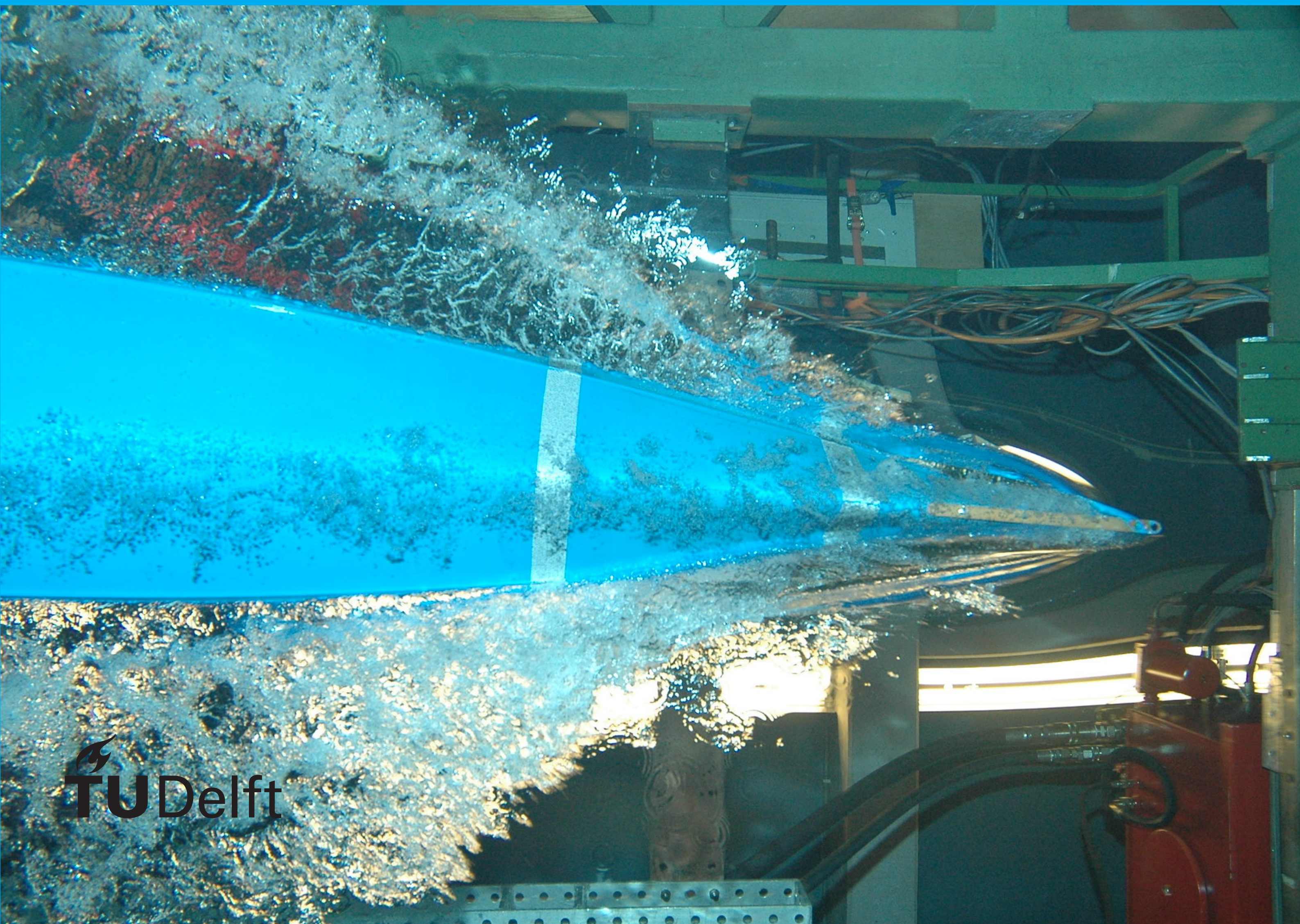
Assessment of spatio-temporal traffic dynamics in mobile networks and its exploitation towards operational cost reduction

Rochal Saxena

TU Delft
Network Architectures and Services
KPN



Assessment of spatio-temporal traffic dynamics in mobile networks and its exploitation towards operational cost reduction

by

Rochal Saxena

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday November 29, 2018 at 2:00 PM.

Student number:	4576551	
Project duration:	January 1, 2018 – October 18, 2018	
Thesis committee:	Dr. Remco Litjens MSc	TNO, TU Delft
	Dr. ir. Gerard Janssen	TU Delft
	Ir. Rob Buckers	KPN

This version of the thesis is public.

TU Delft

Important Note

This is the public version of the thesis. Proprietary information which belongs to KPN has been redacted in this version of the thesis. If you have any question(s) related to the thesis, please contact either:

Rochal Saxena (TU Delft)	**email:** r.saxena@student.tudelft.nl
or	
Rob Buckers (KPN)	**email:** rob.buckers@kpn.com

Preface

This thesis marks the completion of my master studies at TU Delft. This research project was carried out at KPN which is one of the leading mobile service providers in The Netherlands. It took me approximately twelve months to complete this research project and it has been an incredible journey throughout.

This thesis would not have been possible without the help and support from various people. Firstly, I would like to thank my supervisor Remco Litjens for introducing me to this research project at KPN and for his critical feedback and suggestions for improving the quality of the research project. I would also like to thank Rob Buckers at KPN for agreeing to guide me through my graduation assignment, for the long and interesting discussions on whether the scientific theory matches the reality as observed in the KPN network and for his patience in explaining to me the ENIQ tooling used to access the network data. I would also like to thank Gerard Janssen for productive discussions related to the working of a MIMO system and for agreeing to be a part of the thesis committee. I would also like to thank everyone in the Mobile Quality department and various other people at KPN for helping me with practical matters or explaining the inner working of some subsystems, some of these people are Wim van Blitterswijk, Ashley Beynon, Gerard de Groot, Zoubir Irahhauten, Daan Helming, Rob Hendriks, Marcon de Vrede, Aad Westerman, Royston Wesley.

Finally, I would like to thank my parents for their support throughout this journey and for believing in me

Rochal Saxena
Delft, November 2018

Abstract

The environmental policy of KPN has an objective of reducing CO2 emissions and energy consumption to contribute in making the planet more sustainable. This thesis project researches the potential for reducing energy consumption and, consequently, the operational costs of the KPN LTE network. There are various resources in a mobile network which consume energy and switching off these resources will lead to a reduced energy consumption. The trade-off in switching off these resources involves a potential impact on the capacity and the quality (throughput) offered by the mobile network. In this thesis we quantify the attainable savings in operational cost and/or energy consumption while still satisfying the quality of service requirements set by KPN. Five different operational cost saving schemes are analyzed on three different timescales in this research. The results of this research indicate that there is ample opportunity present for KPN to reduce their operational cost and/or energy consumption with minimal impact on the quality of service. The results further indicate that maximum savings are attainable in the live KPN network by effectuating the 'turning off carriers' scheme, in which the capacity carrier is turned off at the base stations. The reduction in energy consumption and operational cost on effectuating this energy reduction scheme is in the range of 21-70% with respect to the reference scenario for each of the timescale. Furthermore, the results also indicate that of the five operational cost saving schemes investigated, the 'turning off carriers' scheme has the lowest impact on the experienced quality of service.

Acronyms

3GPP	3^{rd} Generation Partnership Project
AA	Active Antennas
AC	Alternating Current
AM	Acknowledged Mode
BSC	Base Station Controller
CDMA	Code Division Multiple Access
CLSM	Closed Loop Spatial Multiplexing
CoMP	Coordinated Multipoint Transmission
CSI	Channel State Information
DC	Direct Current
DL	Downlink
DRB	Data Radio Bearers
DTX	Discontinuous Transmission
EARTH	Energy Aware Radio and neTwork tecHnologies
eNB	eNodeB
EPC	Evolved Packet Core
EU	European Union
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FDD	Frequency Division Duplex
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat Request
HSDPA	High Speed Downlink Packet Access
HSS	Home Subscription Server
IMS	Internet Protocol Multimedia Subsystem
KPI	Key Performance Indicator
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
MBSFN	Multicast-broadcast Single Frequency Network
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
MRC	Maximal Ratio Combining
MS	Mobile Station
OFDMA	Orthogonal Frequency Division Multiple Access
OLSM	Open Loop Spatial Multiplexing
PDCP	Packet Data Convergence Protocol
PDN-GW	Packer Data Network Gateway
PMI	Precoding Matrix Indicator
PRB	Physical Resource Block
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RNC	Radio Network Controller
Rx	Receiver
SDU	Service Data Unit

SFBC	Space Frequency Block Code
S-GW	Serving Gateway
SIMO	Single Input Multiple Output
SINR	Signal to Interference plus Noise Ratio
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SRB	Signaling Radio Bearers
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TMA	Tower Mounted Amplifier
TTI	Transmission Time Interval
Tx	Transmitter
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunications System

Contents

1 Introduction

Mobile traffic has been growing exponentially over the past years for many reasons, such as, an increase in number of mobile devices and connections (phones, tablets and wearable devices), and increase in video streaming traffic. Globally, the increase in mobile data traffic in 2017 was 57% [1]. The ever-increasing mobile data traffic poses various challenges to network operators as they have to ensure that there is sufficient capacity in their respective networks to meet the minimum quality of service requirements, which in turn are also increasing over time, e.g. related to user throughput and call setup success rates. Some methods to increase capacity in a network require investments in the form of addition of new sites, increasing bandwidth, or deploying new radio features to enhance spectral efficiency.

At the same time, the network operators also incur operating expenses to maintain the mobile network and its components. Energy costs compose a significant portion of the operating expenses incurred by the mobile network operator [2]. The energy consumption at the base stations is up to 70% of the total energy consumption in the whole network [3], and is dependent on factors like cooling, traffic volume and energy efficiency of base station components.

The massive increase in mobile data traffic in a network as depicted in [1] is typically seen on a large time scale varying from a few months to a year. However, also on the smaller time scales there is a significant variation observed in the dimensions of space and time. Network operators have various resources (sites, carriers, antenna ports) which can be dynamically allocated in response to these variations.

Spatio-temporal variation in traffic volume in an LTE mobile network and its exploitation towards operational cost reduction is the core focus of this report. Mobile traffic volume in a mobile network varies in space and time differently on different timescales. In this report the focus is on the hourly, daily, and monthly timescale. At these timescales, the resource allocation may be meaningfully adapted to reduce operational cost. The relevance of the higher timescales of daily and monthly is primarily in the assessment of the attainable gains, and not in how to achieve them. The tooling to adapt resource allocation remains the same across all three timescales. Some of the schemes which will be discussed in the following sections in this report are applicable only at a specific timescale as the allocation of some resources, e.g. sites, cannot be adapted at lower timescales. A business-use case for the network operator is developed to reduce the operational costs by exploiting the information obtained from the spatio-temporal traffic variations. This is done primarily via reducing the energy consumption of base stations by temporarily switching off resources during periods of low traffic volume, thus saving on energy costs.

1.1. Approach and Scope

Mobile traffic volume is a metric which provides invaluable information to a mobile network operator about the traffic distribution across the network. This information is important feedback to the network operator which can potentially affect the planning of new sites, network configuration (e.g. related to sleep modes), and traffic steering between the different frequency carriers to optimize the mobile network.

Mobile networks are dimensioned according to the busy-hour data usage (peak traffic). However, traffic volume across the network varies with space and time for various reasons, e.g.

1. Hourly, daily, monthly routine of the user, e.g. daily work-sleep schedule, weekly entertainment, shopping schedule, vacations
2. Events such as live sports events or music festivals
3. Pricing of the data bundles
4. Power outage or maintenance at the base station

The scope of this report is limited to periodic traffic volume variations on the hourly, daily and monthly timescale, specifically the variation observed during the daytime-nighttime hours, the weekends-weekdays on the daily timescale, and the Summer-vacation months of July and August on the monthly timescale. The periodic variations are primarily driven by aspects of human behavior (see first two points above). The emphasis is only on the periodic variations as the non-periodic variations are random and impossible to predict [4], and a prediction for mobile traffic data is required as the schemes under consideration for operational cost reduction cannot be deployed instantaneously. Majority of the schemes under consideration can only be deployed not less than 30 minutes after a variation has occurred as there is a 30 minute lag before the traffic volume information becomes available to the network operator (KPN). This research assumes that the energy saving schemes will be deployed and managed by the network operator itself.

There are a few energy saving features offered by the equipment vendor (Ericsson) to the network operators (KPN), e.g. cell sleep mode which switches off the cell after a certain period of inactivity in the cell. These energy saving features can change the network resource allocation faster as the information about traffic volume variation is available to the equipment vendor before it becomes available to the operator (exact time duration is unknown as the detailed working of these features is not made public by the vendors). However, these features have a license fee associated with them which are significant. If the objective of the operator is to reduce operating costs, and the costs savings arising from the implementation of these features is lower than the license fee, this does not lead to any gain for the operator. However, if the objective of the operator is to reduce energy consumption irrespective of the impact on operating costs, then the operator may still purchase the features and implement them. Two of the schemes discussed and analyzed in this thesis are similar to two energy saving features offered by the vendor. It cannot be said with certainty that the schemes are identical to the features offered by the vendor as the detailed workings of the features are not made public. The attainable gains in operational cost reduction are analyzed for the two schemes similar to the features in this thesis. The attainable gains for the two schemes can be compared against the license fee for the two features offered by the operator. However, this part is not covered in this thesis as the license fees are established after negotiations between the operator and the vendor.

Irregular variations in mobile traffic can be predicted to a very small extent by complementing the traffic volume information with another dataset which includes the activities and events occurring at the sites. However, such a dataset would have to be created from scratch by the network operator. It is not feasible to create a dataset listing all the events and activities (e.g. shopping malls closed for maintenance, small scale parties which are hosted at city centers) scheduled to take place at all the radio base station sites. Hence, it is impossible to predict the irregular variations. However, large scale music festivals (for e.g. Defqon 1, Lowlands) are manually tracked by KPN and capacity is temporarily added to the festival sites to cope with the huge influx of users.

The trends observed from the traffic volume variation during these specific time periods are exploited to reduce operating costs for the operator via five different schemes under consideration:

1. Turning off carriers: A site in a mobile network can have more than one frequency carrier present. If multiple carriers are present, then some of the carriers may be turned off during periods of low traffic volume. This leads to a reduction in energy consumption as unused carriers consume more energy than turned-off carriers. Even at zero load, there is a fixed amount of energy consumed and this consumption is lower for a turned-off carrier (see Section 4.1.1). Operational costs savings in this scheme arise from the reduction in energy costs due to reduced energy consumption. This scheme is similar to the ‘cell sleep mode’ energy saving feature offered by the vendor.

2. Antenna muting: LTE employs MIMO (Multiple Input Multiple Output) to improve system performance (in the current implementation using either open-loop spatial multiplexing to increase data rates or transmit diversity to increase link reliability). The KPN network currently employs 2×2 MIMO. In this scheme one of the two antennas is switched off in the cell, and operational costs are reduced due to a reduction in energy consumption. Switching off one of the antennas in the cell may reduce the bit-rate and/or the SINR, but as long as both the QoS reductions are above the minimum target, the reduction is acceptable and the energy consumption reduction welcome. This scheme is similar to the 'MIMO sleep mode' energy saving feature offered by the vendor.

3. Discontinuous transmission (DTX): The frame structure for LTE has provisions for allocating some of the sub-frames to MBSFN (Multicast-broadcast Single Frequency Network). By design, sub-frames allocated to MBSFN do not have to transmit compulsory reference signals. In this scheme, the transmission structure of an LTE frame is changed to include sub-frames allocated to MBSFN so that fewer signals are transmitted. Transmitting fewer signals leads to reduced energy consumption, and therefore a reduction in operational costs.

4. Reducing license capacity: A License from the equipment vendor to the network operator is an artificial resource in the form of a software limitation. It puts a limit on the capacity of the system. A license with a higher capacity limit costs more than a license with a lower limit. Operational costs savings in this scheme arise from reduced license costs.

5. Ericsson's so-called 'Psi Coverage' solution: A site in a mobile network is typically divided into three sectors. Each sector has a radio unit associated with it. A radio unit is essentially a radio transceiver with some processing capabilities. Therefore, in standard practice, each site in a mobile network has three radio units present. In this scheme, at low traffic volumes the site is reconfigured to have a single (rather than three) radio unit serve all three sectors by the means of a splitter. The operational costs savings in this scheme arise from the reduction in energy costs due to reduced hardware. This scheme is a hardware solution, and can only be effectuated at the monthly timescale as the sites cannot be reconfigured at the hourly and daily timescales.

The trade-off for all the five schemes mentioned above is reduced capacity and throughput. In mobile networks there is an inherent trade-off between coverage, capacity, and quality of service metrics. E.g. given the same network deployment and hence resource, an increase in the quality of service targets would lead to a reduction in capacity and/or coverage. The schemes mentioned above impact the capacity and quality of service aspects of the mobile network adversely. If the minimum quality of service target is still met on implementation of the scheme, then the trade-off may be acceptable to the network operator (covered in detail in Chapter 4). The operational cost reduction schemes will be referred to as schemes from this point forward.

To fulfill performance-based quality requirements, KPN has to dimension its LTE network. The dimensioning of the radio network is based on busy hour traffic, and it would lead to adding capacity (e.g. frequency carriers, sites) when the quality requirements are not fulfilled (i.e. throughput requirements, see Section 2.7). Since the offered traffic volume is not constant throughout the day, there will be excess capacity present during the non-busy hours. The traffic volume dynamic varies per site depending on the location (e.g. city centers, rural areas, urban areas) on all three timescales. The objective of KPN is to make their radio network efficient by matching the capacity provided to the actual traffic demand (e.g. switch off resources during low traffic periods) at each time instant.

As the schemes mentioned above cannot be applied instantaneously in response to change in traffic volume, prior information about the trends in traffic volume at the three timescales is needed. From the traffic profile information, it is observed that the traffic volume is periodically at its lowest levels during *fixed time intervals* on all three timescales (covered in detail in Chapter 3). To provide a uniform network-wide solution, the schemes may only be applied during these fixed time intervals (explained in Section 3.1.2). As the traffic dynamics are different for rural and urban sites, for some rural sites, it may be possible to apply the schemes outside this fixed time interval while still meeting minimum QoS requirements, whereas it may not be possible to do the same for an urban site. A different solution for each site (e.g. a different number and/or range of hours of scheme application) makes the analysis less generic and more tailor made. A tailor made solution for each site requires KPN to invest significant time and resources into implementing the tooling required for the schemes and may not lead to any overall gain in operational cost reduction. Presently, a tailor made solution is not feasible for KPN due to the complexity of the tooling. Furthermore, as scheme(s) activation would take 30 minutes at the minimum after a change in traffic volume is observed, the traffic volume levels may have already changed by the time the scheme(s) gets effectuated. Therefore, the application of schemes is limited to the fixed intervals of time (these intervals last from eight hours to ten months, see Chapter 3) which are long enough for the schemes to be implemented without any significant change in traffic volume levels by the time the scheme becomes effectuated. The objective of this research is limited to applying the schemes during the fixed time intervals when traffic volume is periodically at its lowest levels, and effectuating the schemes outside these fixed intervals of time is outside of the scope of this research.

The scope of this research is also limited to the downlink (link from base station to user) as the uplink (link from user to base station) energy consumption is extremely low when compared to downlink (only 3% of total energy consumption of a radio base station when considering power amplifier, baseband, analog, control and supply components). Therefore, the operating cost savings attainable in uplink are very small when compared with downlink.

A mobile network consists of a combination of macro-cells and small cells. In the KPN network, only [redacted] of the sites are small cell sites. Macro-cells provide coverage across a large area while small cells have a range of few tens of meters. Small cells are usually deployed indoors to provide coverage. The radio base station(s) deployed at macro-cells and small cells are different. The small cells employ a different type of radio unit (different from macro-cells), and do not have cell sectorisation, the use of multiple carriers and MIMO. For these reasons, small cells (indoor sites) are excluded from this research as three out of five schemes cannot be deployed at them (turning off carriers, antenna muting, Ericsson psi coverage solution).

4G supports both Frequency Division Duplex (FDD) and Time Division Duplex (TDD). Only [redacted] of the carriers in the KPN Network operate in the TDD mode. Carriers are operated in the TDD mode to mainly meet the spectrum license conditions, and for this reason these carriers cannot be turned off. Furthermore, majority of the carriers [redacted] in the KPN 4G LTE network operate in the FDD mode. Therefore, the scope of this thesis is limited to the FDD carriers.

1.2. Motivation for Research

The research direction of this thesis was motivated by the evolving environmental policy of KPN which has an objective of reducing CO2 emissions and the energy footprint. To realize this goal, observations made by KPN during the summer vacation period in which the traffic volume drastically increased for a few sites was analyzed. This observation by KPN is the primary motive for discovering whether there are periodic changes in traffic volume in the network at different timescales. With this information, the operator can modify the resource allocation in response to the periodic changes and reduce the energy consumption via the schemes mentioned in Section 1.1. Reduction in energy consumption automatically causes a reduction in operational costs and is an added benefit for the operator.

1.3. Research Questions and Methodology

Two main research questions are answered by this research:

1. The significance of periodic traffic volume variation is measured by quantifying the variation for the three timescales under consideration: hourly, daily, and monthly. Quantifying the variation will provide insight into the fraction of total sites which exhibit periodic variations which are big enough to be worthwhile from an energy saving perspective.

2. Analysis of the potential schemes that can be employed by an operator to reduce energy consumption and other aspects of operating costs while meeting the minimum quality of service conditions. Examine the benefits and drawbacks of the different schemes, the trade-off between the provided quality of service and attained cost savings, and total costs saving possible for each scheme at different timescales.

The objective of the analysis is to quantify the trade-off between attainable cost savings and the minimum QoS requirement, and make a recommendation to the operator regarding what would be the most optimal scheme to manage the network resource allocation in response to a change in traffic volume. The different schemes are not combined and analyzed as there is an overlap in the resources that the different schemes turn-off or reduce. For e.g. turning off a carrier also includes turning off [redacted] the antennas (antenna muting turns off [redacted] antenna). Therefore, combining the different schemes will not necessarily result in additional gains in operational cost savings.

For the first research question, a measurement-based analysis of spatio-temporal traffic volume variation in a LTE mobile network is carried out using network counters (see Section 2.6). The methodology for the second research questions consists of a model-based analysis of the inter-relation between traffic volume, resource utilization, throughput and energy consumption.

This research is highly relevant for mobile network operators as it provides them with key insight into how the traffic volume at certain sites varies over the course of day, a week and a year. It also quantifies the significance of the traffic volume variation at these sites. The traffic volume variations provide the operator with an opportunity to reduce operating costs by implementing the five schemes mentioned above. The results of this research will provide the operator with an estimate of attainable cost savings and the trade-offs associated with each scheme.

1.4. Related Work

Mobile traffic has been researched with different objectives and using different methods previously in literature. The primary metric of interest is the 'total traffic volume', i.e the amount of data transmitted in the downlink. In [4] the traffic profile of a mobile network on three different timescales, hourly, daily and weekly is investigated for forecast purposes using a time series approach. An important result in [4] is that the irregular (non-periodic) traffic volume variations in a mobile network are random and thus impossible to predict. In [5], the spatio-temporal variation in mobile traffic is studied to find the bounds on the predictability of mobile traffic. The results obtained using information theory tools in this research shows that in isolation, individual traffic demand can be predicted with an accuracy of 85%.

Extensive research has been done to make the mobile network more energy efficient based on different criteria. A two-year long (2010 to 2012) EU project named "Energy Aware Radio and neTwork tecHnologies" [6] was undertaken by a consortium of major stakeholders with a project target of reducing energy consumption by at least 50%. Various solutions ranging from hardware innovations to radio interface algorithms were investigated and developed in this project. Hardware solutions which improved the energy efficiency of power amplifiers and transceivers were investigated in [7]. The results indicated that the energy efficiency of the components can be improved by 30% to 55% using different techniques. Another hardware solution is investigated in [6, 8], the energy efficiency of low loss foam printed antennas is investigated and the results indicate an average energy efficiency improvement of approximately 15% over commonly printed array solutions. In [6, 8] the concept of "Active Antennas" is also introduced. 'Active antennas' are explained in [6] as "the next step in the evolution of radio access networks (RANs) is related to Active Antennas (AA) where signal conversion steps, filtering, active RF parts and the radiating elements become integrated into a new device fed by optical fibers". Active antennas have improved energy efficiency due to a reduction in coaxial cable losses along the Tx/Rx chain [6]. The implementation of beam-forming on active antennas is investigated in [8] and the results indicate and energy efficiency improvement of 10% to 40%.

Various radio resource management solutions (software solutions) have also been investigated to improve the energy efficiency of mobile networks. In [9], an energy-aware adaptive sectorisation scheme is analyzed in which the sectorisation of the base-stations is adpated in response to a change in traffic volume. The sectorisation is adapted by switching-off some sectors and changing the beam-width for the remaining sectors. The results of this study show that energy consumption can be reduced by at least 21% by using this proposed scheme. In [10], a scheduler is developed with the objective of reducing energy consumption and the simulations indicate the energy consumption can be reduced by 25-40%. An extensive research into the the energy efficiency of mobile networks in 2020 when compared with a 2010 baseline is performed in [11]. The results obtained in this research indicate an energy efficiency improvement factor of approximately 793. Various scenario aspects such as traffic growth, hardware evolutions and dynamic adaptation of network resources are considered in this research. In [12] the concept of 'cell zooming' is introduced. 'Cell zooming' is described as the process of adaptively changing the cell size based on traffic volume and user requirements. Various algorithms which implement cell zooming are developed and analyzed in [12], and the results indicate that energy consumption can be significantly reduced on implementing these algorithms.

The traffic volume variation in a mobile network was considered for switching on-off cells in [13] and the results indicated that energy consumption can be reduced by 4.13% in a 3G mobile network. In [14], an "energy-oriented network optimisation" scheme for an UMTS/HSDPA network was developed and assessed, the scheme switched off sites in off-peak hours and the results indicate that energy savings of up to 40% are attainable. In [15] an extensive model aimed at quantifying the energy efficiency of a mobile network is developed for various base station models and long-term traffic models. In [16] various schemes to improve the energy efficiency in a mobile network are discussed ranging from Ericsson's Psi-Coverage solution to antenna muting. These solutions are targeted at the tail end of the traffic distribution, that is sites which have the lowest volume traffic. In [17] the concept of Discontinuous transmission (DTX) is introduced to reduce energy consumption with potential gains of up to 90% reduced energy consumption possible compared to a system without discontinuous transmission. Discontinuous transmission exploits the adaptable frame structure of LTE to reduce energy consumption by transmitting fewer signals. In [18] the impact on energy consumption via antenna muting is investigated. The results in this study indicate that energy consumption can be reduced up to 50% in a low traffic volume scenario without affecting Quality of Service. Simulation were carried out for 4×4 MIMO and 2×2 MIMO scenario in this research. In [6], the scheme of base station cooperation was also investigated. With the use of Coordinated Multipoint Transmission (CoMP) it was discovered that energy consumption can be reduced by 8% to 15% under different CoMP scenarios.

1.5. Thesis Contribution

The existing research done on the development of energy saving schemes make an assumption that radio management resources can be switched on-off near instantaneously in response to a change in traffic volume, and thus there would be negligible to minimal trade-off in terms of impact on quality of service metrics. While these results may hold in small-scale simulations under controlled environments, it is unknown if these schemes/algorithms will lead to similar results if deployed in a live network. Furthermore, the assumption of near instantaneous switching on-off of resources is not valid in real live networks as changing the resource allocation takes a significant amount of time (at least 30 minutes for KPN), and therefore there is a trade-off present. Also, the existing research is based on simulations in which assumptions are made about the traffic dynamics. The assumptions that are made about the traffic dynamics do not always hold in live networks, and therefore the energy savings attainable in a real live network can be different from the simulations results obtained in the existing research. In some cases, the existing research is also contingent on future developments in hardware technology and this makes some of the results in existing research inapplicable to present day mobile networks. One downside of this research is that it is specific to the KPN LTE network. However, the methodology developed in assessing the operational cost reduction schemes in this report can be applied to any LTE network(s) with changes primarily occurring for the QoS constraints and the traffic dynamics. In this sense, the results of this research can vary considerably for a different LTE network, unless the different LTE network considered was similar to the KPN LTE network with regards to QoS constraints and network layout.

The existing research on energy savings in mobile networks is targeted towards negligible to minimal QoS impact. This is very different from the analysis presented in this report which has an objective of quantifying the trade-off between attainable operational cost (and energy) savings and the QoS targets. The main contribution of this thesis is in assessing the attainable energy savings (and operational cost) gains under QoS constraints in a real live LTE network, and this has not been covered by existing research. This objective is achieved by modeling the relationship between traffic volume, resource utilization and throughput for the five different schemes. Furthermore, this research also presents the cost savings potential of some of the energy saving features offered by the equipment vendor (cell sleep mode, MIMO sleep mode which are similar to schemes 1 and 4 mentioned in Section 1.1). Based on the results of this research, the network operator can decide on whether to purchase these features and implement them.

1.6. Thesis Structure

The remainder of the thesis is organized into four different chapters. Chapter 2 gives a brief overview of LTE (e.g frame structure, MIMO). It also covers the Key Performance Indicators relevant to this research. Chapter 3 and its subsections analyze the spatio-temporal traffic variation at the three different timescales. Chapter 4 analyzes the five different schemes in detail and presents the trade-off between the provided quality of service and the attained cost savings. Chapter 5 presents the conclusion and the future work possibilities.

2
Overview of LTE and KPIs

As this research is only based on the LTE/LTE-A mobile network, it is crucial to understand how a LTE network functions. A high-level overview of the functioning of a LTE network and its key components along with Key Performance Indicators relevant to this research are presented in this chapter.

2.1. LTE Objectives

The acronym LTE stands for Long-Term Evolution with LTE-A standing for Long Term Evolution Advanced. The air interface of LTE is referred to as E-UTRAN which is “Evolved UMTS Terrestrial Radio Access Network”. It is developed according to the requirements laid down by 3GPP (3rd Generation Partnership Project) in [19] with the main goals as follows:

1. Improved spectrum efficiency compared to 3G
2. Increased peak throughput, 100 Mbps in the downlink and 50 Mbps in the uplink
3. Improved cell edge bitrate
4. Reduce radio access network latency to below 10 ms

To meet these goals several changes were made to the architecture of LTE compared to GSM and UMTS.

A brief description of some key terms in a cellular network is given below:

- eNodeB: Base station in LTE/LTE-A
- UE: User Equipment (mobile device)
- Downlink: The radio link from the base station (eNodeB) to the mobile device (UE) is called the downlink.
- Uplink: The radio link from the mobile device to the base station is called the uplink.

2.2. LTE Architecture

LTE is based on the spectrally efficient OFDMA (Orthogonal Frequency Division Multiple Access) technology in which orthogonal sub-carriers with a spacing of Δf=15 [kHz] are assigned to individual users. This allows multiple users to transmit data simultaneously on different sets of sub-carriers with no intra-cell interference. The high spectrum efficiency in LTE is achieved due to a combination of factors, such as, the use of OFDMA, MIMO and high order modulation schemes.

The LTE architecture is also simplified (compared with GSM and UMTS), with the elimination of the Radio Network Control unit, whose functions are transferred to the eNodeB. This also results in a reduction in latency in the radio access network. Furthermore, all services are in the packet-switched domain only with circuit-switching domain eliminated. Figure 2.1 shows the architecture of different radio technologies from

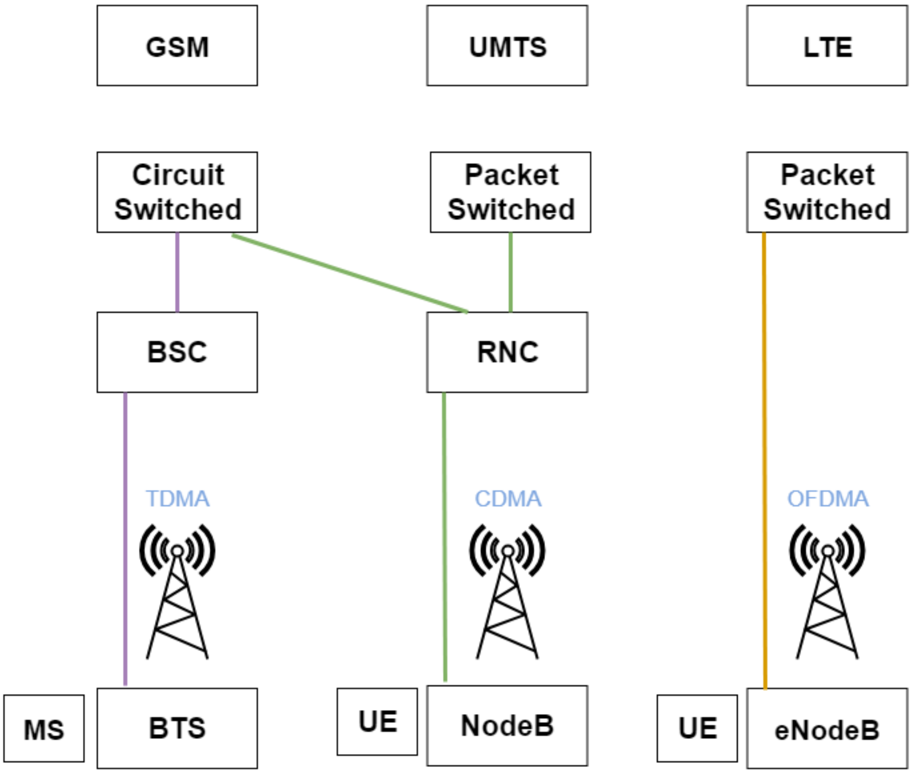


Figure 2.1: Architecture from GSM to LTE

GSM to LTE.
As evident from Figure 2.1, LTE has a more simplified architecture compared to earlier radio technologies with the removal of Radio Network Controller and the circuit switched core. Figure 2.2 shows the components of the Evolved Packet Core in LTE. The main components and their functions in EPC as defined in [21] are:

1. MME: Mobility Management Entity; the MME handles the control plane functions relating to subscriber and session management

2. HSS: Home Subscription Server; the HSS oversees storing and updating user profile information, identification, and addressing

3. PDN-GW: Packet Data Network Gateway; the PDN-GW connects the EPC to the external packet data networks

4. S-GW: Serving Gateway; the S-GW is the termination point of packet data interface towards E-UTRAN, also acts as a local mobility anchor for handovers between different eNodeB's.

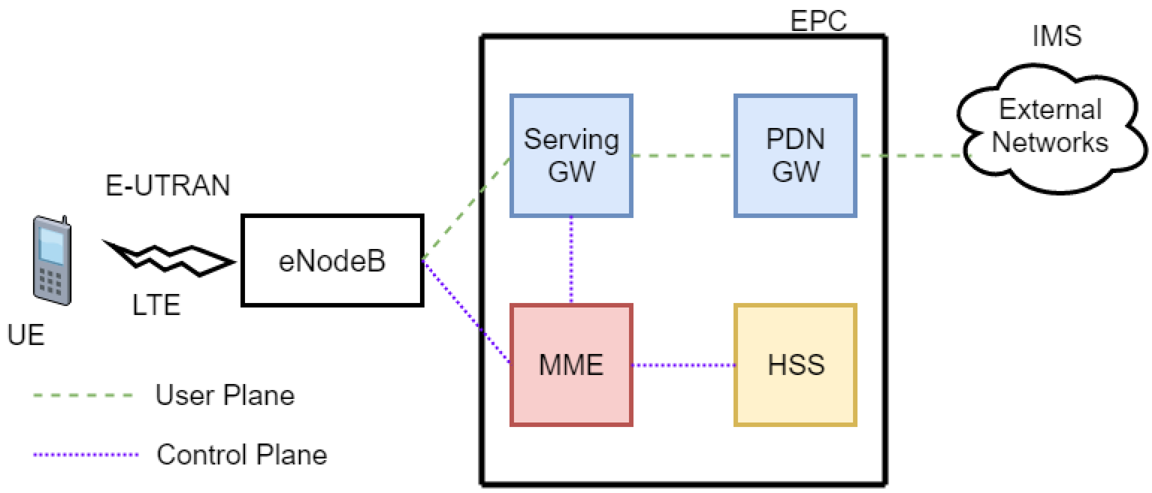


Figure 2.2: Evolved Packet Core in LTE, Source: (Frédéric Firmin - 3GPP MCC, fig 2) [20]

2.3. LTE Frame Structure

The frame structure for transmission in LTE is defined by 3GPP in [22]. Accordingly, downlink and uplink transmissions are organized into radio frames with duration of 10 ms. Each radio frame is divided into twenty time slots, each with a duration of 0.5 ms. A subframe consists of two consecutive time slots, hence the subframe duration is 1 ms.

Effectively this means that there are ten subframes available for transmission in the downlink & uplink in a 10 ms interval. The LTE frame structure has provisions for assigning up to six subframes in a radio frame to MBSFN. Figure 2.3 shows the FDD frame structure in LTE. Each OFDM symbol is prefixed with a cyclic prefix to prevent inter-symbol interference (ISI) due to multipath fading. The normal prefix length leads to seven symbols per time slot.

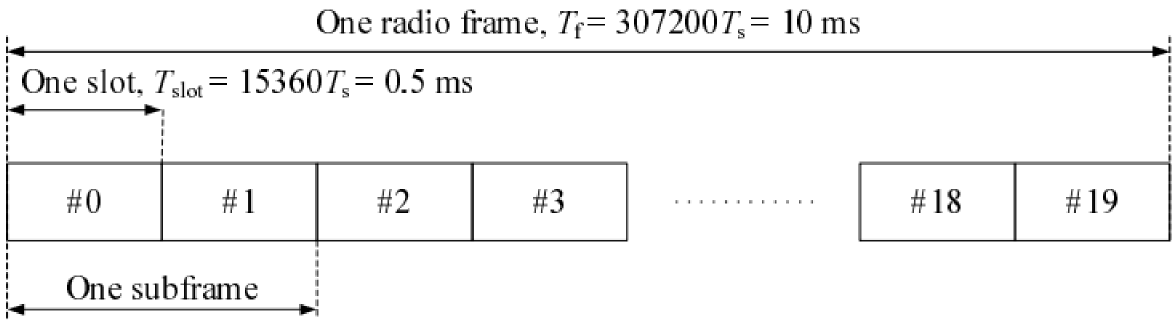


Figure 2.3: LTE Frame structure Type 1 for FDD, Source: 3GPP (fig 4.1-1,p.10) [22]

A Physical Resource Block (PRB) is defined as N_{symbol} (number of symbols) in the time domain and $N_{subcarrier}$ (number of subcarriers) in the frequency domain. A PRB thus consists of $N_{symbol} \times N_{subcarrier}$ resource elements. Each PRB consists of twelve subcarriers in the frequency domain and seven symbols in the time domain, which leads to 12×7=84 resource elements in a single PRB. Figure 2.4 shows the resource grid structure consisting of a resource block and resource elements.

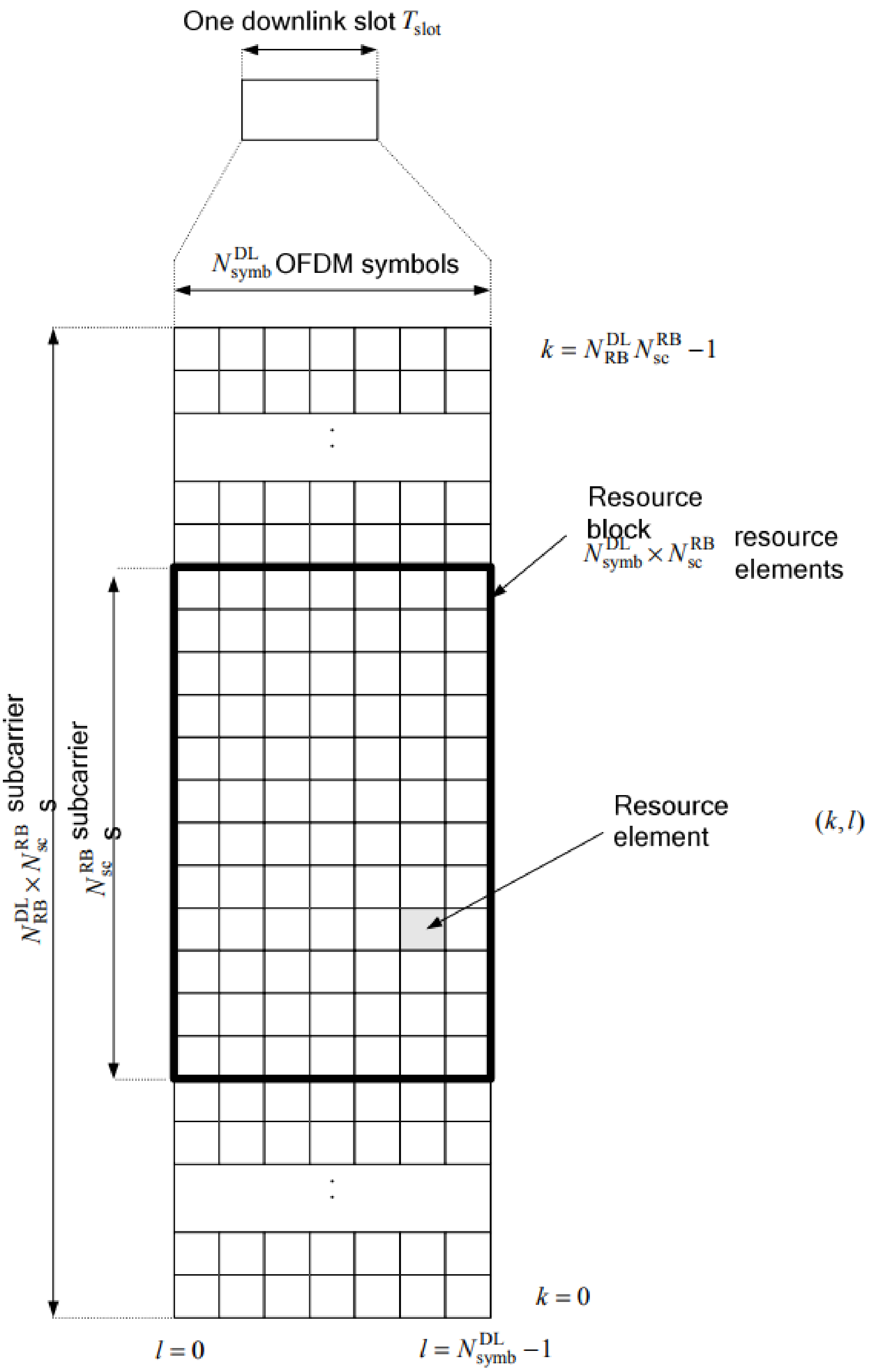


Figure 2.4: Resource grid displaying the PRB and resource elements. Source: 3GPP (fig. 6.2.2-1,p.52) [22]

2.4. MIMO in LTE

MIMO stands for Multiple Input Multiple Output. It is a technique used either to increase the data rate of the transmission in downlink and/or uplink or towards the improvement of link reliability (SINR, coverage). As the scope of this project is limited to the downlink, the focus will be on downlink MIMO system.

In a MIMO system, multiple antennas are used at both the transmitter (N number of antennas) and receiver (M number of antennas). The system performance through these multiple antennas can be improved by the following ways:

1. Spatial multiplexing: Independent data streams are transmitted from all the transmitting antennas which leads to an increase in data rates (ideally by a factor $\min\{N, M\}$). There are two types of spatial multiplexing, open-loop (OLSM) and closed-loop (CLSM). The primary difference between OLSM and CLSM is the feedback mechanism from the UE which provides channel state information (CSI) at the transmitter. This feedback is absent in OLSM and present in CLSM.

2. Transmit Diversity: Same data is transmitted from all the transmitting antennas which leads to an improvement in coverage (SINR)

3. Beamforming: The antenna beam is focused towards the users to improve system performance. This is achieved through signal processing techniques

Figure 2.5 shows the structure of a MIMO system with multiple antennas and receivers.

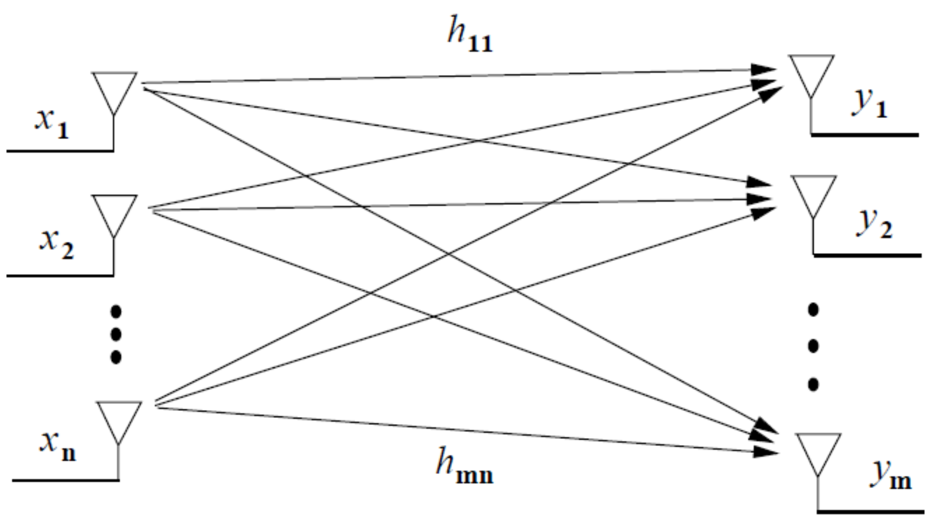


Figure 2.5: MIMO System, Source: (Andrea Goldsmith, fig. 10.1, p. 298) [23]

Mathematically, a MIMO system can be represented by the following equation:

$$y = Hx + n \tag{2.1}$$

where y and x denote the received and transmitted symbol vectors of length M and N respectively, H denotes the channel matrix of dimension $M \times N$ and n denotes the noise vector of length M. Assuming there are N transmitters and M receivers, the rank R_H of the channel matrix H is $\leq \min\{N, M\}$.

The full potential of spatial multiplexing is achieved only when $R_H = \min\{N,M\}$, which implies uncorrelated channels, and occurs in rich scattering environments. If there are correlated channels present, then the rank of H matrix R_H is $< \min\{N, M\}$, and the increase in data rate is less than the ideal factor of $\min\{N, M\}$.

The KPN LTE network currently employs 2×2 MIMO, and it operates either in the open-loop spatial multiplexing mode or the transmit diversity mode. Therefore, the R_H is ≤2 for the KPN LTE network. Depending on the rank indicator (rank of the H matrix) reported by the UE, which depends on actual channel conditions, the eNodeB either transmits one stream (transmit diversity) or two streams (OLSM) of data.

As will be discussed in more detail below, in the scheme of antenna muting, MIMO operation is ceased as one of the two transmitting antennas (2×2 MIMO) is muted (switched off) to reduce energy consumption, and this has an impact on system performance (see Chapter 4).

2.5. Multi-Carrier Mobile Network

An eNodeB (also referred to as a site) provides services in a given cell. The operator decides on the site location, sectorisation and antenna tilts (azimuth and downtilt). The size of the cells is jointly determined by the site grid, the antenna configurations and radio propagation aspects. Each cell is associated with a specific frequency carrier. Generally, a mobile network consists of multiple frequency carriers (at a single eNodeB), which results in multiple cells associated with multiple frequency carriers at a site. Traditionally, the cell(s) associated with the lowest frequency carrier is (are) called the coverage layer whereas the cell(s) associated with the higher frequency carriers is (are) called the capacity layer(s).

Figure 2.6 shows the typical layout of a mobile network site with two different frequency carriers (layers), 800 MHz and 1800 MHz.

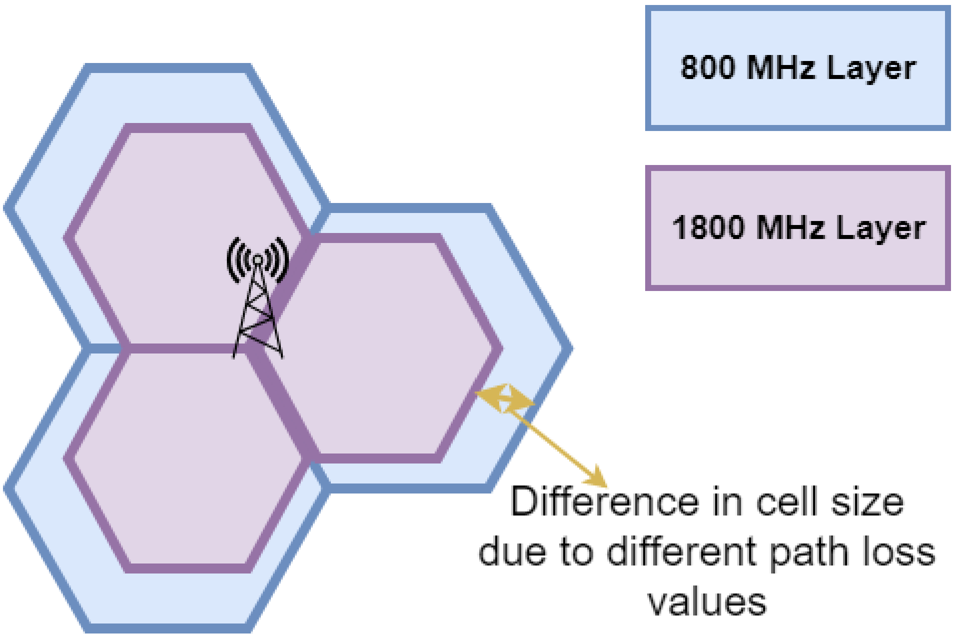


Figure 2.6: Mobile Network Site Layout

As shown in Figure 2.6, the 800 MHz frequency layer which is also known as the coverage layer, has a greater cell size (coverage area) than the higher frequency 1800 MHz layer. This is due to the fact that the path-loss at a given distance d_o is directly proportional to frequency of the electromagnetic wave as shown by the following equation:

$$Path - Loss = (\frac{4\pi d_o}{\lambda})^2$$
$$\lambda = \frac{c}{f} \tag{2.2}$$

where c denotes the speed of light, λ denotes the wavelength, f denotes the frequency, d_o denotes the distance between the transmitter and receiver. In reality, the path loss is larger as there are many obstacles which can cause scattering, shadowing, and reflection(s) of the electromagnetic wave. One of the commonly used models to predict path loss in outdoor environments is the Okumura-Hata model.

The site layout shown in Figure 2.6 represents a three-sector site, with six cells in total. The advantage of sectorisation is that it improves the SINR (Signal to Interference plus Noise Ratio) performance at the cell edge, and that only a single site needs to be acquired to deploy multiple cells, which makes sectorisation an economical way of cellular densification. Typically, a mobile network site has three sectors (cells) associated with each frequency carrier.

As will be discussed in more detail below, in the scheme of turning off carriers, the capacity layer associated with the 1800 MHz frequency carrier is turned off and its traffic transferred to the corresponding coverage layer associated with the 800 MHz frequency carrier (see Chapter 4).

2.6. Key Performance Indicators (KPIs)

Key Performance Indicators provide information about the system performance of the Radio Access Network and quality of service as experienced by users [24]. They can be used to test the impact of changes in network parameters/settings, detect unsatisfactory performance levels in the network, monitor the traffic volume and resource utilization in the network, and provide triggers for targeted and timely capacity expansions.

The International Telecommunications Union - Telecommunications (ITU-T) has a standard model for Quality of Service (QOS) [25], the different QOS categories in the standard model are:

1. Accessibility: "The ability of a service to be obtained within specified tolerances and other given conditions when requested by the user" (e.g. ERAB setup success rate)
2. Retainability: "The ability of a service, once obtained, to continue to be provided under given conditions for a given time duration." (e.g. call drop rate)
3. Integrity: "The degree to which a service is provided without excessive impairments, once obtained" (e.g. throughput)

The KPIs are calculated from performance management counters. The network counters collect data continuously at different granularity levels. A variety of network counters associated with multiple KPIs are available for every cell (lowest granularity level in the spatial domain) in the network. In the time domain, the lowest granularity level is fifteen minutes and data collected at the fifteen-minute level is aggregated for the higher granularity scales of hourly, daily, weekly, monthly and yearly. The data at the hourly and lower granularity levels is only available for a period of 30 days from the start of measurement reporting period, while at the daily and higher granularity scales it is available for up to two years from the start of measurement reporting period. The measurement data becomes available in the performance management application after 30 minutes from the start of the measurement reporting period. This 30 minute lag is an important factor due to which the schemes cannot be applied instantaneously in response to a change in traffic volume.

The metric (e.g. traffic volume) and KPIs (e.g. PRB utilization, throughput) relevant for this research are:

1. Total volume traffic: The amount of data traffic transferred in the downlink. Unit is Megabit.
2. UE throughput: The throughput experienced by a UE. The data is collected in the form of a distribution.
3. Active users in downlink: The average number of users considered active per unit time (ms) when data was required to be scheduled in the downlink (time-averaged over time when there are active sessions, defined from counters below).
4. Average PRB utilization: Fraction of time-frequency resources used in a given period of time.
5. License utilization: Fraction of license capacity used in a given period of time. The data is collected in the form of a distribution.

The KPIs are calculated from the following counters which are defined in the Ericsson Network IQ statistics performance management application [26]:

1. pmPdcpVolDlDrb: "The total volume (PDCP SDU) on Data Radio Bearers that has been transferred (UM and AM) in the downlink direction. Unit is kilobit"

2. pmPrbUsedDlSrbFirstTrans: "The total number of Physical Resource Block (PRB) pairs for HARQ initial transmissions used for Signaling Radio Bearers (SRB) in the downlink"

3. pmPrbUsedDlPcch: "The total number of Physical Resource Block (PRB) pairs used for paging"

4. pmPrbUsedDlDtch: "The total number of Physical Resource Block (PRB) pairs used for Data Radio Bearers (DRB) in the downlink"

5. pmPrbUsedDlBcch: "The total number of Physical Resource Block (PRB) pairs used for System Information"

6. pmPrbAvailDl: "The total number of Physical Resource Block (PRB) pairs available for transmission in the downlink"

7. pmActiveUeDlSum: "Number of UE's considered active in the downlink direction. Sampling rate is per transmission time interval". The active UEs are summed in a fifteen minute-interval.

8. pmSchedActivityCellDl: "The aggregated number of ms in which DRB data was required to be scheduled in the downlink. Unit is 1 ms"

9. pmLicDlCapDistr: "Shows the utilization of the downlink baseband capacity relative to the minimum of installed license for downlink baseband capacity or hardware capacity limit. Expressed as a percentage of the minimum of license limit or hardware capacity limit. Sampling rate is 10 ms"

10. pmUeThpDlDistr:"Distribution of the downlink UE throughput"

From these counters the relevant KPI's are defined as follows:

1. Total volume traffic (Megabit) = pmPdcpVolDlDrb * 10^{-3}
2. PRB Utilization = $\frac{pmPrbUsedDlSrbFirstTrans + pmPrbUsedDlPcch + pmPrbUsedDlDtch + pmPrbUsedDlBcch}{pmPrbAvailDl}$
3. Time-average number of active users in the downlink = $\frac{pmActiveUeDlSum}{pmSchedActivityCellDl}$
4. UE throughput (full distribution)= pmUeThpDlDistr (UE throughput data is available in the form of a histogram, each sample in the histogram represents 1 session)
5. License utilization (full distribution)= pmLicDlCapDistr (license utilization data is available in the form of a histogram, each sample in the histogram represents the license utilization for a period of 10 ms)

A small snipper of data from the counters is shown above in Figure 2.7. Each row in the above figure shows

[redacted]

Figure 2.7: Data from the Counter(s)

the metric of traffic volume, and KPIs PRB utilization and Active users in downlink for a period of one hour for one cell. [redacted]. In the following sections in this report, each row in Figure 2.7 will be referred to as one sample.

2.7. KPN Performance Targets

The minimum QoS target set by KPN is [redacted]. Based on a study carried out by TNO [27] for KPN (under default operational conditions i.e no resource reduction/switch-off), it was found out that this requirement is satisfied if:

1. The average PRB utilization should remain below [redacted] for a cell with 10 MHz bandwidth
2. The average PRB utilization should remain below [redacted] for a cell with 20 MHz bandwidth

From the above results it is observed that the PRB utilization threshold for meeting the minimum QoS requirements is higher for a cell with more bandwidth. An increase of resources, e.g. an increase in bandwidth (number of PRBs) leads to an increase in traffic handling capacity and therefore the minimum QoS threshold on PRB utilization is increased.

The study carried out by TNO [27] was based on the assumptions that the users (UE) are uniformly distributed over the cell, and that the traffic offered in all the cells is the same. Although these assumptions are generally not valid for all the cells in the KPN LTE network, the capacity management and planning process in the KPN LTE network for all the cells is nevertheless based on these results.

A cell is said to be in congestion if the minimum QoS target of [redacted] is not met. At KPN, the check for minimum QoS conditions are [redacted].

As the KPN capacity management and planning process assumes the results in [27] to be valid for all the cells in the network, exceeding the PRB utilization threshold given above also, by definition, implies that a cell is in congestion. An additional condition is also imposed to define congestion. The time-average number of active users in the downlink should be [redacted] (set by KPN, on average there are [redacted] users when sessions are active) for a cell to have a congestion label. This condition ensures that the congestion label is not applied due to a single active user. PRB utilization threshold exceeded due to a single user is ignored by the operator as the operator will not expand capacity at a site in response to a single active user. Note that the amount (number of hours) of congestion is upper bounded by the number of samples which meet the requirement of [redacted] on the time-average number of active users in the downlink.

The magnitude of congestion i.e. the magnitude by which the PRB utilization exceeds the PRB utilization-threshold limit defined above is inconsequential. However, the time-interval for which the PRB utilization exceeds the PRB utilization-threshold limit is not inconsequential. E.g. Assume that there are two cells with a bandwidth of 10 MHz each and meet the condition on the time-average number of active users in the downlink. The first cell has a PRB utilization of [redacted] for a time-interval of two hours and the other cell has a PRB utilization of [redacted] for a timer-interval of four hours. Both these cells exceed the PRB utilization threshold defined above and are defined to be in congestion. The cell which experiences congestion for a time-interval of four hours will receive priority for capacity expansion over the cell which experiences congestion for a time-interval of two hours.

2.8. Background information about the KPN 4G LTE Network

A cellular network consists of multiple carriers (or layers) as explained in Section 2.5. The KPN LTE Network has the following carriers/layers (scope limited to FDD carriers only).
When a new site is setup, it only has the coverage ([redacted] MHz frequency) layer. Over time, in response to the growth in traffic volume at a site, the higher frequency capacity layers are added. As a consequence of this, there are only a small fraction of sites in the KPN LTE network that have the higher frequency layers of [redacted] MHz present. However, the capacity layer of frequency [redacted] MHz is more widely present in the KPN network. Approximately [redacted] of the sites in the KPN Network have both the [redacted] MHz frequency carriers present. The [redacted] MHz carrier is present at only [redacted] of the sites and the [redacted] MHz carrier is present at [redacted] of the sites. Majority of the higher frequency carriers ([redacted]) are operated to satisfy the

[redacted]

Table 2.1: KPN LTE Network Parameters

spectrum license conditions. Therefore, for the analysis, only sites with [redacted] frequency carriers are considered. The [redacted] MHz frequency carrier will be referred to as the coverage layer, and the [redacted] MHz frequency carrier will be referred to as the capacity layer in the following sections.

3

Spatio-Temporal Traffic Variation

The mobile traffic volume variation at different locations at different timescales is presented in this chapter. Periodic mobile traffic variation is primarily driven by user behavior. The routine of the user and what time of the day, day of the week and/ or month of the year it is, impacts the traffic volume offered to a given site. In the following sections the variation in mobile traffic volume is analyzed for all three timescales under consideration. The three timescales are defined as follows:

1. Hourly timescale: The traffic volume per hour. This timescale is analyzed by looking at the daily 24-hour traffic volume pattern.
2. Daily timescale: The traffic volume per day, obtained from aggregating hourly traffic volume for all 24 hours in a day. This timescale is analyzed by looking at the weekly traffic volume pattern of workdays (Monday to Friday) versus weekend days (Saturday and Sunday)
3. Monthly timescale: Traffic volume per month, obtained from aggregating daily traffic volume for all the days in a month. This timescale is analyzed by looking at the monthly traffic volume pattern of summer-vacation months (July and August) versus non-summer vacation months (January to March).

3.1. Hourly Timescale

The primary purpose of the analysis presented in this section is to identify sites where the operational cost reduction schemes may be applied/effectuated during part of the day. On the hourly timescale the mobile traffic volume at a site is in the order of [redacted] mainly depending on what kind of site is considered and on the time of day. The traffic volume at both the site level and the network level closely mirrors the daily work and sleep schedule of the users. For this timescale, the nighttime hours are defined to be the time interval between 23:00-07:00 hours and the daytime hours are defined to be the time interval between 07:00-23:00 hours.

3.1.1. Network-Level Hourly Variation

The hourly traffic volume for the whole network is obtained from aggregating the traffic volume for each site in the network for each given hour. The hourly traffic volume for the whole KPN LTE network for a period of one week is shown in Figure 3.1. This figure is representative for any week as at the hourly timescale, the daily work-sleep/entertainment-sleep schedule of users is a universal activity which takes place every week. As is evident from the figure, the traffic volume during the nighttime hours (23:00-07:00 hours) is significantly lower than the traffic volume during the daytime hours (7:00-23:00 hours) for all days of the week. The split between nighttime hours and daytime hours is chosen with the aim of maximizing the potential of operational cost savings, and as observed in Figure 3.1, the traffic volume begins to significantly decrease from 23:00 hours onwards for all days of the week. This difference in the traffic volume between night hours and daytime hours provides an opportunity to potentially apply operational cost reduction schemes during the night hours. The magnitude of traffic volume during the nighttime hours varies from [redacted]. This matches expectations as it is a reasonable assumption than during the nighttime hours most users would be asleep, and hence there would only be

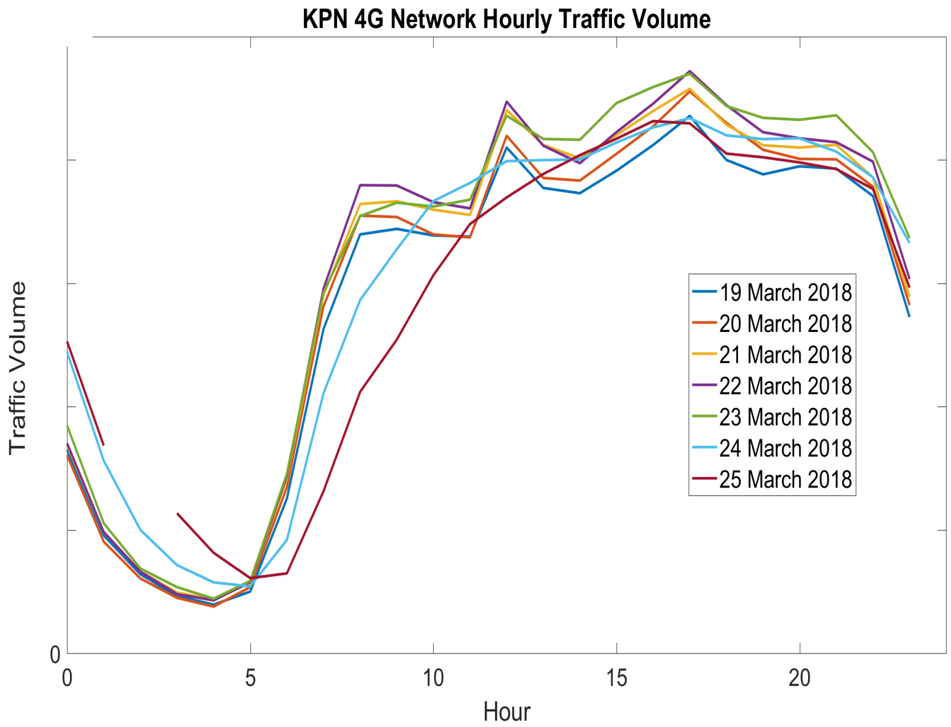


Figure 3.1: Hourly variation of the network-wide downlink traffic volume, as observed in the period from 19-25 March, 2018.

background activity on their phones which consumes less traffic when compared with activities like video streaming or checking emails.

There is a slight difference between the weekdays (19-23 March, first 120 hours) and the weekend (24-25 March, 120-168 hours in the figure). As the figure shows, people do tend to get up a bit later in the weekend, and the distinct peaks at 8 am and noon (lunch time) that exist during the working days, do not exist on the weekend, when things are more smoothed out, since people have less of a fixed time schedule On the weekend, the peak traffic volume between 17:00-18:00 hours is lower than the corresponding peak traffic volume on weekdays. Also, on the weekends the traffic volume between 00:00-05:00 hours is higher than the corresponding traffic volume on weekdays. A possible reason for these observations is the inherent difference between weekdays (workdays) and weekends due to different user behavior on the weekends. Note that in Figure 3.1, there is a break in the curve for 25 March between 02:00-03:00 hours as the daylight savings time began in The Netherlands at 02:00 on 25 March, 2018.

3.1.2. Site-Level Hourly Variation

As measures for reducing operational costs will be applicable at the site-level, it is necessary to quantify the hourly traffic volume variation at the site level. The objective of the analysis in this section is to discover the fraction of total sites in the network where traffic volume during the nighttime hours is significantly lower than the traffic volume during daytime hours, and to quantify how much lower the traffic volume is. The analysis is also carried out to find sites for a pilot study (for further analysis in Chapter 4) to implement the most promising operational cost reduction schemes. The sites which have the highest levels of excess capacity provided during the least busy hour, hence the sites whose traffic loads are lowest compared to the available resources, are chosen for the pilot.

The shape of the hourly traffic volume curve at a given site is similar to the hourly network traffic volume curve shown in Figure 3.1, and the main difference between them is the presence of abundant small-scale irregular variations during the daytime hours for the site-level traffic curve (see Figure 3.2 for two specific sites).
Hourly traffic volume at a site level is more susceptible to irregular variations as a small change in user-level

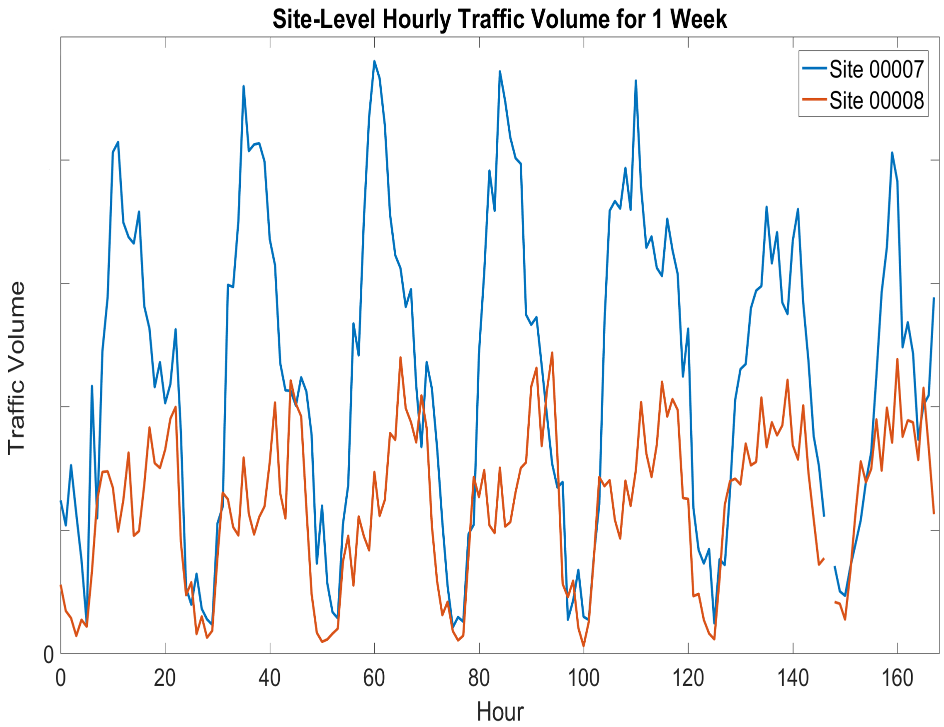


Figure 3.2: Hourly variation of the site-level downlink traffic volume for two sites, as observed in the period 19-25 March, 2018.

behavior (in the form of more users, more data usage per user or both) can impact the hourly traffic volume metric. Such irregularities tend to average out when applying a network-wide scope, and a significant (structural rather than incidental) change in user-level behavior is required to impact the hourly traffic volume metric.

As evident from Figure 3.2, traffic volume is periodically significantly lower during nighttime hours at both the network-level and site-level. The traffic dynamics at each site are different, however, all the sites shown an equivalent periodicity and have the lowest traffic volume during the nighttime hours. To provide a uniform network-wide solution to KPN, the schemes may only be applied during the fixed time interval of nighttime hours. As already mentioned, sites in a cellular network are dimensioned (in terms of site density and assigned resources per site) according to the peak hour traffic demand. The maximum excess capacity at a site during the nighttime hours is defined as a percentage change in traffic volume between the peak hour (17:00-18:00 hour) and the least busy hour (04:00-05:00 hour). This definition of excess capacity maximizes operational cost savings, while another way of defining excess capacity is the percentage change in traffic volume during the peak hour (17:00-18:00 hour) and the *peak* hour during the night, which minimizes number of hours of congestion experienced. As there are two options feasible for defining 'excess capacity', it is up to the operator to decide on the definition depending on the objective of maximizing cost savings or minimizing number of hours of congestion. As will be shown in Chapter 4, the definition of 'excess capacity' used in this report (maximize cost savings) leads to minimal to negligible number of hours of congestion. Furthermore, this definition of 'excess capacity' ignores the fact that in the peak-hour there may be excess capacity as well, which is primarily the case in rural areas where planning may be coverage oriented rather than capacity oriented, as a generic network-wide solution is assessed in this report instead of a site-specific solution.

A percentage change instead of an absolute magnitude change is calculated because sites in a cellular network have inherently different traffic dynamics, e.g. as reported in [16], " Typically, half of the network sites carry only 15 percent of the total traffic, while 5 percent of sites carry 20 percent of the traffic." Sites in busier regions (for e.g. city centers) carry much more traffic (and hence generally also have more assigned resources) than sites in rural areas, and thus an absolute magnitude change metric can overshadow the rural area sites completely. This fact is exhibited below in Table 3.1. As seen from the table, an absolute magnitude change metric for traffic volume for an urban site ([redacted]) is significantly higher when compared with the corresponding metric for a rural site ([redacted]), whereas the percentage change metric is comparable for both the sites ([redacted] vs [redacted] for an urban and a rural site respectively).

[redacted]

Table 3.1: Comparision between absolute and percentage change in traffic volume metric using a rural and urban site as an example

A higher percentage change value indicates that a higher level of excess capacity is provided at a site. The procedure for discovering the fraction of sites which have traffic volume significantly lower during night-time hours when compared with daytime hours, and quantifying how much lower the traffic volume is, is described below:

1. Exclude indoor sites and sites which have extremely low traffic volume (out of scope, see Section 1.1).
2. Calculate the total traffic volume metric for each site for the hour between 04:00-05:00 and 17:00-18:00 for one week (seven days) from the network counter data. If a site was under maintenance or if data was unavailable for these hours, that site was excluded from consideration.
3. Calculate the mean of the total traffic volume metric for the hour between 04:00-05:00 and 17:00-18:00 hours for each site over all days of the week (data collected for one week only). The mean is calculated to average out the potential irregular variations that may be present at a site, and to take into account the inherent difference between the traffic volume during weekdays and weekends. Denote the mean traffic volume between 04:00-05:00 and 17:00-18:00 hours for a given site by V_{4-5} and V_{17-18} respectively.
4. Calculate the percentage change between the mean traffic volume metrics V_{17-18} and V_{4-5} which were calculated in the previous step for each site. The value of percentage change (for each site) is calculated from the following equation:

$$Percentage - Change = \frac{V_{17-18} - V_{4-5}}{V_{17-18}} \times 100\%$$

5. The sites are sorted in descending order of the value of percentage change. Positive valued percentage change means that $V_{17-18} > V_{4-5}$ and negative valued percentage change means that $V_{17-18} < V_{4-5}$.
6. Steps 1-5 above are repeated for a different week to verify that the result obtained in the first iteration was not an anomaly.

The result of this exercise is the list of sites (presented as a CDF in Figure 3.3) which are sorted in a descending order by the values of percentage change which indicate the levels of excess capacity that are provided at a site. The following observations were made from the result obtained from the above exercise:

1. On investigating the physical environment of the top 100 sites in the list by manually checking the site location(s) on KPN systems, it was discovered that sites serving schools and office buildings are abundantly present at the top of the list.
2. All the sites except for two have a positive valued percentage change (calculated in step 4 of the above exercise). This indicates that the mean traffic volume metric $V_{17-18} > V_{4-5}$ almost everywhere. The two sites which do not show a positive valued percentage change seem to be sites where tests were carried out by KPN. This observation is expected as during the hour between 04:00-05:00 majority of the users would be inactive or asleep, whereas between 17:00-18:00 majority of the users would be active.

3. As almost all the sites (> 99% sites) show a positive valued percentage change, at the hourly timescale it can be argued that there is only a *temporal* variation in traffic volume between daytime and nighttime hours. The location (space) of the sites does not appear to play a role in the traffic dynamics, as at all the sites, traffic volume increases from nighttime to daytime and then falls back at nighttime.

4. Sites present at the top of the ordered list were selected for the analysis of the operational cost reduction schemes in Chapter 4 (pilot study). The value of percentage change for the selected pilot sites is ≥ 95%, indicating that the traffic volume between 04:00-05:00 is less than 5% (100-95=5%) of the traffic volume between 17:00-18:00 hours (see Figure 3.3).

Sites serving schools and offices exhibit the highest value of percentage change which is intuitively logical. These locations are inactive to a high degree during nighttime as there would not be any schoolchildren or office workers present at night. Within the top 100 sites, there is a considerable dissimilarity in the traffic volume curve(s) during daytime hours due to different user routines at these sites. However, as the operational cost reduction schemes may only be applied at nighttime, the daytime curve(s) at these sites are of no consequence as long as there is a significant change between the daytime and nighttime traffic volume (and hence a lot of excess capacity at night).

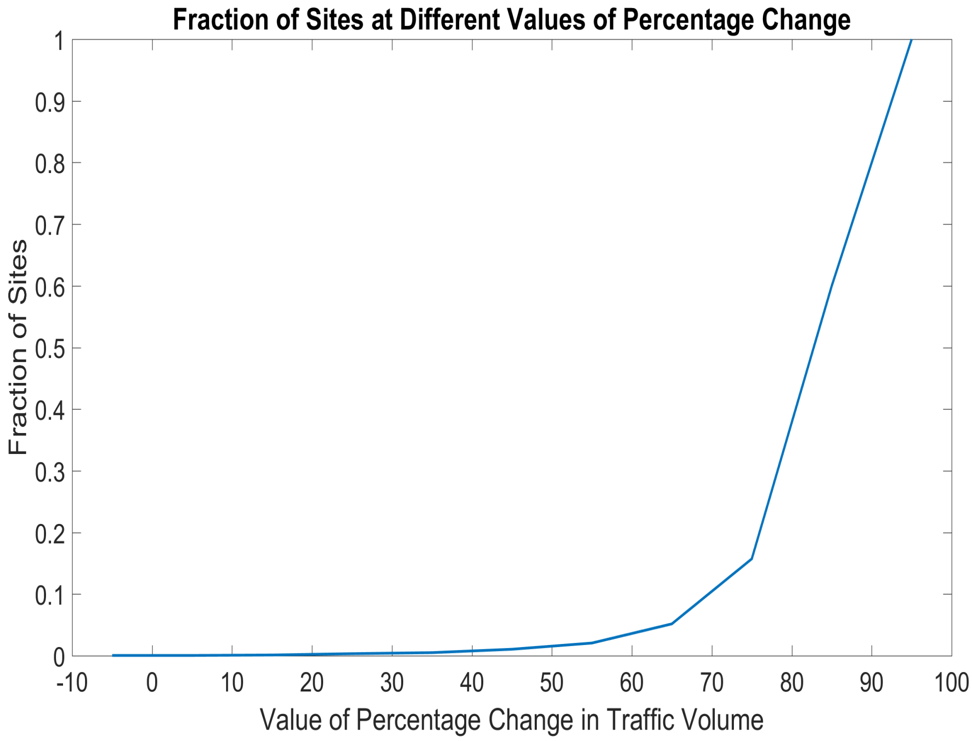


Figure 3.3: Fraction of sites at different values of percentage change in traffic volume presented in the form of a CDF for the hourly timescale.

Figure 3.3 shows the value of percentage change in traffic volume observed at all the sites (at step 4 of the calculation exercise) in the form of a CDF. The horizontal axis shows the value of percentage change between V_{17-18} and V_{4-5}, as defined above. A positive horizontal axis value means that $V_{17-18} > V_{4-5}$, and a negative horizontal axis value would reflect the opposite. This figure serves as an indicator of the number of sites where the schemes can be applied. The percentage of sites with a traffic volume change >95% is 9.73%.

3.2. Daily Timescale

The analysis on the daily timescale is carried out in a manner similar to the one in the previous section for the hourly timescale. Therefore, only brief remarks will be mentioned about the procedure. On the daily timescale the mobile traffic profile at a site is in the order of [redacted] depending on the type of site and the day of the week it is. For this timescale, weekdays are defined to be the days from Monday to Thursday and weekends are the days of Saturday and Sunday. Friday is excluded from the analysis in Section 3.2.2 as it is a transitional day from weekday to weekend, Friday evening can be assumed to be the start of the weekend.

3.2.1. Network Level Daily Variation

At the network level, it is unclear how the daily user profile affects the traffic volume metric. Figure 3.4 shows the daily traffic volume for KPN's LTE Network for a period of two continuous weeks. As can be seen from the figure, traffic volume dips on Wednesday(s), Saturday(s) and Sunday(s) when compared with the day(s) preceding it i.e. Tuesday(s, Friday(s) and Saturday(s). Note that Figure 3.4 is zoomed in as the vertical axis range is from [redacted] TB, if the vertical axis would be plotted from [redacted], then the daily variations in the network-wide daily traffic volume would seem a lot less significant than they do now. Note that the vertical axis range in figure 3.4 is from [redacted] TB,

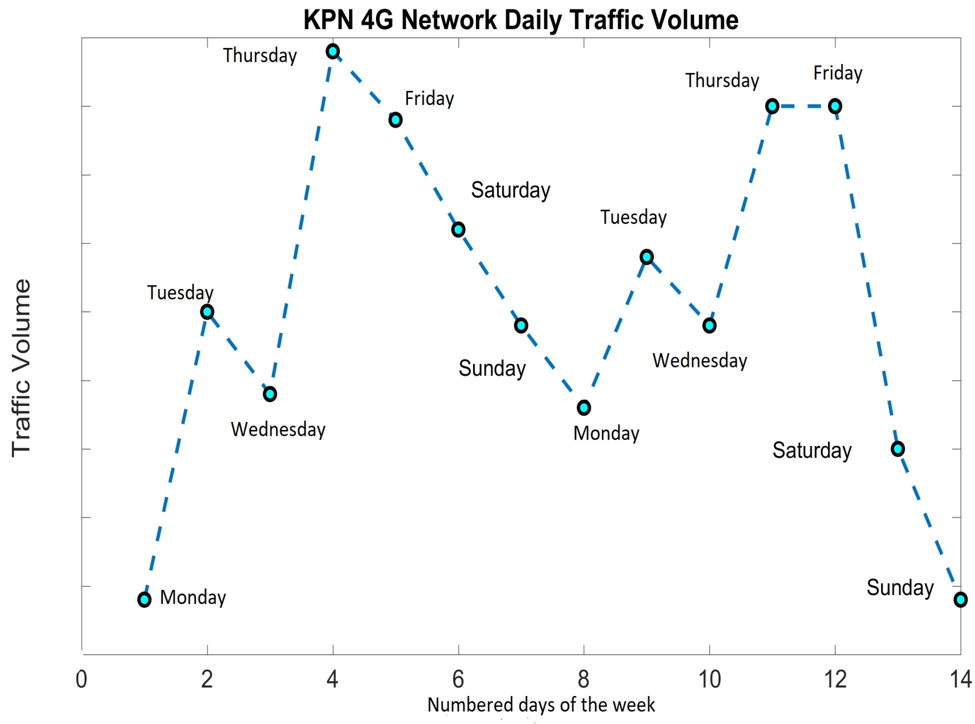


Figure 3.4: Daily variation of the network-wide daily traffic volume, as observed in the period from March 5-18, 2018

3.2.2. Site level daily variation

As the operational cost reduction schemes will be applied at the site level, it is necessary to quantify the daily variation in traffic volume at the site level. As it is observed from Figure 3.4 the traffic volume for the network dips on Saturday (compared to Friday), the change in traffic volume is calculated between weekdays and weekends. The procedure for the analysis is identical to the one described in Section 3.1.2. In this timescale, traffic volume metric for each site for a period of two weeks is calculated from the counter data. For this timescale, V_{04-05} and V_{17-18} from the hourly timescale are replaced by V_{WKND} and V_{WKD} respectively for each site. The traffic volume (per site) for the weekdays (Monday to Thursday) for two weeks will be averaged to obtain V_{WKD}, and the traffic volume (per site) for the weekends (Saturday and Sunday) for two weeks will be averaged to obtain V_{WKND}. The rest of the procedure is identical.

The result of this exercise (analysis similar to hourly timescale) is the list of sites which are sorted in a descending order by the values of percentage change which indicate the levels of excess capacity that are provided at a site. The following observations were made from the result obtained from the exercise:

1. On investigating the physical environment of the top 100 sites in the list by manually checking the site location(s) on KPN systems, it was discovered that sites serving schools, universities and offices are once again abundantly present at the top of the list. The physical environment of the sites present at the bottom of the list was discovered to be entertainment and recreational areas.

2. Approximately 47% the sites have a positive valued percentage, and 53% of the sites have a negative valued percentage change (see Figure 3.5). This indicates that for some sites traffic volume decreases during the weekend, and for some sites it increases during the weekend. |5|

3. Sites present at the top of the ordered list were selected for the pilot study (for further analysis in Chapter 4). 48 sites are selected for the pilot and these sites exhibit a percentage change in traffic volume >80%.

Sites serving schools and offices exhibit the highest value of percentage change, which is once again intuitively logical. These sites are inactive to a high degree during weekends as there would not be any students or office workers present during the weekends. Sites serving entertainment and recreational area exhibit negative valued percentage change, and this is also intuitively logical as these sites are generally busy during the weekends.

Figure 3.5 shows the value of percentage change in traffic volume observed at all the sites (at step 4 of the calculation exercise) in the form of a CDF. The horizontal axis shows the value of percentage change between V_{WKD} and V_{WKND}. A positive horizontal axis value means that $V_{WKD} > V_{WKND}$, and a negative horizontal axis value would reflect the opposite. At the sites which show a positive valued percentage change, the operational cost reduction schemes may be effectuated during weekends (full 48 hours). Whereas, at sites which show a negative valued percentage change, the operational cost reduction schemes may be effectuated during all the weekdays (whole days). Approximately 14% of the sites exhibit a minimum value of percentage change (magnitude) of |50%| (Figure 3.5). As the procedure for the analysis of the operational cost reductions schemes on the daily timescale will be similar, for sites which exhibit positive valued percentage change and sites which exhibit negative valued percentage change, the detailed analysis in Chapter 4 is only carried out for sites which exhibit positive valued percentage change.

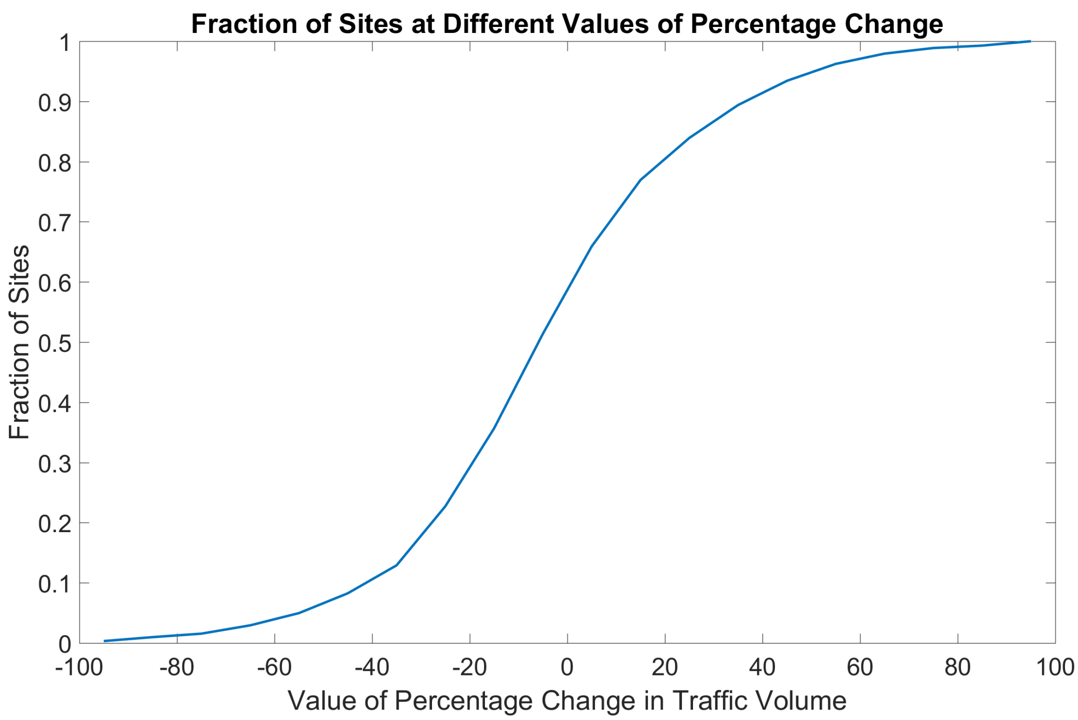


Figure 3.5: Fraction of sites at different values of percentage change in traffic volume presented in the form of a CDF for the daily timescale.

Figure 3.6 shows the daily variation for four sites, two sites were selected from the top of the list and the remaining two sites were selected from the bottom of the list. The traffic volume at Sites 1 and 2 decreases on weekends, and increases on weekends for Sites 3 and 4. Traffic volume either dips or peaks during weekends at these sites periodically (every week). From Figures 3.5 and 3.6 it can be seen there is a potential opportunity

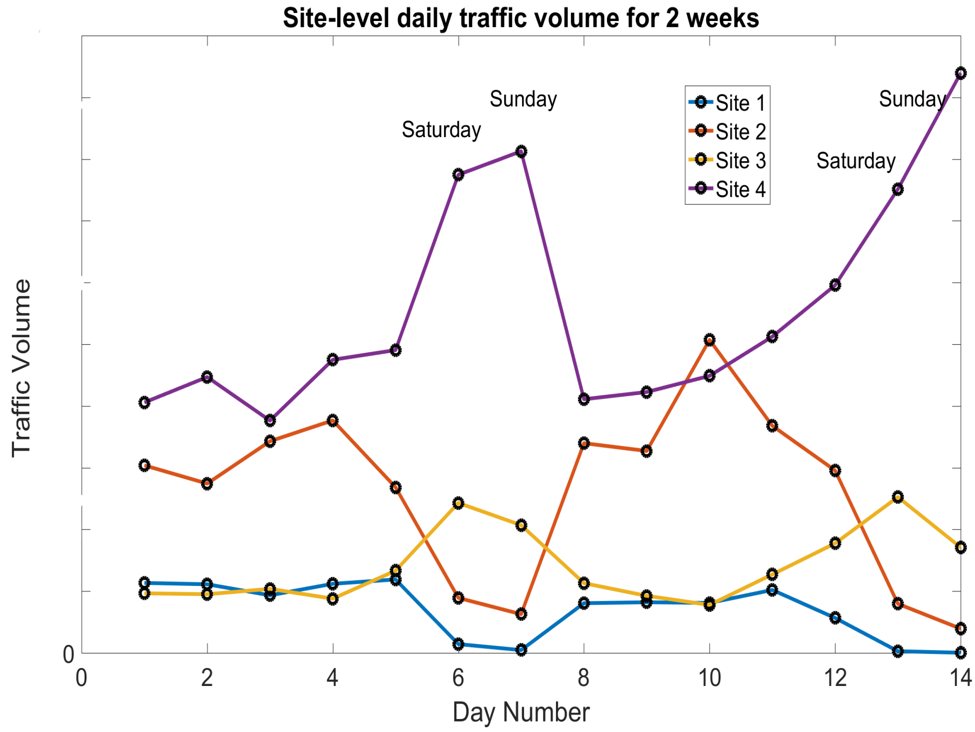


Figure 3.6: Daily variation of the site-level downlink traffic volume for four sites, as observed in the period March 5-18, 2018

available to the operator to effectuate the operational cost reduction schemes for a period of whole weekends or weekdays depending on the type of site.

3.3. Monthly Timescale

The analysis on the monthly timescale is carried out in a manner similar to that presented in the previous subsections. At the monthly timescale the mobile traffic volume at a site is in the order of [redacted] mainly depending on the kind of site and the month of the year it is. At the monthly level, the traffic volume at the network level is always increasing due to an increase in subscribers, connected devices and average data usage per device. On top of this increasing trend, there are structural fluctuations observed at the site level, and these structural fluctuations are the focus of the analysis in this section. For this timescale, the summer vacation months are defined to be the months of July and August and the non-summer vacation months are defined to be the months from January to March.

3.3.1. Network Level Monthly Variation

There is no distinctive variation in the network-level monthly traffic volume. The only definitive observation that can be made is that the traffic volume at the monthly level is always increasing for the network as a whole. Figure 3.7 below shows the network-wide monthly level traffic volume from January 2016 to July 2018. The traffic volume increases each month, which is likely due to an increase in the number of connected subscribers/devices and an increase in smartphone data usage. An above-proportional increase is observed during the month of December during the last two years in 2016 and 2017, which is likely due to the fact that KPN has been offering a 5GB data bundle for free to all its subscribers in December (every year). Figure 3.7 shows the monthly traffic volume for the whole of KPN's LTE network for last 3 years. As is clear from it, the traffic volume increases each month.

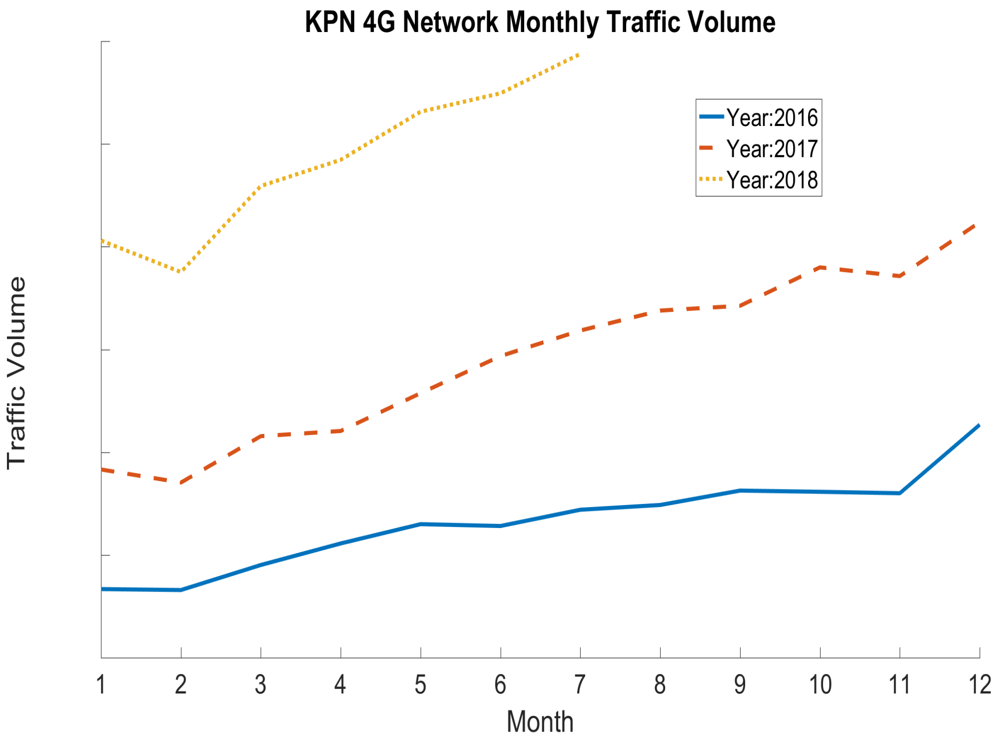


Figure 3.7: Monthly variation of the network-wide daily traffic volume, as observed in the period from January 2016- July 2018.

The magnitude of mobile traffic volume at the monthly level is in the range of a few thousand Tera-bytes. The traffic volume curve for the monthly timescale is extremely different from those presented in the previous Sections 3.1 and 3.2 as there are no obvious reasons (e.g. sleep schedule or other user-level behavioral aspects) for significant variation in the network-wide traffic volume at this timescale.

3.3.2. Site level monthly variation

The analysis at the site level is carried out in a manner identical to that outlined in previous subsections. As said, there are no definitive peaks or dips present in the network monthly traffic volume metric. However, as the observation that traffic volume increases significantly for a fraction of sites during the summer vacation months of July and August was made by KPN and served as motivation for this research, the focus of the analysis on this timescale is narrowed to the variation in traffic volume between the months of July-August when compared with the other months in a year. For this timescale, V_{04-05} and V_{17-18} from the hourly timescale are replaced by V_{NSMR} and V_{SMR} respectively. V_{NSMR} and V_{SMR} are the average traffic volume (per site) for the months of January to March and July-August. The months of April to June are transitional months to the summer-vacation period and are therefore left out from the analysis. There is no significant variation(s) observed in the site-level traffic volume for the months of September to December, and thus these months are also left out from the analysis. The analysis is only carried out using data from the year 2017 as a significant number of sites were added in 2017. The result of this exercise (analysis similar to hourly timescale) is the list of sites which are sorted in a descending order by the values of percentage change which indicate the levels of excess capacity that are provided at a site. The following observations were made from the result obtained from the exercise:

1. Sites serving camping grounds, beaches and marinas are abundantly present in the top 100 sites list.
2. The traffic profile of the sites at the top of the list was investigated and it was discovered that after July and August, the traffic volume at these sites falls back to the levels observed at the start of the year apart from the effects of the structural traffic growth.
3. Majority of the sites (≈96%) show a positive valued percentage change, that is the traffic volume during July and August is higher than the traffic volume during the months of January to March. This is likely

due to the structural growth in traffic volume. As the monthly timescale is much larger than the hourly and daily timescale, there is a substantial growth in subscribers and devices connected to the network and thus the traffic volume always increases month to month ([1]).

4. There are approximately 4% of the sites which exhibit a negative valued percentage change. On investigation, these sites were discovered to be serving universities and schools.

5. Approximately 11% of the sites show a >50% change in traffic volume, that is the traffic volume during the months of January-March is < 50% of the traffic volume during the months of July and August.

Sites serving camping grounds, marinas, and beaches are abundantly present in the top 100 sites, and this is intuitively logical. Recreational sites like these have peak traffic during the summer vacation period due to reasons like a huge influx of tourists, people on vacation. Sites serving universities and schools exhibiting negative valued percentage change is also expected. Students are on holiday during the summer vacation period and therefore the universities and schools will be relatively empty. Figure 3.8 shows the value of percentage change in traffic volume observed at all the sites (at step 4 of the calculation exercise) in the form of a CDF. The horizontal axis shows the value of percentage change between V_{SMR} and V_{NSMR}. A positive horizontal axis value means that $V_{SMR} > V_{NSMR}$ and a negative horizontal axis value(s) would reflect the opposite.

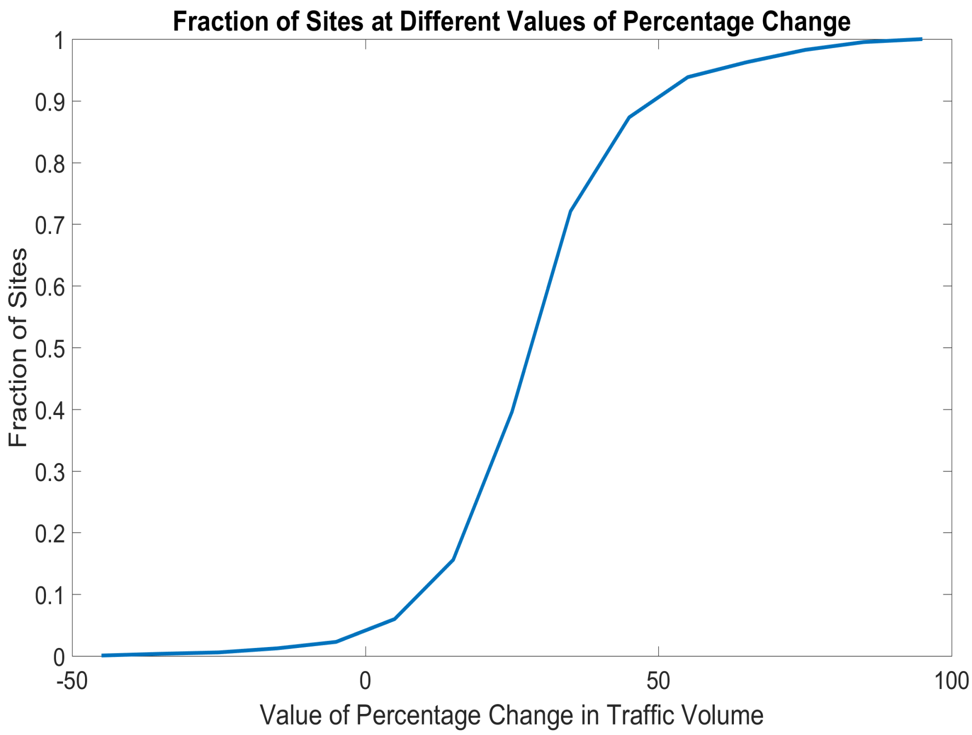


Figure 3.8: Fraction of sites at different values of percentage change in traffic volume presented in the form of a CDF for the monthly timescale.

The traffic profile(s) for a few sites which exhibit the highest and lowest value of percentage change in traffic volume is shown in Figure 3.9 below. As evident from it, there is a huge increase in traffic volume between the months of June-August and then the traffic volume falls to levels observed at the start of the year. Sites 1 to 4 in the figure exhibit a dramatic increase in traffic volume during the months of July and August, whereas Site 5 shows a dramatic decrease in traffic volume during the same time period. From Figures 3.8 and 3.9 it can be seen that there is an opportunity for the operator to effectuate the operational cost reduction schemes on the monthly timescale. The schemes can be effectuated all year round except for the months of July and August for sites which show positive valued percentage changed, and can be effectuated only during the months of July and August for sites which show a negative valued percentage change.

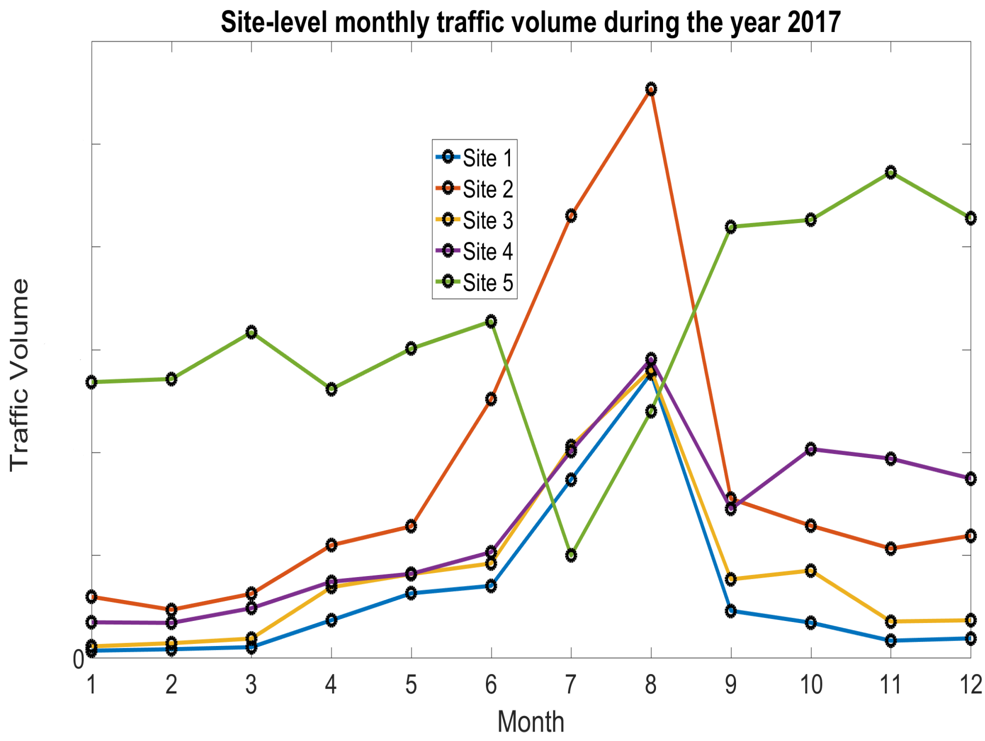


Figure 3.9: Monthly variation of the site-level downlink traffic volume for four sites, as observed in the period of January-December 2017.

3.4. Overview of the three timescales

In this chapter we have analyzed mobile traffic variation at both the network and site level and considered the three distinct hourly, daily and monthly timescales. From the results presented, the main findings are as follows:

1. Hourly timescale: Traffic volume is periodically at its lowest levels during nighttime hours (23:00-07:00 hours) for almost all sites. 9.73% of the sites exhibit a percentage change in traffic volume of >95%.

2. Daily timescale: Periodically, traffic volume dips during weekends for some sites and peaks for other sites at the same time. 14% of the sites exhibit a percentage change (±) of at least |50%|.

3. Monthly timescale: Traffic volume is structurally increasing from month to month. Besides this trend several sites also exhibit the effects of seasonal variations in the handled traffic volume. More specifically for some sites it dramatically increases during the months of July and August and then falls back to the traffic levels observed during January to March along with the effects of the structural traffic growth. At the same time, for a few sites traffic volume dramatically decreases during the months of July and August. 11% of the sites exhibit a percentage change in traffic volume >50%.

The fraction of sites at different values of percentage change in traffic volume is also analyzed for the three timescales. An operator may decide to set some threshold value on percentage change in traffic volume for effectuating the operational cost reduction schemes. However, this is completely the prerogative of the operator. Depending on the threshold assigned by the operator, the operational cost reductions schemes will be effectuated for a different number of sites in the network. From Figure 3.3 it is for example observed that more than 99% of the sites show greater than 50% change in traffic volume at the hourly timescale. Therefore, if the threshold set by operator is 50%, 99% of the sites in the network will be available for effectuation of the operational cost reduction schemes.

The potential congestion that may occur on effectuating the operational cost reduction schemes is lower for sites with a higher 'percentage change value'. The lower the threshold, the higher the number of sites at which the schemes will be effectuated, but the potential for congestion will also be higher. Nothing concrete can be said about the potential cost reduction as it depends on the energy consumption (PRB utilization) in the baseline scenario (normal operation mode when no resources are switched-off or reduced). A site having a higher or lower 'percentage change value' reveals nothing about its baseline energy consumption (PRB utilization) and thus nothing can be said about the potential cost reduction before performing calculations.

The pilot-sites selected for the assessment of the operational cost reductions schemes do not necessarily have the highest potential for savings in operational cost and/or energy consumption. The pilot-sites represent a fraction of the total number of sites where the operational cost reduction schemes may be effectuated. Thus, it may be possible that the sites which were not selected for the pilot (lower percentage change value compared to the pilot-sites) for the assessment of the operational cost reduction schemes, may have higher savings in operational cost and/or energy consumption when compared with the pilot-sites due to a higher baseline operational cost and/or operation while still meeting the minimum QoS requirements.

It is concluded that there is ample opportunity for the operator to reduce operational cost by effectuating the scheme(s). The attainable savings in the operational cost and/or energy consumption on effectuating the different schemes is discussed in the following chapter.

4

Operational Cost Reduction Schemes

The operational cost reduction schemes are assessed and discussed in this chapter. There are five schemes under consideration, and the objective is to quantify the trade-off between operational cost savings and the minimum quality of service requirements. The five different schemes are turning off carriers, antenna muting, discontinuous transmission, Ericsson's so called Psi-Coverage solution and reduce license capacity, and these schemes will be described and assessed in the following sections. The trade-off is presented in the form of the number of hours of congestion and hence sub-target throughput performance that may occur if the scheme is activated. The decision on whether the trade-off is acceptable and hence whether or not to effectuate the scheme rests with KPN management. The reference scenario(s) for calculating operational cost savings will be described in the following sections. Note that the schemes are applied at the site level and not the cell level, while the minimum QoS requirements are checked at the cell level. Schemes are applied at the site level due to the fact that not all schemes can be applied at the cell level (e.g. reducing license capacity) and an energy model at the cell level is unavailable.

4.1. The Framework

From the perspective of the operational cost reduction schemes, the three different timescales are identical with respect to the assessment of the schemes. The only difference between the three timescales is the number of hours when the scheme(s) is (are) active and the number of sites where the scheme(s) is (are) active. The analysis for the schemes is carried out within the following framework:

1. The mechanism to effectuate the scheme(s) for the three different timescales is the same. A different tooling is not required for the different timescales, only the time duration and the affected number of sites depend on the considered timescale. For e.g. on the hourly timescale the scheme(s) may be activated for eight hours every night, on the daily timescale the scheme(s) may be activated for 48 and/or 120 hours every week during the weekends and/or weekdays and on the monthly timescale the scheme(s) may be activated for ten months in a year outside of the summer vacation period. The affected number of sites depends on the 'percentage change value' threshold set by the operator and can be calculated from Figures 3.3, 3.5 and 3.8 for the hourly, daily and monthly timescale respectively.

2. The analysis is presented from the viewpoint that the decision to whether activate the scheme(s) for a fixed duration is taken every day, week, or month (hourly, daily, monthly timescale) based on the savings and congestion experienced the previous day, week or month. The scheme(s) will remain active for the whole time duration corresponding with the considered timescale. For e.g. on the hourly timescale, the scheme will remain active for all eight hours (23:00-07:00 hours, fixed duration) in the night. The trade-off will be the number of hours of congestion experienced every day, week or year. Note that the minimum QoS conditions are still checked every hour for all the three timescales (during the time period when the schemes are active).

3. The check(s) for congestion can be carried out via a throughput-based check and/or a PRB utilization-based check. The throughput-based checks are carried by checking whether the [redacted] meets the target of [redacted] and the PRB utilization-based checks is carried out by comparing the PRB

utilization with the threshold limit(s) mentioned in Section 2.7. For the operational cost reduction schemes of 'reducing license capacity' and 'Ericsson's Psi-coverage solution', no congestion checks are carried out for the reasons mentioned in Sections 4.5 and 4.6.

4. The different schemes are analyzed by applying them only to the pilot sites (all cells) across the three different timescales. This is done to compare the schemes with each other. In some cases, effectuating the operational cost reduction schemes on one frequency band may affect another frequency band, this will be explicitly accounted for in the analysis in the following sections.

As the network counter-data is continuously logged (every fifteen minutes) for all the cells in the network, the assessment of the operational cost reduction schemes may be repeated every day, week or month (point 2 above) with fresh data samples obtained the previous day, week or month. Executing this exercise every day, week or month will automatically account for traffic growth and addition of new sites in the network, and adapt the cost savings attainable and trade-off of number of hours of congestion. For the assessment of the operational cost reduction schemes, it is assumed that the future fluctuation(s) in traffic volume will mirror the fluctuations observed in the past. This assumption is reasonable as it has been shown in [4] that when traffic volume is decomposed into a regular and irregular component, the regular component has a well defined structure (strong periodicity) and the irregular component is random. Therefore, in general, the future fluctuations in traffic volume will mirror the past fluctuations, but this cannot be guaranteed due to the presence of irregular variations in traffic volume.

For the throughput-based check (point 3 above), an assumption is made that the UE sessions in the throughput-distribution are meaningful file-transfers and not text-messages. Text-message sessions inherently have lower throughput than file-transfer sessions due to TCP slow start [28]. If majority of the sessions are text-message sessions, then the [redacted] would likely fall below the target of [redacted] and lead to a false-positive congestion label. Whereas, in actuality, there would be no congestion as the UEs would be experiencing low throughput due to the inherent nature of the text-message sessions and not due to capacity related issues.

As the throughput data is available from the network-counters in the form of a histogram, an assumption that the sessions (samples) are uniformly distributed with the bins of the histogram is made to facilitate the throughput-based check for minimum QoS conditions.

As the methodology for the assessment of the operational cost reduction schemes is identical for all three timescales, the detailed analysis is only presented for the hourly timescale in the following sections. Only the final cost/energy savings and the trade-off of number of hours of congestion is presented for the two other timescales.

The cost/energy savings results are presented on a per-day, per-weekend and a per-week basis for the hourly, daily and monthly timescale respectively. The results are not expanded to an yearly figure as it is unknown what the savings may be in the future. As traffic volume grows in the future (due to the reasons mentioned in Chapter 1), the 'fixed time duration' during which the operational cost reduction schemes may be effectuated may also change. However, the increase in traffic volume will be accompanied with an increase in PRB utilization as well, and thus higher savings in energy/cost (energy consumption and saving increases with PRB utilization, discussed below in the power model). Therefore, in the future, we may have higher savings but for a shorter time-duration. Thus, detailed calculations would need to be done to know whether the savings in the future are higher or lower than the present-day savings.

If the time-duration for the effectuation of the operational cost reduction schemes remains the same in the future, there will be higher savings in energy/cost. However, these higher savings would also be accompanied with an increase in the number of hours of congestion experienced.

4.1.1. Reference Point in the KPN LTE Network

The energy supplied to an antenna varies from carrier to carrier in the KPN LTE network (depending e.g. on the presence of tower mounted amplifiers (TMA), different cable losses). The parameter which sets the total energy supplied to the two transmitting antennas is referred to as the 'reference point' by KPN. In the KPN network, the 'reference point' parameter can have the following values:

1. For the [redacted] MHz carrier, the ‘reference point’ is set to [redacted] Watt ([redacted] Watt to each antenna port, [redacted] MIMO operation) which equals [redacted] dBm ([redacted][mW]).
2. For the [redacted] MHz carrier, the reference point energy setting can either be [redacted] Watt ([redacted] dBm) or [redacted] Watt ([redacted] dBm). The [redacted] MHz carriers which have the reference point set to [redacted] while the remaining [redacted] MHz carriers do not have a TMA present.

Note that the ‘reference point’ is the total energy at the output of the radio unit, and this energy is decreased by the magnitude of the jumper cable losses towards the antenna ports. Also note that for the [redacted] carrier, the reference point is defined at the output of the TMA and not at the output of the radio unit. Thus, for the [redacted] carrier, the effective output at the radio unit is equal to the reference point plus feeder losses (generally 3 dB) as the TMA provides no gain in the downlink. For the illustration of the ‘reference point’ of the [redacted] carrier, a simplified diagram is shown in Figure 4.1.

[redacted]

Figure 4.1: Reference point(s) for the 800 MHz and the 1800 MHz carrier

Note that all the pilot-sites selected for the assessment of the operational cost reduction schemes have TMAs present. For this research, the total energy consumption at a site is defined to be a sum of the energy consumption of each carrier present at the site. The parameters reference point, the bandwidth of the carrier, the number of antenna ports, the PRB utilization and some default settings are input into the IMEC power model [29] to calculate the energy consumption of the carrier(s) (see Appendix A). The detailed working of the IMEC power model is described in [30–32]. The IMEC power model was also validated by running an experiment in the live network (switching off one antenna) and the observations are presented in Appendix A.

The energy consumption of a carrier is a combination of a fixed and a variable component. The fixed component is the amount of energy which is always consumed (even at zero PRB utilization), and the variable component scales approximately linearly with system load (PRB utilization). The IMEC power model only considers the five main components of the radio base station for energy consumption calculation purposes. These five components as defined in [32] are:

1. Power Amplifier (PA): Energy consumed by the amplifier circuitry. PA consumes the highest energy out of all the components at full load (100% PRB utilization). Assuming KPN base-station parameters for the IMEC power model, at full load for a bandwidth of 10 MHz, the power amplifier consumes [redacted] of the total energy consumption.
2. Analog Front End: Energy consumed by the analog circuitry present in a radio base station. Assuming KPN base-station parameters for the IMEC power model, at full load for a bandwidth of 10 MHz, the analog front end consumes [redacted] of the total energy consumption.
3. Digital Baseband: Energy consumed by the digital processing operations (for e.g. MIMO Precoding, channel estimation). Assuming KPN base-station parameters for the IMEC power model, at full load for a bandwidth of 10 MHz, the digital baseband consumes [redacted] of the total energy consumption.
4. Digital Control: Energy consumed by processing operations such as backhaul and network processing. At full load for a bandwidth of 10 MHz, the digital consumes [redacted] of the total energy consumption.
5. Power Supply System: Energy consumed by AC/DC and DC/DC converters and it also accounts for integrated cooling. Assuming KPN base-station parameters for the IMEC power model, at full load for a bandwidth of 10 MHz, the power supply system consumes [redacted] of the total energy consumption.

Some base stations may have external air conditioner(s) present at the site which consume a significant amount of energy. However, macro sites in the KPN Network do not have external air conditioner(s) present, and the indoor (small cell) sites may have an air conditioner present but the indoor sites are out of scope. Therefore, the energy consumed by an air conditioner does not have any impact on this research. There is no other non-carrier specific component in the IMEC power model.

4.1.2. Reference Scenarios

The energy consumption and operational costs on effectuating the scheme(s) will be compared to the energy consumption and operational cost of the reference scenario(s) to calculate the energy and operational cost savings. The reference scenario(s) is the default operation mode i.e. when no resource(s) is (are) reduced or switched off. In Chapter 3, sites were selected for a pilot study on all three timescales. The energy consumption for default operation and on effectuating any scheme(s) for these sites is calculated from the IMEC power model [29] by inputting the mean site-carrier PRB utilization (per carrier, over all cells(sites) and hours under consideration, see Appendix B) and other default parameters into the power model (see Appendix A). The value of energy consumption per carrier in KWh on each timescale is obtained by inputting the energy consumption data from the IMEC power model into the following equation:

$$Energy(per-carrier) = \frac{N_S \times P \times T}{1000} kWh \tag{4.1}$$

where N_S denotes the number of sites (pilot sites), P denotes the average energy consumption (per site in W) at any instant and is obtained from the IMEC power model (Appendix A), and T denotes the number of hours for which the energy consumption is calculated. The total energy consumption at a site is equal to the sum of the energy consumption of each carrier present at the site.

For KPN, the cost of energy is [redacted] per kWh of consumption. The operational costs (energy costs) will be obtained by multiplying the energy consumption in kWh (from Equation (4.1)) with the energy cost per kWh ([redacted]).

The value of N_S and T for all the three timescales is shown in Table 4.1 below. On the hourly timescale, for the assessment of the schemes, samples (defined in Section 2.6) were collected from the pilot sites (N_S=47) for a period of seven days (one week, 16 - 22 July 2018) during the nighttime hours (8 hours between 23:00-07:00 hours). Therefore, in total, there are 7896 samples per carrier (47x3×7×8=7896). From these samples (per carrier), the mean PRB utilization (reference scenario) was calculated and input into the IMEC power model. The output of the IMEC power model was input into Equation (4.1) and the value of energy consumption in kWh was obtained and is shown in Table 4.2. Similarly, samples were also collected for the daily and monthly timescale and the reference scenario energy consumption is calculated and also shown in Table 4.2. Note that on the hourly timescale, while the schemes may only be effectuated for a period of eight hours, the energy consumption is shown for a period of 24 hours.

On the daily timescale, samples were collected (N_S=49) for a period of three weekends (six days, 7,8,14,15,21 and 22 July 2018) during all the hours of the day (24). Therefore, in total, there are 21168 samples (49×3×6×24=21168) per carrier. On the monthly timescale (N_S=48), samples were collected for a period of one week (seven days, 24-30 June 2018) during all the hours of the day (24). Therefore, in total, there are 24192 samples (48×3×7×24=24192) per carrier.

Timescale	N_S	T
Hourly	47	24
Daily	49	48
Monthly	48	168

Table 4.1: Value of N_S and T across all the three timescales

Timescale	Energy consumption per carrier in kWh		Total energy consumption (coverage+capacity) in kWh
	Coverage	Capacity	
Hourly	102.64	342.12	444.76
Daily	203.82	685.37	889.19
Monthly	822.52	2628.86	3451.38

Table 4.2: Reference scenario energy consumption for all the three timescales

From the samples collected across all the three timescales, the initial number of hours of congestion (before any scheme effectuation) was calculated using a throughput-based check and a PRB utilization-based check (described in Section 4.1) and is shown in Tables 4.3 and 4.4. Note that for the hourly timescale, the reference scenario congestion level is shown for a period of eight hours only as the scheme(s) at the hourly timescale may only be effectuated for a time-interval of eight hours.

Note that for the monthly timescale in Table 4.4, the throughput distribution was unavailable for one day (24 June 2018) for all the sites. Therefore, the congestion levels calculated by the throughput-based check for the reference scenario and as well as on effectuating schemes, are representative for six days (25-30 June 2018) only.

Timescale	Number of hours of congestion per carrier		Total number of hours of congestion (coverage+capacity)
	Coverage	Capacity	
Hourly	0	0	0
Daily	0	0	0
Monthly	121	22	143

Table 4.3: Reference scenario: Initial number of hours of congestion calculated from a PRB utilization-based check across all sites samples under consideration

As seen from Tables 4.3 and 4.4, the PRB utilization-based and the throughput-based checks on minimum QoS conditions give significantly different results. The throughput-based check reports a significantly higher number of hours of congestion experienced when compared with the PRB utilization-based check (for all the three timescales). A reason for this dissimilarity could be that, in actuality, majority of the sessions are

Timescale	Number of hours of congestion per carrier		Total number of hours of congestion (coverage+capacity)
	Coverage	Capacity	
Hourly	12	0	12
Daily	54	0	54
Monthly	913	45	958

Table 4.4: Reference scenario: Initial number of hours of congestion calculated from a throughput-based check across all samples under consideration

text-message sessions which have lower throughputs by nature, and this leads to a false-positive congestion label. As the session type (whether file-transfer or text-message) cannot be checked by the network operators, the assumption that all sessions are meaningful-file transfers is made so that there is no missed-detection of (genuine) congestion.

As the PRB utilization-based check is based on the PRB utilization threshold which was derived under the assumption of file-transfers sessions ([27]), the PRB utilization-based check is more reliable for quantifying the number of hours of congestion experienced as it does not lead to false-positive congestion labels nor missed-detections of congestion. Note that the dissimilarity between the congestion levels reported by the two methods has no impact on energy/cost savings. The congestion levels reported by both the methods is quantified and presented and it is up to the operator to deem which check (method) is more reliable. The recommendation we give in this thesis is that the PRB utilization-based check is more reliable than the throughput-based check due to the reasons mentioned above.

4.2. Turning off carriers

As already mentioned, sites in a cellular network may have multiple carriers present. This scheme can only be applied at sites which have multiple carriers present. In this scheme, the capacity layer cells are switched off and their traffic is transferred to the sister cells of the coverage layer. The source of operational cost savings in this scheme is a reduction in energy consumption at a site. On turning off the capacity layer (all three cells) at a site, the fixed energy consumption of the capacity layer reduces significantly. The fixed energy consumption drops from 128 W to 7.51 W for carriers with TMA present at zero PRB utilization, as obtained from [29]) and is the source of the reduction in energy consumption at a site. On offloading the traffic from the capacity to the coverage layer cells (after turning off the capacity layer), the energy consumption of the coverage layer increases. This increase occurs as the energy consumption scales approximately linearly with PRB utilization.

However, this increase in energy consumption for the coverage layer is of a smaller magnitude than that of the reduction in energy consumption for the capacity layer. For e.g. assuming a *worst case* scenario with an average PRB utilization of 50% (50 PRBs used out of 100 available PRBs) for the capacity carrier, average PRB utilization of 0% for the coverage carrier (0 PRBs used out of 50 available PRBs) and no improvement in spectral efficiency (in reality there will be an improvement in spectral efficiency due to improved propagation conditions as the coverage carrier is operated at a lower frequency) on offloading the traffic. The new avaerage PRB utilization would be 100% (50 PRBs used out of 50 available PRBs) for the coverage carrier and undefined for the capacity carrier as it is turned off. In this *worst case* scenario, the energy consumption for the capacity carrier reduces from [redacted] W reduction) and for the coverage carrier it increases from [redacted] W increase). Overall, the energy consumption reduces by [redacted] W). In real-scenarios, the energy consumption reduces by a significantly higher magnitude as the increase in average PRB utilization of the coverage carrier will be lower that what is illustrated in the *worst case* scenario here.

In [33], the file size (bytes) is mapped to PRB utilization by the following equation:

$$\rho = \frac{1}{180} \frac{\lambda S}{\eta R_{prb}} \tag{4.2}$$

where ρ denotes the PRB utilization, λ denotes the flow arrival rate (1/sec), S denotes the average file size, η denotes the spectral efficiency and R_{prb} denotes the PRB rate (number of PRBs per second).

From the above equation, it can be seen that the the overall impact on ρ is unclear when traffic is offloaded from the capacity to the coverage cell(s). While spectral efficiency (η) would increase on offloading, the PRB rate (R_{prb}) would decrease (halved) as there are fewer PRBs available per second on the coverage cell(s). Let Equation (4.2) represent the PRB utilization for the capacity cell(s) (20 MHz bandwidth).

On offloading the traffic (same λ and S), let ρ' denote the new PRB utilization (coverage cell(s)), R'_{prb} denotes the new PRB rate which is equal to $\frac{R_{prb}}{2}$ and η' denotes the new spectral efficiency ($\eta' \geq \eta$). Then from Equation (4.2),

$$\begin{aligned} \rho' &= \frac{1}{180}\frac{\lambda S}{\eta' R'_{prb}} \\ &= \frac{2}{180}\frac{\lambda S}{\eta' R_{prb}} \end{aligned} \tag{4.3}$$

On dividing Equation (4.3) by Equation (4.2),

$$\begin{aligned} \frac{\rho'}{\rho} &= \frac{2\eta}{\eta'} \\ \rho' &= \rho\frac{2\eta}{\eta'} \end{aligned} \tag{4.4}$$

From Equation (4.4), it can be seen that the new PRB utilization ρ' is a function of ρ, η, η' and ρ' is greater than ρ when $\eta' < 2\eta$. The *worst case* scenario illustrated earlier in this section had the condition that $\eta' = \eta$, and thus resulted in a two-fold increase in the PRB utilization (50% to 100%).

From Equation (4.2), for a constant η, it can be seen that the PRB utilization is a linear function of data traffic volume (λS). However, η is not a constant in live-networks, and as a consequence PRB utilization (ρ) is not a linear function of data traffic volume (λS). As discussed in [33], there are two important effects which impact spectral efficiency which in turn impacts PRB utilization. The first effect is the change in traffic volume in the neighboring cell(s). An increase in interference due to an increase in traffic volume in the neighboring cell(s) can degrade the spectral efficiency in the serving cell, and thus lead to a higher PRB utilization. The second effect is that an increase in traffic volume in the serving cell leads to an improvement in the spectral efficiency due to multi-user diversity gain. Multi-user diversity gain is defined in [34] as “the gain in cell or user throughput that can be achieved by intelligently choosing which of multiple present users to serve in any transmission time interval, based on their individual instantaneous radio link qualities”. The collective impact of both these effects, as described in [33] “is that if the serving cell load increases at the same time with the neighbour cell load, degradation in spectral efficiency is partially compensated by multiuser scheduling gain.”

As the improvement in spectral efficiency (on offloading η increases to η') is complex to establish in a live-network, and to account for the variation in η due to the two effects mentioned above, a different approach is used for the assessment of the scheme of *turning off carriers*. The assessment of this scheme is carried out by modeling the relationship (curve-fitting) between the average PRB utilization and the average traffic volume for the coverage layer cell(s) from the live-network data samples.

Note that coverage is not impacted by this scheme as the coverage layer remains intact (switched on). The tooling to *turn off carriers* is already available at KPN, and the cost to implement this scheme is negligible in monetary terms. Time spent by KPN to prepare the tooling for execution by setting N_S and T (see Table 4.1) is the only cost incurred in implementing this scheme.

A drawback of this scheme is that the user throughputs would reduce in general as the coverage layer has fewer resources available (half of capacity layer). However, this reduction is acceptable as long as the minimum QoS target is met.

The analysis is only carried out for the sites that were selected for the pilot study (see Chapter 3). These sites have both the coverage and the capacity carrier present. The assessment of the scheme of *turning off carriers* requires a necessary assumption to be made. The assumption is as follows:

1. The relationship modeled (curve-fitting) between PRB utilization and traffic volume is identical for all the sites selected for the pilot study. A different relationship is not modeled for each site but instead a single relationship is modeled using data samples from all the sites collectively.

This assumption is necessary for the following two reasons:

1. The relationship between PRB utilization and traffic volume is modeled only using samples where the time-average number of active users in the downlink (KPI Active users) are [redacted]. This condition is scarcely met as the scheme may only be applied during periods when traffic volume is at its lowest levels (as a result the time-average number of active users in the downlink are also low) and there would not be enough data samples present to model the relationship if a per site solution is desired.

2. To provide a uniform network-wide solution. A different model between PRB utilization and traffic volume for each site would considerably increase the time spent by KPN on the tooling to execute this scheme.

The procedure for calculating the trade-off between the operational cost (energy consumption) savings and the experienced number of hours of congestion is defined below. The following steps carried out for each hour and each cell separately, and aggregated in the end.

1. The metric and KPIs of traffic volume, PRB utilization and the time-average number of active users in the downlink are calculated per hour for all the coverage and capacity layer cells of the sites selected for the pilot. The metrics and KPIs are calculated only for the fixed time-period when the scheme may be active (for e.g. between 23:00-07:00 hours on the hourly timescale).

2. The data traffic volume (in Megabits) is available per hour. The data traffic volume metric for all the samples is divided by 3600 (3600 seconds in one hour) to obtain the average data traffic volume with a unit of Megabits per second (Mbps).

3. The PRB utilization for all the coverage layer cells is used to calculate the mean site-carrier PRB utilization for the coverage carrier (one carrier component at a site). Similarly the mean site-carrier PRB utilization is also calculated for the capacity carrier using all the capacity layer cells (second carrier component at a site). Both the calculated mean site-carrier PRB utilizations (coverage and capacity layer) are input individually into the IMEC power model to obtain the mean energy consumption which is then input into Equation (4.1). Note that the mean of the PRB utilization of the three cells associated with a carrier at site equals the carrier PRB utilization for that site. The carrier PRB utilization of different sites is then used to calculate a mean site-carrier PRB utilization. For e.g. the number of PRBs available is identical for all the the three cells (same carrier) at a site for a period of one hour, the only difference is in the number of utilized PRBs. X denotes the number of PRBs available for each cell, A,B and C denote the number of PRBs utilized by the three cells at the site, then:

$$\begin{aligned} Mean-PRB-Utilization-Cells &= \frac{\frac{A}{X}+\frac{B}{X}+\frac{C}{X}}{3} \\ &= \frac{A+B+C}{3X} \\ Carrier-PRB-Available &= X+X+X \\ &= 3X \\ Carrier-PRB-Used &= A+B+C \\ Carrier-PRB-Utilization &= \frac{A+B+C}{3X} \end{aligned}$$

Note that the energy consumption and operational cost calculated in this step is presented in the reference scenario in Section 4.1.2.

4. All the samples are filtered with the condition that the time-average number of active users in the downlink are [redacted] (samples defined in Section 2.6).

5. The number of filtered samples (Step 4) which exceed the PRB utilization threshold defined in Section 2.7 are calculated. This number represents the initial number of hours of congestion experienced before scheme activation and is also presented in the reference scenario in Section 4.1.2.

6. Using all the filtered coverage layer samples, a relationship is modeled between PRB utilization and traffic volume (curve fitting using MATLAB [35]). A similar model is not needed for the capacity layer as the capacity layer will be turned off and its PRB utilization will be zero (or undefined as there would be zero resources available).

7. The average traffic volume (Mbps) carried on the capacity layer cell(s) is added to the traffic volume carried on the sister coverage layer cell(s) for all the samples (not just filtered samples). The reason for taking all the samples is given further below.

8. The PRB utilization for the coverage layer cell(s) when the capacity layer traffic is added to the coverage layer (Step 7) is calculated from the relationship modeled in Step 6 above (all samples and not just filtered samples).

9. The number of samples which exceed the PRB utilization threshold (Section 2.7) are calculated along with the mean site-carrier PRB utilization. This number represents the number of hours of congestion that may be experienced on effectuating the scheme.

10. The mean-carrier PRB utilization calculated in Step 9 above is input into the IMEC power model, the output of the power model is input into Equation (4.1) to obtain the energy consumption in kWh for the coverage carrier. The capacity layer cells are assumed to be in shut-down mode for calculating the energy consumption of the capacity carrier (from the IMEC power model and Equation (4.1)). Note that all samples are used in this step, not just filtered ones.

11. By comparing the number of hours of congestion experienced, energy consumption and operational costs (Step 10) with the value of these metrics presented in Section 4.1.2, the operational cost and energy savings along with number of hours of congestion experienced due to *turning off carriers* is calculated.

Note that the minimum QoS condition cannot be explicitly checked in terms of the [redacted]) as the throughput distribution is unknown once traffic is added from the capacity layer to the coverage layer cell(s). For congestion, there is a condition that the time-average number of active users in the downlink are [redacted]. When traffic from the capacity layer cells is added to the coverage layer cells (assuming capacity layer is turned off), the time-average number of active users in the downlink for the coverage layer cell(s) becomes unknown. There is no information available about at which time slots the users were active on the capacity layer cell(s) and the coverage layer cell(s). Therefore, a worst case assumption is made, and the assumption is that for the coverage layer cell(s), the time-average number of active users in the downlink are [redacted] for all samples when traffic from capacity layer cell(s) is added to the coverage layer cell(s).

The relationship between PRB utilization and traffic volume is modeled using only the filtered samples (the time-average number of active users in the downlink are [redacted]). This is done as the other samples (the time-average number of active users in the downlink are [redacted]) are not representative of the traffic dynamics in the cell(s) as very few active users were present in the cell(s) during those hours (1 sample = 1 hour). The mean PRB utilization is always calculated using all the samples. The reason for this is that this metric is used (only) for determining the energy consumption and energy is consumed irrespective of whether the cell(s) (sample(s)) is (are) in congestion or not.

A simplified illustrative example for the above described exercise (Steps 1-11) is as follows. Assume that we have 90 samples per carrier, each with PRB utilization ρ_i^c, traffic volume (Mb) v_i^c, the time-average number of active users in the downlink τ_i^c where $i \in \{1,2,3,\ldots,90\}$ (number of samples) and $c \in$ {coverage, capacity} (two carriers).
In step two of the exercise, v_i^c is divided by 3600 to obtain the average traffic volume (Mbps) represented by μ_i^c.

In step three, the mean site-carrier PRB utilization (coverage) is calculated from $\rho_i^{coverage}$ and the mean site-carrier PRB utilization (capacity) is calculated from $\rho_i^{capacity}$, and these metrics are used to establish the reference scenario energy consumption.
In step four, samples are filtered with the condition that the time-average number of active users in the down-link are [redacted]. Assume that only the first 25 samples (on each carrier) meet this condition. Then we would be left with ρ_i^c, μ_i^c and τ_i^c where $i \,\epsilon\, \{1,2,3,\ldots,25\}$ and $c \,\epsilon\,$ {coverage, capacity}.
In step five, ρ_i^c where $i \,\epsilon\, \{1,2,3,\ldots,25\}$ and $c \,\epsilon\,$ {coverage, capacity}, it is measured whether ρ_i^c exceeds the PRB utilization threshold limits (see Section 2.7) and this establishes the initial number of hours of congestion for the reference scenario.
In step six, $\rho_i^{coverage}$ and $\mu_i^{coverage}$ where $i \,\epsilon\, \{1,2,3,\ldots,25\}$ are used to model a relationship (curve-fitting).
In step seven, $\mu_i^{coverage}$ and $\mu_i^{capacity}$ are added together to obtain $\xi_i^{coverage}$, where $i \,\epsilon\, \{1,2,3,\ldots,90\}$.
In step eight, $\xi_i^{coverage}$ where $i \,\epsilon\, \{1,2,3,\ldots,90\}$ is input into the model (step six) to obtain the new PRB utilization $\varrho_i^{coverage}$.
In step nine, $\varrho_i^{coverage}$ where $i \,\epsilon\, \{1,2,3,\ldots,90\}$, it is measured whether $\varrho_i^{coverage}$ exceeds the PRB utilization threshold limits (see Section 2.7) and this establishes the number of hours of congestion that may be experienced on *turning off carriers*.
In step ten, the mean site-carrier PRB utilization (coverage) is calculated from $\varrho_i^{coverage}$ where $i \,\epsilon\, \{1,2,3,\ldots,90\}$ and this metric is used to obtain the energy consumption. The energy consumption for the capacity carrier is also calculated assuming shut down mode in the power model.
In step eleven, the results obtain in steps nine and ten are compared with the reference scenario to establish the trade-off between the operational cost savings and the experienced number of hours of congestion.

4.2.1. Hourly Timescale

On the hourly timescale, there are 47 sites selected for the pilot study. All these sites have both the coverage and the capacity layer present. Effectively, there are 47×3= 141 cells per layer (282 cells in total) at all the sites combined. The analysis is carried out using the procedure defined in the previous section. For the analysis, samples were collected from these 47 sites for a period of seven days (one week, 16 July- 22 July 2018) during the nighttime hours (eight hours between 23:00-07:00 hours). Therefore, in total, there are 7896 samples per layer (141×7×8=7896).

As explained in Section 4.2, a throughput-based check for minimum QoS conditions cannot be carried out for the scheme *turning off carriers*. Only PRB utilization-based check(s) will be carried out for establishing the number of hours of congestion that may be experienced on effectuating the scheme. From the data samples obtained, zero hours of congestion were observed in total (including both the coverage and the capacity layer cell(s)). Therefore, during default operation before turning off the capacity layer, there was no congestion experienced.

On filtering the 7896 samples of the coverage layer with the condition that the time-average number of active users in the downlink are [redacted], 61 samples remained. These samples were used to model the relationship between PRB utilization and traffic volume (for the coverage layer). The scatter plot of these samples is shown in Figure 4.2 along with a cubic fit and its 95% confidence bounds. The curve was fitted to the scatter plot using the MATLAB curve fitting toolbox [35], the toolbox uses the method of linear least squares to fit a curve. For the fitted curve, as defined in [35], the confidence bounds are calculated using the equation "$C = b \pm t\sqrt{S}$ where b denotes the coefficients produced by the fit, t is calculated using the Student's t cumulative distribution function, and S is a vector of the diagonal elements from the estimated covariance matrix of the coefficient estimates."

[redacted]

Figure 4.2: Cubic fit with 95% confidence bounds for PRB Utilization vs Traffic Volume on the hourly timescale

The coefficients of the cubic fit and its 95% confidence bounds (in brackets) are as follows:

[redacted] (4.5)

The R^2 value of the cubic curve fitted in the above figure is calculated to be 0.8221. This indicates that the fitted curve is good as 82% of the variance in data around the mean is accounted for by the fit. As the sample-size is small (61 samples), the confidence interval is wide (as seen from Figure 4.2). The sample size may be increased by considering a larger set of sites and/or collecting samples over a longer period of time. As the procedure for the assessment of the schemes is planned to be executed every day/week/month (see Section 4.1) with the latest data-samples, the sample-size may only be increased by considering a larger set of sites. The confidence bounds (Figure 4.2) are also used for the assessment of the scheme to find the bounds on number of hours of congestion (sensitivity analysis) that may be experienced. The cubic fit will be referred to as $f(x)$, the lower confidence bound will be referred to as $f_l(x)$ and the upper confidence bound curve will be referred to as $f_u(x)$ from this point forward.

As seen from Figure 4.2, for values of the traffic volume < 5 Mbps, the fitted curve is almost linear. This indicates that at low traffic volumes, an increase in PRB utilization would lead to a proportional increase in traffic volume. At values of the traffic volume between 5 to 15 Mbps, the fitted curve is not linear anymore. An increase in PRB utilization is accompanied with a more than proportional increase in traffic volume. A reason for this observation could be that an increase in traffic volume improves the system performance via multi-user diversity gain [33]. At values of the traffic volume > 15 Mbps, an increase in PRB utilization is accompanied with a less than proportional increase in traffic volume. A reason for this observation could be that as the traffic volume in the serving cell increases, the traffic volume in the neighboring cell(s) also increases. This can lead to increased interference for the serving cell and thus a degradation in spectral efficiency [33]. Therefore, more PRBs have to be utilized for the same increase in traffic volume as before (0-15 Mbps regime). Another reason for this observation could be that this is just an artifact of the cubic-fit trying to account for the points around 15 Mbps.

On executing the procedure described in the previous section (Steps 8-9) with the model represented by Equation (4.5) and Figure 4.2, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.5.

Model	Reference Scenario: Initial number of hours of congestion	Number of hours of congestion on scheme effectuation
$f(x)$	0	7
$f_l(x)$	0	0
$f_u(x)$	0	65

Table 4.5: The number of hours of congestion that may be experienced on offloading the traffic carried on the capacity layer to the coverage layer cell(s) on the hourly timescale. Calculated by carrying out a PRB utilization-based check on the output of the model in Equation (4.5)

As seen from Table 4.5, the number of hours of congestion experienced on scheme effectuation ranges from 0-65 hours out of total of 7896 hours of operation. Therefore, the minimum QoS requirements are met for 7831-7896 hours with a 95% likelihood.

Initially, the mean site-carrier PRB utilization for the coverage layer cells (using all samples) was found out to be [redacted] and for capacity layer cells it was [redacted]. This mean site-carrier PRB utilization is input into the power model to obtain the reference energy consumption and operational cost. On offloading the traffic volume originally carried on the capacity layer to the sister coverage layer cell(s), the mean site-carrier PRB utilization using $f(X)$ was calculated to be [redacted]. The mean site-carrier PRB utilization using $f_u(x)$ was calculated to be [redacted]. The increase in PRB utilization of the coverage layer is expected due to an increase in traffic volume of the coverage layer cell(s).

The PDF of PRB utilization (obtained from the data-samples) for the coverage and the capacity layer initially (before traffic was offloaded to the coverage layer from the capacity layer), and for the coverage layer after scheme effectuation (using $f(x)$) is shown in Figure 4.3 below. From the figure, it can be seen that more than 90% (0.18×5, bin-width is 5) of the samples have PRB utilization between 0-5% before scheme effectuation. After scheme effectuation, more than 70% of the samples have PRB utilization between 0-5%.

The mean site-carrier PRB utilization of the coverage layer obtained after offloading the traffic carried on the capacity layer to the coverage layer is used to obtain the energy consumption in kWh using the IMEC power model [29] and Equation (4.1). The mean site-carrier PRB utilization is calculated using $f(x)$ and $f_u(x)$ only. The optimistic lower confidence bound $f_l(x)$ is not considered as we want to be conservative in estimating the energy savings. The capacity layer is assumed to be in shut-down mode for calculating the energy consumption in kWh (using the IMEC power model and Equation (4.1)). Note that the scheme may only be applied for a period of eight hours, however the energy consumption is shown for a period of 24 hours. The 24 hour energy consumption on scheme effectuation is obtained by subtracting the reduction(or gain) in energy consumption for the eight hour time period from the 24 hour reference scenario. The energy consumption in kWh and operational costs is presented for a period of 24 hours in Table 4.6.

On effectuating the scheme of *turning off carriers*, the total energy consumption and operational cost are 343.14 kWh and [redacted] using $f(x)$ and 348.85 kWh and [redacted] using curve $f_u(x)$. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 95.91 to 101.62 kWh are attainable with a 95% likelihood. Consequently, operational cost savings of [redacted] are also attainable with the same likelihood. The energy consumption and operational cost both reduce by 21.56% to

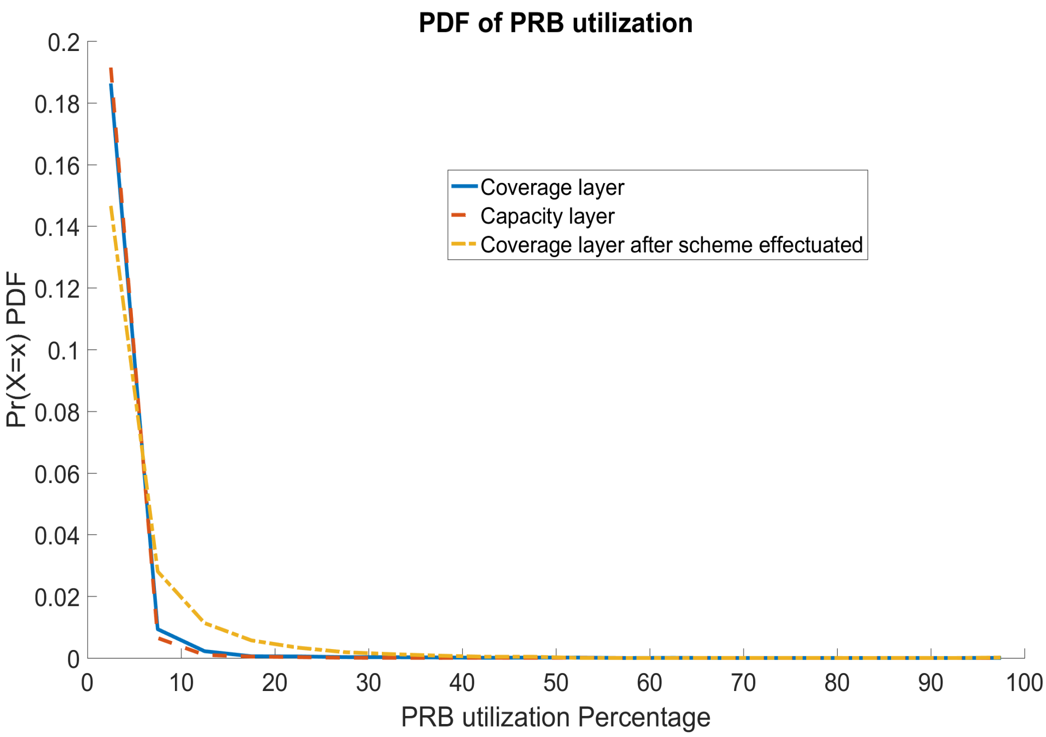


Figure 4.3: PDF of initial PRB utilization for i) Coverage layer ii) Capacity layer, and PDF of PRB utilization for iii) coverage layer after scheme effectuation using $f(x)$, on the hourly timescale

	Reference Scenario	Cubit fit $f(x)$	Upper confidence bound $f_u(x)$
Energy Consumption Coverage Layer	102.64 kWh	103.63 kWh	109.34 kWh
Energy Consumption Capacity Layer	342.12 kWh	239.51 kWh	239.51 kWh
Total Energy Consumption	444.76 kWh	343.14 kWh	348.85 kWh
Total Operational Cost	[redacted]	[redacted]	[redacted]

Table 4.6: Total energy consumption and operational cost after effectuating the scheme *turning off carriers* on the hourly timescale. Shown for 47 sites and for a time period of 24 hours (one day) for $f(x)$ and $f_u(x)$

22.84% compared to the reference scenario values. Note that these savings are for 47 sites (pilot sites) and for a duration of 24 hours only (one day).

Presently, the tooling that is available at KPN to *turn off carriers*, will turn-off the capacity carrier(s) for all eight hours in the night irrespective of whether there would be congestion on the coverage carrier on turning-off the capacity carrier or not. However, the savings in energy consumption and operational cost will occur for all eight hours as the capacity carrier is turned-off for all eight hours.

In the future, there may be tooling available which can make a decision on whether to turn-off capacity carriers or not per hour every night (eight hours). If such a tooling is employed, there would be no (additional) congestion experienced as the capacity carrier will not be turned-off if it leads to congestion on the coverage carrier. However, the savings in energy consumption and operational cost may be less than the savings attainable with the presently available tooling, as savings will not occur when turning-off the capacity carrier would lead to congestion on the coverage carrier.

4.2.2. Daily Timescale

The analysis for the daily timescale is carried out in an identical manner. Therefore, only the relevant and final results and figures are presented in this section. On the daily timescale there are 49 sites selected for the pilot study. All these sites have both the coverage and the capacity layer present. Effectively, there are 49×3=147 cells per layer (294 cells in total). The data samples were collected from these sites for a period of three weekends (six days, 7,8,14,15,21 and 22 July 2018) during all the hours of the day (24). Therefore, in total, there are 21168 samples (49×3×6×24=21168) per layer.

There was no congestion experienced in total (including both the coverage and the capacity layer) during default operation (before scheme effectuation). On filtering the coverage layer samples with the condition that the time-average number of active users in the downlink are [redacted], 217 samples remained. These samples were used to model the relationship between PRB utilization and traffic volume. A relationship is modeled again as the pilot sites on the daily timescale are different from the pilot sites on the hourly timescale (on all three timescales the pilot sites are different with little overlap). The scatter plot of these samples along with a power law fit and its 95% confidence bounds are shown below in Figure 4.4.

[redacted]

Figure 4.4: Power law fit with 95% confidence bounds for PRB Utilization vs Traffic Volume on the daily timescale

The coefficients of the power law fit and its 95% confidence bounds (in brackets) are as follows:

[redacted] (4.6)

The power law fit be referred to as $f(x)$, the lower confidence bound curve will be referred to as $f_l(x)$ and the upper confidence bound curve will be referred to as $f_u(x)$.

On executing the procedure described in the previous section (Steps 8-9) with the model represented by Equation (4.1) and Figure 4.4, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.7.

Model	Reference Scenario: Initial number of hours of congestion	Number of hours of congestion on scheme effectuation
$f(x)$	0	6
$f_l(x)$	0	0
$f_u(x)$	0	16

Table 4.7: The number of hours of congestion that may be experienced on offloading the traffic carried on the capacity layer to the coverage layer cell(s) on the daily timescale. Calculated by carrying out a PRB utilization-based check on the output of the model in Equation (4.6)

Initially, the mean site-carrier PRB utilization for the coverage layer cells (using all samples) was found out to be [redacted] and for capacity layer cells it was [redacted]. On offloading the traffic volume originally carried on the capacity layer to the sister coverage layer cell(s), the mean site-carrier PRB utilization using $f(x)$ was calculated to be [redacted]. The mean site-carrier PRB utilization using $f_u(x)$ was calculated to be [redacted]. The optimistic lower confidence bound $f_l(x)$ is not considered as we want to be conservative in estimating the energy savings. The PDF of initial PRB utilization of the coverage and the capacity layer, and of the PRB utilization of the coverage layer after scheme effectuation (using $f(x)$) is shown below in Figure 4.5.

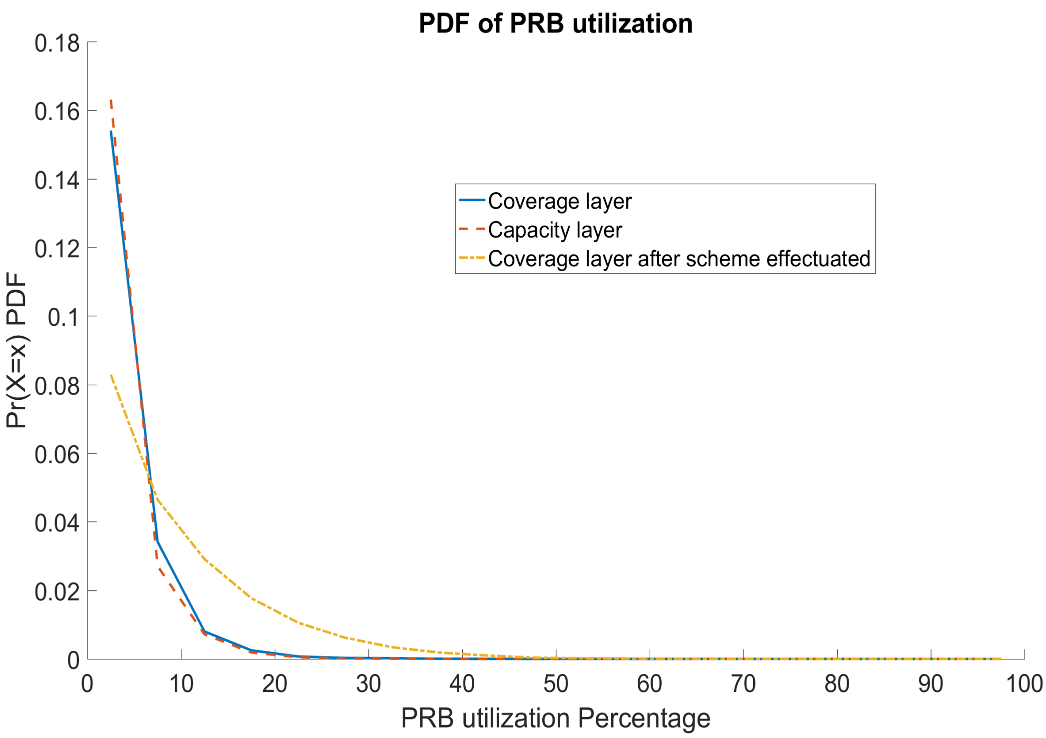


Figure 4.5: PDF of initial PRB utilization for i) Coverage layer ii) Capacity layer, and PDF of PRB utilization for iii) coverage layer after scheme effectuation using $f(x)$, on the daily timescale

The mean site-carrier PRB utilization of the coverage layer obtained after offloading the traffic carried on the capacity layer to the coverage layer is used to obtain the energy consumption in kWh using the IMEC power model [29] and Equation (4.1). The energy consumption in kWh and operational costs is presented for

	Reference Scenario	Power law fit $f(x)$	Upper confidence bound $f_u(x)$
Energy Consumption Coverage Layer	203.82 kWh	233.08 kWh	239.90 kWh
Energy Consumption Capacity Layer	685.37 kWh	27.75 kWh	27.75 kWh
Total Energy Consumption	889.19 kWh	260.83 kWh	267.65 kWh
Total Operational Cost	[redacted]	[redacted]	[redacted]

Table 4.8: Total energy consumption and operational cost after effectuating the scheme *turning off carriers* on the daily timescale. Shown for 49 sites and for a time period of 48 hours (one weekend) for $f(x)$ and $f_u(x)$

a period of 48 hours in Table 4.8.

On effectuating the scheme of *turning off carriers*, the total energy consumption and operational cost are 260.83 kWh and [redacted] using $f(x)$ and 267.65 kWh and [redacted] using curve $f_u(x)$. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 621.54 to 628.36 kWh are attainable with a 95% likelihood. Consequently, operational cost savings of [redacted] are also attainable with the same likelihood. The energy consumption and operational cost both reduce by 69.89% to 70.66% compared to the reference scenario values. Note that these savings are for 49 sites (pilot sites) and for a duration of 48 hours only (one weekend).

4.2.3. Monthly timescale

The analysis for the daily timescale is carried out in an identical manner. Therefore, only the relevant and final results and figures are presented in this section. On the monthly timescale there are 48 sites selected for the pilot study. All these sites have both the coverage and the capacity layer present. Effectively, there are 48×3=144 cells per layer (288 cells in total). Data samples were collected from these sites for a period of one week (seven days, 24-30 June 2018) during all the hours of the day (24). Therefore, in total, there are 24192 samples (48×3×7×24=24192) per layer.

Initially, there was 143 hours of congestion experienced in total (121 hours on the coverage layer and 22 hours on the capacity layer) during default operation (before scheme effectuation). On filtering the coverage layer samples with the condition that the time-average number of active users in the downlink are [redacted], 2857 samples remained. These samples were used to model the relationship between PRB utilization and traffic volume. A relationship is modeled again as the pilot sites on the monthly timescale are different from the pilot sites on the hourly/daily timescale (on all three timescales the pilot sites are different with little overlap). The scatter plot of these samples along with a power law fit and its 95% confidence bounds are shown below in Figure 4.6. The coefficients of the power law fit and its 95% confidence bounds (in brackets) are as follows:

[redacted] (4.7)

The R^2 value of the fitted curve is 0.5937, that is only 59% of the variance in the samples around the mean is accounted for by the fit. As the curve is fitted on a real live network dataset, it is reasonable to expect that the model would not be perfectly fitted with high values of R^2. The power law curve fitted and shown in Figure 4.6 and represented by Equation (4.7) is the best estimate for the relationship between PRB utilization and traffic volume at the monthly timescale. As seen from the figure, the confidence intervals are narrower when compared with the other two timescales. As there are a high number of samples (2857), it is expected that the confidence interval would be narrow. Note that for calculating the coefficients of the fit, the values of

[redacted]

Figure 4.6: Power law fit with 95% confidence bounds for PRB Utilization vs Traffic Volume on the monthly timescale

PRB utilization exceeding [redacted] have been excluded, as by definition, these samples are already in congestion.

Compared to the previous subsections on the hourly and daily timescale, the number of filtered samples is significantly higher for the monthly timescale (2857 samples compared to 61 and 217 samples for the hourly and daily timescale). This indicates that there are a higher time-average number of active users in the downlink at the monthly timescale when compared to the other two timescales.

The power law fit will be referred to as $f(x)$, the lower confidence bound curve will be referred to as $f_l(x)$ and the upper confidence bound curve will be referred to as $f_u(x)$.

On executing the procedure described in the previous section (Steps 8-9) with the model represented by Equation (4.1) and Figure 4.6, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.9. As during the normal operation mode

Model	Reference Scenario: Initial number of hours of congestion (total)	Number of hours of congestion on scheme effectuation
$f(x)$	143 (121+22)	600
$f_l(x)$	143 (121+22)	375
$f_u(x)$	143 (121+22)	823

Table 4.9: The number of hours of congestion that may be experienced on offloading the traffic carried on the capacity layer to the coverage layer cell(s) on the monthly timescale. Calculated by carrying out a PRB utilization-based check on the output of the model in Equation (4.7)

(before adding traffic from capacity layer to coverage layer) there was 143 hours of congestion experienced in total, the effective number of hours of congestion experienced on scheme effectuation is obtained by subtracting 143 from the number of hours of congestion experienced on scheme effectuation.

Initially, the mean site-carrier PRB utilization for the coverage layer cells (using all samples) was found out to be [redacted] and for capacity layer cells it was [redacted]. On offloading the traffic volume originally carried on the capacity layer to the sister coverage layer cell(s), the mean site-carrier PRB utilization using $f(x)$ was calculated to be [redacted]. The mean site-carrier PRB utilization using $f_u(x)$ was calculated to be [redacted]. The optimistic lower confidence bound $f_l(x)$ is not considered as we want to be conservative in estimating the energy savings. The PDF of initial PRB utilization of the coverage and the capacity layer, and of the PRB utilization of the coverage layer after scheme effectuation (using $f(x)$) is shown below in Figure 4.7.

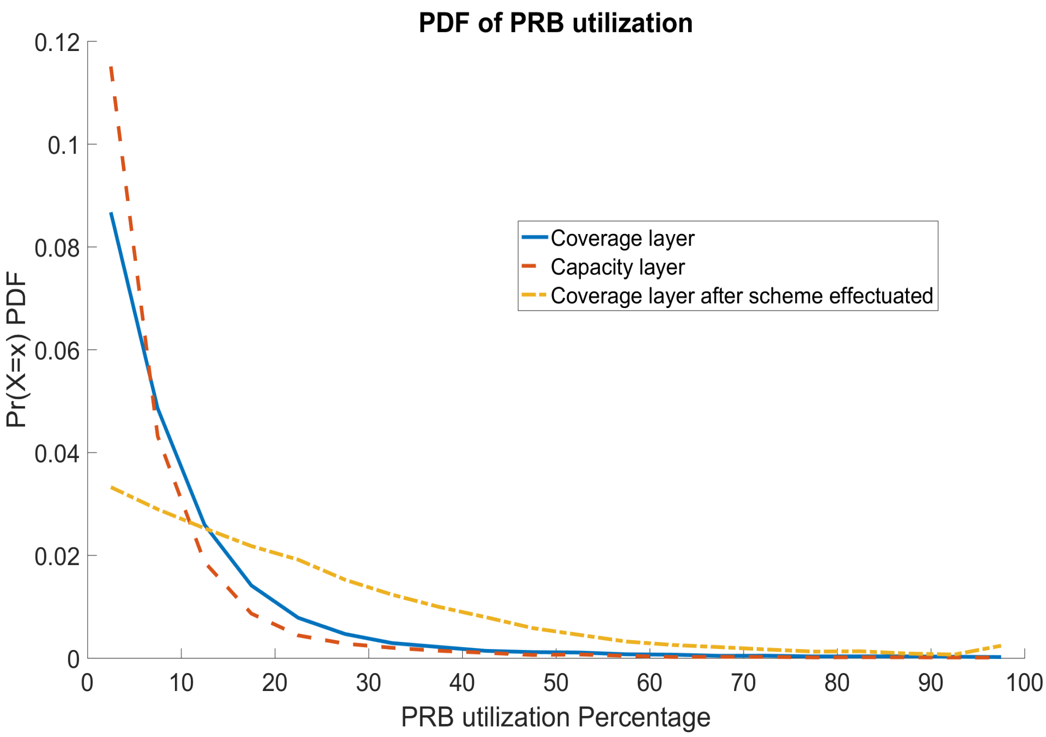


Figure 4.7: PDF of initial PRB utilization for i) Coverage layer ii) Capacity layer and PRB utilization after scheme effectuation using curve C of iii) Coverage layer, on the monthly timescale

The mean site-carrier PRB utilization of the coverage layer obtained after offloading the traffic carried on the capacity layer to the coverage layer is used to obtain the energy consumption in kWh using the IMEC power model [29] and Equation (4.1). The energy consumption in kWh and operational costs is presented for a period of 168 hours in Table 4.10.

	Reference Scenario	Power law fit $f(x)$	Upper confidence bound $f_u(x)$
Energy Consumption Coverage Layer	822.52 kWh	1153.15 kWh	1177.34 kWh
Energy Consumption Capacity Layer	2628.86 kWh	95.15 kWh	95.15 kWh
Total Energy Consumption	3451.38 kWh	1248.3 kWh	1272.49 kWh
Total Operational Cost	[redacted]	[redacted]	[redacted]

Table 4.10: Total energy consumption and operational cost after effectuating the scheme *turning off carriers* on the monthly timescale. Shown for 48 sites and for a time period of 168 hours (one week) for $f(x)$ and $f_u(x)$

On effectuating the scheme of *turning off carriers*, the total energy consumption and operational cost are 1248.3 kWh and [redacted] using $f(x)$ and 1272.49 kWh and [redacted] using curve $f_u(x)$. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 2178.89 to 2203.08 kWh are attainable with a 95% likelihood. Consequently, operational cost savings of [redacted] are also attainable with the same likelihood. The energy consumption and operational cost both reduce by 63.13% to 63.83% compared to the reference scenario values. Note that these savings are for 48 sites (pilot sites) and for a duration of 168 hours only (one week).

4.2.4. Conclusion

The scheme of turning off carrier(s) was described and analyzed in the previous subsections. As seen from the results presented, effectuating this scheme can lead to significant energy and operational cost savings. The percentage savings vary between the three different timescales. Highest savings are observed for the daily timescale. The energy consumption (and savings) across all the three timescales is presented in the form of a bar chart in Figure 4.8.

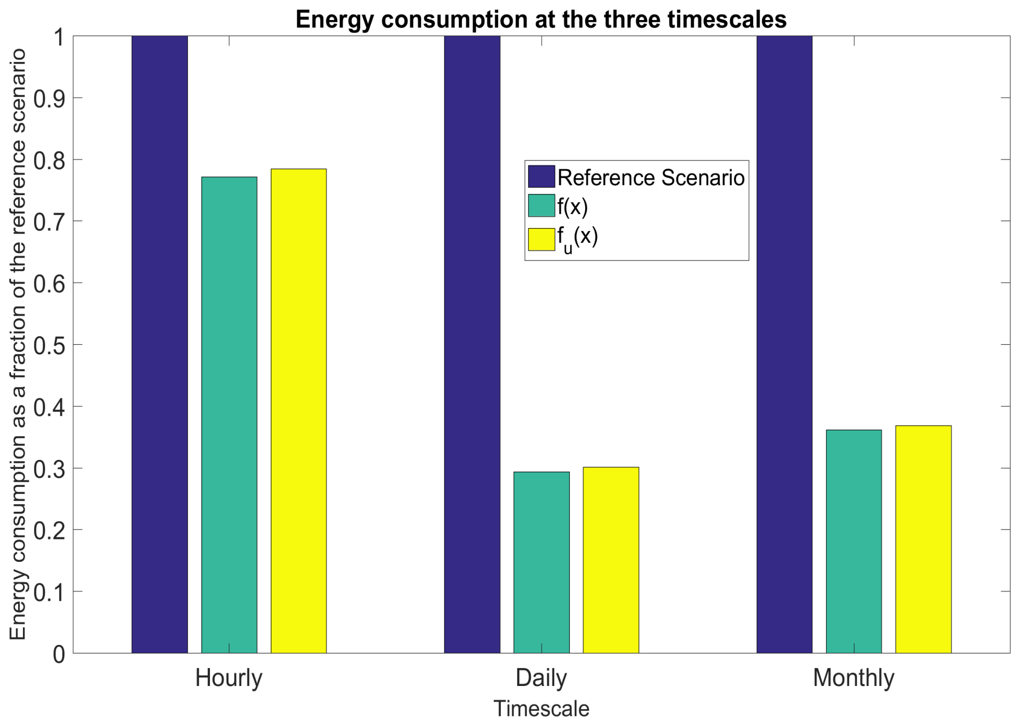


Figure 4.8: Energy consumption across all the three timescales on *turning off carriers*. Shown for $f(x)$ and $f_u(x)$ as a fraction of the reference scenario energy consumption

The lowest savings are observed for the hourly timescale. The reason for this observation is that, for the hourly timescale, the energy consumption is presented for a time-interval of 24 hours whereas the scheme *turning off carriers* was only effectuated for a time-interval of eight hours. For the other two timescales, daily and monthly, the time-interval for which the energy consumption is shown corresponds to the the time-interval when the scheme *turning off carriers* was effectuated.

The savings for the daily timescale are slightly higher than the savings for the monthly timescale. This observation can be partially explained by the fact that for the monthly timescale, the increase in PRB utilization (coverage layer) on turning off the capacity layer was significantly higher than the increase in PRB utilization (coverage layer) for the daily timescale. As a larger increase in PRB utilization (coverage layer) leads to a higher energy consumption and consequently lower savings, the savings for the daily timescale are higher than the savings for the monthly timescale.

The total congestion levels observed across all the three timescales is presented in the form of a bar chart in Figure 4.9. For readability, the vertical axis in the figure is on a log scale. As observed from the figure, there

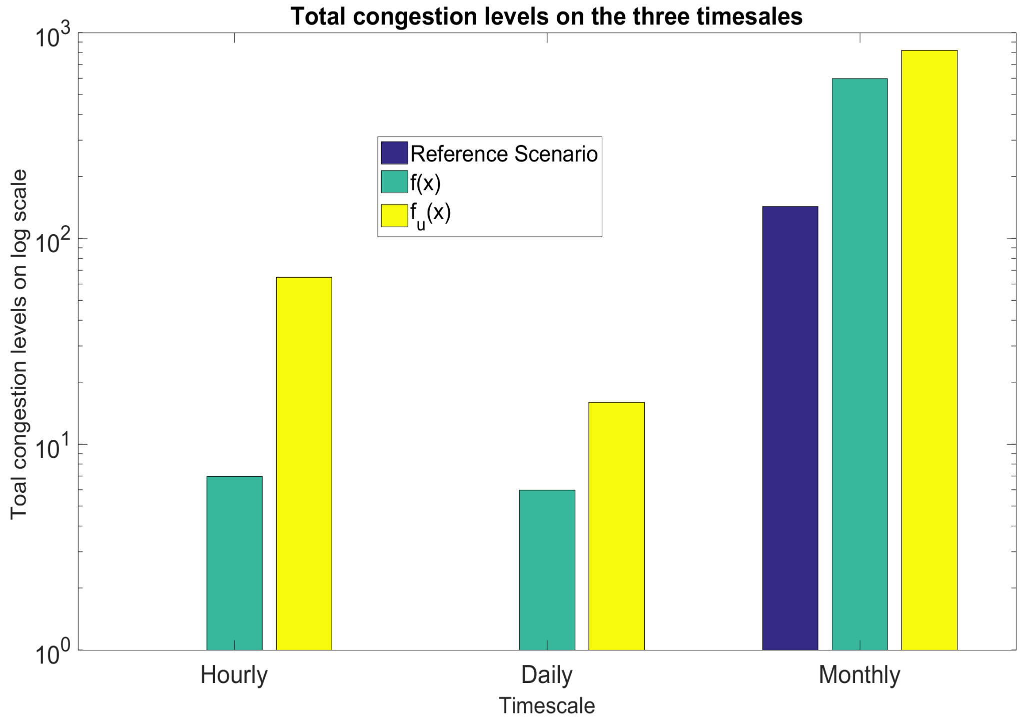


Figure 4.9: Total congestion levels across all the three timescales on *turning off carriers*. Shown for the reference scenario, $f(x)$ and $f_u(x)$

is a significant increase in congestion levels for the monthly timescale. For the reference scenario on the hourly and daily timescale, there was no congestion observed. Insignificant congestion levels are observed on the hourly and daily timescale due to *turning off carriers*. A possible reason for the monthly timescale exhibiting the highest increase in congestion levels could be that the pilot sites selected for the monthly timescale have significantly higher traffic activity (and PRB utilization) than the pilot sites selected for the other two timescales, and this leads to a significant increase in the congestion levels on *turning off carriers* on the monthly timescale.

4.3. Antenna Muting

In this scheme of Antenna Muting, MIMO operation effectively ceases as one of the two transmitting antenna ports (KPN operates 2×2 MIMO) is shut down. This effectively leads to a 1×2 SIMO (Single Input Multiple Output) operation. The energy and operational cost savings in this scheme arise from the reduction in energy consumption as SIMO operation consumes less energy than 2×2 MIMO. The mode of MIMO operation (OLSM or transmit diversity) depends on the rank indicator (R_H) reported by the UE. This mode of transmission is defined as 'Transmission mode 3' by 3GPP [36] in which the transmission mode is adapted based on R_H. As mentioned in Section 2.4, the rank indicator indicates the rank of the channel matrix H, and can take a value of $R_H \leq 2$. The rank reported by the UE depends on the channel correlation.

In LTE, 'transmit diversity' is implemented by using SFBC (Space-Frequency Block Codes)[37] which are an adaptation of the 'Alamouti Scheme' [38] (Space-Time Block Codes). An illustration of SFBC is given below in Table 4.11 using a two antenna and two sub-carriers example, where S_1 denotes the symbol transmitted by antenna 1 on sub-carrier 1 and S_1^* is the complex conjugate of S_1. The rest of the symbols are similarly defined in the table below.
In OLSM, there is no CSI (specifically PMI: Precoding Matrix Indicator) feedback from the UE to the eNodeB. However, the precoding matrices are predefined in the 3GPP standard [22, 36].

	Antenna 1	Antenna 2
Sub-carrier 1	S_1	$-S_2^*$
Sub-Carrier 2	S_2	S_1^*

Table 4.11: SFBC Illustration

The channel capacity for spatial-multiplexing MIMO operation when the transmitter does not have CSI, under uniform power allocation assumption is given in [23] as:

$$C = \sum_{i=1}^{R_H} B \log_2(1 + \frac{\gamma_i}{M_t}) \tag{4.8}$$

where C denotes the channel capacity, B denotes the channel bandwidth, R_H denotes the rank of the H matrix (see Section 2.4), M_t denotes the number of transmitting antennas and $\gamma_i = \frac{\sigma_i^2 P}{\sigma_n^2}$ where P denotes the total transmitted power, σ_i is the power gain of the i^{th}channel and σ_n^2 is the noise power. As R_H is full rank (2) in the scenario under consideration (OLSM mode is operated only when R_H is full rank), from the above equation it can be seen that there would be a two-fold increase in the channel capacity when compared with a SISO operation mode.

A graphical illustration of this scheme is given in Figure 4.10.

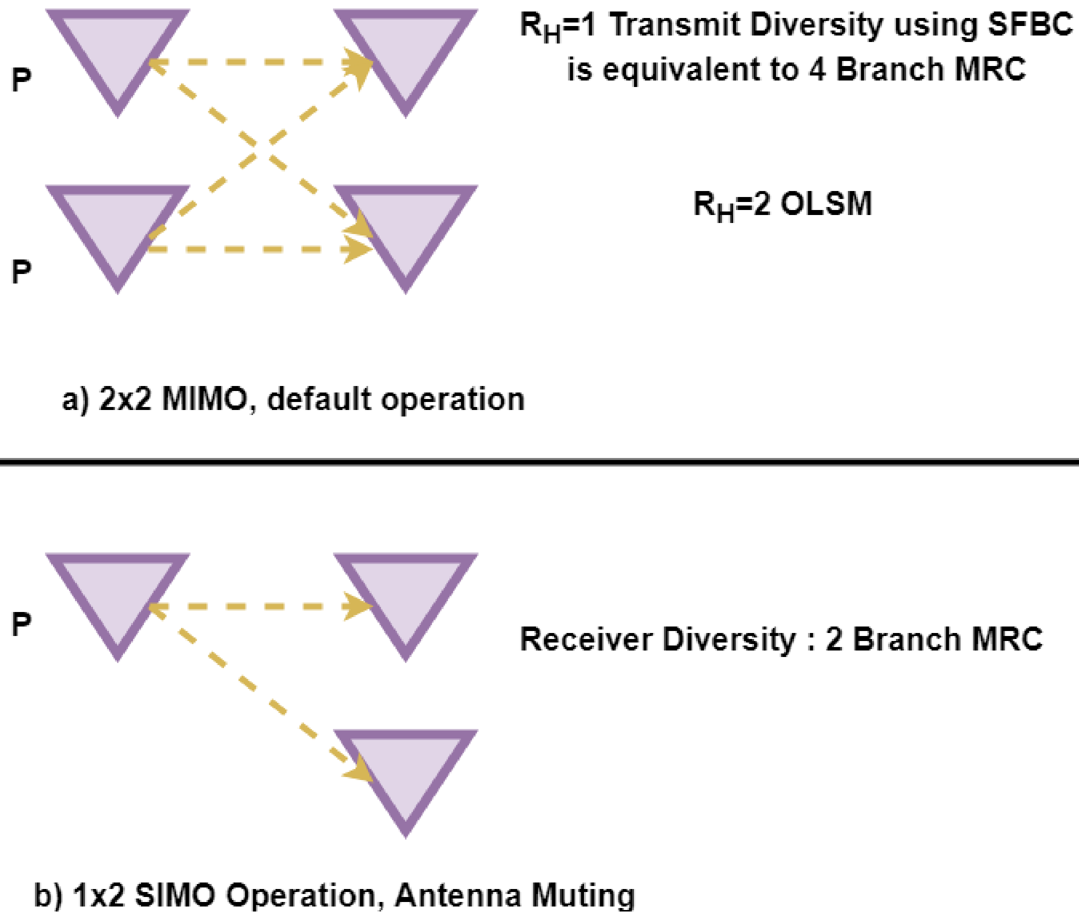


Figure 4.10: a) MIMO, Deafult Operation Mode and b) SIMO, Antenna Muting

Note that the total transmit power is not constant when one of the two transmitting antennas is muted. The total transmit power is $P + P = 2P$ during the default operation mode, and it is P when the scheme *antenna muting* is effectuated. However, the power transmitted per antenna remains the same (P).

The three configurations illustrated in the figure above, when compared with a SISO (Single Input Single Output) reference model with total transmitted power P, lead to the following results:

1. In the default operation mode when R_H=1, transmit diversity mode (using SFBC) results in a 6 dB gain in SNR. As shown in [38], a 2×2 transmit diversity scheme is equivalent to a four branch MRC (Maximal Ratio Combining)(receiver diversity) when the transmitted power per antenna is the same. In MRC, as shown in [23], the SNR at the receiver is the sum of the SNRs of each individual branch i.e. the SNR increases linearly with the number of branches. In the scenario under consideration, we have four branches (2×2 transmit diversity $\equiv fourbranchMRC), thusthereisa10\log_{10}4$=6 dB gain in SNR when compared with the SISO reference model.

2. In the default operation mode when R_H=2, OLSM ideally results in a two-fold increase in throughput (see Equation (4.8)) when compared with the SISO reference model. The increase in throughput may not always be two-fold as the precoding matrix is not based on the real-time channel conditions.

3. In the scheme of *antenna muting*, receiver diversity in the form of a two branch MRC is employed. Proceeding similarly to point 1 above, this results in a $10\log_{10}2$=3 dB gain in SNR when compared with the SISO reference model.

On effectuating the scheme *antenna muting*, from the above three points it can be observed that:

1. If we transform from the default operation mode of 2×2 MIMO when R_H=1 (transmit diversity using SFBC) to the receiver diversity (two branch MRC) SIMO operation mode (Antenna Muting), the SNR reduces by 3 dB (diversity loss, points 1 and 3 above, 6 dB -3 dB= 3 dB).

2. If we transform from the default operation mode of 2×2 MIMO when R_H=2 (OLSM) to the receiver diversity (two branch MRC) SIMO operation mode (Antenna Muting), the throughput reduces by a factor slightly less than two. In the receiver diversity (two branch MRC) mode, the SNR would increase (compared to OLSM) which may improve the throughput slightly. Thus, overall, the throughput reduces by a factor slightly less than two.

The channel capacity (throughput) is related to the SINR by the Shannon capacity equation [39]:

$$C = B\log_2(1 + SINR) \tag{4.9}$$

where C denotes throughput, B denotes the channel bandwidth, and $SINR$ denotes the signal to interference plus noise ratio. From point 1 above, on effectuating the scheme *antenna muting* and when the initial default operation mode is transmit diversity (R_H=1), the reduction in throughout due to the 3 dB reduction in SNR (diversity) depends on what the initial SNR was. In the worst case, the bit rate reduces by a factor of ≈ 2 due to the the 3 dB reduction in SNR. Let C1 denote the initial throughput during the default operation mode R_H=1 (transmit diversity) and C2 denotes the throughput when the scheme *Antenna Muting* is effectuated and the SNR reduces by 3 dB. Figure 4.11 shows the ratio C1/C2 vs SNR for a channel with 10 MHz bandwidth.

As seen from Figure 4.11, at higher SNR values, the throughput reduces by a factor of ≈ 1.1 (C1/C2). At lower SNR values, in the worst case, the throughput reduces by a factor of ≈ 2 (C1/C2).

Assuming the worst case scenario, the throughput for all the UE sessions (at zero load) would reduce by a factor of two on transforming from the default 2×2 MIMO operation to SIMO operation (*antenna muting*). The trade-off of number of hours of congestion that may be experienced is calculated using the PRB utilization-based check. A throughput-based check cannot be carried out as the throughput distribution becomes unknown on effectuating the scheme of *antenna muting*. The reduction in the throughput for each session (at different loads) in the distribution is not known.

Note that on effectuating the scheme *antenna muting*, the PRB utilization would double as halving the throughput would lead to twice as long session-times, and this leads to a doubling of the the use of resources (PRB utilization doubled). A thing to note is that if there are PRBs available in the frequency domain, due to

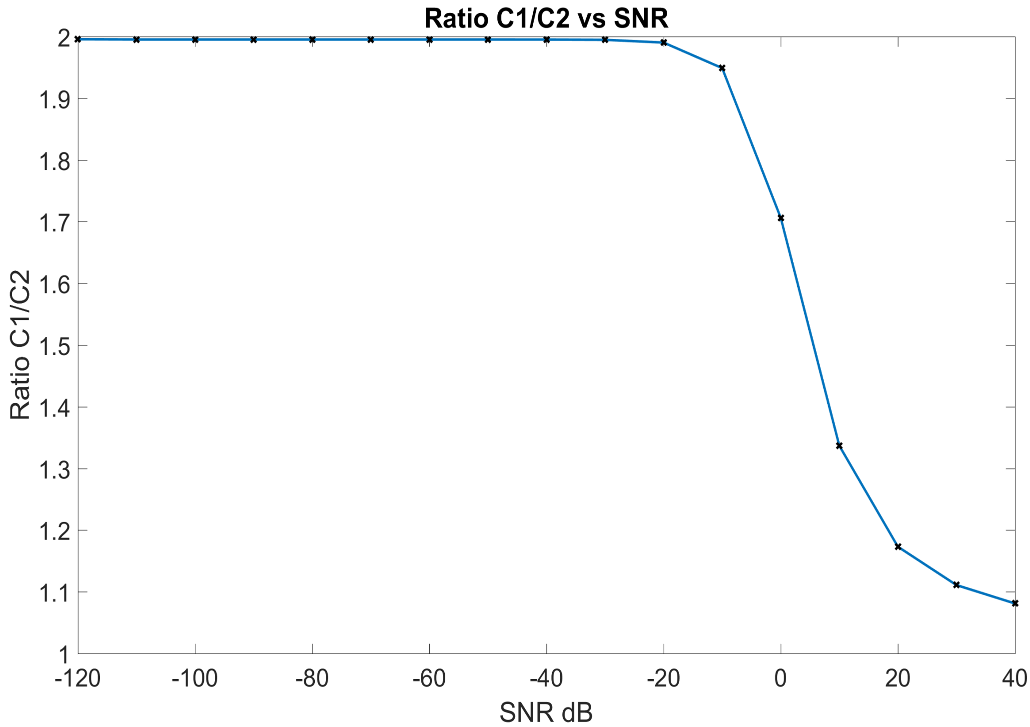


Figure 4.11: Ratio C1/C2 vs SNR (dB), C1 and C2 are calculated using the Shannon equation Equation (4.9) for a channel with 10 MHz bandwidth

small (text-message) UE session and/or small number of sessions, the PRB utilization would double but the throughput(s) would not be halved. The PRBs available in the frequency domain would be assigned to the UEs, and this would lead to the UEs having the same throughput as before at double the PRB utilization. In this scenario, the increase in PRB utilization compensates for the reduction in throughput caused by the scheme effectuation. Note that while the PRB utilization increases (doubles), there is still an overall reduction in energy consumption as the PRB utilization increases for the SIMO mode. There is a significant reduction in energy consumption on muting one antenna port and this is shown in the following section(s). The mean site-carrier PRB utilization which is inserted into the IMEC power model [29] on effectuating the scheme *antenna muting* is shown in Appendix B.

In [33], it is approximated that the average throughput in a radio network decreases linearly with an increase in PRB utilization by the following equation:

$$T_{avg} = T_{max} \times (1 - \rho) \tag{4.10}$$

where T_{avg} denotes the average throughput, T_{max} denotes the average throughput at zero load, ρ denotes the PRB utilization. Assuming that now one of the two antenna ports is muted, the new average throughput at zero load halves (worst-case approximation), and the new PRB utilization is double that of the original PRB utilization as the user session-time doubles. Effectively $T'_{max} = 0.5 \times T_{max}$ and $\rho' = 2 \times \rho$ where T'_{max} is the new average throughput at zero load and ρ' is the new PRB utilization. Then, from Equation (4.10), the new average throughput T'_{avg} is:

$$\begin{aligned} T'_{avg} &= T'_{max}(1 - \rho') \\ T'_{avg} &= 0.5 \times T_{max}(1 - 2 \times \rho) \end{aligned} \tag{4.11}$$

To calculate the reduction in average throughput, Equation (4.11) is subtracted from Equation (4.10) which leads to:

$$T_{avg} - T'_{avg} = 0.5 \times T_{max} \tag{4.12}$$

As seen from Equation (4.12), the reduction in average throughput ($T_{avg} - T'_{avg}$) on muting one antenna port is constant and independent of traffic load. For a given PRB utilization, $\rho = \rho'$, the ratio T_{avg} / T'_{avg} is constant (=two in the example above). As minimum QoS checks are based on the PRB utilization and not the traffic load, we will focus on the scenario that the ratio T_{avg} / T'_{avg} is constant. Figure 4.12 illustrates the constant reduction in throughput (against traffic load, 10 Mbps reduction) and the constant ratio of throughput (against PRB utilization, ratio=two) on reducing the number of antenna ports (two to one). In the figure, example values of average throughput at zero load equal to 20 and 10 Mbps for MIMO and SIMO respectively is used, along with the example value that a 1 Kbps (0.5 Kbps) traffic load equals a PRB utilization of one (100%) for MIMO (SIMO). .

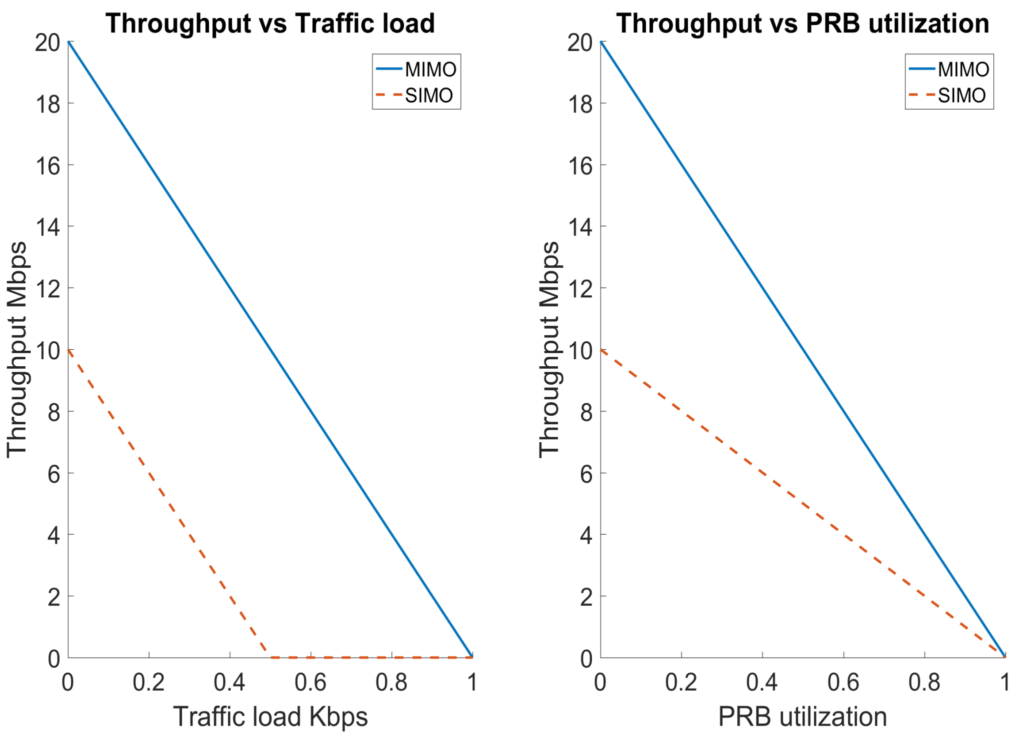


Figure 4.12: Throughput vs Traffic load and Throughput vs PRB utilization, shown to illustrate the constant reduction in throughput and the constant ratio of throughput on effectuating the scheme of antenna muting.

As in the worst-case approximation all users have their throughput halved (at zero load), the entire throughput CDF also shifts accordingly. Therefore, the new [redacted] target on effectuating the scheme *antenna muting* would be [redacted] ([redacted]). We can derive the corresponding (lower) PRB utilization thresholds accordingly based on the results in [27]. The new (lower) PRB utilization threshold (using [redacted] Mbps [redacted] to check whether the minimum QoS conditions are met are:

1. The PRB utilization should remain below [redacted] for a cell with 10 MHz bandwidth.
2. The PRB utilization should remain below [redacted] for a cell with 20 MHz bandwidth.

The tooling to effectuate the scheme *antenna muting* is already available at KPN. The cost to implement is negligible in monetary terms. Time spent by KPN to prepare the tooling for execution by setting N_S and T (see Table 2.1) is the only cost incurred in implementing this scheme. There is no impact on coverage on effectuating this scheme, only the throughput(s) is (are) impacted.

The analysis is carried out for the sites selected for the pilot study again. The procedure to calculate the energy consumption, the operational cost, and the trade-off of number of hours of congestion that may be experienced on effectuating the scheme *antenna muting* is described below. Note that the reference-scenario energy consumption and congestion levels are already known and thus the procedure described below does not cover it.

1. The KPIs of PRB utilization, the time-average number of active users in the downlink, and the throughput distribution are calculated per hour for all the coverage and capacity layer cells of the sites selected for the pilot.

2. The PRB utilization of all the samples (not just filtered) is doubled. The mean-site carrier PRB utilization is calculated for both the coverage and the capacity carrier, and the energy consumption (kWH) is calculated via the IMEC power model [29] and Equation (4.1) assuming SIMO operation (antenna muting).

3. The samples are filtered with the condition the the time-average number of active users in the downlink are [redacted].

4. The PRB utilization (filtered samples) is checked using the new threshold limits defined above (using [redacted] target). Samples which exceed this threshold are said to be in congestion.

5. By comparing the number of hours of congestion experienced, energy consumption and operational costs (Step 4) with the reference scenario value of these metrics presented in Section 4.1.2, the operational cost and energy savings along with the number of hours of congestion that may be experienced due to *antenna muting* is calculated.

Note that the data samples used in the analysis in Section 4.2 are used again for the analysis for the scheme of antenna muting. This is done to enable comparison between the two schemes. Furthermore, the same samples will be using in the analysis of the other scheme(s).

4.3.1. Hourly Timescale

On executing the procedure described in the previous section (Step 4) on the data samples, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.12.

Method	Reference Scenario Initial number of hours of congestion		Number of hours of congestion on scheme effectuation	
	Coverage	Capacity	Coverage	Capacity
PRB utilization-based	0	0	21	7

Table 4.12: The number of hours of congestion that may be experienced on effectuating the scheme antenna muting on the hourly timescale. Calculated using the PRB utilization-based check and the throughput-based check

On executing Step 2 in the procedure described in the previous section, the total energy consumption and operational costs are calculated and shown below in Table 4.13 (for 47 sites for a duration of 8 hours).

	Reference Scenario	Antenna Muting
Energy Consumption Coverage Layer	102.64 kWh	87.31 kWh
Energy Consumption Capacity Layer	342.12 kWh	285.61 kWh
Total Energy Consumption	444.76 kWh	372.92 kWh
Total Operational Cost	[redacted]	[redacted]

Table 4.13: Total energy consumption and operational cost after effectuating the scheme *antenna muting* on the hourly timescale. Shown for 47 sites and for a time period of 24 hours (one day)

On effectuating the scheme of *antenna muting*, the total energy consumption and operational cost are 372.92 kWh and [redacted]. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 71.84 kWh are attainable. Consequently, operational cost savings of [redacted] are also attainable. The energy consumption and operational cost both reduce by 16.15% compared to the reference scenario values. Note that these savings are for 47 sites (pilot sites) and for a duration of 24 hours only (one day).

4.3.2. Daily Timescale

On executing the procedure described in Section 4.3 (Step 4) on the data samples, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.14.

Method	Reference Scenario Initial number of hours of congestion		Number of hours of congestion on scheme effectuation	
	Coverage	Capacity	Coverage	Capacity
PRB utilization-based	0	0	18	4

Table 4.14: The number of hours of congestion that may be experienced on effectuating the scheme antenna muting on the daily timescale. Calculated using the PRB utilization-based check and the throughput-based check

On executing Step 2 in the procedure described in the previous section, the total energy consumption and operational costs are calculated and shown in Table 4.17 (for 49 sites for a duration of 48 hours).

	Reference Scenario	Antenna Muting
Energy Consumption Coverage Layer	203.82 kWh	112.89 kWh
Energy Consumption Capacity Layer	685.37 kWh	343.39 kWh
Total Energy Consumption	889.19 kWh	456.28 kWh
Total Operational Cost	[redacted]	[redacted]

Table 4.15: Total energy consumption and operational cost after effectuating the scheme *antenna muting* on the daily timescale. Shown for 49 sites and for a time period of 48 hours (one weekend)

On effectuating the scheme of *antenna muting*, the total energy consumption and operational cost are 456.28 kWh and [redacted]. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 432.91 kWh are attainable. Consequently, operational cost savings of [redacted] are also attainable. The energy consumption and operational cost both reduce by 48.68% compared to the reference scenario values. Note that these savings are for 49 sites (pilot sites) and for a duration of 48 hours only (one weekend).

4.3.3. Monthly Timescale

On executing the procedure described in Section 4.3 (Step 4) on the data samples, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.16.

On executing Step 2 in the procedure described in the previous section, the total energy consumption and operational costs are calculated and shown in Table 4.17 (for 48 sites for a duration of 168 hours).

On effectuating the scheme of *antenna muting*, the total energy consumption and operational cost are 2062.77 kWh and [redacted]. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 1388.61 kWh are attainable. Consequently, operational cost savings of [redacted] are also

Method	Reference Scenario Initial number of hours of congestion		Number of hours of congestion on scheme effectuation	
	Coverage	Capacity	Coverage	Capacity
PRB utilization-based	121	22	1437	433

Table 4.16: The number of hours of congestion that may be experienced on effectuating the scheme antenna muting on the monthly timescale. Calculated using the PRB utilization-based check and the throughput-based check

	Reference Scenario	Antenna Muting
Energy Consumption Coverage Layer	822.52 kWh	530.61 kWh
Energy Consumption Capacity Layer	2628.86 kWh	1532.16 kWh
Total Energy Consumption	3451.38 kWh	2062.77 kWh
Total Operational Cost	[redacted]	[redacted]

Table 4.17: Total energy consumption and operational cost after effectuating the scheme *antenna muting* on the monthly timescale. Shown for 48 sites and for a time period of 168 hours (one week)

attainable. The energy consumption and operational cost both reduce by 40.23% compared to the reference scenario values. Note that these savings are for 48 sites (pilot sites) and for a duration of 168 hours only (one weekend).

4.3.4. Conclusion

The scheme of *antenna muting* was described and analyzed in the previous subsections. As seen from the results presented, effectuating this scheme can lead to significant energy and operational cost savings. The percentage savings vary between the three different timescales. Once more, the highest savings are observed for the daily timescale. The energy consumption (and savings) across all the three timescales is presented in the form of a bar chart in Figure 4.13.

The lowest energy savings are attained for the hourly timescale for the reason mentioned in Section 4.2.4. The savings in the energy consumption for the daily timescale is higher than the savings in the energy consumption for the monthly timescale for the reason mentioned in Section 4.2.4. The total congestion levels observed on *antenna muting* are shown in the form of a bar chart in Figure 4.14. The observations from the figure are similar to the observations made in Section 4.2.4.

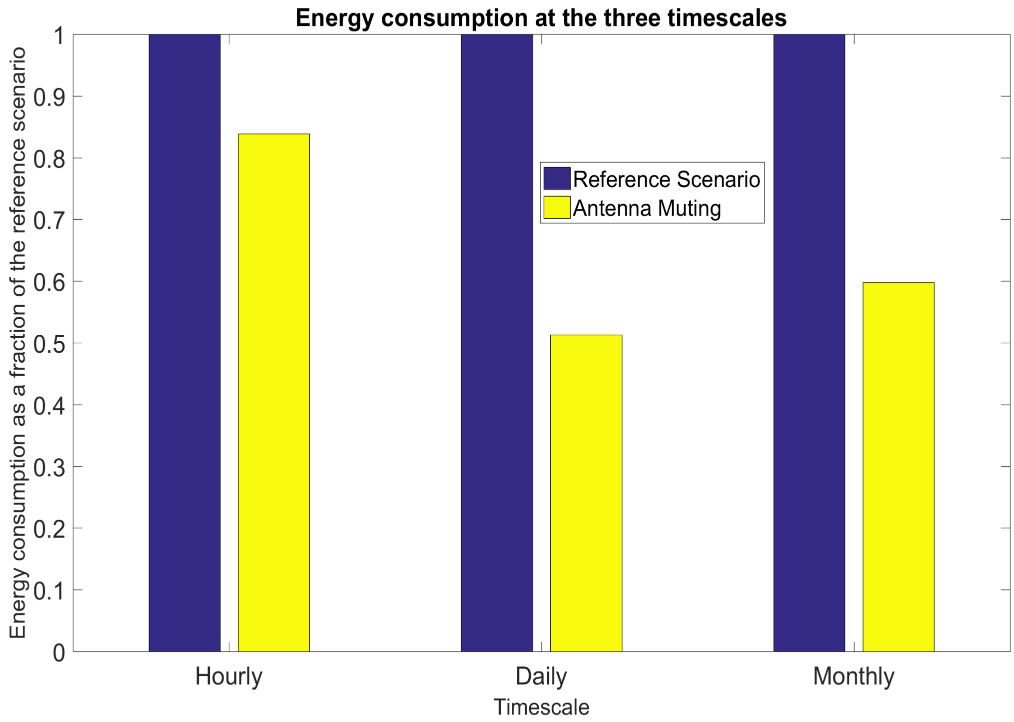


Figure 4.13: Energy consumption across all the three timescales on effectuating the scheme *antenna muting*. Shown for *antenna muting* as a fraction of the reference scenario energy consumption

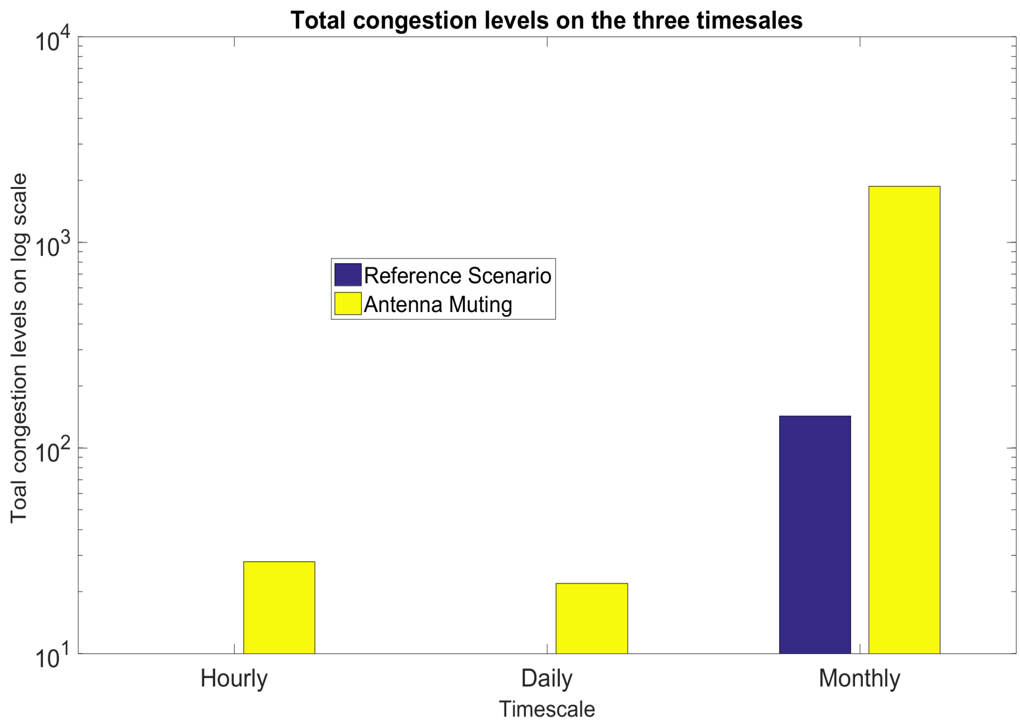


Figure 4.14: Total congestion levels across all the three timescales on effectuating the scheme *antenna muting*. Shown for the reference scenario and *antenna muting*

4.4. Discontinuous Transmission (DTX)

In this scheme of *discontinuous transmission* (DTX), the LTE frame structure is modified to transmit fewer symbols and this leads to a reduction in energy consumption. This scheme is described in detail in [17] with the assumption of near instantaneous on-off switching of resources.

The structure of a LTE radio frame (for the downlink) was described in Section 2.3 and it was seen that one radio frame consists of ten sub-frames. There is some compulsory signalling at the start of each sub-frame (control region) and some compulsory reference and synchronization signals have to be transmitted at certain sub-frames (sub-frame 0 and 5 for synchronization signals) in the six center PRBs. LTE also supports a different type of sub-frame, viz. the MBSFN (Multicast-Broadcast Single-Frequency Network) sub-frame. A maximum of six out of the ten sub-frames in a radio frame can be assigned to MBSFN. The MBSFN sub-frame also has a control region at the start but the rest of it may be set empty ([17]) i.e. there is no transmission of the compulsory reference and synchronization signals. A frame with six MBSFN sub-frames assigned is shown below in Figure 4.15.

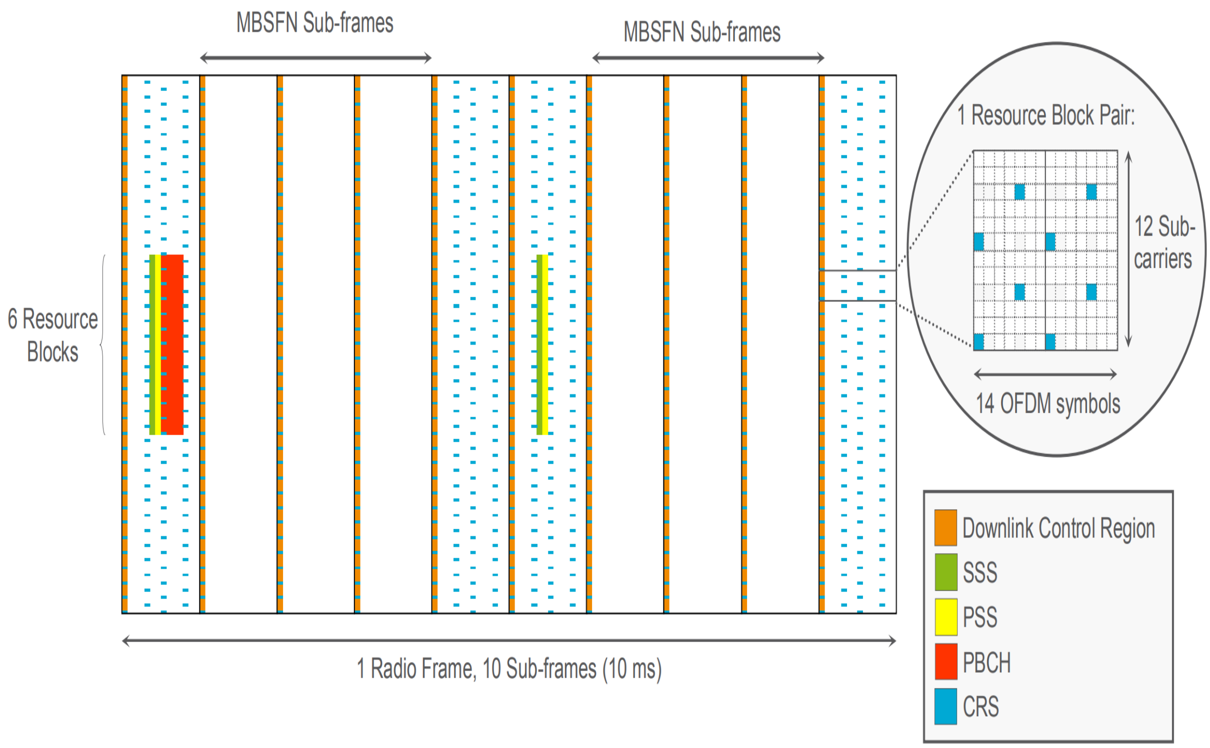


Figure 4.15: Radio frame of LTE with six MBSFN sub-frames assigned. Source:(Pål et al., figure 2) [17]

This scheme consists of modifying the LTE frame structure to include MBSFN sub-frames and setting the MBSFN sub-frames empty. This scheme is essentially a trick as we are not doing any multi-broadcasts, and we are just using the MBSFN feature to keep the sub-frame(s) empty. Each sub-frame consists of a fixed number of PRBs depending on the bandwidth (50 for 10 MHz and 100 for 20 MHz). Effectively, this scheme reduces the total number of PRBs available in a frame. The reduction in PRBs depends on the number of sub-frames assigned to MBSFN, more sub-frames assigned to MBSFN will lead to a higher reduction.

While the resources (in this case PRBs) are in actuality reduced in the time domain (MBSFN sub-frames are assigned at fixed positions in the time domain), the analysis is carried out from the viewpoint that the resources are reduced in the frequency domain (number of PRBs reduced across the whole frame) for simplicity. This is equivalent as the total number of PRBs in a frame remains the same whether the resources are reduced in the frequency domain or time domain (a UE session may consist of hundreds of frames). A dissimilarity between the time-domain implementation and frequency domain implementation is that a UE may experience a higher latency during the setup procedure in the time domain implementation (some sub-frames would be empty), and this would not occur in the frequency domain implementation. Another dissimilarity is that if the UE session-duration is only a few sub-frames long (shorter than a 10 ms frame), the UE throughput in the frequency-domain implementation would be smaller than the time-domain implementation as the frequency-domain implementation has fewer resources available throughout the frame length. However, both these dissimilarities are extremely minor and have negligible effect(s) on the analysis as session-duration is generally hundreds of frames long, and an added latency of 1 ms is insignificant.

To illustrate that the number of PRBs remains equal in both the time-domain and the frequency-domain implementation, consider the following example. E.g. assume that for a bandwidth of 20 MHz, five out of ten sub-frames (50% reduction) in a frame are assigned to MBSFN. Effectively, in the time domain, this leads to a reduction in the total number of PRBs in a frame from 100×10=1000 (100 PRBs in each sub-frame originally) to 100×5=500 (100 PRBs only for 5 sub-frames, other 5 sub-frames are set empty). In the frequency domain, assume that the bandwidth is reduced from 20 MHz to 10 MHz (50% reduction), this leads to a reduction in total number of PRBs in a frame from 100×10=1000 to 50×10=500 (50 PRBs in a sub-frame for 10 MHz bandwidth). As seen, the total number of PRBs in a frame is equal for both the time domain reduction and frequency domain implementation.

For a given PRB utilization when no sub-frame(s) is (are) assigned to MBSFN, on assigning some of the sub-frames to MBSFN, the PRB utilization for the sub-frames not assigned to MBSFN increases (effectively the PRB utilization increases as fewer resources available). The PRB utilization for the sub-frames assigned to MBSFN is zero as the MBSFN sub-frame is set empty in this scheme. On average, this leads to a lower energy consumption. For e.g. assume that five sub-frames out of ten have been assigned to MBSFN, then 50% of the sub-frames have a higher energy consumption due to higher PRB utilization (PRB utilization doubles as number of available PRBs halves) while the other 50% of the sub-frames (MBSFN) have a lower energy consumption (sleep mode energy consumption). Note that for the sub-frames assigned to MBSFN, the energy consumption is calculated using the sleep-mode feature in the IMEC power model, as for these sub-frames there is a partial switching-off of the eNodeB components which leads to a much lower sleep-mode energy consumption. The sleep-mode has a latency of 71 μs ([17, 29]).

The feature MBSFN and the tooling to effectuate this scheme is not available at KPN. The license fees to buy the MBSFN feature and as well as the tooling costs will have to be incurred by KPN to effectuate this scheme. As this scheme may also be effectuated remotely, the time required to activate this scheme is six minutes to account for the fact that the system information (frame structure), on average, can be changed only once every six minutes [17].

Note that a throughput-based check for congestion levels cannot be carried out in this scheme as only the impact on average throughput and the [redacted] on effectuating the scheme is known. The impact on the throughput of individual UE sessions is unknown. Therefore, only a PRB utilization-based check for congestion levels is carried out using the PRB utilization threshold limits which are derived below. As explained in Section 4.1.2, the PRB utilization-based check is more reliable than the throughput-based check.

The congestion level(s) and the attainable energy savings on effectuating *discontinuous transmission* are bounded by the options of assigning one and six sub-frames in a radio frame to MBSFN. This is intuitively logical as reducing more resources (assigning more sub-frames to MBSFN) would lead to higher congestion levels (higher energy savings). Thus, the highest congestion levels (energy savings) would be observed when assigning six sub-frames to MBSFN, and lowest congestion levels (energy savings) would be observed when assigning only one sub-frame to MBSFN. Therefore, the detailed analysis in the following sections is only shown for these two options.

The PRB utilization increases on effectuating *discontinuous transmission*, as overall, fewer PRBs are available at the same traffic load. E.g. assume that initially there are 100×10=1000 PRBs per frame, out of which 300 PRBs are utilized. The initial PRB utilization is then 30%. On assigning two sub-frames to MBSFN, overall, there will be 800 PRBs available. The new PRB utilization is then 37.5% (30×1.25=37.5). The factor by which the PRB utilization increases is shown in table 4.18 for the two options which are analyzed in the following sections, assigning one and six sub-frames to MBSFN. Note that the final PRB utilization in the table is cut-off at 100%. The mean site-carrier PRB utilization which is inserted into the IMEC power model [29] on effectuating the scheme *discontinuous transmission* is shown in Appendix B.

Number of sub-frames assigned to MBSFN	Initial PRB utilization	Final PRB utilization
$n = 1$	ρ_i	$1.11 \times \rho_i$
$n = 6$	ρ_i	$2.5 \times \rho_i$

Table 4.18: Factor by which PRB Utilization increases for $n = 1, 6$

To carry out PRB utilization-based checks for minimum QoS conditions, new PRB utilization thresholds need to be derived for the six different cases (six sub-frames may be assigned to MBSFN). For $n = \{1, 2, 3, \ldots, 6\}$ where n denotes the number of sub-frames assigned to MBSFN in a frame, the analysis will be carried out by reducing the bandwidth (frequency domain) by a factor of $f = \{10\%, 20\%, \ldots, 60\%\}$ where f is the percentage reduction in bandwidth. On assigning $n = \{1, 2, 3, 4, 6\}$ sub-frames to MBSFN, in frequency domain, the new bandwidth (assuming original bandwidth of 20 MHz) would be $b = \{18, 16, 14, 12, 8\}$ MHz. As these band- widths are not permissible in the LTE system, there are no (existing) results (for throughput versus PRB utilization) available for these bandwidths. However, for the case of $n = 5$ ($b = 10$ MHz), results are available as a bandwidth of 10 MHz is permissible in a LTE system. From [27], the [redacted] reduces from [redacted] (at zero load, constant ratio at a given PRB utilization) on halving the bandwidth (20 MHz to 10 MHz). The procedure is similar to the procedure presented in Section 4.3 (see Figure 4.12 and Equation (4.10), instead of MIMO vs SIMO there would be 20 MHz vs 10 MHz bandwidth). The remaining scenarios of $n = \{1, 2, 3, 4, 6\}$, are assumed to be uniformly distributed along the scale of [redacted]. For e.g. assuming 20 MHz bandwidth, as for n=5 the [redacted] is reduced by a factor of [redacted], for $n = \{1, 2, 3, 4, 6\}$ it is assumed that the [redacted] is reduced from [redacted] respectively, reduction by a factor of [redacted]. The same result is also assumed to hold for a cell with 10 MHz bandwidth as the new bandwidth (on assigning sub-frames to MBSFN) in this case is also not permissible in the LTE system. Note that the detailed analysis is only shown for $n = 1, 6$.

As the shift in [redacted] CDF is known on halving the bandwidth (20 MHz to 10 MHz), we can derive the (lower) PRB utilization threshold accordingly based on the results in [27]. For e.g. for the case of halving the bandwidth from 20 MHz to 10 MHz, the [redacted] reduces by a factor of [redacted] (from [redacted] Mbps to [redacted] at zero load, constant ratio). The new [redacted] target would be [redacted] and the corresponding PRB utilization threshold is [redacted] (from [27]). Similarly, using the different factors mentioned above for $n = 1, 6$, the PRB utilization thresholds are derived from [27] and are shown in Table 4.19. The reference case when no sub-frames are assigned to MBSFN ($n = 0$) is also shown in the table.

[redacted]

Table 4.19: PRB Utilization threshold(s) $n = 0, 1, 6$

From Table 4.19, it is observed that as n increases, the PRB utilization threshold decreases. This is intuitively logical as on increasing n, fewer resources are available, and thus the minimum QoS target will be reached earlier (at lower PRB utilization). The procedure to calculate the energy consumption, the operational cost, and the trade-off of number of hours of congestion experienced on effectuating *discontinuous transmission*is similar to the previous sections:

1. The KPIs of PRB utilization and the time-average number of active users in the downlink are calculated per hour for all the coverage and capacity layer cells of the sites selected for the pilot.

2. The PRB utilization is increased based on the value of n (see Table 4.18).

3. The samples are filtered with the condition that the time-average number of active users in the downlink ([redacted]).

4. The number of hours of congestion experienced is calculated from the PRB utilization threshold limits defined in Table 4.19. If a sample exceeds the threshold, then that sample is said to be in congestion.

5. The energy consumption is calculated by inserting the new mean site-carrier PRB utilization into the power model for a fraction of the time (non MBSFN sub-frames) and for the other fraction of time (MBSFN sub0frames) the sleep-mode energy consumption is assumed. The overall energy consumption is calculated by taking a weighted mean (0.9,0.1 and 0.4,0.6 as weights for $n = 1, 6$ respectively) of the fractions calculated above. E.g. for $n = 1$, there are 90% non-MBSFN sub-frames and 10% MBSFN sub-frames. The overall energy consumption is then, $0.9\times$(non-MBSFN energy consumption)+$0.1\times$(sleep mode energy consumption).

6. By comparing the number of hours of congestion experienced, energy consumption and operational costs (Step 5) with the reference scenario value of these metrics presented in Section 4.1.2, the operational cost and energy savings along with the number of hours of congestion that may be experienced due to *discontinuous transmission* is calculated.

4.4.1. Hourly Timescale

On executing the procedure described in the previous section (Steps 1-4) on the data samples, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.20.

Number of sub-frames assigned to MBSFN	Reference Scenario Initial number of hours of congestion		Number of hours of congestion on scheme effectuation	
	Coverage	Capacity	Coverage	Capacity
$n = 1$	0	0	0	0
$n = 6$	0	0	28	9

Table 4.20: The number of hours of congestion that may be experienced on effectuating the scheme discontinuous transmission on the hourly timescale. Calculated using the PRB utilization-based check

On executing Step 5 in the procedure described in the previous section, the total energy consumption and operational costs are calculated and shown below in Table 4.21 (for 47 sites for a duration of 8 hours).

On effectuating the scheme of *discontinuous transmission*, the total energy consumption and operational cost are 439.28 kWh and [redacted] for $n = 1$, and 412.46 kWh and [redacted] for $n = 6$. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 5.48 to 32.3 kWh are attainable. Consequently, operational cost savings of [redacted] are also attainable. The energy consumption and operational cost both reduce by 1.23% to 7.26% compared to the reference scenario values. Note that these savings are for 47 sites (pilot sites) and for a duration of 24 hours only (one day).

4.4.2. Daily Timescale

On executing the procedure described in the previous section (Steps 1-4) on the data samples, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.22.
On executing Step 5 in the procedure described in the previous section, the total energy consumption and operational costs are calculated and shown below in Table 4.23 (for 49 sites for a duration of 48 hours).

	Reference Scenario	*n=1*	*n=6*
Energy Consumption Coverage Layer	102.64 kWh	101.39 kWh	95.72 kWh
Energy Consumption Capacity Layer	342.12 kWh	337.89 kWh	316.74 kWh
Total Energy Consumption	444.76 kWh	439.28 kWh	412.46 kWh
Total Operational Cost	[redacted]	[redacted]	[redacted]

Table 4.21: Total energy consumption and operational cost after effectuating the scheme discontinuous transmission on the hourly timescale. Shown for 47 sites and for a time period of 24 hours (one day)

Number of sub-frames assigned to MBSFN	Reference Scenario Initial number of hours of congestion		Number of hours of congestion on scheme effectuation	
	Coverage	Capacity	Coverage	Capacity
$n = 1$	0	0	0	0
$n = 6$	0	0	43	8

Table 4.22: The number of hours of congestion that may be experienced on effectuating the scheme discontinuous transmission on the daily timescale. Calculated using the PRB utilization-based check

	Reference Scenario	*n=1*	*n=6*
Energy Consumption Coverage Layer	203.82 kWh	196.74 kWh	166.52 kWh
Energy Consumption Capacity Layer	685.37 kWh	659.03 kWh	547.07 kWh
Total Energy Consumption	889.19 kWh	855.77 kWh	713.59 kWh
Total Operational Cost	[redacted]	[redacted]	[redacted]

Table 4.23: Total energy consumption and operational cost after effectuating the scheme discontinuous transmission on the daily timescale. Shown for 49 sites and for a time period of 48 hours (one weekend)

On effectuating the scheme of *discontinuous transmission*, the total energy consumption and operational cost are 855.77 kWh and [redacted] for $n = 1$, and 713.59 kWh and [redacted] for $n = 6$. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 33.42 to 175.6 kWh are attainable. Consequently, operational cost savings of [redacted] are also attainable. The energy consumption and operational cost both reduce by 3.75% to 19.74% compared to the reference scenario values. Note that these savings are for 49 sites (pilot sites) and for a duration of 48 hours only (one weekend).

4.4.3. Monthly Timescale

On executing the procedure described in the previous section (Steps 1-4) on the data samples, the trade-off of number of hours of congestion that may be experienced on effectuating the scheme was obtained and is presented in Table 4.24.

On executing Step 5 in the procedure described in the previous section, the total energy consumption and

Number of sub-frames assigned to MBSFN	Reference Scenario Initial number of hours of congestion		Number of hours of congestion on scheme effectuation	
	Coverage	Capacity	Coverage	Capacity
$n = 1$	121	22	211	59
$n = 6$	121	22	1982	682

Table 4.24: The number of hours of congestion that may be experienced on effectuating the scheme discontinuous transmission on the monthly timescale. Calculated using the PRB utilization-based check

operational costs are calculated and shown below in Table 4.25 (for 48 sites for a duration of 168 hours).

	Reference Scenario	*n=1*	*n=6*
Energy Consumption Coverage Layer	822.52 kWh	803.01 kWh	710.27 kWh
Energy Consumption Capacity Layer	2628.86 kWh	2558.55 kWh	2290.18 kWh
Total Energy Consumption	3451.38 kWh	3361.56 kWh	3000.45 kWh
Total Operational Cost	[redacted]	[redacted]	[redacted]

Table 4.25: Total energy consumption and operational cost after effectuating the scheme discontinuous transmission on the monthly timescale. Shown for 48 sites and for a time period of 168 hours (one week)

On effectuating the scheme of *discontinuous transmission*, the total energy consumption and operational cost are 3361.56 kWh and [redacted] for $n = 1$, and 3000.45 kWh and [redacted] for $n = 6$. On comparing the energy consumption on effectuating the scheme with the reference scenario, energy savings of 89.82 to 450.93 kWh are attainable. Consequently, operational cost savings of [redacted] are also attainable. The energy consumption and operational cost both reduce by 2.60% to 13.06% compared to the reference scenario values. Note that these savings are for 48 sites (pilot sites) and for a duration of 168 hours only (one week).

4.4.4. Conclusion

The scheme of *discontinuous transmission* was described and analyzed in the previous subsections. As seen from the results presented, effectuating this scheme leads to little energy and operational cost savings (compared to the other schemes presented in the previous sections). The percentage savings vary little between the three different timescales. Once more, the highest savings are observed for the daily timescale for the reason mentioned in Section 4.2.4. The energy consumption (and savings) across all the three timescales is presented in the form of a bar chart in Figure 4.16.

The total congestion levels observed in *discontinuous transmission* are shown in the form of a bar chart in Figure 4.17. For readability, the vertical axis in the figure is on a log scale. The observations from the figure are similar to the observations made in Section 4.2.4.

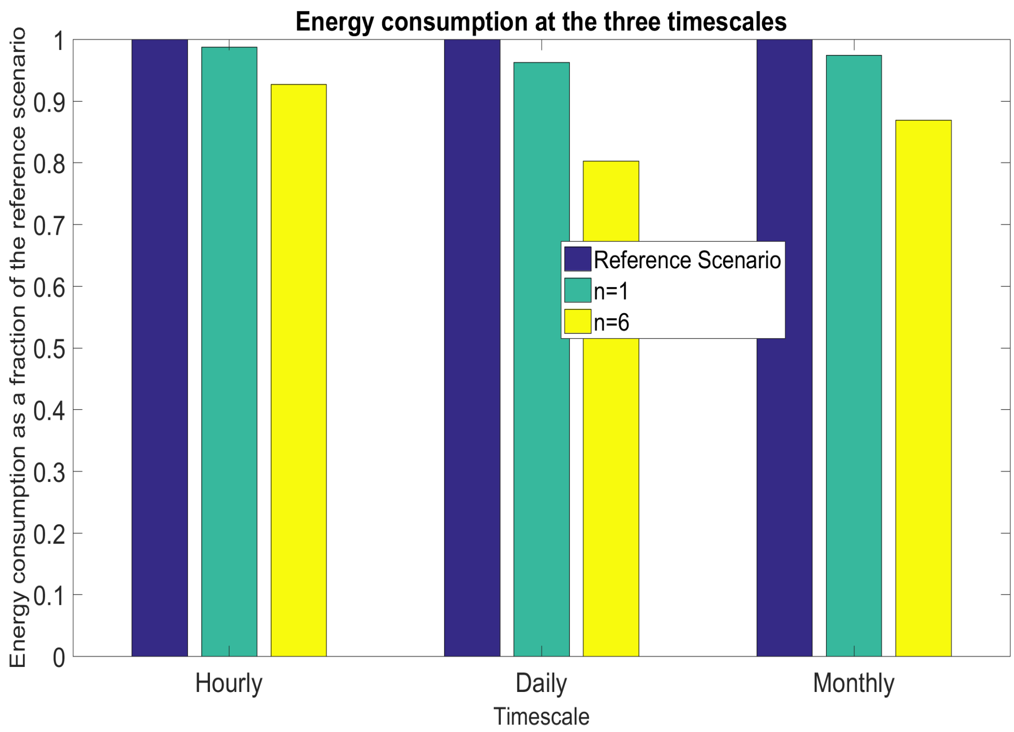


Figure 4.16: Energy consumption across all the three timescales on effectuating the scheme *discontinuous transmission*. Shown for $n = 1, 6$ as a fraction of the reference scenario energy consumption

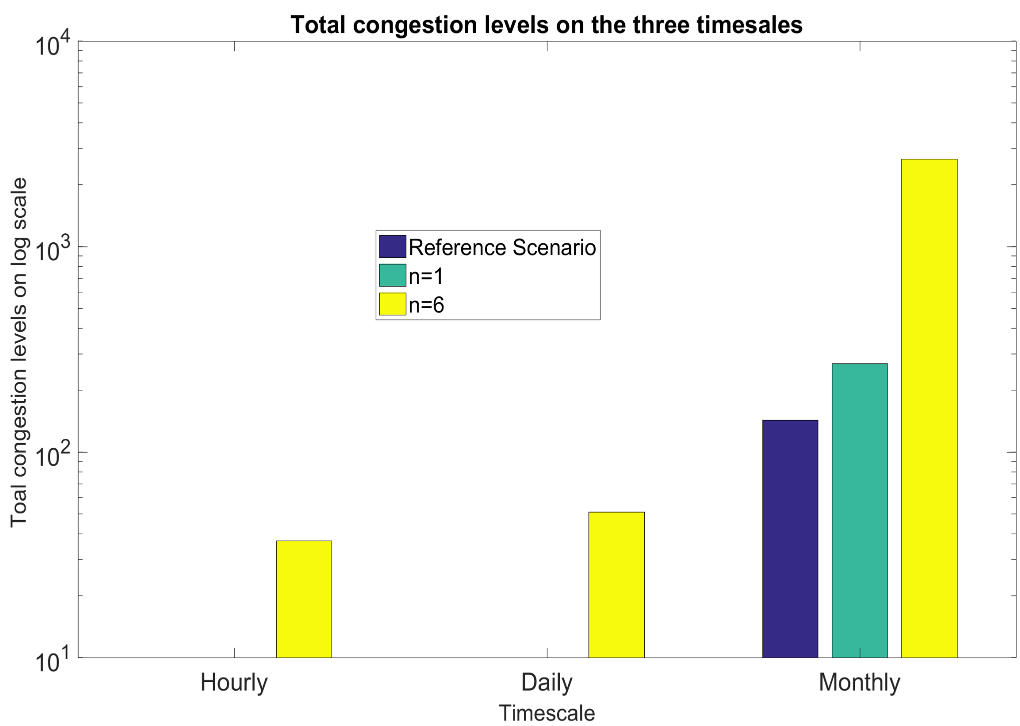


Figure 4.17: Total congestion levels across all the three timescales on effectuating the scheme *discontinuous transmission*. Shown for the reference scenario and $n = 1, 6$.

4.5. Reducing license capacity

In this section the scheme of *reducing license capacity* is discussed. A license is an artificial resource present in the form of a software limitation. There are various different types of licenses deployed in a radio base station and some of the features they limit are the transmission power of the base station, the number of connected users, inter-radio access technology handover capability and downlink capacity (Mbps cap).

This scheme is based on the license of downlink capacity. This license is applied at the site level and not the cell level. It limits the downlink capacity (across all layers) at any instant (per TTI) to the licensed capacity which can be either [redacted] Mbps. Presently, there is some non-uniformity in the licensed capacity across different sites. In the future, it is expected that sites with only the coverage layer will have a licensed capacity of [redacted] Mbps, sites with coverage and capacity layer will have a licensed capacity of [redacted] Mbps, sites with coverage and two capacity layers (1800, 2100 MHz carriers) will have a licensed capacity of [redacted] Mbps and sites with coverage and three capacity layers (1800, 2100 and 2600 MHz carriers) will have a licensed capacity of [redacted] Mbps.

The licensed capacity indirectly affects the downlink user throughput(s). Exceeding the licensed capacity leads to throttling of the user throughput(s) until the licensed capacity is not exceeded. The license capacity, by default, is set to a high value due to the benchmark(s) testing and the (indirect) impact of licensed capacity on the peak user throughput.

The license usage is available from the counter pmLicDlCapDistr. This counter shows the license usage in the form of a histogram (with bins of 0-5%, 5-10%..95-100%) with a sampling rate of 10 ms (every frame as a frame is 10 ms long), i.e. every 10 ms a sample is added to the histogram which represents the license usage for that time-interval. For license usage below 100%, the downlink user throughput is not limited (throttled) due to the license, as there is still spare license capacity available. As the licensed capacity threshold is enforced per TTI (ten TTIs in a frame), it may be possible that some of the user throughput(s) are throttled in some TTI(s) even when the license usage is below 100% (license usage has a sampling rate of 10 ms). However, nothing concrete can be said about this as the network-counter data does not provide enough information relating to this.

As the license usage is available in the form of a histogram and license capacity limits user throughput(s) once the licensed capacity is exceeded, only user session(s) that are active during the time-interval(s) (10 ms) when the license usage is 95-100% (last bin in the license usage histogram) may be limited. In this scheme, the licensed capacity may arbitrarily reduced and a different level of congestion may be experienced. The licensed capacity can be arbitrarily reduced as the licensed capacity can be set to [redacted] (Mbps). Quantifying the cost savings due to this strategy is challenging as licenses are bought in bulk after protracted negotiations with the equipment vendor, and the existing licenses cannot be exchanged for a new lower limit license due to the contract conditions.

KPN has to pay a fee for the license but is allowed to reassign the licenses among the different sites to different values. The licensed capacity can be changed remotely, however it is unclear how long it takes for this scheme to effectuate changes. It is unknown whether a new license software key has to be sourced from the vendor and entered into the system. The cost of the license is [redacted] per Mbps per site. It is shown below (Figure 4.18) that more than 95% of the time, only 0-10% of license capacity is utilized. The findings presented here may motivate KPN to negotiate a change from the fixed cost model to a per usage basis cost model (like energy which is charged per kWh). Unlike the other schemes, energy has no part in this scheme, since no actual hardware is (partially) switched off. Note that in Figure 4.18, none of the sites have a licensed capacity of [redacted].

For this scheme, the congestion level(s) on reducing the licensed capacity cannot be calculated for the following reasons:

1. The licensed capacity is defined at the site level and consequently the license usage is only available at the site level, whereas the QoS check(s) is (are) carried out at the cell level. Thus, from the network-counter data, it is not possible to map license usage to a particular cell at the site.
2. On reducing the licensed capacity, how the 'reduced license capacity' impacts the user throughput(s)

[redacted]

Figure 4.18: PDF of license usage in Mbps (log scale) for the different licensed capacities on the hourly timescale. Obtained by aggregating the license usage data for all sites for all hours under consideration and filtered by the different licensed capacities.

at the cell level is not known. Consequently, the impact on PRB utilization is also not known as the potentially longer session-duration(s) due to the lower throughput(s) will impact the PRB utilization.

3. For the PRB utilization-based check for congestion level(s), the PRB utilization threshold on reducing the licensed capacity is not known.

For the above mentioned reasons, the drawback on *reducing license capacity* cannot be quantified. However, from Figure 4.18, it can be observed that there is a potential for operational cost savings as only a fraction of the licensed capacity is utilized for a majority of the time.

4.6. Ericsson's 'Psi-coverage' solution

The Psi Coverage Solution offered by Ericsson will be briefly discussed in this section. As mentioned, a site in a cellular network generally has three sectors/cells. Each of these sector has its own radio unit and antenna system, hence effectively there are three radio units and antenna system present at each site (associated with each layer). In the PSI coverage solution offered by Ericsson, only one radio unit is used instead of three. The single radio unit connects to the three antenna system by the means of a splitter [16]. At low traffic volumes, it is advertised that this solution provides the same coverage, throughputs similar to a regular system and a reduced energy consumption. A graphical representation of this solution is shown in Figure 4.19 below.
This solution is targeted towards rural area sites which are coverage-oriented rather than capacity-oriented. Rural sites, generally, always have low-traffic volumes. However, in this section its feasibility towards operational cost reduction is discussed. As the solution is close-sourced, not enough information is available about it. The energy savings possible on effectuating this scheme is assumed to be 240 W as advertised. Based on the limited literature available about this solution, it is hypothesized that in the Psi Coverage Solution, the capacity of the system is reduced via hardware changes (different radio units).

As this solution involves the use of a different type of radio units, investments will have to be made by KPN to purchase these radio units. Furthermore, as this solution involves a hardware change, this solution cannot be effectuated at the hourly and daily timescale. It takes one to two days to send engineers to the site, change the radio unit layout and run the standard diagnostics to check whether the site is functioning as intended.

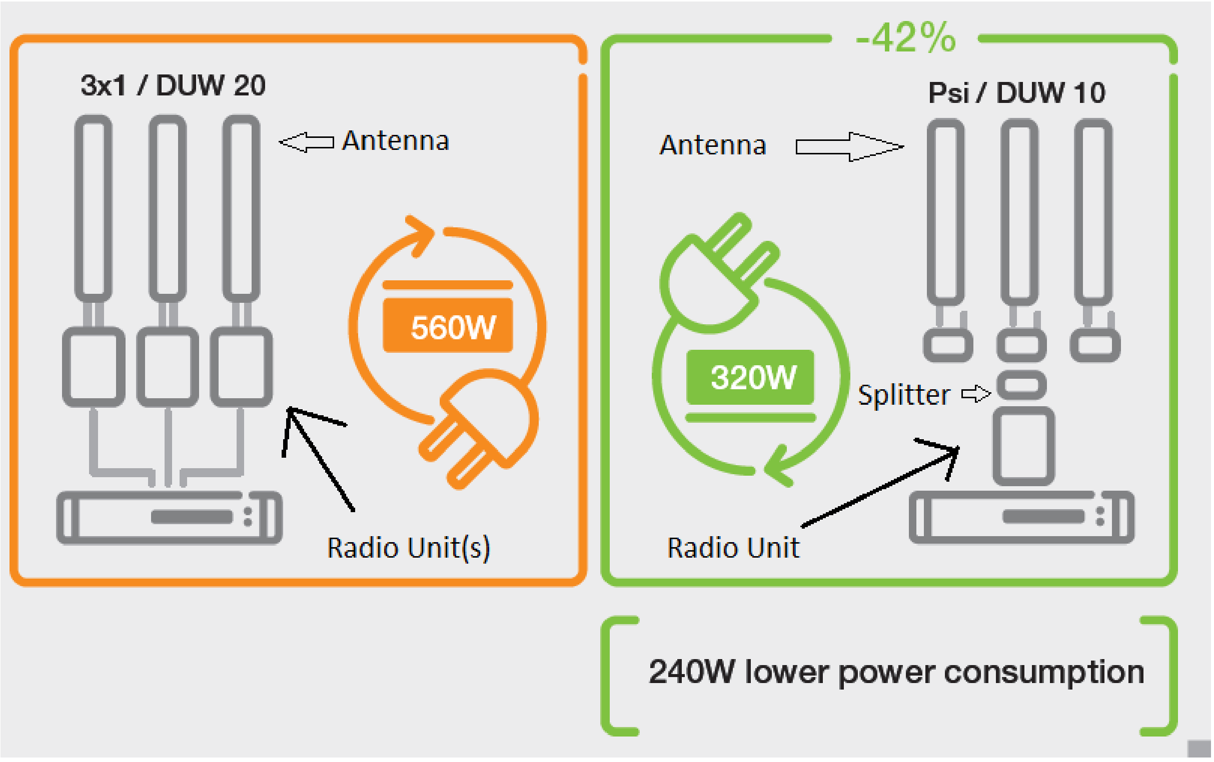


Figure 4.19: A traditional base station system on the left and the Ericsson Psi Coverage Solution on the right. Source:(Pål et al., figure 7) [16]

For KPN, just the cost of sending engineers to a site is [redacted]. This cost is complemented with the costs of the new radio units required for this solution. As advertised, this solution reduces energy consumption by 240 W. Assuming this solution is effectuated at a single site for a period of ten months, the savings in operational cost for KPN would be [redacted] ([redacted]). The savings in operational cost are overshadowed by the operational and investment costs required to effectuate this solution ([redacted]). Therefore, in a dynamic approach, this solution is not viable at all to reduce operational costs. Unless the savings increase substantially in the future (due to increase in energy costs), this scheme will never be a viable option in a dynamic approach. The operator will lose a substantial amount of operating capital on effectuating this scheme with no gains in operational cost. As this solution is significantly cost prohibitive, the congestion levels on effectuating this solution are not analyzed.

This solution provides significant gains when setting up new (rural) sites which are coverage-oriented. The hardware costs will be significantly less (compared to a traditional solution) as only one radio unit is used instead of three, and the energy consumption is reduced as well.

5 Conclusion and Future Work

5.1. Conclusion

In this thesis the spatio-temporal traffic variation in a mobile network and various operational cost reduction schemes were analyzed. The spatio-temporal traffic variation was analyzed for the three different timescales of hourly, daily and monthly. For these three timescales, (pilot) sites which carried significantly less traffic during the night, at weekends and non-summer periods were selected for the analysis of the operational cost reduction schemes. As these sites carried less traffic and hence needed less resources (e.g. antenna, PRBs) during fixed time-intervals, resources could be switched-off during the fixed time-intervals to reduce the energy consumption and consequently the operational cost.

For the different timescales, it was observed that a significant number of sites exhibit 'significant potential' for energy savings. The term 'significant potential' here means that, at these sites, the resources can be switched-off and/or reduced while still satisfying the minimum QoS targets for a majority of the time. At the hourly timescale, 99% of the sites exhibit 'significant potential'. At the daily timescale, overall, 14% of the sites exhibit 'significant potential' (combining weekday/weekend sites). At the monthly timescale, 11% of the sites exhibit 'significant potential'. Note that the term 'significant potential' reveals nothing about the energy and operational cost saving potential of the site. Sites where turning off the resources would lead to the highest savings in energy and operational cost, are generally the sites where there would be highest congestion levels experienced on turning off the resources. The term 'significant potential' only provides the information that turning off and/or reducing the resources for a fixed time-interval would not lead to congestion most of the time.

The five different operational cost reduction schemes analyzed in this thesis are:

1. Turning off carriers:A site in a mobile network can have more than one frequency carrier present. If multiple carriers are present, then some of the carriers may be turned off during periods of low traffic volume. This leads to a reduction in energy consumption as fewer hardware components are active. The tooling required to effectuate this scheme is available at KPN.

2. Antenna Muting: LTE employs MIMO (Multiple Input Multiple Output) to improve system performance, to increase data-rates or to improve link reliability. In this scheme, one of the two antenna ports (KPN employs 2×2 MIMO) is muted (switched-off). This leads to a reduction in energy consumption as fewer hardware components are active. The tooling required to effectuate this scheme is available at KPN.

3. Discontinuous Transmission (DTX): The frame structure for LTE has provisions for allocating some of the sub-frames to MBSFN (Multicast-broadcast Single Frequency Network). A maximum of six sub-frames out of ten may be assigned to MBSFN. The sub-frames assigned to MBSFN do not have to transmit compulsory reference signals. Assigning some of the sub-frames to MBSFN leads to a reduction in energy consumption as the hardware remains inactive for a longer duration (sleep-mode). The tooling required to effectuate this scheme is not available at KPN.

4. Reducing License Capacity: A license is an artificial resource in the form of a software limitation. It puts a limit on the capacity of the system. Reducing the license capacity leads to a reduction in operational costs as a license with a lower capacity limit costs less than a license with a higher capacity limit. Energy consumption has no role in this scheme as no hardware is (partially) switched off. The tooling required to effectuate this scheme is available at KPN.

5. Ericsson's so-called 'Psi Coverage' solution: In this scheme, the site layout is transformed from a traditional three radio unit structure to a single radio unit structure. Transforming the site layout to a single radio unit structure leads to a reduction in energy consumption as fewer hardware components are active. The tooling required to effectuate this scheme is not available at KPN.

As seen from the results presented in Chapter 3, there is ample opportunity present for KPN to reduce their operational cost and/or energy consumption. While there could be more opportunities in the future to reduce operational costs via faster on-off switching of resource(s) mechanisms and quicker traffic volume reporting times, in the current live-network, resources may be switched off and/or reduced for a set of sites during every night, every week/weekend or all round the year except the months of July and August. The analysis of the operational cost reduction schemes in Chapter 4 was carried out to discover the savings in operational cost (€) and/or energy consumption (kWh). Pilot-sites were selected for each timescale to analyze the operational cost reduction schemes. The drawback of effectuating the operational cost reduction schemes was presented in the form of number of hours of congestion (congestion level) that may be experienced. The drawback of the number of hours of congestion was calculated based on a PRB utilization-based check and the threshold limits for PRB utilization were obtained from [27].

The decision to whether effectuate the above mentioned operational cost reductions schemes rests with the KPN management. The assessment of the operational cost reduction schemes is planned to be continuous, i.e the methodology developed in this thesis is to be applied continuously every day, week or month to adapt the energy savings and congestion level estimates. An adaptive mechanism is needed as the (constant) growth in traffic and changes in user behavior need to be accounted for, as these factors impact the energy savings and congestion level estimates directly.

For each timescale, to quantify the savings in operational cost and/or energy consumption, the operational cost and/or energy consumption on effectuating the operational cost reduction schemes was compared with the reference scenario operational cost and/or energy consumption. The reference scenario operational cost and/or energy consumption was calculated for N_S number of sites (pilot sites) and T number of hours (collectively).

Timescale	N_S	T
Hourly	47	24
Daily	49	48
Monthly	48	168

Table 5.1: Value of N_S and T across all the three timescales

The above mentioned values of N_S and T were also used to calculate the operational cost and/or energy consumption on effectuating the operational cost reduction schemes (collectively). Based on the operational cost and/or energy consumption savings results presented in Chapter 4, across all the three timescales, the highest operational costs and energy consumption savings are attained on *turning off carriers.* Compared to the reference scenarios presented in Section 4.1.2, energy and operational cost savings of 20-70% are attainable on *turning off carriers.* The exact savings (percentage) attainable depends on the timescale, and on the site(s) where the carriers are turned off.

The second highest savings are attained on effectuating the operational cost reduction scheme of *antenna muting.* Compared to the reference scenarios, energy and operational cost savings of 16-49% are attainable on effectuating *antenna muting.*

Effectuating the scheme of *discontinuous transmission* would lead to energy and operational cost savings which are significantly less than the savings attainable for *turning off carriers* and *antenna muting*. Compared to the reference scenarios,energy and operational cost savings of 1-20% are attainable on effectuating *discontinuous transmission*.

While there is a potential for significant operational cost savings on effectuating the scheme *reducing license capacity*, it is challenging to quantify the savings as the license costs are established after negotiations between the operator (KPN) and the vendor (Ericsson). Furthermore, the congestion level(s) that may be experienced on effectuating the scheme *reducing license capacity* cannot be quantified for the reasons mentioned in Section 4.5.

The 'Psi Coverage' solution offered by Ericsson does lead to a reduction in energy consumption, however, it is significantly cost-prohibitive in a dynamic approach. KPN would need to make significant investments and incur personnel costs to effectuate this scheme, and these costs completely overshadow the savings attained on effectuating this scheme. This scheme may provide significant gains when setting up new coverage-oriented sites as it requires fewer hardware components. The congestion level(s) on effectuating this scheme were not analyzed due to the cost-prohibitive nature of this scheme.

The highest energy and operational cost savings are attained on the daily timescale for all the operational cost reduction schemes, and the minimum energy and operational cost savings are attained on the hourly timescale. The observation that the energy and operational cost savings are minimum for the hourly timescale can be explained by the fact that for the hourly timescale, the operational cost reductions schemes may only be effectuated for a time-interval of eight hours, but the savings in energy consumption and operational cost are presented with respect to a 24 hour time-interval.

The congestion levels that may be experienced on effectuating the various operational cost reduction schemes are also presented in Chapter 4. Comparing the congestion levels on effectuating the various operational cost reduction schemes with the reference scenarios, it is observed that comparable levels of congestion are obtained for the operational cost reduction scheme of *antenna muting* and *discontinuous transmission*. The lowest congestion level(s) are observed for the operational cost reduction scheme of *turning off carriers*.

The energy consumption for the reference scenario and all the operational cost reduction schemes (as a fraction of the reference scenario) is shown in Figure 5.1. Note that the most optimistic options (highest energy savings) are selected and shown for the scheme *turning off carriers* and *discontinuous transmission*. The highest energy saving for *discontinuous transmission* occurs when six sub-frames are assigned to MBSFN, and for *turning off carriers* it occurs using the most optimistic curve (curve $f(x)$, see Chapter 4).

The total congestion levels observed on effectuating all the operational cost reduction schemes are shown in Figure 5.2. For readability, the vertical axis in the figure is on a log scale.

From Figure 5.1, it is clearly observed that the highest energy and operational cost savings are attained on effectuating scheme of *turning off carriers* for all the three timescales. From Figure 5.2, it is observed that effectuating the operational cost reduction scheme of *turning off carriers* leads to the lowest total congestion levels.

Therefore, based on the results presented in Figures 5.1 and 5.2, it is recommended that KPN effectuates the scheme of *turning off carriers* in response to a change in traffic volume, as it leads to the highest energy and operational cost savings and the lowest congestion levels.

Effectuating the operational cost reduction scheme of *turning off carriers* leading to highest energy and operational cost savings is intuitively logical. The capacity carrier, at zero load, consumes almost three times as much energy as the coverage carrier. Thus, on turning off the capacity carrier, there is a huge reduction in the overall energy consumption as the energy consumption of the capacity carrier reduces to the shut-down mode energy consumption.

The observation that the lowest congestion levels are experienced on effectuating the scheme *turning off*

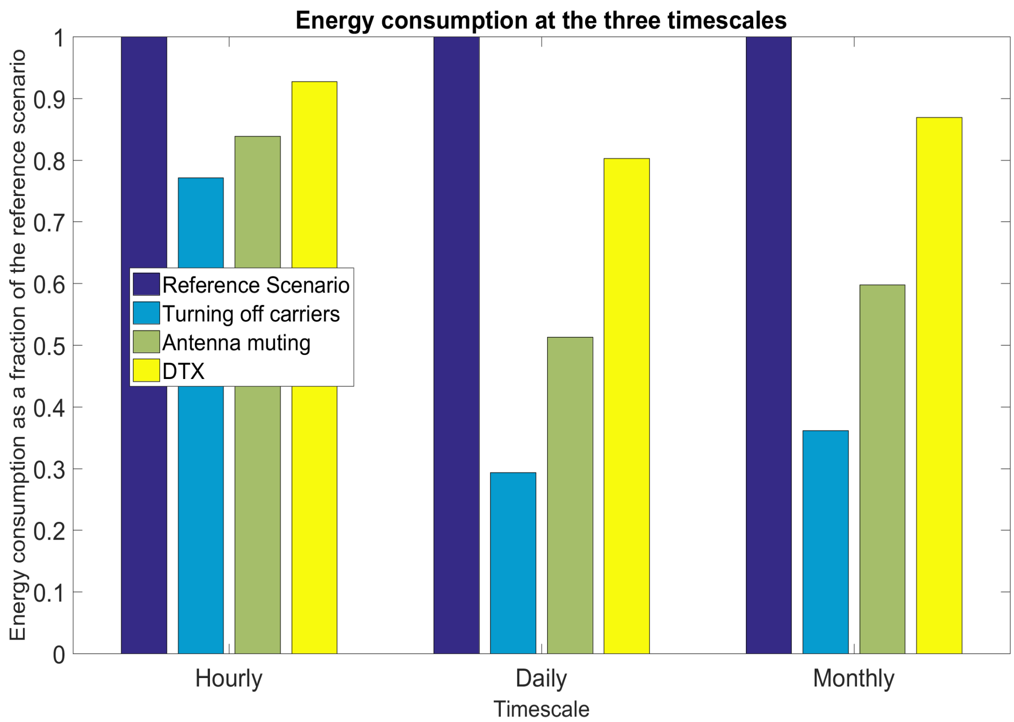


Figure 5.1: Energy consumption for all the operational cost reduction schemes. Shown for all the three timescales using the most optimistic option(s) of $n = 6$ and curve $f(x)$.

carriers can be partially explained by the fact that on *turning off the capacity carrier*, there are fewer cells overall and thus lower congestion levels. E.g. assume that there are three coverage and three capacity cells, and all the cells exceed the (corresponding) PRB utilization threshold and are experiencing congestion. On turning off the three capacity cells, congestion can be experienced only in the remaining three coverage cells. Whereas for the other operational cost reduction schemes, as they do not involve a complete shut-down of the cell(s), all the six cells would experience congestion.

The highest energy and operational cost savings are attainable on the daily timescale. Thus, it is also recommended to focus on the daily timescale sites and to effectuate the operational cost reduction scheme of *turning of carriers* at these sites.

Note that the operational cost reduction scheme of *turning off carriers* may only be applied at sites where multiple carriers are present. Presently, approximately [redacted] of the sites in the KPN network have multiple carriers present and this number is increasing. While this thesis focused on the 1800 MHz capacity carrier, KPN may also focus on the 2600 MHz carrier to reduce energy consumption. However, KPN may need to seek regulatory approval to turn off the 2600 MHz carrier(s) as these sites are mainly operated to meet the spectrum license conditions.

On comparing the operational cost reduction schemes of *antenna muting* and *discontinuous transmission*, it is clearly observed that *antenna muting* leads to higher savings at similar congestion levels. Thus, it is recommended to effectuate the operational cost reduction scheme of *antenna muting* instead of *discontinuous transmission* at sites which do not have multiple carriers.

The (absolute) savings in energy consumption (kWh) on effectuating the operational cost reduction schemes at the three timescales is shown below in Table 5.2
The (absolute) savings in operational cost (€) on effectuating the operational cost reduction schemes at the three timescales is shown below in Table 5.3

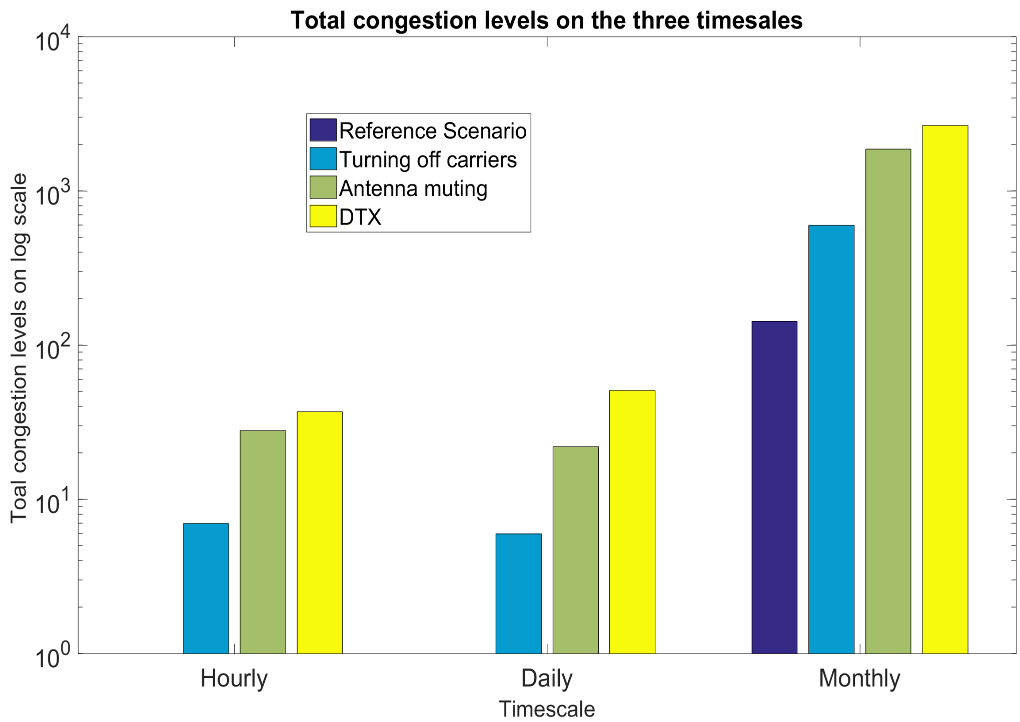


Figure 5.2: Total congestion levels observed for all the operational cost reduction schemes. Shown for all the three timescales using the most optimistic option(s) of $n = 6$ and curve $f(x)$.

Scheme \ $Timescale$	Hourly (kWh)	Daily (kWh)	Monthly (kWh)
Turning off carriers	95.91 to 101.62	621.54 to 628.36	2178.89 to 2203.08
Antenna muting	71.84	432.91	1388.61
Discontinuous transmission	5.48 to 32.3	33.42 to 175.6	89.82 to 450.93

Table 5.2: Absolute magnitude of energy savings (kWh). Shown for the operational cost reduction schemes at all the three timescales.

The savings shown in Tables 5.2 and 5.3 above are for a small set of pilot sites (N_S). A rough estimate of the savings for a larger set of sites (e.g 99% of the sites at the hourly timescale) can be found by appropriately scaling the results in the table above. The reason for which the results in Table 5.2 can be scaled for the energy savings estimate is that more than 90% of the time, the PRB utilization for the fixed time-intervals when the operational cost reduction schemes may be effectuated is comparable (0-10% PRB utilization) for all the sites in the network. Thus, a rough estimate may be found out by scaling the results in Table 5.2. However, congestion levels cannot be scaled and detailed calculations need to be carried out to establish the congestion levels on effectuating the schemes for a larger set of sites. Congestion levels cannot be scaled as congestion is defined based on the time-average number of active users in the downlink.

The results presented in Chapter 4 may also be compared against the license fees of the energy saving features offered by the equipment vendor (Ericsson) and a decision can be made by the KPN management on whether to purchase the features and implement them in the live-network.

[redacted]

Table 5.3: Absolute magnitude of operational cost savings (€). Shown for the operational cost reduction schemes at all the three timescales.

5.2. Future Work

The possible directions for the future research are as follows:

1. This research was based on the LTE Network. The schemes discussed in this thesis will have to be analyzed again for the 5G and the ultra- dense networks that are envisaged for the future.
2. An assessment of the energy savings attainable on combining the different operational cost reduction schemes.
3. Presently, the congestion analysis is based on a heuristic approach to some extent. The approach for congestion analysis may be improved in future research.
4. Quantifying the savings and congestion for the operational cost reduction scheme of *reducing license capacity*.
5. An investigation into a pilot of the analyzed operational cost reduction schemes in the live-network.
6. The development of near-instantaneous mechanisms to on-off switch resources in response to a traffic variation as the variation in traffic cannot be completely predicted, the operators can only react to the traffic fluctuations and cannot be proactive.
7. Effectuating some of the schemes discussed in this thesis will lead to gains in the performance on the neighboring cell(s) for e.g. turning off the carrier will lead to reduced interference for the neighboring cell. These gains have not been investigated.
8. This research can be carried out for the other operators in the region (T-Mobile, Vodafone) and can provide an overview for the energy savings attainable in The Netherlands as a whole.

A
Appendix

The IMEC power model [29] is briefly covered in this appendix. As explained in Section 4.1.1, there are two different 'reference points' in the KPN LTE network.

1. For the 800 MHz carrier: [redacted], defined at the output of the radio unit.
2. For the 1800 MHz carrier: [redacted], defined at the output of the tower mounted amplifier (TMA).

As already mentioned in the thesis (Section 4.1.1), the TMA provides no gain in the downlink. For the calculation of the energy consumption (kWh) for the reference scenarios and the energy saving schemes, various parameters and default settings were inserted into the IMEC power model. The parameter values that were inserted for the energy consumption (kWh) are presented below. The 800 MHz carrier and the 1800 MHz carrier column(s) represent the parameters that were input for the reference scenario energy consumption. A denotes the energy saving scheme of *turning off carriers*, B denotes the energy saving scheme of *antenna muting* and C denotes the energy saving scheme of *discontinuous transmission*. The '-' in the following table represents no change from the reference scenario parameters. For A, B and C, unless explicitly mentioned in brackets, the parameter value presented in the table is inserted for both the 800 MHz and the 1800 MHz carrier.

[redacted]

Table A.1: Parameter values inserted into the IMEC power model [29]

Note that the 'Output power per PA' is set to [redacted] (per antenna) for the 1800 MHz carrier whereas the reference point is [redacted] (per antenna). The power model does not account for the TMA. As the reference point is defined at the output of the TMA, the feeder loss (3 dB) is added to the reference point and inserted into the power model (see Figure 4.1). The TMA provides no gain in the downlink. For the parameters that are not explicitly mentioned in the table above, default values that are already inserted for them are used to calculate the energy consumption. The output of the power model is a graph where the vertical axis is the energy consumption and the horizontal axis is the PRB utilization. The output of the power model is shown below in Figure A.1 for an example configuration (not actual KPN configuration).

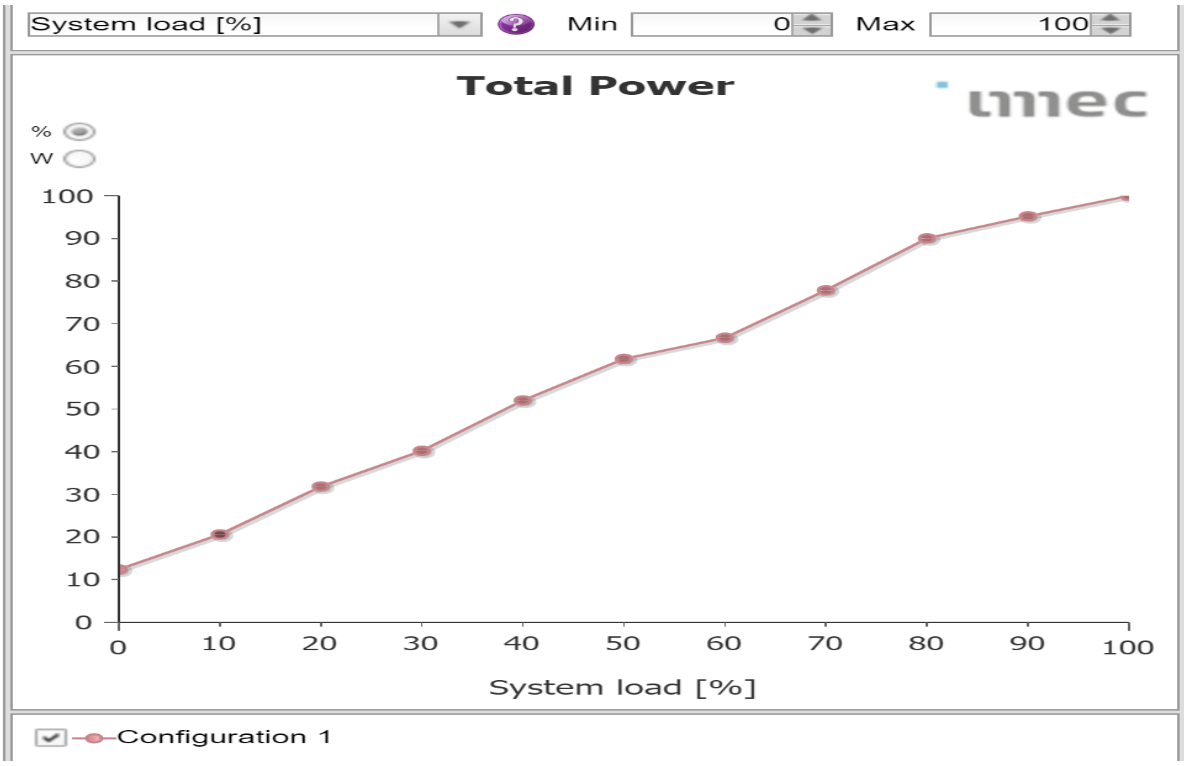


Figure A.1: Energy consumption vs PRB utilization for an example configuration. Source: IMEC [29]

An experiment was conducted in the live-network to check whether the magnitude of energy saving reported by the IMEC power model is comparable. In the experiment, the energy saving scheme *antenna muting* was effectuated for the capacity carrier at a site. As the energy consumption data available for the live-network includes all the components (2G, 3G, LTE, data-logging components) present at a site, only the overall change in energy consumption can be tracked. The site selected for the experiment is one of the pilot sites used for the assessment of the energy saving schemes in this thesis. The site generally carried very low traffic during the night. The scheme *antenna muting* was effectuated for eight hours in a night. For comparison, the energy consumption during the night when the scheme *antenna muting* was effectuated was compared to the energy consumption the night before and the night after when the scheme was not effectuated. A reasonable assumption that the overall traffic characteristics at the site during the three nights did not vary much is made to enable comparison of the energy consumption data. The energy consumption in the live-network for these three nights is shown in Figure A.2 below. The reduction in energy consumption for the capacity carrier on *antenna muting* is reported by the IMEC power model to be [redacted] (at a PRB utilization of [redacted]). In the live-network, a reduction of [redacted] is observed on *antenna muting* for the capacity carrier. Thus, the IMEC power model underestimates the savings attained in the live-network. The experiment was repeated for a different site and the results were similar. The energy savings reported by the IMEC power model is of a similar magnitude and thus the IMEC power model is validated.

Figure A.2: Energy consumption in the live-network for one site

B
Appendix

The mean site-carrier PRB utilization for the reference scenarios and on effectuating the operational cost reduction schemes is presented in this appendix. The mean site-carrier PRB utilization metric is inserted into the IMEC power model [29] to calculate the energy consumption (kWh). Note that for the hourly timescale, all the PRB utilization metrics presented are for a time-interval of eight hours. However, the energy consumption at the hourly timescale is presented for a time-interval of 24 hours in this thesis. For the reference scenario of the hourly timescale (24 hours), the PRB utilization inserted into the IMEC power model was [redacted] for the coverage carrier and [redacted] for the capacity carrier. The initial PRB utilization for the coverage and the capacity carrier (before switching off and/or reducing resources) is shown in Table B.1.

[redacted]

Table B.1: Initial mean site-carrier PRB utilization for the coverage and the capacity carrier for all the three timescales

The mean site-carrier PRB utilization on effectuating the scheme *turning off carriers* is shown in Table B.2.

[redacted]

Table B.2: The mean site-carrier PRB utilization for the coverage carrier on effectuating the scheme of *turning off carriers*

The mean site-carrier PRB utilization on effectuating the scheme of *antenna muting* is shown in Table B.3.

The mean site-carrier PRB utilization on effectuating the scheme of *discontinuous transmission* is shown in Table B.4

[redacted]

Table B.3: The mean site-carrier PRB utilization for the coverage and the capacity carrier on effectuating the scheme of *antenna muting*

[redacted]

Table B.4: The mean site-carrier PRB utilization for the coverage and the capacity carrier on effectuating the scheme of *discontinuous transmission*

Bibliography

[1] Cisco Visual Networking Index. Global mobile data traffic forecast update, 2016-2021, 2017.

[2] Nokia Siemens Networks. Corporate responsibility report, 2009.

[3] Huawei Whitepaper. Improving energy efficiency, Lower CO2 emission and TCO, 2009.

[4] Fengli Xu, Yuyun Lin, Jiaxin Huang, Di Wu, Hongzhi Shi, Jeungeun Song, and Yong Li. Big data driven mobile traffic understanding and forecasting: A time series approach. *IEEE transactions on services computing*, 9(5):796–805, 2016.

[5] Guangshuo Chen, Sahar Hoteit, Aline Carneiro Viana, Marco Fiore, and Carlos Sarraute. Spatio-temporal predictability of cellular data traffic. Technical report, RT-0483 INRIA Saclay-Ile-de-France, 2017.

[6] EU EARTH. Energy Aware Radio and neTwork tecHnologies. *"D6.4: Final Integrated Concept"*, 2012.

[7] Dieter Ferling, Thomas Bitzer, Thomas Bohn, Dirk Wiegner, and Andreas Pascht. Power efficient transceivers to enable energy-efficient mobile radio systems. *Bell Labs Technical Journal*, 15(2):59–76, 2010.

[8] EU EARTH. Energy Aware Radio and neTwork tecHnologies. *"D4.2: Green Radio Technologies"*, 2012.

[9] Yinan Qi, Muhammad Ali Imran, and Rahim Tafazolli. Energy-aware adaptive sectorisation in LTE systems. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pages 2402–2406. IEEE, 2011.

[10] Hauke Holtkamp. Enhancing the energy efficiency of radio base stations. *arXiv preprint arXiv:1311.3534*, 2013.

[11] Remco Litjens, Yohan Toh, Haibin Zhang, and Oliver Blume. Assessment of the energy efficiency enhancement of future mobile networks. In *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pages 2336–2341. IEEE, 2014.

[12] Zhisheng Niu, Yiqun Wu, Jie Gong, and Zexi Yang. Cell zooming for cost-efficient green cellular networks. *IEEE communications magazine*, 48(11), 2010.

[13] Andreas Eisenblätter, Rafał Pisz, and Szymon Stefański. Greennets-white paper. 2013.

[14] Remco Litjens and Ljupco Jorguseski. Potential of energy-oriented network optimisation: Switching off over-capacity in off-peak hours. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, pages 1660–1664. IEEE, 2010.

[15] Gunther Auer, Vito Giannini, Claude Desset, Istvan Godor, Per Skillermark, Magnus Olsson, Muhammad Ali Imran, Dario Sabella, Manuel J Gonzalez, Oliver Blume, et al. How much energy is needed to run a wireless network? *IEEE Wireless Communications*, 18(5), 2011.

[16] Pål Frenger, Christer Friberg, Ylva Jading, Magnus Olsson, and O Persson. Radio network energy performance: Shifting focus from power to precision. *Ericsson Review*, 2014.

[17] Pål Frenger, Peter Moberg, Jens Malmodin, Ylva Jading, and István Gódor. Reducing energy consumption in LTE with cell DTX. In *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pages 1–5. IEEE, 2011.

[18] Per Skillermark and Pål Frenger. Enhancing energy efficiency in LTE with antenna muting. In *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pages 1–5. IEEE, 2012.

[19] 3GPP TR. 25.913. *Universal Mobile Telecommunications System (UMTS) LTE "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)",ETSI*, 25.913, 2009.

[20] Frédéric Firmin. The evolved packet core. *url=http://www.3gpp.org/technologies/keywords-acronyms/100-the-evolved-packet-core.*

[21] Lte network infrastructure and elements - lteencyclopedia, url= https://sites.google.com/site/lteencyclopedia/lte-network-infrastructure-and-elements.

[22] 3GPP TS. 36.211. *LTE, Evolved Universal Terrestrial Radio Access (E-UTRA), Physical channels and modulation", ETSI*, 36.211, 2011.

[23] Andrea Goldsmith. *Wireless communications.* Cambridge university press, 2005.

[24] Ericsson. Key performance indicators. 37/1553-HSC 105 50/1, 2017.

[25] ITU-T E Series. E.800. *Quality of service and dependability vocabulary*, 11/88, 1988.

[26] Ericsson Network IQ Statistics, url= https://www.ericsson.com/ourportfolio/network-management/network-iq-statistics, 2018.

[27] K. Trichias, R. Litjens, and A. Pais. LTE RAN features carrier aggregation study. *TNO Report TNO 2015 R10708*, 2015.

[28] Van Jacobson. Congestion avoidance and control. In *ACM SIGCOMM computer communication review*, volume 18, pages 314–329. ACM, 1988.

[29] IMEC. Power model for wireless base stations, url=http://www.imec.be/powermodel, 2018.

[30] Claude Desset, Björn Debaillie, and Filip Louagie. Towards a flexible and future-proof power model for cellular base stations. In *2013 24th Tyrrhenian International Workshop on Digital Communications-Green ICT (TIWDC)*, pages 1–6. IEEE, 2013.

[31] Claude Desset, Bjorn Debaillie, and Filip Louagie. Modeling the hardware power consumption of large scale antenna systems. In *Green Communications (OnlineGreencomm), 2014 IEEE Online Conference on*, pages 1–6. IEEE, 2014.

[32] Bjorn Debaillie, Claude Desset, and Filip Louagie. A flexible and future-proof power model for cellular base stations. In *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, pages 1–7. IEEE, 2015.

[33] Jari Salo. LTE radio load versus user throughput.

[34] Remco Litjens. Advanced topics in mobile communications (ET4396). *"Lecture Slides 5-7: UMTS/HSDPA"*, 2017.

[35] Mathworks. Curve fitting toolbox users guide. *MATLAB*, 2018.

[36] 3GPP TS. 36.213. *LTE Evolved Universal Terrestrial Radio Access (E-UTRA) "Physical layer procedures";Version 14,ETSI*, 36.213, 2018.

[37] Helmut Bolcskei and Arogyaswami J Paulraj. Space-frequency coded broadband OFDM systems. In *Wireless Communications and Networking Confernce, 2000. WCNC. 2000 IEEE*, volume 1, pages 1–6. IEEE, 2000.

[38] Siavash M Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on selected areas in communications*, 16(8):1451–1458, 1998.

[39] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.