, Mohammed Atiquzzaman, and Daniyal Alghazzawi. Uav assistance paradigm: State-of-the-art in applications and challenges. *Journal of Network and Computer Applications*, 166:102706, 2020.

[82] Ursula Challita and Walid Saad. Network formation in the sky: Unmanned aerial vehicles for multi-hop wireless backhauling. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.

[83] Boris Galkin, Jacek Kibilda, and Luiz A DaSilva. Backhaul for low-altitude uavs in urban environments. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.

[84] Mohamed Alzenad, Muhammad Z Shakir, Halim Yanikomeroglu, and Mohamed-Slim Alouini. Fso-based vertical backhaul/fronthaul framework for 5g+ wireless networks. *IEEE Communications Magazine*, 56(1):218–224, 2018.

[85] Jiri Pokorny, Aleksandr Ometov, Pablo Pascual, Carlos Baquero, Pavel Masek, Alexander Pyattaev, Ainoa Garcia, Carlos Castillo, Sergey Andreev, Jiri Hosek, et al. Concept design and performance evaluation of uav-based backhaul link with antenna steering. *Journal of Communications and Networks*, 20(5):473–483, 2018.

[86] Mikhail Gerasimenko, Jiri Pokorny, Tibor Schneider, Jakub Sirjov, Sergey Andreev, and Jiri Hosek. Prototyping directional uav-based wireless access and backhaul systems. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019.

[87] Chen Qiu, Zhiqing Wei, Xin Yuan, Zhiyong Feng, and Ping Zhang. Multiple uav-mounted base station placement and user association with joint fronthaul and backhaul optimization. *IEEE Transactions on Communications*, 68(9):5864–5877, 2020.

[88] Chen Qiu, Zhiqing Wei, Zhiyong Feng, and Ping Zhang. Backhaul-aware trajectory optimization of fixed-wing uav-mounted base station for continuous available wireless service. *IEEE Access*, 8:60940–60950, 2020.

[89] Marie-Josepha Youssef, Joumana Farah, Charbel Abdel Nour, and Catherine Douillard. Full-duplex and backhaul-constrained uav-enabled networks using noma. *IEEE Transactions on Vehicular Technology*, 69(9):9667–9681, 2020.

[90] Md Moin Uddin Chowdhury, Sung Joon Maeng, Eyuphan Bulut, and Ismail Güvenç. 3-d trajectory optimization in uav-assisted cellular networks considering antenna radiation pattern and backhaul constraint. *IEEE Transactions on Aerospace and Electronic Systems*, 56(5):3735–3750, 2020.

[91] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage. *IEEE Communications Letters*, 20(8):1647–1650, 2016.

[92] Achiel Colpaert, Evgenii Vinogradov, and Sofie Pollin. Aerial coverage analysis of cellular systems at lte and mmwave frequencies using 3d city models. *Sensors*, 18(12):4311, 2018.

[93] Sibren De Bast, Evgenii Vinogradov, and Sofie Pollin. Cellular coverage-aware path planning for uavs. In *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. ieee, 2019.

[94] Andrey V Savkin and Hailong Huang. Deployment of unmanned aerial vehicle base stations for optimal quality of coverage. *IEEE Wireless Communications Letters*, 8(1):321–324, 2018.

[95] Haixin Wang and Jianxin Shen. Analysis of the characteristics of solar cell array based on matlab/simulink in solar unmanned aerial vehicle. *IEEE Access*, 6:21195–21201, 2018.

[96] Yan Sun, Derrick Wing Kwan Ng, Dongfang Xu, Linglong Dai, and Robert Schober. Resource allocation for solar powered uav communication systems. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2018.

[97] Yan Sun, Dongfang Xu, Derrick Wing Kwan Ng, Linglong Dai, and Robert Schober. Optimal 3d-trajectory design and resource allocation for solar-powered uav communication systems. *IEEE Transactions on Communications*, 67(6):4281–4298, 2019.

[98] Erol Gelenbe and Omer H Abdelrahman. An energy packet network model for mobile networks with energy harvesting. *Nonlinear Theory and Its Applications, IEICE*, 9(3):322–336, 2018.

[99] Erol Gelenbe. Energy packet networks: Ict based energy allocation and storage. In *International Conference on Green Communications and Networking*, pages 186–195. Springer, 2011.

[100] Ioannis Dimitriou, Sara Alouf, and Alain Jean-Marie. A markovian queueing system for modeling a smart green base station. In *European Workshop on Performance Engineering*, pages 3–18. Springer, 2015.

[101] Eline De Cuypere, Koen De Turck, and Dieter Fiems. A queueing model of an energy harvesting sensor node with data buffering. *Telecommunication Systems*, 67(2):281–295, 2018.

[102] Kishor Patil, Koen De Turck, and Dieter Fiems. Optimal data collection in wireless sensor networks with correlated energy harvesting. *Annals of Telecommunications*, 74(5):299–310, 2019.

[103] Sami Akın and M Cenk Gursoy. On the energy and data storage management in energy harvesting wireless communications. *IEEE Transactions on Communications*, 67(11):8056–8071, 2019.

[104] Vinay Chamola and Biplab Sikdar. Solar powered cellular base stations: Current scenario, issues and proposed solutions. *IEEE Communications magazine*, 54(5):108–114, 2016.

[105] Hussein Al Haj Hassan, Loutfi Nuaymi, and Alexander Pelov. Renewable energy in cellular networks: A survey. In *2013 IEEE online conference on green communications (OnlineGreenComm)*, pages 1–7. IEEE, 2013.

[106] KY Lee, YT Cha, and JH Park. Short-term load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 7(1):124–132, 1992.

[107] Pedro Torres, Hugo Marques, Paulo Marques, and Jonathan Rodriguez. Using deep neural networks for forecasting cell congestion on lte networks: a simple approach. In *International Conference on Cognitive Radio Oriented Wireless Networks*, pages 276–286. Springer, 2017.

[108] Fatma Ezzahra Salem, Azeddine Gati, Zwi Altman, and Tijani Chahed. Advanced sleep modes and their impact on flow-level performance of 5g networks. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pages 1–7. IEEE, 2017.

[109] Gunther Auer, Oliver Blume, Vito Giannini, Istvan Godor, M Imran, Ylva Jading, Efstathios Katranaras, Magnus Olsson, Dario Sabella, Per Skillermark, et al. D2. 3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown. *Earth*, 20(10), 2010.

[110] Luca Chiaraviglio, Pawel Wiatr, Paolo Monti, Jiajia Chen, Josip Lorincz, Filip Idzikowski, Marco Listanti, and Lena Wosinska. Is green networking beneficial in terms of device lifetime? *IEEE Communications Magazine*, 53(5):232–240, 2015.

[111] Svante Arrhenius. About the reaction speed during the inversion of cane sugar by acidic acids. *magazine for physical chemistry*, 4(1):226–248, 1889.

[112] Yung-Li Lee, Jwo Pan, Richard Hathaway, and Mark Barkey. *Fatigue testing and analysis: theory and practice*, volume 13. Butterworth-Heinemann, 2005.

[113] Margot Deruyck, Wout Joseph, and Luc Martens. Power consumption model for macrocell and microcell base stations. *Transactions on Emerging Telecommunications Technologies*, 25(3):320–333, 2014.

[114] Chris Mi and M Abul Masrur. *Hybrid electric vehicles: principles and applications with practical perspectives*. John Wiley & Sons, 2017.

[115] Mohammad Jafari, Glenn Platt, Zahra Malekjamshidi, and Jian Guo Zhu. Technical issues of sizing lead-acid batteries for application in residential renewable energy systems. In *2015 4th International Conference on Electric Power and Energy Conversion Systems (EPECS)*, pages 1–6. IEEE, 2015.

[116] H Gharavi and R Ghafurian. Ieee recommended practice for sizing lead-acid batteries for stand-alone photovoltaic (pv) systems ieee std 1013–2007. In *Proc. IEEE*, volume 99, pages 917–921, 2011.

[117] Aron P Dobos. Pvwatts version 5 manual. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.

[118] Margot Deruyck, Emmeric Tanghe, David Plets, Luc Martens, and Wout Joseph. Optimizing lte wireless access networks towards power consumption and electromagnetic exposure of human beings. *Computer Networks*, 94:29–40, 2016.

[119] Greta Vallero, Margot Deruyck, Michela Meo, and Wout Joseph. Accounting for energy cost when designing energy-efficient wireless access networks. *Energies*, 11(3), 2018.

[120] Preben Elgaard Mogensen and Jeroen Wigard. Cost action 231: Digital mobile radio towards future generation system, final report. In *Section 5.2: On antenna and frequency diversity in GSM. Section 5.3: Capacity study of frequency hopping GSM network*. 1999.

[121] Karthikeyan Shanmugam, Negin Golrezaei, Alexandros G Dimakis, Andreas F Molisch, and Giuseppe Caire. Femtocaching: Wireless video content delivery through distributed caching helpers. *arXiv preprint arXiv:1109.4179*, 2011.

[122] Jie Dai, Zhan Hu, Bo Li, Jiangchuan Liu, and Baochun Li. Collaborative hierarchical caching with dynamic request routing for massive content distribution. In *2012 Proceedings IEEE INFOCOM*, pages 2444–2452. IEEE, 2012.

[123] You-Chiun Wang and Kai-Chung Chien. A load-aware small-cell management mechanism to support green communications in 5g networks. In *2018 27th Wireless and Optical Communication Conference (WOCC)*, pages 1–5. IEEE, 2018.

[124] Tuyen X Tran, Abolfazl Hajisami, and Dario Pompili. Cooperative hierarchical caching in 5g cloud radio access networks. *IEEE Network*, 31(4):35–41, 2017.

[125] Imtiaz Parvez, Ali Rahmati, Ismail Guvenc, Arif I Sarwat, and Huaiyu Dai. A survey on low latency towards 5g: Ran, core network and caching solutions. *IEEE Communications Surveys & Tutorials*, 20(4):3098–3130, 2018.

[126] Hao Wu, Hancheng Lu, Feng Wu, and Chang Wen Chen. Energy and delay optimization for cache-enabled dense small cell networks. *IEEE Transactions on Vehicular Technology*, 69(7):7663–7678, 2020.

[127] Qiaoyang Ye, Beiyu Rong, Yudong Chen, Mazin Al-Shalash, Constantine Caramanis, and Jeffrey G Andrews. User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, 12(6):2706–2716, 2013.

[128] Tao Han and Nirwan Ansari. A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources. *IEEE/ACM Transactions on Networking*, 24(2):1038–1051, 2015.

[129] Lime Microsystem. Lime demonstrates fprf transceivers at mobile world congress shanghai, 2015.

[130] Microdornes. md4-1000: Robust and powerful – UAV / drone model from Microdrones, 2019.

[131] ETSI. ETSI TS 136 212 v14.2.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding (3GPP TS 36.212 version 14.2.0 Release 14). Technical report, ETSI, April 2017.

[132] ETSI. ETSI TS 136 213 v14.6.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 14.6.0 Release 14). Technical report, ETSI, April 2018.

[133] AGCOM Autorita per le Garanzie Nelle Comunicazioni. OSSERVATORIO SULLE COMUNICAZIONI N.4/ 2018. Technical Report N.4/2018, AGCOM - Autorita per le Garanzie Nelle Comunicazioni, 2018.

[134] Wahab Khawaja, Ismail Guvenc, David W Matolak, Uwe-Carsten Fiebig, and Nicolas Schneckenburger. A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles. *IEEE Communications Surveys & Tutorials*, 21(3):2361–2391, 2019.

[135] Akram Al-Hourani, Sithamparanathan Kandeepan, and Abbas Jamalipour. Modeling air-to-ground path loss for low altitude platforms in urban environments. In *2014 IEEE global communications conference*, pages 2898–2904. IEEE, 2014.

[136] 3GPP. 3GPP TR 36.777. 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on Enhanced LTE Support for Aerial Vehicles (Release 15). Technical report, Valbonne - France, 2017.

[137] Sanshan Sun, Wei Jiang, Gang Feng, Shuang Qin, and Ye Yuan. Cooperative caching with content popularity prediction for mobile edge caching. *Tehnički vjesnik*, 26(2):503–509, 2019.

[138] Jianfa Wu, Honglun Wang, Na Li, Peng Yao, Yu Huang, and Hemeng Yang. Path planning for solar-powered uav in urban environment. *Neurocomputing*, 275:2055–2065, 2018.

[139] ETSI. ETSI TS 136 101 v14.5.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (3GPP TS 36.101 version 14.5.0 Release 14). Technical report, ETSI, November 2017.

[140] Erol Gelenbe. A sensor node with energy harvesting. *ACM SIGMETRICS Performance Evaluation Review*, 42(2):37–39, 2014.

[141] PE Oguntunde, OA Odetunmibi, and AO Adejumo. On the sum of exponentially distributed random variables: A convolution approach. *European Journal of Statistics and Probability*, 2(1):1–8, 2014.

[142] Daniela Renga and Michela Meo. Modeling renewable energy production for base stations power supply. In *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 716–722. IEEE, 2016.

[143] Manoj Kumar Panjwani and Ghous Bukshsh Narejo. Effect of altitude on the efficiency of solar panel. *International Journal of Engineering Research and General Science*, 2(4), 2014.

[144] Albert Espinal, Rebeca Estrada, and Carlos Monsalve. Traffic model using a novel sniffer that ensures the user data privacy. In *MATEC Web of Conferences*, volume 292, page 03002. EDP Sciences, 2019.

[145] Margot Deruyck, Emmeric Tanghe, Wout Joseph, Willem Vereecken, Mario Pickavet, Bart Dhoedt, and Luc Martens. Towards a deployment tool for wireless access networks with minimal power consumption. In *2010 IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*, pages 295–300. IEEE, 2010.

[146] Commission implementing regulation (eu) 2019/947 of 24 may 2019 on the rules and procedures for the operation of unmanned aircraft. *OL 152*, L 218:45–71.

[147] Margot Deruyck, Jorg Wyckmans, Luc Martens, and Wout Joseph. Emergency ad-hoc networks by using drone mounted base stations for a disaster scenario. In *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–7. IEEE, 2016.

[148] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*, 2019.

Appendix A

List of Acronyms

3GPP Third Generation Partnership Project

AF Accelerator Factor

AI Artificial Intelligence

AME Average Mean Error

ANN Artificial Neural Network

AP Access Point

ARE Average Relative Error

ARIMA Auto-Regressive Integrated Moving Average

BH Backhaul

BLR Block Linear Regression

BS Base Station

CN Core Network

CTMC Continuous Time Markov Chain

DFD Descending Front Detection

DOD Depth of Discharge

DP Data Packet

DRAM Dynamic Random Access Memory

EP Energy Packet

HW Hardware

ICT Information and Communication Technology

KPI Key Performance Indicator

LFU Least Frequently Used

LSTMC Long Short Term Memory Cell

LTE Long Term Evolution

LoS Line-of-Sight

MANO Management and Orchestration System

ME Mean Error

MEC Multi-access Edge Computing

ML Machine Learning

MNO Mobile Network Operators

MSE Mean Squared Error

MU-MIMO Multi User - Multiple Input Multiple Output

NLoS Non-Line-of-Sight

OPEX Operational Expenditure

PV Photovoltaic

QoS Quality of Service

RAN Radio Access Network

RB Resource Block

RE Relative Error

RES Renewable Energy Source

RNN Recurrent Neural Network

RoD Resource on Demand

RoPE Resource on Produced Energy

SA Seasonal ARIMA

SISO Single Input Single Output

UAV Unmanned Aerial Vehicles

UAV-BS Unmanned Aerial Vehicles - Base Station

WI Walfish Ikegami

Appendix B

List of Co-Authored Publications

Journal Publications

- Vallero, G., Deruyck, M., Meo, M., & Joseph, W. (2018). Accounting for energy cost when designing energy-efficient wireless access networks. *Energies*, 11(3), 617.
- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2019). Greener RAN operation through machine learning. *IEEE Transactions on Network and Service Management, 16*(3), 896-908.
- Vallero, G., Renga, D., Meo, M., & Marsan, M. A. (2021). RAN energy efficiency and failure rate through ANN traffic predictions processing. *Computer Communications*.
- Vallero, G., Deruyck, M., Meo, M., & Joseph, W. (2021). Base Station switching and edge caching optimisation in high energy-efficiency wireless access network. *Computer Networks*, 192, 108100.
- Castellanos, G., Vallero, G., Deruyck, M., Martens, L., Meo, M., & Joseph, W. (2021). Evaluation of flying caching servers in UAV-BS based realistic environment. *Vehicular Communications*, 32, 100390.

Conference Publications

- Donevski, I., Vallero, G., & Marsan, M. A. (2019, April). Neural networks for cellular base station switching. In I*EEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 738-743). IEEE.

- Vallero, G., Deruyck, M., Joseph, W., & Meo, M. (2020, June). Caching at the edge in high energy-efficient wireless access networks. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.

- Vallero, G., Pristeri, E., & Meo, M. (2020, June). Coping with power outages in mobile networks. In *2020 Mediterranean Communication and Computer Networking Conference (MedComNet)* (pp. 1-4). IEEE.

- Vallero, G., Renga, D., Meo, M., & Ajmone Marsan, M. (2020, November). Processing ANN Traffic Predictions for RAN Energy Efficiency. In *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (pp. 235-244).

- Vallero, G., & Meo, M. (2021, June). Modelling Solar Powered UAV-BS for 5G and Beyond. In *2021 19th Mediterranean Communication and Computer Networking Conference (MedComNet)* (pp. 1-8). IEEE.

- Vallero, G., & Meo, M. (2021). Hybrid Energy Production Analysis and Modelling for Radio Access Network Supply. In *SMARTGREENS* (pp. 131-141).