TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

# Common Radio Resource Management in Virtual Heterogeneous Networks

**Maria Luísa Pedro Brito da Torre Caeiro**

**Supervisor:** Doctor Luís Manuel de Jesus Sousa Correia

**Co-Supervisor:** Doctor Filipe Duarte dos Santos Cardoso

Thesis approved in public session to obtain the
Ph.D. Degree in Electrical and Computer Engineering

Jury final classification: Pass with Merit

**Jury**

**Chairperson:** Chairman of the IST Scientific Board

**Members of the Committee:**

Doctor Edmundo Heitor Silva Monteiro
Doctor Ramón Agüero Calvo
Doctor Luís Manuel de Jesus Sousa Correia
Doctor Rui Manuel Rodrigues Rocha
Doctor Filipe Duarte dos Santos Cardoso
Doctor António José Castelo Branco Rodrigues

2014

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

# Common Radio Resource Management in Virtual Heterogeneous Networks

**Maria Luísa Pedro Brito da Torre Caeiro**

**Supervisor:** Doctor Luís Manuel de Jesus Sousa Correia
**Co-Supervisor:** Doctor Filipe Duarte dos Santos Cardoso

Thesis approved in public session to obtain the
Ph.D. Degree in Electrical and Computer Engineering
Jury final classification: Pass with Merit

**Jury**

**Chairperson:** Chairman of the IST Scientific Board

**Members of the Committee:**

Doctor Edmundo Heitor Silva Monteiro, Professor Catedrático da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Doctor Ramón Agüero Calvo, Professor Associado da Universidad de Cantabria, Espanha

Doctor Luís Manuel de Jesus Sousa Correia, Professor Associado do Instituto Superior Técnico, da Universidade de Lisboa

Doctor Rui Manuel Rodrigues Rocha, Professor Associado do Instituto Superior Técnico, da Universidade de Lisboa

Doctor Filipe Duarte dos Santos Cardoso, Professor Coordenador da Escola Superior de Tecnologia do Instituto Politécnico de Setúbal

Doctor António José Castelo Branco Rodrigues, Professor Auxiliar do Instituto Superior Técnico, da Universidade de Lisboa

2014

*To my beloved daughters Susana and Catarina*

# Acknowledgements

Above all, I wish to express my deep and sincere gratitude to Prof. Luis M. Correia and Prof. Filipe D. Cardoso for their guidance, support, motivation, corrections, productive ideas and discussions throughout this work.

Several people have contributed to this work by inspiring discussions and fruitful collaboration. I want to mention namely our Group for Research on Wireless (GROW), in particular Lúcio Ferreira and António Serrador, who work close to me in the European projects, in which this work was partially developed.

Special thanks go to my family and friends who have spared my company on many occasions when I has committed to this work. Above all, I thank my daughters Susana and Catarina and my father for their patience, love and support.

I would like to express a special acknowledgement to Escola Superior de Tecnologia de Setúbal and Instituto Politécnico de Setúbal, for all the support and for giving me the opportunity to be granted by "Programa de Apoio à Formação Avançada de Docentes do Ensino Superior Politécnico" (PROTEC).

Finally, I would like to mention the support given by Instituto de Telecomunicações, INOV – INESC Inovação, as well as the European Commission through the ICT-NEWCOM++, ICT-4WARD, ICT-SAIL, COST 2100 and IC 1004 projects, for partially funding this research.

# Abstract

This thesis addresses the virtualisation of wireless access in order to provide the required capacity to a set of Virtual Base Stations (VBSs) with diverse requirements, instantiated in a given geographical area. A novel network architecture, based on a generic network virtualisation environment, in which both physical and virtual perspectives are considered and the main stakeholders are taken into account, is proposed. A new tier of Radio Resource Management (RRM) is proposed for inter-VNets (Virtual Networks) RRM aiming at transposing the cooperative set of functionalities to the virtualisation environment. Two novel algorithms for VRRA are also proposed, taking the variability of the wireless medium into account, and continuously influencing RRM mechanisms according to VBSs' requirements. Algorithms' performance is evaluated through simulation, differentiation and isolation among VBSs being verified independently of the changes in usage profile and number of VBSs. Compared with current network operators' business model, the on demand proposed algorithm provides an increase of more than 45% of the cluster performance, guaranteeing the contracted minimum capacity to the guaranteed VBSs and maintaining the best effort VBSs close to the contracted reference one.

## Keywords

Virtual Networks, Wireless Access Virtualisation, Heterogeneous Wireless Networks, Virtual Radio Resource Allocation, Quality of Service.

# Resumo

Esta tese aborda a virtualização do acesso sem fios com o objetivo de fornecer capacidade a um conjunto de estações de base virtuais, com diferentes requisitos, criadas numa determinada área geográfica. É proposta uma nova arquitetura de rede, com base num ambiente genérico de virtualização da rede, onde as perspetivas física e virtual são consideradas e os principais intervenientes no novo modelo de negócio são identificados. Um novo nível de gestão de recursos rádio é proposto para gerir a atribuição dos recursos às redes virtuais, cujo principal objetivo é transpor o conjunto de funcionalidades da gestão cooperativa de recursos rádio para o ambiente de virtualização da rede. Propõem-se ainda dois algoritmos para a atribuição virtual de recursos rádio, que consideram a variabilidade do meio sem fios e influenciam continuamente os mecanismos de gestão de recursos rádio, próprios de cada tecnologia de acesso, para atuarem de acordo com o nível de satisfação dos requisitos dos recursos virtuais. O desempenho dos dois algoritmos foi avaliado através de simulação, tendo-se verificado diferenciação e isolamento entre recursos virtuais independentemente das alterações ao perfil de utilização, e quantidade de recursos virtuais. Comparando com o modelo de negócio atual dos operadores de comunicações móveis, o algoritmo proposto permite aumentar o desempenho do *cluster* em mais de 45%, garantindo a capacidade mínima contratada para os recursos virtuais do tipo garantido e mantendo o fornecimento de serviço próximo da capacidade de referência contratada.

# Palavras-chave

Redes Virtuais, Virtualização do Acesso sem Fios, Redes sem Fios Heterogéneas, Atribuição Virtual de Recursos Rádio, Qualidade de Serviço.

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| 2G | 2nd Generation |
| 3G | 3rd Generation |
| 3GPP | 3rd Generation Partnership Project |
| 4G | 4th Generation |
| 5G | 5th Generation |
| AC | Access Category |
| AN | Ambient Networks |
| AP | Access Point |
| ARP | Allocation and Retention Priority |
| ARQ | Automatic Repeat Request |
| AS | Access System |
| BE | Best Effort |
| BER | Bit Error Ratio |
| BLER | BLock Error Ratio |
| BS | Base Station |
| BSC | Base Station Controller |
| BTS | Base Transceiver Station |
| CDMA | Code Division Multiple Access |
| CF | Cost Function |
| CFQ | Channel Quality Feedback |
| CoRRM | Cooperative Radio Resource Management |
| CPU | Central Processing Unit |
| CRRM | Common RRM |
| CVRRM | Cooperative VNet RRM |
| CVRRM-C | CVRRM Coordination |
| CVRRM-d | CVRRM distributed |
| DE | Decision Element |
| DL | DownLink |
| EDGE | Enhanced Data rates for GSM Evolution |
| EE | Enforcement Element |
| EGPRS | Enhanced GPRS |
| eNodeB | Evolved NodeB |
| ETSI | European Telecommunications Standards Institute |
| eUTRAN | Enhanced UTRAN |

| | |
|---|---|
| FDMA | Frequency Division Multiple Access |
| FP | Forwarding Point |
| FTP | File Transfer Protocol |
| GBR | Guaranteed Bit Rate |
| GENI | Global Environment for Network Innovations |
| GERAN | GSM/Edge Radio Access Network |
| GLL | Generic Link Layer |
| GPRS | General Packet Radio Service |
| GRT | Guaranteed |
| GSM | Global System for Mobile Communications |
| GW | Gateway |
| HARQ | Hybrid Automatic Repeat reQuest |
| HIPERLAN | High Performance Local Area Network |
| HSDPA | High Speed Downlink Packet Access |
| HSPA | High Speed Packet Access |
| ICI | Inter-Carrier Interference |
| IE | Information Element |
| IEEE | Institute of Electrical and Electronics Engineers |
| IFFT | Inverse Fast Fourier Transform |
| IMS | IP Multimedia System |
| IMT | International Mobile Telecommunications |
| INC | Intra-/Inter-Node Communication |
| InP | Infrastructure Provider |
| IP | Internet Protocol |
| IP2W | IP to Wireless |
| ISP | Internet Service Provider |
| ITU | International Telecommunications Union |
| IVS | Initial VNet Selection |
| JRRM | Joint RRM |
| KPI | Key Performance Indicator |
| L2 | Layer 2 |
| L3 | Layer 3 |
| LAN | Local Area Network |
| LPFC | Linear Proportional Feedback Control |
| LTE | Long Term Evolution |
| MAC | Medium Access Control |
| MBR | Maximum Bit Rate |
| MCS | Modulation and Coding Scheme |
| MIMO | Multiple Input Multiple Output |
| MME | Mobility Management Entity |

| | |
|---|---|
| MN | Mobile Node |
| MRA | Multi-Radio Access |
| MRRM | Multi-Radio Resource Management |
| MSC | Mobile Switching Center |
| MVNO | Mobile VNet Operator |
| NVS | Network Virtualisation Substrate |
| OConS | Open Connectivity Services |
| OFDM | Orthogonal Frequency Division Multiplexing |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| OR | Orchestration Registry |
| OSAP | Orchestration Service Access Point |
| PAPR | Peak to Average Power Ratio |
| PCRF | Policy and Charging Rule Function |
| PRB | Physical Resource Block |
| PSMM | Pilot Strength Measurement Message |
| QCI | QoS Class Identifier |
| QoS | Quality of Service |
| RAC | Resource Allocation Control |
| RAN | Radio Access Network |
| RAT | Radio Access Technology |
| RRM | Radio Resource Management |
| RRP | Radio Resource Pools |
| RRU | Radio Resource Unit |
| SAE | System Architecture Evolution |
| SC | Single Carrier |
| SCRM | Supplemental Channel Request Message |
| SDMA | Space Division Multiple Access |
| SER | Symbol Error Rate |
| SINR | Signal to Interference plus Noise Ratio |
| SOP | Service Orchestration Process |
| SP | Service Provider |
| TDMA | Time Division Multiple Access |
| UE-MAC | User Equipment Medium Access Control |
| UL | UpLink |
| UMTS | Universal Mobile Telecommunications System |
| UPE | User Plane Entity |
| UTRAN | UMTS Terrestrial Radio Access Network |
| VBS | Virtual Base Station |
| VHO | Vertical HandOver |
| VHOS | Vertical HandOver Support |

| | |
|---|---|
| VINI | VNet Infrastructure |
| VLink | Virtual Link |
| VNet | Virtual Network |
| VNO | VNet Operator |
| VNode | Virtual Node |
| VNP | VNet Provider |
| VoIP | Voice over IP |
| VRRA | Virtual Radio Resource Allocation |
| VVS | VNet Valid Set |
| WCDMA | Wideband Code Division Multiple Access |
| Wi-Fi | Wireless Fidelity |
| WiMAX | Worldwide Interoperability for Microwave Access |
| WLAN | Wireless Local Area Network |
| WMAN | Wireless Metropolitan Area Network |

# List of Symbols

| | |
|---|---|
| $\gamma$ | SINR |
| $\gamma_{eff}$ | Effective SINR |
| $\gamma_{thr}^{RAT_r}$ | SINR threshold for RAT *r* |
| $\eta_{BS}$ | BS utilisation |
| $\eta_{Cl\,n}$ | Cluster utilisation in time frame *n* |
| $\overline{\eta_{Cl}}$ | Average cluster utilisation |
| $\eta_{VBS}$ | VBS utilisation |
| $\tau$ | Delay on service request |
| $\tau_{InP\,n}^{EU}$ | Total delay experienced by end-user *n* when the VBS has not reached the total amount of data rate contracted |
| $\tau_{VNO\,n}^{EU}$ | Total delay experienced by end-user *n* when the minimum VBS contracted data rate has been already reached |
| $\tau^{VBS}$ | VBS delay |
| $\overline{\tau_{InP}^{VBS}}$ | Average delay on service request InP |
| $\tau_{max}^{VBS}$ | VBS maximum delay |
| $\overline{\tau_{VNO}^{VBS}}$ | Average delay on service request VNO |
| $\upsilon$ | Modulation and coding scheme |
| $\xi_{R_{ref}}$ | Percentage of time frames allowed to be out of contract |
| $\xi_R^{RAT_r}$ | Percentage of contracted data rate allocated to the VBS on RAT *r* |
| $\Delta t_{TF}$ | Time frame |
| $BS_{RAT_r}^{Cl}$ | Set of BSs in the cluster from RAT *r* |
| $BS_n^{RAT_r}$ | BS *n* from $RAT_r$ in the cluster |
| $c_m$ | Monetary cost |
| $c_{BS}$ | BS cost |
| $c_{NT}$ | Network total cost |
| $c_o$ | Operator cost |
| $c_{o_{r,b}}$ | Cost for BS *b* of RAT *r* |

| | |
|---|---|
| $c_{u\,n}$ | End-user *n* cost |
| $E_b$ | Energy per bit Noise power spectral density ratio |
| $I_i$ | Interference power on end-user *i* |
| $I_c^{max}$ | Maximum interference power value |
| $k_{b,i}$ | Normalised value of KPI *i* for BS *b* |
| $k_{u,i}$ | Normalised user/VNO KPI *i* |
| $L_{max}^{VBS}$ | VBS maximum loss rate |
| $L_{p\,j}$ | Path loss for end-user *j* |
| $N_0$ | Noise power spectral density ratio |
| $N_{EU}$ | Number of active end-users |
| $N_{KPI_r}$ | Total number of KPIs for RAT *r* |
| $N_{KPI_u}$ | Total number of KPIs for users/VNOs |
| $N_{RAT}$ | Number of existing RATs in the cluster |
| $N_{TF}$ | Total number of time frames in the observation interval |
| $N_{RRU\,occ}^{BS}$ | Number of RRUs occupied by end-users in BS |
| $N_{BS}^{Cl}$ | Total number of cluster BSs |
| $N_{BE}^{Cl}$ | Total number of cluster BE VBSs |
| $N_{EU}^{Cl}$ | Total number of cluster end-users |
| $N_{GRT}^{Cl}$ | Total number of cluster GRT VBSs |
| $N_{VBS}^{Cl}$ | Total number of cluster VBSs |
| $N_{con}^{EU}$ | Number of end-users connected during the observation time interval |
| $N_{nc_{InP}}^{EU}$ | Number of end-users not connected during the observation time interval |
| $N_{nc_{ovl}}^{EU}$ | Number of end-users that are not connected when the capacity contracted for the VBS is already reached (VBS overloaded) |
| $N_{RRU}^{EU}$ | Number of RRUs assigned to the end-user |
| $N_{EU}^{max}$ | Maximum number of end-users in the cell |
| $N_{TF}^{out}$ | Number of time frames out of contract |
| $N_{BS}^{RAT_r}$ | Total number of BSs of RAT *r* in the cluster |
| $N_{RRU}^{RAT_r}$ | Total number of RRUs per BS of RAT *r* |

| | |
|---|---|
| $N_{RRU_{occ}}^{RAT_r}$ | Total number of RRUs occupied by end-users in RAT *r* |
| $N_{EU}^{VBS}$ | Number of end-users connected to the VBS over the observation interval |
| $N_{EU\ n}^{VBS}$ | Number of end-users connected to the VBS in time frame *n* |
| $N_{EU_{nc}}^{VBS}$ | Total number of end-users that tried to enter the network but have not been connected during the observation time interval |
| $N_{EU_T}^{VBS}$ | Total number of end-users that tried to connect to the VBS during the observation interval |
| $\overline{N_{EU}^{VBS}}$ | Average number of connected end-users during the observation interval |
| $N_{RRU_0}^{VBS}$ | Initial number of RRUs allocated to the VBS |
| $N_{RRU_r}^{VBS}$ | Total number of RRUs allocated to the VBS on RAT *r* |
| $N_{RRU_{r,occ}}^{VBS}$ | Total number of RRUs occupied by VBS end-users on RAT *r* |
| $p$ | Penalty |
| $p^{BE}$ | BE VBS Penalty |
| $p_i^{BE}$ | BE VBS penalty in time frame *i* |
| $p_m^{BE}$ | Penalty for BE VBS *m* |
| $p^{Cl}$ | Cluster Penalty |
| $p^{GRT}$ | GRT VBS penalty |
| $p_i^{GRT}$ | GRT VBS penalty in time frame *i* |
| $p_n^{GRT}$ | Penalty for GRT VBS *n* |
| $P^{av}$ | Power availability |
| $P_j$ | Transmitted power for/by end-user *j* |
| $r_{serv}^{VBS}$ | Ratio of served data rate |
| $r_{TF}^{out}$ | Out of contract |
| $R$ | Data rate |
| $R_{\upsilon_n}$ | RRU achieved data rate for modulation and coding scheme $\upsilon_n$ |
| $R_{av}^{BS}$ | Available data rate per BS |
| $R_{max}^{BS}$ | Maximum BS data rate |
| $R_{max}^{BS_n}$ | Maximum data rate for BS *n* |
| $R_{serv}^{BS}$ | BS serving data rate |
| $R_{serv}^{Cl}$ | Cluster serving data rate |

| | |
|---|---|
| $R_{serv\,n}^{Cl}$ | Cluster serving data rate in time frame *n* |
| $\overline{R_{serv}^{Cl}}$ | Average cluster serving data rate |
| $R_{max}^{Cl}$ | Maximum cluster data rate |
| $R_{serv}^{EU}$ | End-user data rate |
| $R_{serv\,n}^{EU}$ | End-user data rate for end-user *n* |
| $\overline{R_{serv}^{EU}}$ | Average end-user data rate |
| $R_{typ_n}^{EU}$ | Typical service data rate for end-user *n* |
| $R_{\gamma_n}^{RAT_r}$ | Data rate per RRU of RAT *r* associated with SINR threshold $\gamma_n$ |
| $R_{MCS_n}^{RAT_r}$ | Data rate per RRU of RAT *r* associated with MCS *n* |
| $R_{RRU_{max}}^{RAT_r}$ | Maximum RRU data rate for RAT *r* |
| $R_{min}^{VBS}$ | Minimum VBS contracted data rate |
| $R_{min}^{VBS_i}$ | Minimum contracted data rate for VBS *i* |
| $R_{ref}^{VBS}$ | Reference VBS contracted data rate |
| $R_{ref}^{VBS_j}$ | Reference contracted data rate for VBS *j* |
| $R_{req}^{VBS}$ | VBS requested data rate |
| $R_{req}^{VBS_i}$ | total data rate requested by end-users in VBS *i* |
| $R_{req\ n}^{VBS}$ | VBS requested data rate in time frame *n* |
| $\overline{R_{req}^{VBS}}$ | Average VBS requested data rate |
| $R_{serv}^{VBS}$ | VBS serving data rate |
| $R_{serv}^{VBS_i}$ | VBS serving data rate for VBS *i* |
| $\overline{R_{serv}^{VBS}}$ | Average VBS serving data rate |
| $R_{min}^{VNet}$ | Minimum VNet contracted data rate |
| $R_{serv}^{VNet}$ | VNet serving data rate |
| $\overline{R_{serv}^{VNet}}$ | Average VNet serving data rate |
| $RAT_{ID}$ | RAT identification |
| $RAT^{Cl}$ | Set of RATs in the cluster |
| $s_{av}$ | Service availability |

| | |
|---|---|
| $s_{ID}$ | Service identification |
| $s_{tos}$ | Type of service |
| $s_{Vol}$ | Service session data volume |
| $s_{\lambda}$ | Service inter-arrival time |
| $s_{\Delta t}$ | Service session time |
| $s_{min}^{R}$ | Minimum service data rate |
| $s_{typ}^{R}$ | Typical service data rate |
| $s_{max}^{\tau}$ | Maximum service delay |
| $S_{InP}^{VNO}$ | Satisfaction level on the InP |
| $S_{ovl}^{VNO}$ | Satisfaction level on extra capacity requested |
| $t_i$ | Time at which MN tries to enter the network |
| $t_{int}$ | Service time delayed |
| $t_{int\ n}^{EU}$ | Time duration the service has been delayed for end-user *n* when the VBS operates within contract |
| $t_{VNO\,n}^{EU}$ | Time duration the service has been delayed for end-user *n,* when the VBS operates out of contract |
| $\overline{t_{int}^{VBS}}$ | Average VBS time service delayed InP |
| $\overline{t_{VNO}^{VBS}}$ | Average VBS time service delayed VNO |
| $VBS^{BE}$ | Set of BE VBSs in the cluster |
| $VBS^{Cl}$ | Set of VBSs in the cluster |
| $VBS^{GRT}$ | Set of GRT VBSs in the cluster |
| $VNet_{ID}$ | VNet identification |
| $VNet_{mob}$ | VNet type of mobility |
| $VNet_{tos}$ | VNet type of service |
| $VNO_{ID}$ | VNO identification |
| $w_i$ | Weight of each user/VNO *i-th* KPI |
| $w_o$ | Operator/InP's weight |
| $w_{r,i}$ | Weight of *i-th* KPI of RAT *r* |
| $w_u$ | User/VNO's weight |

# List of Software

| | |
|---|---|
| **MS Excel 2007** | Excel is used for implementing analytical models and to process results. |
| **MS Visio 2003** | Visio is used to edit the figures presented in this thesis. |
| **MS Word 2007** | Word is used to edit this thesis and associated documents, such as publications and projects' technical reports. |
| **VRRA Simulator** | Thesis simulator, developed in Visual C++, and used to evaluate the majority of the proposed models. |

# Chapter 1

# Introduction

This chapter provides an introduction to the thesis, presenting first a brief history of mobile and wireless networks and network virtualisation. The motivation and main objectives of the thesis are described, and the novel aspects and concepts are explored and highlighted. An overview of the research strategy followed is done next, the involvement in European Projects and published work being presented. Finally, the structure of the thesis is described.

## 1.1 Brief History

In the last decades, wireless and mobile networks have experienced an extraordinary evolution, driven by the increasing number of mobile users, the emergence of new applications and services, and the innovations on radio access technologies. Nowadays, a range of mobile and wireless networks is available with different radio interfaces and network architectures, the focus of next generations of mobile networks being on the access to different wireless technologies at the same time and combining different flows from different technologies, Figure 1.1.

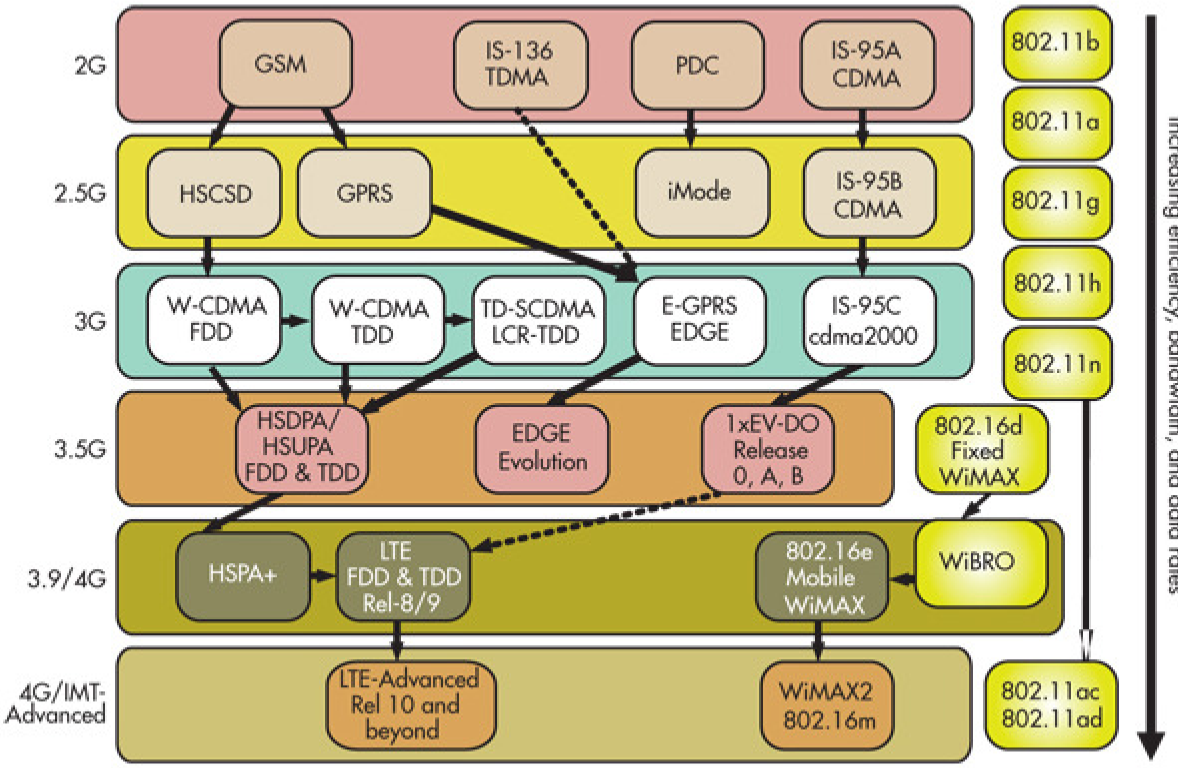



Figure 1.1. Wireless and Mobile Systems Evolution (extracted from [Fren11]).

The 1st generation of mobile communications systems was based on analogue cellular technologies, introduced by most of the European countries in the 1980's. The explosion of mobile communications took place with the introduction of the Global System for Mobile Communications (GSM), the digital 2nd Generation (2G). GSM was standardised by the European Telecommunications Standard Institute (ETSI), avoiding the fragmented and inefficient mobile market of the 1st generation. By creating a single market and introducing competition, and GSM as the *de facto* global standard for cellular phones, the European Union achieved economies of scale in the mobile sector and has been its world's leader. GSM networks are based on a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) in order to provide voice and data services over circuit switch. Packet switching services were integrated later, first with the introduction of General Packet Radio Service (GPRS) and after by Enhanced Data rates for GSM Evolution (EDGE), which allow increasing the bit rate by optimising the way that radio resources are used.

The 3rd Generation (3G) of mobile communications systems was both intended to meet the challenge of providing higher-speed data services and to further increase the radio network capacity. Universal Mobile Telecommunications System (UMTS) started as the joint European and Japanese system for 3G, being standardised by 3G Partnership Projects (3GPP). It is part of the International Mobile Telecommunications 2000 (IMT-2000) family of International Telecommunications Union (ITU), which defined the system's requirements, aiming at unifying the existing wireless access systems into a flexible radio infrastructure. The main multiple access technique used for 3G is based on Code Division Multiple Access (CDMA), Wideband CDMA (WCDMA), adopted for UMTS, becoming the dominant technology. In UMTS, circuit switching is maintained together with packet one, however, with the introduction of High Speed Packet Access (HSPA) schemes in UMTS Release 5 (2002) [HoTo04] and thereafter, the new releases of UMTS are focused on packet switching.

In addition to cellular systems, current wireless technologies include Wireless Local Area Networks (WLAN) Institute of Electrical and Electronics Engineers (IEEE) 802.11 [IEEE09] and Wireless Metropolitan Area Networks (WMAN) IEEE 802.16 [IEEE04b]. These systems are being also considered to support the evolution path of mobile communications, as part of the global wireless and mobile systems family, contributing to the optimisation of the diverse wireless resources utilisation and the increase of the overall systems capacity.

There were some 3G transitional cellular and wireless systems towards the 4th Generation (4G), comprising mainly Long Term Evolution (LTE, 3GPP Release 8) [3GPP09a] and Mobile Worldwide Interoperability for Microwave Access (WiMAX), IEEE802.16e [IEEE06]. LTE introduces a new highly simplified flatter network architecture, in order to allow seamless mobility with minimal latency. Only two types of nodes were defined, the evolved NodeB (eNodeB) and the Mobility Management Entity/Gateway (MME/GW), all interfaces being based on the Internet Protocol (IP). Switching was approved to be packet only, and the radio access changed from CDMA to Orthogonal FDMA (OFDMA) and Single Carrier FDMA (SC-FDMA). Mobile WiMAX (802.16e) is also based on OFDMA, supporting different bandwidths for both UpLink (UL) and DownLink (DL).

The most relevant factors that distinguish 4G networks are roaming across networks, IP interoperability, and higher data rates. In 4G, different access technologies, such as WLAN, WMAN and cellular, are combined on a common platform and interoperate to offer different services in different radio environments. The 4G IMT-advanced systems comprise mainly LTE advanced (LTE Release 10) and Mobile WiMAX advanced (IEEE802.16m) [Mous12]. LTE-Advanced is an evolution of LTE Releases 8 and 9, while IEEE802.16m is, in many respects, a new radio-access technology although retaining several of the basic characteristics of 802.16e, including the basic OFDMA. 802.16m introduces many features similar to LTE Release 10, such as the use of multi-carrier transmission (carrier aggregation) or the introduction of physical resource units for resource assignment.

5th Generation (5G) is a name used in some research papers and projects to denote the next major phase of mobile telecommunications standards beyond 4G/IMT-Advanced standards. At present, 5G is not a term officially used for any particular specification or in any official document yet made public by telecommunication companies or standardisation bodies. Meanwhile, some proposals for 5G are emerging, e.g., [Mous12], where a network architecture consisting of reconfigurable multi-technologies core and a single fully reconfigurable terminal able to autonomously operate in different heterogeneous access networks is proposed. The proposed network is enforced by nanotechnology, cloud computing and All IP Platforms.

It is envisaged that future networks will be networks of networks, consisting of multiple-access technologies, multiple bands, widely-varying coverage areas, all self-organised and optimised. Key to the future generations of mobile communications are multimedia communications, wireless access to broadband fixed networks, and seamless roaming among different systems.

Network virtualisation is a key enabler to support this holistic approach. A brief overview of network virtualisation and its evolution is made in what follows.

Network virtualisation derives from the concept of virtualisation in computer science further applied to memory, storage, and machine virtualisations. In communications, the terms “virtualisation” and “virtual” are associated with several technologies. Among these, one can note Asynchronous Transfer Mode (ATM) virtual circuits, Multi-protocol Label Switching (MPLS) virtual paths, Virtual Private Networks (VPN), Virtual Local Area Networks (VLAN), and Virtual overlay networks.

Most known network virtualisation techniques, such as ATM, MPLS, or other similar technologies, are regarded as link virtualisation techniques rather than full-fledged solutions for network virtualisation. More advanced solutions, such as VPNs, are also mere techniques using public networks for enhanced security and compatible execution environments for shared and legacy applications. VPNs are typically limited to providing simple Virtual Links (VLinks) or IP forwarding, and are not used for end-to-end deployments or full virtualisation of the underlying infrastructure.

The Internet architecture, developed over 30 years ago, has demonstrated its merits by the vast collection of applications it now supports, and the wide variety of network technologies over which it currently runs [APST05]. However, the Internet has suffered from its own tremendous success. The interplay of the end-to-end design of IP and the particular interests of competing stakeholders has led to its growing ossification. Changes to the Internet architecture that address its fundamental deficiencies or enable new services have been restricted to incremental changes. The main reasons are the required changes in routers and host software, and the Internet’s multi-provider nature, which requires that all Internet Service Providers (ISPs) jointly agree on any architectural change.

The inability to adapt to new pressures and requirements has led to an increasing number of ad-hoc workarounds, many of which violate the Internet's canonical architecture. While derided by architectural purists, these modifications have usually arisen to meet legitimate needs that the architecture itself could not support. These architectural add-ons can serve a valuable short-term purpose, but they significantly impair the Internet's long-term flexibility, reliability, and manageability.

A way to overcome the Internet impasse, as proposed in [APST05], is through virtualisation, whereby several network instances can co-exist on a common physical network infrastructure. Network virtualisation is an abstraction process aiming at isolating the logical network functionality from the underlying physical network resources. It enables the aggregation and provision of the network by combining different physical networks into a single virtual one, or splitting a physical network into multiple virtual ones, which are isolated from each other. By enabling a plurality of diverse network architectures to coexist on a shared physical substrate, virtualisation mitigates the ossifying forces in the current Internet and allows the continuous development of innovative network technologies.

Network virtualisation covers aspects like resource virtualisation and slicing. The virtualisation of the physical resources consists of implementing multiple instances of a required logical resource on a single machine/node within the same or different set of physical resources allocated to the Virtual Network (VNet), the slice. When compared to wired ones, wireless resources introduce some new challenges to virtualisation due the specific characteristics of the wireless environment. On the one hand, the isolation of traffic cannot be guaranteed due the scarcity of the radio spectrum, which cannot be over provisioned, while on the other hand, the radio signal propagation is a very node-specific property, being difficult to control, and has a significant impact on most VNets [SaBa08]. Slicing consists of allocating a coherent subset of physical resources to a specific VNet. The slicing process in wireless networks has also some specific issues derived from the characteristics of the medium; the provisioning of slices to multiple VNets with different radio links requires the capability to share radio resources, while at the same time avoiding interference among the different VNets [SaBa08].

Two main virtualisation approaches have been introduced in [APST05]: the "purist" and the "pluralist".

The architectural "purist" views virtualisation as a tool for architecture evaluation and the periodic deployment of successive, singular Internet architectures. Purists aim at architectural flexibility, because the architecture will remain in place a long time. Often, however, this flexibility does not result in immediate user benefits.

The "pluralist" view seeks to make virtualisation an architectural attribute of the Internet. Pluralists put more emphasis on short-term performance improvements, arguing that the desired flexibility derives from adding or augmenting overlays, rather than from the nature of each individual overlay. By enabling a plurality of diverse network architectures to coexist on a shared physical substrate, virtualisation mitigates the ossifying forces at work in the current Internet and enables continual introduction of innovative network technologies. Such a diversified Internet would allow the existence of architectural deficiencies to be holistically addressed, as well as enable the introduction of new architectures supporting new types of applications and services.

Network virtualisation is one of the key technologies for proceeding further in defining new generation networks. Network virtualisation refers to the instantiation of the logical entities (e.g., nodes, links etc.) on top of shared physical entities. Research projects all over the world, e.g., PlanetLab (USA) [Plan10], 4WARD (EU) [4WAR10] and Akari (Japan) [AKAR06] have adopted network virtualisation as a core technology. Therefore, various proposals for the architecture of the virtualisation of networks have emerged in recent years, some being related to virtualisation for test-beds, e.g., PlanetLab [PMRK06] or Orbit [RSOG05], and others for future networks, e.g., [SWPF09] and [ZZRR08].

The virtualisation of wireless networks has gained increasingly attention in the last few years. A set of wireless virtualisation techniques has been introduced and discussed in [SaSr06], essentially based on the main multiple access technologies referred for mobile and wireless systems. The idea behind the virtualisation of the physical resources is to allow a single machine/node to implement multiple instances of a required logical resource within the same or different set of physical resources allocated to the VNet (slice). Wireless virtualisation for specific Wireless and Mobile Networks has been recently addressed in literature, e.g., LTE [ZLGT10], WiMAX [BSMR10], and WLAN [BVSR10], [XKYG11], the majority of these approaches addressing mainly wireless resources virtualisation.

The definition and benefits of network virtualisation is a common understanding among all architecture proposals. The main goals to achieve through virtualisation are the isolation of physical network resources, and the holding of multiple independent and programmable logical networks. Implementing multiple network architectures on top of the isolated logical networks allows for a meta-architecture, where multiple architectures and test-beds for experimenting new architectures are enabled. The operation of such multiple networks leads to user and application specific logical networks, as well as to new business models for operators.

## 1.2 Motivation and Objectives

The rapid growth of Internet data traffic continues to be a reality, with mobile users having a significant contribution to this increase. It has been predicted that in the near future, mobile Internet usage will increasingly surpass the fixed one. It is forecasted that in 2017 wired devices will account for 45% of IP traffic, while Wi-Fi (Wireless Fidelity) and mobile devices will take the remaining 55% [Cisc13]. This increase of traffic cannot be followed by the expansion of network resources, since it will not be cost-efficient anymore. In wireless networks, more than cost, the problem is the inherently limited resources. In fact, the available radio resources are scarce with variable performance and there is lack of spare spectrum.

On the other hand, the diversity of Internet services and applications, e.g., voice, web browsing, file transfer and video streaming, with totally different requirements, as well as the arise of new business models for service provision, lead to changes in network operation. The integration of all these diverse services into a single protocol is not always feasible. Network virtualisation makes it possible to offer multiple optimised transport services, allowing clean-slate and legacy protocols to be deployed on separate VNets. The concept of network virtualisation provides the basis for an architecture that enables the deployment of multiple network solutions on top of a common network infrastructure, being considered the approach to be adopted for the Future Internet.

To overcome these issues, capacity sharing becomes a hot topic nowadays, allowing operators to split the high costs of network infrastructure by sharing the available wireless capacity. An adequate distribution of capacity should be done in order to satisfy the diverse service requirements, providing the requested capacity to the multiple operators, and maintaining a high level of resource utilisation without affecting applications performance. Wireless network sharing and wireless virtualisation are being proposed as the main approaches to deal with it.

Sharing radio resources in multiple access schemes for mobile and wireless systems has been intensively investigated, concerning the separation of the radio links for different end-users of the same system. The topic has also been explored by the research community for the introduction of Mobile VNet Operators (MVNOs) in 3G networks, [JoKS04], [AlBa05], [AMSE06], [HeWh06], and is now a hot topic for LTE. LTE network sharing standards can be found in [3GPP13a], and some proposals have been presented for active Radio Access Network (RAN) sharing, e.g., [Alca12] and [KMZR13]. More recently, the 3GPP group for RAN Sharing Enhancements identified a set of use cases to allow a more flexible and efficient RAN sharing [3GPP13b]; still, radio resource sharing is being confined to the same system.

The virtualisation of the wireless access, as a component of VNets, introduces some challenges, due to the specific characteristics of the wireless environment. The isolation of traffic cannot be guaranteed due to the scarcity of the radio spectrum, which cannot be over provisioned, and is variable in capacity due to channel conditions, interference, and end-users mobility. Furthermore, in addition to sharing infrastructure resources, if the Infrastructure Provider (InP) wants to provide some level of guarantees to the VNet operator (VNO), e.g., minimum data rate or delay, the resources to be shared and allocated to different VNOs with diverse requirements should be monitored and reallocated, to satisfy the established settings. The slicing process in wireless networks has also some specific issues, derived from the characteristics of the medium; the provisioning of slices to multiple VNets with different radio links requires the capability to share radio resources, while at the same time avoiding interference among the different VNets [SaBa08]. Managing radio resources sharing to provide VNet's requirements, such as contracted capacity, abstracting the involved wireless systems, becomes a major concern.

In the context of network virtualisation, the majority of the proposed approaches mainly address wireless resources virtualisation, which is not the focus of this work. Only some of them tackle the management of radio resources to be shared among the several VNets. Furthermore, in these approaches, the assignment of radio resources to VNet end-users is handled within one physical resource, in which the virtual resources are instantiated. Still, most of the current work does not address the allocation of radio resources based on the capacity required by the virtual resources, but rather based on a required amount of radio resources, which may perform differently according to the wireless medium conditions, possibly not providing the requested capacity.

In the wireless networks sharing approach, operators are forced to use similar network functions, as defined by 3GPP specifications, hence, the possibility of having different multiple VNets with their own functions and communication protocols, isolated from each other (the main advantage of network virtualisation), cannot be achieved. Still, without having an integrated perspective on the multiple radio access technologies, the abstraction of the wireless access is only partially achieved, preventing one to take full advantage of all available wireless infrastructures. Moreover, the several models proposed for radio resources sharing are not based on capacity request, the allocation of radio resources being more or less fixed, hence, not being dynamically adapted to the network state, in order to satisfy the requested capacity. This may lead to situations in which VNets are running out of contract, denying service to their end-users, even when some radio resources are available.

The main objective of this thesis is to identify, develop and propose new approaches to improve the Cooperative RRM (CoRRM) in environments where VNets are extended over heterogeneous wireless networks, Figure 1.2: the development of new mechanisms and policies for managing VNets radio resources sharing, accounting for the dynamicity introduced by VNets creation, variations in services demand, traffic load, physical resources, and capacity variability inherent to the wireless medium.

One fundamental requirement is to achieve an efficient integration of different Radio Access Technologies (RATs), making the problem a matter of CoRRM [SWLM04] with an additional dimension related to the multi-VNet environment. New stakeholders are expected in the market, like VNet Enablers, e.g., VNet Providers (VNPs) and VNOs, InPs, besides the existing Service Providers (SP). Since new relations and inter-dependencies must be considered, the interaction among these new stakeholders must be taken into account by CoRRM policies. Furthermore, the allocation of physical resources to different VNets introduces new constraints that need to be addressed. At the RRM level, these constraints should also be taken in account, since the controlled Radio Resource Units (RRUs) pools are not static (from the VNO viewpoint). In fact, they are grouped according to the allocation of VNets, and may be reallocated to another VNet, or simply do not belong to any VNet.

The main question that is addressed is how to cooperatively manage the radio resources sharing in the new VNets framework, in order to satisfy VNets' requirements.

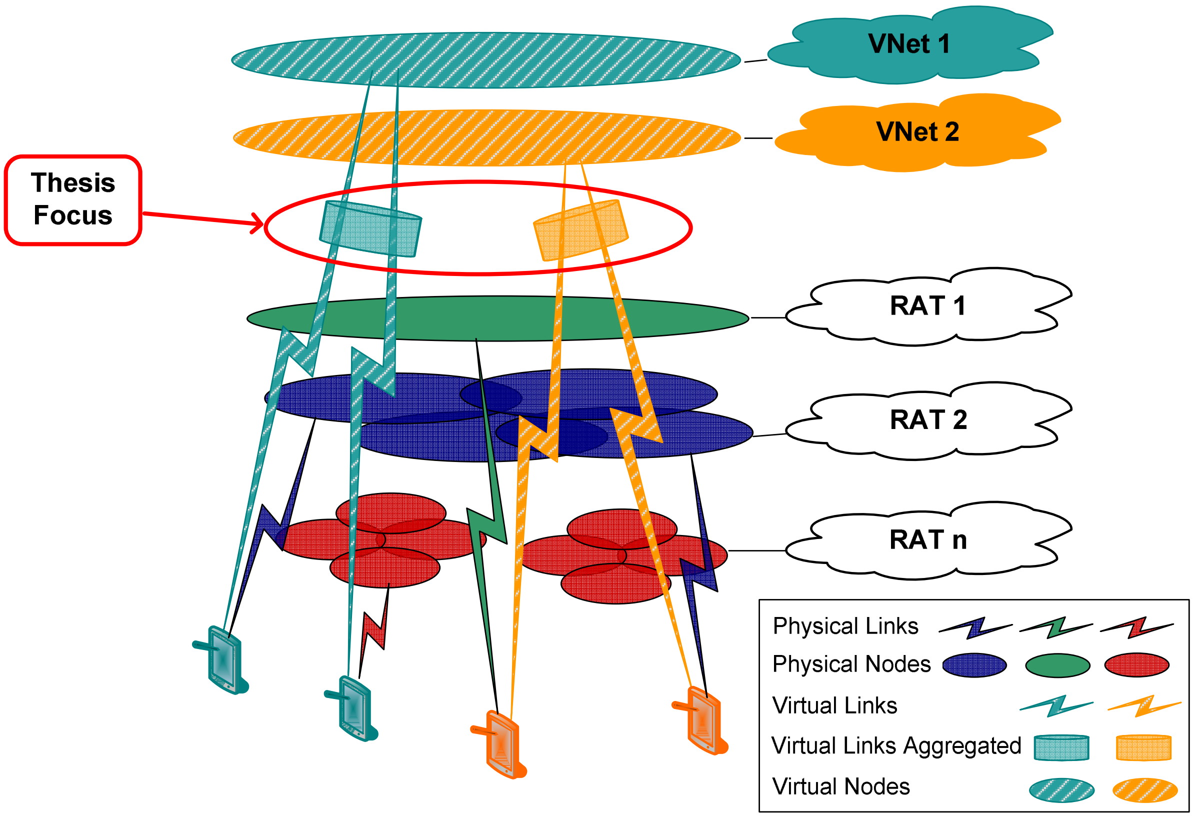



Figure 1.2. Thesis focus.

## 1.3 Novelty

The novelty of this work is the new approach for capacity sharing in wireless and mobile networks by the virtualisation of the wireless access, allowing to extend the classical VNet concept to the wireless access, while giving the possibility of deploying different logical network functionalities on top of it. Our proposal is to virtualise the wireless access by managing, in a common way, the radio resources available from heterogeneous wireless systems, in order to provide capacity to the several virtual resources, the VBSs, created over it.

Wireless virtualisation can introduce some additional promising aspects relative to wireless network sharing. In fact, besides the share of physical resources, network virtualisation main targets are the possibility of simultaneously running different network protocols over multiple VNets, which are isolated from each other, and with independent management functions. This way, by setting the VNet type of service with adequate requirements, e.g., minimum data rate and/or maximum delay, the VNets may satisfy their end-users' QoS. Thus, the differentiation among VNets can be achieved, enabling to deploy different levels of VNets guarantees.

The generalisation of the problem as a Cooperative Radio Resources Management problem with an additional level of abstraction, the virtual RRM one, allows following an approach of integration of the several levels of RRM, which needs to be adapted, but that actively participates in the process to achieve the main target of provision of the contracted level of service for all VNOs operating over the common infrastructure. Naturally, the added virtual RRM level needs to assume the coordination role of all underlying RRM levels, as it is aware of VNets requirements and has the responsibility to satisfy them. Still, the specific algorithms to implement the needed functionality at underlying RRM levels can evolve without overthrowing the outlined approach.

The proposed network model adds a virtual resource managing level, which according to VNets demands and the dynamic performance of the heterogeneous wireless network, enforces the existing RRM and/or CoRRM mechanisms to differently handle the end-users of the several deployed VNets. The proposed algorithms allows managing the allocation of capacity to the virtual resources, adapting the allocation of radio resources to wireless medium conditions and end-users demand, i.e., the amount of radio resources allocated to the virtual resource is not an issue, since the amount of contracted capacity is provided to the virtual resource. It is worthwhile to note that RRM scheduling strategies are out of the scope of this work. However, it is envisaged that setting scheduling parameters per VNet by modifying the existing Medium Access Contol (MAC) scheduling mechanisms is possible.

A claimed innovation is the broader perspective of virtual resources as an aggregated connectivity resource abstracted from a pool of RRUs of different RATs, allowing to benefit from CoRRM, i.e., managing radio resources across different technologies, and overcoming the limited bandwidth availability of wireless technologies. Instead of looking at the wireless virtualisation from the perspective of the instantiation of virtual machines in the wireless nodes, the proposed vision is the virtualisation of the wireless access to provide a required capacity to the VNet in order to serve its end-users. Hence, the proposed approach is agnostic to the point where the virtual node instantiation takes place, being possible to have virtual nodes in each physical wireless node, or somewhere in the cloud requesting virtual access over a given geographic area covered by a set of wireless nodes. Furthermore, the fact that this capacity can be modified on demand, without manually changing the configuration of the network, is another important innovation.

## 1.4 Research Strategy and Impact

The work developed in this thesis was done within the scope of different research European frameworks and projects, such as the Seventh Framework Programme (FP7-ICT) and Cooperation in the field of Scientific and Technical Research (COST), namely, ICT-4WARD [4WAR10], ICT-NEWCOM++ [NEWC11], ICT-SAIL [SAIL13], and COST Action IC 1004 [IC1013]. Although all these projects had a considerable work overhead beyond this thesis, they enabled sharing knowledge, visions and experience with multiple researchers of international institutions, namely networks' manufactures, cellular operators, research centres and universities, resulting in multiple cooperative activities and publications.

In the development of this thesis, these projects naturally had a considerable influence over many decisions taken. Reciprocally, the impact of the research activity carried in this thesis had impact on these projects. In this thesis, an Open Connectivity Services (OConS) functional architecture is proposed, which opened new potentialities for the integration of novel mechanisms in a modular approach. It was widely adopted within the ICT-SAIL project, to integrate several novel mechanisms, which were evaluated and demonstrated as a key result of the project. The proposed OnDemand Virtual Radio Resource Allocation (VRRA) algorithm was modelled according to the OConS architecture within the ICT-SAIL project, in order to demonstrate the advantages of its use within the OConS framework. Besides this, the Cooperative VNet RRM (CVRRM) approach was defined as part of the VNets architecture designed within the ICT-4WARD project, being evaluated for the provision of capacity for different types of virtual resources instantiated in heterogeneous wireless networks.

The work presented in this thesis was disseminated in several papers that have been published or submitted to various conferences and journals:

- International Journals:
  - L. Caeiro, F.D. Cardoso and L.M. Correia, "OnDemand Radio Resource Allocation for Virtual Wireless Access", submitted to *Wireless Personal Communications*, Feb. 2014.
- International Conferences:
  - Caeiro, L., Cardoso, F. and Correia, L.M., "Wireless Access Virtualisation: Adrdressing Virtual Resources with different Types of Requirements", in *Proc. of EuCNC'2014 - 23rd European Conference on Networks and Communications*, Bologna, Italy, Jun. 2014.
  - L.S. Ferreira, R. Agüero, L. Caeiro, A. Miron, M. Soellner, P. Schoo, L. Suciu, A. Timm-Giel and A. Udugama, "Open Connectivity Services for the Future Internet", in *Proc. of WCNC 2013 - IEEE Wireless Communications and Networking Conference*, Shanghai, China, Apr. 2013.
  - L. Caeiro, F.D. Cardoso and L.M. Correia, "OConS Supported On Demand Radio Resource Allocation for Virtual Connectivity", in *Proc. of MONAMI 2012 - 4th ICST International Conference on Mobile Network Management*, Hamburg, Germany, Sep. 2012.
  - L. Caeiro, F.D. Cardoso and L.M. Correia, "Adaptive Allocation of Virtual Radio Resources over Heterogeneous Wireless Networks", in *Proc. of EW2012 - European Wireless 2012*, Poznan, Poland, Apr. 2012.

- R. Agüero, L. Caeiro, L.M. Correia, L.S. Ferreira, M. García-Arranz, L. Suciu and A. Timm-Giel, "OConS: Towards Open Connectivity Services in the Future Internet", in *Proc. of MONAMI 2011 - 3rd ICST International Conference on Mobile Network Management*, Aveiro, Portugal, Sep. 2011.
- L. Caeiro, A. Serrador, F.D. Cardoso and L.M. Correia, "A Generic Service Interface for Cloud Networks", in *Proc. of Future Network & MobileSummit 2010*, Florence, Italy, June 2010.
- L.S. Ferreira, L. Caeiro, M. Ferreira and A.S. Nunes, "QoS performance evaluation of a WLAN mesh versus WIMAX network for an isolated village scenario", in *Proc. of EuroFGI Workshop on IP Quality of Service and Traffic Control*, Lisbon, Portugal, Dec. 2007.

The main contributions made within the European research projects were the following ones:

- Scalable and Adaptive Internet Solutions (ICT-SAIL) [SAIL13], an FP7-ICT Large-scale integrating project (2010-2013):
  - Applications for connectivity services and evaluation [MiSu13].
  - Architecture and mechanisms for connectivity services [TiSu13].
  - Architectural concepts of connectivity services [Suci11], [SuTi12].
- Architecture and Design for the Future Internet (ICT-4WARD) [4WAR10], an FP7-ICT Large-scale integrating project (2008-2010):
  - Virtualisation Approach: Evaluation and Integration Update [Bauc10b].
  - Virtualisation Approach: Evaluation and Integration [Bauc10a].
  - Virtualisation Approach: Evaluation and Integration (Draft) [Bauc09b].
  - Virtualisation Approach: Concept [Bauc09a].
  - Virtualisation Approach: Concept (Draft) [Bauc08b].
  - Milestone Report [Bauc08a].
- Network of Excellence in Wireless Communications ++ (NEWCOM++) [NEWC11], an FP7-ICT specific targeted research project (2008-2010):
  - Final report of the JRRM and ASM activities [Rome10].
- COST Action IC 1004 [IC1013] (2011-2015):
  - L. Caeiro, F.D. Cardoso and L.M. Correia, "Adaptive Allocation of Virtual Radio Resources over Heterogeneous Wireless Networks", IC1004 TD(11) 02018 on COST IC1004 2nd MC and Scientific Meeting, Lisbon, Portugal, Oct. 2011.
  - L. Caeiro, F.D. Cardoso and L.M. Correia, "Wireless Access Virtualisation: Physical versus Virtual Capacity", IC1004 TD(13) 08036 02018 on COST IC1004 8th MC and Scientific Meeting, Ghent, Belgium, Sep. 2013.

## 1.5 Contents

This thesis is structured into 8 chapters, and 3 appendixes. Their content is summarised below.

An introduction to the thesis is provided in the current Chapter 1, presenting a brief history of wireless and mobile networks as well as the evolution of network virtualisation in Section 1.1. In Section 1.2, the thesis motivation and objectives are presented, and in Section 1.3, the novel aspects and concepts explored in the thesis are highlighted. Section 1.4 provides an overview on the research strategy, and European Projects contributions and published work are identified. Finally, the dissertation contents are detailed in Section 1.5.

An overview of RRM in the several RATs within the scope of the thesis, and also the cooperative radio resource strategies proposed in literature, is given in Chapter 2. In Section 2.1, the fundamentals of RRM are described, and the main RRM functions for TDMA, CDMA, OFDM for WLANs and OFDMA based networks are identified. Section 2.2 addresses CoRRM in heterogeneous networks. Finally, in Section 2.1 the services and applications are presented.

An overview of network virtualisation is presented in Chapter 3, focusing on wireless virtualisation and related approaches, and on a novel framework for the provision of connectivity services. In Section 3.1, both experimental oriented and future concept architectures are presented. Wireless virtualisation challenges and proposals are presented in Section 3.2. In Section 3.3, the current RAN sharing is presented as an alternative to wireless virtualisation. An open connectivity framework is presented in Section 3.4, allowing to offer novel connectivity services, flexibly orchestrating legacy and novel mechanisms.

Novel models and algorithms to manage radio resources in virtualised environments are proposed in Chapter 4. The reference network architecture for the so called CVRRM is described in Section 4.1. The strategies used for RRM in virtualised environments, as well as the main CVRRM functions are presented in Section 4.2. The main assumptions and inputs considered in the model are presented in Section 4.3. Section 4.4 describes the analytical model and the approach used for data rate estimation. The evaluation metrics are presented in Section 4.5. Finally, in Section 4.6, the strategies and algorithms proposed for managing the radio resources sharing in VNets is presented, namely, the Cost Function (CF) for resources evaluation, the virtual radio resources allocation algorithms, the initial VNet selection, and the VNet handover support.

The implementation of models in a simulator is described in Chapter 5. In Section 5.1, the main assumptions taken for the simulator development are presented, giving a general overview of the simulator, presenting its main features and describing the main blocks. In Section 5.2, the details of the implementation are described, namely, traffic and end-users generation and the algorithms being used. Finally in Section 5.3, the simulator assessment is presented.

The scenarios and theoretical results for the proposed VRRA algorithms are presented in Chapter 6. In Section 6.1, the identification of the scenarios used for evaluation is made. In Sections 6.2 to 6.8, the theoretical reference values are presented and the comparison with simulated ones is done.

An analysis of simulation results for the different scenarios and use cases being considered is done in Chapter 7. After initial considerations made in Section 7.1, in Section 7.2, the performance of the proposed Adaptive-VRRA algorithm is evaluated for different strategies for instantiation of several virtual resources in the same physical cluster. Section 7.3 compares the virtualisation approach with the actual network deployment in which there are different operators each with its own infrastructure and clients. A comparison between the OnDemand-VRRA algorithm and the fixed allocation of radio resources is made in Section 7.4. In Section 7.5, several strategies for the provision of virtual capacity are compared. The variation of the number of virtual resources of each type is addressed in Section 7.6, and the impact of changing the amount of virtual resources created in one physical cluster is analysed in Section 7.7. Finally, in Section 7.8, the OnDemand-VRRA algorithm is evaluated for different service profiles.

The main conclusions of the thesis are presented in Chapter 8. Section 8.1 presents a summary of the thesis. Section 8.2 presents the main results including the novelty of the work developed within this thesis. Section 8.3 points out aspects to be addressed in future work.

Appendix A presents an overview of the data rate adaptation performed for several mobile and wireless systems, as well as the data rate values per RRU as a function of Signal to Interference plus Noise Ratio (SINR) for the mobile and wireless network systems under study. The traffic profiles generated by the simulator are represented graphically in Appendix B. In Appendix C, the simulator assessment results are presented, in particular, the ones related to the simulator transitory interval and the sensitivity to the number of simulations.

# Chapter 2

# RRM Basic Aspects

RRM and cooperative RRM basic aspects are introduced in this chapter, as they are key topics for the work developed in the thesis. An overview of network services and applications is provided first. The fundamental RRM concepts and the main functions applied for each of the considered radio access technologies are introduced next. Finally, the architectural proposals for cooperative RRM are presented.

## 2.1 RRM strategies for different Radio Access Technologies

### 2.1.1 Fundamental Concepts

One of the objectives of a network operator is to deploy a network able to support its customers with the required QoS. Focusing on the radio component, the output of radio network planning should be the provision of RRUs along the service area, by means of a certain radio network topology and a given configuration of the cell sites [RSAD05]. However, the amount of RRUs to be provisioned varies with service penetration and usage profile, which change in time and space. The most basic way to overcome these issues is by means of network over dimensioning, and RRUs overprovision, in order to guarantee QoS to end-users; but, radio resources are limited, and this alternative is not a cost-efficient one. The challenge is to be able to provide the desired QoS level with a minimum of resources, therefore, minimising operator's investment, while meeting network design requirements.

Wireless communications are dynamic in nature due to several varying conditions, including propagation, traffic generation, and interference, among others. Hence, the management of the provisioned RRUs should be also dynamic, in order to maintain end-users' QoS. RRM is in charge of allocating and managing the RRUs provided by the radio network, Figure 2.1.

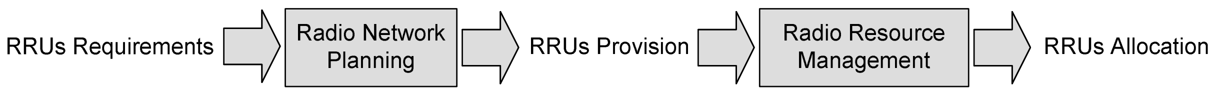



Figure 2.1. Relationship between Radio Network Planning and RRM (extracted from [RSAD05]).

RRM functions are responsible for taking decisions regarding the setting of different parameters influencing air interface behaviour [RSAD05]. The overall behaviour of the air interface at any given time results from the decisions taken by different RRM functions. However, consistency needs to be ensured among the different actions that will be undertaken by the different functions and mechanisms to solve conflicts deriving from contradictory actions/reactions. The correct design of RRM functions considers that some functionalities rely on actions/reactions of other functionalities to achieve a global performance.

RRM functions gather information and measurements related to the general radio environment and QoS. This can include Signal-to-Noise Ratio (SNR), throughput, delay of radio bearers, handover, and admission statistics, as well as technology dependent values, such as, channel allocation, orthogonal coding, and intra- and inter-cell interference values. Different RRM functions target different radio interface elements and effects, hence, they can be classified according to the time scales they use to be activated and executed. A set of RRM functions with the corresponding typical time scales between consecutive activations of the different algorithms are [RSAD05]:

- inner loop power control, e.g., 1 slot (less than 1 ms) in CDMA;
- packet scheduling and MAC algorithms, in around 1 frame;
- admission control, handover, congestion control, outer loop power control in CDMA transmission, from tens to thousands of frames.

RRM strategies are not subject of standardisation, hence, being a differentiation issue among manufacturers and operators. The RRM strategies of legacy networks, e.g., TDMA ones, only need a few parameters to tune their optimality. As the complexity increased in the radio interface, e.g., CDMA and OFDMA based networks, the need to increase and harmonise the general knowledge on RRM strategies has been an issue, since multiple dimensions appear in the problem. An RRU is the set of basic physical parameters necessary to support transmission between the mobile terminal and the base station for a given reference service, therefore, they differ according to the multiple access technique, and RRM is specific of each one.

### 2.1.2 TDMA based Networks

GSM is used as an example of TDMA, as it is a widely adopted cellular network. The role of RRM functions in GSM is to establish, maintain, and release communication links between end-users and the Mobile Switching Centre (MSC). The elements that are mainly concerned with RRM functions are the Mobile Node (MN) and the Base Transceiver Station (BTS). However, since the RRM component performs connection management also during cell handover, it also affects the MSC in its handover management component. In TDMA, the RRU corresponds to a time-slot.

The main RRM functions are Channel Assignment, Channel Allocation, Power Control, Frequency Hopping, and Handover.

Channel Assignment’s main target is to assign a required number of channels to each cell to achieve both efficient frequency spectrum utilisation and minimisation of interference effects. Channel assignment algorithms can be classified as static or dynamic. In a static approach, channels are allocated and prefixed to each cell during setup, according to the traffic intensity estimated for the cell. The dynamic one takes into account that traffic varies dynamically from cell to cell, assignment being performed in accordance to the actual traffic load. From a resource utilisation viewpoint, the dynamic approach is preferable over the fixed one, as it is designed to adjust resource assignment according to traffic demand, hence, supporting a higher capacity and lowering call blocking [SiSD12].

Efficient Channel Allocation is important for meeting the QoS requirements of both voice calls and packet connections in integrated GSM/GPRS networks. Several resource allocation algorithms were proposed, considering fix and dynamic resource allocation, with and without queue capability deallocation of channels from on-going packets and reallocation of released idling channels to GPRS data. In dynamic schemes, partial resources can be allocated to packet request, which can substantially reduce the GPRS dropping probability, and also indicates that the voice queuing mechanism could significantly lower call blocking. However, in channel re-allocation schemes, the decrease in call blocking is done at the expenses of the slight increment of packet dropping. Finally, a composed dynamic channel allocation scheme with guard channels, channel deallocation/re-allocation for voice call and packet queues can adapt to different QoS requirements, by adjusting the number of guard channels and the size of packet queues, achieving a better performance of QoS provisioning [ZhRe05].

Power Control is needed to control adjacent channel interference, and mitigate the interference caused by the near-far problem. To minimise co-channel interference and to conserve power, both MNs and BTSs operate at the lowest power level that maintains an acceptable signal quality. The MN measures the signal strength or signal quality, based on the Bit Error Ratio (BER), and passes the information onto the Base Station Controller (BSC), which ultimately decides if and when the power level should be changed. Power control should be handled carefully, in order to avoid instability, arising from having MNs in co-channel cells alternating increasing their power in response to higher co-channel interference caused by other MNs increasing their power [Scou97].

Frequency Hoping allows the MN and BTS to transmit each TDMA frame on a different carrier frequency. As multipath fading depends on the carrier frequency, slow frequency hopping helps to compensate the problem, since co-channel interference is actually randomised.

Concerning Handover, there are four different types in GSM, transferring a call between:

- logical channels (time-slots) in the same cell;
- cells under the control of the same BSC;
- cells under the control of different BSCs, but belonging to the same MSC;
- cells under the control of different MSCs.

Handovers can be initiated by either the MN or the BSC, as a means of traffic load balancing, being mainly controlled by the MSC. However, to avoid unnecessary signalling, the first two types of handovers are managed by the respective BSC. Although handover algorithms are not specified in the standards, there are two basic algorithms, both closely tied in with power control [Maca91]: the Minimum Acceptable Performance algorithm, which gives precedence to power control over handover, so that when the signal degrades beyond a certain point, the power level of the MN is increased; the Power Budget one uses handover to maintain or improve a certain level of signal quality, giving precedence to handover over power control, avoiding the “smeared” cell boundary problem, and reducing co-channel interference.

### 2.1.3 CDMA Based Networks

In CDMA based networks, and in particular in UMTS (WCDMA), capacity is tightly coupled to the degree of interference on the air interface. On the one hand, the performance of a given connection depends on the behaviour of the other users sharing the radio access interface, due to the use of different spread sequences, which in most cases are not perfectly orthogonal. On the other hand, the relevant parameter for interference is the number of simultaneous users (transmitting in a given moment), because it strongly impacts on the total interference. Additionally, in DL, the air interface capacity is directly determined by the required transmission power, which determines the transmitted interference, hence, in order to maximise capacity, the transmission power needed by one link is minimised. In UL, the transmission power determines the amount of interference to the adjacent cells, and the received power determines the amount of interference to other MNs in the same cell. In CDMA networks, the RRU is defined by a carrier frequency, a code sequence and a power level, [HoTo04].

According to the RRM target, the following functions are identified: Admission Control, Congestion Control, Code Management, Handover, User Equipment Medium Access Control (UE-MAC) and Packet Scheduling, and Power Control.

Admission Control is particularly relevant, because there is no hard limit on the maximum capacity. Since the maximum cell capacity is intrinsically connected to the amount of interference or, equivalently, to the cell load level, algorithms are based on measurements and/or estimation of the status of network load, as well as on the estimation of the load increase that the acceptance of a connection request causes. A request is admitted when QoS requirements can be met, provided that the already existing connections do not have these requirements affected by this request. The algorithms used in Admission Control are executed separately for UL and DL, because of different issues impacting on both links. However, a connection request can be admitted only after gaining permission from the corresponding UL and DL algorithms. These algorithms must take QoS requirements in terms of real- and non-real-time transmissions, and the variation of the resources needed for each connection along the time interval under consideration. Admission conditions for non-real-time traffic can be more relaxed if RRM mechanisms complementing admission control are able to limit this type of traffic when the air interface load is excessive.

Congestion Control faces situations in which QoS guarantees are at risk due to the evolution of system dynamics (mobility aspects, increase in interference, traffic variability, etc.). Congestion algorithms need to continuously monitor the network, in order to correct overload situations that are caused by excessive interference. The main network measurements that must be taken into account are DL transmitted power and UL cell load factor, which need to be suitably averaged, to avoid congestion problems by either false or no detections. These algorithms must be able to quickly react to overload conditions, in order to prevent degradation of links quality, residing on the network side.

Code Management is devoted to managing the DL Orthogonal Variable Spreading Factor code tree used to allocate physical channels among different users. Given that the number of available codes is limited, it is important to be able to allocate/reallocate codes in an efficient way. In general, a code allocation strategy aims at minimising code tree fragmentation, preserving the maximum number of high rate codes, and eliminating code blocking.

The Handover mechanism is controlled by the network, which is responsible for the measurements regarding UL and the overall system status, with the assistance of measurements regarding DL. These measurements can be done either periodically or event-triggered. The handover decision is carried out by means of an algorithm that is not standardised; nevertheless, some examples of algorithms are presented in 3GPP specifications. The pilot channel plays an important role in handover decisions: by increasing or decreasing its transmitted power level, more users can be attracted to or refrained from joining the network.

User Equipment Medium Access Control (UE-MAC) and Packet Scheduling algorithms are devoted to deciding the suitable radio transmission parameters for each connection in a reduced time scale and in a very dynamic way. In UL, this functionality is decentralised, so that the MAC layer of every MN executes the so-called UE-MAC algorithm to select the instantaneous bit rate to be applied in each transmission time interval for a given radio access bearer; however, to ensure specific QoS figures, the eligible transmission rates are only those defined by the network in a centralised way. In DL, the operation is naturally centralised and carried out by the packet scheduling algorithm; it is responsible for scheduling non-real-time transmissions over shared channels, following a time-based approach (i.e., multiplex of a low number of users simultaneously with relatively high bit rates), a code-based one (i.e., multiplex of a high number of users simultaneously with relatively low bit rates), or combinations of both. Prioritisation mechanisms can be considered in scheduling algorithms.

Tight and fast Power Control is perhaps the most important aspect in WCDMA, in particular in UL, since without it a single overpowered mobile terminal would block a whole cell. Two main mechanisms are used: the inner loop, which is responsible for adjusting the transmitted power, on a short time basis (on the order of 1 ms), so that the receiver get the required energy per bit to noise power spectral density ratio ($E_b/N_0$); the outer loop, which is responsible for selecting a suitable $E_b/N_0$ target, depending on BLock Error Ratio (BLER) or BER requirements, and operating on a slower time basis (on the order of seconds), adapting power control to changing environments.

### 2.1.4 OFDM for WLANs

OFDM divides a channel into a number of equally spaced, but mutually independent, frequency sub-carriers, each one transmitting lower data rates of a high-rate information stream. OFDM, used in WiFi, takes the coded signal for each sub-carrier and uses the Inverse Fast Fourier Transform (IFFT) to create a composite waveform. RRUs are then the sub-carriers.

The main advantages of OFDM are high spectral efficiency and inter-symbol interference prevention, this being a major problem in wideband transmission over multipath fading channels. Spectral efficiency is obtained since sub-carriers are chosen to be orthogonal to each other, i.e., cross-talk between the sub-carriers is eliminated and sub-carrier overlapping is allowed. Inter-symbol interference prevention results from, on the one hand, the transmission parallel low-rate streams, and on the other hand, the relatively long duration of symbols compared with the time characteristics of the channel, allowing for the insertion of a guard interval between the symbols.

Still, some drawbacks could be encountered in OFDM, like the sensitivity to Doppler shift and high Peak to Average Power Ratio (PAPR). The former arises from the sensitiveness to channel variations, destroying the orthogonality between sub-carriers and generating Inter-Carrier Interference (ICI). The latter occurs because the independent phases of the sub-carriers imply that they will often combine constructively, requiring more expensive transmitter circuitry and leading to poor power efficiency.

From an RRM perspective, two main aspects of wireless environments are relevant, both causing performance degradation, with impact on the corresponding QoS: one is the time-varying nature of the propagation channel, also of great importance to the other RATs; the other is the fact that several sub-carriers may experience interference or attenuation different from one another. Channel variations within one OFDM symbol introduce inter-carrier and co-channel interferences; which can be overcome by transmitting the lowest amount of power while satisfying rate constraints, or using the minimum bandwidth required to satisfy user constraints. Concerning the different interference and attenuation experienced by sub-carriers, adaptive modulation and channel coding is introduced, applied across all sub-carriers or individually to each one; a Dynamic OFDM concept, based on the principle that at any given time the individual sub-carriers do not have an identical gain or are subjected to a different level of interference, has been proposed [Bing90], followed by a family of approaches in which the transmitter adaptively controls the modulation type, the transmit power and the coding scheme applied on a per packet and sub-carrier bases, e.g., [GEPW07].

The following RRM functionalities are important: Dynamic Frequency Selection, Transmit Power Control, Admission Control, Congestion Control/Load Balancing, and Handover.

Dynamic Frequency Selection is used to dynamically switch the operational frequency channel from an AP to another, allowing the selection of a new frequency channel when the current one is under interference of neighbour devices or noise. Since the propagation environment changes over time, a procedure that senses the radio environment periodically is implemented in order to allocate the best frequency channel. An optimal channel assignment for a given WLAN should minimise the overlap between coverage areas of co-channel APs.

Co-channel interference is a major factor affecting link outage probability, thus, the use of Transmit Power Control to have the lowest level that satisfies rate constraints is a key function. In an extended deployment of WLANs, composed of several APs in a given area, inter-cell interference may generate severe transmission errors, and the use of power control may cause gaps in coverage. The APs transmit power should be set in a way that compensates for those coverage gaps, without excessive coverage overlap.

Admission Control is an important function, since the bandwidth of each AP is shared among all the users associated with it. New users cannot be accept when the majority of the AP resources are already allocated. Although the best received signal is one of the main conditions for admission control, the usage and availability of resources must be considered to guarantee that users already connected to the network maintain the targeted requirements for their services.

Load Balancing/Congestion Control continuously monitors the network in order to analyse the received signal quality and the load of each AP in a given area. Based on this evaluation, users can be associated with another AP in order to avoid QoS degradation and to maximise the utilisation of the overall radio resources. This feature is used in addition to admission control, to best distribute the users across APs, due to users' mobility, resulting in some APs experiencing low utilisation whilst others being near congestion.

Handover is directly associated with Load Balancing/Congestion Control, i.e., when a near congestion situation is detected by the latter, and the decision is to switch a user from one AP to another, the former is triggered. Handover is also related to Admission Control, given that after the decision to switch a user to another AP, the association acceptance of the new user needs to be performed.

### 2.1.5 OFDMA Based Networks

OFDMA, being based on OFDM, provides a multi-access scheme for a multi-user communication system, inheriting its immunity to inter-symbol interference and frequency selective fading. In fact, OFDMA is much more than just a physical layer solution, exploiting the unique physical properties of OFDM by enabling significantly higher layer advantages that contribute to very efficient packet data transmission [ChGu06], being used in LTE.

In OFDMA, there is a structure of both time-slots and sub-carriers, assigned to different users for multiple access, therefore, the channel is shared among various users, each owning a mutually disjoint set of time-slots and sub-carriers. Sub-carriers are allocated to a user according to the amount of information to be sent, controlled by the MAC layer, by scheduling resource assignment based on user demands. RRUs are defined by a set of sub-carriers and time-slot, also designated by Resource Blocks.

RRM functions are based on link adaptation and sub-carrier allocation. The former is already proposed for a single user OFDM, in order to adjust the allocation of sub-carriers. The latter considers the instantaneous information of the radio channel parameters obtained for each user, in order to assign the best set of sub-carriers that allows the optimum use of the resources.

RRM functions like Admission Control, Congestion Control, Packet Scheduling and Handover are used to optimise radio resources utilisation, maintaining a good perceived service quality for users through the whole network.

When a new user initiates a connection request, Admission Control computes how many resources are needed to support the requested QoS, then checks whether the required resources can be satisfied, providing that the service quality of existing connections is not reduced [LiNi05]. It takes into account the resource allocation varies substantially as the number of users and allocated sub-carriers changes. The number of supported users is determined by finding the number of sub-carriers to be assigned using the required data rate, BER and average channel gain of each user [HJSY06].

Congestion Control deals with the dynamism of the subcarrier allocation process and the adaptive modulation characteristics of the system. It monitors the resources available, in order to prevent the overloading of the system and users' QoS degradation. Actions for this feature result in refusing new connections, or in switching users to another BS.

Packet scheduling is performed at the MAC sub-layer, being responsible for defining the transmission order of packets, from different competing flows [RRSS05]. At a given instant, the packet scheduler tries to maximise system performance in terms of different QoS requirements, such as delay, loss rate, throughput, and utilisation of limited radio resources, in response to bursty data traffic and time-varying channel conditions. In order to meet a required minimum throughput or a maximum delay for a given user, it assigns a set of sub-carriers that in each time instant allows accomplishing the given requirements.

Again, Handover is related to admission control and congestion control. The former when the handover process is initiated, because admission control must be used to accept or reject the new connection. The latter because a decision made by congestion control in near overloaded situations can be to trigger a handover.

## 2.2 Cooperative RRM in Heterogeneous Wireless Networks

The mobile communications network is composed of various types of RATs that constitute a global heterogeneous wireless network. New RATs may appear in future generation mobile communications, enforcing the need for cooperation among them, in order to provide users the best connectivity anytime and anywhere. The heterogeneous wireless networks concept is intended to propose a flexible and open architecture for a large variety of different wireless access technologies, for applications and services with different QoS demands, and different protocols. The main goal is to make the heterogeneous network transparent to users, a secondary one being to design an architecture that is independent of the wireless access technology.

In order to accomplish these objectives, and to optimise the global radio resources utilisation, cooperation among the specific RRMs of each air interface technology is needed. The complementary characteristics of the different RATs allow achieving a more efficient use of the overall resources with CoRRM, rather than with the usage of the various RRMs independently, the so-called trunking gain. CoRRM must take into consideration the overall resources in all available RATs, and dynamically select the best RAT, in order to guarantee at each moment the most efficient use of the available radio resources. A vertical handover procedure must be considered, in order to enable a number of necessary features: avoiding disconnections due to lack of coverage in the current RAT; avoiding blocking due to overload in the current RAT; improvement of QoS by changing RAT; supporting user's and operator's preferences in terms of RATs usage or load balance among RATs. Inter-RRM signalling among RATs should be also required, in order to transfer information among RRM entities upon which resource allocation and admission control decisions can be taken.

A number of architectures and algorithms to implement CoRRM have been studied and proposed in the last few years. An overview of the main ones is presented in what follows.

Several approaches to the cooperation among different RATs in a heterogeneous network were made. A Common RRM (CRRM) approach was proposed by 3GPP to enable the cooperation in between UMTS and GSM [RSAD05]. CRRM is a mechanism for an intelligent distribution of traffic among systems, offering the possibility to increase the overall network capacity and user perceived QoS, thereby reducing network costs. 3GPP introduced a new entity designated by CRRM, defining Radio Resource Pools (RRPs), which are sub-sets of the whole set of resources from an operator. These RRPs are controlled by RRM entities, which are responsible for RRM inside an RRP, and by CRRM entities, whose function is to coordinate RRM entities, Figure 2.2.

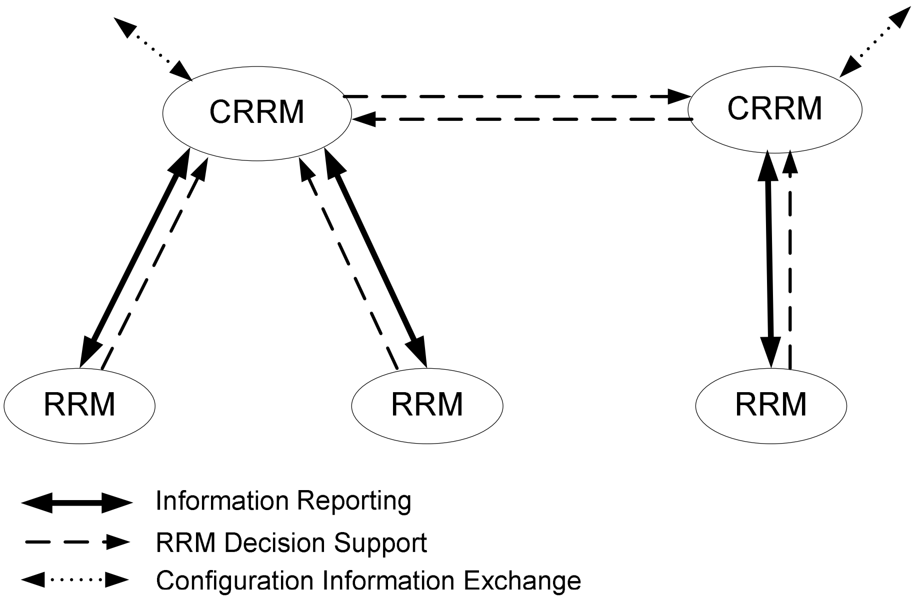
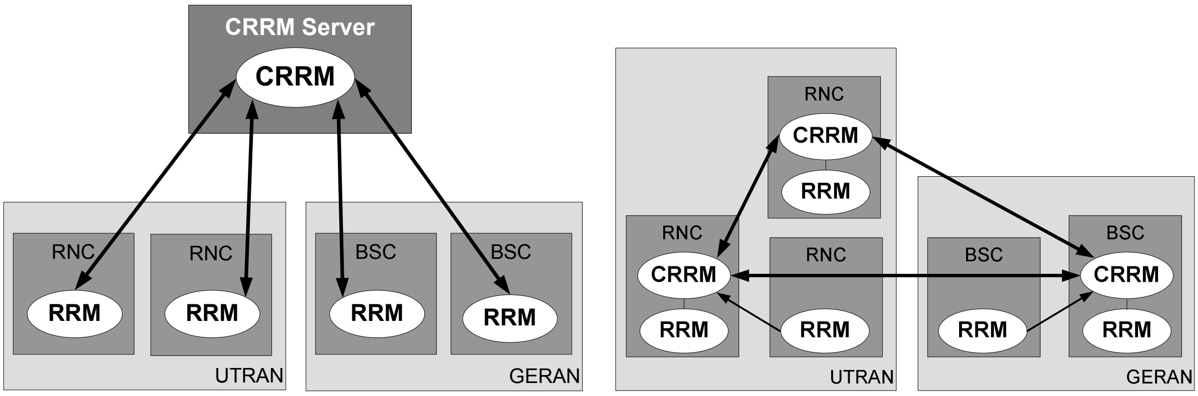



Figure 2.2. Coordination among different radio resource pools (based on [RSAD05]).

CRRM should direct users to the RRP that is the most suitable in terms of users' service and network constraints, such as minimising interference or fostering load balancing. Two main architectures have been proposed for CRRM [3GPP03]: CRRM server and integrated CRRM, Figure 2.3. The former implements RRM and CRRM entities into separate nodes, CRRM being a stand-alone server, while the latter integrates the CRRM functionality into the existing UMTS Terrestrial RAN (UTRAN)/GSM RAN (GERAN) nodes. This last approach has the main benefit of requiring limited changes to achieve optimal system performance, since almost all the required ingredients to support the CRRM functionality already exists.

Within the scope of the European COST273 project, a policy based CRRM approach [Meag02] was proposed to 3GPP for Release 6. This would allow a centralised CRRM entity to provide policies to the RRM ones, thus, enabling traffic to be dynamically adjusted in the network on the basis of a common strategy. The CRRM server can act as a policy manager for the access to the radio bearer resources within UTRAN/GERAN, by performing the RRM algorithms that are based on dynamic status information per cell, from all cells in the network. The CRRM server is also connected to other RANs, allowing dynamic inter-system RRM.




(a) CRRM server. (b) Integrated CRRM.

Figure 2.3. CRRM approaches (Based on [RSAD05]).

Joint RRM (JRRM) was introduced by the European IST-SCOUT project for inter-working between High Performance Local Area Network (HIPERLAN/2) and UMTS [HLPS05]. JRRM is similar to CRRM, but complements it with additional features and algorithms, and further radio systems. The architecture of JRRM is quite similar to the one of CRRM, whereby resources are centrally allocated for all involved RATs by a single entity. Moreover, JRRM complements CRRM by several modifications and additional features, its architecture corresponding to a very tight coupling, where traffic is split among RATs. Optimal QoS can be achieved with traffic splitting supported by adaptive radio multi-homing, which provides multiple radio access for a single terminal in order to allow the mobile terminal to maintain simultaneous links to different RATs.

Two main entities are defined to optimise spectral efficiency for coupled systems, Joint Radio Resource Scheduling and Joint Session Admission Control, handling various bearer types (e.g., voice and video) and different users' and services' QoS constraints, and scheduling traffic adaptively for mixed traffic types. With the information of the estimated load in all sub-networks, the Joint Load Control entity located together with Joint Admission Control distributes traffic based on the characteristics of the co-existing RATs. The joint scheduling is important for mobile terminals having simultaneous connections to several networks. The amount of resources to be offered, based on user traffic's QoS requirements, is assigned to contributing networks.

A multi-layered RRM scheme was introduced by the European IST-MIND project [MIND02] for the cooperation among various RATs [SHHS02]. The need for a multi-layered approach appears as a consequence of the proposed network, which could be formed not only of multiple technologies but also of multiple domains. When multiple technologies are introduced, different Layer 2 (L2) will interact with each other, and there should be a layer that is the bridge among technologies. In the Multiple Domain-Multiple Technology, it is considered that an area of commonality is the IP layer, Layer 3 (L3), which is used as the bridge among technologies through the IP to Wireless (IP2W) interface. At L3, a decision can then be made on the best resource management across multiple technologies, removing the inter-domain management conflicts. The generic framework for the multi-layered approach is presented in Figure 2.4, where RRM entities at both L2 and L3 can be seen. The approach is hierarchical, L2 having self-contained resource management for the Single Technology-Single Domain case, which improves the design efficiency. The L3 RRM entity takes inputs from the specific access technology with generic messages across IP2W. It is worthwhile noting that the L2 RRM entity has the same architecture for different access technologies. A manager function is also included in the multi-layered architecture, which is responsible for the interactions management among access technologies specific RRM entities, such as coordinating handover.

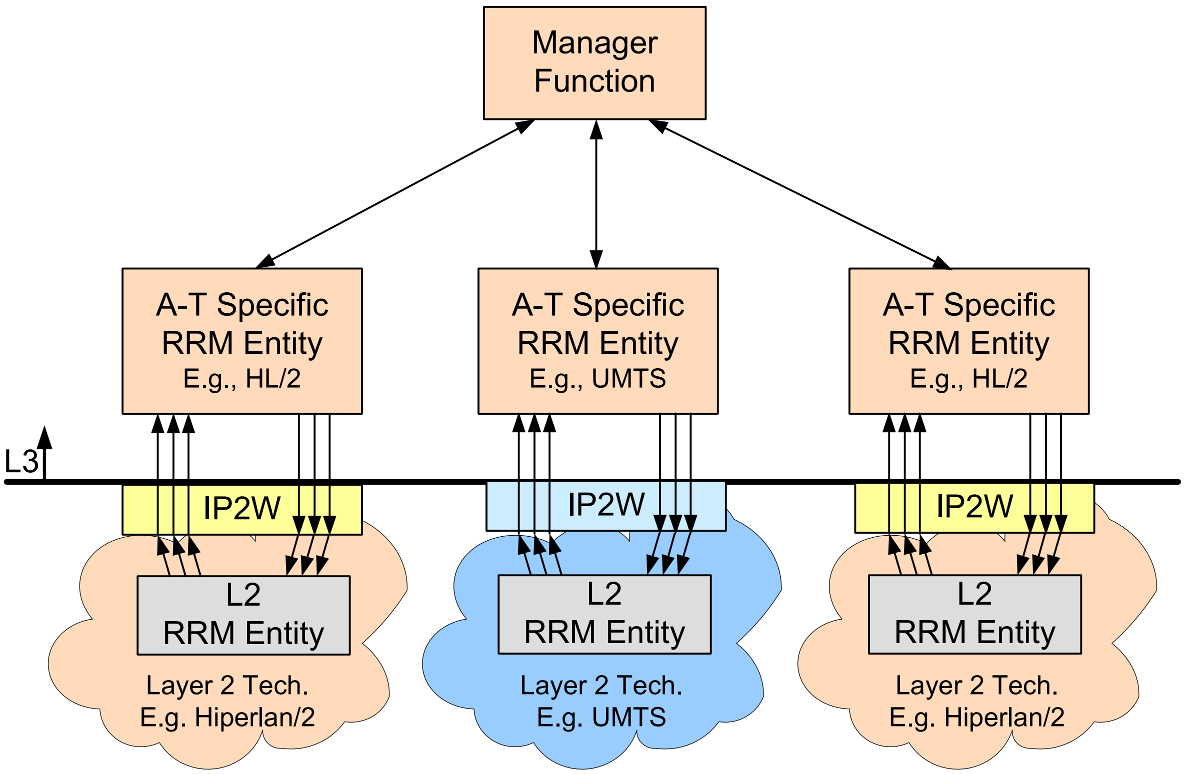



Figure 2.4. The multi-layered architecture (extracted from [SHHS02]).

The European IST Ambient Networks (AN) project [Ambi07] defined the Multi-Radio Access (MRA) architecture, Figure 2.5, consisting of Multi-RRM (MRRM) and Generic Link Layer (GLL) functionalities [SABG06]. One of the key objectives of this architecture is the efficient use of the multi-radio resources by means of effective radio access selection mechanisms. MRRM is the key control entity in the MRA architecture, complementing the radio technology specific RRM, such that the selection to activate a RAT for a user session is decided in the MRRM. Decisions are based on the link state information provided by the GLL entity in the radio links, as well as on other available information, e.g., service requirements, resource costs and current resource availability, operator and user policies, network and cell load, and terminal capabilities. However, the execution to activate an alternative access is done by a Handover Control function, which controls data handling in a Forwarding Point (FP) when switching over the user session to a new access flow. FP is a routing decision point that maps higher level flows to access flows, being located in the MRA Anchor point.

The MRRM functionality is distributed among multiple MRRM entities that may take on different roles in their joint operation. The Access Selection Function in the core network is the master MRRM entity responsible for deciding on the best-suited access for a bearer, the Access Network Control Function located in each access network monitors access network related parameters, and the Connection Management Function in each mobile terminal monitors its access flow quality. MRRM tasks can be distributed in a centralised or decentralised way, among MRRM entities of different ANs. The distribution aspects define the roles of the MRRM entities, which depend on network composition agreements (e.g., master-slave relation), provided services, and properties of the constituting RANs.

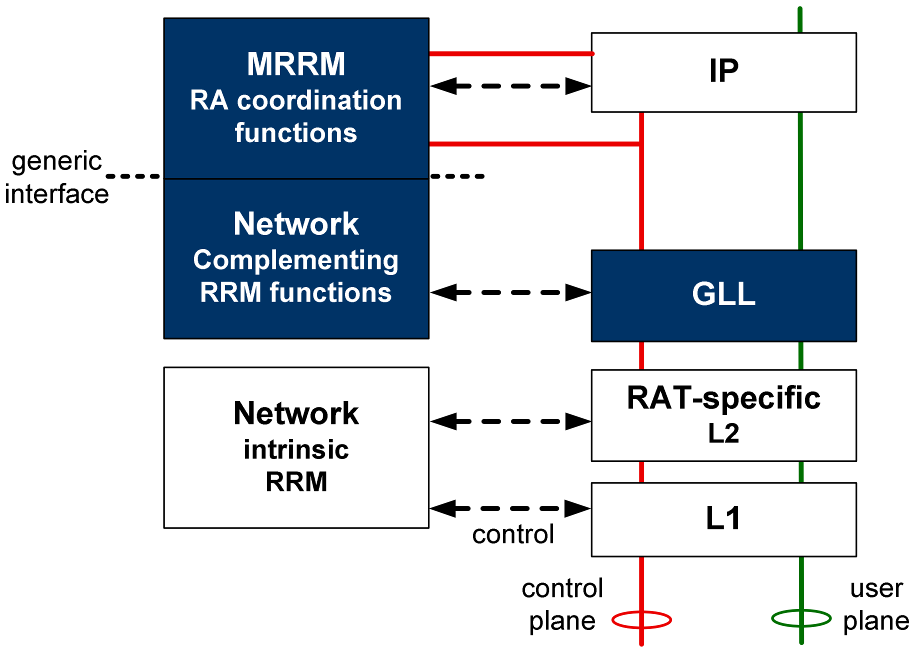



Figure 2.5. MRA Architecture (extracted from [SABG06]).

Although not being a strict CoRRM architecture, the 3GPP System Architecture Evolution (SAE) integrates different access technologies into a common packet core network, allowing for inter-system handover [3GPP08]. One of the main objectives of this architecture, Figure 2.6, is the support of mobility between multiple heterogeneous RATs. Some new entities and interfaces have been defined in order to reach this target: MME, User Plane Entity (UPE), and 3GPP and SAE anchors. The MME main function is to manage MN mobility and MN identity, performing MN authentication and authorisation, idle mode MN tracking and reachability, and security negotiations. The UPE manages the user data path, including parameters of the IP service and routing. The 3GPP and SAE anchors are respectively the mobility anchor between GSM/UMTS and LTE, and the mobility anchor between

3GPP and non-3GPP networks.

In terms of interfaces, seven new ones were defined:

- control-Plane between eNodeB and MME;
- user-Plane between eNodeB and UPE;
- mobility support between WLAN, 3GPP IP access or non-3GPP IP access, and Inter Access System (AS) Anchor;
- user and bearer information exchange;
- inter 3GPP access system mobility or mobility support between GPRS Core and Inter AS Anchor;
- transfer of subscription and authentication data for user access to the evolved system;
- transfer of QoS policy and charging rules from Policy and Charging Rule Function (PCRF).

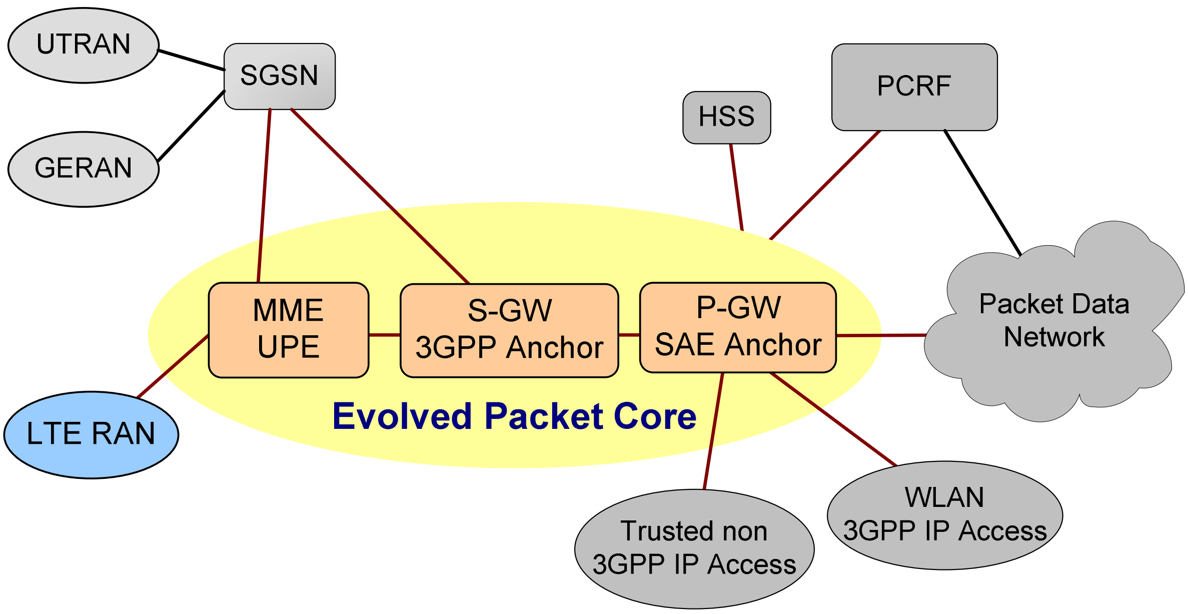



Figure 2.6. Logical high level architecture for the evolved network (based on [3GPP08]).

The main RRM functions, e.g., admission control, power control, scheduling and handover, are common to all RATs, the relative importance of each one depending of the specificities of each RAT. Due to the current coexistence on the same serving area of several RATs, and the trend for integration of several wireless networks, as defined by SAE, the implementation of CoRRM strategies enables to benefit from the specificities of each RAT, the diversity of services, and the flexibility of end-users. The different proposed architectures for CoRRM shows the diverse degrees of RRM functions' splitting, e.g., by different OSI layers or by functionality, the introduced abstraction layer being common in order to decouple the specificities of the radio interface of each RAT. Of course the inherent functions of CoRRM, e.g., initial RAT selection or VHO, are always running at a high level, as they need to have the global knowledge of all RATs. This kind of approach, where the global functionality is performed at a high and common level among all RATs, the abstraction layer generalises or unifies the comparison among RATs, and RAT dependent RRM functions are performed at the RAT level, are the basis to transpose CoRRM to a virtualised network environment.

## 2.3 Services and Applications

A basic overview on services and applications is presented, since their different characteristics determine the use of the most appropriate RRM strategies.

Within the scope of CDMA based networks, particularly in UMTS, 3GPP has classified services as shown in Table 2.1, where one may find four main service classes: Conversational, Streaming, Interactive and Background. These classes may be described as follows [HoTo04]:

- Conversational – real-time applications, characterised by a strict low end-to-end delay (<400 ms, preferable <150 ms). Services have near-symmetric two-way traffic. Voice, video telephony, and some games that need very low delay, are the best examples for this category.
- Streaming - streaming data transfer applications (e.g., web broadcast, audio streaming, and video streaming on demand) are characterised by high asymmetric traffic (DL being the most significant one) and some delay tolerance. Information is transported in a continuous stream, allowing its processing by the MN (e.g., visualisation) before the reception of the entire file is finished. The usage of buffers allows this class of services to be more delay tolerant (<5 s), compared to the Conversational one.
- Interactive - client-server applications (e.g., web browsing, database access, and games), where a low round trip delay is required, are in this class. The user can ask for different kinds of information from a certain remote server. Services in this class are generally more tolerant to delays, and generate an asymmetric traffic. However, there is no tolerance to errors, thus, the error probability in the received data must be low to prevent too many retransmissions.
- Background - highly delay tolerant applications (e.g., short messages, e-mail, and database download). The common aspect relies on that the user does not have a limited time to receive the information, hence, the network does not need to process it immediately, which allows for high delays. Applications in this class only use the network for information transmission when its resources are not being used by applications from the other service classes. Despite the delays, information should not have errors.

Table 2.1. 3GPP service class classification (extracted from [HoTo04]).

| | | **Traffic class** | | | |
|---|---|---|---|---|---|
| | | **Conversational** | **Streaming** | **Interactive** | **Background** |
| **Fundamental characteristics** | **Connection delay (main attribute)** | Minimum fixed | Minimum variable | Moderate variable | Big variable |
| | **Buffering** | No | Allowed | Allowed | Allowed |
| | **Nature of traffic** | Symmetric | Asymmetric | Asymmetric | Asymmetric |
| | **Bandwidth** | Guaranteed bit rate | Guaranteed bit rate | No guaranteed bit rate | No guaranteed bit rate |

The characteristics of a WLAN are somehow different from cellular networks, but even so, the 3GPP classification can be used as a starting point to analyse services differentiation. However, related to the IEEE802.11 WLAN standard, another classification is proposed in the IEEE802.1D standard [IEEE04a] based on wired LANs, some new traffic types being defined:

- Network Control - the most important traffic that must have priority over the rest;
- Voice - very stringent regarding delay, the maximum delay being as low as possible;
- Video - with some limitations regarding the maximum delay, but not as severe as with the Voice traffic type;
- Controlled Load - having important applications subject to admission control and with controlled throughput;
- Excellent Effort - a best effort traffic with higher priority than lower classes;
- Best Effort (BE) - the normal LAN traffic;
- Background - traffic that is allowed on the network, but that should not interfere with the traffic from any of the other classes.

The correspondence between these traffic types and User Priority is presented in Table 2.2.

Table 2.2. Mapping between User Priority and traffic types (extracted from [IEEE04a]).

| **User Priority/Traffic Classes** | **Traffic type** |
|---|---|
| 1 | Background |
| 2 | *Spare* |
| 0 (Default) | Best Effort |
| 3 | Excellent Effort |
| 4 | Controlled Load |
| 5 | Video |
| 6 | Voice |
| 7 | Network Control |

Considering the IEEE802.11 WLAN set of standards, User Priority is then mapped onto the Traffic or Access Categories (ACs) according to the 802.11e standard [IEEE05a], Table 2.2. Best Effort (0), Video Probe (1), Video (2) and Voice (3) are the four ACs that have been defined for the Enhanced Distributed Channel Access, one of the new channel access mechanisms introduced by 802.11e, allowing traffic from different applications to have different priorities while accessing the channel.

Concerning OFDMA for cellular networks, in particular LTE, nine classes of service were standardised by 3GPP [3GPP12], being identified by a QoS Class Identifier (QCI). These are targeted at common services types: conversational voice and video, streaming video, gaming, IP Multimedia System (IMS) signalling, and differentiated access. The QCI indicates a specific priority, maximum delay, and packet error rate. Table 2.4 presents the standardised class of services and the associated QoS characteristics. The goal of standardising a QCI with corresponding characteristics is to ensure that applications/services mapped onto that QCI receive the same minimum level of QoS in multi-vendor network deployments and in case of roaming.

Table 2.3. Mapping between Traffic Classes and ACs (extracted from [IEEE05a]).

| User Priority/Traffic Classes | Access Category | Traffic Type defined in 802.11e |
|---|---|---|
| 1 | 0 | Background |
| 2 | 0 | Background |
| 0 (Default) | 1 | Best Effort |
| 3 | 1 | Best Effort |
| 4 | 2 | Video |
| 5 | 2 | Video |
| 6 | 3 | Voice |
| 7 | 3 | Voice |

To support end-to-end QoS for IP-based traffic, LTE uses the concept of service data flows and bearers. Each service data flow is associated with one and only one QCI, which is then mapped onto bearers, enabling a differential treatment for traffic with differing QoS requirements. Each bearer is associated with a set of QoS parameters that describe the properties of the transport channel, including bit rate, packet delay, packet loss, BER, and scheduling policy in the BS. A bearer has two or four QoS parameters, including QCI, depending on whether it is a real-time or best-effort service:

- QCI;
- Allocation and Retention Priority (ARP);
- Guaranteed Bit Rate (GBR) - real-time services only;
- Maximum Bit Rate (MBR) - real-time services only.

Table 2.4. LTE Standardised QCI Characteristics (extracted from [3GPP12]).

| QCI | Resource Type | Priority | Packet Delay Budget [ms] | Packet Error Loss Rate | Example Services |
|---|---|---|---|---|---|
| 1 | GBR | 2 | 100 | $10^{-2}$ | Conversational Voice |
| 2 | | 4 | 150 | $10^{-3}$ | Conversational Video (Live Streaming) |
| 3 | | 3 | 50 | | Real-time Gaming |
| 4 | | 5 | 300 | $10^{-6}$ | Non-Conversational Video (Buffered Streaming) |
| 5 | Non-GBR | 1 | 100 | | IMS Signalling |
| 6 | | 6 | 300 | | Video (Buffered Streaming); TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc). |
| 7 | | 7 | 100 | $10^{-3}$ | Voice; Video (Live Streaming); Interactive Gaming |
| 8 | | 8 | 300 | $10^{-6}$ | Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.) |
| 9 | | 9 | | | |

The ARP is used in bearer establishment, and can become a particularly important parameter in handover situations where a mobile terminal roams to a cell that is heavily congested. The network looks at the ARP when determining if new dedicated bearers can be established through the radio BS. The 3GPP standards provide mechanisms to drop or downgrade lower-priority bearers in situations where the network becomes congested.

A bearer can have GBR or not (non-GBR). A GBR bearer has a minimum amount of bandwidth that is reserved by the network, and always consumes resources in a BS, regardless of whether it is used or not. Non-GBR bearers are used for best-effort services, such as file downloads, email, and Internet browsing, and do not have specific network bandwidth allocation requirements. The MBR is not specified on a per-bearer basis for non-GBR bearers, however, an Aggregate MBR is specified on a per-subscriber basis for all non-GBR bearers.

# Chapter 3

# Virtual Networks Overview

An introduction to network virtualisation, presenting the major architectural approaches and the new related key actors is initially done. The basic techniques for the virtualisation of the wireless resources and for wireless network slicing are described, and an overview of the main proposals presented in the scientific community for wireless virtualisation is made. Finally, a novel open and flexible framework for the provision and management of new and legacy connectivity services is introduced.

## 3.1 Network Virtualisation

As previously referred in Chapter 1, there are two main approaches for network virtualisation, the purist and the pluralist. The former considers the virtualisation of the network as a tool for network architecture evaluation, while the latter takes virtualisation as an architectural attribute of the global network. In this section, a summary of the main architectures proposed for both views is made, referring to the former as experimental oriented architectures and to the latter as future concept ones.

Experimental oriented architectures do not consider several key factors relevant for virtualising the (commercial) Internet; they assume a hierarchical trust model that centres on a universally trusted entity. To overcome this limitation, competing players with individual administrative zones that have only limited trust and also have the desire to hide information, e.g., their topologies, are considered.

PlanetLab is a highly successful example of a distributed, large scale testbed [BBCC04]. PlanetLab has a hierarchical model of trust, which is realised by Planet Lab Central (PLC). PLC is operated by the PlanetLab organisation, being the ultimately trusted entity that authorises access to resources. Other actors are the infrastructure owners and the users that run their research experiments. For each experiment, virtual machines on various nodes are grouped into slices that can be managed and bootstrapped together. As the deployed virtualisation mechanism offers only container based virtualisation capabilities at the system level, and does not virtualises the network stack, PlanetLab offers no network virtualisation as such.

VNet Infrastructure (VINI) is a testbed platform that extends the concept of virtualisation to the network infrastructure [BFHP06]. In VINI, routers are virtualised and interconnected by VLinks. As such, VINI allows researchers to deploy and evaluate new network architectures with real routing software, traffic loads, and network events. VINI supports simultaneous experiments with arbitrary network topologies on a shared physical infrastructure. VINI builds on the architecture and management framework introduced by PlanetLab, by extending the management with interfaces to configure VLinks. The most updated implementation, being based on Trellis, allows for a higher forwarding performance. It introduces a lower level system virtualisation architecture, which uses container based virtualisation techniques for both system and network stack virtualisation [BMMM08], thus, virtualisation flexibility is limited to the user space. VINI provides rudimentary concepts for end-user attachments, being a solution that would not scale to a wide scale Internet [SWPF09].

Emulab is also a very popular testbed platform. It offers sophisticated management and life-cycle processes, but not that much of a network architecture [Dike00]. Emulab offers a virtual topology configuration and automatic bootstrapping of experiment nodes. Initially, Emulab focused on dedicated servers, virtualisation capabilities being added later.

Global Environment for Network Innovations (GENI) is a large-scale initiative in the United States, for building a federated virtualised testbed, aiming at providing a powerful set-up for experimental purposes. In GENI, all operations are signed off and managed by a central Geni Clearing House, which can be regarded as being similar to a VNP [ElFa09]. Key GENI concepts include [Muss12]:

- Programmability - the deep programmability of all resources, including computational and network resources. Researchers may download software into GENI-compatible nodes to control how those nodes behave.
- Virtualisation and Other Forms of Resource Sharing - a virtualised (shared) infrastructure, where each GENI experimenter (or other GENI user) gathers resources into their own isolated "slice" of resources, and then configures them to support experiments; whenever feasible, nodes implement virtual machines, allowing multiple researchers to simultaneously share the infrastructure and each experiment running on its own, and isolated slices created end-to-end.
- Federation - a federated heterogeneous infrastructure that is evolving over time. Different parts of the GENI suite are owned and/or operated by different organisations, and the National Science Foundation portion of the GENI suite forms only a part of the overall 'ecosystem'.
- Slice-based Experimentation - GENI experiments will be an interconnected set of reserved resources on platforms in diverse locations. Researchers will remotely discover, reserve, configure, program, debug, operate, manage, and teardown distributed systems established across parts of the GENI suite.

During the first phase of the development, both VINI/Planetlab and Emulab were used as GENI prototypes (ProtoGeni). GENI is currently in the new extended "Meso-scale Deployment" infrastructure stage, which started in October 2009; after that, GENI enabled commercial hardware was deployed across 13 university campuses, including 10 clusters of servers and Virtual Machines, plus 8 WiMAX sites, linked by build-outs through two US national research backbones. In October 2011, new projects were approved to extend the infrastructure within the next 3 years [Muss12].

Concerning future concept architectures, a generic network virtualisation environment is depicted in Figure 3.1. Three views of the network are represented:

- Physical infrastructure - physical networking resources, such as routers and link infrastructure.
- Virtualised Substrate - set of slices created in the physical resources.
- Virtual Networks (VNets) - logic networks instantiated on demand by a virtualisation enabler, using the virtualised substrate.

These three network views involve the need of three processes, Figure 3.1:

- Virtualisation of Resources - The process of partitioning the physical resources into slices. It should be noted that the physical infrastructure must enable the virtualisation of physical networking resources.
- Provisioning of VNets - The process of building VNets using virtual resources and partial topologies, triggered by a request for a VNet, which may be instantiated on demand;
- Management of VNets - Once the VNet has been constructed, it must be managed in order to offer services, which involves access to the virtual resources composing the VNet.

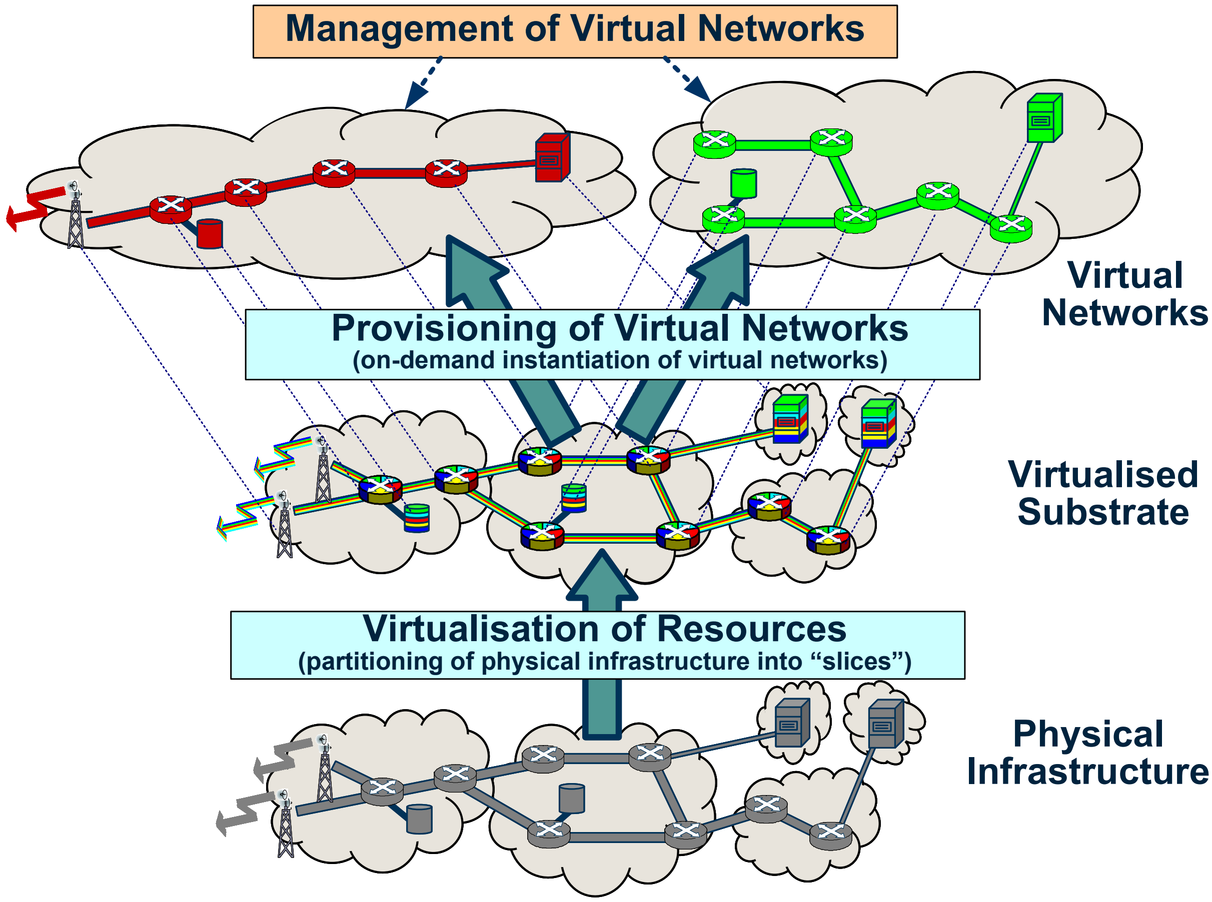



Figure 3.1. Network virtualisation overview (extracted from [4WARD]).

Economic models and use cases are not critical for testbed design, but are crucial for the adoption of Internet-wide virtualisation architectures. Several business models have been proposed to this extent. In [ChBo10], the players in the network virtualisation model are the end-user, the SP, the InP and a Broker, Figure 3.2.

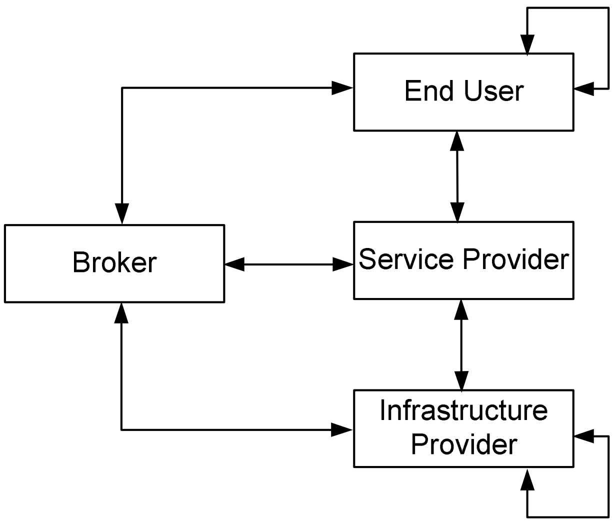



Figure 3.2. Network virtualisation business model (extracted from [ChBo10]).

End Users are similar to the end-users in the existing Internet, except that the existence of multiple VNets from competing SPs enables them to choose from a wide range of services. InPs deploy and actually manage the underlying physical network resources. This role has been decoupled from the conventional roles for SPs, which, in this virtualisation environment, lease resources from multiple InPs to create VNets and deploy customised protocols, by programming the allocated network resources to offer end-to-end services to end-users. Finally, Brokers act as mediators between InPs, SPs, and end-users in the network virtualisation marketplace, allowing them to select the desired services from a wide range of SPs.

CABO shares the same roles definition as the above for InPs and SPs [FeGR07], but in Cabernet [ZZRR08] a "Connectivity Layer" is introduced between them, Figure 3.3. This connectivity layer uses VLinks purchased from InPs to run VNets with the necessary geographic footprint, reliability, and performance for the SPs. It facilitates the entry of new SPs, by abstracting the negotiations with different InPs, and allows for aggregation of several partial networks into one set of infrastructure level resource reservations.

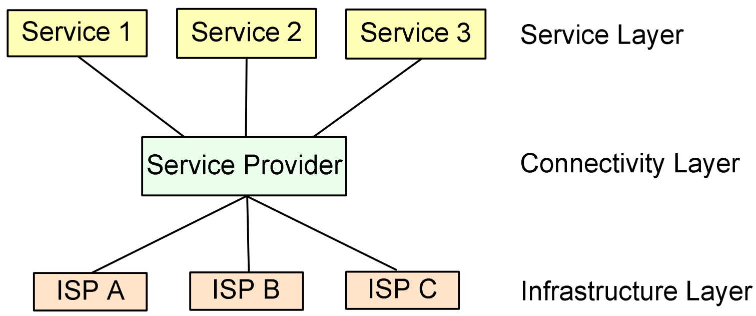



Figure 3.3. Hourglass model of Cabernet architecture (extracted from [ZLGT10]).

In line with this structure, [SWPF09] split the SP and Connectivity Provider roles into three different roles of VNP, VNO, and SP, Figure 3.4. These roles allow for a more granular splitting of responsibilities with respect to network provisioning, network operation, and service specific operations that may be mapped onto different business entities, according to various different business models. The following business rules are defined:

- Physical Infrastructure Provider (InP), which owns and manages the physical infrastructure (the substrate) and provides wholesale of raw bit and processing services (i.e., slices), hence, supporting network virtualisation.
- VNet Provider (VNP), which is responsible for assembling virtual resources from one or multiple InPs into a virtual topology.
- VNet Operator (VNO), which is responsible for the installation and operation of a VNet over the virtual topology provided by the VNP according to the needs of the SP, and thus realises a tailored connectivity service.
- Service Provider (SP), using the VNet to offer a service, which can be an added value one where the SP acts as an application, or a transport one with the SP acting as a network.

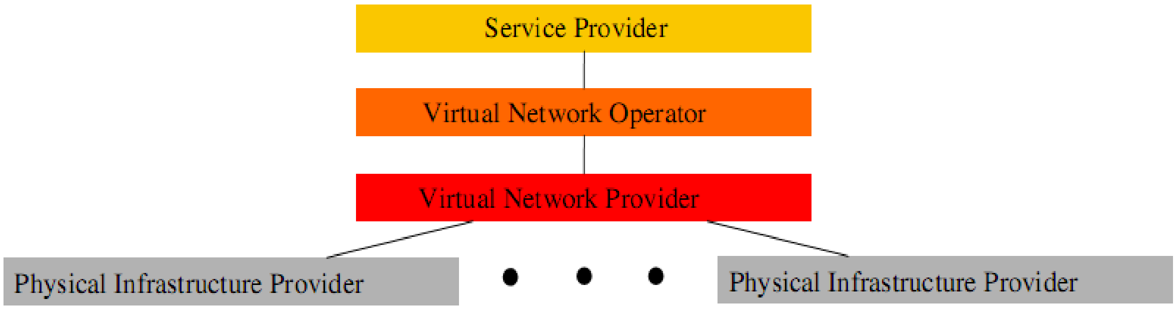



Figure 3.4. Network virtualisation business rules (extracted from [SWPF09]).

This work depicts the network virtualisation approach developed within 4WARD project [4WAR10], in which the main idea is the possible support it can provide for the coexistence of different network architectures in a secure and isolated manner, but maintaining their interoperability. The virtualisation of network resources originates VNets composed of several slices (sets of virtual resources) for each of the different network architectures. The current Internet might be considered as one of the different network architectures, and thus it might exist and interoperate with the others as a VNet. In order to have a global network, in which the user must be always best connected, the interoperability between the different VNets is to be done by using the so called folding points, which is the point where two VNets meet and can communicate across each other. Physically, folding points could be located in one physical site in which two or more Virtual Nodes (VNodes) are hosted, or between two adjacent physical sites and even more, and could be the user equipment itself with the capability to access different VNets, Figure 3.5.

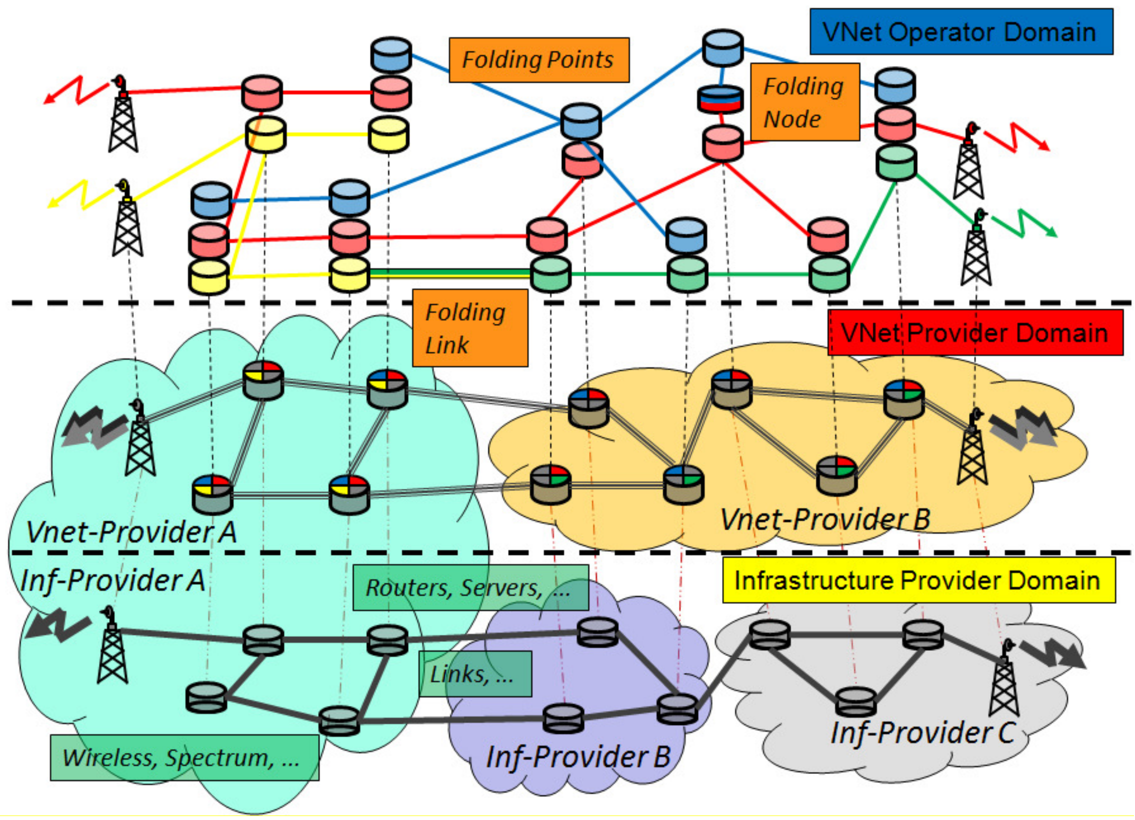



Figure 3.5. Network virtualisation environment (extracted from [4WAR10]).

A relevant aspect on 4WARD's approach is the adaptive virtual resource provisioning to maintain VNets, allocated initially on demand, in response to a VNet creation request. A distributed fault-tolerant embedding algorithm, which relies on substrate node agents to cope with failures and severe performance degradation, is proposed [HLZP10]. Other architectures for network virtualisation has been proposed, e.g., [HZLL08] and [LCBG12]. The target of considering multiple heterogeneous technologies together in an integrated environment, as considered in 4WARD, is pursuit in this thesis, within the scope of mobile and wireless technologies.

## 3.2 Wireless Virtualisation

### 3.2.1 Resource Virtualisation Techniques

Wireless resources, like spectrum, MNs and wireless infrastructure, introduce some new challenges to resource virtualisation, compared to wired ones, due the specific characteristics of the wireless medium. On the one hand, the isolation of the experiments cannot be guaranteed, due to the scarcity of spectrum, which cannot be over provisioned, and on the other hand, signal propagation is a very node-specific property, difficult to control, and with a significant impact on most experiments [SaSr06].

The main access techniques that have been discussed for the virtualisation of wireless networks are presented in [SaSr06], which include FDMA, TDMA, CDMA, a combination of FDMA and TDMA, and Frequency Hopping (FH). The application of each one depends of the RRU of each technology.

FDMA refers to switching channels, or using multiple cards on different frequency partitions, within a physical node to emulate multiple VNodes, Figure 3.6. This technique can only be applied in wireless resources that have frequency as one of the RRUs, such as WLANs. The main problem related to this technique is that the switching from one VNode to another is not instantaneous, and when multiple VNodes are implemented in one physical, each VNode will get their turn to transmit, if a round robin scheduler is used, only after the sum of the switching time with the active time for each VNode multiplied by the number of VNodes. When multiple cards are used, some co-channel interference might affect transmission. The scalability is limited by the number of available orthogonal frequencies, the mix of radio nodes from different technologies, the switching time between frequencies, the active time for each VNode, the co-channel interference if a multiple card-based approach is used, and the mix of applications with varying throughput/latency needs.

TDMA refers to switching time-slots within a physical node to emulate multiple VNodes, Figure 3.7, different users getting a given frequency partition in different time-slots. In this case, there is a finite context switching time, hence, VNodes will get their turn to transmit in a round robin manner, the waiting time for the next turn to transmit being given by the sum of the context switching time with the active time for each VNode multiplied by the number of VNodes. The scalability is limited by the switching time between time-slices, the channel acquisition time assuming existing MAC, the active time for each VNode, and the mix of applications with varying throughput/latency needs.

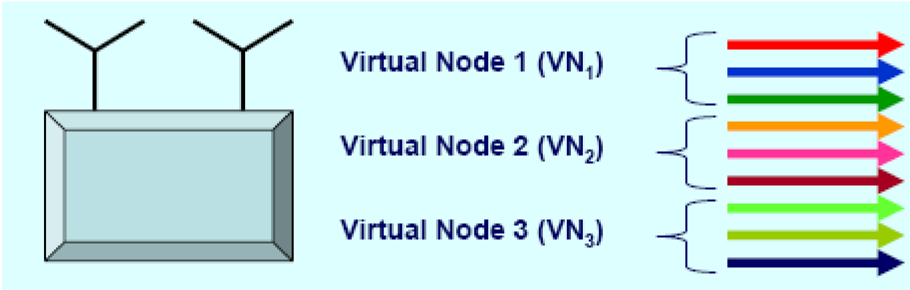



Figure 3.6. FDMA based virtualisation (extracted from [SaSr06]).

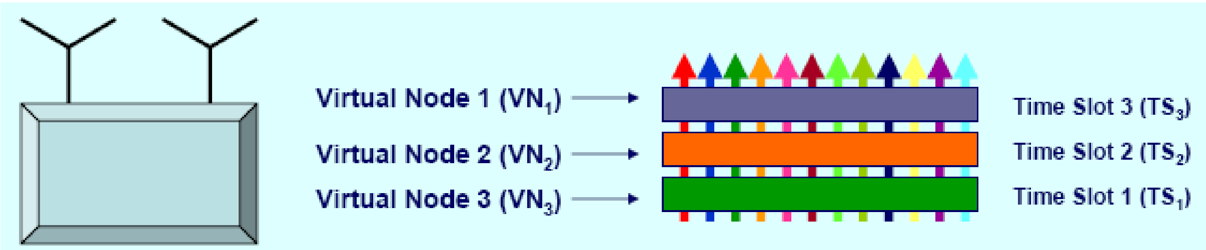



Figure 3.7. TDMA based virtualisation (extracted from [SaSr06]).

In combined FDMA and TDMA, Figure 3.8, a VNode is identified by a unique combination of Frequency Partition and Time-slot. The switching time consists of the channel switching time caused by frequency switching, and the context switching time from one process (running on a VNode) to another due to the time division multiplexing. Therefore, VNodes get their turn to transmit in a round robin manner over a cycle time of the product between the number of VNodes, and the sum of the context switching multiplied by the number of time-slots, and the frequency switching time. The duration over which a VNode is active can be pre-configured and be made known to every node in the system. This approach can only be applied to wireless technologies where frequency is one of the RRUs. Scalability is limited by the number of available orthogonal frequencies, the switching time between frequencies, the switching time between time-slices, the active time for each VNode, and the mix of applications with varying throughput/latency needs.

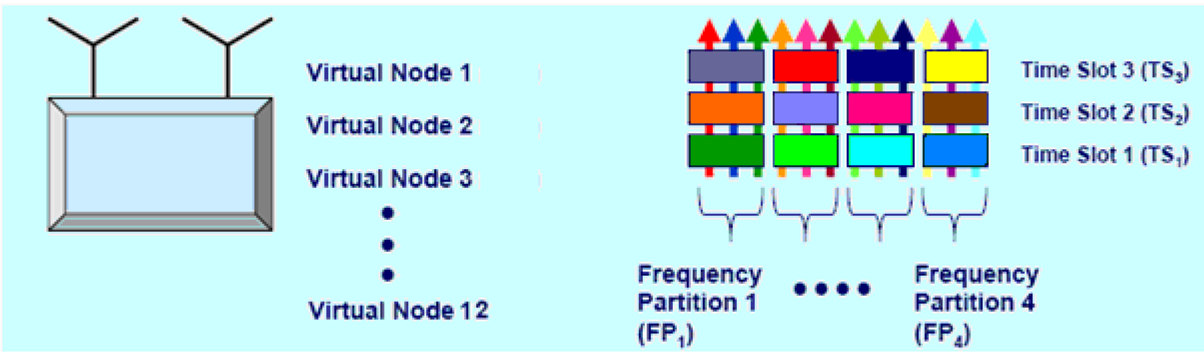



Figure 3.8. Combined FDMA and TDMA based virtualisation (extracted from [SaSr06]).

FH refers to partitioning of a node by allocating a unique sequence of frequency and time-slots to a VNode, Figure 3.9. This approach is very similar to the combined FDMA and TDMA, with the difference that in this case a pre-defined sequence of frequency and time-slots is allocated per VNode, instead of a unique pair frequency/time-slot. Scalability is similar to the FDMA and TDMA combination. As in previous cases, in which frequency is partitioned and allocated to VNodes, FH can only be used for wireless technologies where frequency is a RRU.

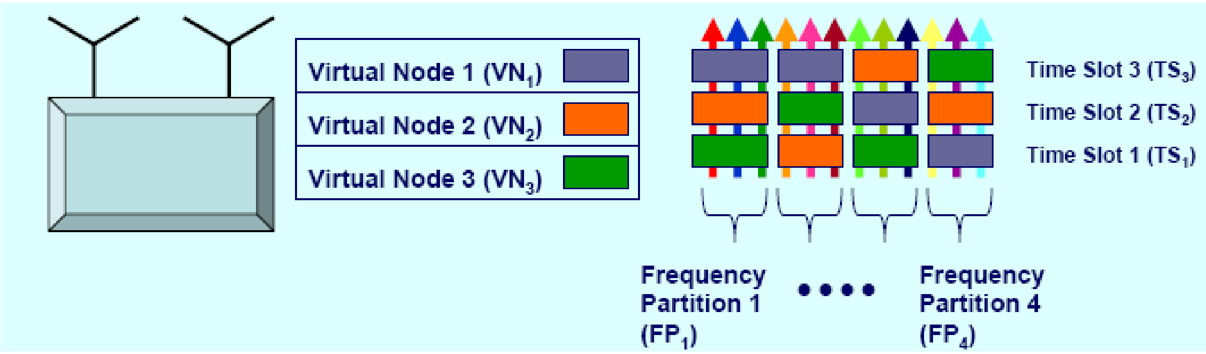



Figure 3.9. FH based virtualisation (extracted from [SaSr06]).

CDMA refers to switching codes within a physical node to emulate multiple VNodes, Figure 3.10. This type of wireless virtualisation can be applied to wireless resources in which the RRU is the code, such as in UMTS. Each VNode has a set of orthogonal codes assigned, allowing for the simultaneous use of the physical node by multiple VNodes. A hybrid UMTS/Wi-Fi network is shown in Figure 3.10. According to the different virtualisation techniques presented above, a slice composed of a set of orthogonal codes between the UMTS BS Router and GW and a frequency partition between GW and end-user device might be created.

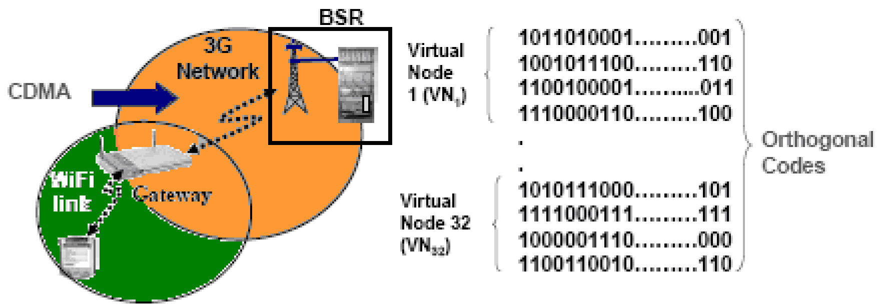



Figure 3.10. CDMA based virtualisation in the 3G network (extracted from [SaSr06]).

### 3.2.2 Slicing Techniques

The slicing process in wireless networks, i.e., the allocation of a coherent subset (a slice) of physical resources to a specific experiment or VNet, has also some specific issues derived from the characteristics of the medium. Firstly, mobility and handover, and emerging applications, such as vehicular networks, have stringent latency requirements that are very difficult to achieve with an abstraction of a VNode in a shared wireless network. Furthermore, the topology concept is very important in wireless networks, because of frequency planning and the differences between nodes. In fact, VNets with wireless components are very much topology-dependent, and this dependence must be incorporated into the slice allocated to a VNet. Finally, when two or more slices coexist in the same hardware, a coherence requirement needs to be addressed, because in order to establish the wireless link a transmitter-receiver pair has to be configured to the same channel parameters. This means that when a transmitter of one slice is active, all of the corresponding receivers and potential sources of interference, as defined by the slicing process, should be simultaneously active on their appropriate channels of operation [SaSr06].

Slicing techniques for wireless networks are mainly based on a combination of the wireless virtualisation techniques enumerated above, since a slice is an association of several VNodes with same specific requirements. The only exception is the Space Division Multiple Access (SDMA) that appears in slicing techniques and is not presented yet, because SDMA only makes sense if a spatial grouping of the wireless resources is to be addressed in the creation of an experiment or VNet, not in the virtualisation of the wireless resources themselves. In SDMA, a full node is allocated to a given user, and no virtualisation of the wireless resource is done. The partitioning of the total wireless nodes is done using "spatial" separation, so that nodes within a given partition do not interfere with nodes in another partition, Figure 3.11. SDMA enables the simultaneous use of a wireless network by partitioning the network and allowing each VNet to run its experiments in its assigned set of nodes (which is its "slice"). SDMA partitioning is limited by the physical span for a given transmit power. When the nodes are in the physical proximity of each other, an "artificial stretching" technique is used to logically, and not physically, stretch space; consisting of controlling the transmit power and using "noise injection" in order to put up artificial barriers among the nodes belonging to different partitions.

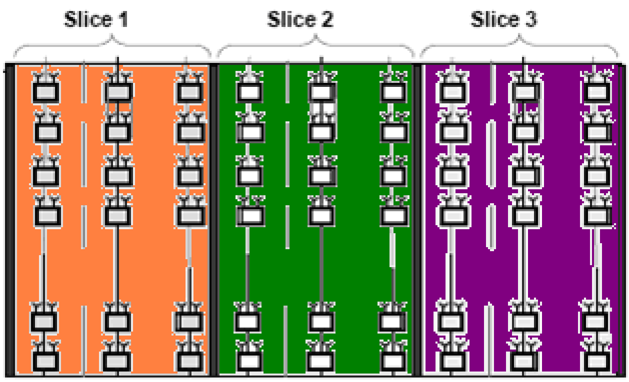



Figure 3.11. SDMA based slicing (extracted from [SaSr06]).

SDMA can be combined with TDMA, FDMA, or both, the additional dimension to the problem being added. The combination of SDMA, FDMA and TDMA nodes is presented as an example of these combinations, Figure 3.12. The partitions using "spatial" separation are further partitioned in the frequency domain, by creating "frequency partitions" which in turn are partitioned in the time domain by creating "time-slots". This kind of technique can only be used in networks where frequency switching is available and can accomplish the requirement of stringent time synchronisation.

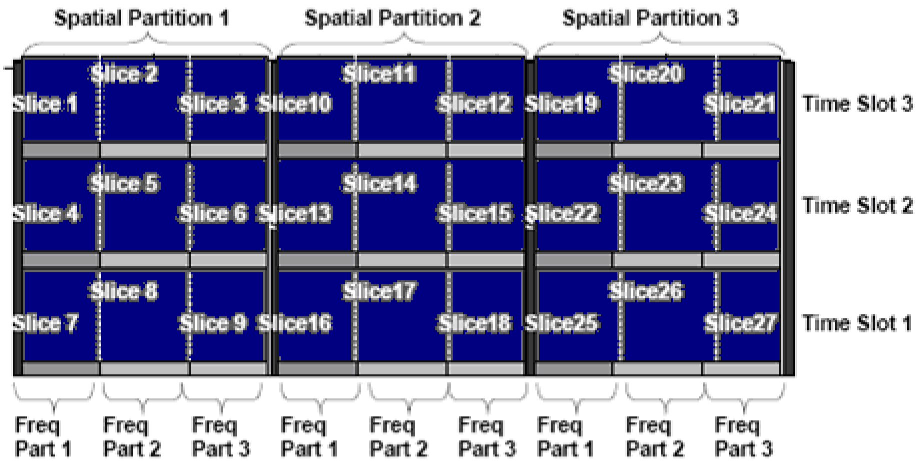



Figure 3.12. Combined SDMA, FDMA and TDMA based Slicing (extracted from [SaSr06]).

### 3.2.3 Virtual Radio

An interesting proposal for wireless resources virtualisation is the framework designated by Virtual Radio as presented in [SaBa08]. Based on the background from network virtualisation, the concept is extended into the wireless domain, being denoted as radio virtualisation. Radio virtualisation is the process of sharing and allocating resources belonging to a physical radio link (i.e., a radio resource).

This approach is based on centralised spectrum usage coordination for different radio systems, where the different radio systems are realised as VNets on a commonly shared physical network infrastructure. The access to the transmission resources are managed according to a multiple access scheme, like CDMA, TDMA, or FDMA. Several multiple access schemes can also be combined. In order to avoid interference among transmissions in different virtual radio networks, the access of the different virtual radio nodes to the radio resources is coordinated by a common Resource Allocation Control (RAC) function, Figure 3.13. This central coordination function provides an efficient usage of resources with low overhead and without contention, and avoids interference and collisions among the different virtual radio networks, providing a high level of predictability of the resources available to each virtual radio node.

A complementary aspect is the needed configuration of radio networks in order to establish the specific functionality of a virtual radio network in a physical node. It is mentioned that the only theoretical restriction in configurability is given by the coordinated sharing of the physical resources, and the proposed RAC function can fulfil this requirement. In fact, it provides the use of a common structure of radio resource partitioning for all virtual radio nodes, and a coordinated access to radio resource blocks. Apart from the data-plane, also control functions are pointed as configurable per virtual radio node, e.g., local routing and mobility management (including mesh and ad-hoc routing, mobility management optimisation, and context transfer), RRM and scheduling (within the virtual radio), cross-layer design and optimisation, authentication and authorisation schemes, as well as battery-saving schemes, like discontinuous transmission/reception and sleep modes.

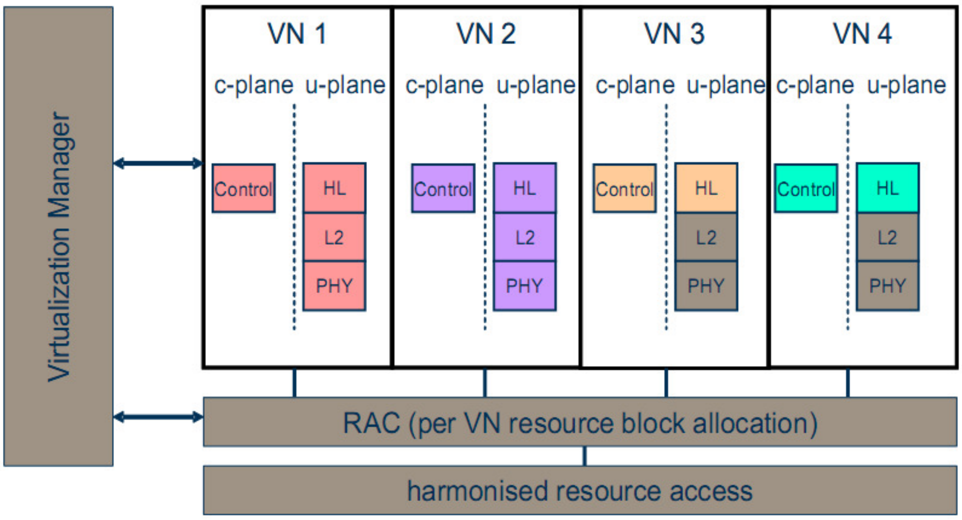



Figure 3.13. VNodes sharing the radio resources of the physical node (extracted from [SaBa08]).

Some ways for the implementation of these configurable functions are also discussed in this work. A virtual radio network is defined as a VNet in an edge network of the Future Internet that comprises multiple inter-connected virtual radio nodes. The virtual radio nodes are VNodes, i.e., virtual instantiations of node functionality running on a physical network node, with their own transmission procedures and protocols, the virtual radios. As an example, a physical radio node can then be part of different virtual radio networks, Figure 3.14. According to the authors, radio virtualisation provides flexibility in the design and deployment of new wireless networking concepts, allowing customisation of radio networks for dedicated networking services at reduced deployment costs.

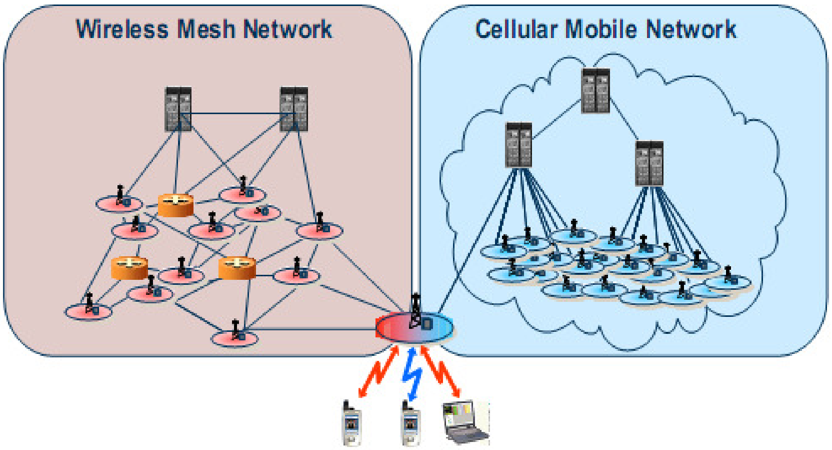



Figure 3.14. Physical radio node being part of two virtual radio networks (extracted from [SaBa08]).

## 3.2.4 Wireless Virtualisation for Wi-Fi, WiMax and LTE

Wireless virtualisation for specific wireless and mobile networks has been recently addressed in literature. In this subsection, recent work is briefly presented, in order to give an overview of the proposed algorithms for wireless virtualisation in Wi-Fi, WiMax and LTE networks.

In [ZLGT10], the authors have chosen LTE as a case study to extend network virtualisation into the wireless area. Following the principle used by Xen [Will07], a well-known computer virtualisation software, their proposal is to add a hypervisor to the eNodeB in order to perform the scheduling of physical resources onto the virtual ones, Figure 3.15. The LTE Hypervisor is responsible for virtualising the eNodeB into a number of virtual eNodeBs, each one used by a different virtual operator, physical resources being scheduled among the different virtual instances via the hypervisor. In addition, the LTE hypervisor is also responsible for scheduling the air interface resources (i.e., OFDMA sub-carriers) between virtual eNodeBs.

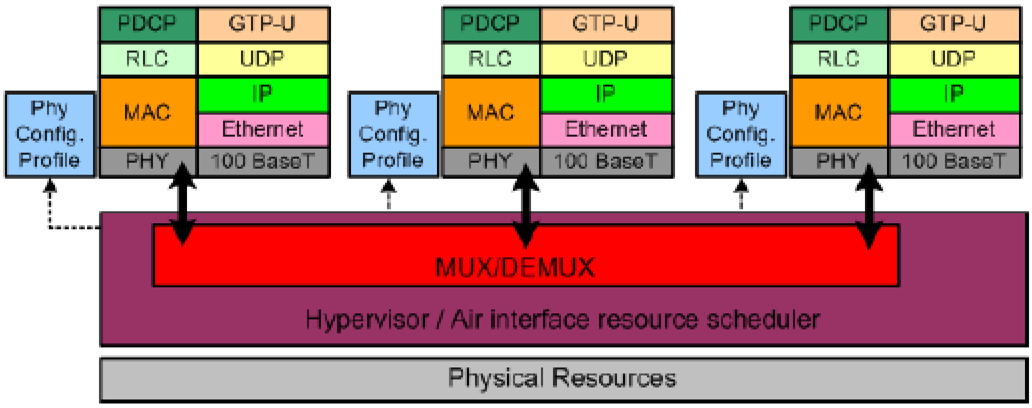



Figure 3.15. Virtualised LTE eNodeB protocol stack (extracted from [ZLGT10]).

The proposed hypervisor algorithm divides the spectrum among the virtual operators based on predefined contracts that the virtual operators have made with the InP. Four types of contracts have been defined: *Fixed guarantees* - the virtual operator requests a fixed bandwidth that would be allocated all the time, whether it is used or not; *Dynamic guarantees* - the virtual operator requests a guaranteed maximum bandwidth that is allocated if required, otherwise only the actual needed one is allocated; *BE with min guarantees* - the virtual operator specifies a minimum guaranteed bandwidth that is allocated at all time, the allocation being done in a BE manner; *BE with no guarantees* - the virtual operator is allocated only part of the bandwidth, if the current load allows.

It is worthwhile noting that the operator contract is expressed in terms of the number of Physical Resource Blocks (PRBs), each operator being responsible for the estimation of the PRBs needed for a time interval. By sharing the air interface resources among the virtual operators based on their contracts, and the traffic load, the overall resources utilisation is enhanced and the performance of both network and end-user is improved.

Concerning WiMAX, the challenges for virtualisation of resources in a cellular BS is addressed by presenting an architecture and performance evaluation of a virtualised wide-area cellular wireless network [BSMR10]. The main purpose is to enable the shared use of BS resources by multiple independent slice users (experimenters or MVNOs), each with possibly distinct flow types and network layer protocols. The proposed virtual BS architecture, Figure 3.16, is based on an external substrate, which uses a layer-2 switched data path, and an arbitrated control path to the WiMAX BS.

This architecture is capable of supporting multiple virtual BS substrates, which could be either local or remotely located. The framework implements virtualisation of BS's radio resources to achieve isolation among multiple VNets. An algorithm for weighted fair sharing among multiple slices based on an airtime fairness metric has been implemented for the first release. Preliminary experimental results from the virtual BS prototype are given, demonstrating mobile network performance, isolation across slices with different flow types, and custom flow scheduling capabilities.

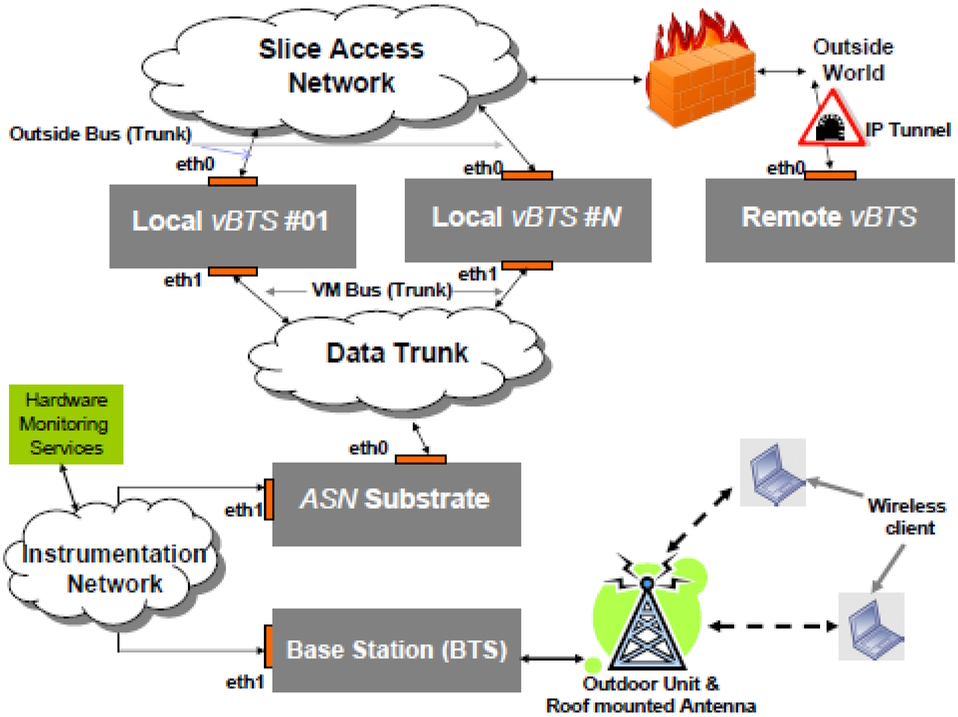



Figure 3.16. Generic architecture for WiMAX deployment (extracted from [BSMR10]).

Other approaches related to WiMAX virtualisation have been recently proposed, [LYLZ12] and [KMZR12]. In the former, an elastic resource allocation algorithm enabling wireless network virtualisation, aiming at achieving VNet isolation and resource efficiency, is proposed. Although the motivation and algorithm design are based on WiMAX, the principle and algorithmic essence are also applicable to other OFDM access-based networks. The focus of this work is to streamline the MVNO business model in order to allow them to pay just what they use. The scenario considers local and foreign virtual mobile networks, where the local is the owner of the infrastructure. The target for local networks is to use as much bandwidth as possible after the foreign traffic has been served satisfactorily. The presented algorithm involves firstly the virtualisation of the physical wireless network into multiple slices, each representing a VNet, and secondly the allocation of the physical resources within each VNet. In the latter, the design and implementation of a Network Virtualisation Substrate (NVS) for effective virtualisation of wireless resources in cellular networks is presented. NVS introduces a slice scheduler that allows the existence of slices with bandwidth-based and resource-based reservations, simultaneously, and includes a generic framework for efficiently enabling customised flow scheduling within the BS on a per-slice basis. Through a prototype implementation on a testbed, the authors show that different flow schedulers in different slices can run for both DL and UL, different slices simultaneously run with different types of reservations, and perform slice-specific application optimisations for providing customised services.

The SplitAP architecture to address the problem of sharing UL airtime across groups of users by extending the idea of network virtualisation is proposed in [BVSR10]. The proposed architecture, allows deploying different algorithms for enforcing airtime fairness across client groups. In this study, the authors highlight the design features of the SplitAP architecture, and present results from evaluation on a prototype deployed with two algorithms, Linear Proportional Feedback Control (LPFC) and LPFC+, for controlling group fairness. Performance comparisons on the ORBIT testbed [Orbi12] show that the proposed algorithms are capable of providing group air-time fairness across wireless clients irrespective of the network volume, and traffic type.

Another proposal for WLAN virtualisation is presented in [XKYG11]. With the proposed solution, named virtual Wi-Fi, the full WLAN functionalities are supported inside virtual machines, each one establishing its own connection with self-supplied credentials, and multiple separate WLAN connections are supported through one physical WLAN network interface. Results, based on a designed and implemented prototype, show that with conventional virtualisation overhead mitigation mechanisms, the proposed approach can support fully functional wireless functions inside the virtual machine, and achieve close to native performance of WLAN with moderately increased CPU usage.

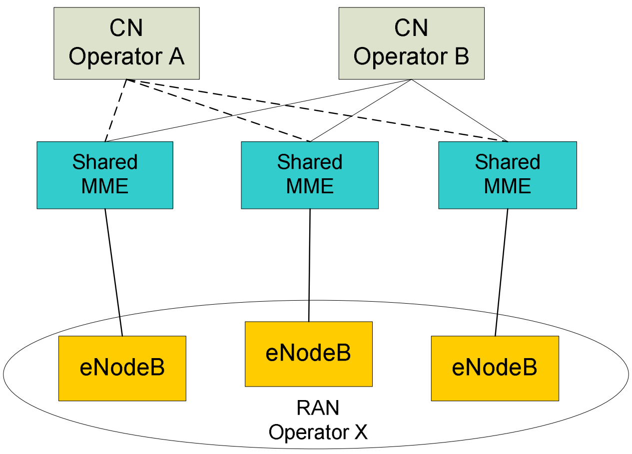
The several techniques for wireless resources virtualisation and slicing of wireless networks presented in this section intend to provide the fundamentals of wireless network virtualisation. Furthermore, the state of the art in wireless virtualisation allows saying that most of the work proposed in literature are related to the wireless node virtualisation and focused on one specific wireless network. Although sharing radio resources among virtual nodes is considered in some of the proposals, it is confined to one RAT, the concept of providing VNet requirements based on the differentiation among them, as infrastructure users, never being explored.

## 3.3 Radio Access Network Sharing

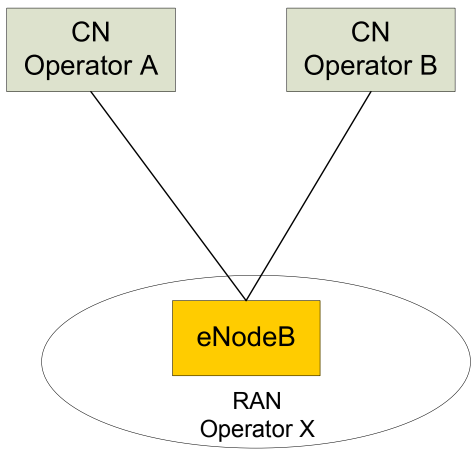
RAN sharing has become an important issue for 3G and beyond operators. Sharing network infrastructure amongst operators offers an alternative solution to reduce the investment in the coverage phase, allows increased coverage, reduces time to market, and allows earlier user acceptance for its related services. MVNOs typically do not have their own infrastructure, rather making use of operators' infrastructure, and treating it as a commodity to offer added-value services. The goal of offering these services is to differentiate from the incumbent operator, allowing for customer acquisition and preventing the MVNO from competing on the basis of price alone. Some MVNOs are actually deploying their own MSCs and even Service Control Points, providing advanced and differentiated services based on the exploitation of their own intelligent network infrastructure.

Several RRM strategies for 3G multi-operator networks have been proposed in the literature, since there is a critical need for radio resources control among multiple operators. Different solutions for how radio resources may be allocated to sharing operators in a roaming based multi-operator UMTS network are discussed in [JoKS04]: a particular method based on RRM with non-preemptive priority queuing in the admission control is presented in detail, providing an attractive trade-off between fairness and total system capacity. Al-Jarbou, [AlBa05], studied the effect of heavy data traffic like Web and File Transfer Protocol (FTP) on the shared network, and how radio resources with roaming based mechanisms may be allocated to the sharing operators; a mechanism based on RRM with preemptive priority queuing in admission control was presented. An RRM strategy proposal, known as adaptive partitioning with borrowing, to cope with the architectural changes introduced by MVNOs is presented in [AMSE06]; according to simulation results, the proposed resource allocation strategy provides higher resource utilisation under load conditions leading to increased revenue. A model for cooperative resources allocation game in shared networks and a set of bargaining solutions based on the concept of preference functions, which depends on the weight the players place on their own gain and the losses of others, is presented in [HeWh06].

Network Sharing in LTE is standardised in [3GPP13a], with two architectures to be: the Gateway Core Network configuration, Figure 3.17(a), in which, besides shared RAN nodes, the core network operators also share core network nodes; and the Multi-Operator Core Network, Figure 3.17(b), in which multiple core network nodes, operated by different operators, are connected to the same radio network controller.



(a) Gateway Core Network



(b) Multi-Operator Core Network

Figure 3.17. Proposed architectures for LTE Network Sharing (based on [3GPP13a]).

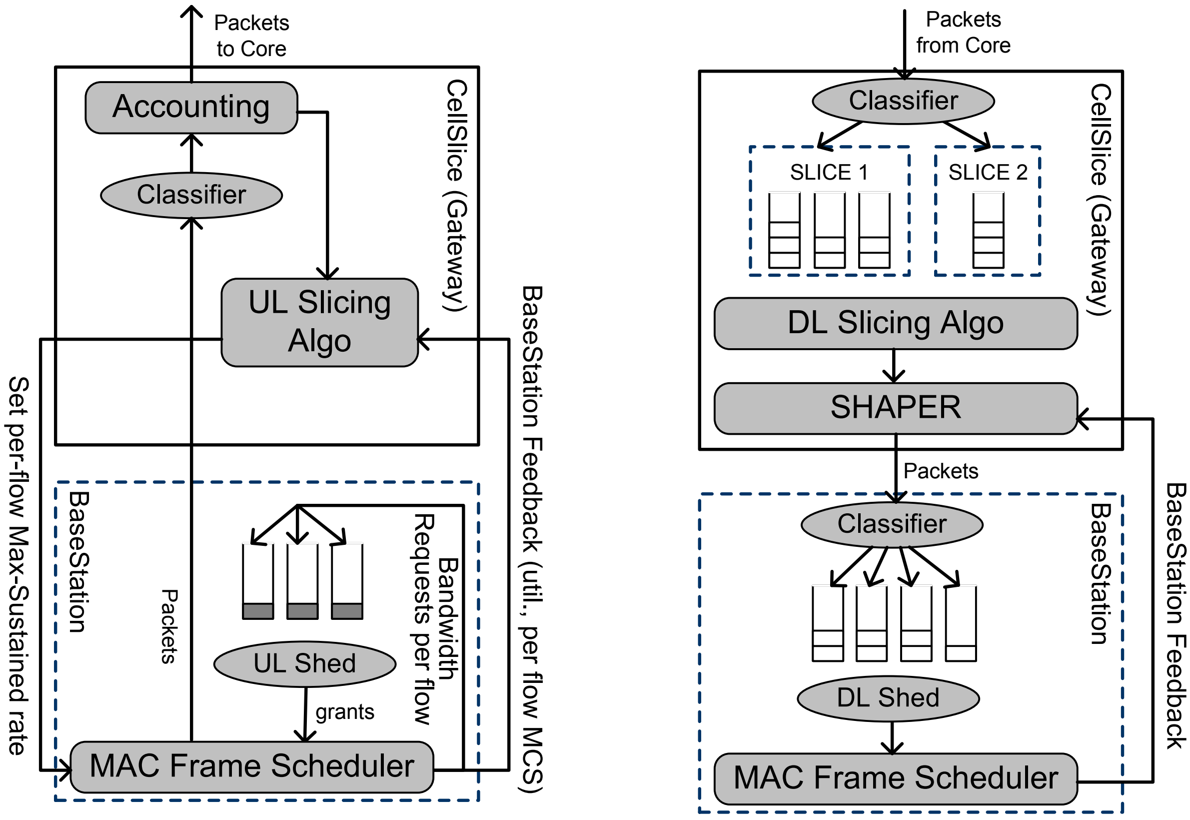
Although RAN is shared in both architectures, the standard does not specify how capacity is shared among the several core networks operators competing for radio access. Besides national roaming, in which a standard roaming agreement is established among operators, or passive sharing, where only the sites are shared, an active sharing is needed to support those network sharing architectures. In active sharing, the LTE evolved UTRAN (eUTRAN) is common to several operators and shared among them. Several core network operators are connected to the common shared eUTRAN.

One proposal for eUTRAN sharing is made in [Alca12], taking spectrum usage, QoS, and capacity sharing into account, among other aspects. Two strategies are proposed for spectrum usage, dedicated and shared spectrum per operator. Sharing spectrum is more efficient, as it does not create a strict split of the radio resources among operators. Strict split means that if the subscribers of one operator are using its whole bandwidth, then no additional subscribers of this operator can enter the network in this cell even if there is still bandwidth available from another operator.

Concerning QoS, an end-to-end model referring several mechanisms used to control it within the shared eUTRAN are presented, being distributed among different segments of the network. At the eNodeB level, there is Call Admission Control, Policing per radio bearers, Traffic shaping per operator, and Marking based on QCI specified at radio bearer establishment. At the eUTRAN edge router, IP QoS features can be used to perform policing and shaping at aggregate level, to control the amount of traffic coming from each core network operator in DL. Within the transport network between the eNodeB and the eUTRAN edge router, the transport network will support QoS to provide the correct priority to IP packets or Ethernet frames marked by the eUTRAN edge router or the eNodeB. Finally, regarding the sharing of capacity, several strategies are proposed at the eNodeB level, ranging from “fully pooled” to “fully split”. In the former, there are no resources reserved per operator, hence, a fair access to resources for each operator cannot be guaranteed. The latter allows for a strict reservation of resources per operator, which may lead to an inefficient use of the available resources, denying service to end-users when some resources are still available. In between, there are the “partial reservation” and “unbalanced” strategies, where operators have a partial amount of resources reserved, the remaining being shared among all.

The main drawbacks of this proposal are the direct mapping onto the amount of radio resources reservation for capacity provisioning, and the static configuration of these strategies. In fact, the capacity provided by the reserved radio resources may vary due to the wireless medium variability, and if this amount is not dynamically adapted to the network state, the capacity contracted by the operator may not be provided. On the other hand, the strategies for RAN sharing are configured at the network management level and per eNodeB, as well as the admission control parameters, which may involve a great effort, e.g., when a new operator wants to enter the business, or an existing operator wants to change the contracted capacity, denoting some kind of inflexibility.

A novel system for slicing wireless resources in a cellular network for effective RAN sharing is proposed in [KMZR13]. CellSlice, the system designation, is a gateway-level solution that achieves the slicing without modifying the BSs' MAC schedulers, thereby, significantly reducing the barrier for its adoption, Figure 3.18. According to the authors, CellSlice's design is access-technology independent, hence, being applicable to LTE, LTE-Advanced, and WiMAX networks, among others. The network can work by using resource-based or bandwidth-based reservation strategies. In the former, slices are allowed to make reservations in terms of fraction of resources needed per unit time, though, they can make use of additional resources beyond their reservations to improve end-user experience. In latter, slices reserve bandwidth, the dependency of the effective bandwidth achieved on the channel conditions of users within the slice being avoided. CellSlice was evaluated on a WiMAX prototype and by simulation for larger-scale scenarios. It is concluded that by overriding the scheduling decisions taken by the BSs, in order to impose slice-specific resource allocation, it is possible to achieve the slicing of wireless resources remotely from gateways with simple algorithms in both UL and DL. This work is very similar to the one in [KMZR12], but virtualisation is not applied.




(a) UL Slicing (b) DL Slicing

Figure 3.18. CellSlice (based on [KMZR13]).

Recently, the 3GPP group for RAN Sharing Enhancements identified a set of use cases, and the corresponding requirements, in order to allow a more flexible and efficient RAN sharing [3GPP13b]:

- *RAN Sharing Monitoring* implies that the Hosting RAN provider will be able to provide to the Participating Operators to retrieve Operation, Administration and Management status information at the same level of detail as in non-shared RAN for the share of their resources.
- *Asymmetric RAN Resource Allocation* states that it will be possible to establish each Participating Operator's pre-agreed usage portion of the Hosting RAN. A shared RAN element measures network resource usage at all times, separately for each Participating Operator, and identifies whether the Participating Operator's pre-agreed usage portion of the Hosting RAN is being used. The Hosting RAN will be capable to apportion among Participating Operators reduced resource allocations when QoS objectives cannot be met, due to excessive traffic load, distributed according to Participating Operators pre-agreed usage portion of the Hosting RAN.

- *Dynamic RAN Sharing Enhancements* and *On-Demand Automated Capacity Brokering*, which state that Participating Operator may require varying network capacities during different time periods of the day or the week, and the Hosting RAN Provider might share by automatic means some designated portion of its RAN capacity with other Participating Operators (e.g., MVNOs). Load Balancing in shared RAN highlights the situation of a certain shared coverage area consisting of several cells, which are shared by multiple operators. The agreed shares are predefined among operators. In this case, Load Balancing among these cells needs to take the network sharing ratio per operator into account.

In this section, several proposals for RAN sharing from the initial MVNO deployments to the definition of RAN sharing Enhancements are presented. It can be noted that the trend is to a more flexible share of the RAN, use cases and the set of requirements to address it being shown. However, the integration of these proposals into a virtualised network environment is not envisaged, thus, preventing the possibility of running simultaneously different network protocols over multiple VNets, which are isolated from each other and with independent management functions.

## 3.4 OConS Architectural Framework Overview

A novel architectural framework, designated by OConS, has been developed within the European Project ICT-SAIL, in order to provide enhanced and new connectivity mechanisms that improve end-users experience and operator's network performance, by providing adaptive, flexible, heterogeneous, and multi-protocol solutions to better cope with the dynamics of networks and the continuous evolution of technology [SAIL13]. Connectivity services and architectural framework build on existing Internet foundations, support different transport paradigms, and provide a unified and abstract access to connectivity services, on demand, based on the proposed orchestration functionalities. OConS offers an open architecture for connectivity services, which provides a flexible framework, supporting both legacy and enhanced connectivity mechanisms. It is able to dynamically adapt the operation of the involved mechanisms according to the particular requirements of the services and applications. Generally speaking, OConS is a control framework that provides the capability to orchestrate a set of connectivity services, running on one or more interconnected nodes.

A brief overview of the OConS architecture is presented in this section, with its key components. An OConS connectivity service is formed by a specific combination of OConS connectivity mechanisms. In order to make the design of new mechanisms easy, to be able to "compose" them together and to share and reuse their functionalities, OConS mechanisms are modelled following a mechanism-level architecture that decomposes them into information monitoring (i.e., Information Element (IE)), decision making (i.e., Decision Element (DE)), and execution and enforcement (i.e., Enforcement Element (EE)) functional entities. The abstractions of the functional entities are independent from any layer or protocol. By having a common way of representing current and future mechanisms, they ease the instantiation, launch and interconnection of mechanisms through clearly defined interfaces, forming OConS connectivity services at link, network and flow levels. Furthermore, these basic functional bricks facilitate sharing and reusability whenever possible, i.e., the information and measurements collection, the decision, and the execution entities.

A representation of the OConS functional architecture is presented in Figure 3.19. The Service Orchestration Process (SOP), at the centre of the OConS functional architecture, is capable of orchestrating an OConS service composed of one or several OConS mechanisms. OConS users, i.e., generic applications, communicate with the SOP by means of the Orchestration Service Access Point (OSAP). Through OSAP, users communicate their requests regarding the desired connectivity services, to be set-up by SOP, and receive notifications about the status of the requested connectivity services. In order to store the data of the various mechanisms, rules and policies, as well as the network state, SOP is connected to a database, named the Orchestration Registry (OR). The Intra-/Inter-Node Communication (INC) functionality takes care of exchanging messages among architecture components, both locally and remotely.

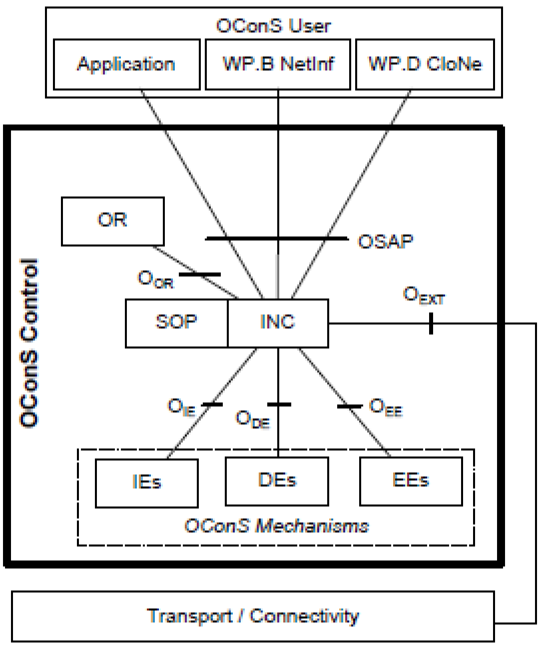



Figure 3.19. OConS functional architecture (extracted from [SuTi12]).

A key functionality of the OConS architecture is Orchestration, being responsible to provide on-demand connectivity-as-a-service. The orchestration function embeds the knowledge of the available networking resources of the Nodes and Links used within a given network, but also of available OConS entities and mechanisms. Orchestration is triggered either explicitly by a user/application or implicitly by monitoring the network state. An application has specific requirements, which are sent to OSAP. The orchestration functionality dynamically identifies and launches, from a set of available mechanisms, the most appropriate ones to answer the specific connectivity requirements. Mechanisms can be distributed over several OConS nodes (e.g., end-terminals, access-routers, per domain controllers, etc.), spanning one or several links or OConS domains. The orchestration applies at several levels, each of them having specific functionalities, as represented in Figure 3.20:

- Orchestration Register - during the bootstrapping and discovery process of local entities, they are registered, so that the orchestration becomes aware of their existence and location.
- Orchestration Monitoring - OConS has the knowledge of available networking resources of the Nodes and Links used within a given network. By means of the IEs, it monitors the network state, implicitly communicating the need of triggering appropriate mechanisms to answer specific adverse situations (e.g., network congestion).
- Entities, Resources and Mechanisms Orchestration - it is responsible for the orchestration among OConS entities, i.e., discovery/bootstrapping/configuration within a node, as well as the allocation and management of OConS entities' resources (processing capabilities, memory, etc.). Upon request from the corresponding connectivity requirements, it identifies, from the Orchestration Register, the most appropriate mechanisms (legacy or OConS ones) that need to be launched. This is communicated to the Link/Flow/Network Orchestration functionalities, responsible for the launch and management of particular mechanisms;
- Link, Network and Flow Connectivity Services Orchestration - it is in charge of the instantiation, composition, and launching of OConS mechanisms (within a single or multiple nodes), and their later control and management, supported by specific OConS signalling.

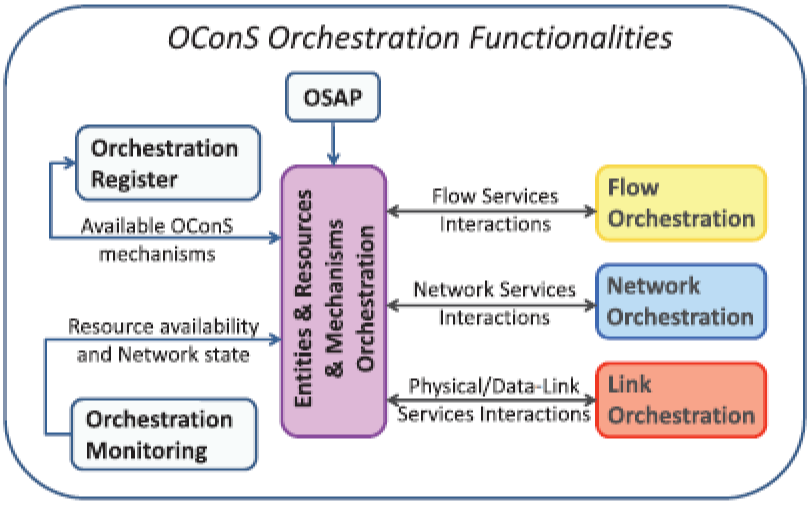



Figure 3.20. OConS Orchestration Functionality [SAIL13].

In brief, the presented OConS architecture is a novel framework for connectivity services, in which some software defined network concepts can be found. In fact, the orchestration entity acts according to a programmable set of rules, allowing to manage several connectivity mechanisms, deployed at diverse layers and possibly decoupled from the physical network nodes. This set of mechanisms can cooperate among them, as an answer to a service request and a given network state, in accordance to pre-defined orchestration rules. Moreover, the modular description of the mechanisms allows modelling any connectivity decision algorithm, which can take advantage of this flexible operational environment. The proposed models have been modelled according to this framework, to benefit from its functionality.

# Chapter 4

# Models and Algorithms

Novel models and algorithms, to manage radio resources in virtualised environments, are proposed in this chapter. The approach for RRM in virtualised environments, and the characterisation of the proposed Cooperative VNet RRM (CVRRM), are presented. Furthermore, the proposed network architecture, the main assumptions and inputs, the analytical model, the strategies and algorithms for virtual radio resources allocation algorithms, the initial VNet selection and the VNet handover support, and the metrics for evaluation are presented. The modelling of the OnDemand-VRRA algorithm for integration on the OConS Architecture is also done in this chapter.

## 4.1 Network Architecture

The proposed network architecture refers to the virtualisation of the wireless access as part of VNets, being based on the generic network virtualisation environment presented in Figure 3.1. Hence, the considered network environment envisages the existence of multiple VNets created by a VNet Enabler, which can be a VNP and/or VNOs. SPs use these VNets, settled on demand to satisfy their service requirements, in order to deliver services to their customers. This way, the physical infrastructure owned by InPs is shared among several VNets, providing services with different requirements, and to multiple SPs. The physical view of the proposed network architecture is depicted in Figure 4.1.

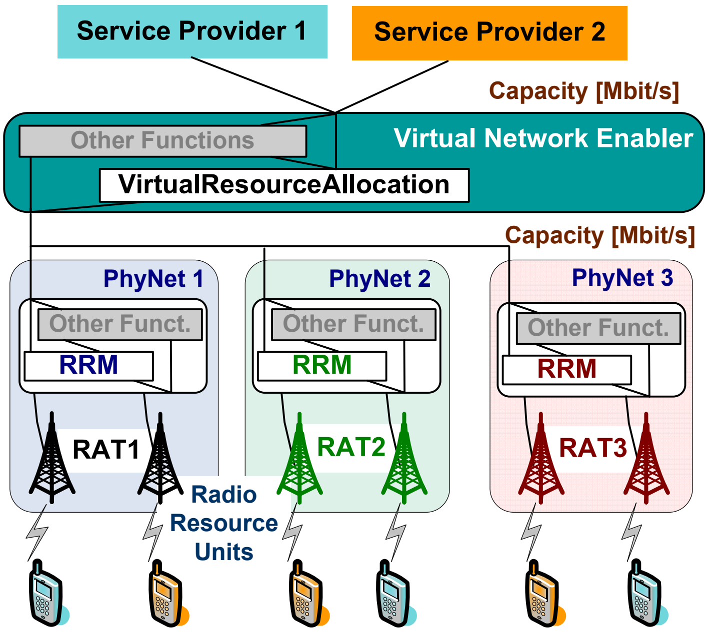



Figure 4.1. Physical Network Architecture.

The following elements are represented in Figure 4.1:

- Service Providers - entity that provides/delivers services/applications over the (virtual) network for a set of clients.
- Virtual Network Enabler - entity enabling network virtualisation. Several functions are under the responsibility of this entity: receiving and processing the requirements for virtualisation; negotiation with InPs, to use their physical resources in order to provide capacity; VNets' creation and delegation. Although it can be an external role, the operation and maintenance of VNets is considered within this block for simplicity. Virtual Resources Allocation, besides other type of resources allocation, e.g., computational ones, includes the mechanism that

manages the allocation of radio resources from the physical networks to the virtual resources created by the virtualisation process;

- Physical Networks (PhyNets) - set of physical resources of each RAT, e.g., BSs or other nodes of the network architecture, owned by InPs. The physical resources should allow for the instantiation of virtual ones, i.e., should be capable of sharing their physical components. RRM is composed of a set of specific mechanisms from each physical network, performing the well-know RRM functions, e.g., admission control, scheduling, radio resources allocation, and handover, among others.

For simplicity, SPs' requests are illustrated as being just a capacity demand, although their requirements cannot be limited to that. Based on the request for capacity and infrastructure availability, the VNet Enabler defines a VNet adequate to service delivery, performing the Virtual Resource Allocation. Virtual resources composing the VNet are then created on top of the network infrastructure, by sharing the available physical network capacity.

Within the scope of this thesis, the virtual resources deployed over physical infrastructures are designated, from now on, as Virtual BSs (VBSs). It is worth to note that though the instantiation of a VBS involves the virtualisation of processing and memory resources, in order to run the inherent functions of a BS, the details of this instantiation process are not within the scope of this thesis, the main focus being rather the radio part of the VBS, i.e., the set of radio resources allocated to VBSs. In this sense, VBSs are assumed to be implemented on top of a group of BSs from heterogeneous networks serving a given geographic area over which capacity demand is issued; this group of BSs serving a delimited geographical area is designated as a cluster. The requested capacity may be split over one or several VBSs, by the Virtual Resource Allocation function; in the case that several VBSs coexist, a partial capacity requirement is established for each one. VBSs' capacity is then provided by the allocation of RRUs over the several BSs deployed in the cluster; the RRU is the minimum radio unit that can be allocated to an end-user in a physical BS, depending on the RAT, e.g., a time-slot in TDMA or a code in CDMA.

Figure 4.2 depicts the VNOs' view of the network, the logical one, with the following elements:

Virtual Base Stations (VBSs) - virtual resources created to provide the capacity required by an SP over a given geographical area. VBSs capacity is collected from the available radio resources of all the BSs in that area.

- Virtual Resource Management - process that manages the use of VBS's capacity, enabling to perform RM functions, e.g., to adapt the capacity required to the VBSs utilisation.
- Virtual Networks (VNets) - characterised by the type of contract, the amount of required capacity and other kind of requirements, like location and topology. Within the scope of this thesis, virtual resources sharing the physical infrastructure are the VBSs.

VNOs are the players that manage and operate the VNets, including their virtual resources, to satisfy Service Providers' requests. They know only the virtual resources that are part of the VNet with their associated capacity, the set of physical resources being hidden from them.

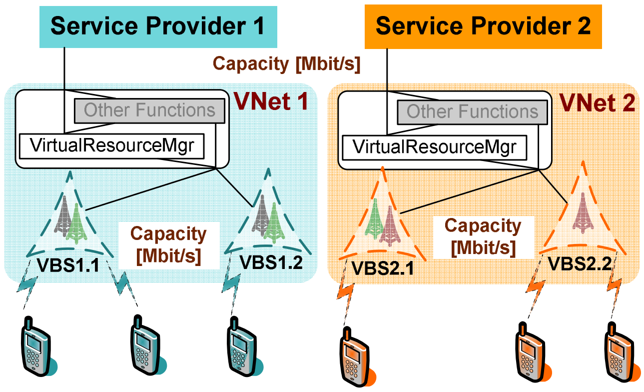


Figure 4.2. Logical Network Architecture.

To make use of a service, the end-user connects physically to the BSs, Figure 4.1, but the connection to the VNet providing the service is made logically via a VBS, through a VLink, Figure 4.2. The physical link is the group of RRUs allocated to the end-user, whereas the VLink is the capacity, in bit/s, allocated from the VBS, Figure 4.3. The mapping between the physical links onto VLinks is essential to compute the VBS aggregated capacity, allowing satisfaction monitoring, and consequently the trustiness between the VNO and the InP.

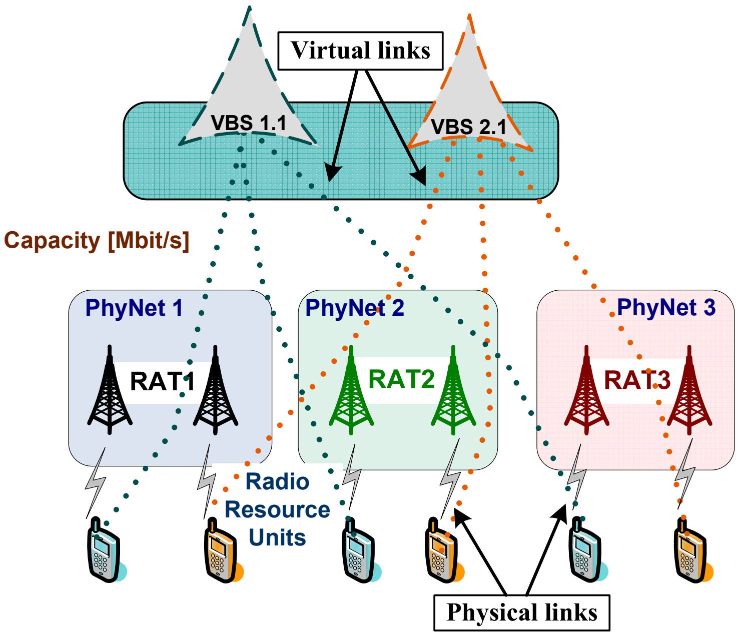


Figure 4.3. Mapping between virtual and physical links.

## 4.2 RRM for Virtualised Environments

### 4.2.1 Relationship between Physical and Virtual Levels

Network virtualisation introduces new concepts that imply various adaptations on the operation of networks. Focusing on wireless and mobile networks, it is important to define the RRM relationship between physical and virtual levels, i.e., which RRM functions must be split or joint, and what scope each one must have. In this section, a framework to deal with it is proposed.

The following considerations, based on VNet concepts and terminology, are worthwhile noting:

- VNets are constituted by VNodes and VLinks that may belong to different technologies, namely, wireless ones;
- VNets have their own QoS requirements, e.g., capacity, which must be specified when the VNet is created;
- physical nodes have an entity that must be aware of the available resources, and the resources occupied by VNodes and VLinks instantiated within the physical resources in terms of bandwidth, processor capacity, memory, etc.;
- when wireless medium changes occur, resources availability must be updated to reflect the changes, e.g., reduction of bandwidth and burst errors.

Concerning the virtualisation of wireless resources, in particular radio resources virtualisation, two levels of RRM functions must be considered, Figure 4.4: Intra- and Inter-VNet ones. The former allows managing how end-users of a VNet share the radio resources of that particular VNet; it is the VNO that can freely define the kind of RRM it uses within its VNet. One can have, e.g., two different VNets using heterogeneous technologies: one uses CoRRM for efficiency, the other does not for simplicity. The latter, designated as Cooperative VNet RRM (CVRRM), is responsible for managing how physical resources are allocated to different VNets. CVRRM ensures that every VNet gets the amount of resources negotiated in the VNet establishment phase. It should be stressed that it does not operate on the resources that are required by an individual end-user; instead, it considers the aggregated resources demand of different VNets; nevertheless, it can be triggered by individual demands that potentially may affect the aggregated ones. In a multi-access analogy, CVRRM and Intra-VNet RRM are equivalent to MRRM or CRRM, with the difference of the operational context.

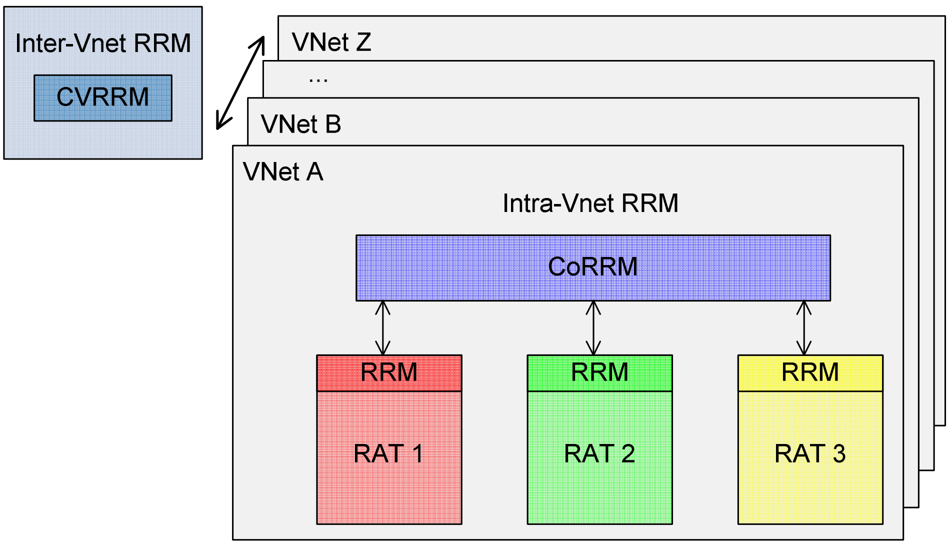



Figure 4.4. Inter-VNet RRM and Intra-VNet RRM.

One of the novelties of this thesis is the development of algorithms within the scope of the second level of RRM referred above, CVRRM, in order to optimise the physical resources utilisation, providing the required capacity to the virtual resources and maintaining isolation among them. Since the CVRRM target is similar to the CoRRM one, the main CoRRM functions, like vertical handover, access selection, and scheduling among RATs is included in the CVRRM set of functionalities, but in a higher level perspective or abstraction view. Thus, the CVRRM functionality is devoted to the characteristics abstraction of heterogeneous environments, from the virtualisation process, keeping the main CoRRM target, i.e., to optimise network resources usage and to provide the always best connectivity, ensuring VNets QoS requirements.

The resources considered in the CVRRM context are the physical nodes and links, and the VNodes and VLinks. Still radio resources, abstracted by RRUs, are central resources within this scope. CVRRM strategies are based on a global knowledge of physical resources, their partial allocation to VNets, the co-located resource mapping, and the fundamental VNets characteristics to which resources are allocated. The VNet "owners" agreements (inter-VNPs, inter-InPs and VNOs) are also important information to be known. CVRRM can be centralised covering a given area or infrastructure provider, and/or located in the physical nodes, depending on the VNet deployment strategy.

Figure 4.5 shows a logical diagram where the relations between CVRRM and other entities are depicted. The relation with the VNet Enabler Management consists of the exchange of information concerning the VNet creation to CVRRM, e.g., VNet requirements and virtual resource allocation, and from CVRRM in the exchange of feedback information about, e.g., VNet sub-utilisation or need of VNet expansion. The thresholds for the referred exchange of information from CVRRM should be configurable at VNet establishment. CVRRM should also exchange information with the Intra-VNet RRM, in order to coordinate the allocation of radio resources to each VNet and receive information about VNet operation.

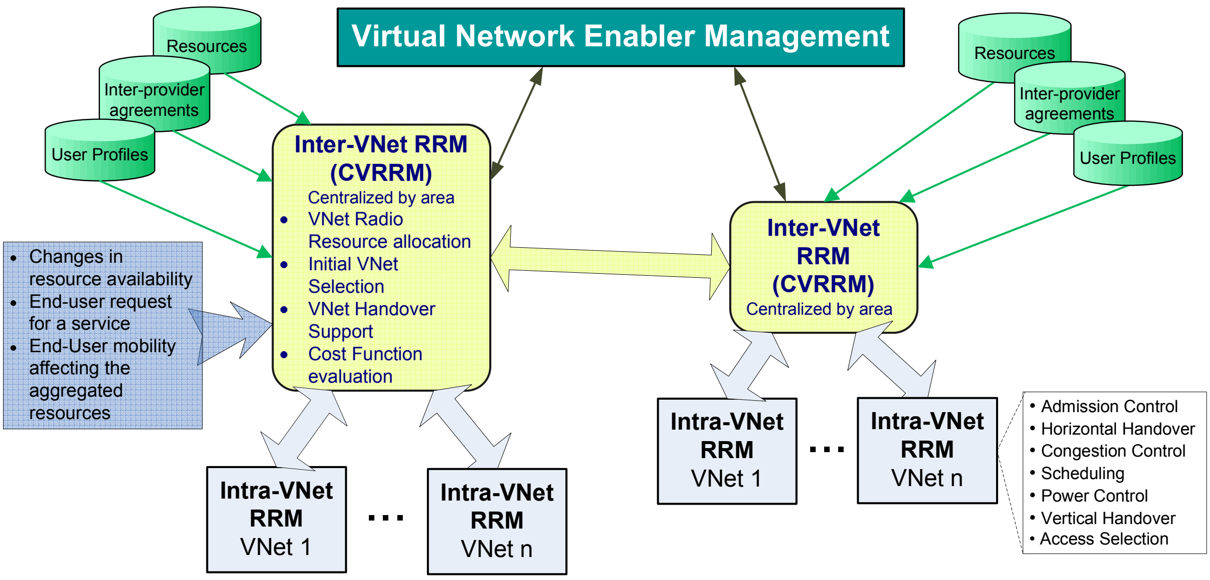



Figure 4.5. CVRRM and relations with other entities.

The access to VNet's global information is always considered, Figure 4.5. Therefore, CVRRM can react not only to the changes in the amount of resources allocated to a VNet, but also to the user requests that affect the aggregated virtual resources. Appropriate monitoring of Key Performance Indicators (KPIs) determines changes in the allocation of radio resources, in order to maintain VNet requirements. These changes are supported by the computation of a Cost Function (CF), which allows a unified comparison among all physical resources in the heterogeneous environment, according to a given management policy. This CF, described in Section 4.6.1, is an abstraction or simplification used to integrate a set of radio resources KPIs into a single one, the cost of a given resource. The obtained CF value reflects network conditions and management strategies in the virtual and heterogeneous network environment.

Based on CoRRM concerns, namely, initial RAT access selection, vertical handover, and resources scheduling/allocation, three CVRRM main functions are identified:

- Virtual Radio Resource Allocation (VRRA) - manages RRUs allocation to different VNets/VBSs, in order to ensure the amount of capacity negotiated at the VNet establishment; it takes the possible changes in capacity/availability of radio resources that affect VNet requirements into account, e.g., data rate and delay.
- Initial VNet Selection (IVS) - allowing transparency to end-users in the process of VNet attachment and optimising VNets utilisation.
- VNet Handover Support (VHOS) - providing the always best connectivity, even when the VNet coverage is impossible, therefore, allowing handover between different VNets.

CVRRM functions are distributed among BSs and central nodes in the network architecture. From the BSs, the VRRA collects KPIs and radio resources utilisation information, interacting with the MAC scheduling for parameters configuration per VBS. Being responsible for the VBS aggregated capacity management, VRRA is also implemented at the cluster level, performing the coordination among all distributed VRRA functionalities within the physical resources. An Access Broker should be considered as a special VNO or SP, in which IVS accesses VNets information and end-users profile, in order to evaluate the best VNet to select. An interface to the VNet Enabler Management is also considered, since a VHO decision can trigger VNet adaptation or extension, ensuring the best connectivity.

Figure 4.6 illustrates CVRRM within the network architecture defined in Section 4.1. The CVRRM coordination (CVRRM-C) role is represented at the VNet Enabler level, in which VBSs are created. It is through this coordination function that CVRRM decides and informs the adapted RRM of each physical network, by the additional CVRRM distributed (CVRRM-d) function, how to share radio resources in order to provide the VBSs' requested capacity. CVRRM-d, represented in the RRM of each physical network, depicts the capability that the physical networks must provide to allow the configuration of the MAC scheduler per VBS, and to report relevant information about the use of the radio resources to the CVRRM coordination.

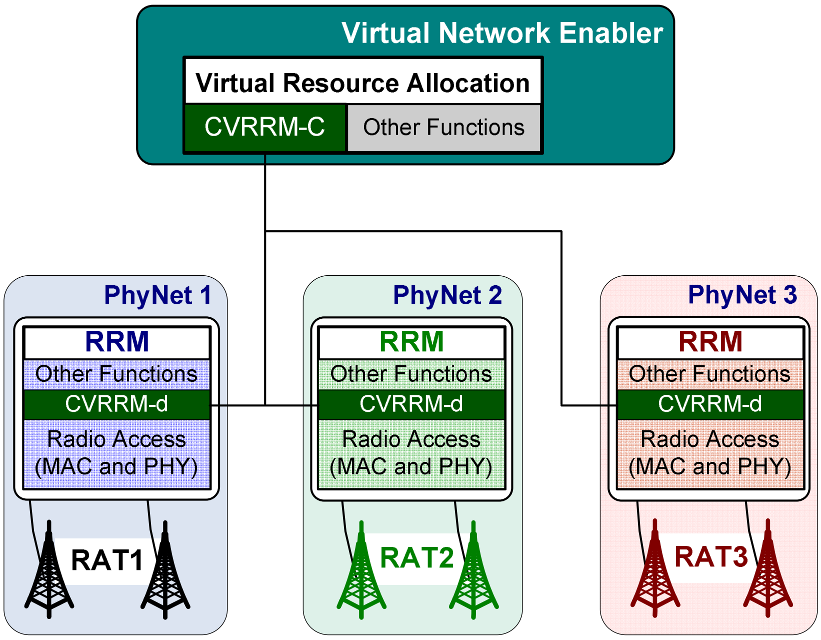



Figure 4.6. CVRRM within the network architecture.

### 4.2.2 CVRRM Functions

VRRA is the function that manages how physical resources are allocated to different VNets. It ensures that every VNet receives the amount of capacity (data rate) negotiated in the VNet establishment phase. The VRRA function compares the updated resource capacity with the VNet requirements, and if it is below a certain threshold (depending on the type of VNet requirements) it must discover, among the co-located BSs (including its own), the extra amount of resources to be allocated in order to guarantee the VNet requirements.

VRRA is developed to perform the mapping between virtual and physical links, dynamically adapting the allocation of RRUs to the network conditions and VNet utilisation. These functionalities are distributed between the virtual resource allocation and the RRM. Since one is dealing with heterogeneous networks, it is implemented at the cluster level, taking all heterogeneous networks in the area into account, and at RRM one, being locally implemented at the BSs. At the cluster level, it manages the aggregated capacity provided to the virtual resource, by sharing the set of available RRUs from all RATs; at the RRM level, it maps the capacity requested to a RAT onto RRUs assigned to end-users, which are restricted by the number of RRUs allocated to all end-users in the VNet.

The first consideration on Initial VNet Selection (IVS) procedure is related to which function or entity should execute this procedure. Users' management is done by Intra-VNet RRM, which is a function controlled by the VNO; however, when a service is started, an attach point must be selected, according mainly to service requirements, the available VNet capacity, and users' preferences and contracts. In an initial phase, one needs to decide which VNet ensures the requirements of the service to be started, from the ones handled by the VNOs with which the SP has contract. According to these hypotheses, it can be concluded that Intra-VNet RRM cannot select the best VNet for a given user/service, since it is a management function within a specific VNet, and it does not have the knowledge of the whole VNet framework. The IVS function must be delegated to an entity that has a global view of the VNets and of the agreements between users and providers/operators or inter-providers/operators. Hence, in order to execute this function, an instance of CVRRM must be located in a “Broker”, which grants access to its VNets’ global information.

The IVS. procedure provides mechanisms to obtain detailed information about different VNets, such as provider information, network protocols, QoS guarantees, security mechanisms, and virtual resources in the end-user vicinity. With this type of VNet information, a set of possible VNets can be identified by comparison with the service requirements and the user contract. By similarly to Ambient Networks, it is called the VNet Valid Set (VVS). After this step, the VNets’ cost, computed according to pre-defined strategies reflected in the CF, is the basis for deciding the most efficient VNet to select.

The VHOS procedure is not a standard handover procedure, since a single user is not directly moved from one attach point to another, but he/she might be moved due to an extension, adaptation, or migration of the VNode where he/she is attached to the network. The VHOS procedure is executed when the Intra-VNet RRM identifies the situation in which the mobility of the user requires extra resources, due to the lack of coverage/capacity of the actual VNet resources. In this situation, the Intra-VNet RRM should trigger the VHOS that is responsible to find, within the end-user’s neighbourhood, the best BS with physical resources available, and request for an extension/adaptation/migration of the VNet (procedure similar to the one described for Adaptive-VRRA). In the case that no physical resources are available, VHOS requests the means to use another VNet with lower load (if the agreements and the characteristics of this VNet allow for it). Additionally, VHOS can be triggered by the “Broker”, when a VNet providing a similar service, but with a better cost value, is identified.

## 4.3 Assumptions and Inputs

### 4.3.1 Main Assumptions

In order to establish a network model, some assumptions are taken: uniform coverage by all wireless systems under analysis and the inexistence of a specific requirement from the VNO related to the wireless technology in use. It is considered that VNOs do not care about the specific wireless technology being used, as long as the contractual requirements are ensured. Moreover, it is assumed that end-user terminals are mobile and capable of supporting different radio interfaces, so that they can connect to any available network.

Concerning the wireless access technologies involved, one considers TDMA/FDMA, CDMA, OFDM, and OFDMA, as they cover most of the current wireless systems (GSM, UMTS, Wi-Fi, and LTE), which from now on are considered as examples of such access technologies. Although, the RRU definition for each access technology is different, a level of abstraction is added, enabling a common approach to manage all radio resources. It is considered that each wireless link is generically composed of RRUs, which vary in number and capacity, according to the technology involved, Figure 4.7. However, the characteristics of each technology are taken into account, in order to emphasise the specific factors that influence RRU data rate.

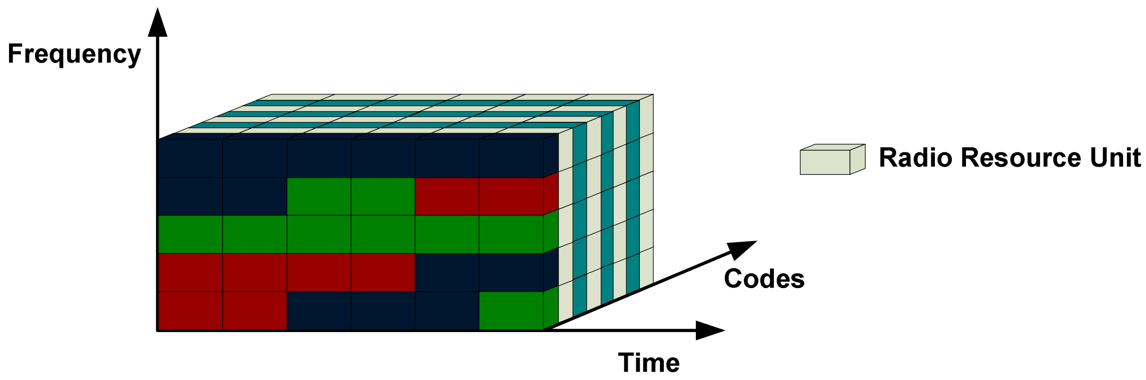



Figure 4.7. RRUs (based on [SaBa08]).

Propagation channel impairments, dynamics and end-users mobility are implicitly considered by continuously changing the end-user SINR, influencing the applied MCS, and thus the amount of RRUs needed to perform the service in a given RAT.

VNets are classified according to their contractual requirements, namely, QoS requirements. Different VNets may have quite different QoS requirements. Each VNet must be able to define its own QoS policies, so within a given VNet, different classes of traffic must be handled differently. At the infrastructure level, handling QoS and isolation between VNets may follow several approaches with different degrees of resource optimisation. For scalability reasons, QoS mechanisms in the substrate should try to use aggregation mechanisms. On the other hand, inter-VNet isolation requires a segregation of resources between different VNets. Thus, there is a trade-off between scalability and capability to guarantee strict QoS isolation between VNets [Bauc10a].

The kind of assurance, considered in the thesis, to be provided by a VNet is Guaranteed (GRT) QoS and BE. The former ensures that the requested constraints, per VLink aggregate, are not violated at any time. The latter provides a best-effort service, i.e., no guarantee at all is given if or when data are delivered, though indicative performance parameters may be followed. Statistical multiplexing should also be used to efficiently manage available resources. Assuming that not every VNet at all times uses its full amount of allocated physical resources, and viewing each single VNet as a stochastic process, an InP can cautiously overbook its substrate resources trying to increase its revenue.

The proposed generic network structure is illustrated in Figure 4.8.

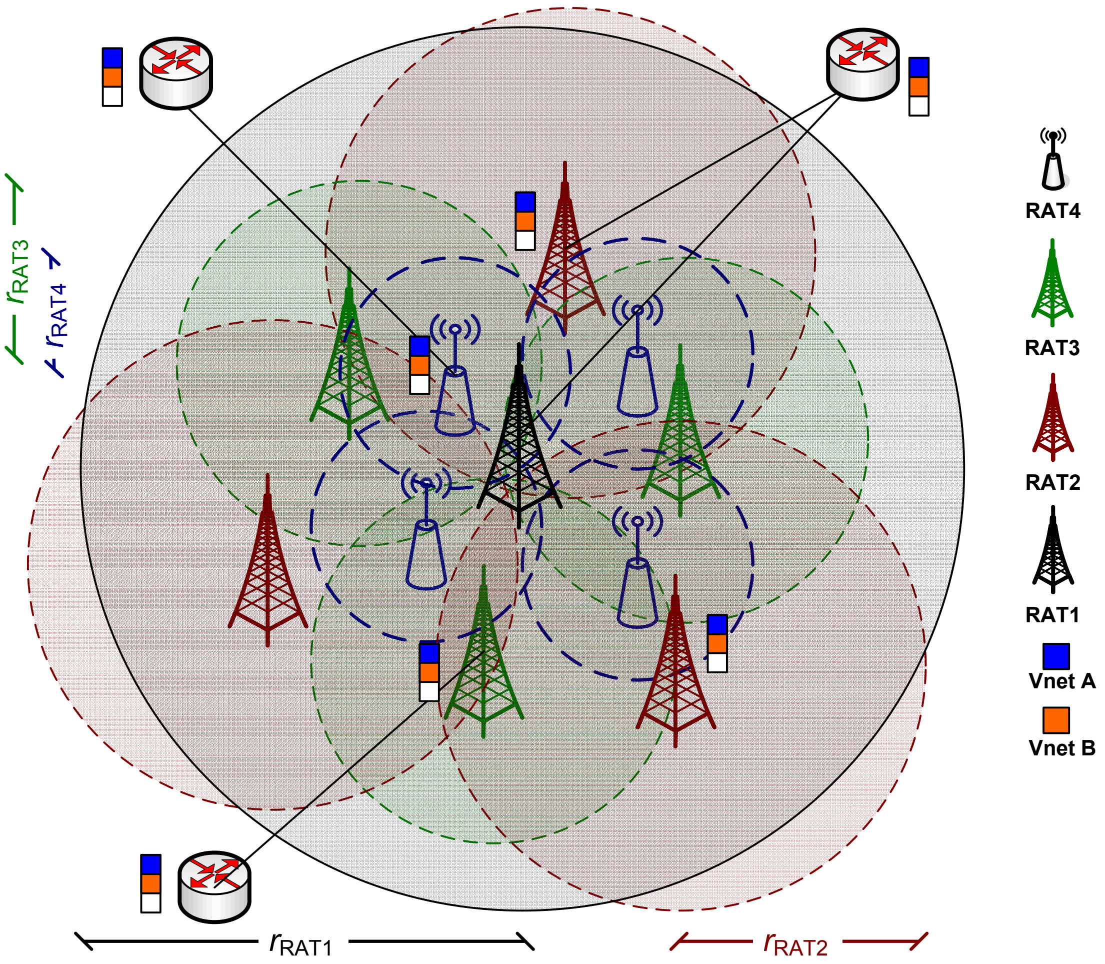



Figure 4.8. Network Structure.

### 4.3.2 Model Inputs

Different types of inputs are needed to feed the proposed model, which are grouped according to their nature into Scenarios, Network, System and Performance ones, Figure 4.9.

Scenario inputs are all the parameters that characterise the environment and the usage of the network for a given operational situation. The following sub-groups are identified:

- Global network utilisation - service penetration;
- End-users related information - quantity and usage profile;
- Physical resources - quantity, location and RAT of each BS;
- Virtual resources - quantity, type of requirements and composition of VNets;
- Inter-Provider agreements - rules for network sharing;
- Clients/Providers Strategies - KPIs used for CF computation;

In particular, for the virtual resources (VNets and VBSs), the main information needed as input is:

- $VNet_{ID}$ - VNet identification;
- $VNet_{tos}$ - VNet type of service, e.g., BE or GRT;

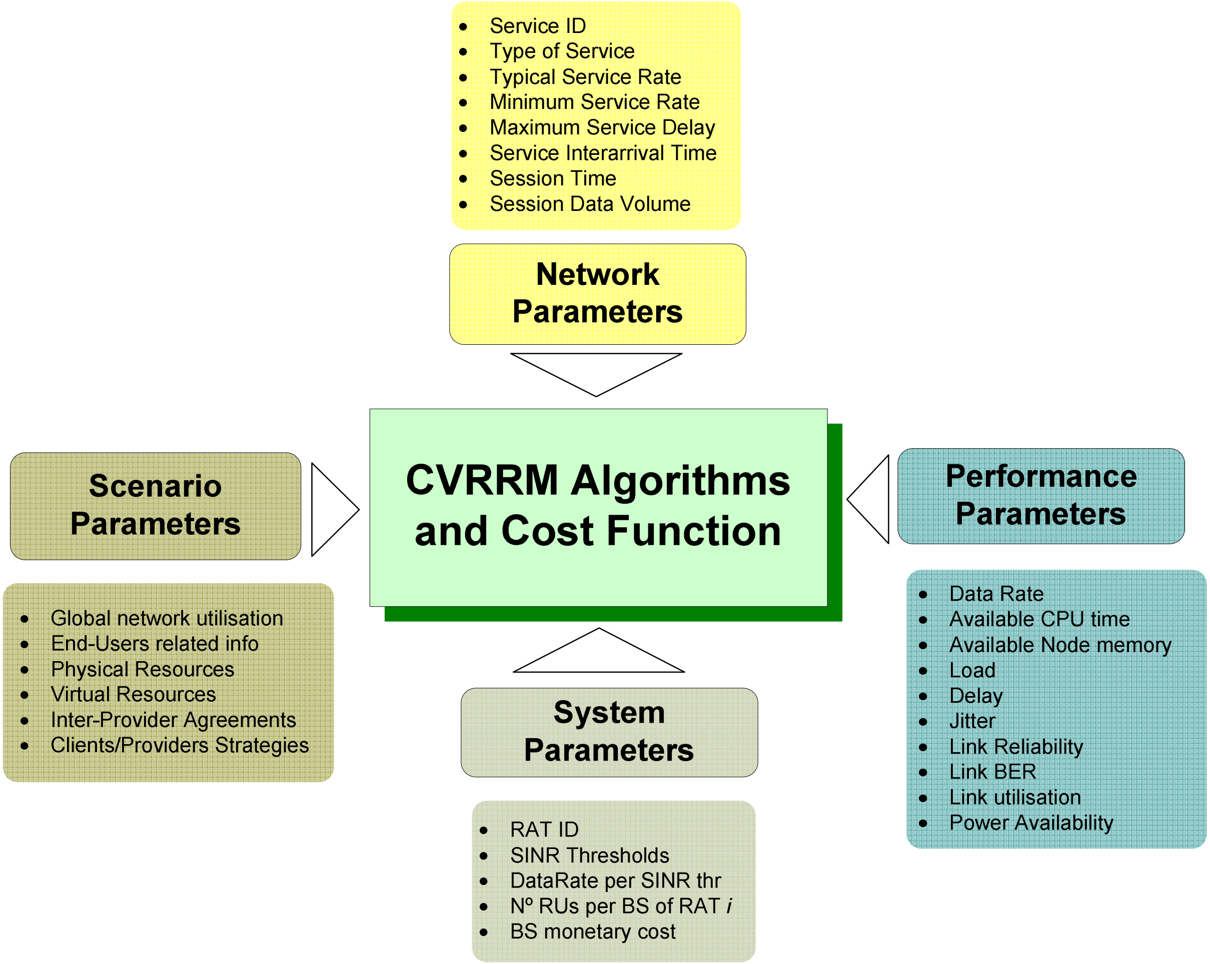



Figure 4.9. Model inputs.

- $VNO_{ID}$ - VNO identification, which should be associated with each VNet, in order to infer the agreements between VNOs, e.g., allowing end-users from others VNOs to access the VNet;
- $R_{min}^{VBS}$ - Minimum VBS Contracted Data Rate, the data rate contracted by the VNO as the minimum value InPs should provide when requested;
- $R_{ref}^{VBS}$ - Reference VBS Contracted Data Rate, the data rate contracted by the VNO as a reference value to be provided by InPs to the VBS;
- $\tau_{max}^{VBS}$ - VBS maximum delay.

$VNet_{ID}$, $VNO_{ID}$ and $VNet_{tos}$ defined for a VNet are inherited by the VBSs composing the VNet.

The Network inputs are the parameters that typify the provided network services:

- $s_{ID}$ - Service identification, e.g., Voice over IP (VoIP), Video, Web;
- $s_{tos}$ - Type of service, i.e., BE or GRT;
- $s_{typ}^{R}$ - Typical Service Data Rate, i.e., the typical data rate for service performing;
- $s_{min}^{R}$ - Minimum Service Data Rate, i.e., the minimum data rate to perform the service;

- $s^{\tau}_{max}$ - Maximum Service Delay, i.e., the maximum delay the service can support;
- $s_{\lambda}$ - Service Inter-arrival Time, i.e., the average time between two service sessions;
- $s_{\Delta t}$ - Service Session Time, i.e., the average time duration of the service;
- $s_{Vol}$ - Service Session Data Volume, i.e., the average amount of data volume per service session.

The wireless physical resource, the BS, can be generically characterised by its total number of RRUs, $N^{RAT_r}_{RRU}$, which for simplification is considered fixed for a given $RAT_r$; it is assumed that signalling and control channels are omitted. The meaning of each RRU, and its associated data rate, depends mainly on the technology in use. Furthermore, channel impairments and sources of interference, which imply changes in SINR values, are also technology dependent, and should be taken into account to evaluate the RRU data rate over time. Other characteristics of great importance for the CF calculation, related to the type of RAT in use, are the type of allowed end-users mobility and the monetary cost of the BS that should include not only the initial BS cost (CAPEX) but also its operational one (OPEX). This set of parameters has been considered as System inputs, since they are specific of the wireless/mobile systems in use. The characterisation of each system is done according to:

- $RAT_{ID}$ - RAT identification;
- $\gamma^{RAT_r}_{thr}$ - SINR threshold for RAT *r*;
- $R^{RAT_r}_{\gamma_n}$ - Data rate per RRU of RAT *r* associated with SINR threshold $\gamma_n$;
- $R^{RAT_r}_{RRU_{max}}$ - Data rate the RRU of RAT *r* can provide, if the most favourable modulation and coding scheme is applied.
- $N^{RAT_r}_{RRU}$ - Number of RRUs per BS of a given RAT *r*;
- $c_{BS}$ - BS cost, i.e., the cost associated with the BS, both initial and operational ones.

The physical networks/resources utilisation as well as the virtual ones instantiated on them, allows analysing the networks/resources availability to provide the required and/or additional capacity. Performance parameters are the inputs that enable CVRRM algorithms to determine the network state. This way, by computing the physical and virtual resources' cost, defined in Section 4.6.1, they can adapt dynamically the decisions to be taken according to the network state. The following have been identified:

- $N^{EU}_{RRU}$ - Number of RRUs assigned to the end-user;
- $R_{\upsilon_n}$ - Data rate achieved by each RRU assigned to the end-user, according to the applied modulation and coding scheme $\upsilon_n$;
- $R^{EU}_{serv}$ - End-user served data rate;
- $R^{EU}_{typ_n}$ - Typical service data rate for end-user *n*;

- $\eta_{BS}$ - BS utilisation;
- $\eta_{VBS}$ - VBS utilisation;
- $\tau$ - Delay on service request.
- $t_{int}$ - Duration of time the service is delayed.

These parameters are collected by a monitoring process, which is assumed to be an independent block from the CVRRM viewpoint, hence, they are considered as inputs to the model.

# 4.4 Model Description

## 4.4.1 Analytical Model

An analytical model is presented in what follows, in order to obtain the VNet's capacity for the proposed network architecture. From the physical viewpoint, a cluster with a set of BSs from various RATs is considered as the small management unit in terms of VRRA:

$$BS_{RAT_r}^{Cl} = \left\{ BS_1^{RAT_r}, ..., BS_{N_{BS}^{RAT_r}}^{RAT_r} \right\} \quad (4.1)$$

where:

- $N_{BS}^{RAT_r}$ - Total number of BSs of $RAT_r$ in the cluster;
- $N_{RAT}$ - Total number of RATs in the cluster, defined as:

$$RAT^{Cl} = \left\{ RAT_1, ..., RAT_{N_{RAT}} \right\} \quad (4.2)$$

The BS characterisation is made from the viewpoint of the RAT it belongs to, and the relation to the end-users connected through it. Concerning the RAT viewpoint, besides the number of RRUs specific of that RAT, the BS is characterised by its maximum capacity or *Maximum BS Data Rate*, i.e., the total capacity (bits per second) provided by the RRUs of any given BS, from now on designated as data rate, if the most favourable modulation and coding scheme is applied. Hence, the *Maximum BS Data Rate* is given by:

$$R_{max[\text{bit/s}]}^{BS} = N_{RRU}^{RAT_r} \cdot R_{RRU_{max}}^{RAT_r} \ [\text{bit/s}] \quad (4.3)$$

Regarding the relationship between BS and end-users, the BS is characterised by the *BS Serving Data Rate*:

$$R_{serv[\text{bit/s}]}^{BS} = \sum_{n=1}^{N_{EU}^{BS_j}} R_{serv\,n}^{EU} \ [\text{bit/s}] \quad (4.4)$$

where:

- $R^{EU}_{serv\,n}$ - *End-user Data Rate* for end-user $n$, i.e., the data rate with which the end-user is being served, which depends of the number of RRUs assigned to him/her and the data rate the RRUs are achieving, being obtained by:

$$R^{EU}_{serv\,[\mathrm{bit/s}]} = N^{EU}_{RRU} \cdot R_{\upsilon_n}\ [\mathrm{bit/s}] \tag{4.5}$$

It is assumed that the distribution of end-users among BSs is uniform.

The cluster, being a group of BSs, can inherit the BS characterisation, i.e., be described by its maximum capacity and serving data rate. Hence, two other parameters have been defined: *Maximum Cluster Data Rate* and *Cluster Serving Data Rate*. The *Maximum Cluster Data Rate* is the maximum capacity of the cluster, i.e., the sum of the *Maximum BS Data Rate* of all BSs of that cluster:

$$R^{Cl}_{max[\mathrm{bit/s}]} = \sum_{n=1}^{N^{Cl}_{BS}} R^{BS_n}_{max\,[\mathrm{bit/s}]} \tag{4.6}$$

where:

- $N^{Cl}_{BS}$ - Total number of BSs within the cluster;
- $R^{BS_n}_{max}$ - Maximum data rate for BS $n$, given by (4.3).

The *Cluster Serving Data Rate* is the sum of the serving data rates of all BSs composing the cluster:

$$R^{Cl}_{serv\,[\mathrm{bit/s}]} = \sum_{n=1}^{N^{Cl}_{EU}} R^{EU}_{serv\,n}\ [\mathrm{bit/s}] \tag{4.7}$$

where:

- $N^{Cl}_{EU}$ - Total number of end-users in the cluster.

Concerning the VNet, several VBSs from various VNets may exist in the cluster, being identified by:

$$VBS^{Cl} = \left\{VBS_1, ..., VBS_{N^{Cl}_{VBS}}\right\} \tag{4.8}$$

where:

- $N^{Cl}_{VBS}$ - Total number of VBSs in the cluster.

The VBS can be defined according to the contracted capacity, $R^{VBS}_{min}$ or $R^{VBS}_{ref}$, and the data rate provided to all end-users connected to the VBS, $R^{VBS}_{serv}$, designated by *VBS Serving Data Rate*:

$$R^{VBS}_{serv\,[\mathrm{bit/s}]} = \sum_{n=1}^{N^{VBS}_{EU}} R^{EU}_{serv\,n}\ [\mathrm{bit/s}] \tag{4.9}$$

where:

- $N_{EU}^{VBS}$ - Number of end-users connected to the VBS.

The data rate requested by end-users to the VBS, i.e., the *VBS Requested Data Rate*, $R_{req}^{VBS}$, is also important information, since it allows knowing if the VBS is running in under- or overloaded conditions. It is computed as the aggregation of the typical data rates of all end-users services in the VBS*:*

$$R_{req}^{VBS}{}_{[\text{bit/s}]} = \sum_{n=1}^{N_{EU_T}^{VBS}} R_{typ_n}^{EU} \; [\text{bit/s}] \tag{4.10}$$

To express the relation between VNOs and InPs, which allows evaluating the established Serving Level Agreement (SLA), two parameters have been defined:

- Penalty, $p$ - the amount the InP should pay to the VNO, when the VBS is operating out of contract, i.e., when SLAs are not satisfied;
- Time frame, $\Delta t_{TF}$ - the interval of time of the same order of magnitude of the time scale defined for common/joint RRM algorithms.

Concerning the description of the VBSs according to the two types considered in this work, GRT and BE, the GRT VBS, $VBS^{GRT}$, is characterised by a *Minimum Contracted Data Rate*, $R_{min}^{VBS}$, which should be guaranteed for all time frames, and a *Penalty* computed as the total number of time frames the VBS is out of contract:

$$p^{GRT} = \sum_{i=1}^{N_{TF}} p_i^{GRT} \tag{4.11}$$

where:

- $N_{TF}$ - Total number of time frames in the observation interval;
- $p_i^{GRT}$ - Penalty of a given GRT VBS in time frame *i*, according to:

$$\begin{cases} p_i^{GRT} = 0, & R_{serv}^{VBS} \geq R_{min}^{VBS} \text{ in } \Delta t_{TF\,i} \\ p_i^{GRT} = 1, & R_{serv}^{VBS} < R_{min}^{VBS} \text{ in } \Delta t_{TF\,i} \end{cases} \tag{4.12}$$

The BE VBS, $VBS^{BE}$, is defined by its *Reference Contracted Data Rate*, $R_{ref}^{VBS}$, which is indicative and should be defined as a percentage $\xi_{R_{ref}}$ of the total number of samples, i.e., the minimum fraction of time frames InPs should make available the reference contracted data rate to the VNO in order to avoid penalties. An associated *Penalty* accounts for the number of time frames the VBS is out of contract above $\xi_{R_{ref}}$ percentage of the total:

$$p^{BE} = \sum_{i=1}^{N_{TF}} p_i^{BE} - \xi_{R_{ref}} \cdot N_{TF} \tag{4.13}$$

subject to:

$$\frac{\sum_{i=1}^{N_{TF}} p_i^{BE}}{N_{TF}} \geq \xi_{R_{ref}} \quad (4.14)$$

where:

- $p_i^{BE}$ - penalty of a given BE VBS in time frame *i*, according to:

$$\begin{cases} p_i^{BE} = 0, & R_{serv}^{VBS} \geq R_{ref}^{VBS} \ \ in\ \Delta t_{TF\,i} \\ p_i^{BE} = 1, & R_{serv}^{VBS} < R_{ref}^{VBS} \ \ in\ \Delta t_{TF\,i} \end{cases} \quad (4.15)$$

In order to account for the global profit, one considers the target of maximising $R_{serv}^{Cl}$,

$$max\left(R_{serv\,[\text{bit/s}]}^{Cl}\right) = max\left(\sum_{n=1}^{N_{EU}^{Cl}} R_{serv\,n[\text{bit/s}]}^{EU}\right) \quad (4.16)$$

and minimising the penalties within the all cluster,

$$min\,(p^{Cl}) = min\left(\sum_{n=1}^{N_{GRT}^{Cl}} p_n^{GRT} + \sum_{m=1}^{N_{BE}^{Cl}} p_m^{BE}\right) \quad (4.17)$$

where:

- $N_{GRT}^{Cl}$ - total number of GRT VBSs in the cluster;
- $N_{BE}^{Cl}$ - total number of BE VBSs in the cluster.

through an adequate allocation of RRUs to the VBSs. The former, (4.16), considers that VNOs pay the service based on used capacity; the latter, (4.17), assumes that an amount of money must be paid back to the VNO if the contract is not fulfilled.

## 4.4.2 Data Rate Estimation

In a heterogeneous network environment, where different technologies are used, one needs to know which are the factors that influence QoS parameters, like changes in modulation due to lower signal quality, interference (influenced by the number of users, cells, carriers, etc.), power, capacity, and speed. The main factors for a particular technology must be identified, however, the interaction between physical layer and MAC sub-layer must also be considered.

Each of the QoS parameters must be related to each RRU, in order to know how many RRUs should be assigned to guarantee users' requirements. However, the capacity of the RRUs is not static, since wireless medium channel impairments, interference conditions, coding and modulation waveforms vary in time and space. The relationship between QoS parameters and RRUs for each RAT is presented in Table 4.1.

Table 4.1. Relation between QoS parameters and RRUs.

| | TDMA | CDMA | OFDM/OFDMA |
|---|---|---|---|
| **RRUs** / **QoS** | # Time-slots | # Codes and Power | # frequencies and # sub-channels |
| **Data rate** | SNR + Modulation + Coding | SINR + Modulation + Coding | |
| **Delay** | Propagation time + Frame width | Propagation time | Propagation time + MAC |
| **Packet error rate** | SINR | | |

The data rate is the main QoS parameter to be analysed in this work, which is highly dependent on the SINR. For packet data services, a larger SINR can be used to provide higher data rates by reducing coding or spreading, and/or increasing the constellation density. It is straightforward to see that cellular spectral efficiency, in terms of bit/s/Hz, can be increased by a factor of two or more if users with better links are served at higher data rates [NaBK00]. To achieve optimal data rates, fast rate adaptation is required on fast fading channels; data rate adaptation techniques adjust the coding and modulation schemes based on SINR values.

A generic approach is as follows: when the received SINR is high, the radio link BER is low, hence, a coding scheme with a small number of parity bits may offer adequate protection; at low SINR, "stronger" codes may be needed to protect data against radio link errors, since these codes add more parity bits to each block. The error performance of a cellular radio link varies as end-users move within a cell. To make the most efficient use of the radio link, coding schemes are dynamically selected in response to changes in the quality of the radio link.

It is possible to estimate the data rate at which payload bits are carried over the radio link as a function of SINR. Throughput-versus-SINR curves for all MCSs available in a wireless data network show at which values of SINR it is advantageous to switch MCS. A generic relation throughput-versus-SINR is depicted in Figure 4.10 for three MCSs. The basic principle is to use higher modulation levels and “weaker” channel coding when the channel condition is good, and on the other hand, to use lower modulation levels and “stronger” channel coding when the channel is not so good. SINR switch points are often hard-coded at the transmitter, and correspond to points for which the throughput follows the *Optimum* curve.

In this thesis, the data rate is considered constant by intervals, which is a good approximation, except for lower values of SINR:

$$R_{[\text{bit/s}]} = f(\upsilon) = g(\gamma) = \begin{cases} R_{1[\text{bit/s}]} & \text{if} \quad \gamma_{-1[\text{dB}]} \leq \gamma_{[\text{dB}]} \leq \gamma_{0[\text{dB}]} \\ R_{2[\text{bit/s}]} & \text{if} \quad \gamma_{0[\text{dB}]} < \gamma_{[\text{dB}]} \leq \gamma_{1[\text{dB}]} \\ R_{3[\text{bit/s}]} & \text{if} \quad \gamma_{1[\text{dB}]} < \gamma_{[\text{dB}]} \leq \gamma_{2[\text{dB}]} \\ \ldots & \\ R_{n[\text{bit/s}]} & \text{if} \quad \gamma_{n-2[\text{dB}]} < \gamma_{[\text{dB}]} \leq \gamma_{n-1[\text{dB}]} \end{cases} \tag{4.18}$$

where:

- $R$ - Data rate;
- $\upsilon$ - Modulation and coding scheme;
- $\gamma$ - SINR value.

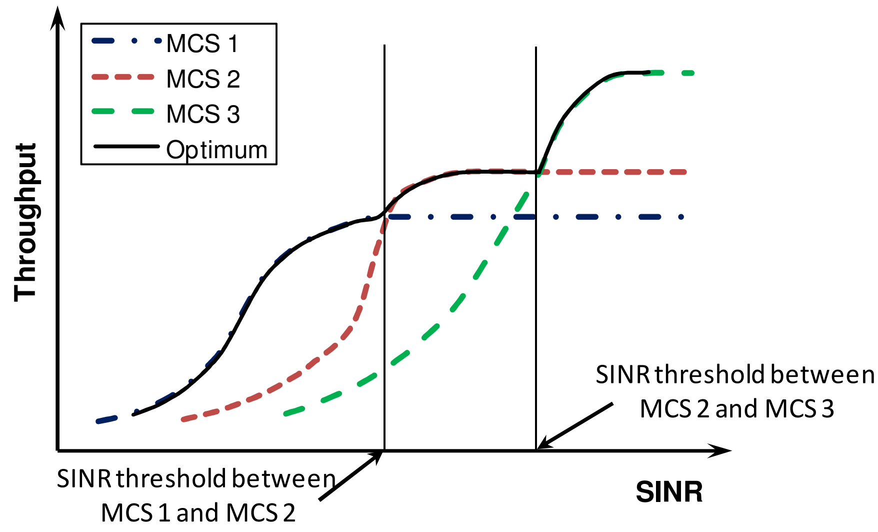



Figure 4.10. Relation between throughput and SINR for different MCSs (based on [TCST04]).

The selection of each MCS, modulation and coding pair $\upsilon$, is performed based on defined SINR thresholds, enabling the availability of higher individual data rate, therefore, increasing the average data rate per cell, and the adaptable robustness to cope with errors introduced while transmitting over the fading radio channel. The threshold values $\gamma_i$ to switch the MCS, are largely studied in the literature in order to optimise spectrum efficiency, and depend on the particular system and the procedure used to determine the SINR value [ZhVi05]. The value of $R$ achieved for each selected MCS depends also on the system.

The metrics used to determine the SINR value and the parameters that are considered for rate adaptation depend on the wireless technology and system characteristics.

In TDMA, a slot-by-slot data rate adaptation is achieved through adaptive coding and modulation, while the symbol rate and block size are left unchanged. Additionally, higher data rates are obtained in some systems by time-slot aggregation or incremental redundancy. This last procedure effectively matches the coding rate to the channel SINR without requiring SINR estimation and feedback. In addition, the transmission of redundant information dispersed in time provides a diversity advantage during decoding [NaBK00]. A summary of rate adaptation for some systems is presented in Table A.1. In these systems, channel quality is estimated at the receiver, and information is provided to the transmitter through appropriately defined messages. Some metrics have been proposed to estimate channel quality: frame error rate; mean and standard deviation of Symbol Error Ratio (SER) or BER; average SINR. The interference value depends on the reuse factor, which can be considered as a constant, and rate adaptation is based on variations of the received signal. SINR thresholds and respective data rates used for GSM's Enhanced GPRS (EGPRS) are presented in Table A.4.

For CDMA, the basic RRU is the code and the associated power. Rate adaptation in these systems is achieved through a combination of variable spreading, coding, and code aggregation. Higher rates are achieved differently for the various systems. In the cdmaOne, it is done through Walsh code aggregation, one to eight codes being assigned to each data user, each of which supporting a data rate of 9.6 kbit/s. UMTS and CDMA2000 achieve higher rates through a combination of variable spreading and coding. Incremental redundancy is also being considered in UMTS [NaBK00]. Pilot strength measurements are used to estimate the SINR, e.g., in cdmaOne and CDMA2000, these measurements are provided to the BS through the Pilot Strength Measurement Message (PSMM) or included in the Supplemental Channel Request Message (SCRM); in UMTS, the measurement report message can additionally include BLER, BER, received power, path loss, and DL SINR measurements. A summary of rate adaptation for these systems is presented in Table A.2.

Interference depends on the number of users and on the active service data rate per user, which determines the power allocated to each user. Assuming that the main contribution to interference is caused by intra-cell one, the interference power on user *i* is computed by considering all active users receiving or transmitting in the cell [PrCJ02]:

$$I_{i[\mathrm{W}]} = \sum_{\substack{j=1 \\ i \neq j}}^{N_u} L_{p\,j} P_{j[\mathrm{W}]} \tag{4.19}$$

where:

- $L_{p\,j}$ - path loss for end-user *j;*
- $P_j$ - transmitted power for/by end-user *j.*

This dependency is considered linear, in order to simplify the model, being a percentage of the maximum possible interference, corresponding to the maximum number of users in the cell [Lope08]:

$$I_{i[\mathrm{W}]} = \frac{N_{EU}}{N_{EU}^{max}} \cdot I_c^{max}{}_{[\mathrm{W}]} \tag{4.20}$$

where:

- $I_c^{max}$ - maximum interference power value;
- $N_{EU}$ - number of active end-users;
- $N_{EU}^{max}$ - maximum number of end-users in the cell.

SINR can then be derived from this interference value in conjunction with the power allocated to the user. Thus, the data rate may be obtained by using the threshold method referred above. In Appendix A, SINR thresholds and respective data rates for UMTS/HSPA are presented as an example.

In multi-carrier systems, like OFDM, data symbols are modulated onto sub-carriers, and the SINR on each sub-carrier is measurable, its value indicating the channel quality during symbol transmission. Due to this property, it is well known that adaptive modulation and coding can be easily implemented on a sub-carrier/sub-channel basis, based on the SINR measurement on sub-carriers/sub-channels. The decision to determine the suitable physical mode within each OFDM sub-channel is based on the SINR level expected; it requires Channel State Information (CSI), which can be obtained by Channel Quality Indication (CQI); Dynamic Subcarrier Assignment (DSA) also requires this information to choose the right sub-frequency for each end-user [RaWa07]. A summary of rate adaptation for OFDM based systems is presented in Table A.3.

In order to determine each end-user data rate, a link layer abstraction procedure is used, Figure 4.11, consisting of evaluating the effective SINR based on the SINR level of each sub-carrier allocated to the end-user. The effective SINR maps an instantaneous multi–state channel – described by a set of subcarrier SINR samples – onto an instantaneous scalar value, the effective SINR. The effective SINR, $\gamma_{eff}$ , is then used to find an estimate of the BLER probability from basic AWGN link-level-performance. The accuracy of the Effective SINR Mapping (ESM) is validated through an adjustment of the predicted BER ( $\gamma_{eff}$ ) for an AWGN channel to the measured instantaneous BER ( $\gamma_i$ ) derived from link level simulations. This equalisation of BER samples for all instantaneous channel states allows using AWGN mapping tables for various channel models in system level simulations [MoOb06].

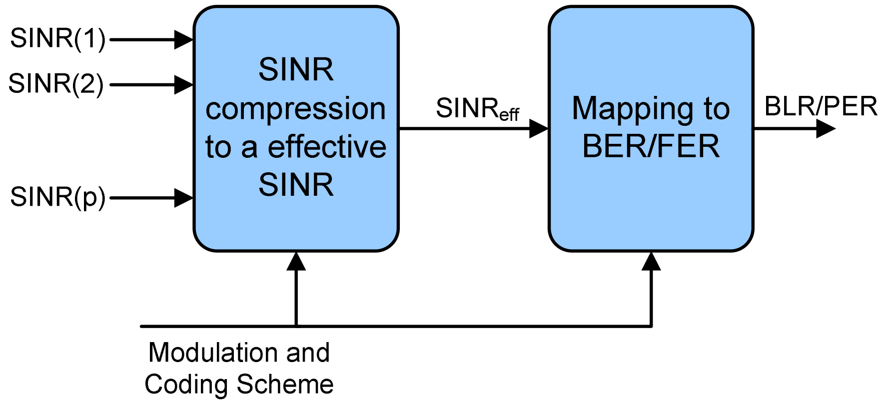



Figure 4.11. Link layer abstraction procedure (based on [TuWa05]).

There are several popular mapping functions proposed in literature to perform the effective SINR calculation, such as, Exponential Effective SINR Mapping (EESM) [HSHL97], and Mutual Information (MI) Effective SINR Mapping (MI-ESM) based link quality model [TsSo03]. EESM is a simple mapping method, in which all the sub-carriers for an end-user have to use the same modulation and coding scheme. MI-ESM is more advanced, since the use of the same MCS for all the sub-carriers of an end-user is not imposed. The basic idea of all methods is to find a compression function that maps the sequence of varying SINRs onto a single value that is strongly correlated with the actual BLER/PER [TuWa05]. The effective SINR determination is not within the scope of this work, since one assumes the existence of a monitor entity providing this information; hence, one derives the sub-channel data rate by multiplying it by the bandwidth of the sub-channel. SINR thresholds and respective data rates used for OFDM and OFDMA (Wi-Fi and LTE, respectively) are presented in Tables A.8 and A.10.

## 4.5 Evaluation Metrics

Network performance indicators are essential to assess algorithms in different scenarios and network conditions. Those depicted bellow are used, allowing a proper validation of the proposed model, by assessing critical issues related to the virtualisation process, such as Virtual Access with QoS guarantees.

The performance of the physical resources is assessed through the *Average Cluster Serving Data Rate* and the *Cluster Utilisation,* which gives an indication about the efficiency of the overall radio resources available in the cluster.

To access the performance of the virtual resources, the VBSs, one defined the *Average VBS Serving Data Rate, VBS Utilisation, Average End-user Data Rate, Ratio of Data Rate Served*, *Average Cluster Serving Data Rate*, *Cluster Utilisation*, *Out of Contract*, *Satisfaction Level on the InP, Satisfaction Level on extra Capacity Requested, Average VBS Time Service Delayed InP*, *Average VBS Time Service Delayed VNO, Average Delay on Service Request InP*, and *Average Delay on Service Request VNO:*

- *Average VBS Serving Data Rate* - average of the VBS serving data rate over the observation time interval:

$$\overline{R_{serv}^{VBS}}_{[\text{bit/s}]} = \frac{\sum_{n=1}^{N_{TF}} R_{serv\,n[\text{bit/s}]}^{VBS}}{N_{TF}} \tag{4.21}$$

  $\overline{R_{serv}^{VBS}}$ allows evaluating the algorithm ability to allocate the adequate quantity of RRUs to the VBS, in order to satisfy the VBS contracted data rate.

- *VBS Utilisation* - ratio between the *Average VBS Serving Data Rate* and the minimum contracted data rate:

$$\eta_{VBS} = \frac{\overline{R_{serv}^{VBS}}_{[\text{bit/s}]}}{R_{min\,[\text{bit/s}]}^{VBS}} \tag{4.22}$$

  A value of $\eta_{VBS}$ greater than 1.0 means that the $R_{serv}^{VBS} \geq R_{min}^{VBS}$ for GRT VBSs or $R_{serv}^{VBS} \geq R_{ref}^{VBS}$ for BE VBSs.

- *Average End-user Data Rate* - average data rate the end-user has been served during the observation time interval:

$$\overline{R_{serv}^{EU}}_{[\text{bit/s}]} = \frac{\overline{R_{serv}^{VBS}}_{[\text{bit/s}]}}{\overline{N_{EU}^{VBS}}} \tag{4.23}$$

  where:

- $\overline{N_{EU}^{VBS}}$ - average number of connected end-users during the observation interval:

$$\overline{N_{EU}^{VBS}} = \frac{\sum_{n=1}^{N_{TF}} N_{EU\ n}^{VBS}}{N_{TF}} \tag{4.24}$$

where:

- $N_{EU\ n}^{VBS}$ - number of end-users connected to the VBS in time frame *n*.

$\overline{R_{serv}^{EU}}$ allows evaluating how the algorithm influences the handling of overall end-users, in order to maintain the end-user data rate of GRT services between certain limits, maximising the end-user data rate for BE services.

- *Ratio of Data Rate Served* - VBS served data rate relative to the VBS requested data rate:

$$r_{serv}^{VBS} = \frac{\overline{R_{serv}^{VBS}}[\text{bit/s}]}{\overline{R_{req}^{VBS}}[\text{bit/s}]} \tag{4.25}$$

where:

- $\overline{R_{req}^{VBS}}$ - average of the VBS requested data rate over the total number of time frames in the observation time interval, given by:

$$\overline{R_{req}^{VBS}}_{[\text{bit/s}]} = \frac{\sum_{n=1}^{N_{TF}} R_{req\ n[\text{bit/s}]}^{VBS}}{N_{TF}} \tag{4.26}$$

  where:

  - $R_{req\ n}^{VBS}$ - VBS requested data rate in time frame *n*, given by (4.44).

$r_{serv}^{VBS}$ takes values from 0 to 1, depicting situations of heavy or light traffic, respectively. It is used to support the evaluation of the VBS response to the amount of requested data rate. Analysed together with $\overline{R_{serv}^{VBS}}$, it allows to determine if the amount of contracted capacity by the VNO is adequate to end-users demand.

- *Average Cluster Serving Data Rate* - average of the cluster serving data rate over the observation time interval:

$$\overline{R_{serv}^{Cl}}_{[\text{bit/s}]} = \frac{\sum_{n=1}^{N_{TF}} R_{serv\,n[\text{bit/s}]}^{Cl}}{N_{TF}} \tag{4.27}$$

where:

- $R_{serv\,n}^{Cl}$ - Cluster serving data rate in time frame *n*.

$\overline{R_{serv}^{Cl}}$ is defined to evaluate the performance of the overall cluster, allowing one to observe the impact of using VRRA algorithms for different use cases.

- *Average Cluster Utilisation* – average cluster utilisation over the observation interval:

$$\overline{\eta_{Cl}} = \frac{\sum_{n=1}^{N_{TF}} \eta_{Cl\,n}}{N_{TF}} \tag{4.28}$$

where:

  - $\eta_{Cl\,n}$ - ratio between the maximum data rate corresponding to the RRUs occupied by end-users in time frame *n* and the maximum data rate the cluster can provide, given by:

$$\eta_{Cl} = \frac{\sum_{r=1}^{N_{RAT}} N_{RRU_{occ}}^{RAT_r} \cdot R_{RRU_{max}\,[\text{bit/s}]}^{RAT_r}}{\sum_{r=1}^{N_{RAT}} N_{RRU}^{RAT_r} \cdot R_{RRU_{max}\,[\text{bit/s}]}^{RAT_r}} \tag{4.29}$$

  where:

    - $N_{RRU_{occ}}^{RAT_r}$ - total number of RRUs occupied by end-users for RAT *r*.

$\overline{\eta_{Cl}}$ is a measure of the RRUs utilisation within the cluster. It should be analysed together with $\overline{R_{serv}^{Cl}}$, since the efficiency of the use of the RRUs is as important as maximising their use.

- *Out of Contract* - total number of time frames out of contract over the observation time interval:

$$r_{TF}^{out} = \frac{N_{TF}^{out}}{N_{TF}} \tag{4.30}$$

where:

  - $N_{TF}^{out}$ - number of time frames out of contract, i.e., the number of time frames in which $R_{serv}^{VBS} < R_{min}^{VBS} < R_{req}^{VBS}$.

- *Satisfaction Level on the InP* - VNO satisfaction level regarding the service provided by the InP, concerning the provision of enough physical resources to fulfil the contracted capacity:

$$S_{InP}^{VNO} = 1 - \frac{N_{nc_{InP}}^{EU}}{N_{nc_{InP}}^{EU} + N_{con}^{EU}} \tag{4.31}$$

where:

  - $N_{nc_{InP}}^{EU}$ - number of end-users not connected during the observation time interval when $R_{serv}^{VBS} < R_{min}^{VBS} < R_{req}^{VBS}$;
  - $N_{con}^{EU}$ - number of end-users connected during the observation time interval.

$S_{InP}^{VNO}$ accounts for the effective decrease in the amount of contracted capacity perceived by the VNO, hence, it can be used to detect contract violations. It is worthwhile to note that $S_{VNO}$ can be considered as a user satisfaction measure, since VNOs are indirectly the "users" from the VRRA viewpoint.

- *Satisfaction Level on extra Capacity Requested* - satisfaction level regarding the service provided by the InP when the VBS is already running with the contracted capacity, but VBS end-users request additional capacity:

$$S_{ovl}^{VNO} = 1 - \frac{N_{nc_{ovl}}^{EU}}{N_{nc_{ovl}}^{EU} + N_{con}^{EU}} \qquad (4.32)$$

where:

- $N_{nc_{ovl}}^{EU}$ - number of end-users that are not connected when the capacity contracted for the VBS is already reached (VBS overloaded), $R_{min}^{VBS} < R_{serv}^{VBS} < R_{req}^{VBS}$ .

The metrics related to the contract established between VNOs and InPs are the *Out of Contract* and the *Satisfaction Level,* which allow measuring the contract failure from the InP viewpoint, and the grade of satisfaction of the VNO in the service provided by the InP, respectively.

For performance indicators related to the delays experienced by VNet end-users, two main measures are considered, Figure 4.12. : the *Delay on Service Request*, $\tau$, which is the time elapsed between the instant an end-user tries to enter the network, $t_i$, and the instant at which the connection is established; and the *Service Time Delayed*, $t_{int}$, which is the time the end-user is delayed during the session, due to lack of RRUs to achieve the minimum service data rate.

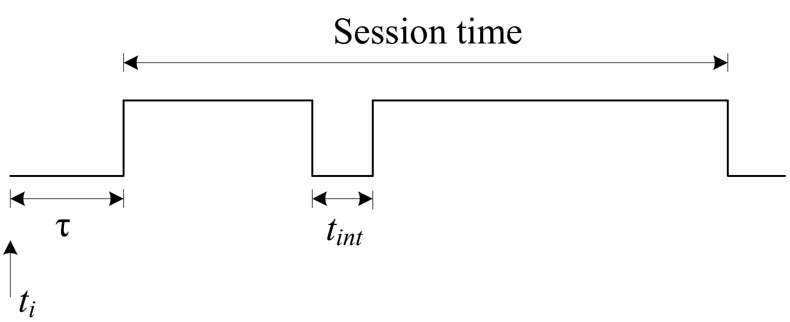



Figure 4.12. End-user Delays.

- *Average VBS Time Service Delayed InP* - average time end-users performing a service are delayed during the total observation time interval when the VBS is running within the contracted capacity:

$$\overline{t_{int}^{VBS}}[\mathrm{s}] = \frac{\sum_{n=1}^{N_{EU}^{VBS}} t_{int\,n}^{EU}[\mathrm{s}]}{N_{EU}^{VBS}} \qquad (4.33)$$

where:

- $N_{EU}^{VBS}$ - Number of end-users connected to the VBS over the observation interval;

- $t_{int_n}^{EU}$ - Time the service has been delayed for end-user *n,* when the VBS is running within the contracted capacity.

The amount of time the service is delayed corresponds to the duration of time the end-user cannot achieve the minimum service data rate. In particular, for BE services, it is the time the end-user is connected to the network with data rate equal to zero.

- *Average VBS Time Service Delayed VNO* - average time end-users are delayed during the total observation time interval when the VBS is running over the contracted capacity:

$$\overline{t_{VNO}^{VBS}}[\mathrm{s}] = \frac{\sum_{n=1}^{N_{EU}^{VBS}} t_{VNO_n}^{EU}[\mathrm{s}]}{N_{EU}^{VBS}} \tag{4.34}$$

where:

- $t_{VNO_n}^{EU}$ - time the service has been delayed for end-user *n,* when the VBS is running over the contracted capacity.

- *Average Delay on Service Request InP* - average delay end-users experience on service request when the VBS has not reached the total amount of contracted data rate:

$$\overline{\tau_{InP}^{VBS}}[\mathrm{s}] = \frac{\sum_{n=1}^{N_{EU_T}^{VBS}} \tau_{InP_n}^{EU}[\mathrm{s}]}{N_{EU}^{VBS} + N_{EU_{nc}}^{VBS}} \tag{4.35}$$

where:

- $\tau_{InP_n}^{EU}$ - total delay experienced by end-user *n* when the VBS has not reached the total amount of data rate contracted, $R_{serv}^{VBS} < R_{min}^{VBS} < R_{req}^{VBS}$;
- $N_{EU_{nc}}^{VBS}$ - total number of end-users that tried to enter the network but have not been connected during the observation time interval;
- $N_{EU_T}^{VBS}$ - total number of end-users trying to connect to the VBS in the observation interval.

- *Average Delay on Service Request VNO* - average delay end-users experienced on service request when the VBS has been served with the minimum VBS contracted data rate:

$$\overline{\tau_{VNO}^{VBS}}[\mathrm{s}] = \frac{\sum_{n=1}^{N_{EU_{nc}}^{VBS}} \tau_{VNO_n}^{EU}[\mathrm{s}]}{N_{EU}^{VBS} + N_{EU_{nc}}^{VBS}} \tag{4.36}$$

where:

- $\tau_{VNO}^{EU}$ - total delay experienced by each end-user when the minimum VBS contracted data rate has been already reached, $R_{min}^{VBS} < R_{serv}^{VBS} \leq R_{req}^{VBS}$.

This metric allows to detect situations in which the VNO has contracted a lower capacity than needed, or a peak of traffic occurs in the considered VBS.

## 4.6 Strategies and Algorithms

### 4.6.1 Parameters

CVRRM functions will interact with a Monitoring Entity (ME), which provides real-time measurements, like available resources quantity and quality, co-located resources and failure detection. Furthermore, it is assumed that an ME instance exists in the physical node, providing global monitoring information, and in each VNode, collecting its own monitoring information. It is assumed that the ME monitors the wireless medium and the node, therefore, providing the inputs to CF computation, based on [SeCo07], in order to allow the comparison among resources, and among VNets.

The strategies used by CVRRM to select the “best” VNet or to support handover among VNets are related to the contractual VNet requirements, being reflected by KPIs’ weights in the CF computation for VNet comparison. The CF performed per VNet or VBS allows integrating a set of KPIs into a single one, in this case, the cost of a given VNet. The cost value is then a common metric that allows identifying the usefulness of a VNet to provide a given service.

A CF performed per BS is also considered for the selection of the “best” BS to connect end-users, additionally to the preferred list of RATs for the requested service. The KPIs’ weights for this CF computation are defined according to VNet requirements, e.g., in terms of capacity, delay, energy consumption or mobility from the user viewpoint, which is the VNO in this case. Of course these weights are combined with the ones defined by the InP strategy for managing the physical infrastructure, the operator viewpoint.

The approach considered for CF calculation is based to [SeCo07], though it is adapted to the VNet environment. The total CF of a resource is divided into two sub-CFs, one being related to the InP and the other to the VNO. Each of this sub-CFs is weighted with different values, enabling the implementation and evaluation of different policies on the CVRRM, according to the type of VNet. The operator/InP cost for BS $b$ of type of RAT $r$, $c_{o_{r,b}}$, is computed as:

$$c_{o_{r,b}} = \frac{1}{\sum_{i=1}^{N_{KPI_r}} w_{r,i}} \sum_{i=1}^{N_{KPI_r}} w_{r,i} \cdot k_{b,i} \qquad (4.37)$$

where:

- $N_{KPI_r}$ - total number of KPIs of a given RAT $r$;

- $w_{r,i}$ - weight of *i* KPI of RAT *r*;
- $k_{b,i}$ - normalised value of each KPI *I* for BS *b* ($0 \le k_{b,i} \le 1$).

The cost for each user/VNO *u*, $c_{u_n}$, is calculated according to:

$$c_{u_n} = \frac{1}{\sum_{i=1}^{N_{KPI_u}} w_i} \sum_{i=1}^{N_{KPI_u}} w_i \cdot k_{u,i} \tag{4.38}$$

where:

- $N_{KPI_u}$ - total number of user/VNO KPIs, defined as a function of the VNet type and requirements;
- $k_{u,i}$ - normalised value of user/VNO KPI *i*;
- $w_i$ - weight for the user/VNO KPI *i*.

Both $c_{o_{r,b}}$ and $c_{u_n}$ are normalised parameters, thus, in normal situations, these two should be in between 0 and 1. The cost of the BS *b* to attach an end-user of a given VNet, is given by:

$$c_b = \frac{1}{w_o + w_u}\left(w_o \cdot c_{o_{r,b}} + w_u \cdot c_{u_n}\right) \tag{4.39}$$

where:

- $w_o$ - Operator/InP's weight;
- $w_u$ - User/VNO's weight;

In this thesis, the perspective of the user/VNO is not considered for the sake of simplicity, thus, $w_o = 1$ and $w_u = 0$. The BS cost is based on the maximum data rate available on the BS, in order to perform load balance among the BSs of the cluster, which is normalised over the maximum $R_{max}^{BS}$ among all the RATs in the cluster.

$$R_{av}^{BS}{}_{[\text{bit/s}]} = \frac{\left(N_{RRU}^{RAT_r} - N_{RRU_{occ}}^{BS}\right) \cdot R_{RRU_{max}}^{RAT_r}{}_{[\text{bit/s}]}}{max(R_{max}^{BS})\big|_{RAT^{Cl}}} \tag{4.40}$$

where:

- $R_{RRU_{max}}^{RAT_r}$ - maximum RRU data rate for $RAT_r$;
- $N_{RRU}^{RAT_r}$ - total number of RRUs per BS of $RAT_r$;
- $N_{RRU_{occ}}^{BS}$ - number of RRUs occupied by end-users in BS;
- $R_{max}^{BS}$ - Maximum BS data rate.

The BS to be selected is then the one that, from the more adequate RAT to perform the service, have the minimum cost, i.e., the maximum available data rate.

The detailed algorithms description for each of the VRRA functions is presented in Sections 4.6.2 and 4.6.3. The Adaptive Virtual Radio Resource Allocation (Adaptive-VRRA), Section 4.6.2, does a pre--

allocation of the RRUs to the VBSs according to the contracted capacity over the set of heterogeneous wireless systems available, adapting it to compensate wireless link variations. The OnDemand Virtual Radio Resource Allocation (OnDemand-VRRA), Section 4.6.3, allocates RRUs only if they are requested by VNet end-users, still adapting the RRUs allocation to reach the VBS contracted capacity. As the primary issue arising from the virtualisation of the wireless access is concerned to the infrastructure sharing and isolation among multiple VNOs, VRRA is considered as the main function of CVRRM, hence, Initial VNet Selection and VNet Handover Support were not further developed in the context of this thesis.

The main target of VRRA is to provide the required capacity to VBSs, optimising radio resources utilisation. The VRRA algorithms presented in this thesis are heuristic ones, which manage the allocation of RRUs among VBSs when they are requested by VNet end-users. The management of radio resource allocation from VBSs is coordinated to provide different levels of service to the various VNOs or SPs. This is achieved by taking the variability of the wireless medium and the diversity of the existing RATs into account.

The VRRA algorithms work on a time frame basis, larger than all time frames associated with each of the RATs under consideration, hence, all allocation decisions taken at the VBS level are implemented at RAT one. OnDemand-VRRA is responsible for dynamically (re)allocating RRUs, satisfying the *Minimum Contracted Data Rate* for GRT VNets (4.41), and aiming at the *Reference Contracted Data Rate* for BE VNets (4.42):

$$R_{serv}^{VBS_i}{}_{[\text{bit/s}]} \geq R_{min}^{VBS_i}{}_{[\text{bit/s}]} \quad , \quad R_{req}^{VBS_i} \geq R_{min}^{VBS_i} \, , \forall \, VBS_i \in VBS^{GRT} \tag{4.41}$$

$$min\left( R_{ref}^{VBS_j}{}_{[\text{bit/s}]} - R_{serv}^{VBS_j}{}_{[\text{bit/s}]} \right) \quad , \; R_{ref}^{VBS_j} > R_{serv}^{VBS_j} \, , \; \forall \, VBS_j \in VBS^{BE} \tag{4.42}$$

subject to:

$$\sum_{i=1}^{N_{VBS}} R_{serv}^{VBS_i}{}_{[\text{bit/s}]} < \sum_{r=1}^{N_{RAT}} N_{RRU}^{RAT_r} \cdot R_{RRU_{max}\,[\text{bit/s}]}^{RAT_r} \tag{4.43}$$

where:

- $R_{req}^{VBS_i}$ - *VBS Requested Data Rate*, i.e., the total data rate requested by end-users in VBS $i$, given by:

$$R_{req}^{VBS_i}{}_{[\text{bit/s}]} = \sum_{n=1}^{N_{EU}^{VBS_i}} R_{req\,n[\text{bit/s}]}^{EU} \tag{4.44}$$

  where:

  - $R_{req\,n}^{EU}$ - data rate requested by end-user $n$.

One should note that if a GRT VBS is not using all the contracted capacity, its end-users must be served with the capacity they are requesting, i.e., if a given GRT VBS serving data rate is below the contracted capacity, the RRUs allocated to its end-users must correspond to the data rate requested by them. The optimisation of radio resources utilisation is indirectly achieved by allowing the allocation of RRUs to any VBS, after all other VBSs in the cluster have their contracted capacity satisfied. This means that all available RRUs in the cluster are allocated to any VBS, as long as they have been requested, avoiding the waste of radio resources, e.g., due to a previous allocation to VBSs that did not use them. In fact, one is not dealing directly with the scheduling of the radio resources to the end-users, but rather indirectly, by enforcing the decisions taken from the cluster viewpoint to be considered by RRM algorithms.

## 4.6.2 Adaptive Virtual Radio Resource Allocation

Radio resource allocation is initially made by the pre-allocation of RRUs to VBSs, over the set of heterogeneous wireless systems available, with the aim of providing the minimum contracted capacity. It is assumed that the allocation of RRUs in a BS implies the instantiation of the VBS onto the BS with own requirements, although the VBS may be part of a VNet created within the cluster. The initial number of RRUs to allocate to a VBS (4.45), in order to match the VBS contracted data rate, is based on the maximum achievable data rate for the RRUs in each RAT, Figure 4.13, which corresponds to the maximum data rate the RRUs can perform, without interference or channel impairments, for the RAT in use.

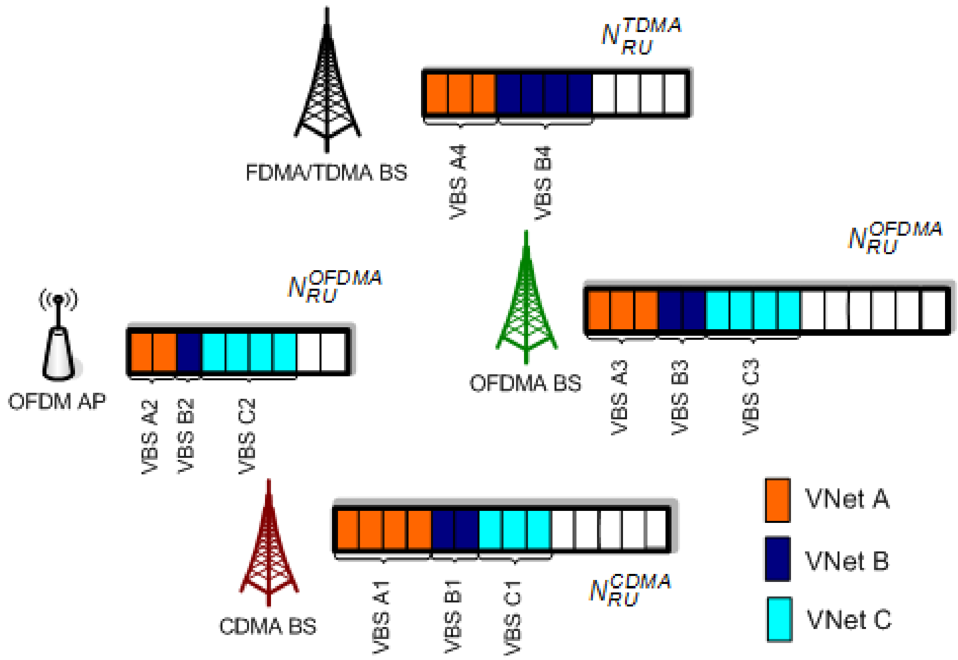



Figure 4.13. Initial number of RRUs per system and number of RRUs allocated to VNets.

The percentage of the VBS contracted data rate to be allocated on each RAT depends on the strategy used to instantiate the VBS, which should determine if the VBS is to be instantiated in all the BSs in the cluster, only in a limited number of BSs, or even in the BSs of specific RATs.

$$N^{VBS}_{RRU_0} = \sum_{n=1}^{N_{RAT}} \frac{\xi^{RAT_r}_R \cdot R^{VBS}_{min}\,[\text{bit/s}]}{R^{RAT_r}_{RRU_{max}}\,[\text{bit/s}]} \tag{4.45}$$

where:

- $N^{VBS}_{RRU_0}$ - initial number of RRUs allocated to the VBS;
- $\xi^{RAT_r}_R$ - percentage of contracted data rate to be provided by RAT *r;*
- $R^{RAT_r}_{RRU_{max}}$ - maximum data rate provided by one RRU of RAT *r*.

Knowing that RRUs data rate may change over time, the VBS data rate has to be evaluated periodically. The AdaptiveRRA is responsible for dynamically reallocating RRUs to reflect the network's operation condition, satisfying the VBS minimum capacity. The strategy used for the selection of the BS to reallocate the additional RRUs is based on two main criteria: the most adequate RAT for the provided services, and the BS with maximum available capacity.

The Adaptive-VRRA algorithm reacts to changes in capacity/availability of RRUs that affect VNet/VBS requirements, e.g., data rate, delay, and error rates. These changes are mainly caused by adaptive modulation and coding, to increase data rate for reliable transmission. The aggregated data rate of the VBS strongly depends on mobile terminals mobility, RAT type, distance to the BSs, and channel impairments, among other parameters. The computation of the VBS data rate capacity over time is then the sum of all the individual ones achieved in the RRUs assigned to end-users, added to the unused RRUs pre-allocated to the VBS according to its demand:

$$R^{VBS}_{serv}[\text{bit/s}] = \sum_{n=1}^{N^{VBS}_{EU}} R^{EU}_{serv\,n}\,[\text{bit/s}] + \sum_{r=1}^{N_{RAT}} \left( N^{VBS}_{RRU_r} - N^{VBS}_{RRU_{r,occ}} \right) \cdot R^{RAT_r}_{RRU_{max}}\,[\text{bit/s}] \tag{4.46}$$

where:

- $N^{VBS}_{RRU_r}$ - number of RRUs allocated to the VBS;
- $N^{VBS}_{RRU_{r,occ}}$ - number of RRUs of RAT *r* assigned to VBS end-users.

RRUs allocated to the VBS but not assigned to end-users are considered independently of the environment, and so the maximum RRU data rate, according to the specific RAT, is used for computation.

The Adaptive-VRRA algorithm uses monitoring information, SINR for data rate determination, in order to compare the current capacity with the contractual one, then, deciding on RRUs (re)allocation to a given VBS. The knowledge of the BSs in the cluster, the RRUs allocation to the VBSs, the co-located BSs, and the fundamental VBS's characteristics should be available. Operators/providers agreements are also important information that should be known indirectly, through the granted access to co-located BSs.

The number of occupied RRUs varies inversely to their data rate, for a constant offered traffic. The RRU data rate for the set of RRUs allocated to each end-user depends on its SINR, being smaller as the SINR decrease. Thus, the number of RRUs occupied by all end-users in the VBS changes, to follow the data rate requested to the VBS, being limited by the number of RRUs corresponding to the VBS contracted data rate. The maximum VBS serving data rate is also changing, as it is calculated by the sum of the data rates of end-users plus the unused RRUs pre-allocated to the VBS.

Whenever the VBS capacity is below the contracted minimum one, a compensation mechanism is evoked to perform the selection of additional RRUs. The selection is made among the co-located BSs, according to the BSs' availability and cost, Figure 4.14.

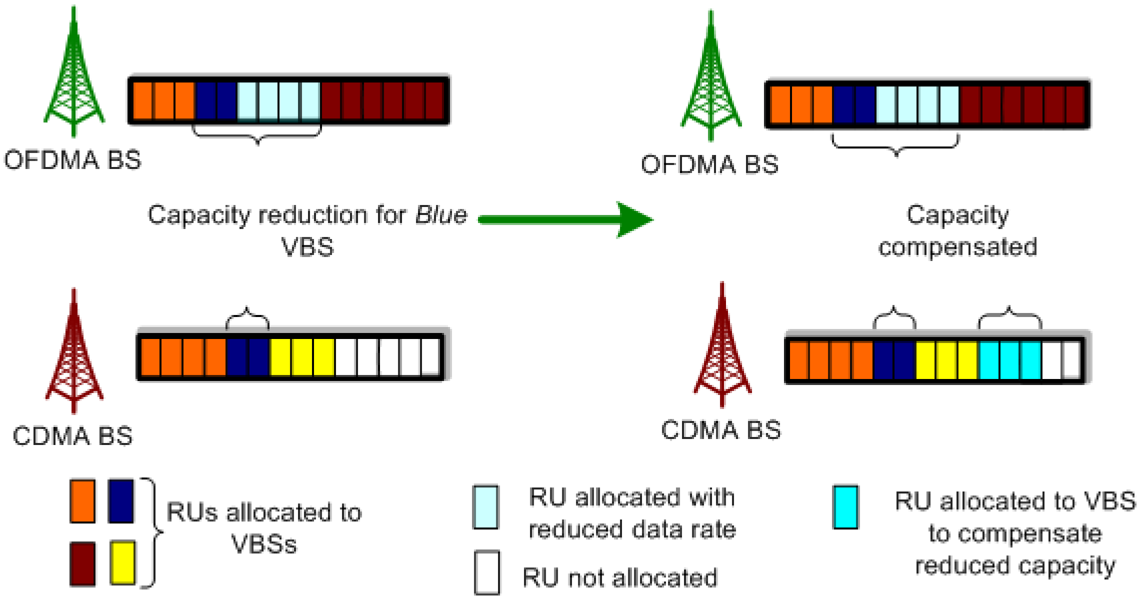
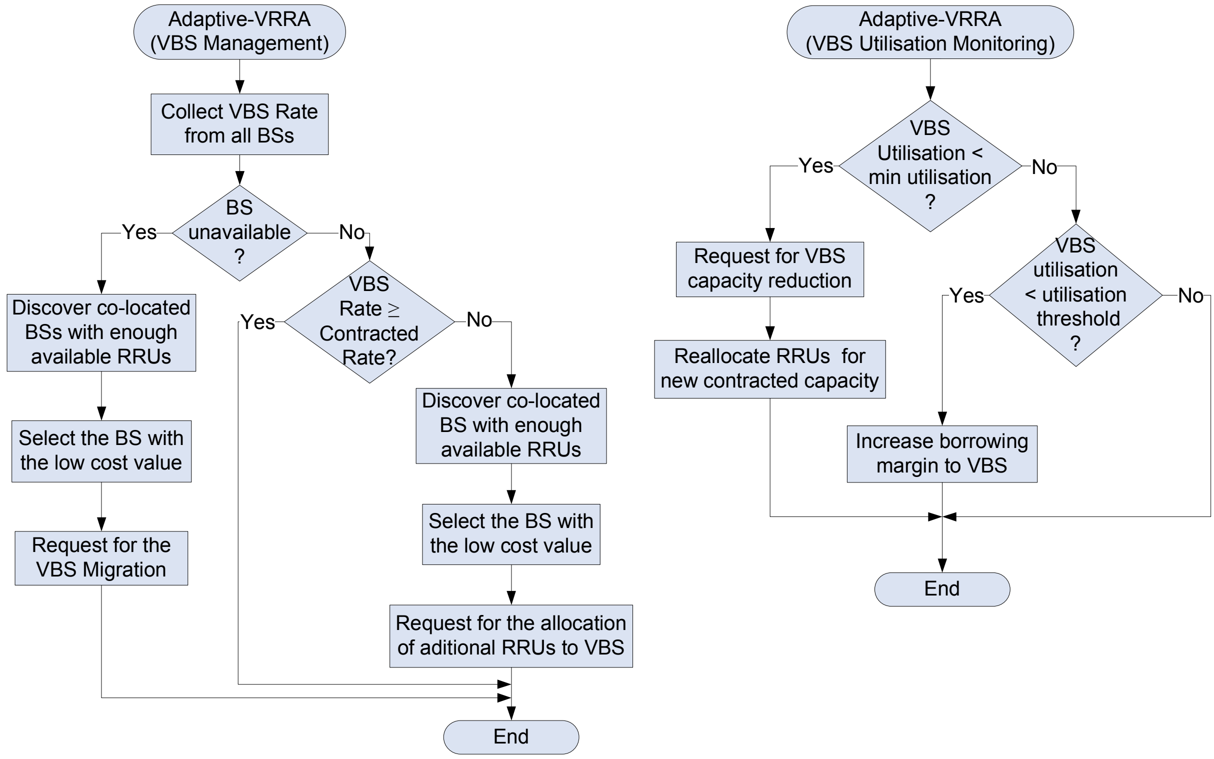



Figure 4.14. Radio channel reallocation in a neighbour physical resource.

Two main situations can happen at this point: the VBS can be span to multiple co-located BS (allowed by the existent policies) in order to get the VBS required capacity, or if the BS in which the VBS is instantiated becomes unavailable, the total capacity required for the VBS should migrate to another BS(s). These changes can affect VNet-RRM, which must be informed.

Concerning the availability computation, besides the unallocated RRUs, a VNet borrowing margin, similar to the one defined in [AMSE06], is considered, and determined by the VBS type, which is adapted according to the VBS usage. As an example, in a BE VBS, RRUs may be transferred (borrowed) to perform the total amount of data rate required by a GRT VBS, if no other RRUs are available. The opposite is only possible if the GRT VBS is running on low usage.

It is important to note that this evaluation is performed by InPs, essentially to support the decision to select the best BS, in which RRUs will be allocated to VBSs. The scanning time of this decision process is adapted dynamically, depending on resources utilisation, variability of the radio interface, and VNets characteristics. Depending on VBS utilisation and VBS type, Adaptive-VRRA may also decide on the migration or adaptation of the amount of RRUs allocated to the VBSs in order to optimise radio resource usage, e.g., when the VBS operates on low usage over a long period of time. The flowchart presented in Figure 4.15 depicts the Adaptive-VRRA algorithm: Figure 4.15(a) presents the VBS Management procedure and Figure 4.15(b) the process related to the VBS utilisation monitoring.




(a) VBSs Management (b) VBS Utilisation Monitoring

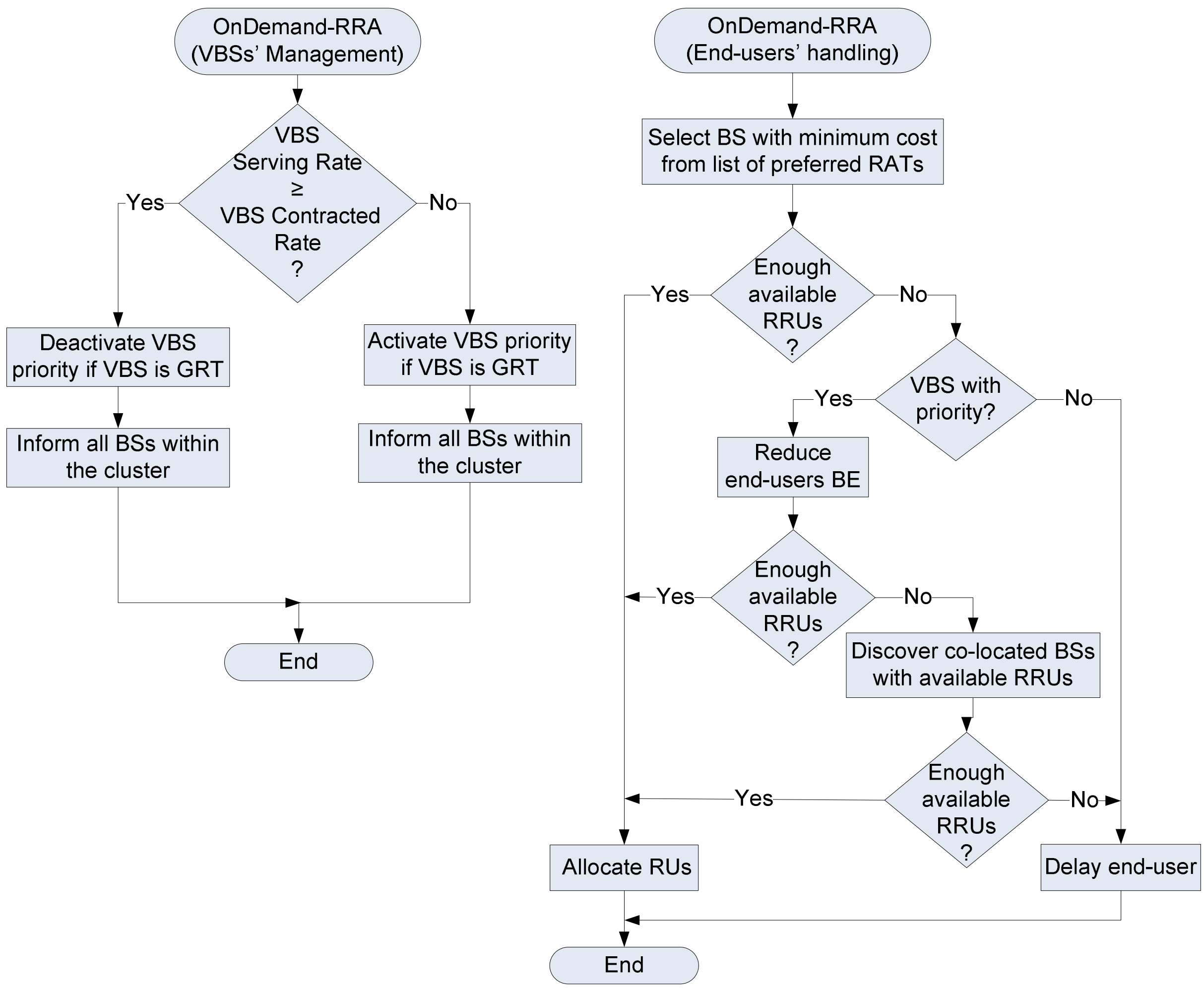
Figure 4.15. Adaptive-VRRA.

### 4.6.3 On Demand Virtual Radio Resource Allocation

OnDemand-VRRA is responsible for dynamically (re)allocating RRUs to reflect the network operation condition, satisfying the VNet minimum capacity. This is supported by a VNet priority scheme and a data rate reduction strategy, besides the access selection mechanism.

Concerning access selection, end-users are connected to the different VBSs according to the requested service and their contract with the VNOs. The physical connection is established over one of the existing RATs in the coverage area, according to a list of preferences related to the requested service, the available capacity, and the strategy defined for resource evaluation. This strategy, e.g., minimum load and/or cost, is based on the BS cost, where several KPIs are weighted. Within the scope of this thesis, the strategy used is the minimum load or maximum availability.

The VNet priority scheme, running at cluster level, assumes a coordination role and enables to set differentiated end-users according to the type of VNet and $R_{serv}^{VBS}$, Figure 4.16 (a). VBSs are initialised to be handled with priority, all BSs in the cluster being informed of this, to activate the data rate reduction process. When $R_{min}^{VBS}$ is reached, the priority to be given to end-users who wish to connect to this VBS is deactivated. This priority scheme based on $R_{serv}^{VBS}$, allows one to implement a data rate reduction strategy whenever GRT VBSs have priority, preventing starvation on BE VBSs when the contracted data rate in GRT VBSs is reached.



(a) VBSs Management (b) End-users handling (adapted RRM functions)

Figure 4.16. OnDemand-VRRA algorithm.

The data rate reduction strategy is essential to compensate possible end-user data rate decrease due to degradation of medium conditions, being applied to services with a minimum required data rate when the VBS operates within the contracted capacity, i.e., when the VBS priority is activated. The adopted data rate reduction strategy is as follows. Whenever the VBS priority is activated for a GRT VBS, and the end-user tries to connect to a BS in which there are not enough RRUs for providing the required services, BE end-users connected to the BS are reduced according to:

- the QoS priority class of the performed service [IEEE05b], end-users performing services with lower priority being the first to be reduced;

- SINR, end-users with lower one being reduced first, allowing to optimise radio resource usage.

Still, if there are not enough RRUs to reach the requested data rate, the RRM or the Cooperative RRM is requested to do the evaluation of co-located BSs, in order to select the one with enough RRUs available and with the minimum cost to handover end-users. The end-users handling process is depicted in Figure 4.16(b).

It is worthwhile to note the difference between OnDemand-VRRA and the radio resource allocation and adaptation mechanisms at the MAC level, which deal with end-user performance instead of the VBS one. OnDemand-VRRA acts as a coordinator that enforces its VRRA decisions onto RRM functions, namely, RRA and admission control, for the RATs within the cluster that should be adapted to receive these settings. Information, such as end-user VBS and priority of the VBS should be known to those RRM functions, in order to be taken into account on admission and assignment of RRUs to end-users.

## 4.7 OnDemand-VRRA Model on OConS Architecture

OnDemand-VRRA was modelled according to the OConS architecture, in order to take advantage of its flexible approach, e.g., concerning the activation and configuration during network operation. One DE has been identified in the Cluster Manager (CM) that is responsible to manage a given set of BSs, and local resource management is performed by other DEs per BS. The former is responsible to apply the priority scheme described in Section 3.1, and to reallocate RRUs in co-located BSs for vertical handovers; the latter, based on the VNet priority scheme, implements the OConS Supported on Demand Radio Resource Allocation for Virtual Connectivity data rate reduction strategy. An additional DE is taken at the User Equipment (UE), to deal with the access selection mechanism; although it can be external to OnDemand-VRRA, it has been also considered within this work. Figure 4.17 illustrates the mechanism mapping, the numbers in the boxes being a possible sequence of steps produced.

When an OConS user connectivity request is received, via OSAP, the Service Orchestration Process handles and instantiates or (re)configures the OnDemand-VRRA mechanism for the new connectivity requirements (1), e.g., QoS type for the virtual resource, minimum data rate contracted or delay. Connectivity requirements are passed onto the cluster manager DE (2), which activates the priority of all VNets in the cluster by sending this information to the several DEs in the BSs (3). An end-user requests to initiate a service/application (4), activates the access selection mechanism on the UE, which according to link performance indicators (5) and the RAT priority list (6), information gathered from the IE, decides the initial RAT selection (7) and establishes the connection (8) enforcing the decision in the corresponding EE.

The CM receives data rates requests from all end-users in the VNets (9), compares the VNet serving data rate and the contracted one, with the information in the IE (10), and decides the VNet priority (11). The result of this decision is then sent to all BSs to set the VNet priority (12), accordingly. The BS DEs use the KPIs from the IEs (13), e.g., wireless rate and usage, to run the data rate reduction mechanism, and decide the reallocation of RRUs to end-users connected to the BS (14). The decision can be to keep the end-user in the same BS (15a), or to request the CM to try the radio resource allocation in a co-located BS (15b). To support the decision for reallocation of RRUs in co-located BSs (17), the CM requests information from the co-located BSs (16), in order to evaluate the best one to reallocate the RRUs (18a), informing the UE to change the connection to the new BS (18b).

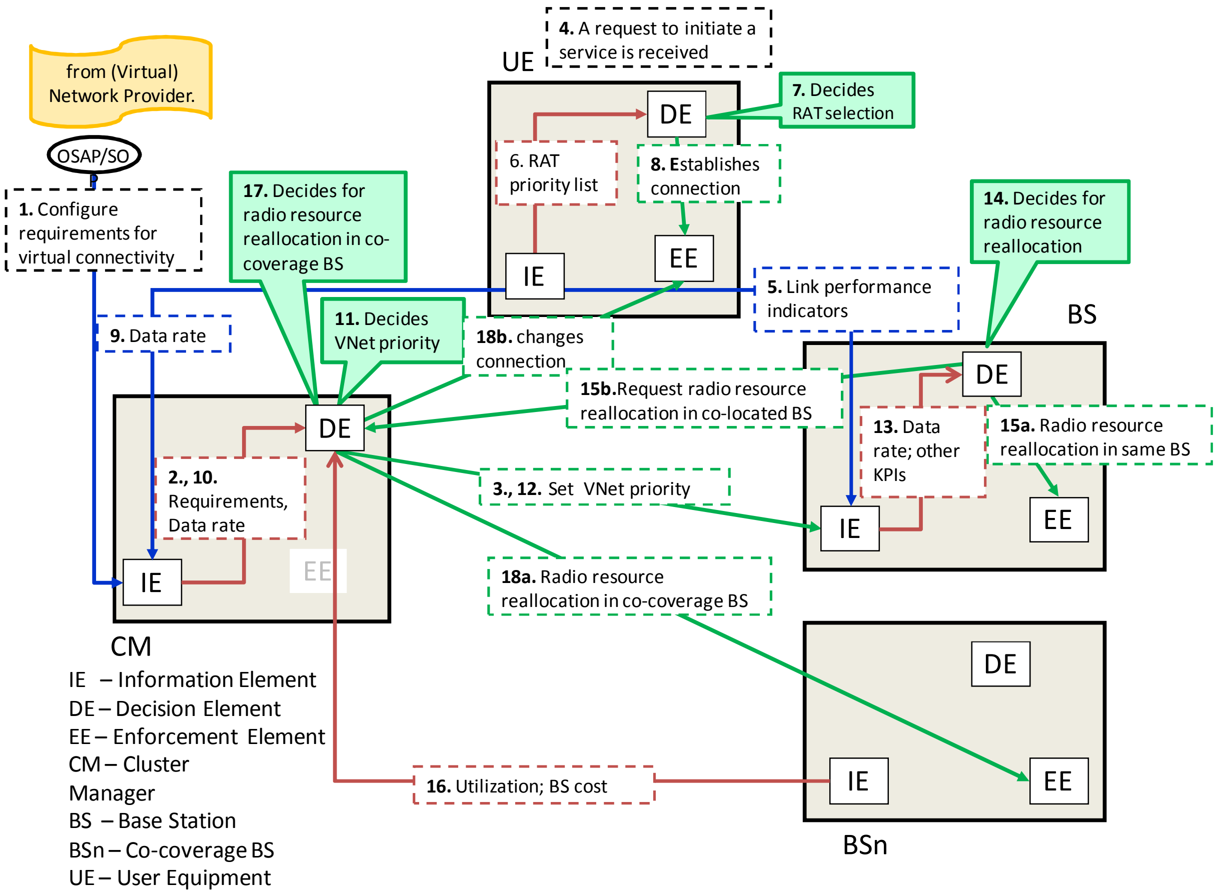



Figure 4.17. OnDemand-VRRA modelled according to the OConS Architecture.

The implementation of VRRA within the OConS architecture potentially brings the set of benefits as follows. Reconfiguration mechanisms allow streamlining the process of adapting at runtime the changes of requested capacity for the virtual resources. The communication capabilities among functional entities inherent to OConS nodes, allow to set triggers in the IEs of the group of nodes in the cluster, actuate over the several schedulers, and also inform the cluster manager automatically about changes occurring within the cluster. The OConS capability to launch a service composed of several mechanisms allows to instantiate, e.g., VRRA with an enhanced access selection mechanism to optimise the resource utilisation according to end-users policies while providing the capacity requested for the virtual resources.

## 4.8 Integrated View

This section summarises the various concepts and proposals introduced throughout Chapter 4, so that one can have an integrated view of what is being proposed.

In Section 4.1, the network architecture is depicted. Focused on the virtualisation of the wireless access, it is based on the generic network virtualisation environment presented in Section 3.1. The physical and logical perspectives of the network have been introduced, as well as the mapping between the physical and VLinks, which mainly aims at translating the demand for capacity in the allocation of RRUs. The physical networks under study and the differentiation of the VNets based on its requirements, as they are considered in this work, are also presented in this section.

The approach used for RRM in virtualised environments is presented in Section 4.2. Two levels of RRM are considered: the Intra-VNet RRM concerns on how the radio resources of a particular VNet are shared among its end-users, while the Inter-VNet RRM aims at managing the set of radio resources shared among VNets. The latter is the main topic of this thesis, and considers a cooperative management of the radio resources from all the heterogeneous wireless networks serving a given area and being shared among the VNets instantiated on that area. This cooperative RRM strategy, designated as CVRRM, brings the main functions of CRRM to the virtualisation context namely, the initial access selection, the scheduling of the radio resources among the heterogeneous wireless networks, and the vertical handover. This set of functionalities is applied to handle the various VNets as an aggregated resource instead of the individual end-users, resulting on the *Initial VNet Selection*, the *Virtual Radio Resource Allocation* and the *VNet Handover Support* functions of CVRRM.

The main assumptions taken for the model are presented and the needed inputs identified in Section 4.3. In brief, the RATs currently used in the most common systems, namely, FD/TDMA, CDMA, OFDM and OFDMA are assumed. The specificities of the RRUs of each RAT are abstracted by a generic RRU, although the diverse characteristics are taken into account. Regarding the VNets, GRT and BE VNets are considered according to the agreed QoS guarantees per aggregated VLink.

The description of the proposed model is made in Section 4.4. First, the analytical model is presented, by defining the main physical and virtual components and associated parameters in order to obtain VNet's capacity for the network architecture. The main theoretical assumptions for data rate estimation are also presented in this section.

Section 4.5 identifies the evaluation metrics that allow quantifying the benefits of introducing the proposed algorithms. Metrics to evaluate the performance of the virtual and the physical resources have been defined, e.g., *Average VBS*, *Cluster Serving Data Rate*, and VBS *Out of Contract*.

Section 4.6 is devoted to the strategies and algorithms proposed for the implementation of CVRRM. For the VRRA, two algorithms are proposed and implemented in the developed simulator. A first one, designated by Adaptive-VRRA, does a pre-allocation of the RRUs to the VBSs according to the contracted capacity over the set of heterogeneous systems available, adapting it to compensate the wireless links' variations. A second, called OnDemand-VRRA, allocates the RRUs only if they are requested by the VNet end-users, still adapting the RRUs allocation to reach the VBS contracted capacity. The strategies used by CVRRM are related to the contractual VNet requirements, and are reflected by KPIs weights in the CF computation for resources evaluation.

An algorithm to manage the allocation of radio resources from different RATs to the VBSs, OnDemand-VRRA, is proposed. The main target is to provide the required capacity (data rate) to the VBSs, according to the type of guarantees of the VBSs and the VBSs' utilisation, maintaining isolation among the VBSs. Taking into account the variability of the wireless medium, the algorithm continuously influences RRM mechanisms, namely admission control and MAC scheduling, to be aware of the VBSs' state relative to the service level agreement.

Finally, in Section 4.7, the integration of OnDemand-VRRA on the OConS architecture is presented, with the purpose of highlighting the advantages of being deployed over such kind of open and flexible approaches. The OConS capability to orchestrate an adequate connectivity service, the ability to accept changes in the service configuration, and the continuous availability of the network state makes it ideal to support our wireless access virtualisation approach. In fact, on the one hand, virtual capacity requests can be issued and modified at any time, and on the other hand, the state of the network must be known in order to allow the dynamic adaptation of RRUs allocation to the VNets.

It should be referred that the basis of our approach is aligned with the set of use cases defined by the 3GPP group for RAN Sharing Enhancements presented in Section 3.3, if one maps the Hosting RAN provider onto the InP, the Participating Operators onto the VNOs, and the usage portion of the Hosting RAN onto the contracted capacity. Also, monitoring functions per VNO, reductions in resource allocation per VNO according to its contract, and On-Demand capacity provision, which have been assumed to build our model, are planned in these enhancements.

# Chapter 5

# Models Implementation in a Simulator

This chapter aims at presenting the most relevant functional blocks proposed and implemented into the VRRA simulator. The main assumptions taken and the details of the implementation in a simulator are presented. The simulator assessment strategy is explained and results are presented.

# 5.1 Simulator Overview

The implemented VRRA simulator uses system level principles, since the target is not to evaluate physical layer performance but rather to study higher layer events and interactions, exploring the behaviour of the whole network with a predefined number of BSs and uniformly distributed end-users. It is divided into three main functions identified by the grey, blue, and green blocks in Figure 5.1.

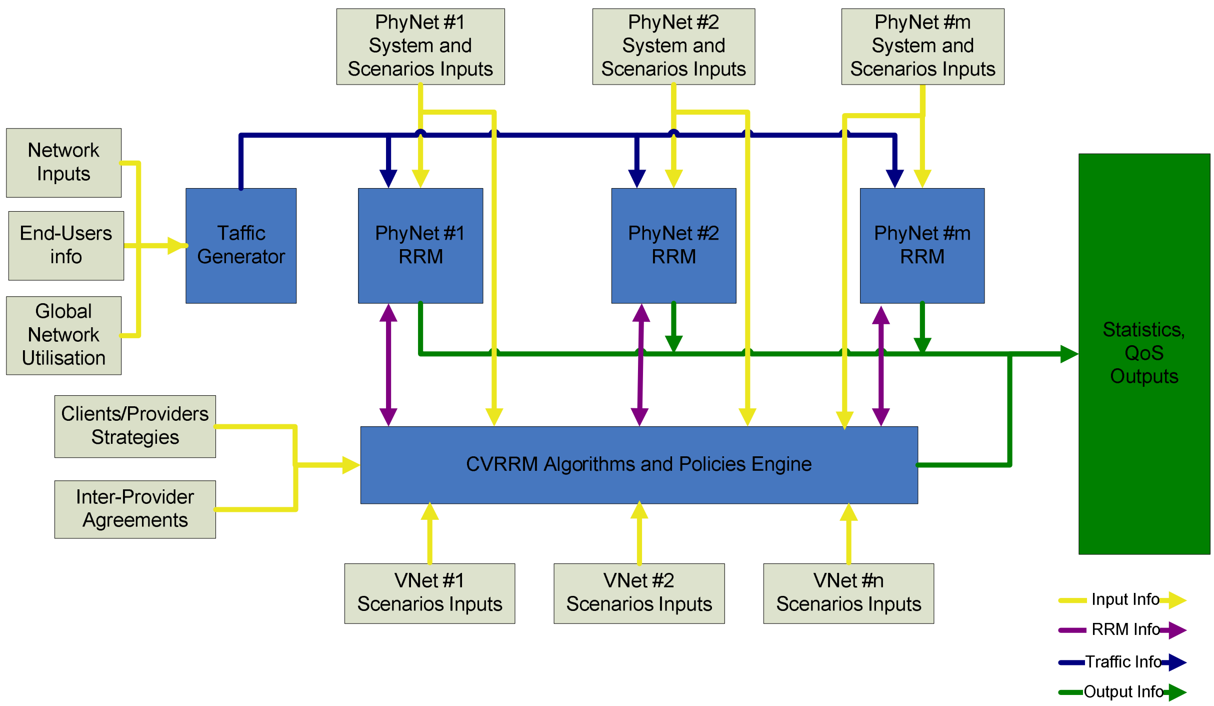



Figure 5.1. Simulator block diagram.

The input blocks, in grey, are dedicated to the following inputs:

- Global Network Utilisation: service penetration.
- Network Inputs: Service profiles, services rates and duration, etc.
- End-Users Info: number of users, type of SLA, preferences, contracts, etc.
- PhyNets Inputs: SINR thresholds, Data Rata as a function of SINR threshold, number and location of BSs, etc.
- CVRRM Strategies/Policies Inputs: CF weights and QoS parameters, for each type of multiple access technology.
- Inter-providers agreements inputs: inter-dependencies among the several providers, i.e., end-users of VNet #1 can use the wireless infrastructure of InP A and B.
- VNet # 1 up to VNet # m: Scenarios inputs per VNet.

The blue set of blocks is where most of the simulation computational effort is performed. These blocks have the following functionalities:

- Traffic Generation - traffic information vectors of all end-users and services are built.
- PhyNet #1 RRM up to PhyNet #m - perform the fundamental functionalities of a specific VNet, by running/managing and monitoring the radio links conditions and services attached (generated by the Traffic Generation block), thus, requesting a significant computational effort.
- CVRRM Algorithms and Policies Engine - being common to all VNets, this block is requested many times for control of VNet requirements, and runs the CF, which is related to all Virtual and physical active resources in the scenario.

Finally, the green block is where the selected output parameters are displayed, most of them being QoS and system statistics at RRM and CVRRM levels. These parameters can be used to extract others, by establishing logical relations among them.

The general algorithm of the simulator is depicted in Figure 5.2. The processes for handling end-users, for managing the allocation of radio resources to the virtual ones, and the computing processes supporting the decisions are identified.

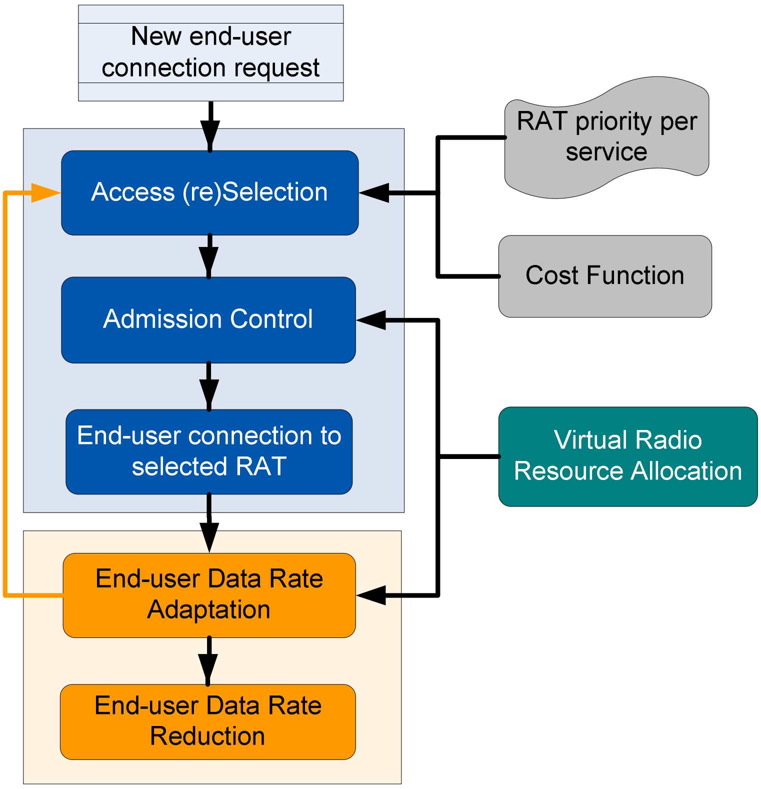



Figure 5.2. Simulator flowchart.

The processes related to “Access Selection” and “Admission Control” for end-user attachment are in blue. These processes are evoked whenever a new end-user connection request is received or implicitly due to network operation needs. The access selection is supported by the “RAT priority per service list” and by the CF process, which is responsible to compute the resource cost according to a pre-defined strategy.

The processes of adapting the RRUs assignment to the end-users in order to overcome the variation in capacity of the wireless systems are presented in yellow: “end-user Data Rate adaptation” and “end-user Data Rate Reduction”. These processes are essential to monitor and adapt the end-user served rate in order to preserve the minimum service rate.

The “Virtual Radio Resource Allocation” is the process in which the VRRA algorithms presented in Sections 4.6.2 and 4.6.3 are implemented. This process influences “Admission Control” and “end-user Data Rate adaptation” processes, according to the utilisation of the virtual resources capacity. It is worth to note that, in Section 4.6.3, end-users handling is presented as part of the OnDemand-VRRA algorithm for better explanation. However, generally, the “End-users handling” process should be considered as a separate process that, in case of OnDemand-VRRA implementation, is strongly influenced by its decisions.

A representation of the VRRA simulator to include the perspective of the network architecture model presented in Section 4.1 is depicted in Figure 5.3. VNOs and InPs are represented, since they are key players in the VNet operation phase. In fact, VRRA should track both virtual and physical resources to reach its main target, support the guaranteed contracts established between InPs, and VNOs, which are the VNOs requirements. Moreover, the simulator diagram depicts the several components (BSs, VBSs and VNets) with their attributes and the processes implemented in the VRRA simulator. Other configurable inputs needed for simulation are also presented.

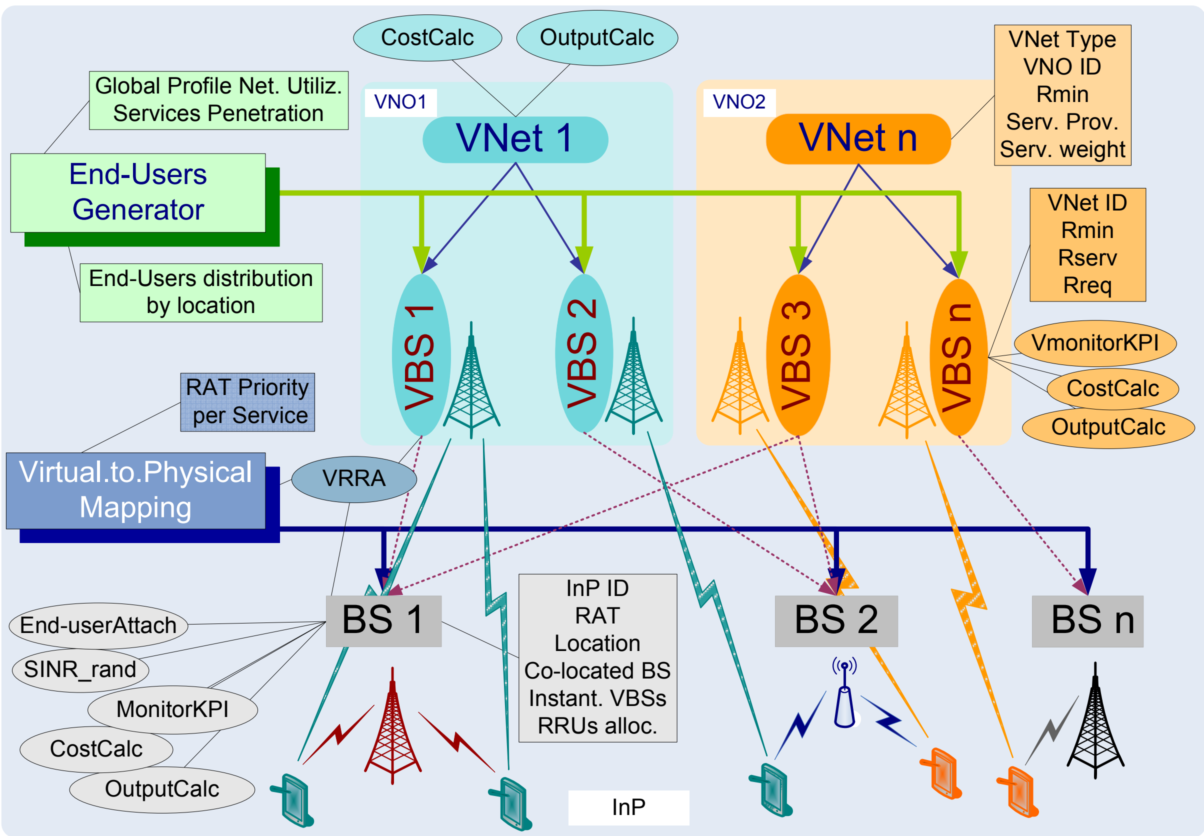



Figure 5.3. VRRA simulator diagram.

At the InP level, BSs' capabilities are mainly identified according to their RAT. Based on the CVRRM model described in Section 4.4, one assumes one cluster of heterogeneous wireless networks serving a given geographical area. BSs are then modelled by the number of RRUs, the SINR thresholds, and the related data rates per RRU, specific for each RAT under consideration. The identification of the VBSs instantiated on the BS, as well as the number of RRUs allocated to each one, are pre-configured attributes that naturally change over time (operation dynamics). As a particular case, one may only define one VBS instantiated for all BSs in the cluster by allocating a residual number of RRUs in each BS. In addition, the location (coverage zones) and the InP ID (which may take the role of network operator in particular scenarios) are also attributes of the BSs.

At the VNO level, VNets are characterised by their contracted capacity (minimum data rate) and their type, intending to reflect QoS requirements, the services they provide, and the weight they have on the global service provision. The VNO identification is also an attribute of the VNet. VBSs are defined by their contracted capacity (minimum serving data rate), and the VNet it belongs to, inheriting from it the type and set of services provided.

In order to define network traffic, the number of end-users is initially set and distributed by service, in accordance to a global network usage profile. End-users are distributed by VNets, taking the services provided by each VNet into account. The end-user location is randomly determined, based on the percentage of end-users per location received as an input to the simulator. Finally, the mapping between VBSs and BSs associates each end-user to the most suitable physical resource, considering the end-user location, the "RAT Priority per Service" list, and the VBSs' and the BSs' costs, which reflect their operational situation.

In order to reduce the complexity of system simulations, one assumes that equal transmit power is allocated to each RRU, the same MCS is applied to all RRUs assigned to each end-user, and all transmitted packets are received correctly. Moreover, dynamic channel variations and mobility of end-users are considered implicitly within the simulator by imposing the variation of the end-users' signal level. Network conditions that can cause reduction in the capacity of virtual resources have been forced, by setting a high percentage of active end-users and dynamically changing their SINR in order to reflect wireless medium variation ("SINR_rand" process in Figure 5.3). End-users are entering and leaving the network according to the inter-arrival time and the mean duration of their services; both time intervals are based on an exponential statistical distribution. The process for end-users handling, "end-user Attach", is biased by the VRRA algorithm according to the operation state of the VBSs.

KPIs monitoring, "MonitorKPI", and CF computation, "CostCalc", are performed periodically at BS and VBS levels. These two processes provide the necessary inputs to "end-user Attach" processes. The "OutputCalc" process is responsible to compute the metrics for assessment presented in Section 4.5, being implemented at BS, VBS and VNet levels.

VRRA is implemented at the physical and virtual levels, in order to allow using the capacity requested among all the BSs within the cluster. At the virtual level, it manages the VBSs based on the type of

VNet and the comparison between the serving rate and the minimum contracted one. At the physical level, it influences end-user attachment according to the VBS utilisation and the situation of the others VBSs instantiated within the cluster.

# 5.2 Simulator Implementation

## 5.2.1 Traffic Generation

Traffic characterisation is beyond the scope of this work; however, since the proposed algorithms consider different types of network services in order to react differently, a simple traffic model is defined for each service. Therefore, instead of the three tiers for service characterisation as defined in [HaGB05], namely, session, activity and packet levels, Figure 5.4, only the session level is used to model the arrival process of users into the network.

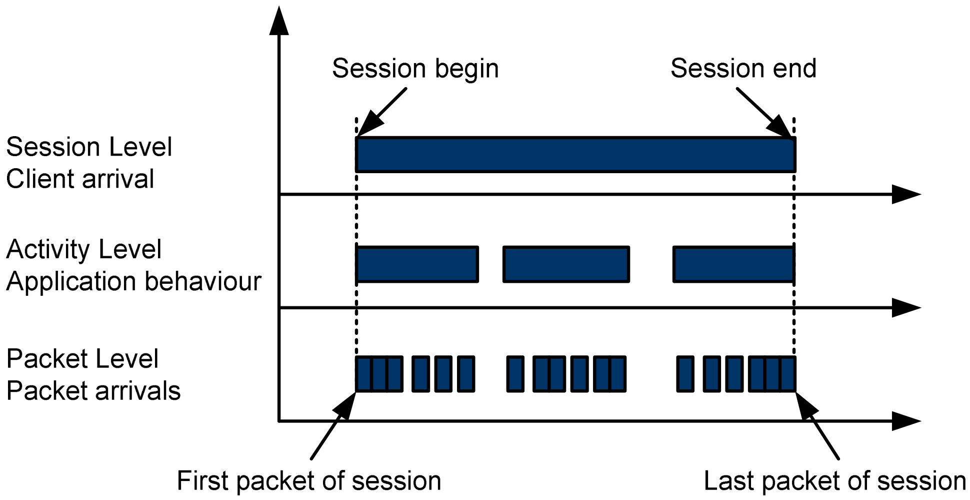



Figure 5.4. Generic Traffic Source Model (extracted from [HaGB05]).

Users start an application/service, use it during some time or to transmit a certain data volume, and then disconnect from the network. This is the typical behaviour of the population of users/clients at the access nodes, being aligned with the defined network model in which access nodes are the BSs in the cluster. The time at which users arrive (beginning of a new session) can be described by a statistical distribution of inter-arrivals into time. The Poisson model (birth death process) is one of the most popular ones to represent the arrival of end-users. This model represents an average rate of arrivals, without memory, independent of previous ones. The session duration in real-time applications, e.g., VoIP, is characterised by a statistical distribution depending on the type of application, while in non-real-time applications, e.g., Web or File Sharing, the traffic source is modelled by the quantity of information to transmit, i.e., the session data volume; in this case, the session duration depends on the network response and not on the source itself [HaGB05]. The total amount of traffic for one session, characteristic of non-real-time services, is assumed to be transmitted continuously after the service is initiated. The same principle is applied to real-time services characterised by its duration, in which the average session time is reduced, because inactivity periods are omitted. The specificity of the services is inferred from the mean service time and the typical rate of service requests, Table 5.1. Although this is not a completely realistic approach, it reflects a diverse utilisation of the network related to the several services under consideration, hence, allowing algorithms assessment.

Table 5.1.Traffic characteristics (based on [KILL01] and [Seba07]).

| Services | Class of service | Inter-arrival time [s] | Data Volume [MB] | Service time [s] |
|---|---|---|---|---|
| **VoIP** | Conversational | Exp(λ) | Not applicable | Uniform [a, b] |
| **Video** | Streaming | Exp(λ) | Lognormal [μ, σ] | Not applicable |
| **File Sharing** | Background | Exp(λ) | Lognormal [μ, σ] | Not applicable |
| **Web/data** | Interactive | Exp(λ) | Lognormal [μ, σ] | Not applicable |

The most usual services from each of the classes defined by 3GPP, Section 2.3, have been grouped in line with the services definition presented in [Cisc12b]. The following groups have been derived:

- Web/Data - Web, and other data traffic (excludes file sharing);
- File sharing - peer-to-peer traffic;
- Video - video calling and video streaming;
- VoIP - traffic from retail VoIP services and PC-based VoIP.

For each group of services, the session data volume average was computed among the several applications included in the group. The statistical distribution used for the data volume generation in each session is the Lognormal distribution [KILL01]. The time duration of a VoIP session is considered to obey to a Uniform distribution. The service inter-arrival time represents the elapsed time till a new session is initiated or a new end-user starts the service.

It must be referred that for VoIP and Video services, a minimum service data rate, $s_{min}^{R}$, has been considered, bellow which the services cannot be provided.

## 5.2.2 End-users Generation

End-users are generated at the beginning of a simulation, in accordance to a pre-defined number; only a percentage of them, also configurable, are active at start-up time. Static end-users' attributes are set randomly at the beginning of simulation, according to the simulation scenario: location, service to be requested, VNet, VNO, and network provider. All these parameters are subject to configurable percentages of service profile and of end-users per location and per operator provider.

The heterogeneous wireless cluster is divided in zones, corresponding to the intersections of BS coverage areas. End-user's location is assigned according to the pre-defined percentage of end-users per zone. This location is unchanged during the simulation, allowing end-users to access the network potentially through all the BSs in their assigned zone, if no other restriction is applied. The service is assigned to the end-user according to the global network service profile, i.e., the percentage of end-users performing each service. The service requested by an end-user keeps the same over the simulation time span. End-users are allocated to one of the VNets providing the requested service, considering the relative percentage of end-users served by each VNet, which is an input for simulation. The end-user's VNO is derived from the allocated VNet, since each VNet belongs to a single VNO.

The end-user's network provider is assigned according to a penetration percentage of each network provider. As the penetration percentage, the number of network providers is also an input parameter for simulation. The end-user behaviour in the network is determined by the profile of the service he/she is performing, being defined by the session inter-arrival time distribution, and the session data volume or duration distributions, in line with the type of service, Section 5.2.1. End-users access the network according to the inter-arrival time and the mean duration of their services. The implementation of this behaviour in the simulator is illustrated by the state diagram of Figure 5.5. Four end-users' states have been defined:

- *Inactive* - end-user is waiting the service inter-arrival time runs out;
- *Active* - end-user is receiving service;
- *Waiting* - end-user is waiting for the next time frame to try to obtain the requested service;
- *Interrupted* - end-user service is delayed.

The transition among states can be summarised as follows. The end-user is *Inactive* during the inter-arrival time of the service; when the inter-arrival time runs out, the end-user goes to *Active* or *Waiting*, according to the availability of RRUs compared to the minimum RRUs required to perform the service. In the *Active* state, the end-user is being served until the service is complete, then returning to *Inactive*; during service the end-user can be delayed, because the minimum contracted data rate is not achieved, for GRT services, or because the data rate becomes zero, for BE ones, due to the data rate reduction strategy. When the service is delayed, the *Interrupted* state, the end-user may return to *Active* if the data rate becomes greater or equal to minimum service data rate. If the end-user is in the *Waiting* state, he/she may go to *Active* if there are enough RRUs to achieve the minimum data rate.

It is assumed that end-users give up from the service, returning to the *Inactive* state, whenever the service delay or the delay on service request is greater than a timeout, configurable by type of service. In both cases, the end-user returns to *Inactive* and the inter-arrival time of the service is initialised. Although a timeout is defined, it is reset each time the end-user receives a number of RRUs greater than the minimum service data rate. Thus, to perform a service, the end-user can be delayed more than this timeout if he/she is alternating from *Interrupted* to *Active*.

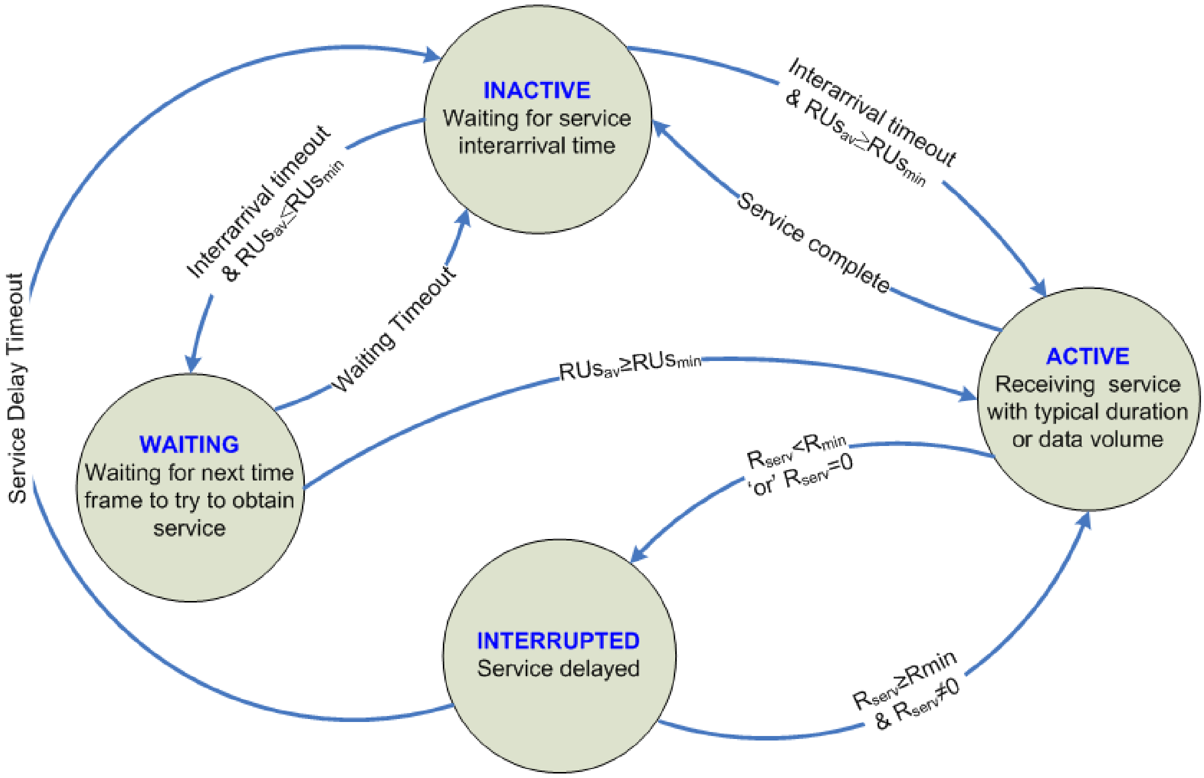



Figure 5.5. End-user state machine.

### 5.2.3 Algorithms' Implementation

As referred in Section 4.2.2, two VRRA algorithms have been implemented: Adaptive-VRRA and OnDemand-VRRA.

Adaptive-VRRA was implemented as described in Section 4.6, including three processes, "VBS Management", "VBS Utilisation Monitoring" and "Compensation Mechanism", contributing to decide the (re)allocation of RRUs to VBSs, Figure 5.6. "VBS Management" and "VBS Utilisation Monitoring" take the aggregated utilisation of the VBS into account, i.e., the capacity in use by all end-users connected to the VBS. This allows adapting the initial allocation of RRUs (based on the VBS contracted capacity) to VBS usage, enabling to optimise the physical resources utilisation. The "Compensation Mechanism" considers the BS capacity in terms of RRUs, the RRUs utilisation within the BS, and the RRUs pre-allocated to the VBSs that are still available, allowing the VBS to reach the contracted capacity.

"VBS Management" runs at the cluster management level, virtual to physical mapping, because it should have the knowledge of all VBS requirements and the VBS instantiation within the cluster, i.e., how many RRUs per BS are allocated to the VBS. It receives from the "MonitorKPI" (see Section 5.1) the VBS serving data rate in each BS, in order to compare $R_{serv}^{VBS}$ with $R_{min}^{VBS}$ or $R_{ref}^{VBS}$, depending on the type of the VBS. "VBS Management" interacts with the "Compensation Mechanism", since it evokes this last process whenever detects that the VBS is out of contract.

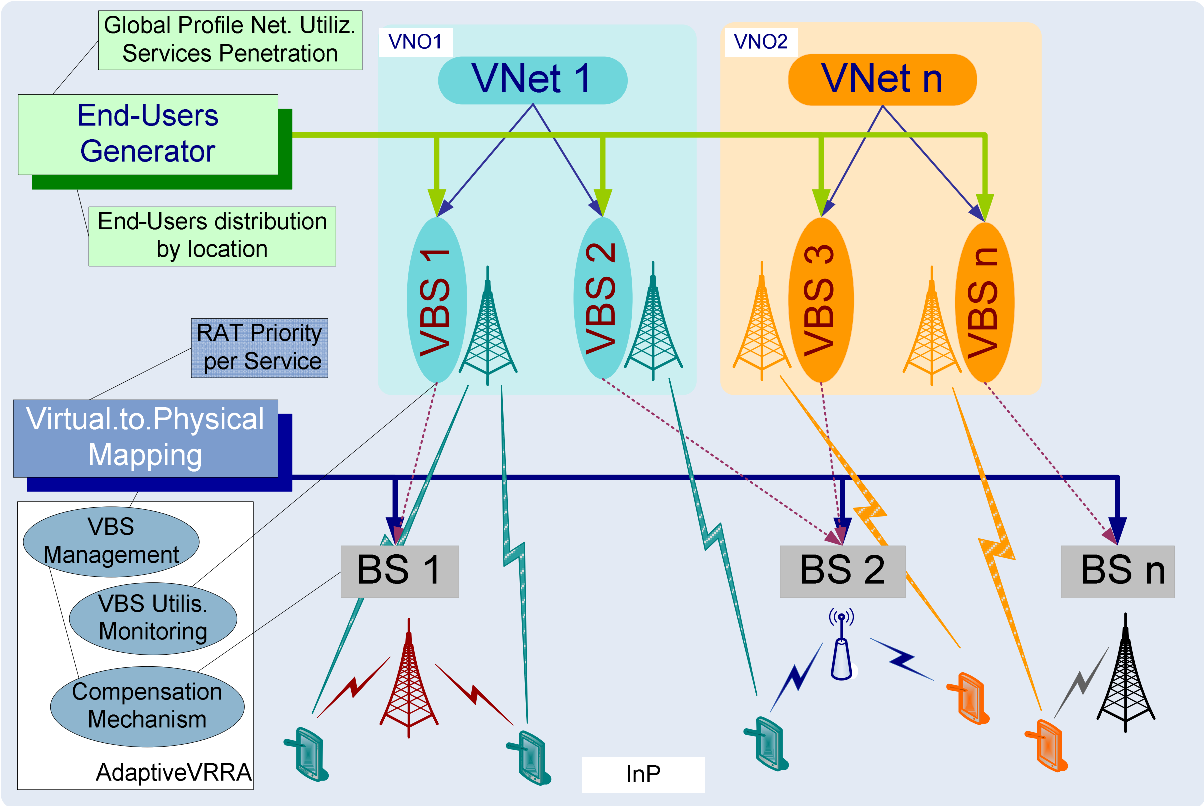



Figure 5.6. Adaptive-VRRA simulator block diagram.

The "Compensation Mechanism", based on the BS cost, evaluates the BSs within the cluster, in order to decide the more adequate one to allocate additional RRUs, in order to maintain the VBS contracted data rate. "VBS Utilisation Monitoring" monitors $\eta_{VBS}$, deciding to extend or shrink the VBS capacity according to its usage. A threshold value enabling to adapt the VBS borrowing margin to its usage level is also created, allowing to assign RRUs from VBSs that are not using them.

The OnDemand-VRRA algorithm was implemented according to the specification presented in Section 4.6. Two main processes, depicted in Figure 5.7, constitute the algorithm: "VBS Management" and "End_users Handling".

"VBS Management" has the same main purpose as the equivalent process of Adaptive-VRRA, i.e., monitors $R_{serv}^{VBS}$ deciding the reaction to take in order to maintain the GRT VBSs within their contract. However, since in OnDemand-VRRA RRUs are not pre-allocated to VBSs, instead of detecting if the VBS is out of contract it checks when $R_{serv}^{VBS} \geq R_{min}^{VBS}$, deactivating the VBS priority on the cluster in such situation. "End_users Handling" runs at the BS level and manages the assignment of RRUs to end-users. The decisions received from "VBS Management", basically the activation/deactivation of VBSs' priority, are taken into account whenever a VBS end-user tries to initiate or is using a service. The "Data Rate Reduction" supports "End_users Handling" by implementing the strategy described in Section 4.6.3, which is applied without limitations in legacy networks, but is only applied in VNet environments for GRT VBSs with priority, i.e., when $R_{serv}^{VBS} < R_{min}^{VBS}$ is verified.

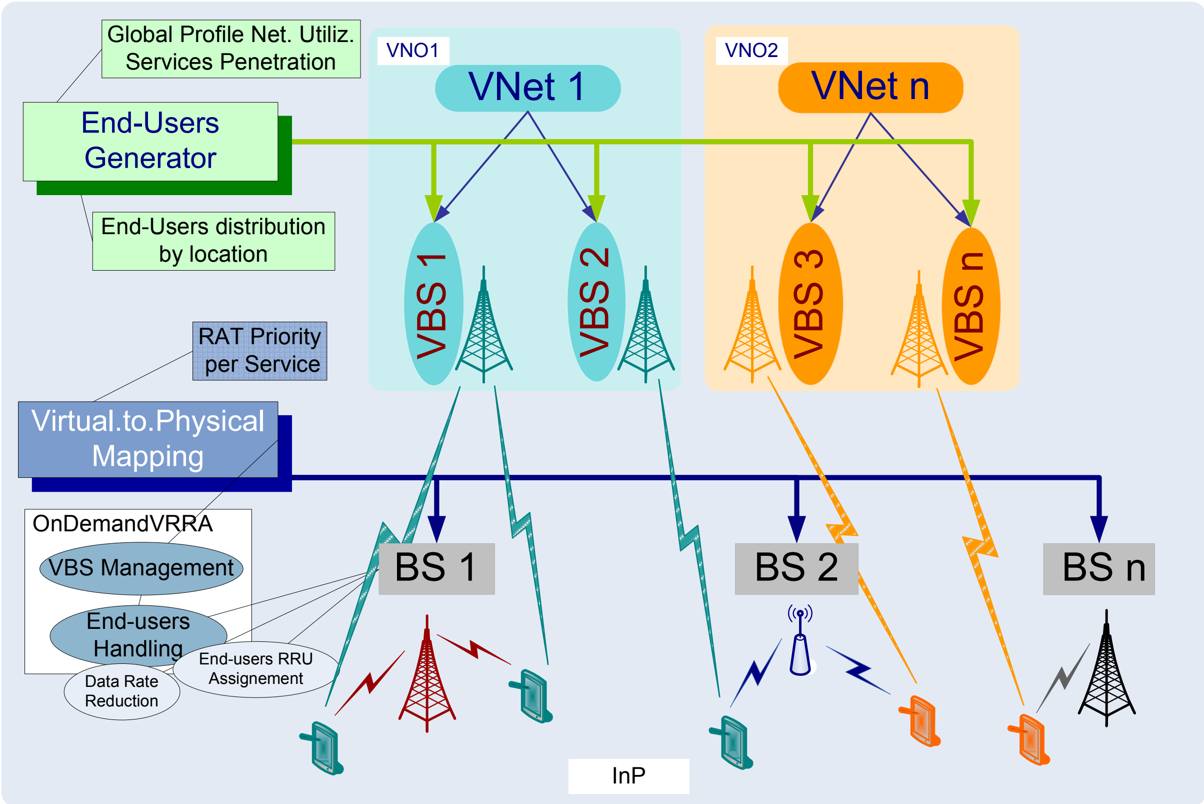



Figure 5.7. OnDemand-VRRA simulator block diagram.

The "End-users RRU Assignment" process is based on the well-known Admission Control and RRA functions, which have been modified to be aware of the dynamics of VBSs environment. In such a way, after the identification of the VBS to the end-user service request, "End-users RRU Assignment" admits or not the end-user, assigning the needed RRUs to perform the service, according to the priority of the VBS at that moment. Moreover, when the end-user is performing the service, the amount of RRUs is being adapted to the variations in the wireless medium, in order to achieve the minimum service data rate. However, this adaptation is constrained by the VBS priority, which reflects the operation state of the VBS.

OnDemand-VRRA can then virtually manage the allocation of RRUs to VBSs according to the VBS type, the operation state of all VBSs within the cluster, and in particular to the VBS in which the end-user request the service. Giving priority to the GRT VBSs operating under the minimum contracted rate, OnDemand-VRRA supports the $R_{min}^{VBS}$ achievement whenever it is requested. On the other hand, if GRT VBSs operate with $R_{serv}^{VBS} \geq R_{min}^{VBS}$, they lose the priority competing for RRUs at the same level as BE VBSs, or even losing priority over them when some $R_{ref}^{VBS}$ are defined for the BE VBSs.

# 5.3 Simulator Assessment

## 5.3.1 Analysis of Simulator Transitory Interval

The simulator assessment, besides the validation of all implemented blocks, was made in two steps: by analysing the transitory interval in the beginning of a simulation, and by investigating how many simulations are needed to have reliable output values.

The Reference scenario, defined in Section 6.1, was taken for simulator assessment. The deployment of a BE and a GRT VBS, with contracted data rates of 1.5 Gbit/s and 1.25 Gbit/s, respectively, and a total quantity of 8 000 end-users in the cluster were considered. The contracted data rates were chosen in order not to exceed the average physical capacity of the cluster and the amount of end-users, because the corresponding amount of requested data rate is greater than the contracted one in both VBSs, forcing the network to operate on extra capacity. The analysis is based on the following performance indicators, described in Section 4.5: *Average VBS Serving Data Rate*, *Out of Contract*, *Satisfaction Level on extra Capacity Requested*, *Average Delay on Service Request InP*, *Average VBS Time Service Delayed*, and *Average Cluster Serving Data Rate*.

The simulator transitory interval was analysed based on the relative deviation percentage computed as the relative difference to the last value of the simulation interval or, for the set of *n* simulations, to the average of all values collected for the total set of simulations:

$$\Delta_{[\%]} = \frac{|X_n - X_E|}{X_E} \quad (5.1)$$

where:

- $X_n$ - Value of parameter $X$ for time *n* or for *n* simulations;
- $X_E$ - Best estimate value of parameter $X$, considered as the last value of the simulation interval or the average of all values collected for the total number of simulations performed.

The evaluation of the simulator transitory interval was made by investigating the first 120 minutes of simulation. Although different numbers of end-users were considered, results are presented for only 8 000 end-users, since the simulator behaviour is identical for all situations.

From the values collected from simulations, two groups of parameters were identified: one having a similar behaviour in relation to the average value, from the beginning and over time, and another presenting an initial phase after which a stable value is reached. *Average VBS Serving Data Rate*, *Average Cluster Serving Data Rate* and *Out of Contract* are in the first group, while *Satisfaction Level on Extra Capacity Requested*, *Average Delay on Service Request InP* and *Average VBS Time Service Delayed* are in the second. *Average VBS Serving Data Rate* and *Average VBS Time Service Delayed* are presented in Figures 5.8 and 5.9, respectively, as example of each group, though the obtained values for the several parameters are represented graphically in Appendix C. The *Average VBS Serving Data Rate* for VNet GRT is not presented, as it is constant for the whole observation time.

It should be stressed that the points plotted in the graphs correspond to the values collected from simulations in each 1s time frame, since this is the granularity considered by VRRA. An exception is made for the *Out of Contract*, which that has been plotted as an average computed every 30 s, corresponding to 30 time frames, because it is an on/off value in each time frame.

For the first group of parameters, it can be said that just after the first time frames of simulation, the network behaviour is maintained. The transitory interval for these parameters is very short, and does not impose a major constraint. The relative deviation computed from (5.1), for the second group of output parameters, is presented in Figures C.5 to C.7; it can be observed that the relative deviation is less than or of the order of 10% for all parameters, except for *Average VBS Time Service Delayed* on VNet GRT when network time is greater than 20 minutes (see also Table C.1). *Average VBS Time Service Delayed* on VNet GRT presents a relative deviation of approximately 20%, even for time frames after the first 20 minutes: however, since the obtained values are in the order of magnitude of units of milliseconds, this accuracy can be considered acceptable for this parameter.

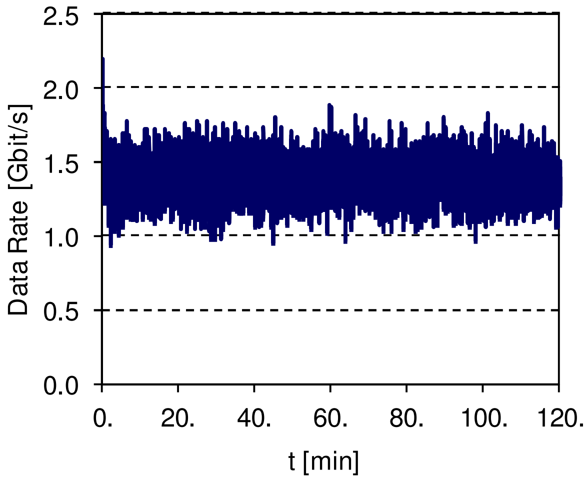
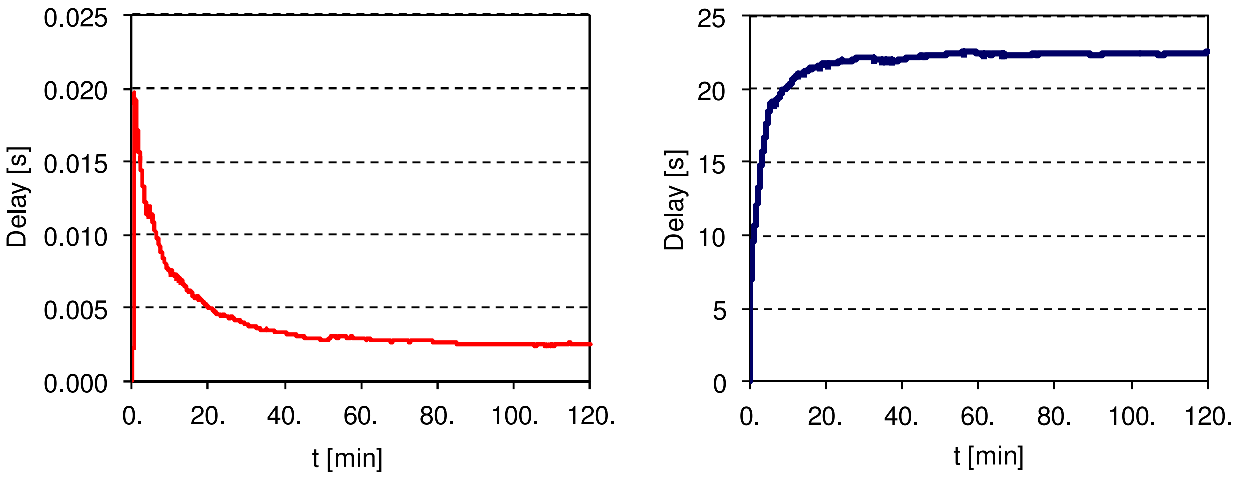



Figure 5.8. *Average VBS Serving Data Rate* over time (VNet BE).




(a) VNet GRT (b) VNet BE

Figure 5.9. *Average VBS Time Service Delayed* over time.

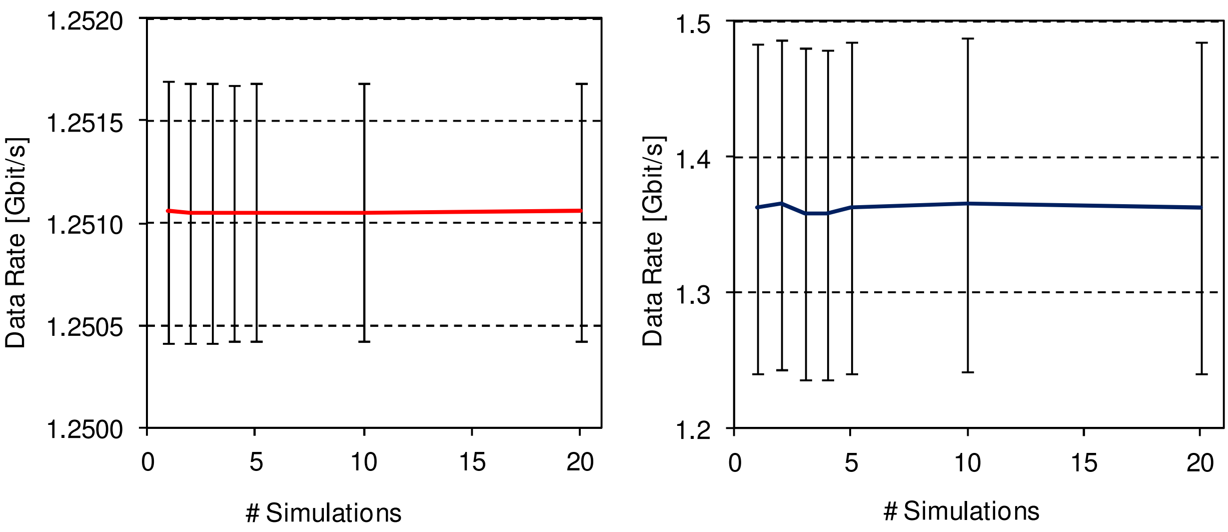
According to the relative deviation values obtained for the several output parameters, 20 minutes can be considered as the simulator transitory interval, hence, this initial time interval was not taken into account for algorithms' assessment through simulations.

### 5.3.2 Sensitivity Analysis as a Function of the Number of Simulations

To evaluate the impact that the number of simulations has on output values, a total of 50 simulations have been performed, with the duration of one hour network time after the initial transitory interval of 20 minutes, the output parameters being registered every second. For each run, the generation of random values is done according to a different seed, affecting the values of the following input variables:

- service inter-arrival time, data volume and service time, which change for each time the end-user request service;
- end-user SINR, changing in each time-frame;
- service assigned to end-users, which is fixed for each simulation;
- end-user location within the cluster, which is fixed for each simulation.

The results for the several parameters are depicted graphically in Section C.2, as a function of the number of simulations. It can be observed that the average values of most of the output parameters are almost constants, independently of the number of simulations performed. *Average VBS Serving Data Rate* and *Average Delay on Service Request InP* are presented in Figures 5.10 and 5.11, as examples. The same is true for the standard deviation.




(a) VNet GRT (b) VNet BE

Figure 5.10. *Average VBS Serving Rate* for different number of simulations.

The exceptions are the average of *Average VBS Time Service Delayed* for VNet GRT, which decreases for a stable value only after 20 simulations, and the standard deviation of *Out of Contract*, also for VNet GRT, which stabilises after the first simulation. In order to quantify these observations, the deviation percentage relative to the average of all values obtained for the simulations was computed from (5.1), for each set of simulations, Table C.2. From Table C.2, it can be observed that, for the analysed parameters, the relative deviation of the average value is less than or equal to 2%. Only for *Average VBS Time Service Delayed*, on VNet GRT, the relative deviation of the average is greater than 2% until 10 simulations have been performed. However, as stressed for the analysis of simulator transitory interval, the order of magnitude of the service delayed is in units of milliseconds, hence not being great significance for this parameter.

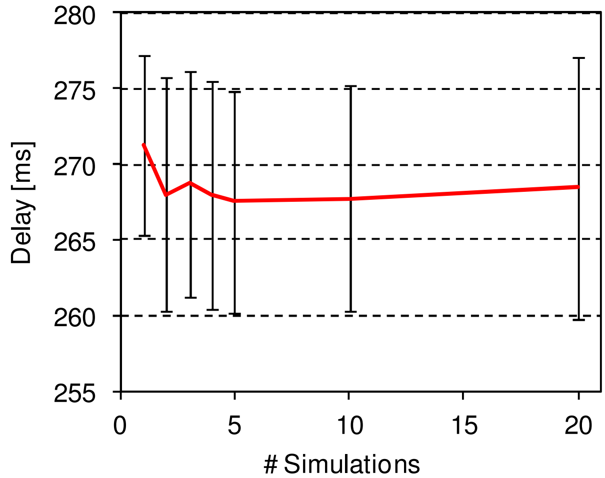



Figure 5.11. *Average Delay on Service Request InP* for different number of simulations.

As a final conclusion, one can say that one simulation is enough to obtain values with the desired accuracy.

# Chapter 6

# Scenarios and Analysis of Theoretical Results

The scenarios used for evaluation and the results achieved by using the models proposed in this thesis are presented. An analysis of the theoretical limits and the comparison with simulation results is performed.

## 6.1 Scenarios

In order to analyse the proposed VRRA algorithms from different perspectives, the assessment was done by starting from a reference scenario over which several changes were applied, by varying a set of relevant parameters. To define the reference scenario, one needs to enumerate the different degree of freedom or input parameters that have been considered in our network model. The input parameters being considered are shown below.

**Physical Capacity (Data Rate)** - The physical capacity or the maximum cluster serving data rate is determined by the number of BSs per RAT, number of RATs, and number of channels/sectors considered per BS. This last dimension was fixed to the maximum possible number of channels for each RAT, since changing the number of channels per RAT is related to the level of interference introduced in the radio access, which is not taken into account in this work, because it is assumed that it is responsibility of the InP to be aware of interference. For the other parameters, the considered scenario is one service area, the cluster, composed of BSs of all RATs, co-located to allow several options on network access. The reference cluster is composed of 2 TDMA, 1 CDMA, 4 OFDMA and 8 OFDM BSs, Figure 6.1.

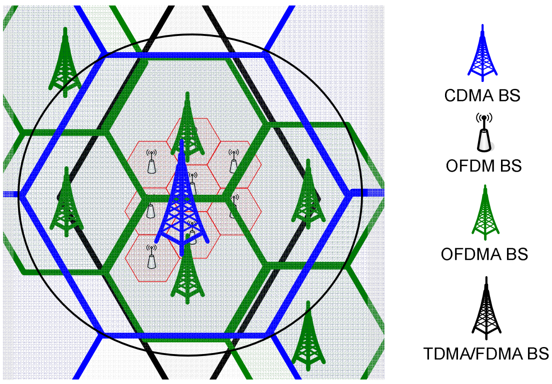



Figure 6.1. Reference cluster.

**Operators** - In a traditional business model, the end-user has a business relationship with a network provider, and can only connect to the network via the infrastructure of his/her network provider, without considering roaming. In the last few years, MVNOs have appeared in the mobile communications market, sharing a given network provider infrastructure to provide specific services for their clients. In this situation, end-users have a relationship with one MVNO, which in turn has a contractual relationship with a network provider. Hence, the MVNO clients can only connect via the infrastructure of the network provider with which his/her MVNO has a contract. With the introduction of network virtualisation, one foresees the introduction of new businesses actors. In particular, the VNO that receives for operation, from a virtualisation enabler, a VNet with specific requirements, according to the SP request, possibly sharing several infrastructures from different network providers or InPs. In this case, the VNO clients may connect via the infrastructure of any NetProv, providing network resources for the VNet. Two types of operators or operator's roles have been considered: the NetProv and the VNO. The NetProv is the owner of the infrastructure, being also the physical network operator, and most of the times providing added value services to the customer. The VNO is the actor that uses the network provider's infrastructures to build and operate a VNet in order to provide a given service. The MVNO, as it is nowadays, can be seen as a particular case of a VNO, with the constraint of using only the infrastructure of one NetProv. Three scenarios have been defined to study the benefits that can be achieved with the introduction of virtualisation in wireless heterogeneous networks: *Standard*, *MVNO* and *VNet* (with VNets). Such scenarios aim at reproducing the actual and foreseen future relations among users and operators, however, since the *MVNO* scenario is a particular case of the *Standard* scenario, it is not considered for evaluation. Regarding the attributes related to operators, the following scenarios were defined:

- **Mono-Operator** - The infrastructure, i.e., BSs, belongs only to one NetProv, all end-users having contract with that operator, allowing access through any BS in the service area;
- **Multi-Operator** - The BSs covering the service area belong to several NetProvs. A pre-configured percentage of end-users have contract with each network operator. One can have also different VNOs to provide the service, the percentage of end-users with contract of each one being also configurable.

**VNets** - The degrees of freedom for this parameter are: the number of VNets, the number of VBSs composing the VNets, the mixing of VNets' type, i.e., the number of GRT and BE VNets, the services each VNet provides, and the data rate contracted values. The strategy used for VBS instantiation over the physical infrastructure is also a parameter that can be changed. Concerning the number of VNets, one considers that there is no virtualisation (zero VNets), which is related to the operator standard scenario, and use cases with 2, 3 and 4 VNets. The number of VBSs composing the VNet is varying from 1, 2 or 6 VBSs. Regarding the VNet type, the used combinations intend to depict situations in which the number of VNets per type is the same, and one type is predominant over the other. Regarding the services provided by VNets, one has considered:

- **Exclusive set of services** - Each service is served only by one VNet; depending on the number of VNets, 1 or 2 services per VNet are considered, grouped by similar QoS requirements;
- **Common set of services** - 2 VNets serving a service, the number of end-users in each VNet being computed according to the relative percentage of VNets providing the service.

The VNet contracted data rate can be set up, being guaranteed if the VNet is GRT or indicative if it is BE, or not, meaning that there is no limitation for VNet serving data rate. In fact, though BE VNets have not associated any priority scheme, the contracted data rate influences the overall operation of the network, because when the minimum data rate is not achieved for all VNets, the VNets already satisfied (even GRT) cannot use more than the contracted data rate. Additionally, if the VNet contracted data rate is set up, the amount of contracted data rate by all VNets sharing the infrastructure can be greater, less than, or on the average cluster serving rate. The average cluster serving rate is considered as the rate the cluster can provide if an intermediary MCS is applied to all RRUs of the cluster. Finally, the strategies for VBS instantiation can be to pre-allocate RRUs over BSs in the serving area, Adaptive-RRA algorithm, or to allow the allocation of RRUs on demand over all BSs, OnDemand-VRRA algorithm. In the former case, the pre-allocation of RRUs can be done over all BSs in the serving area, *MaxBS*, or over the minimum number of BSs, *minBS*.

**Services** - Concerning services, the attribute that can be changed is their penetration. Several profiles of global network utilisation have been defined, Table 6.1.

Table 6.1. Global service penetration profiles in percentage of end-users performing the service.

| Services | Default | DVoW | Balanced | FSW | ViVo |
|---|---|---|---|---|---|
| **VoIP** | 4 | 36 | 28 | 11 | 46 |
| **Video Streaming** | 35 | 17 | 22 | 9 | 34 |
| **File Sharing** | 3 | 14 | 15 | 23 | 6 |
| **Web/Data** | 58 | 33 | 35 | 57 | 14 |

The default profile is derived from the service penetration of Cisco mobile data traffic forecast for 2016 [Cisc12a]. In Distributed VoIP and Web (*DVoW*) and *Balanced* scenarios, the total number of end-users in Conversational/Streaming, traffic with more stringent requirements, and Interactive/Background are balanced, even if the first one has more end-users in VoIP and Web/Data. The remaining scenarios intend to depict use cases in which the number of end-users is unbalanced among stringent and elastic applications. In File Sharing and Web (*FSW*), more weight is given to services with elastic requirements, File Sharing and Web/Data, and in Video and VoIP (*ViVo*) the highest percentage of end-users is performing Video Streaming and VoIP applications.

The values of the parameters for the statistical distributions of Inter-arrival time, Data Volume and Service time defined in Table 5.1 are presented in Table 6.2. The average value for each service is defined in order to generate a high activity in the network. It should be kept in mind that the goal is to introduce minimum data rate support for GRT VNets for different network utilisation conditions, thus, the validation for a scenario of heavy network usage is a main concern.

The typical and minimum data rate is also included in Table 6.2, since they are important parameters for service processing. The typical data rate is the one to be provided by the network if there are enough radio resources available in the coverage area. The minimum data rate is the minimum value that will be provided to the end-user in order to allow performing the service, which is equal to zero for services without data rate requirements.

Values obtained from simulaton for inter-arrival time, service time and data volume per service are presented in Appendix B.

**End-users** - End-users are characterised by the service they are performing, their location, their SINR and by the operators with whom they have contract. Regarding end-users generated traffic, one can say that they are directly related to the data rate requested to the VBSs. The number of end-users per service is computed according to the total number of end-users in the network and the service penetration, the distribution per VNet being calculated through the relative percentage of end-users in each VNet providing the service. The network performance conditions end-users are subjected to, translated by their SINR, is an important issue to take into account for scenarios variation, since they can significantly influence the performance of the network. Four network performance scenarios were considered, *Dynamic*, *Good*, *Reference* and *Poor*, to illustrate different wireless medium conditions causing changes in cluster and VBSs' capacity. In *Dynamic* end-users, SINRs are varying over time to reflect wireless medium impairments and possible movements of the end-users. The other three scenarios depict extreme situations in which all end-users are receiving the same SINR: in *Good* the SINR corresponds to the higher order MCS, in *Reference* an intermediate value of SINR is used, and in *Poor* the end-users' SINR only allow them to use the lower order MCS.

Table 6.2. Service's statistical distributions parameters.

| Services | Typical Data Rate [Mbit/s] | Minimum Data Rate [Mbit/s] | Inter-arrival time [s] | Data Volume [MB] | Service time [s] |
|---|---|---|---|---|---|
| **VoIP** | 0.064 | 0.032 | Exp(60) | Not applied | Uniform [100, 140] |
| **Video** | 2 | 0.512 | Exp(120) | Lognormal [17.5, 10] | Not applied |
| **File Sharing** | 10 | 0 | Exp(120) | Lognormal [12.5, 5] | Not applied |
| **Web/data** | 1 | 0 | Exp(36) | Lognormal [3, 5] | Not applied |

**Strategies for resource evaluation -** The physical resources, BSs, comparison is based on their cost, computed from (4.39), as a function of end-user attachment purpose. The evaluation is done for an initial attachment and for vertical handover, by comparing the cost of the available co-located BSs in the end-user area. The strategy is the maximum absolute capacity available, in which BSs are evaluated through the data rate availability computed according to (4.40). The BS with more available data rate, i.e., with less RRUs assigned to end-users, is the selected one.

A summary of the input parameters, as well as the range of values they can take, is presented in Table 6.3. The Reference scenario is highlighted, the values of the several attributes being indicated in the corresponding column. It should be added that the amount of data rate contracted in this Reference scenario is 1.25 Gbit/s and 1.5 Gbit/s for GRT VBS and BE VBS, respectively. These values are chosen by considering the percentage of BE versus GRT services for the Default service profile, the typical data rates of the services, and the average physical capacity of the cluster.

Table 6.3. Input parameters and considered changes of their attributes.

| Parameter | Attribute to change | Attribute definition | Reference scenario |
|---|---|---|---|
| **Operators** | # NetProv | 1 | * |
| | | 2 (each with all RATs serving the serving area) | |
| | # VNO | 1 | |
| | | 2 (1 per set of services) | * |
| | Type of OPScenario | *Standard* | |
| | | *VNet* | * |
| **VNets** | Service mixing | Exclusive set of services provided (1, 2 services per VNet) | * |
| | | Common set of services (2 VNets providing same set of services) | |
| | VNet type mixing | Not applicable | |
| | | BE/BE | |
| | | GRT/BE | * |
| | | 2GRT/2BE, 3GRT/1BE, 1GRT/3BE | |
| | Quantity | 0 | |
| | | 2 | * |
| | | 3 | |
| | | 4 | |
| | Data rate contracted | *Average* cluster capacity | * |
| | | *Over* cluster capacity | |
| | | *Under* cluster capacity | |
| | | *Limited* - All VNets with limits | * |
| | | *VNet-unLimited* - GRT VNets with limits and BE without | |
| **Service Profile** | % of end-users per service | Default - VoIP(4%); Video(35%); FS(3%); Web(58%) | * |
| | | DVoW - VoIP(36%) Video(17%) FS (14%) Web(33%) | |
| | | Balanced - VoIP(28%) Video(22%) FS (15%) Web(35%) | |
| | | FSW - VoIP(11%) Video(9%) FS (23%) Web(57%) | |
| | | ViVo - VoIP(46%) Video(34%) FS (6%) Web(14%) | |
| **End-users** | Network Operator | 50% of end-users in each network operator | |
| | | 80% of end-users in one operator and 20% on another | |
| | | All end-users in the same network operator | * |
| | Quantity | 8000 | * |
| | | From 1000 to 20000 | |
| | Network Performance Conditions | *Dynamic* - MCS can vary | * |
| | | *Good* – All end-users with higher order MCS | |
| | | *Average* - All end-users with intermediate MCS | |
| | | *Poor* - All end-users with lower order MCS | |

In this section, the set of parameters that were used for the evaluation of the proposed VRRA algorithms are identified, and the considered values of their attributes are defined. The Reference scenario is also defined.

## 6.2 Cluster and VBS Serving Data Rate

The *Average Cluster Serving Data Rate,* $\overline{R_{serv}^{Cl}}$, was computed from (4.27), based on the average RRU data rate and the number of RRUs assigned to end-users in the Cluster; the values are depicted in Figure 6.2. As expected, while the maximum capacity of the cluster is not reached, for each performance scenario, $\overline{R_{serv}^{Cl}}$ increases with the *Cluster Requested Data Rate,* $R_{req}^{Cl}$. The large variability of this performance indicator should be noted, which can range from 1.2 Gbit/s in *Poor* to 6.2 Gbit/s in *Good*, for $R_{req}^{Cl} > 8$ Gbit/s. The average cluster capacity is approximately 2.9 Gbit/s.

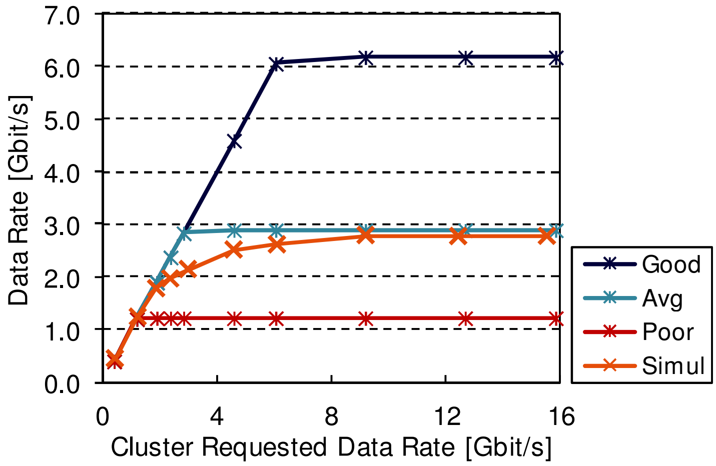
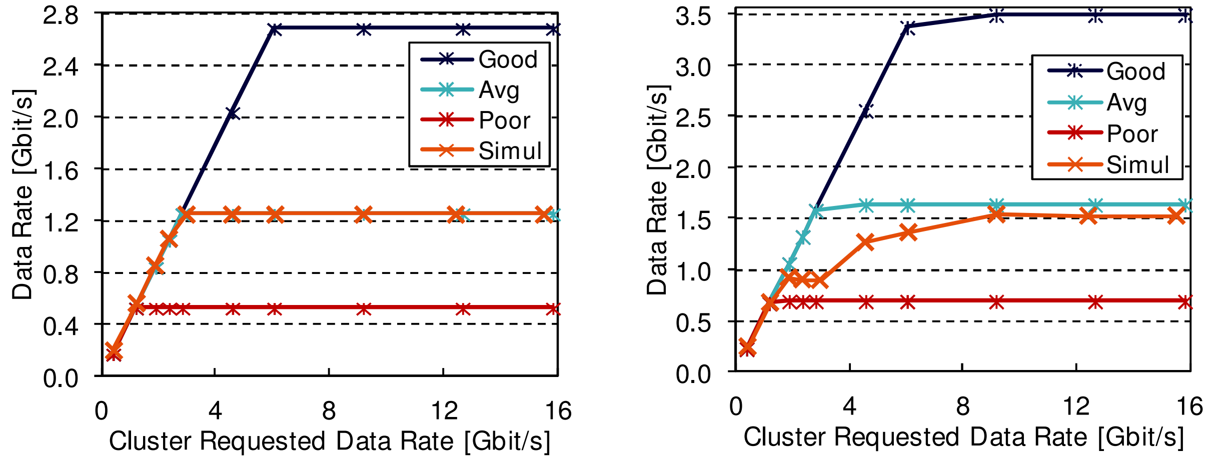



Figure 6.2. *Average Cluster Serving Data Rate* for the several performance scenarios.

Results obtained from simulations, also presented in Figure 6.2, tend to follow the values achieved for the *Average* use case. It should be stressed that instead of maintaining all end-users in the same network performance conditions, as it is the case for the theoretical approach, in simulations each end-user is independently handled, being subjected to particular network performance conditions. Hence, it can be concluded that, by applying OnDemand-VRRA, the cluster behaves as if all end-users are subjected to the *Average* performance conditions, even when network conditions for each one are varying independently from *Poor* to *Good*.

The *Average VBS Serving Data Rate*, $\overline{R_{serv}^{VBS}}$, computed from (4.21), has an analogous behaviour to that of $\overline{R_{serv}^{Cl}}$, depending only on the partial capacity allocated to VBSs, Figure 6.3. The same phenomena relative to capacity variations according to performance scenarios is perceived for both VBSs. The reduction observed from *Good* to *Poor* use cases, for high $R_{req}^{Cl}$ values, is greater for a BE than for a GRT VBS, because the number of RRUs allocated for a GRT VBS is less than the one for a BE VBS, according to the typical data rate of the provided services, and for this computation all RRUs are subject to the same performance conditions. This fact suggests that VBSs providing services with higher typical data rate, i.e., when more RRUs are assigned per end-user, will be more sensitive to the capacity variations motivated by data rate adaptations to the wireless medium conditions.




(a) GRT VBS

(b) BE VBS

Figure 6.3. *Average VBS Serving Data Rate* for the several performance scenarios.

Regarding results from simulations, it is worth to note that while for a GRT VBS the $\overline{R_{serv}^{VBS}}$ follows the *Average* use case, being superimposed in Figure 6.3(a), the achievable values for a BE VBS are higher when $R_{req}^{Cl} > 1.9$ Gbit/s, Figure 6.3(b). In a first phase, for $1.9 < R_{req}^{Cl} < 3$ Gbit/s, $\overline{R_{serv}^{VBS}}$ of BE VBS is almost constant due to the reductions imposed by OnDemand-VRRA, since $R_{req}^{VBS}$ is close to $R_{min}^{VBS}$ for a GRT VBS and BE end-users are reduced. For $3 \leq R_{req}^{Cl} < 6.1$ Gbit/s, $\overline{R_{serv}^{VBS}}$ increases until $R_{ref}^{VBS}$, as the cluster average capacity is enough to satisfy both VBSs' contracted data rates.

## 6.3 Cluster and VBS Utilisation

*Cluster Utilisation*, $\eta_{Cl}$, computed from (4.29) is depicted in Figure 6.4. It can be verified that $\eta_{Cl} = 1$, meaning that all the RRUs in the cluster are assigned to end-users, as soon as $R_{req}^{Cl}$ achieves the maximum $\overline{R_{serv}^{Cl}}$ for the given performance use case, e.g., in *Poor,* $\eta_{Cl} = 1$ for $R_{req}^{Cl} \geq 1.2$ Gbit/s, which is the maximum value achieved for $\overline{R_{serv}^{Cl}}$ in this use case. When $R_{req}^{Cl} \leq max\left(\overline{R_{serv}^{Cl}}\right)$ for each use case, $\eta_{Cl}$ increases with the increase of $R_{req}^{Cl}$.

From simulations, one can observe a reduction of cluster utilisation for $2.4 < R_{req}^{Cl} < 6.1$ Gbit/s, the same range of values in which $\overline{R_{serv}^{VBS}}$ reduction in BE VBS can be detected, Figure 6.3(b).

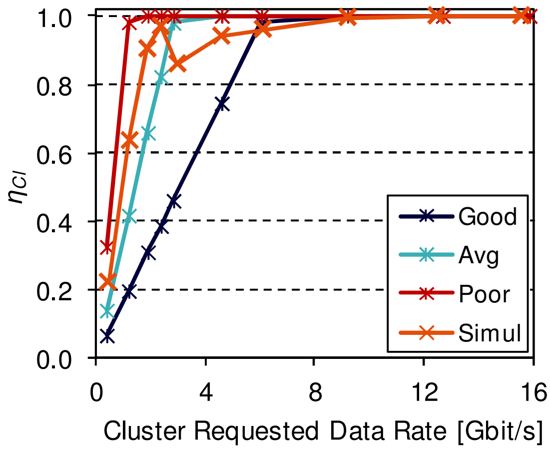
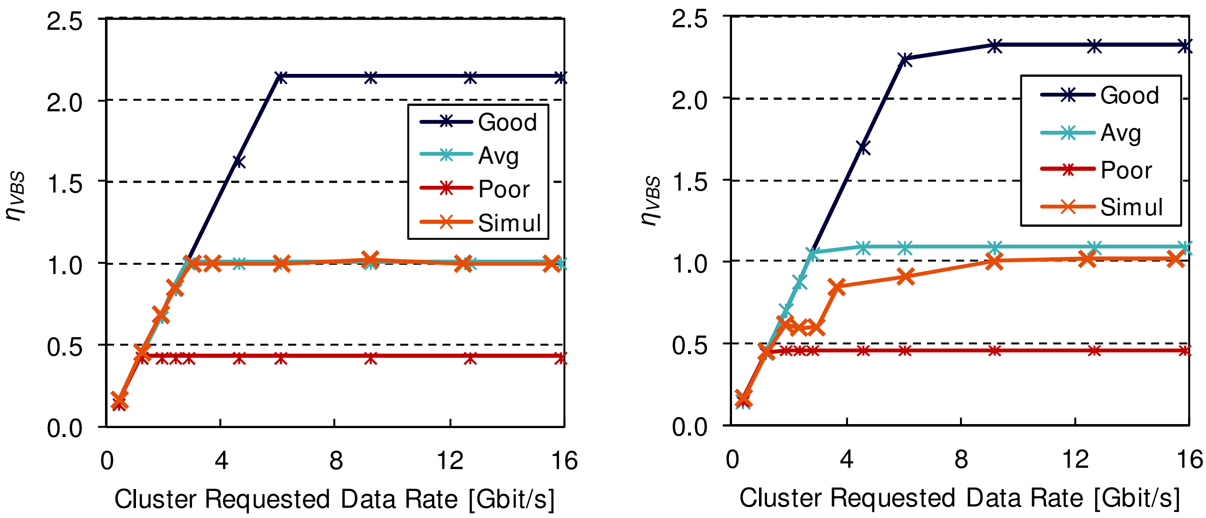



Figure 6.4. *Cluster Utilisation* for the several performance scenarios.

*VBS Utilisation*, $\eta_{VBS}$, theoretical *Average* values are evaluated from (4.22). From Figure 6.5, one concludes that for the *Good* performance use case, the VBS utilisation can be greater than 1 for high $R_{req}^{Cl}$ values, e.g., $\eta_{VBS} = 2.3$ in a BE VBS, Figure 6.5(b), meaning that the VBS is serving more than the contracted data rate. However, for the *Poor* use case, VBSs are operating below the contracted data rate, e.g., $\eta_{VBS} =$ 0.5 in a BE VBS, and near the contract if the performance use case is *Average*, $\eta_{VBS} = 1.1$ in a BE VBS.




(a) GRT VBS (b) BE VBS

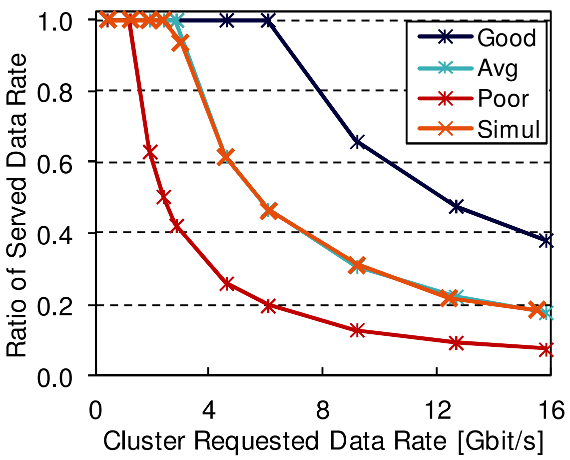
Figure 6.5. *VBS Utilisation* for the several performance scenarios.

A similar behaviour is observed for a GRT VBS, Figure 6.5(a). Furthermore, results show that the variation in $\eta_{VBS}$ between *Good* and *Poor* scenarios becomes larger as the Cluster requested data rate increases. For example, in a GRT VBS, when $R_{req}^{Cl} = 0.7$ Gbit/s, $\eta_{VBS}$ varies from 0.4 to 0.5, and when $R_{req}^{Cl} = 2.4$ Gbit/s it varies from 0.1 to 1.6. In fact, the high values of $R_{req}^{Cl}$ correspond to a greater utilisation of the RRUs allocated to the VBS, making it more sensitive to the deterioration of the radio interface, as it implies the data rate reduction of a higher number of RRUs. It is worthwhile to note that the "Avg" curve is superimposed to the "Simul" one.

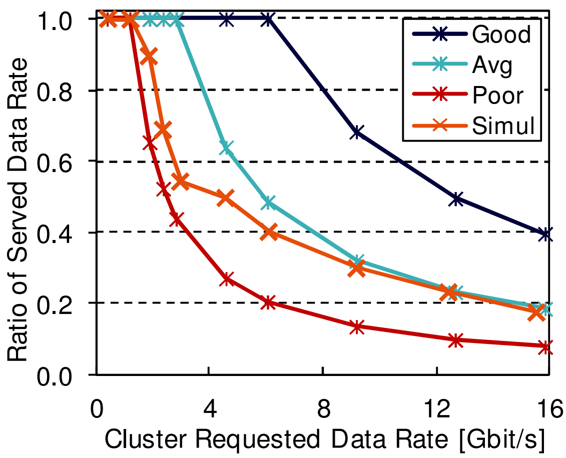
Results from simulations confirm the trend already referred for $\overline{R_{serv}^{VBS}}$, $\eta_{VBS}$ tends to follows the *Average* use case, independently of the cluster requested data rate, in a GRT VBS, and for $R_{req}^{Cl} \geq 6.1$ Gbit/s for a BE VBS.

## 6.4 Ratio of Data Rate Served

The *Ratio of Data Rate Served*, $r_{serv}^{VBS}$, obtained from (4.25), is illustrated in Figure 6.6. Results show that the VBS requested data rate is completely served for a GRT VBS, $r_{serv}^{VBS} = 1$, until the VBS contracted data rate, for the *Poor* use case and maximum achievable serving data rate, Figure 6.6(a). Above it, $r_{serv}^{VBS}$ decreases as $R_{req}^{Cl}$ increases, the minimum value being achieved for *Poor*, $r_{serv}^{VBS} = 0.1$, for $R_{req}^{Cl} \approx 16$ Gbit/s. It can be seen that for $R_{req}^{Cl} = 6.1$ Gbit/s, $r_{serv}^{VBS}$ can vary from 1.0 in *Good* to around 0.2 in *Poor*, which means that also the ratio of served data rate is strongly dependent on the radio interfaces conditions.




(a) GRT VBS




(b) BE VBS

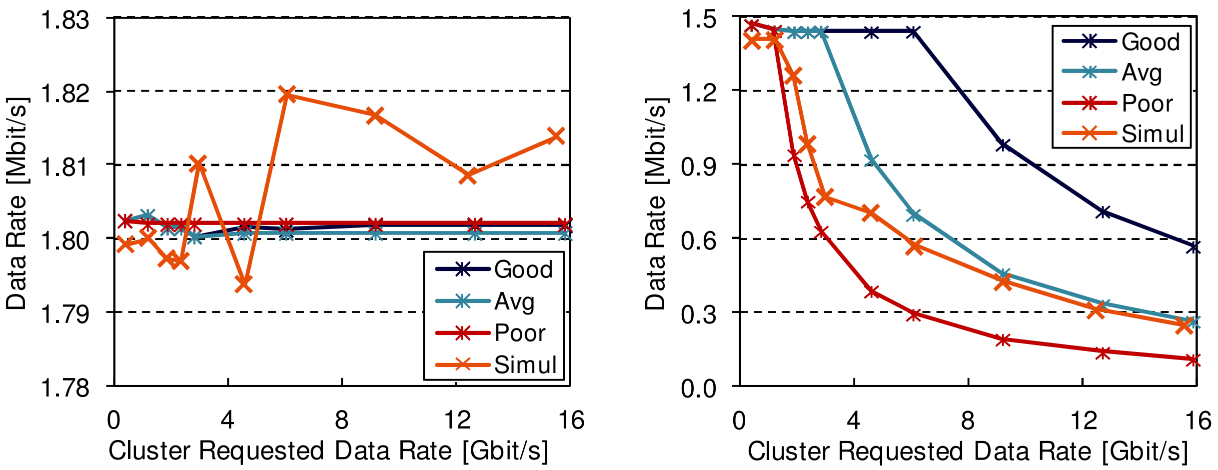
Figure 6.6. *Ratio of Served Data Rate* for the several performance scenarios.

For a BE VBS, Figure 6.6(b), the obtained theoretical values are very similar to the ones achieved for a GRT VBS, because, on the one hand, the capacity allocated to each VBS was based on average values, for both the data rate of the provided services and the RRUs data rate, and on the other hand, the data rate requested to each VBS is proportional to the VBS capacity.

Regarding simulation results, one can observe that $r_{serv}^{VBS}$ follows the ratio achieved in the *Average* use case for a GRT VBS, Figure 6.6(a). For a BE VBS, $r_{serv}^{VBS}$ is near the results achieved for the *Average* and *Poor* use cases in a first phase, $R_{req}^{Cl}$ < 6.1 Gbit/s, tending to the values obtained for the *Average* use case for $R_{req}^{Cl} > 6.1$ Gbit/s. The former phase reflects the situation in which the GRT VBS requesting data rate raises until $R_{min}^{VBS}$, forcing the BE VBS data rate to be reduced. The latter depicts the situation in which both VBSs are running within the contract, allowing the algorithm to manage the allocation of RRUs in order to compensate for variations of the wireless medium, thus, maintaining the ratio of served data rate near the *Average*.

## 6.5 End-user Data Rate

The *Average End-user Data Rate*, $\overline{R_{serv}^{EU}}$, was computed from (4.23). Theoretical results, Figure 6.7(a), show that $\overline{R_{serv}^{EU}}$ is almost constant for end-users in GRT VBS, $\overline{R_{serv}^{EU}} \approx 1.81$ Mbit/s. In fact, it is considered that the number of end-users performing service in the VBS corresponds to the VBS requested data rate that equals the maximum achievable data rate of the allocated RRUs to the VBS, for a given performance use case. From simulation results, $\overline{R_{serv}^{EU}}$ is also approximately constant, a maximum difference of the order of 0.7% being observed relative to the average value obtained for the several use cases.




(a) GRT VBS (b) BE VBS

Figure 6.7. *Average End-user Data Rate* in the cluster for the several performance scenarios.

Concerning the BE VBS, as all end-users trying to enter the network are accounted for the metric computation, $\overline{R_{serv}^{EU}}$ decreases when the cluster requested data rate increases, Figure 6.7(b). From simulations, it must be stressed that the trend of $\overline{R_{serv}^{EU}}$ is similar to the one found for $r_{serv}^{VBS}$, for the same reasons. In fact, for $R_{req}^{Cl} < 6.1$ Gbit/s, $\overline{R_{serv}^{EU}}$ is near the values achieved for the *Poor* use case, and for $R_{req}^{Cl} > 6.1$ Gbit/s, when both VBSs operate with $\overline{R_{serv}^{VBS}}$ greater than the contracted data rate, $\overline{R_{serv}^{EU}}$ follows the *Average* use case.

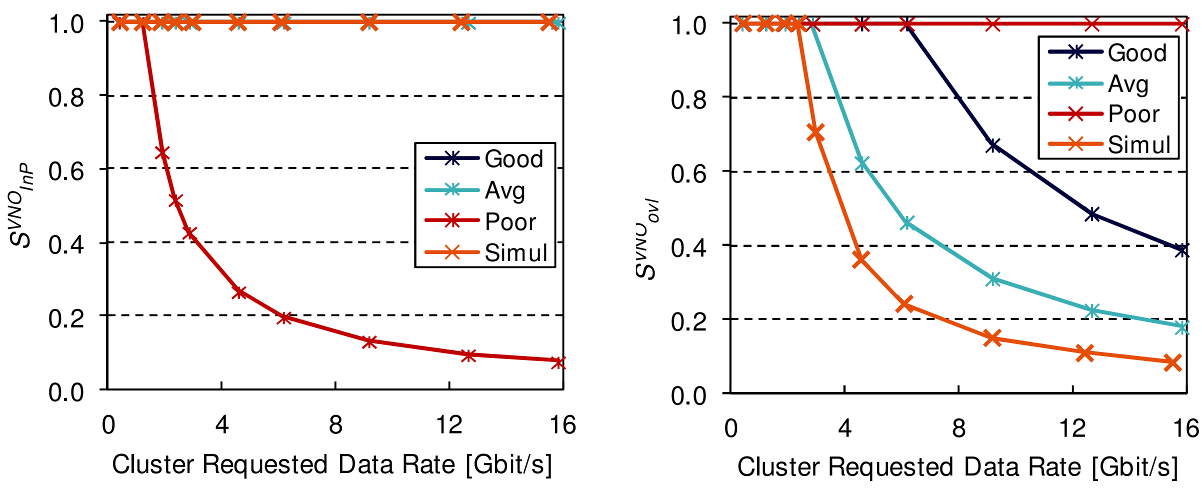
## 6.6 VNO Satisfaction Level

Results for *Satisfaction Level on the InP*, $S_{InP}^{VNO}$, obtained from (4.31), and *Satisfaction Level on extra Capacity Requested*, $S_{ovl}^{VNO}$, obtained from (4.32), are presented for a GRT VBS in Figure 6.8. For a BE VBS the *Satisfaction Level* is not presented, as it reflects the number of end-users that are delayed on service request, which is not applicable to BE end-users.

It can be observed that when network conditions degrade, the *Poor* use case, $S_{InP}^{VNO} < 1$, because some end-users may be not connected, even if the VBS is operating below the contracted data rate, Figure 6.8(a). For the other use cases, $S_{InP}^{VNO} = 1$, because the cluster has enough capacity to connect all end-users when $\overline{R_{serv}^{VBS}} \leq R_{min}^{VBS}$. Only when $\overline{R_{serv}^{VBS}} \geq R_{min}^{VBS}$ some end-users are not connected, which is accounted for $S_{ovl}^{VNO}$ computation as they are requesting service on extra capacity Figure 6.8(b). The same rational can be applied for the *Poor* use case regarding $S_{ovl}^{VNO}$. In fact, it is verified that $S_{ovl}^{VNO} = 1$ whatever the $R_{req}^{Cl}$ value, because for *Poor*, end-users are never requesting service on extra capacity as it is always $\overline{R_{serv}^{VBS}} \leq R_{min}^{VBS}$.

Moreover, one can verify that, generally, the satisfaction level, $S_{InP}^{VNO}$ or $S_{ovl}^{VNO}$ according to the use case, decreases as $R_{req}^{Cl}$ increases, because more end-users cannot be connected due to the unavailability of RRUs to be assigned.

Concerning simulation results, one can conclude that, independently of the wireless medium condition, $S_{InP}^{VNO}$ is always 1, meaning that all end-users are connected if $\overline{R_{serv}^{VBS}} \leq R_{min}^{VBS}$, Figure 6.8(a). When the VBS is running over the contracted capacity, Figure 6.8(b), a decrease in the satisfaction level is perceived, the achieved values being slightly lower than the ones obtained for the *Average* use case. This is because the OnDemand-VRRA algorithm limits the admission of more end-users after the GRT VBS contracted capacity is reached and while the BE VBS is not achieving its contracted data rate. Still, even when the BE VBS is running within contract, $\overline{R_{serv}^{VBS}} \geq R_{ref}^{VBS}$, the $S_{ovl}^{VNO}$ value is very low, $S_{ovl}^{VNO} \approx 0.1$, now due to the physical capacity constraint.




(a) on infrastructure Provider

(b) on extra capacity

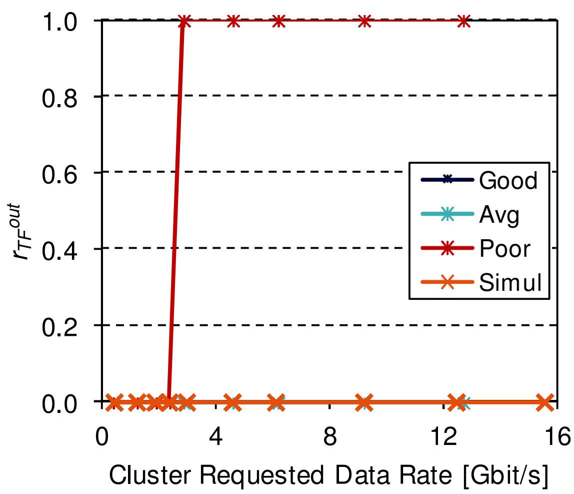
Figure 6.8. *Satisfaction Level* for the several performance use cases.

Summarising, there is a great dependency of *Satisfaction Level* on the network performance use case, reflecting the radio interface conditions. It can be observed that, for the same amount of cluster requested data rate, $R_{req}^{Cl} = 6.1$ Gbit/s, *Satisfaction Level* can vary from 1 in *Good* to 0.2 in Poor, Figure 6.8(a) and (b). In the latter case, although the VNO might not be serving all the contracted capacity, $\overline{R_{serv}^{VBS}} \leq R_{min}^{VBS}$, it cannot assign more capacity to its end-users due to the severe degradation of the radio interface. Applying the OnDemand-VRRA algorithm allows maintaining the maximum value of *Satisfaction Level* whenever the VBS operates within the contracted data rate. However, if the VBS operates over the contracted capacity, the *Satisfaction Level* decreases taking values lower than the achieved for the *Average* use case.

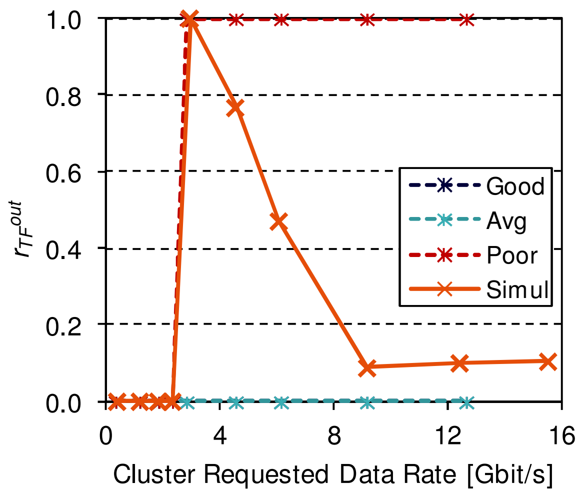
## 6.7 Out of Contract Ratio

Figure 6.9 illustrates the *Out of Contract Ratio*, $r_{TF}^{out}$, variation computed from (4.30). In the *Poor* use case, the GRT VBS runs all the time frames out of contract for $R_{req}^{Cl} > 2.4$ Gbit/s, which corresponds to $R_{req}^{VBS} > R_{min}^{VBS}$, Figure 6.9(a), because the maximum $\overline{R_{serv}^{VBS}}$ the GRT VBS can achieve in this scenario is less than $R_{min}^{VBS}$. Note that VBSs are considered within contract for $R_{req}^{VBS} \leq R_{min}^{VBS}$, $r_{TF}^{out} = 0$ for $R_{req}^{Cl} \leq 2.4$ Gbit/s. The introduction of the OnDemand-VRRA algorithm allows compensating for capacity reduction maintaining the VBS always within contract, $r_{TF}^{out} = 0$, Figure 6.9(a).

For BE VBS theoretical values, the same phenomenon arises in the *Poor* use case, because the VBS can never reach $R_{ref}^{VBS}$. It can be observed that for $R_{req}^{Cl} > 2.4$ Gbit/s, corresponding to the situation in which $R_{req}^{VBS} > R_{min}^{VBS}$, the value of $r_{TF}^{out}$ is always 1 independently of the amount of $R_{req}^{Cl}$, Figure 6.9(b). From simulation results, $r_{TF}^{out} = 1$ for $R_{req}^{Cl} = 2.4$ Gbit/s, denoting the situation referred for the theoretical approach. In this case, the OnDemand algorithm gives more priority to GRT end-users to allow the GRT VBS to operate within contract, reducing BE end-users when some lack of capacity is detected. However, for $R_{req}^{Cl} > 2.4$ Gbit/s, $r_{TF}^{out}$ decreases, since the GRT $\overline{R_{serv}^{VBS}}$ becomes limited to its contracted data rate; as the total contracted data rate in the cluster is near the average cluster capacity, also a BE VBS tends to achieve its $R_{ref}^{VBS}$. It can be observed that for $R_{req}^{Cl} > 9.2$ Gbit/s, $r_{TF}^{out} \approx 0.1$, meaning that for 10% of the time frames the BE VBS cannot achieve $R_{ref}^{VBS}$.




(a) GRT VBS



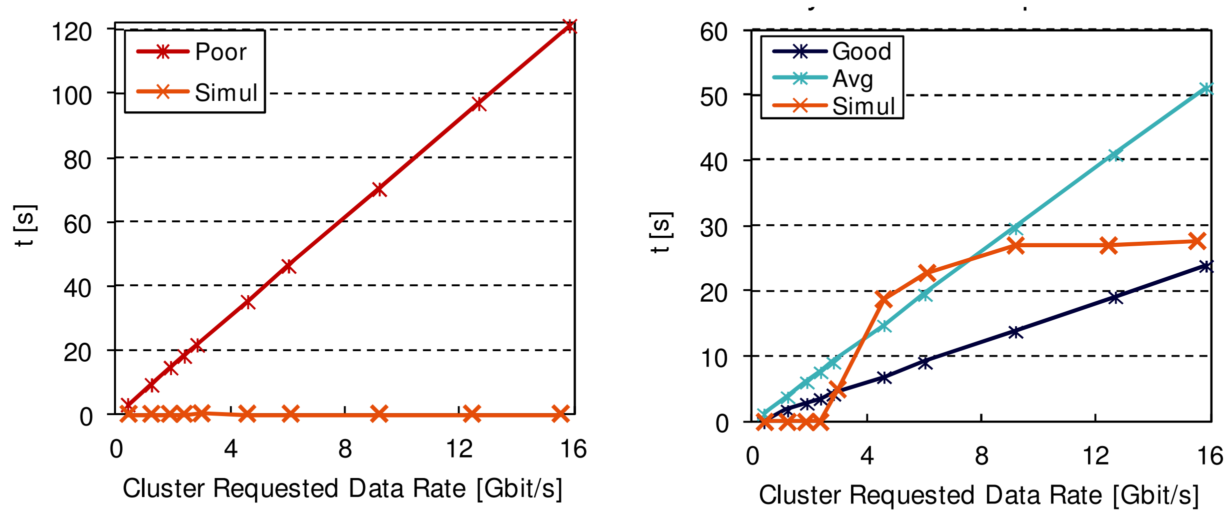

(b) BE VBS

Figure 6.9. *Out of Contract* for the *Average* performance scenario.

## 6.8 Delay on Service Request and Service Delayed

The *Average Delay on Service Request InP*, $\overline{\tau_{InP}^{VBS}}$, calculated from (4.35), and the *Average Delay on*

*Service Request VNO*, $\overline{\tau_{VNO}^{VBS}}$ , calculated from (4.36), are presented in Figure 6.10. $\overline{\tau_{InP}^{VBS}}$ and $\overline{\tau_{VNO}^{VBS}}$ are only applied for GRT services, the theoretical values being based on the number of RRUs an end-user needs to reach the typical service data rate and the average time RRUs are occupied in each VBS for a given use case. To achieve an average value per VBS, provided that two different services are provided, one also considers the percentage of end-users performing each service and the service time duration.




(a) on Infrastructure Provider (b) on extra capacity

Figure 6.10. *Delay on Service request* in GRT VBS for different performance scenarios.

It can be observed, Figure 6.10(a), that in the *Poor* use case, $\overline{\tau_{InP}^{VBS}}$ increases with the cluster requested data rate, and may reach values of the order of hundreds of seconds, $\overline{\tau_{InP}^{VBS}} \approx 120$ s for $R_{req}^{Cl} = 16$ Gbit/s, because the cluster capacity is much reduced in the *Poor* use case, and as soon as the requested data rate is greater than it, end-users are delayed before admission. Results obtained by simulation present values of the order of 100 ms, since end-users in a GRT VBS have priority to enter the network while the VBS is operating under the contracted capacity. For *Average* and *Good* use cases the delays on service request are accounted for, $\overline{\tau_{VNO}^{VBS}}$ , because the achievable data rate in these use cases is greater than the contracted data rate. Figure 6.10(b) shows that $\overline{\tau_{VNO}^{VBS}}$ increases for both use cases with $R_{req}^{Cl}$, though increasing faster in *Average* as the maximum VBS serving data rate is lower than in *Good*. Results from simulation, Figure 6.10(b), show that $\overline{\tau_{VNO}^{VBS}}$ is between the theoretical values achieved for *Average* and *Good* use cases, tending to a given value, since it has been included a timeout to simulate the withdraw of the end-user whenever the waiting time to enter the network is too long, in this case it was set to 30 s.

Results for *Average VBS Time Service Delayed,* obtained from the sum of (4.33) and (4.34), are presented in Figure 6.11. Only results for BE VBS are presented, since it is assumed that the RRUs assigned to GRT end-users are maintained during service time. Although this is true theoretically, because it is assumed that end-users in the network receive the typical data rate until it is completed, in simulations GRT end-users can see their service delayed, the achieved values being of the order of tenths of milliseconds.

From simulation results, it can be concluded that the BE VBS $\overline{t_{int}^{VBS}}$ increases considerably, essentially due to the reduction imposed by the GRT VBS. From Figure 6.11, it is observed that after $R_{req}^{Cl} = 6.1$ Gbit/s the results achieved in simulations become higher than even the *Poor* use case, $\overline{t_{int}^{VBS}} = 11.2$ s in *Poor* and $\overline{t_{int}^{VBS}} = 14.7$ s in simulation, increasingly faster with $R_{req}^{Cl}$.


Figure 6.11. *Average VBS Time Service Delayed* in BE VBS, for different performance scenarios.

Summarising, from theoretical results the performance of the several VNets is strongly dependent on wireless medium conditions, e.g., for GRT VBS the $\overline{R_{serv}^{VBS}}$ can be reduced approximately 75% from *Good* to *Poor* use cases. Comparing the theoretical results with the simulation ones with the OnDemand-VRRA algorithm, it is observed that the VBSs and also the whole cluster behave similarly to the *Average* use case, meaning that OnDemand-VRRA allows mitigating the impact of the wireless medium impairments in the overall performance.

# Chapter 7

# Analysis of Simulation Results

In this chapter, simulation results in different scenarios are presented. First, the results obtained by implementing the Adaptive-VRRA algorithm are discussed. Then, the comparison of the proposed wireless access virtualisation, by implementing the OnDemand-VRRA algorithm, with the actual network models and the fixed allocation of radio resources is made. Finally, the impact of varying the number and type of VBSs, the amount of contracted capacity, and the service profile is evaluated.

## 7.1 Initial Considerations

The results presented in this chapter were obtained from simulations, to assess the proposed algorithms in a dynamic environment, where radio conditions vary randomly, from good to poor performance, as well as the requested data rate. The observation time interval was set to 1 h of network operation, after discarding the first 20 minutes of simulations, for initial convergence. The time frame for network monitoring was 1 s, taken as the major common denominator of all RATs, for the sake of simplicity. In fact, it is important to ensure consistency among the decisions taken at the different levels of RRM, namely, intra-RAT (RRM), inter-RAT (Cooperative RRM) and among VNets, in order to achieve an overall coherent behaviour. Given that, at the VNet level, one should have the perspective of the several RATs, the OnDemandVRRA algorithm takes decisions at a time scale that is defined for Cooperative RRM.

The evaluation metrics considered for the algorithms assessment are the ones defined in Section 4.5, namely:

- *Average VBS Serving Data Rate,* $\overline{R_{serv}^{VBS}}$;
- *Average End-user Data Rate,* $\overline{R_{serv}^{EU}}$;
- *Ratio of Data Rate Served,* $r_{serv}^{VBS}$;
- *Average Cluster Serving Data Rate*, $\overline{R_{serv}^{Cl}}$;
- *Average Cluster Utilisation*, $\overline{\eta_{Cl}}$;
- *Out of Contract*, $r_{TF}^{out}$;
- *Satisfaction Level on the InP,* $S_{InP}^{VNO}$;
- *Satisfaction Level on extra Capacity Requested,* $S_{ovl}^{VNO}$;
- *Average Delay on Service Request InP*, $\overline{\tau_{InP}^{VBS}}$;
- *Average Delay on Service Request VNO,* $\overline{\tau_{VNO}^{VBS}}$
- *Average VBS Time Service Delayed InP*, $\overline{t_{int}^{VBS}}$;
- *Average VBS Time Service Delayed VNO,* $\overline{t_{VNO}^{VBS}}$.

In the management of radio resources, a data rate reduction strategy for services with a minimum required data rate (i.e., GRT services) is applied, to compensate for possible end-user's data rate decrease due to the degradation of medium conditions. However, when considering VNet scenarios, such strategy is only applied if the VBS operates within the contracted capacity, otherwise, the service is delayed as the VBS contracted capacity is already reached. Vertical handover is also considered for all scenarios, when the end-user cannot achieve the minimum service rate in the same BS, because there are not enough RRUs available.

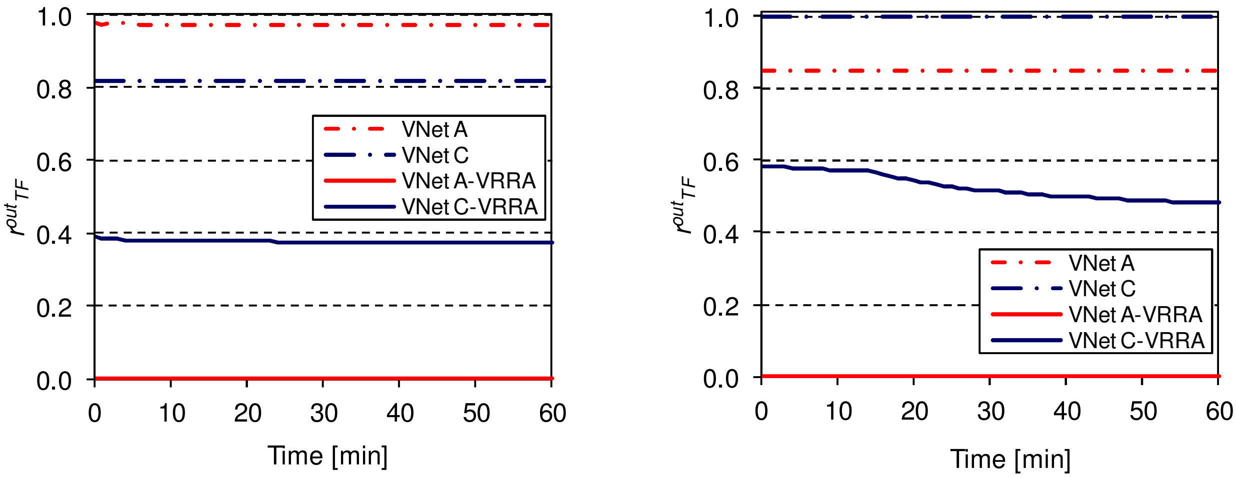
Section 7.2 presents an initial study made for the Adaptive-VRRA algorithm in order to assess different strategies of instantiated VBSs in the BSs. In Section 7.3, the nowadays existing model of operators is compared with the envisage existence of several VNOs sharing a common infrastructure from one or several InPs with a cooperative radio resources management. Two VRRA strategies are compared in Section 7.4, the OnDemand-VRRA and the Fixed Radio Resource allocation per VBS. Section 7.5 addresses the evaluation of several strategies for provision of virtual capacity from the available physical capacity, taking the type and amount of capacity requested by VNets into account. Finally, Sections 7.6 to 7.8 presents the impact of mixing of virtual resources with different requirements, changing the amount of virtual resources created in the cluster and varying the percentage of end-users in each VBS, respectively, in the performance of the proposed OnDemand-VRRA algorithm.

## 7.2 Instantiation of VBSs in the Cluster

In this section, the number of created VBSs and the strategy for allocation of RRUs in the various BSs in the cluster is assessed, for the Adaptive-VRRA algorithm described in Section 4.6.2. In this analysis, each VNet is composed of two or six VBSs within the cluster, the strategies for VBSs' instantiation, *MaxBS* and *minBS*, being defined in Section 6.1. Three VNets are instantiated in the cluster, VNets A and C being of GRT type and VNet B of BE type. VNet A provides VoIP services only, and VNet C provides Video services, data services being provided by VNet B.

Results for a BE VBS are not presented, since it is assumed that for this kind of VNets no compensation is applied when they are operating under the reference contracted data rate, as described in Section 4.6.2. Instead the available capacity in the BE resources can be used by GRT VNets, according to a predefined borrowing margin.

From Figure 7.1, it can be observed that when radio resources are allocated without any compensation mechanism, VNets are operating with $r_{TF}^{out} \geq 0.8$, meaning that the $\overline{R_{serv}^{VBS}}$ achieved for the VBSs is below $R_{min}^{VBS}$, at least 80% of the time, due to impairments of the wireless medium and the limited number of RRUs in the cluster, which is observed for both instantiation strategies, Figure 7.1. On the other hand, when Adaptive-VRRA is introduced, $r_{TF}^{out}$ decreases as the reduction of capacity is partially compensated by the allocation of more RRUs to the VBSs individually, and indirectly for the VNet, as VBS aggregation. Note that VNet C, even with the introduction of Adaptive-VRRA, does not reach the minimum *Out of Contract* value of 0, i.e., it is not running within contract for the whole observation interval. This is mainly because the services provided by VNet C, which are more demanding in terms of data rate, need to use more RRUs per end-user, therefore, increasing the difficulty for compensating the degradation in capacity.



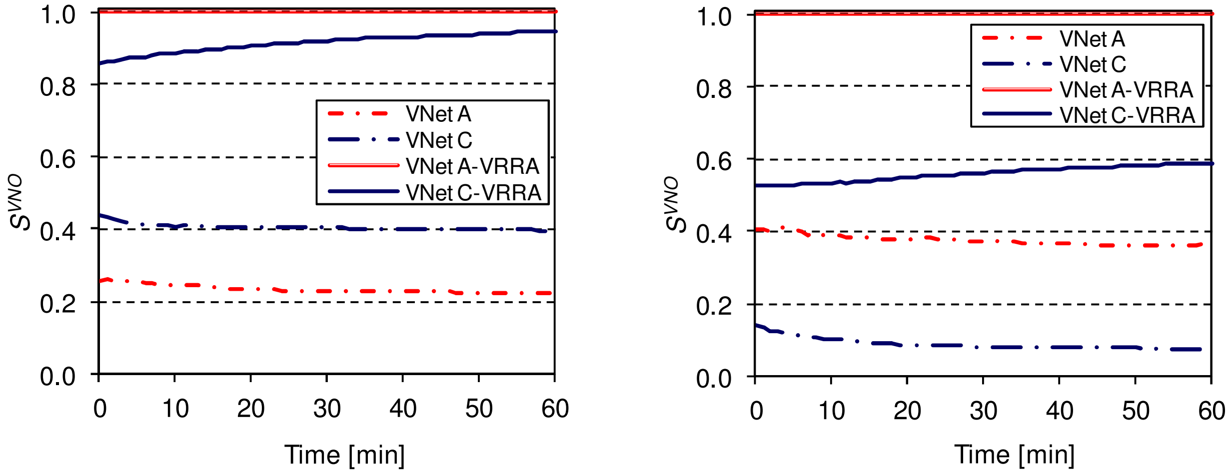

(a) *6VBS-minBS* (b) *6VBS-MaxBS*

Figure 7.1. Out of Contract for *6VBS-minBS* and *6VBS-MaxBS*.

Results for *Satisfaction Level*, Figure 7.2, show that when radio resources are allocated without any compensation mechanism, VNets A and C, the VNets are always below the maximum *Satisfaction Level* of 1, with $S_{InP}^{VNO} < 0.5$, meaning that VNOs are trying to use the contracted rate, but the InP is not providing the agreed minimum, due to degradation in the wireless medium or end-users mobility. As referred for *Out of Contract*, VNet C does not reach the maximum *Satisfaction Level* even with Adaptive-VRRA introduction.

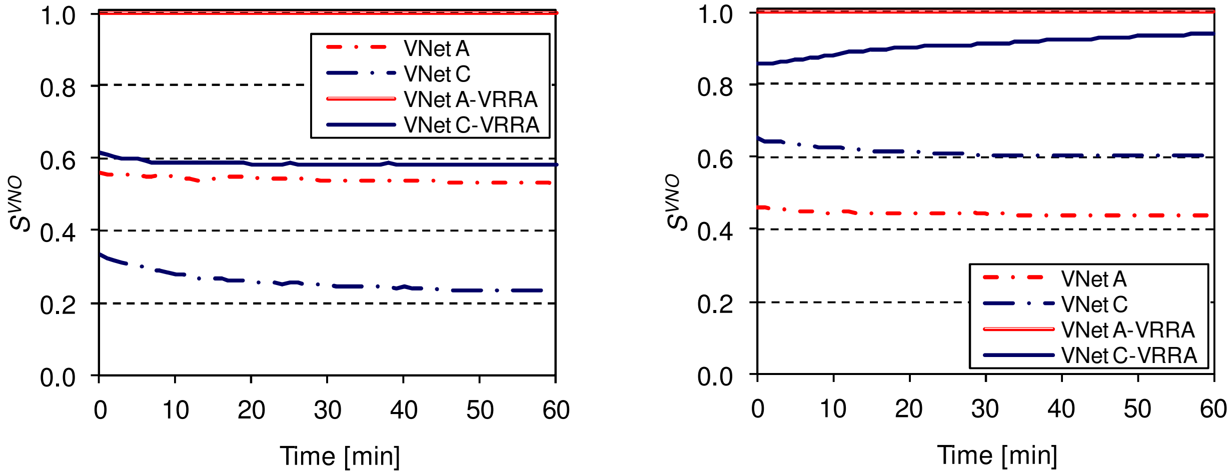
In order to analyse the impact of VBSs quantity in the VNets composition, Figure 7.3(a) depicts *Satisfaction Level* results for *2VBS-minBS*, and Figure 7.3(b) for *2VBS-MaxBS*, the former for the *minBS* instantiation strategy and the latter for the *MaxBS* one. The analysis is done just for VNet C, since VNet A *Satisfaction Level* is fully compensated in both cases by Adaptive-VRRA. From Figures 7.2(a) and 7.3(a), it can be seen that for VNet C without compensation the minimum *Satisfaction Level* is $S_{InP}^{VNO} = 0.40$ for 6 VBSs and $S_{InP}^{VNO} = 0.24$ for 2 VBSs. With Adaptive-VRRA, the maximum *Satisfaction Level* obtained for VNet C is $S_{InP}^{VNO} = 0.94$, for *6VBS-minBS*, which is greater than the obtained for *2VBS-minBS*, $S_{InP}^{VNO} = 0.58$. Therefore, a scenario with a higher number of VBSs achieves better *Satisfaction Level*, with or without Adaptive-VRRA.

Figure 7.2(b) shows that, for VNet C without compensation, the minimum obtained is $S_{InP}^{VNO} = 0.08$ for *6VBS-MaxBS*, while for *2VBS-MaxBS* it is $S_{InP}^{VNO} = 0.6$, Figure 7.3(b). With Adaptive-VRRA, the maximum value of *Satisfaction Level* is $S_{InP}^{VNO} = 0.59$ in VNet C, for *6VBS-MaxBS*, which is lower than the obtained for *2VBS-MaxBS*, $S_{InP}^{VNO} = 0.94$.




(a) *6VBS-minBS* (b) *6VBS-MaxBS*

Figure 7.2. *Satisfaction level* for *6VBS-minBS* and *6VBS-MaxBS*.

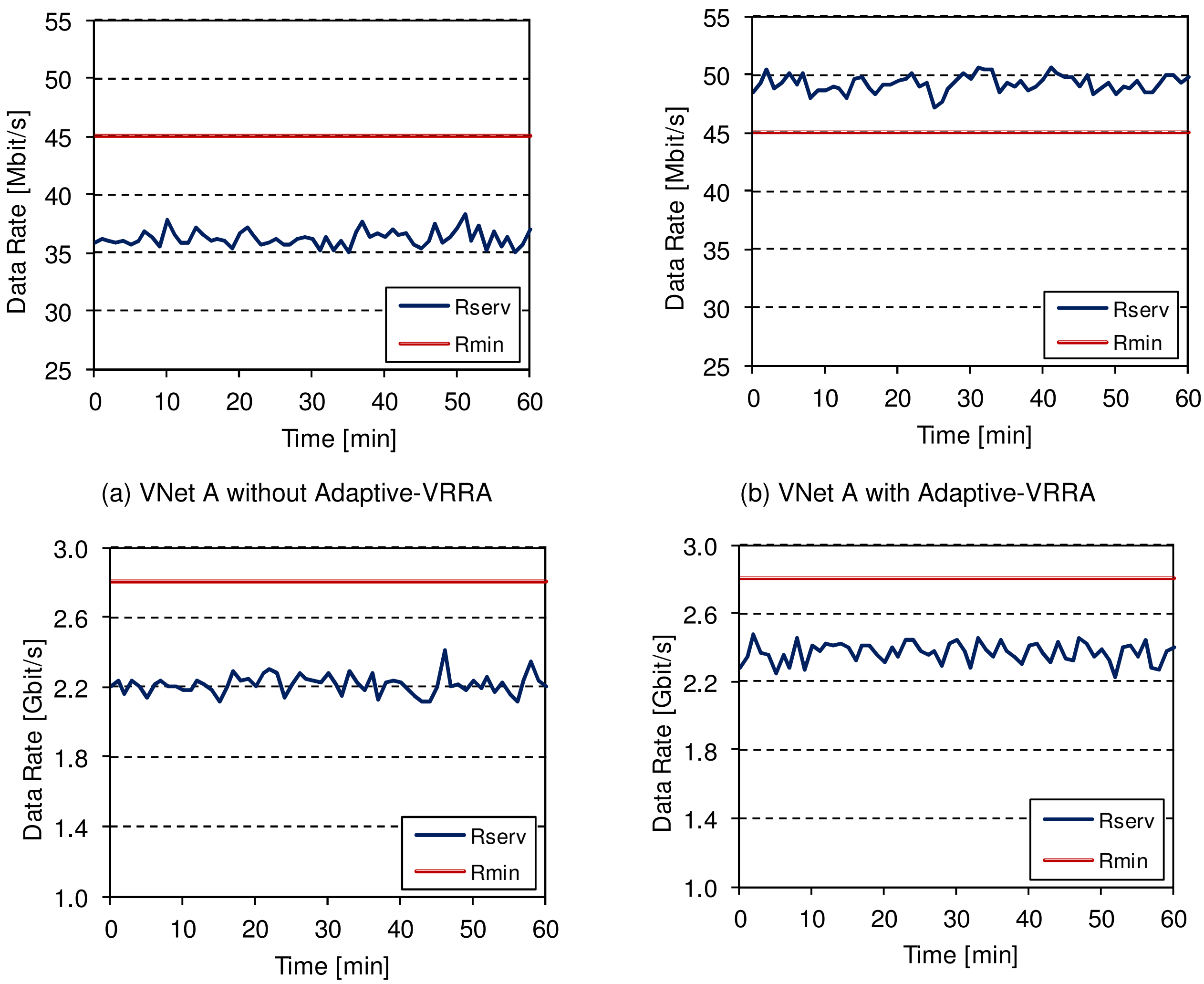



(a) *2VBS-minBS* (b) *2VBS-MaxBS*

Figure 7.3. *Satisfaction level* for *2VBS-minBS* and *2VBS-MaxBS*.

Therefore, concerning the number of VBSs for the same instantiation strategy, the *MaxBS* one, the *Satisfaction Level* achieves better values for VNet with less number of VBSs (with or without Adaptive-VRRA). On the other hand, by applying the *minBS* strategy $S_{InP}^{VNO}$ is higher for the greater amount of VBSs composing the VNet. It can then be concluded that VNet performance depends not only on the quantity of VBSs defined for its implementation, but also on the strategies used to instantiate the VBS in the physical infrastructure.

The *Average VNet Serving Data Rate*, $\overline{R_{serv}^{VNet}}$, and the minimum contracted data rate, $R_{min}^{VNet}$, for *2VBS-MaxBS* is presented in Figure 7.4. It can be seen that, without Adaptive-VRRA, Figure 7.4(a) and (c), VNets are always operating with $\overline{R_{serv}^{VNet}} < R_{min}^{VNet}$. For the case of VNet A, the introduction of the Adaptive-VRRA algorithm allocates unused RRUs in the cluster to compensate the reductions in capacity of assigned RRUs, allowing VNet A to operate with $\overline{R_{serv}^{VNet}} > R_{min}^{VNet}$, Figure 7.4(b). However, for VNet C, Figure 7.4(d), the amount of available RRUs in the cluster is not sufficient to reach $R_{min}^{VNet}$.




(a) VNet A without Adaptive-VRRA

(b) VNet A with Adaptive-VRRA

(c) VNet C without Adaptive-VRRA

(d) VNet C with Adaptive-VRRA

Figure 7.4. *Average VNet Serving Data Rate* for *2VBS-MaxBS*.

The evaluation of the Adaptive-VRRA algorithm has been presented for different VNet compositions and allocation strategies in the physical cluster. As a final consideration, it can be said that the Adaptive-VRRA algorithm may give some support to GRT VNets, by managing the allocation of RRUs to compensate for the degradation on wireless performance conditions; however, the compensation can be insufficient, due to the limited amount of RRUs in the physical cluster. This is more critical when the services provided by VNets are more data rate demanding, as in VNet C, since the performance degradation affects more RRUs simultaneously, mainly because the RRUs are pre-allocated, and the instantiation of the VBSs over the BSs it too strict. On the other hand, the algorithm is trying to allocate RRUs to satisfy the contracted capacity of GRT VNets all over the time, independently of its utilisation, allowing only to borrow RRUs from BE VNets. To overcome the limitations of the Adaptive-VRRA algorithm a new algorithm is proposed, the OnDemand algorithm, presented in Section 4.6.3, which is a more flexible one, as it only allocates RRUs to any type of VNet when it is demanded capacity for service use. Hence, the results presented in the following sections are related to OnDemand-VRRA.

## 7.3 Actual Operators versus VNet Operators Model

The main objective of this section is the comparison between the *Standard* and the *Virtual* scenarios. According to the current and foreseen future relations among users and operators presented in Section 6.1, two scenarios have been defined to study the benefits that can be achieved with the introduction of virtualisation in wireless heterogeneous networks: one without virtualisation, the *Standard* scenario and another one with virtualisation, the *Virtual* scenario. An MVNO scenario, as applied nowadays, is not considered, as it is a particular case of the *Standard* scenario.

In the *Standard* scenario, two NetProvs. are considered, one having contract with 80% of the end-users and the other with 20%. Concerning the *Virtual* scenario, two different use cases are considered, *VNet-Limited* and *VNet-unLimited*, both having one BE VBS and one GRT VBS operating from different VNOs. For the *VNet-unLimited* use case, the BE VBS has a low reference data rate, meaning that it does not impose any additional constraint to the GRT VBS, because the reference rate is always served. For the *VNet-Limited* one, the reference data rate for the BE VBS is defined according to the expected requested data rate from its end-users, becoming an additional constraint for the GRT VBS, as most of the time it cannot be reached, due to the limited capacity of the cluster.

The Reference scenario is used to depict a situation in which the total data rate requested by end-users is above the maximum cluster data rate. A summary of the use case parameters is presented Table 7.1. The achievable *Average VBS Serving Data Rate*, $\overline{R_{serv}^{VBS}}$, is computed from (4.21). For the *Standard* scenario, $\overline{R_{serv}^{VBS}}$ is computed from the *Average End-user Data Rate* for the set of services provided by each VBS in the *Virtual* scenario, for comparison purposes. In the *Standard* scenario, the GRT VBS $\overline{R_{serv}^{VBS}}$ is between the two *Virtual* use cases, Figure 7.5(a), since the share of radio resources is not coordinated among the services or the set of services, end-users being handled independently. Data rate adaptation takes only into account the minimum service data rate, $s_{min}^{R}$, of GRT services, reducing the BE services whenever needed to satisfy that $s_{min}^{R}$. The low value achieved for BE services data rate, $\overline{R_{serv}^{VBS}} = 0.44$ Gbit/s, as well as the relatively higher standard deviation, Figure 7.5(b), is a consequence of this uncoordinated allocation of radio resources.

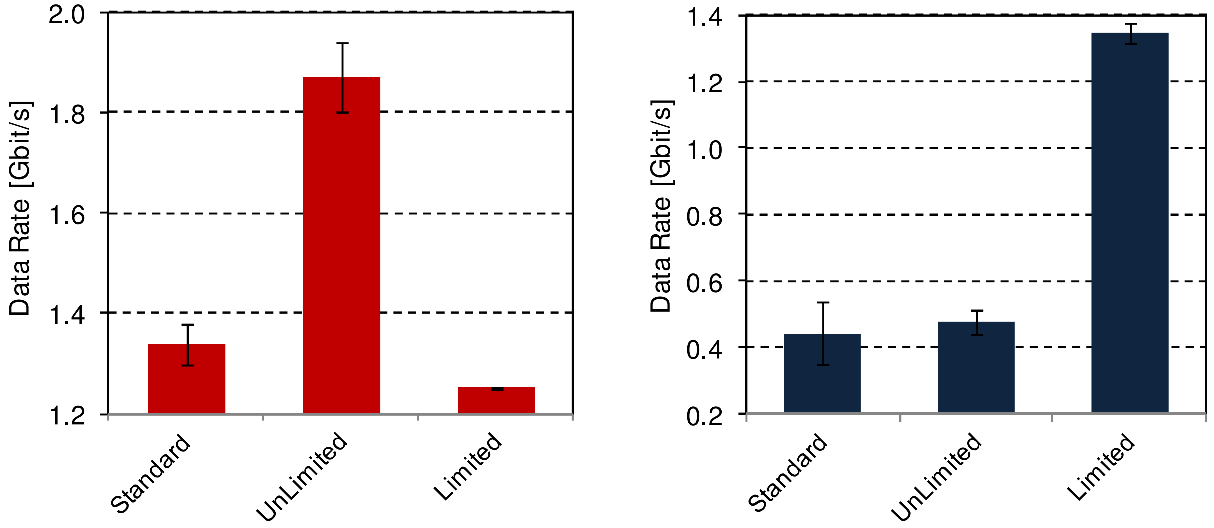
Table 7.1. Operator related use case's parameters.

<table>
<tr><th>Parameter</th><th>Standard</th><th>VNet -unLimited</th><th>VNet-Limited</th></tr>
<tr><td>Operators</td><td>2 NetProvs. providing all the services</td><td colspan="2">2 VNOs - one VNO provides GRT services and the other VNO provides BE services</td></tr>
<tr><td rowspan="5">VNets</td><td rowspan="5">Not applicable</td><td colspan="2">1VNet GRT, 1VNet BE</td></tr>
<tr><td colspan="2">2 VBSs deployed over the cluster</td></tr>
<tr><td colspan="2">GRT VBS provides VoIP and Video<br>BE VBS provides Web and File Sharing</td></tr>
<tr><td colspan="2">Total capacity contracted by VBSs within cluster capacity</td></tr>
<tr><td>GRT VBS with data rate contracted and BE VBS without</td><td>Both VNets with data rate contracted</td></tr>
<tr><td>End-users</td><td>80% in NetProv1, 20% in NetProv2</td><td colspan="2">Distribution of end-users per VBS according to the service they are performing</td></tr>
</table>

Concerning the *Virtual* use cases, one can observe, Figure 7.5(a), that the maximum $\overline{R_{serv}^{VBS}}$ for GRT VBS is reached for *VNet-unLimited*, being 50% greater than for *VNet-Limited*, in which the minimum is achieved. This highlights the importance of configuring a reference data rate for BE VNets, which is also perceived by the results depicted in Figure 7.5(b) for BE services. In fact, though the *VNet-Limited* presents a minimum $\overline{R_{serv}^{VBS}}$ value for GRT VBS, it is always within the contracted $R_{min}^{VBS}$, 1.25 Gbit/s. On the another hand, $\overline{R_{serv}^{VBS}}$ for BE VBS follows $R_{ref}^{VBS} = 1.5$ Gbit/s; nevertheless, as it cannot reach $R_{ref}^{VBS}$, the GRT VBS maintains $\overline{R_{serv}^{VBS}}$ in the contracted $R_{min}^{VBS}$.

For *VNet-unLimited*, as $R_{ref}^{VBS}$ for BE VBS is always satisfied (the defined value, 10 Mbit/s, is too low compared with the total capacity of the cluster), $R_{serv}^{VBS} \geq R_{ref}^{VBS}$, the end-users making GRT services are gradually admitted causing the increase of $\overline{R_{serv}^{VBS}}$ in GRT VBS. BE $\overline{R_{serv}^{VBS}}$ is not limited by its reference data rate, using the remaining radio resources after the admitted GRT end-users are assigned the $s_{min}^{R}$. This results in $\overline{R_{serv}^{VBS}}$ observed values greater than the contracted ones for both GRT and BE VBSs, Figure 7.5(a) and (b), respectively. It can then be said that when the total amount of capacity contracted for both VBSs is less than the total capacity of the physical cluster, the remaining capacity is distributed between both VBSs, optimising the radio resources utilisation.

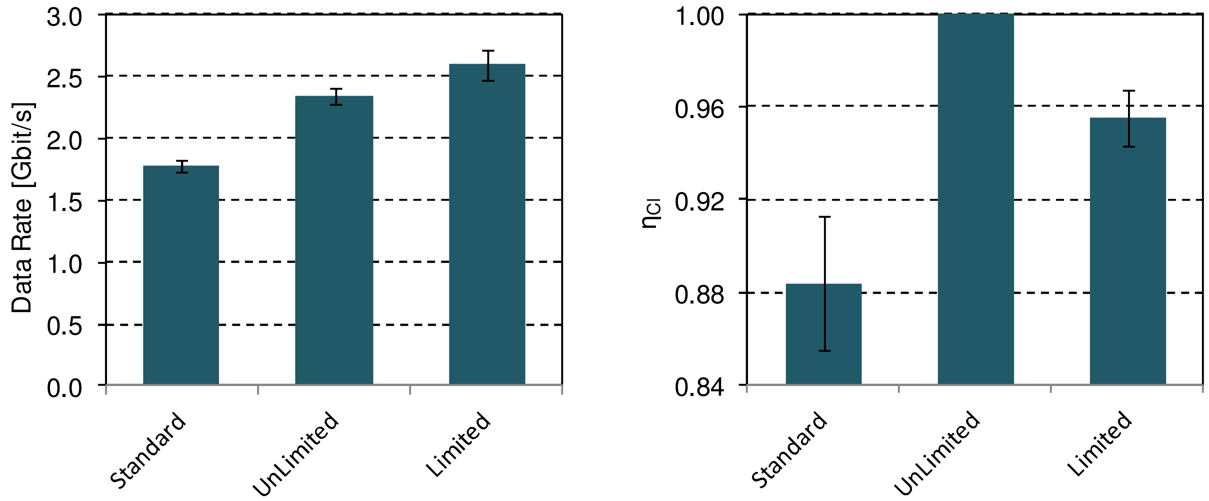
From the analysis of $\overline{R_{serv}^{VBS}}$, one can confirm the achievement of objective (4.41) for the *Virtual* use cases, i.e., $R_{min}^{VBS}$ for GRT VNets is always satisfied, and (4.42) the BE VNets follows $R_{ref}^{VBS}$. However, it should be noted that when the total capacity allocated to VBSs is close to the average capacity of the cluster, *VNet-Limited*, the BE VNets can be penalised to allow the GRT VBSs to get the contracted capacity. As expected, it is also verified that *Virtual* use cases performs better than the *Standard* use case.




(a) GRT services/VBS (b) BE services/VBS

Figure 7.5. *Average VBS Serving Data Rate* for operators' use cases.

Observing the results for the *Average Cluster Data Rate,* $\overline{R_{serv}^{Cl}}$ , obtained from (4.27), and depicted in Figure 7.6(a), one can see that for the *VNet-Limited* use case the cluster achieves the highest value, $\overline{R_{serv}^{Cl}} \approx 2.6$ Gbit/s. On the other hand, the highest *Average Cluster Utilisation,* $\overline{\eta_{Cl}}$ , computed from (4.28), is achieved for *VNet-unLimited*, which denotes a less efficient use of the overall resources than in *VNet-Limited*, since the $\overline{R_{serv}^{Cl}}$ for the former is less than the one for the latter, Figure 7.6(b). In fact, there are more end-users from GRT VBS entering the network and the RRUs may be assigned to end-users in poor wireless performance conditions, because for GRT services a minimum service data rate should be satisfied. For the *Standard* scenario, the lowest value of $\overline{\eta_{Cl}}$ is observed, since there is an additional limitation arising from the NetProv's physical infrastructure partition, avoiding to share the overall RRUs within the cluster.

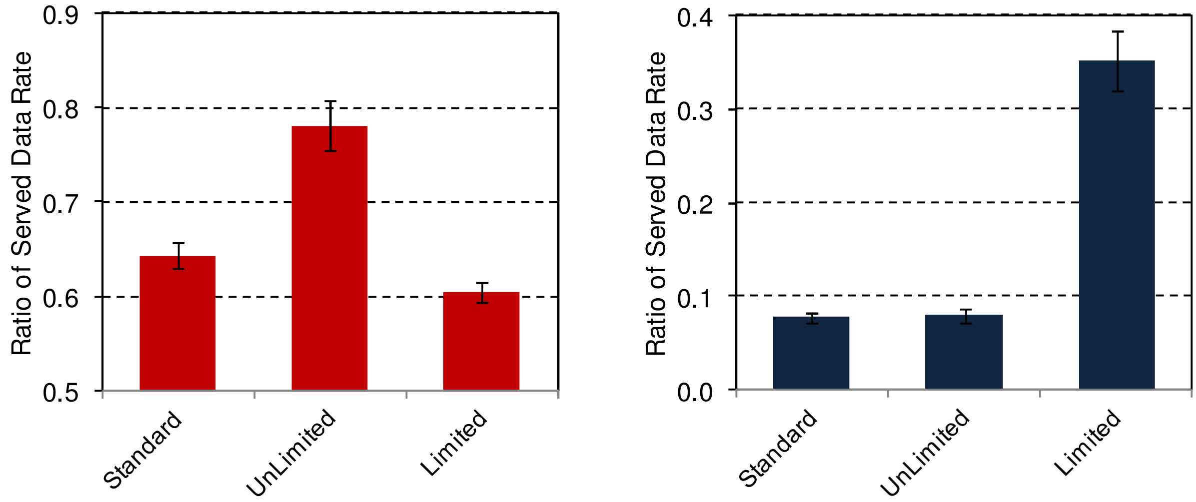



(a) *Average Cluster Data Rate* (b) *Average Cluster Utilisation*

Figure 7.6. *Average Cluster Data Rate* and *Average Cluster Utilisation* for operators' use cases.

The objective of maximising $\overline{R_{serv}^{CI}}$ , (4.16), inherent to any network mechanism, is better achieved when wireless virtualisation is applied, and in particular if a reference data rate is set for BE VBS.

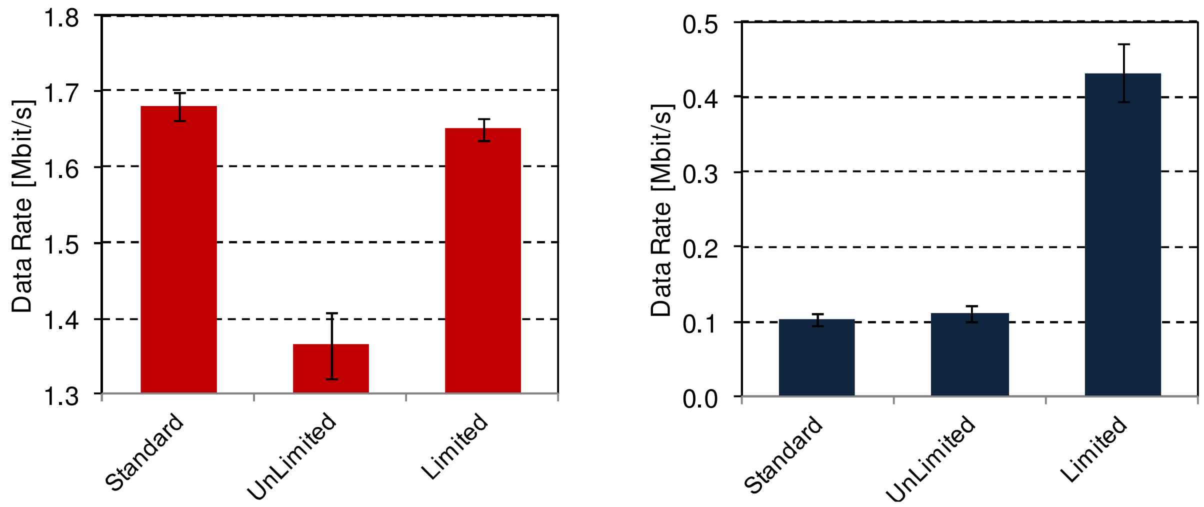
It can also be stressed that *Virtual* use cases are advantageous over *Standard*, in terms of cluster performance. $\overline{R_{serv}^{CI}}$ may increase by approximately 46% in *VNet-Limited*, and the utilisation by 13% in *VNet-unLimited* use case. Concerning the *Ratio of Data Rate Served,* $r_{serv}^{VBS}$ , given by (4.26), it can be concluded that, for GRT VBS, the ratio is minimum for *VNet-Limited*, Figure 7.7(a), though the ratio for BE VBS is the maximum for the same use case, Figure 7.7(b). These results are aligned with the considerations done before for the constraint introduced by setting the contract data rate of BE VBS: *VNet-Limited,* end-users from GRT VBS are no longer admitted as soon as $\overline{R_{serv}^{VBS}} \geq R_{min}^{VBS}$ . For the *Standard* scenario, Figure 7.7(a), although $r_{serv}^{VBS}$ for GRT services is higher than for *VNet-unLimited*, BE services have much lower $r_{serv}^{VBS}$ , Figure 7.7(b), because GRT services have always priority over BE in the *Standard* scenario, the number of BE end-users being always reduced to guarantee the minimum service data rate to GRT services.




(a) GRT services/VBS (b) BE services/VBS

Figure 7.7. *Ratio of Served Data Rate* for operators' use cases.

Analysing the *Average End-user Data Rate*, $\overline{R_{serv}^{EU}}$ , Figure 7.8, it can be concluded that in *VNet-Limited* end-users accepted in the network are served with the highest average data rate both in GRT and BE VBSs. Only in *Standard*, end-users performing GRT services have a slightly better $\overline{R_{serv}^{EU}}$ , approximately 2% higher than for *VNet-Limited*. The *VNet-unLimited* use case, though having a better ratio of served rate for GRT VBS, has $\overline{R_{serv}^{EU}}$ approximately 20% lower than the maximum achieved in *Standard*. Furthermore, for BE services, $\overline{R_{serv}^{EU}}$ in *VNet-Limited* is approximately 3 times higher than for the other two scenarios.




(a) GRT services/VBS (b) BE services/VBS

Figure 7.8. *Average End-user Data Rate* for operators' use cases.

The *Satisfaction Level on extra Capacity Requested*, $S_{ovl}^{VNO}$, calculated from (4.27), is presented in Figure 7.9(a). For the *Standard* scenario, the number of end-users not connected was computed as it is on an extra capacity requested situation, because the notion of contracted capacity cannot be applied. Furthermore, this metric is only meaningful for GRT VBSs, as end-users performing BE services are always connected. From Figure 7.9, it can be verified that $S_{ovl}^{VNO}$ increases if the BE VBS is "not limited", *VNet-unLimited* use case, since more end-users are allowed to enter into the network, because the BE VBS is always within contract. The *VNet-Limited* use case presents the lowest $S_{ovl}^{VNO}$, because $R_{ref}^{VBS}$ of BE VBS is limiting the admission of new GRT end-users, though the contracted capacity, $R_{min}^{VBS}$, is satisfied.

The *Standard* scenario presents $S_{ovl}^{VNO} = 0.56$, higher than the one obtained for *VNet-Limited*, $S_{ovl}^{VNO} = 0.49$. However, one must refer that, for the *Virtual* use cases, $S_{InP}^{VNO} = 1$ as long as the GRT VBS is operating within the contracted capacity, $\overline{R_{serv}^{VBS}} \leq R_{min}^{VBS}$, meaning that end-users of GRT VBS are always connected in this situation.

Values for *Out of Contract* are not presented graphically, as only in the *VNet-Limited* scenario the BE VBS is running out of contract, the GRT VBS always running within contract. The value obtained for *VNet-Limited*, $r_{TF}^{out} = 0.89$, can be explained due to the total amount of contracted data rate, which is close to the average capacity of the cluster.

Concerning results for delays, it must be stressed that, for the *Standard* scenario, the *Average Delay on Service Request* and the *Average Time Service Delayed* were considered as delays on extra capacity requested, since there is no defined contracted data rate.

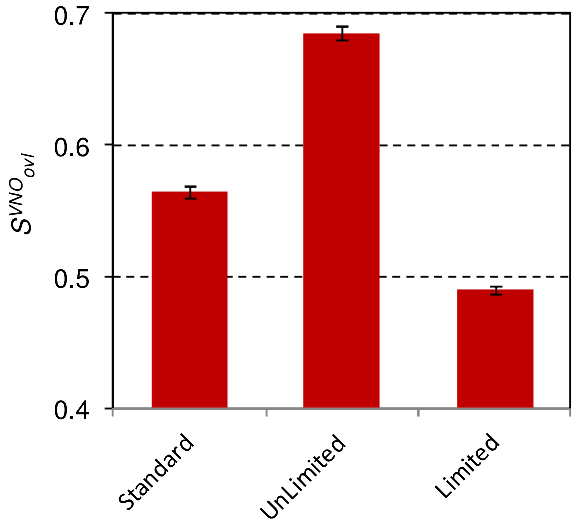



Figure 7.9. *Satisfaction Level on extra Capacity Requested* for operators' use cases.

Regarding *Average Delay on Service Request VNO*, $\overline{\tau_{VNO}^{VBS}}$, Figure 7.10 shows that the minimum value is reached for *VNet-unLimited*. For the *Standard* scenario $\overline{\tau_{VNO}^{VBS}}$ presents approximately the same value achieved for *VNet-Limited*, $\overline{\tau_{VNO}^{VBS}} \approx 50$ s, because the number of end-users that are delayed on service request is approximately the same for both use cases. When $R_{serv}^{VBS} \leq R_{min}^{VBS}$ in GRT VBS, the delay on service request for *Virtual* use cases, $\overline{\tau_{InP}^{VBS}}$, is significantly reduced. being of the order of hundreds of milliseconds in *VNet-Limited*, $\overline{\tau_{InP}^{VBS}} = 270$ ms, and zero for *VNet-unLimited*.

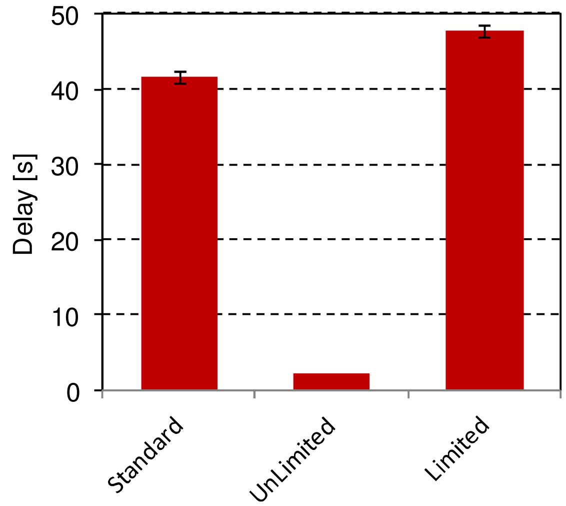
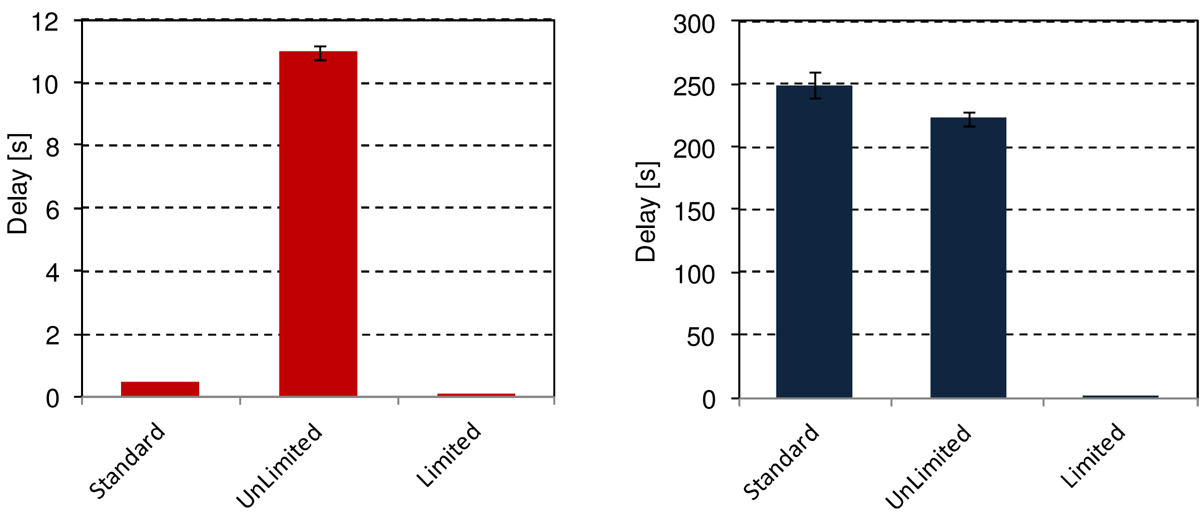



Figure 7.10. *Delay on Service Request VNO* for operators' use cases.

Summarising, it can be concluded that for *Virtual* use cases, the delay on service request is maintained under a certain limit if the VBS is operating within the contracted data rate, such limit depending on the total contracted data rate by the others VBSs instantiated in the cluster. In an extreme situation, if the other VBSs have a low contracted data rate compared to the data rate requested by end-users, the delay on service request can be zero as it is in *VNet-unLimited*.

The *Average VBS Time Service Delayed* on GRT VBS for serving data rates within the contracted capacity is of the order of milliseconds in *VNet-Limited*, $\overline{t_{int}^{VBS}} = 4$ ms, being zero on *VNet-unLimited*. If the VBS achieves the minimum data rate contracted, $\overline{t_{VNO}^{VBS}} \approx 10$ ms for *VNet-Limited*, Figure 7.11(a), reaching values of the order of dozens of seconds, $\overline{t_{VNO}^{VBS}} \approx 11$ s, for *VNet-unLimited*, because the number of end-users in GRT VBS is constrained by the contracted data rate of the BE VBS, being greater for *VNet-unLimited* than for *VNet-Limited*.

For the *Standard* scenario, $\overline{t_{VNO}^{VBS}} \approx 0.5$ s for GRT end-users, Figure 7.11(a), being approximately 250 s for BE ones, Figure 7.11(b). The former value is due to the limited number of GRT end-users, and the latter can be explained by the data rate reduction of BE end-users that is always applied, since there is no contracted data rate. On BE VBS, $\overline{t_{int}^{VBS}}$ is of the order of dozens of seconds, $\overline{t_{int}^{VBS}} = 25$ s, for *VNet-Limited*, and $\overline{t_{int}^{VBS}} = 0$ s for the other use cases. On extra capacity requested, $\overline{t_{VNO}^{VBS}}$ reaches values of hundreds of seconds, $\overline{t_{VNO}^{VBS}} \approx 238$ s for *VNet-unLimited* and $\overline{t_{VNO}^{VBS}} \approx 250$ s for *Standard*, Figure 7.11(b). The increase of *Service Delayed* for the *VNet-unLimited* scenario is also related to the number of GRT end-users in the network that force the data rate reduction of BE end-users. In summary, if the BE VBS has a reference data rate defined, *VNet-Limited* use case, $\overline{t_{int}^{VBS}}$ is under a certain limit at the expenses of degrading the *Average VBS Time Service Delayed* in the BE VBS. Moreover, if a reference data rate is not defined, *VNet-unLimited* use case, $\overline{t_{int}^{VBS}}$ on GRT VBS only increases when the *VBS Serving Data Rate* is greater than the contracted data rate, $\overline{t_{int}^{VBS}}$ for BE VBS being severely increased.




(a) GRT services/VBS (b) BE services/VBS

Figure 7.11. *Average VBS Time Service Delayed VNO* for operators' use cases.

In brief, by comparing the wireless access virtualisation supported by the proposed OnDemand-VRRA with the standard approach, in which there are multiple network operators (each owning part of the physical infrastructure), it can be said that the former allows achieving a better performance of a cluster of BSs from different RATs, enabling the provision of contracted capacity for GRT VNets. It is demonstrated by simulation that, in *Virtual* use cases, the *Average Cluster Serving Data Rate* may increase by approximately 46% and the utilisation by 13%. On the other hand, for *Virtual* use cases, the *Average VBS Serving Data Rate* of GRT VBSs is always greater than the contracted minimum, being constrained by the BE VBSs reference data rate, which tends to be followed. For the *Standard* scenario, the values achieved for *Average VBS Serving Data Rate* are the lowest for BE services, but can be greater than in *Virtual* use cases for GRT services. On the one hand, this denotes the limitation due to the split of the total cluster capacity by the two operators, and on the other hand, the uncoordinated allocation of radio resources since end-users are handled independently.

## 7.4 OnDemand-VRRA versus Fixed Radio Resource Allocation per VBS

In this section, the OnDemand-VRRA algorithm is compared with the allocation of fixed number of radio resources per VBS, one of the basic RAN Sharing strategies, presented in Section 3.3. For it, the number of end-users requesting service in each VBS is changing, meaning that the service profile changes transversely to the use cases. The Reference scenario is taken, only changing the number of VBSs that are four instead of two, and the service profiles. The considered use cases are designated by *4-Default*, *4-FSW* and *4-ViVo,* denoting the existence of 4 VBSs with different service profiles. The strategies for allocation of radio resources to VBSs OnDemand and Fixed are designated as *OnD* and *Fx*, respectively. Table 7.2 summarises the use cases considered for this analysis.

Table 7.2. OnDemand-VRRA versus Fixed Allocation use case's parameters.

<table>
<tr><th>Parameter</th><th>4-Default</th><th>4-FSW</th><th>4-ViVo</th></tr>
<tr><td>Operators</td><td colspan="3">1 NetProv, 2 VNOs, VNet scenario providing all the services</td></tr>
<tr><td rowspan="2">VNets</td><td colspan="3">4 VBSs – 2 VBS GRT, 2 VBS BE<br>1 VBS GRT provides VoIP and 1 GRT VBS provides Video<br>1 BE VBS provides Web and 1 BE VBS provides File Sharing</td></tr>
<tr><td colspan="3">Capacity contracted:<br>VBSs GRT: VoIP VBS - $R_{min}$ = 30 Mbit/s; Video VBS - $R_{min}$ = 1320 Mbit/s<br>VBSs BE : FS VBS - $R_{ref}$ = 250 Mbit/s; Web VBS - $R_{ref}$ = 1500 Mbit/s</td></tr>
<tr><td>Service Profile</td><td>VoIP(4%); Video(35%)<br>FS (3%); Web(58%)</td><td>VoIP(11%); Video(9%)<br>FS (23%); Web(7%)</td><td>VoIP(46%); Video(34%)<br>FS (6%); Web(14%)</td></tr>
</table>

Although in the *Fixed* allocation the VNO does not contract capacity to the VBS but rather an amount of radio resources, the average data rate achieved by the allocated RRUs has been considered as the amount of contracted data rate, thus, one can compare the evaluation parameters for both strategies.

Results for *Average VBS Serving Data Rate* are presented in Figure 7.12 and Figure 7.13 for GRT and BE VBSs, respectively. It can be noted that, for the *4-Default* use case, the obtained values are similar for both *OnDemand* and *Fixed*, since in VoIP VBS $R_{req}^{VBS} < R_{min}^{VBS}$. In Video VBS, *OnDemand* achieves a slightly greater value, $\overline{R_{serv}^{VBS}} = 1.32$ Gbit/s, than for *Fixed*, $\overline{R_{serv}^{VBS}} = 1.28$ Gbit/s, because *OnDemand* assures $R_{min}^{VBS}$ by allocating more RRUs to the VBS, as opposed to *Fixed* where the number of RRU is constant. Moreover, $\overline{R_{serv}^{VBS}}$ achieved by *OnDemand* for Web VBS is smaller than for the *Fixed*, since it is reduced in order to allow the GRT VBSs to reach $R_{min}^{VBS}$.

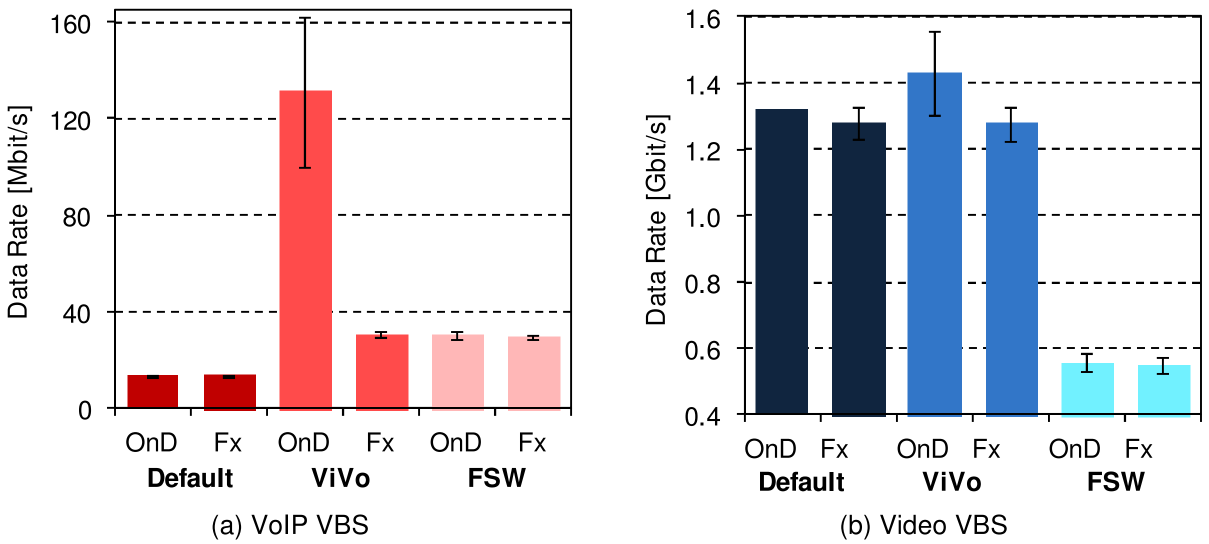



Figure 7.12. *Average VBS Serving Data Rate* of GRT VBSs for *OnDemand* vs. *Fixed*.

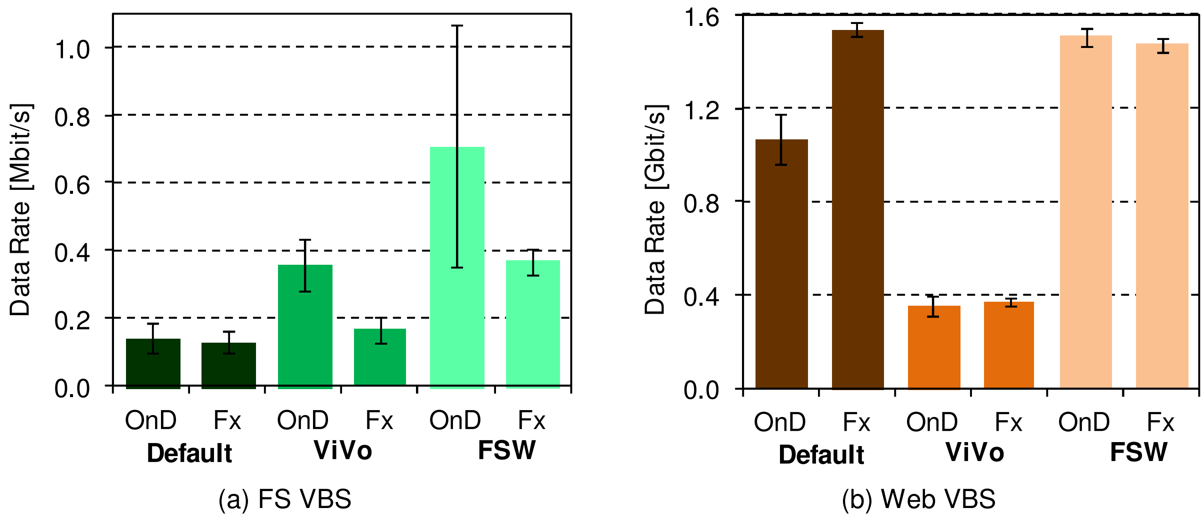



Figure 7.13. *Average VBS Serving Data Rate* of BE *VBSs* for *OnDemand* vs. *Fixed*.

Concerning the more asymmetric use cases, *ViVo* and *FSW*, it can be said that *OnDemand* performs better, because it takes advantage of the RRUs available due to low utilisation of the Web VBS for *ViVo* use case, or the Video VBSs for *FSW* use case. In *ViVo*, the VoIP VBS can achieve around

$\overline{R_{serv}^{VBS}} = 130$ Mbits/s, while *FSW* only reaches $\overline{R_{serv}^{VBS}} = 30$ Mbit/s. A similar behaviour can be found for Video and FS VBSs, the difference being approximately 10% for Video and 50% for FS. In *FSW*, the available RRUs left due to Video VBS low utilisation are available for FS and Web VBSs in *OnDemand* use case. The FS VBS achieves on average 50% more serving data rate in *OnDemand* than in *Fixed,* and the Web VBS more 3%.

Summing up, when the requested data rate for all VBSs is greater than the contracted one, *Default* use case, *OnDemand* and *Fixed* have a similar behaviour though the *OnDemand* one privileges GRT end-users in order to guarantee the data rate contracted, BE VBSs being penalised because of that. When any VBS is underutilised, *OnDemand* performs better than *Fixed*, as it allocates all the available RRUs to any VBS, because the VBSs are operating according to the established contracts.

Results for *Average Cluster Serving Data Rate*, Figure 7.14, reinforce what has been said for the *Average* VBS *Serving Data Rate*. The higher value is achieved for *Fixed* in the *Default* use case, $\overline{R_{serv}^{CI}} = 2.96$ Gbit/s, in which *OnDemand* only achieves $\overline{R_{serv}^{CI}} = 2.54$ Gbit/s. However, when the service profile changes, *ViVo* and *FSW* use cases, the *Fixed* allocation performs worse than *OnDemand* in the same but inverse relation as for *Default*, *OnDemand* being approximately 14% higher than *Fixed*.

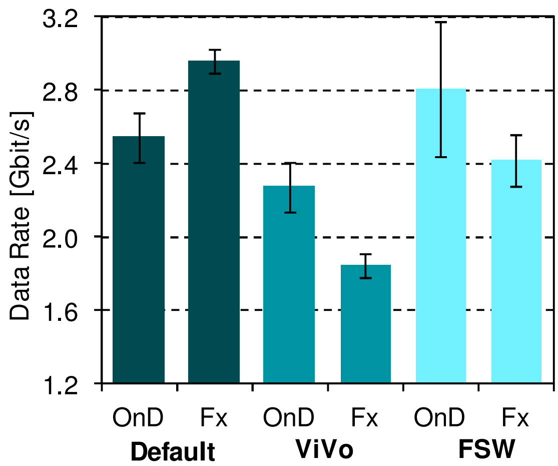


Figure 7.14. *Average Cluster Serving Data Rate* for *OnDemand* vs. *Fixed*.

Regarding *Average Cluster Utilisation*, $\overline{\eta_{CI}}$, it can be verified that *OnDemand* achieves higher values than *Fixed*, Figure 7.15, since the RRUs can be allocated to any VBS whatever the use case. In *Default*, although *OnDemand* present a higher value, $\overline{\eta_{CI}} = 0.967$, than on *Fixed*, $\overline{\eta_{CI}} = 0.896$, the performance is worse, Figure 7.14, because the main target of *OnDemand* is to guarantee the contracted data rate of GRT VBSs, allocating more RRUs to these VBSs to compensate for performance degradation of their end-users whenever the VBS is operating within the contracted data rate. For the other two use cases, both $\overline{\eta_{CI}}$ and $\overline{R_{serv}^{CI}}$ present better values for *OnDemand* than for *Fixed*, in which it is much reduced due to the impossibility of using RRUs of other VBSs. It can be concluded that when the service profile changes, the RRUs efficiency increases for *OnDemand*.

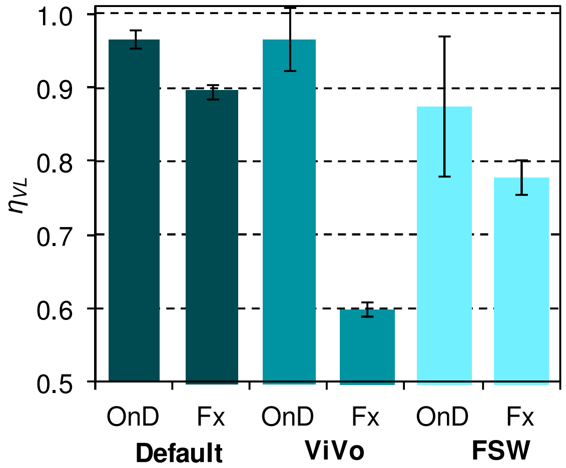



Figure 7.15. *Average Cluster Utilisation* for *OnDemand* vs. *Fixed*.

Concerning *Average End-user Data Rate*, $\overline{R_{serv}^{EU}}$, it can be noted that *OnDemand* maintains the typical service data rate for GRT VBSs, Figure 7.16, except for the *ViVo* use case in which the number of admitted end-users is high and the capacity of the cluster becomes a limitation. Instead, in *Fixed*, only when the requested data rate is below the VBS contracted data rate (VoIP VBS in *Default* and Video VBS in *Vivo*) $\overline{R_{serv}^{EU}}$ reaches the typical service data rate.

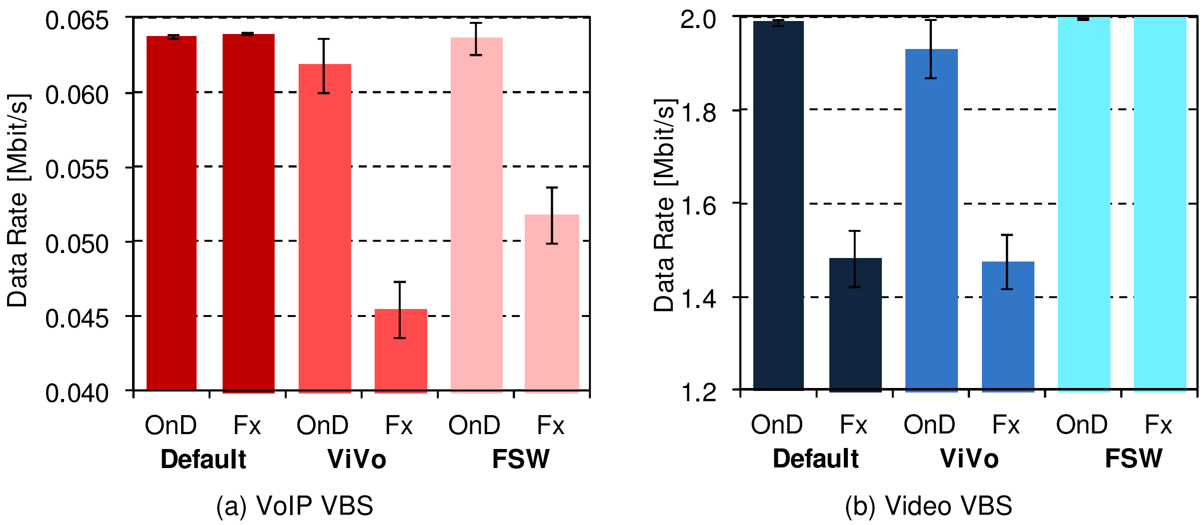



Figure 7.16. *Average End-user Data Rate* of GRT VBSs for *OnDemand* vs. *Fixed*.

For FS VBS, $\overline{R_{serv}^{EU}}$ is higher for *OnDemand* than for *Fixed* whatever the use case, Figure 7.17, which is related to the high number of end-users connected to the VBS, caused by the longer time the end-users take to complete the service. The maximum difference obtained for $\overline{R_{serv}^{EU}}$ is reached in *ViVo*, in which end-users were served with 85% more data rate with *OnDemand* than with *Fixed*.

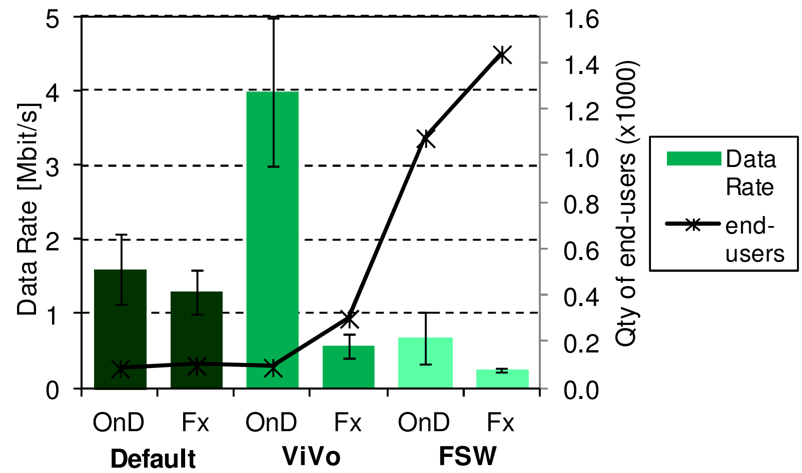



Figure 7.17. *Average End-user Data Rate* and quantity of end-users connected to FS VBS for *OnDemand* vs. *Fixed*.

The same reasoning can be applied to $\overline{R_{serv}^{EU}}$ in Web VBS, though the Web service has a lower average data volume than FS (see Table 6.2), meaning that the time end-users remain in the network should be also lower even for low VBS serving data rates. From Figure 7.18, it can be observed that $\overline{R_{serv}^{EU}}$ in Web VBS follows the trend of the achieved $\overline{R_{serv}^{VBS}}$. The maximum difference between both strategies is obtained for the *Default* use case, in which end-users are served in *OnDemand* with a data rate approximately 50% below the value achieved for *Fixed.*

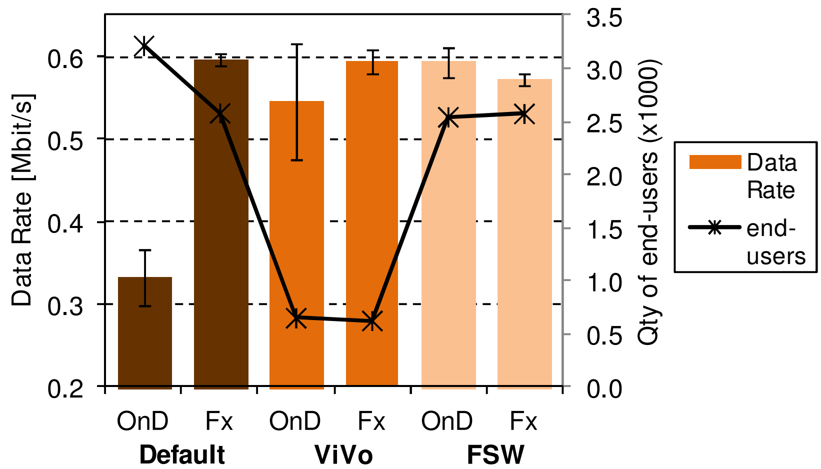



Figure 7.18. *Average End-user Data Rate* and quantity of end-users connected to Web VBS for *OnDemand* vs. *Fixed.*

Results obtained for BE VBSs' *Out of Contract*, $r_{TF}^{out}$, Figure 7.19, allow to say that, when the amount of end-users is enough to assign all RRUs allocated to VBSs in *Fixed*, BE VBSs operate within contract. This is the case of *FSW* for FS VBS, and on all use cases for Web VBS, which are not presented in Figure 7.19 because of that. When the number of end-users connected is low, in *Default* and *ViVo*, the percentage of end-users performing the FS service is 3% and 6% respectively, the percentage of time the VBS is out of contract being very high, $r_{TF}^{out} > 0.9$. This shows the low flexibility of the fixed approach, which restricts the use of RRUs to the BSs' coverage where they are allocated. The explanation for high values of $r_{TF}^{out}$ for the *Default* use case when *OnDemand* is applied, $r_{TF}^{out} > 0.96$ for FS VBS and $r_{TF}^{out} = 1$ for Web VBS (not shown in Figure 7.19), is related to the need of reducing BE end-users in order to guarantee the $\overline{R_{serv}^{VBS}}$ of GRT VBSs.

It is worth to note that for this set of simulations, all the time frames the VBS is operating below the contracted data rate over the total are accounted for *Out of Contract*. As referred in the analytical model, Section 4.4.1, a less stringent contract could be established for BE VBSs in order to reflect the BE behaviour in this type of VBSs, considering only out of contract situations in which the VBS operates more than a percentage of the time frames with $\overline{R_{serv}^{VBS}} < R_{ref}^{VBS}$.

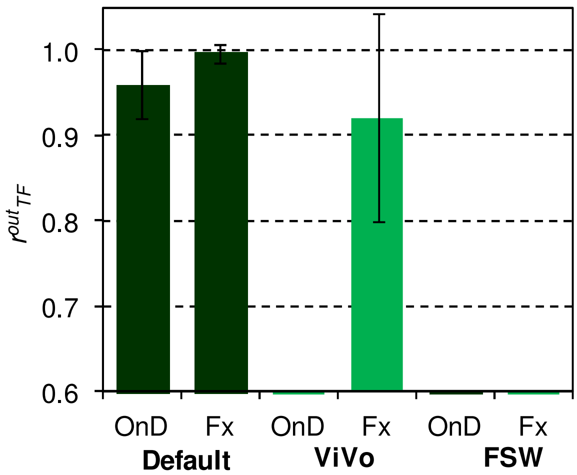
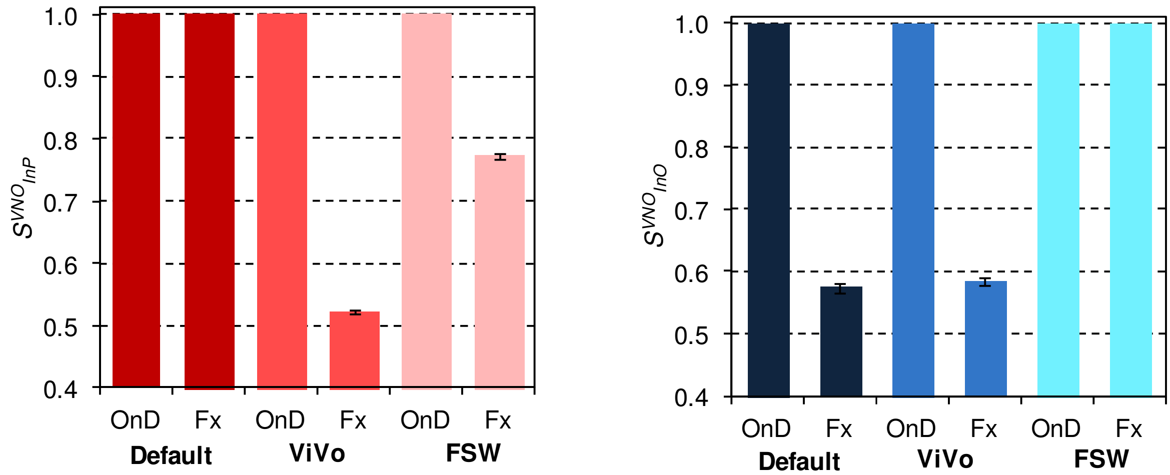



Figure 7.19. *Out of Contract* for FS VBS for *OnDemand* vs. *Fixed*.

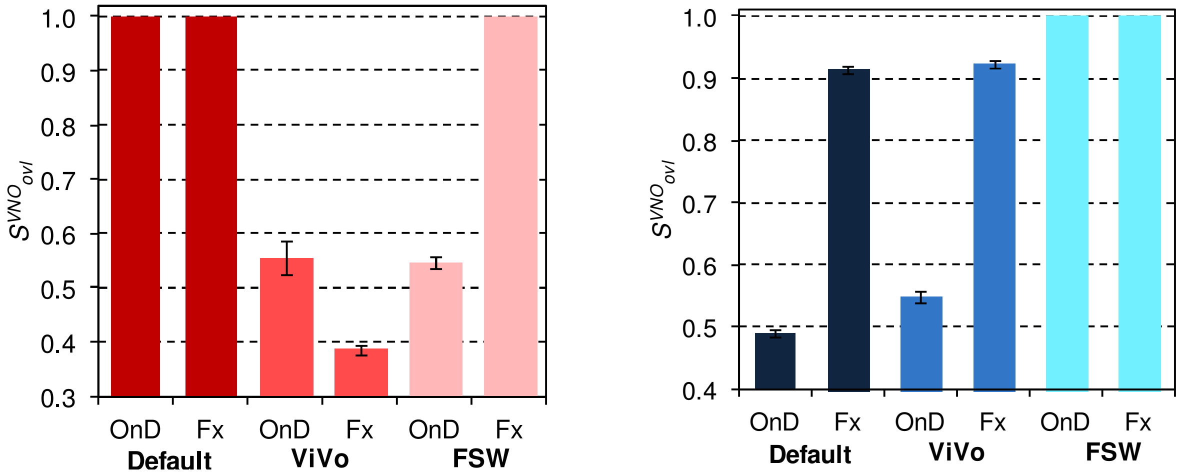
Regarding the *Satisfaction Level on the InP*, it can be observed that $S_{InP}^{VNO} = 1$ for all use cases with *OnDemand*, Figure 7.20. Nevertheless, if the strategy used to allocate RRUs is *Fixed,* it decreases considerably for values within $0.5 < S_{InP}^{VNO} < 0.8$; the unique exception is in VoIP VBS for *Default* and *FSW* in Video VBS, in which VBSs are only requesting less than the data rate contracted. Therefore all end-users are admitted to receive service and $S_{InP}^{VNO} = 1$.

The *Satisfaction Level on Extra Capacity Requested* presents $S_{ovl}^{VNO} = 1$ for *Default* in VoIP VBS and *FSW* in Video VBS, since $R_{req}^{VBS}$ is less than or very close to the contracted data rate, Figure 7.21, hence, VBSs are most of the time operating below the contracted data rate and all end-users are being admitted to the network. In VoIP VBS, for *ViVo* and *FSW* with *OnDemand*, $S_{ovl}^{VNO} \approx 0.55$, Figure 7.21(a), while with *Fixed* for *ViVo* $S_{ovl}^{VNO} = 0.38$ and for *FSW* $S_{ovl}^{VNO} = 1$. For *ViVo*, while with *OnDemand* the VoIP VBS takes advantage of the low number of end-users in BE VBSs as the unique constraint is the overall cluster capacity, with *Fixed* the additional constraint of the fixed amount of allocated RRUs decreases $S_{ovl}^{VNO}$. For *FSW* with *OnDemand*, the amount of $R_{req}^{VBS}$ for BE VBSs is high and the VoIP VBS is limited to $R_{min}^{VBS}$, since it is not possible for the BE VBSs to achieve $R_{ref}^{VBS}$. With *Fixed,* due of the reduced number of end-users in VoIP VBS it can never reach the contracted data rate. In Video VBS, beyond the *FSW* use case commented above, $S_{ovl}^{VNO}$ decreases for values on the order of 0.5 with *OnDemand*, and increases for 0.9 with *Fixed*, which is the expected behaviour for both RRU allocation strategies in *OnDemand*, motivated by the need of limiting the *VBS Serving Data Rate* when the contract is achieved, in order to satisfy the contracts of the other VBSs, or due to the limitation of the physical cluster capacity. With *Fixed*, because most of the time the VBS is not operating on extra capacity, the number of end-users not admitted in this situation is restricted.




(a) VoIP VBS (b) Video VBS

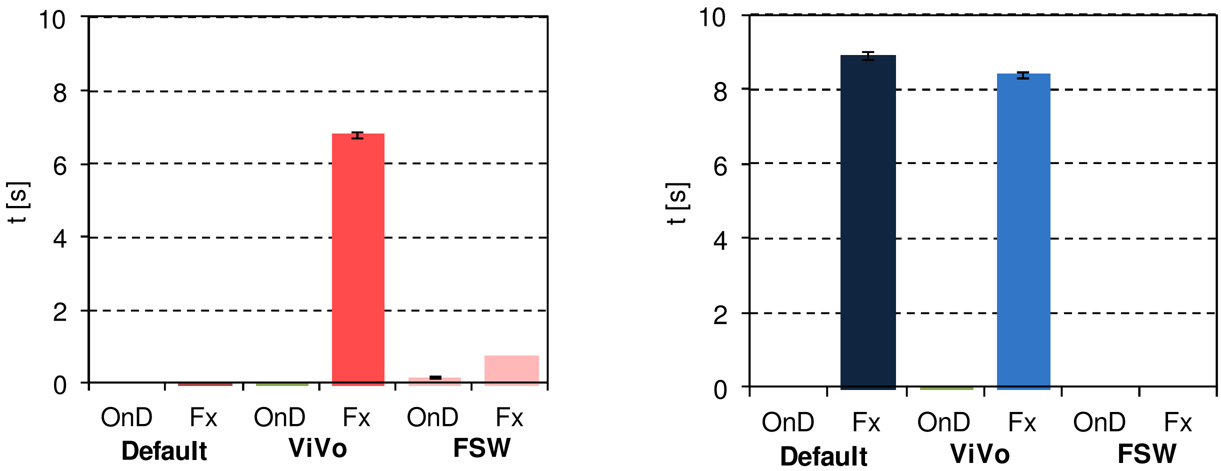
Figure 7.20. *Satisfaction Level on the InP* for *OnDemand* vs. *Fixed*.




(a) VoIP VBS (b) Video VBS

Figure 7.21. *Satisfaction Level on Extra Capacity Requested* for *OnDemand* vs. *Fixed*.

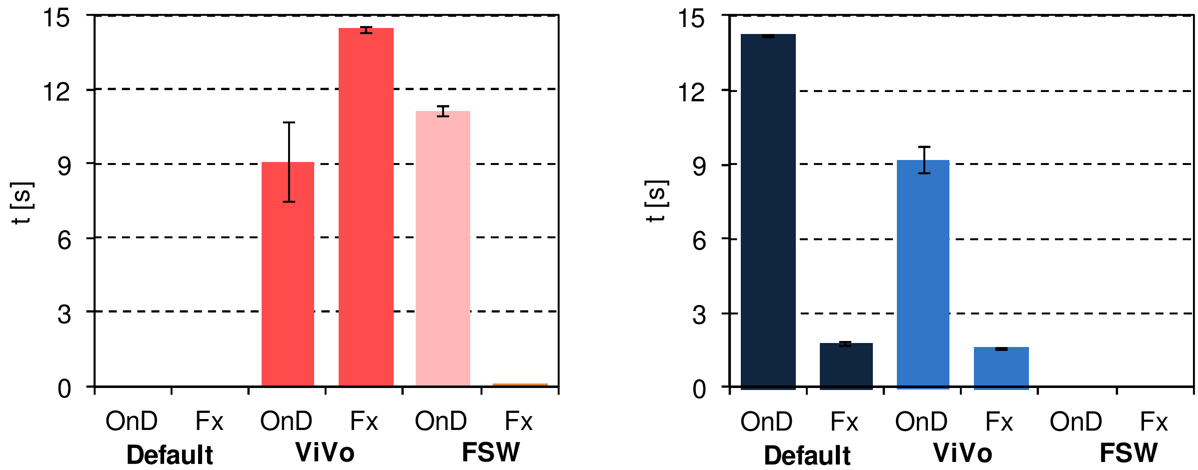
With the *Fixed* allocation of RRUs, the *Average Delay on Service Request InP,* $\overline{\tau_{InP}^{VBS}}$, is in the order of seconds in VoIP and Video VBSs for *ViVo*, and in Video VBS for *Default*, Figure 7.22. For *ViVo*, in VoIP VBS, the reason is related to the variations in $\overline{R_{serv}^{VBS}}$ due to degradation of radio conditions that are not compensated and can cause the VBS to operate with $\overline{R_{serv}^{VBS}} < R_{min}^{VBS}$ in some time frames. The other situations are related to the fact that the VBSs cannot reach $R_{min}^{VBS}$, still they cannot connect all end-users requesting a service. It must be stressed that when using the *OnDemand* strategy for VBS RRUs allocation, end-users are not delayed on service request while $\overline{R_{serv}^{VBS}} \leq R_{min}^{VBS}$.




(a) VoIP VBS (b) Video VBS

Figure 7.22. *Average Delay on Service Request InP* for *OnDemand* vs. *Fixed*.

Regarding the *Average Delay on Service Request VNO*, $\overline{\tau_{VNO}^{VBS}}$, it can be seen that in general delays are greater with *OnDemand* than with *Fixed*, Figure 7.23, because with *OnDemand* end-users are not admitted when $R_{serv}^{VBS} > R_{min}^{VBS}$ and in the other VBSs the contracted capacity is not satisfied. On the other hand, in *Fixed*, end-users are admitted till the RRUs allocated to the VBSs are all assigned.




(a) VoIP VBS (b) Video VBS

Figure 7.23. *Average Delay on Service Request VNO* for *OnDemand* vs. *Fixed*.

It can be verified that, with *OnDemand*, end-users in GRT VBSs experience delays of the order of seconds when accessing the network on extra capacity request, Figure 7.23. Obviously, this is not valid in the use cases in which $R_{req}^{VBS}$ is lower than the contracted data rate, namely, in VoIP VBS for *Default* and Video VBS for *FSW*. With *Fixed*, $\overline{\tau_{VNO}^{VBS}}$ is in the order of 15 s for *ViVo* in VoIP VBS, Figure 7.23(a), being around 1.5 s for *Default* and *ViVo* in Video VBS, Figure 7.23(b). For the other situations, $\overline{\tau_{VNO}^{VBS}} = 0$ s, since VBSs cannot reach the contracted data rate, hence, delays are not accounted for this parameter.

The end-users performing service in VoIP VBS are only delayed with *Fixed*, $\overline{t_{int}^{VBS}}$ being of the order of tens of seconds, when the VBS is operating within the contracted data rate, Figure 7.24. On extra capacity request, $\overline{t_{VNO}^{VBS}} = 3.3$ s in VoIP VBS, for *ViVo* (not shown in Figure 7.24), because the oscillations of end-user data rate, due to wireless medium variability, may cause $R_{serv}^{EU} < s_{min}^{R}$ and the service to be delayed.

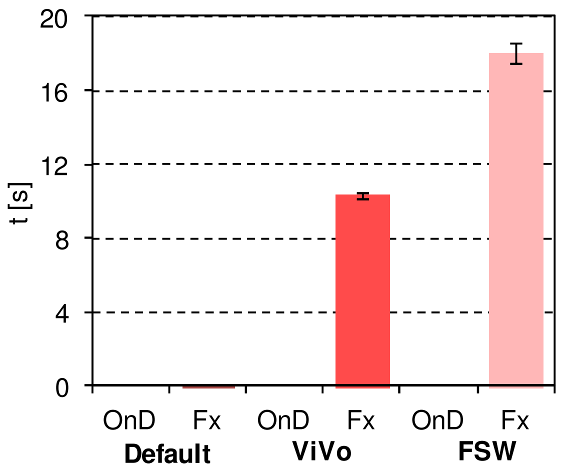
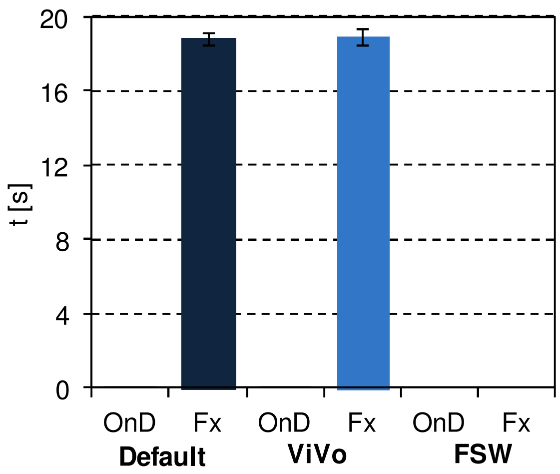


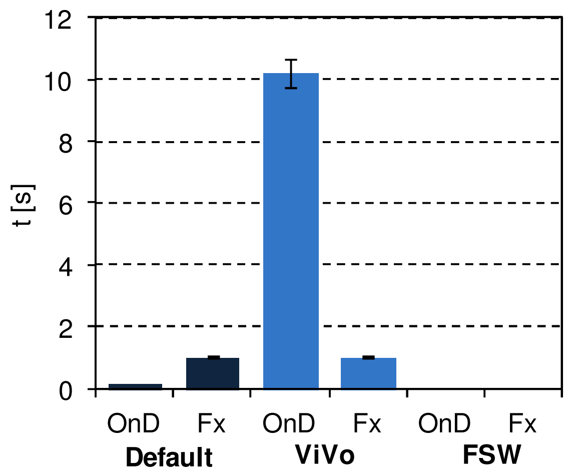

Figure 7.24. *Average VBS Time Service Delayed InP* of VoIP VBS for *OnDemand* vs. *Fixed*.

For Video VBS, an identical behaviour is verified when the VBS is operating within the contracted data rate, Figure 7.25(a). Still, instead of *ViVo* and *FSW*, the delay is observed for *Default* and *ViVo* as they are the use cases in which $R_{serv}^{VBS} < R_{min}^{VBS}$ in Video VBS. On extra capacity, the delay is also perceived for *ViVo* with *OnDemand*, $\overline{t_{VNO}^{VBS}} = 10$ s, Figure 7.25(b). In fact, for *ViVo*, there is a high number of end-users that can enter the network even when the VBS is operating on extra capacity.

In FS VBS both with *Fixed* and *OnDemand*, $\overline{t_{VNO}^{VBS}}$ is considerable, of the order of hundreds of seconds for *FSW*, Figure 7.26. In fact, whenever the number of end-users corresponds to $R_{req}^{VBS} >> R_{min}^{VBS}$, some of the end-users have $R_{serv}^{EU} = 0$ Mbit/s due to lack of capacity, meaning that the service is delayed.




(a) within the contracted data rate, $\overline{t_{int}^{VBS}}$




(b) on extra capacity, $\overline{t_{VNO}^{VBS}}$

Figure 7.25. *Average VBS Time Service Delayed* for Video VBS.

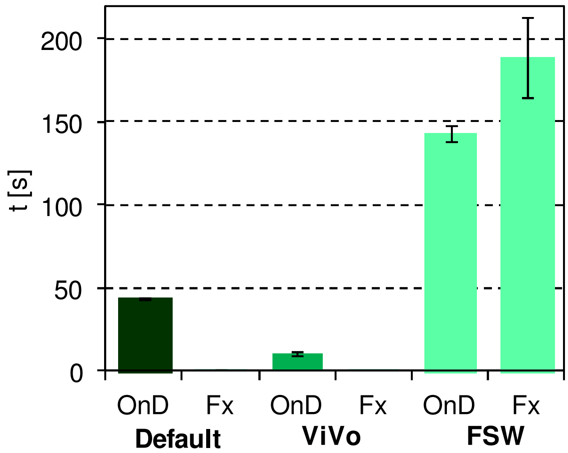



Figure 7.26. *Average VBS Time Service Delayed VNO* of FS *VBS* for *OnDemand* vs. *Fixed*.

With *OnDemand*, $\overline{t_{int}^{VBS}} = 44$ s for *Default* in Web VBS, Figure 7.27, and $\overline{t_{int}^{VBS}} = 11$ s for FS VBS when $R_{serv}^{VBS} < R_{min}^{VBS}$, Figure 7.26. In *ViVo*, although the delays are not as high as in the other use cases they reach values around 10 s for both FS and Web VBSs, Figures 7.26 and 7.27, respectively.

It can then be said that, by using *OnDemand,* end-users in BE VBSs can be considerably delayed, either on demand or when performing service, if the number of end-users in GRT VBSs is high (*Default* and *ViVo*) or when the number of end-users in BE VBSs is high (*FSW*); the former is due to the need to reduce BE end-users in order to maintain the GRT VBSs within contract, and the latter to the huge number of end-users connected to the VBS.

Globally, when comparing the *OnDemand* with the *Fixed* allocation of RRUs to the VBSs, it can be said that *OnDemand* adapts the VBSs' serving data rate to the amount of requested data rate, allowing a higher cluster serving data rate than *Fixed*, if the requested data rate decreases severely in one of the VBSs. In fact, when some of the VBSs are being requested to provide service with less than the contracted data rate, *OnDemand* performs better than *Fixed*, approximately 14% higher, as it can allocate all the available RRUs to any VBSs, since they are operating according to the established contracts. However, if the requested data rate is more than the contracted one in all VBSs the cluster performance is approximately 14% higher in *Fixed*, the main difference being experienced in BE VBSs as *OnDemand* privileges GRT end-users decreasing the data rate of BE end-users in order to guarantee the GRT VBS contracted data rate.

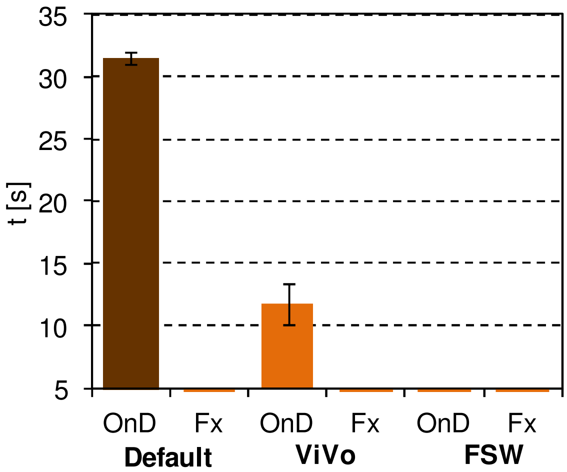



Figure 7.27. Web *Average VBS Time Service Delayed InP* for *OnDemand* vs. *Fixed.*

## 7.5 Physical versus Virtual Capacities

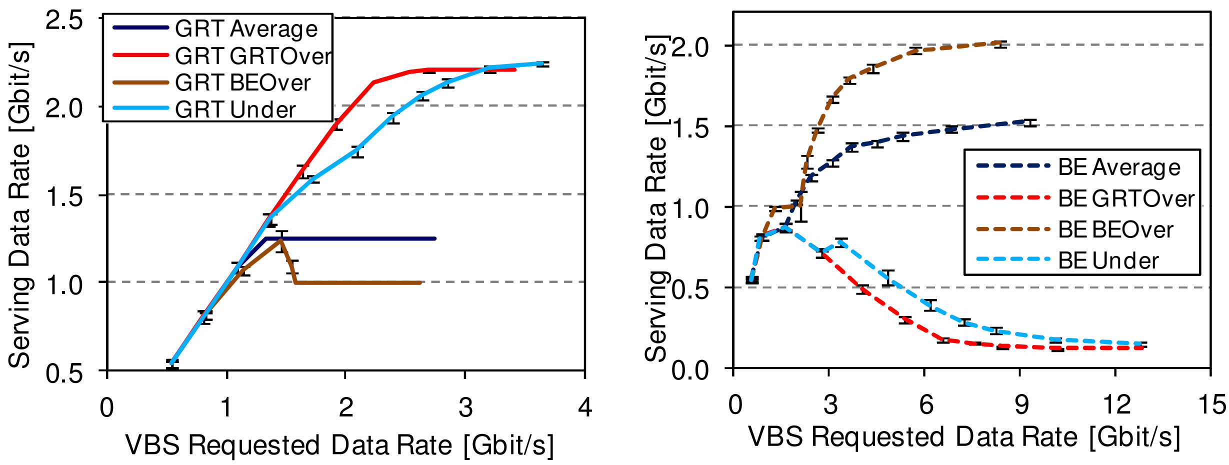
The assessment of the OnDemand-VRRA algorithm regarding situations in which the total amount of capacity contracted by VNOs is under- and over-booked is presented in this section. It is considered that an under booking situation, *Under* use case, occurs when the amount of contracted data rate by all the VBSs instantiated in the cluster is less than the average cluster capacity, i.e., the data rate the cluster can provide when the modulation and coding schemes applied to all the RRUs within the cluster is between the second and third higher data rates. Two over booking situations were considered, *GRTOver* and *BEOver* use cases, in which the total contracted data rate is greater than the average cluster capacity. Finally, an *Average* use case is considered to depict the situation when the contracted capacity is near the average cluster capacity.

For all use cases, the reference scenario is considered, the quantity of end-users increasing from 1 000 to 15 000 to allow simulating situations in which the total data rate requested is below and above the maximum cluster data rate. Table 7.3 presents a summary of the parameters for the four defined use cases.

Table 7.3. Contracted versus Cluster capacity related use case's parameters.

| **Parameter** | ***Average*** | ***Under*** | ***BEOver*** | ***GRTOver*** |
|---|---|---|---|---|
| **Operators** | 2 VNOs - one provides GRT services and the other provides BE services | | | |
| **VNets** | 1VNet GRT, 1VNet BE | | | |
| | 2 VBSs deployed over the cluster | | | |
| | GRT VBS provides VoIP and Video<br>BE VBS provides Web and File Sharing | | | |
| | Total capacity contracted by VBSs within cluster capacity (average) | Total capacity contracted by VBSs under cluster capacity | Total capacity contracted by VBSs over cluster capacity | |
| | GRT:<br>$R_{min}$ = 1.25 Gbit/s<br>BE:<br>$R_{ref}$ = 1.5 Gbit/s | GRT:<br>$R_{min}$ = 1.5 Gbit/s<br>BE:<br>$R_{ref}$ = 50 Mbit/s | GRT:<br>$R_{min}$ = 1.0 Gbit/s<br>BE:<br>$R_{ref}$ = 2.2 Gbit/s | GRT:<br>$R_{min}$ = 2.2 Gbit/s<br>BE:<br>$R_{ref}$ = 1.0 Gbit/s |
| **End-users** | Uniformly distributed | | | |
| | Distribution of end-users per VBS according to the service they are performing | | | |
| | From 1000 to 15000 end-users in the cluster | | | |

From Figure 7.28, it can be observed that the *Average VBS Serving Data Rate,* $\overline{R_{serv}^{VBS}}$, of GRT VBS reaches the contracted data rate for all use cases as soon as the data rate requested by end-users exceeds that value. It can also be seen that BE VBSs follow $R_{ref}^{VBS}$ whenever the GRT VBS is running with at least the minimum contracted capacity. As an example, after the GRT VBS achieves $\overline{R_{serv}^{VBS}} \geq R_{min}^{VBS}$, BE VBS's $\overline{R_{serv}^{VBS}}$ is approximately $R_{ref}^{VBS}$ for both *Average* ($R_{ref}^{VBS}$ = 1.5 Gbit/s) and *Under* ($R_{ref}^{VBS}$ = 50 Mbit/s) use cases.




(a) GRT VBS (b) BE VBS

Figure 7.28. Average VBS serving data rate for *Average*, *Over* and *Under* use cases.

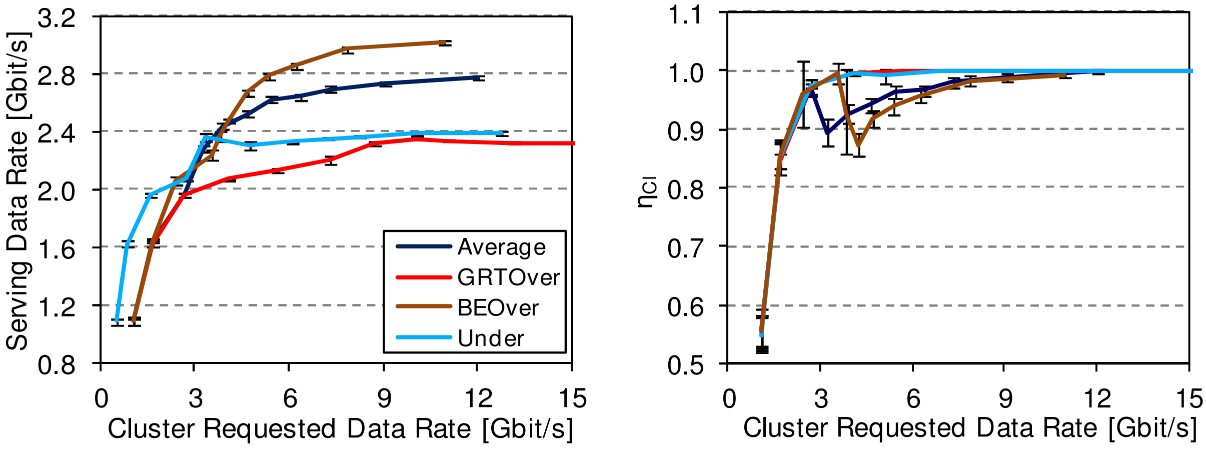
For *BEOver*, although the BE VBS is near $R_{ref}^{VBS}$, it cannot reach that value because the total contracted capacity is above the average one of the cluster. The same reason is underlying the *GRTOver* use case, in which the BE VBS cannot follow $R_{ref}^{VBS}$, because OnDemand-VRRA is allocating RRUs to GRT VBS with priority, to satisfy its minimum contracted data rate, which is near the average capacity of the cluster ( $R_{min}^{VBS} = 2.2$ Gbit/s).

The worth performance of the cluster, i.e., the minimum *Average Cluster Serving Data Rate,* Figure 7.29(a), is obtained for *GRTOver* and *Under* use cases, $\overline{R_{serv}^{Cl}} \approx 2.3$ Gbit/s. In fact, for both use cases a large number of end-users is accepted in GRT VBS, though for different reasons: a high minimum contracted data rate of GRT VBS, $R_{min}^{VBS} = 2.2$ Gbit/s for *GRTOver*, and a low reference data rate of BE VBS for *Under* use case, $R_{ref}^{VBS} = 50$ Mbit/s. The reason is, on the one hand, the large number of RRUs assigned to GRT end-users, due to the high number of GRT connected end-users, and on the other hand, due to the minimum service data rate that must be provided. Given that some GRT end-users are receiving a service in bad performance conditions, the number of RRUs providing low data rate increases, and consequently $\overline{R_{serv}^{Cl}}$ decreases.

The maximum $\overline{R_{serv}^{Cl}}$ is achieved for *BEOver,* $\overline{R_{serv}^{Cl}} \approx 3$ Gbit/s, since all the available capacity, after the provision of the GRT VBS contracted capacity is satisfied, is used by BE end-users. The value obtained for *Average*, $\overline{R_{serv}^{Cl}} = 2.7$ Gbit/s, is also interesting, because in this situation the total contracted data rate, by both VBSs, is approximately the cluster average capacity, being the traffic in each VBS shaped to fit this value.

The *Average Cluster Utilisation*, $\overline{\eta_{Cl}}$, increases with the number of end-users in all uses cases, reaching 100% when $R_{req}^{CL} > 4$ Gbit/s for *Under* and *GRTOver*, and $R_{req}^{CL} > 9$ Gbit/s for *BEOver* and *Average* use cases, Figure 7.29(b). It should be highlighted that for *BEOver* and *Average*, $\overline{\eta_{Cl}}$ may decrease due to the need to assign RRUs to all end-users requesting GRT services, since the reduction of the RRUs allocated to the end-users on BE services is made primarily to those in poor performance conditions. This is the case when the number of end-users in the cluster corresponds to $R_{req}^{VBS} \approx R_{min}^{VBS}$ for GRT VBS and $R_{req}^{VBS} \geq R_{ref}^{VBS}$ in BE VBS. As an example, for the *Average*, the decrease of $\eta_{Cl}$ is verified for $R_{req}^{Cl} = 3.25$ Gbit/s, when $R_{req}^{VBS} = 1.35$ Gbit/s in GRT VBS and $R_{req}^{VBS} = 1.9$ Gbit/s in BE VBS, the contracted capacity in each VBS being $R_{min}^{VBS} = 1.25$ Gbit/s and $R_{ref}^{VBS} = 1.5$ Gbit/s.

Analysing the *Average Cluster Serving Data Rate* and the *Cluster Utilisation* simultaneously, Figure 7.29, one can say that the best RRU efficiency is achieved when the strategy for the overall capacity provision is to limit the capacity contracted by GRT VNets, overbooking the capacity contracted by BE VNets, i.e., the *BEOver* use case. The relative inefficiency for both *Under* and *GRTOver* use cases is related to the quantity of end-users in the GRT VBS. Due to the fact that GRT services have a minimum data rate to be performed, the RRUs may be assigned to end-users in poor performance conditions. For the *Under* use case, the problem is originated by the priority in handling end-users of GRT services whenever all the VBSs in the cluster have their contracted data rate satisfied. For the *GRTOver* use case, the inefficiency is related to the value for contracted capacity of the GRT VBS, which is about 85% of the average cluster capacity, causing most of the connected end-users to be in GRT VBS.

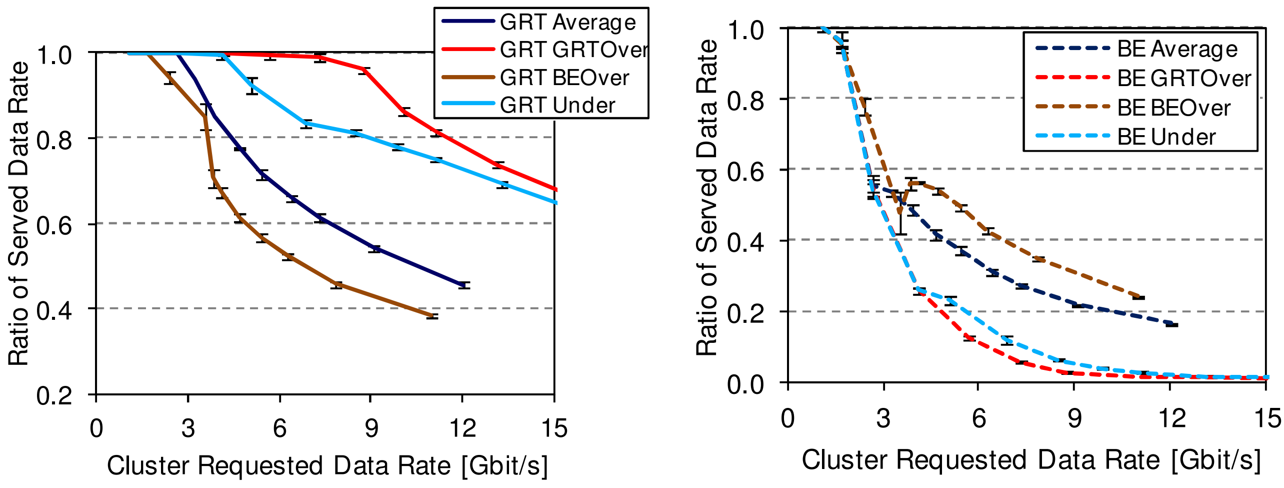



(a) *Average Cluster Serving Data Rate* (b) Average Cluster Utilisation

Figure 7.29. *Average Cluster Serving Data Rate* and *Cluster Utilisation* for *Average*, *Over* and *Under* use cases.

One can then say that a limit for the data rate contracted by GRT VBSs should be established to make efficient the use of RRUs among all VBSs deployed within the cluster. This limit should be defined as a function of the average capacity of the cluster and also of the contracted capacity for BE VBSs. Further studies on this matter are made in the assessment on VNet type mixing, Section 7.6, and on VNet quantity, Section 7.7, to validate this conclusion.

Regarding the behaviour of *Ratio of Served Data Rate*, one can verify that for GRT VBSs the data rate requested by end-users is all served till the contracted capacity, $r_{serv}^{VBS}=1$, Figure 7.30(a). After that point, the decrease of this ratio is more visible for *Average* and *BEOver* use cases, as the VBSs are limited by their contracted capacity rather than by the cluster one, as it happens for the other two use cases (*GRTOver* and *Under*). Concerning the BE VBSs, Figure 7.30(b), the decrease of the $r_{serv}^{VBS}$ is more strict for *Under* and *GRTOver*; in *Under*, this happens because the BE VBS has a very low $R_{ref}^{VBS}$, while in *GRTOver*, the reason is that the GRT VBS has a high $R_{min}^{VBS}$, not allowing the BE VBS to be served due to the limited capacity of the cluster. The minimum value of $r_{serv}^{VBS}$ observed for the *BEOver* use case is achieved when $R_{req}^{CL}$ is around 4.0 Gbit/s, which denotes the situation already highlighted in which the GRT VBS switches from $\overline{R_{serv}^{VBS}} < R_{min}^{VBS}$ to $\overline{R_{serv}^{VBS}} \geq R_{min}^{VBS}$.




(a) GRT VBS (b) BE VBS

Figure 7.30. *Ratio of served data rate* for *Average*, *Over* and *Under* use cases.

Regarding the *Average End-user Data Rate* of GRT VBS, it is maintained approximately constant, $\overline{R_{serv}^{EU}} \approx 1.68$ Mbit/s, for the use cases in which the number of accepted end-users in VBS is limited by the contracted data rate, *Average*, *BEOver* and *GRTOver* use cases, Figure 7.31(a). However, depending on the use case, there is a dependency on the requested data rate:

- in *Average*, it is always verified, since the total contracted data rate is approximately the average capacity of the cluster;
- in *BEOver*, for $R_{req}^{Cl} < 1.5$ Gbit/s and $R_{req}^{Cl} > 4$ Gbit/s;
- for *GRTOver,* till the serving data rate equals the contracted data rate.

It is worth to note that for *BEOver*, although the *Average End-user Data Rate* reaches the maximum obtained value for light traffic, $R_{req}^{Cl} < 1.5$ Gbit/s, a reduction is detected for $1.5 < R_{req}^{Cl} < 4$ Gbit/s. In fact, while the capacity requested by BE end-users is below the data rate contracted by the BE VBS, $R_{req}^{VBS} < R_{ref}^{VBS}$, the GRT VBS keeps accepting end-users. However, as soon as $\overline{R_{serv}^{VBS}} = R_{min}^{VBS} = 1$ Gbit/s in GRT VBS, while $R_{req}^{VBS} < R_{ref}^{VBS} = 2.2$ Gbit/s for BE VBS, the data rate reduction process is deactivated and the unique source to adapt the service data rate of GRT end-users are the RRUs still available, i.e., the RRUs that are not yet assigned to end-users. Hence, $\overline{R_{serv}^{EU}}$ decreases, $\overline{R_{serv}^{EU}} = 1.39$ Mbit/s as minimum, since GRT end-users might just receive the minimum service data rate due to the lack of RRUs available in the cluster. On the other hand, as soon as $R_{req}^{VBS} > R_{ref}^{VBS} = 2.2$ Gbit/s in BE VBS, the GRT VBS is limited also for end-users admission and the $\overline{R_{serv}^{EU}}$ recovers the initial value.

For *GRTOver*, the data reduction process is deactivated when $R_{req}^{VBS} > R_{ref}^{VBS} = 2.2$ Gbit/s and, as most of RRUs are already assigned (heavy traffic in the network), the end-users data rate adaptation to achieve the minimum service data rate is constrained, and some end-users can be delayed due the lack of RRUs to satisfy this value, $\overline{R_{serv}^{EU}}$ being reduced. In *Under*, something similar to *GRTOver* happens, i.e., the physical cluster capacity forces the reduction of $\overline{R_{serv}^{EU}}$ as soon as $\overline{R_{serv}^{VBS}} \geq R_{min}^{VBS}$, deactivating the data rate reduction process; however, the decreasing happens before *GRTOver*, for $R_{req}^{Cl} = 3.5$ Gbit/s, as the $R_{min}^{VBS} = 1.5$ Gbit/s is lower than for *GRTOver* use case, $R_{min}^{VBS} = 2.2$ Gbit/s.

From Figure 7.31(b), it can be observed that in BE VBS $\overline{R_{serv}^{EU}}$ decreases as $R_{req}^{CL}$ increases, mainly because all end-users requesting BE services are accounted as connected, even when there are not available RRUs in the cluster. It can be also noted that for *GRTOver* and *Under* use cases the reduction is more severe when the number of end-users corresponds to $R_{req}^{Cl} > 4$ Gbit/s, since the high number of GRT end-users accepted in the network is getting most of the RRUs in order to satisfy the minimum service data rate. For *Average* and *BEOver* use cases, the decrease of $\overline{R_{serv}^{EU}}$ is attenuated because the GRT VBS serving data rate is constrained by the BE VBS reference data rate.

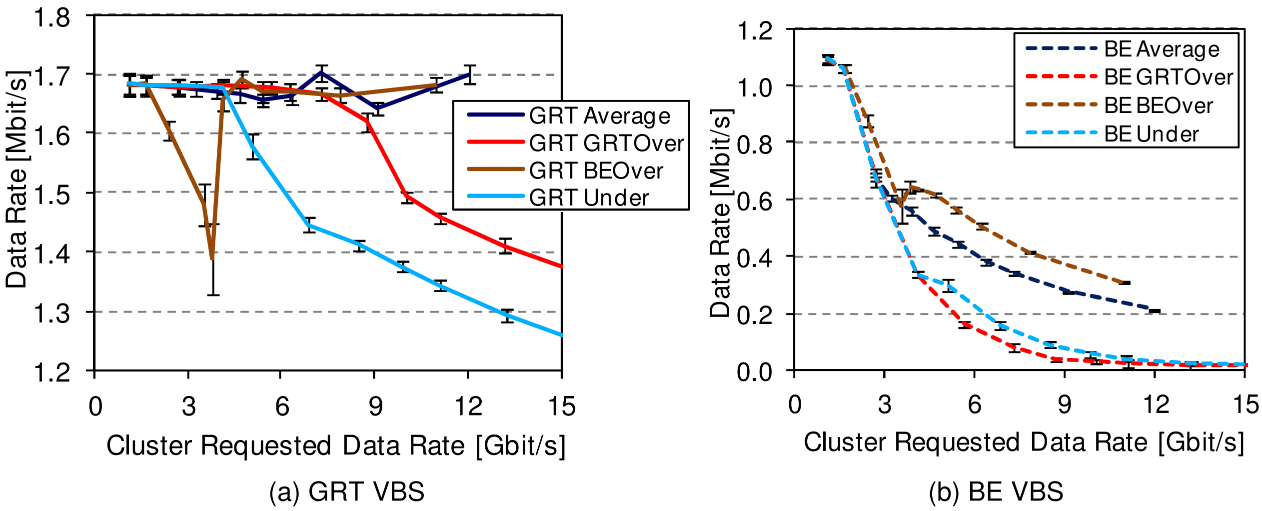
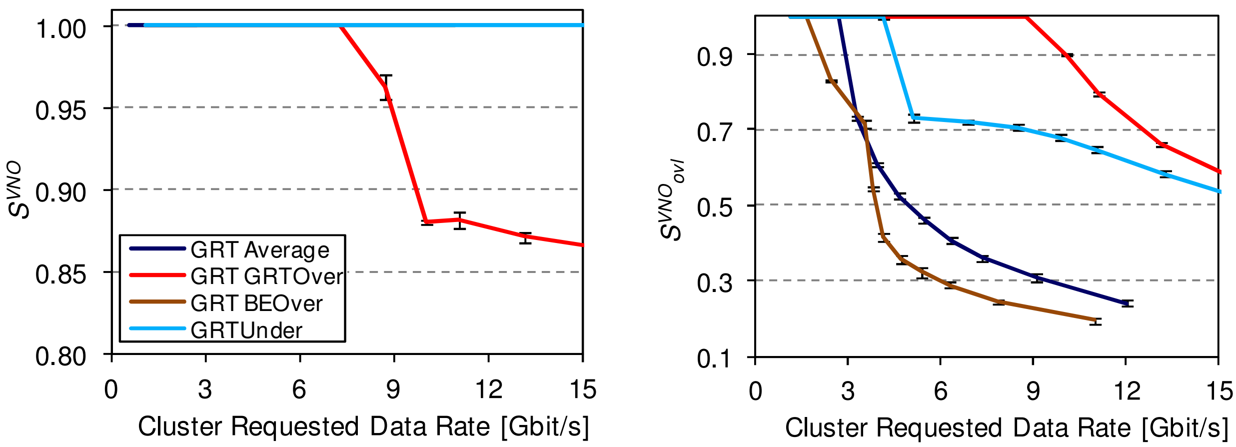



Figure 7.31. *Average End-user Data Rate* for *Average*, *Over* and *Under* use cases.

Concerning the *Satisfaction Level* in GRT VBS, it can be observed that when the VBS is operating within the contracted data rate $S_{InP}^{VNO} = 1$ meaning that all end-users are admitted, Figure 7.32(a). The unique exception is for *GRTOver*, in which the $S_{InP}^{VNO}$ decreases till 0.85 for $R_{req}^{Cl} > 8.5$ Gbit/s, corresponding to $R_{req}^{VBS} > R_{min}^{VBS} = 2.2$ Gbit/s in GRT VBS, because the number of end-users in the GRT VBS is high and, as they must receive the minimum service rate, in some locations most of the RRUs are occupied, not allowing end-users to enter the network. If end-users are requesting a service when the VBS already reaches the contracted capacity, the number of end-users not admitted to the VBS increases as $R_{req}^{Cl}$ increases, hence, $S_{ovl}^{VNO}$ decreases with $R_{req}^{Cl}$, e.g., $S_{ovl}^{VNO} = 0.2$ when $R_{req}^{Cl} = 11$ Gbit/s for *BEOver*.




(a) within the contracted data rate (b) on extra capacity request

Figure 7.32. *Satisfaction Level* of GRT VBS for *Average*, *Over* and *Under* use cases.

Regarding *Out of Contract,* one can say that, by managing the allocation of RRUs with OnDemand-VRRA, the GRT VBS is always within contract whatever the use case considered, $r_{TF}^{out} = 0$ (not presented graphically). To achieve this, the BE VBS may be out of contract for several amounts of $R_{req}^{Cl}$ depending on the use case, the pattern being almost similar for most of the use cases, Figure 7.33. This dependency is related to the value defined for the total contracted data rate compared to the cluster capacity. One can see that *Under* and *GRTOver* are the boundaries for $r_{TF}^{out}$. In *Under*, the BE VBS is always within contract as expected, since the contracted data rate has a very low value, $R_{ref}^{VBS} = 50$ Mbit/s, unlike in *GRTOver*, where only when $R_{req}^{VBS} < R_{ref}^{VBS}$ the BE VBS is within contract, being always out of contract for $R_{req}^{VBS} > 1.5$ Gbit/s in GRT VBS, because end-users in GRT VBS should maintain the minimum service data rate by reducing the data rate of BE end-users when the GRT VBS $\overline{R_{serv}^{VBS}} < R_{min}^{VBS}$ or by physical cluster capacity limitation if $\overline{R_{serv}^{VBS}} \geq R_{min}^{VBS}$.

For the two other use cases, *Average* and *BEOver*, the behaviour is similar: the BE VBS is always within contract till the requested data rate is near the contracted data rate; it presents a peak when end-users in BE VBS are requesting more than the contracted data rate and end-users in GRT VBS are requesting slightly more than the contracted data rate; and it decreases when the GRT VBS reaches the contracted data rate, limiting the number of end-users admitted and allowing the end-users in BE VBS to use the remaining capacity of the cluster and reach the contracted data rate. The main difference among use cases are the values in which the switch from one branch to another occurs, which depends of the contracted data rate for both VBSs. For *Average* $R_{min}^{VBS} = 1.25$ Gbit/s and $R_{ref}^{VBS} = 1.5$ Gbit/s for GRT and BE VBSs, respectively, the peak being of $r_{TF}^{out}$ for $R_{req}^{Cl} \approx 3$ Gbit/s. For *BEOver* $R_{min}^{VBS} = 1$ Gbit/s and $R_{ref}^{VBS} = 2.2$ Gbit/s for GRT and BE VBSs, respectively, the peak being of $r_{TF}^{out}$ for $R_{req}^{Cl}$ between 3.5 Gbit/s and 4.5 Gbit/s.

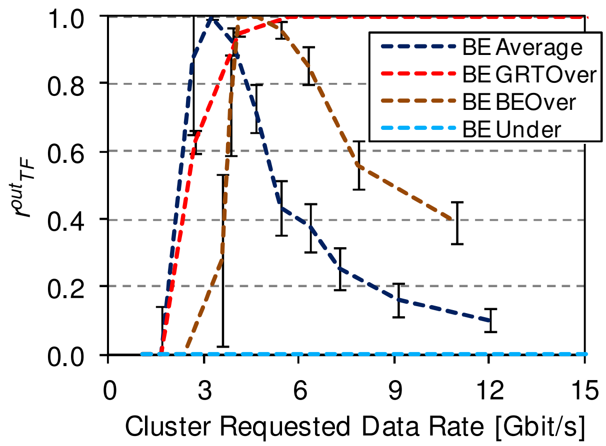
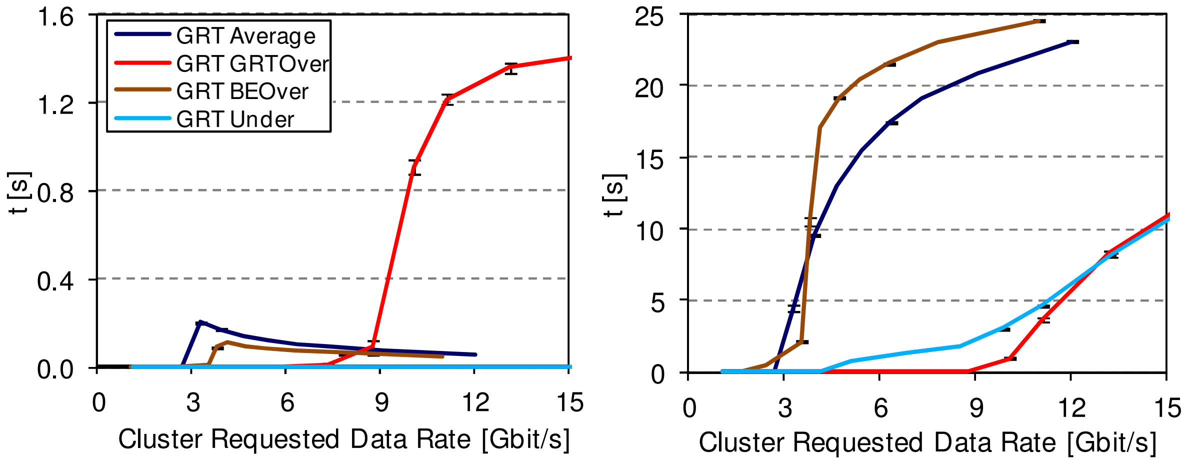



Figure 7.33. *Out of Contract* of BE VBS for *Average*, *Over* and *Under* use cases.

The *Average Delay on Service Request* when the VBS is operating within the contracted data rate, Figure 7.34(a), is of the order of tens of milliseconds, increasing to one or two hundreds of milliseconds for *Average* and *BEOver*, $\overline{\tau_{InP}^{VBS}} = 200$ ms for $R_{req}^{Cl} = 3.2$ Gbit/s in *Average* and $\overline{\tau_{InP}^{VBS}} = 110$ ms for $R_{req}^{Cl} = 4.2$ Gbit/s. The phenomenon behind this delay increase is already identified, and appears when the capacity requested in GRT VBS is near the contracted one, $R_{req}^{VBS} \approx R_{min}^{VBS}$, the capacity requested by BE VBS end-users being greater than the contracted one, $R_{req}^{VBS} > R_{ref}^{VBS}$. For *GRTOver*, $\overline{\tau_{InP}^{VBS}}$ increases up to units of seconds when $R_{req}^{Cl} > 10$ Gbit/s, because the high quantity of end-users in GRT VBS implies the utilisation of most of the RRUs in some locations, to perform the minimum service data rate to end-users already in the network.

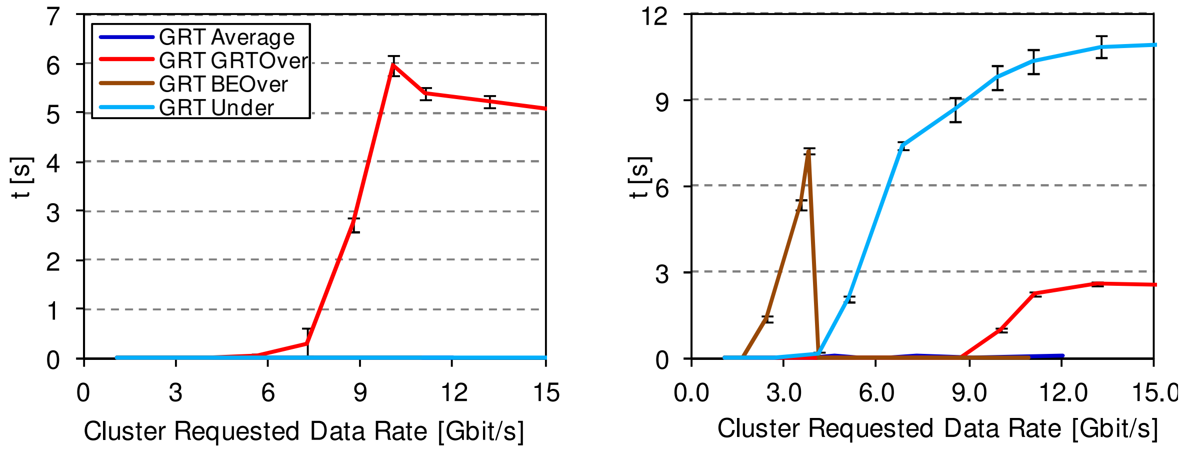
The *Average Delay on Service Request VNO* is of the order of units of seconds, increasing to tens of seconds when $R_{req}^{VBS}$ on GRT VBS is much higher than the contracted capacity, Figure 7.34(b). For *Average* and *BEOver*, this is verified for lower values of $R_{req}^{Cl}$, since the contracted data rate is also lower than for *GRTOver* use case. The phenomenon referred in $\overline{\tau_{InP}^{VBS}}$ for *GRTOver* is observed for the *Under* use case in relation to the *Average Delay on Service Request VNO*, Figure 7.34(b), as in this case the GRT end-users can be accepted behind the contracted data rate.

The *Average Service Delayed* on extra capacity request for GRT VBS presents a peak for *BEOver*, Figure 7.35, related to the fact that till that amount of data rate requested, $R_{req}^{Cl} = 3.8$ Gbit/s, the GRT VBS has admitted all end-users, since the capacity requested on BE VBS was under the contracted one, $R_{req}^{VBS} < R_{ref}^{VBS}$. Furthermore, the $\overline{R_{serv}^{VBS}} > R_{min}^{VBS}$ on GRT VBS and the data rate reduction process is not active causing the end-users in bad performance conditions to be delayed as most of RRUs are already assigned, and BE end-users cannot be reduced. The *Average VBS Service Delayed* for *GRTOver* and *Under* use cases, reflects also the behaviour of *Average Delay on Service Request*.




(a) within the contracted data rate (b) on extra capacity request

Figure 7.34. *Average Delay on Service Request* of GRT VNet for *Average*, *Over* and *Under* use cases.

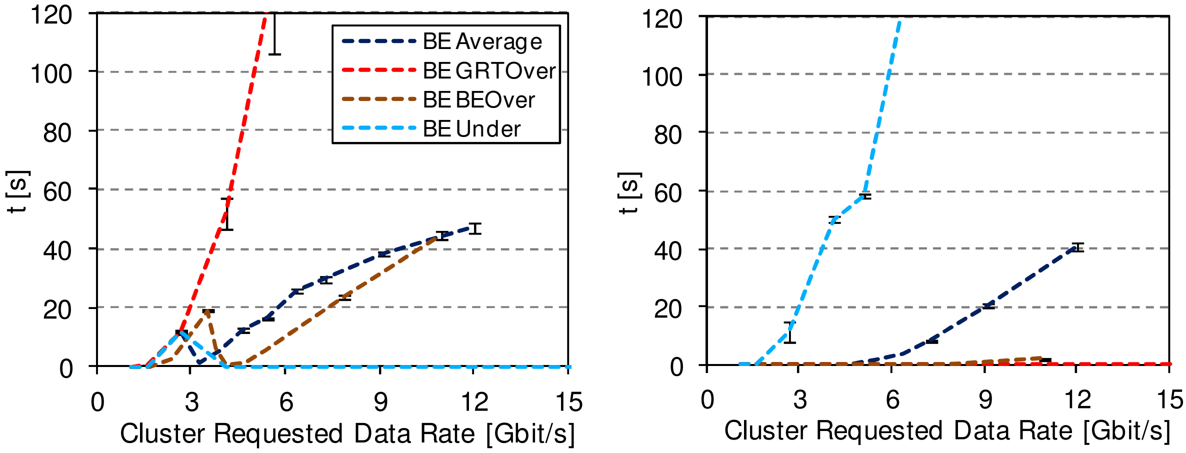



(a) within the contracted data rate (b) on extra capacity request

Figure 7.35. *Average VBS Service Delayed* of GRT VBS for *Average*, *Over* and *Under* use cases.

From Figure 7.36, one can verify that BE end-users are on average considerably delayed, in the order of tens or hundreds of seconds for $R_{req}^{Cl} > 4$ Gbit/s, because there is a high number of end-users entering the network, most of them being delayed due to the limited capacity of the cluster. For a lower number of end-users, $2.5 < R_{req}^{Cl} < 4$ Gbit/s, the *Average VBS Time Service Delayed* for BE VBS presents a peak ( $\overline{t_{int}^{VBS}} \approx 12$ s for *Average* and *Under* use cases, and $\overline{t_{int}^{VBS}} \approx 19$ s for *BEOver*) when the GRT VBS is running near the contracted data rate, $R_{req}^{VBS} \approx R_{min}^{VBS}$, since the data rate reduction process is active and can delay BE end-users to satisfy the minimum service data rate of GRT ones.

Summarising, if the total contracted data rate in the cluster is under the average capacity of the cluster, *Under* use case, the remaining capacity is used by GRT end-users, allowing the GRT VBS to achieve $\overline{R_{serv}^{VBS}} > R_{min}^{VBS}$ whenever requested. This is observed in the studied use case, in which the contracted data rate for GRT VBS, $R_{min}^{VBS} = 1.5$ Gbit/s, is greater than the contracted one for BE VBS, $R_{ref}^{VBS} = 50$ Mbit/s. The GRT VBS achieves a maximum serving data rate of $\overline{R_{serv}^{VBS}} = 2.2$ Gbit/s, since the BE VBS has the contracted data rate always satisfied. The main drawbacks of this approach are the efficiency of the radio resources, and the degradation of GRT end-user experience. Concerning the former, it can be observed that the maximum Cluster Serving rate is around 2.3 Gbit/s, under the average capacity of the cluster, 2.75 Gbit/s, the cluster utilisation being of 100%, Figure 7.29.




(a) within the contracted data rate (b) on extra capacity

Figure 7.36. *Average VBS Time Service Delayed* of BE VBS for *Average*, *Over* and *Under* use cases.

Related to the end-user experience, one can verify that: the average data rate per end-user decreases as soon as the VBS requested data rate reaches the contracted data rate, Figure 7.31(a); the end-user service delay on extra capacity request can be, on average, of the order of ten seconds, Figure 7.35(b); also the delay on service request on extra capacity request can reach values of tens of seconds. On the other hand, the BE VBS presents a BE behaviour, i.e., when the GRT VBS is light loaded and the capacity of the cluster allows it, the BE VBS serving data rate is above the contracted one, Figure 7.31(b); once the GRT VBS requested data rate is enough to increase its serving data rate just the soft constraint related to the contracted data rate of the BE VBS, which is only a reference one, avoids the GRT VBS to use all the capacity of the cluster.

When the total capacity is on the average capacity of the cluster, the *Average* use case, both VBSs can achieve the contracted data rate since the GRT VBS is limited, Figure 7.29. However, when the GRT VBS requested data rate is near the contracted one, the data rate of BE end-users is reduced to allow the GRT VBS to reach the contracted data rate and the BE VBS becomes out of contract, Figure 7.33. It is worth to note that the average data rate per GRT end-user is almost constant, independently of number of end-users in the cluster, because the number of end-users admitted to the GRT VBS is limited by the contracted data rate of both VBSs, Figure 7.31(a). Regarding delays, one can observe that, within the contracted data rate, the delay on service request is of the order of a hundred of milliseconds, Figure 7.34(a), and that the service is never delayed, Figure 7.35(a). On extra capacity request, the delay on service request increases when the number of end-users increases, being of the order of tens of seconds, Figure 7.34(b), and that the service continues to be not delayed, Figure 7.35(a).

For over booking situations, *GRTOver* and *BEOver*, the results obtained for the average capacity of the cluster are extremely different depending on the type of VBS that is contracting more capacity.

For the GRT VBS, *GRTOver* use case, the behaviour is to some extent similar to *Under*, as the number of GRT end-users is also high, though for diverse reasons. In this case, the priority given to GRT end-users, while the VBS *Serving Data Rate* is lower than the contracted capacity, is high, forcing the data rate reduction of BE end-users, Figure 7.28(b), to satisfy the contracted data rate, Figure 7.28(a). Concerning the *Average Cluster Data Rate*, Figure 7.29(a), and *Average Cluster Utilisation*, Figure 7.29(b), it can be observed that, like in *Under*, there is inefficiency in the use of RRUs in this case, related to the value for contracted capacity of the GRT VBS, which is about 85% of the average cluster capacity, leading most of the connected end-users to be in GRT VBS. Also because of that, the contracted data rate of a BE VBS has only influence for situations in which both VBSs are heavy loaded and the GRT VBS is already operating on extra capacity requested, removing the priority of GRT end-users. In the analysed case, due to overbooking, the cluster can never provide the BE VBS contracted capacity and the VBS operates out of contract as soon as the requested data rate is greater than the contracted, Figure 7.33. Regarding delays, one can observe that, within the contracted data rate, the delay on service request is approximately 1 s when the cluster requested data rate is greater than 10 Gbit/s, Figure 7.34(a). A similar behaviour is noted for service delay, Figure 7.35(a), the value being slightly higher. On extra capacity requested, the delay on service request increases when the cluster requested data rate increases above 7.0 Gbit/s, being of the order of tens of seconds, Figure 7.34(b), and that the service delayed increases also in that situation, being of the order of units of seconds, Figure 7.35(a).

For the BE VBS, *BEOver,* while the BE VBS requested data rate is less than the contracted data rate, the GRT VBS achieves *Serving* VBS *Data Rate* above the minimum contracted capacity, which is forced to be the contracted data rate as soon as the BE VBS requested data rate becomes above its contracted data rate, Figure 7.28(a). The BE VBS though does not reach the contracted data rate due to cluster capacity limit is increasing tending to it, Figure 7.28(b). The best RRU efficiency for high values of cluster requested data rate is achieved in this use case, i.e., the ratio of *Average Cluster Data Rate*, Figure 7.29(a), and *Average Cluster Utilisation*, Figure 7.29(b), is greater than for the other use cases. Although the average data rate per GRT end-user is almost constant, a decrease of about 17% is observed when the GRT *VBS Serving Data Rate* is above the VBS contracted data rate and the BE VBS requested data rate is less than the contracted for this VBS, Figure 7.31(a), because the data rate reduction process is deactivated. This fact is also perceived in the results obtained for *Average VBS Time Service Delayed*, Figure 7.35(b). The situation referred for the *Average* use case, in which the GRT VBS requested data rate is near the contracted one, is also detected here by a high value of out of contract, Figure 7.33(b), and by a slightly increase on the *Average Delay on Service Request*, Figure 7.34(a). The behaviour registered for the obtained delays is very similar to the one observed for *Average* use case except for the referred situation for *Average VBS Time Service Delayed*: within the contracted data rate the *Average Delay on Service Request* is of the order of milliseconds, Figure 7.34(a), and the service is never delayed, Figure 7.35(a); on extra capacity requested, the delay on service request increases with the cluster requested data rate, being of the order of tens of seconds, Figure 7.34(b), and the service is not delayed, Figure 7.35(a).

## 7.6 VNet Type Mixing

The OnDemand algorithm assessment, when several combinations of different VNet types are created in the physical cluster, is presented in this section. The total number of VBSs in the cluster was set to 4, the three use cases considered being characterised by the number of VNets of each type: *1GRT_3BE*, *2GRT_2BE*, and *3GRT_1BE*.

The total data rate contracted by each type of VBSs is maintained for all use cases, i.e., the total data rate contracted for GRT VBSs has always the same percentage relative to the total data rate contracted within the physical cluster. From Section 7.5., the *Average* use case was selected, representing one of the most favourable use cases for the total data rate contracted within the average cluster serving one; however, only the results obtained for *Average* are presented. All the VBSs of the GRT type provide both VoIP and Video services, FS and Web services being provided by BE VBSs.

For all use cases, the Reference scenario is considered. This allows depicting one situation in which the total data rate requested, approximately 5.5 Gbit/s on average, is above the average cluster data rate. Table 7.4 summarises the three use cases.

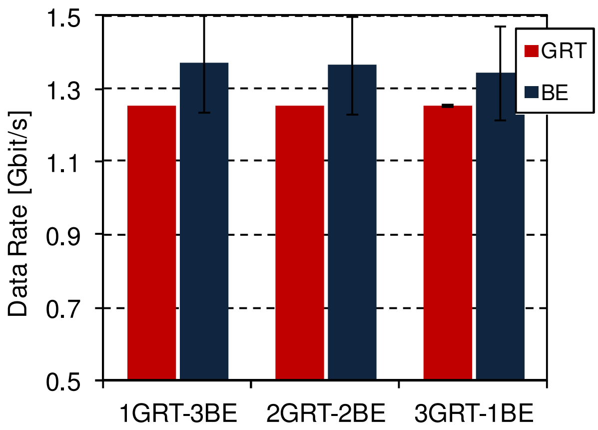
Table 7.4. Use case parameters on VNet Type Mixing.

| **Parameter** | ***2GRT_2BE*** | ***1GRT_3BE*** | ***3GRT_1BE*** |
|---|---|---|---|
| **Operators** | 1 NetProv, 2 VNOs, VNet scenario providing all the services | | |
| **VNets** | 2VNet GRT, 2VNet BE | 1VNet GRT, 3VNet BE | 3VNet GRT, 1VNet BE |
| | 4 VBSs deployed over the cluster (1 VBS per VNet) | | |
| | GRT1, GRT2:<br>VoIP,Video<br>BE1, BE2: FS, Web | GRT1: VoIP,Video<br>BE1, BE2, BE3:<br>FS, Web | GRT1, GRT2, GRT3:<br>VoIP,Video<br>BE1: FS, Web |
| | Total capacity contracted by each type of VBS is fixed | | |
| | GRT1-$R_{min}$=625 Mbit/s<br>GRT2-$R_{min}$=625 Mbit/s<br>BE1- $R_{ref}$ = 750 Mbit/s<br>BE2- $R_{ref}$ = 750 Mbit/s | GRT1-$R_{min}$=1.25 Gbit/s<br>BE1- $R_{ref}$ =500 Mbit/s<br>BE2- $R_{ref}$ =500 Mbit/s<br>BE3- $R_{ref}$ =500 Mbit/s | GRT1-$R_{min}$=425 Mbit/s<br>GRT2-$R_{min}$=425 Mbit/s<br>GRT3-$R_{min}$=400 Mbit/s<br>BE1- $R_{ref}$=1.5 Gbit/s |

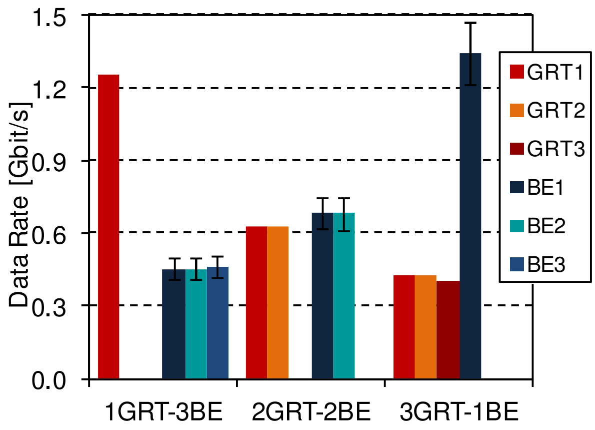
The total *Average VBS Serving Data Rate* for GRT VBSs, i.e., the sum of $\overline{R_{serv}^{VBS}}$ for all GRT VBSs, is slightly increasing with the number of GRT VBSs, Figure 7.37(a), because the adaption is made to reach the $R_{min}^{VBS}$ on each GRT VBS. For the analysed use cases, GRT VBSs are serving a limited data rate, i.e., limited by their $R_{min}^{VBS}$, due to the great amount of data rate requested by end-users, the amount of data rate above $R_{min}^{VBS}$ being residual for each VBS, Figure 7.37(a). Hence, the deviation in the total $\overline{R_{serv}^{VBS}}$ among use cases is not significant, being around 0.2% at maximum, from *1GRT_3BE* to *3GRT_1BE*. It must be stressed that, according to this rationale, increasing the number of GRT VBSs instantiated in the cluster maintaining the total contracted data rate for GRT VBSs, will cause the increase of the total $\overline{R_{serv}^{VBS}}$ for GRT VBSs, as one needs to sum the residual serving data rate of each VBS. However, as $R_{min}^{VBS}$ of VBSs will be smaller, to fit within the capacity of the cluster, the residual serving data rate will also be smaller and the percentage should not increase directly.

Concerning BE VBSs, it can be observed that $\overline{R_{serv}^{VBS}} < R_{ref}^{VBS} = 1.5$ Gbit/s for all use cases, decreasing approximately 1.8% from *1GRT_3BE* to *3GRT_1BE* with the number of GRT VBSs, Figure 7.37(a). As mentioned before, $\overline{R_{serv}^{VBS}}$ for GRT VBSs is achieved at the expense of reducing end-users connected to BE VBSs, and the cluster capacity has not enough radio resources to provide the total data rate contracted when already connected end-users are in poor performance conditions.

Observing the results obtained for the individual VBSs, Figure 7.37(b), it can be said that cluster capacity is distributed equally by the several VBSs, since the contracted data rate as well as the distribution of end-users is also equally distributed. It can be stressed that for BE VBSs, the standard deviation seems to increase when the number of GRT VBSs increases, however, if the sum of all the BE VBSs is considered, Figure 7.37(b), it is almost the same.



(a) Aggregated by VBS type



(b) Per VBS

Figure 7.37. *Average VBS Serving Data Rate* for VNet type mixing.

As a summary, one notes that the total *Average Serving Data Rate* in GRT VBSs is basically the same for all use cases and the individual VBS achieved value is always above $R_{min}^{VBS}$, independently of the use case. To achieve this, the OnDemand-VRRA algorithm manages the allocation of RRUs, causing the reduction of BE end-users and preventing BE VBSs to get their $R_{ref}^{VBS}$. It is also observed that, by increasing the number of GRT VBSs, the total GRT VBSs *Average Serving Data Rate* grows while the total *Average VBS Serving Data Rate* in GRT VBSs decreases, though not significantly in both.

The *Average Cluster Serving Data Rate* is slightly decreasing with the number of GRT VBSs, Figure 7.38, reflecting the analysis done so far for the *Average VBS Serving Data Rate*. In fact, for the *1GRT-3BE* use case $\overline{R_{serv}^{Cl}}$ achieves the maximum value, $\overline{R_{serv}^{Cl}} = 2.62$ Gbit/s, approximately 0.7% above to the *3GRT-1BE* use case, because the increase in GRT VBSs serving data rate is less than the decrease on BE VBSs.

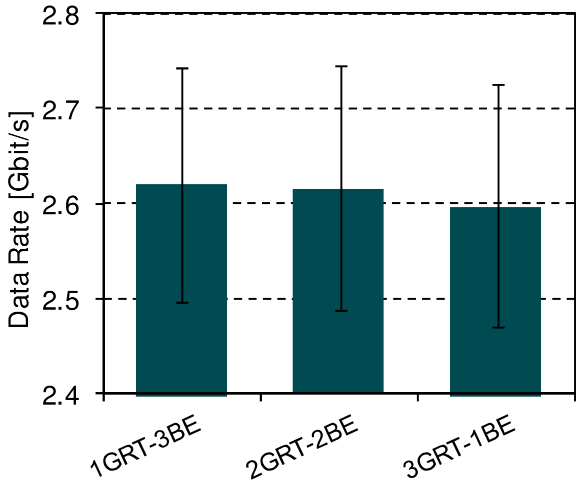



Figure 7.38. *Average Cluster Serving Data Rate* for VNet type mixing.

In accordance with this behaviour the decrease of *Average Cluster Utilisation* with the number of GRT VBSs is also not significant, Figure 7.39, $\overline{\eta_{Cl}} = 0.963$ for the *1GRT-3BE* use case to $\eta_{Cl} = 0.960$ for *3GRT-1BE*, corresponding to 0.3%. Thus, one can say that the RRU efficiency is not affected by the use case, since the decrease of $\overline{R_{serv}^{Cl}}$ is similar to $\overline{\eta_{Cl}}$.

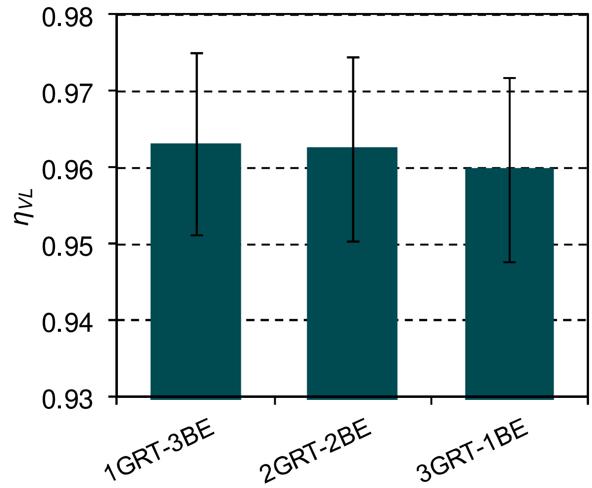



Figure 7.39. *Average Cluster Utilisation* for VNet type mixing.

Regarding the *Average End-user Data Rate*, it can be observed that end-users performing services in GRT VBSs have approximately the same value for all use cases, around 1.65 Mbit/s, Figure 7.40. Only when several GRT VBSs exist, some oscillation is perceived among VBSs, i.e., the *Average End-user Data Rate* is varying in the order of 3%. This is the expected behaviour, since all GRT VBSs provide the same services with a similar service penetration, and end-users are distributed uniformly per VBS. This is also valid for BE VBSs, which, due to the homogeneous distribution of end-users and services, are evenly connected and reduced in VBSs.

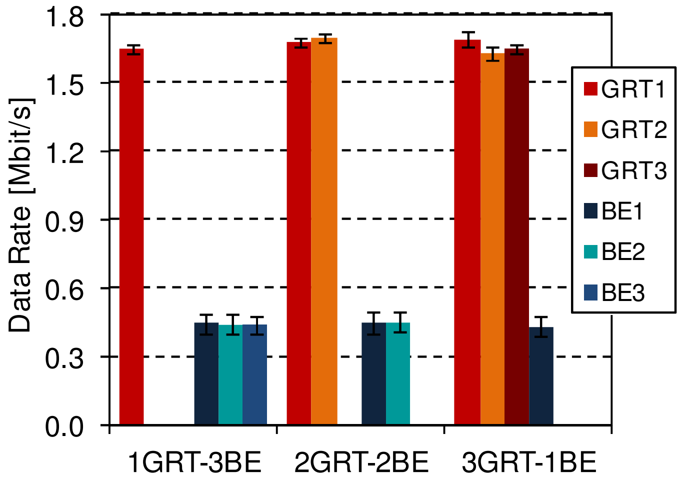



Figure 7.40. *Average End-user Data Rate* in GRT VBSs for VNet type mixing.

As GRT VBSs are always operating within the contract, only results of *Out of Contract* for BE VBSs are presented, Figure 7.41. The *Out of Contract* value is the highest for *3GRT-1BE*, $r_{TF}^{out} = 0.52$, the other use cases presenting some variability among VBSs, being approximately 15% lower on average than in the first case. This is because when there are more than one BE VBS within the cluster, *2GRT-2BE* and *1GRT-3BE* use cases, some diversity is added, allowing the data rate reduction to be made alternatively for end-users in one or another BE VBS, as it is done according to their performance condition. It must be stressed that the achieved values may not correspond directly to penalties to the InP, since it is previewed in our model that for BE VBSs only above certain percentage it is penalised.

Concerning the satisfaction level, in particular the *Satisfaction Level on the InP*, it takes the maximum value ($S_{InP}^{VNO} = 1$) for BE VBSs and for GRT VBSs when $R_{req}^{VBS} < R_{min}^{VBS}$. However, when GRT end-users request service on extra capacity, $\overline{R_{serv}^{VBS}} > R_{min}^{VBS}$, end-users begin to be delayed and the *Satisfaction Level on extra Capacity Requested* of GRT VBSs decreases to $S_{InP}^{VNO} \approx 0.45$, Figure 7.42. In this case, the satisfaction level is high when only one VBS is serving all GRT end-users, decreasing up to 8.5% for the lower value. In fact, as the total data rate contracted is fixed for each type of VBS, when more than one VBS is created in the cluster, the VBSs reaches the contract sooner causing the growth of end-users waiting for service on extra capacity requested in each VBS.

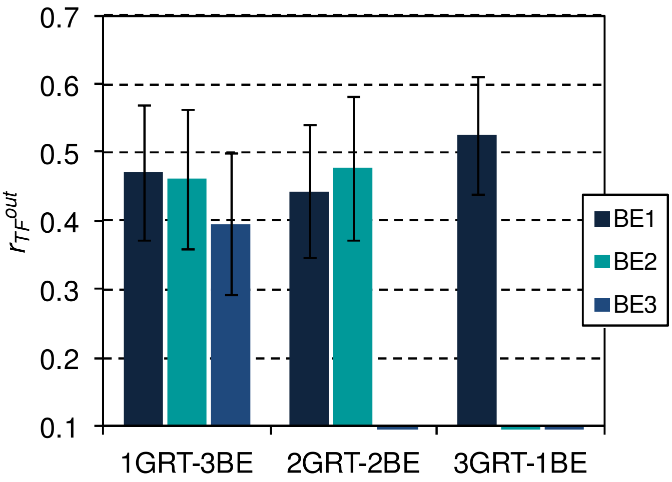



Figure 7.41. *Out of Contract* in BE VBSs for VNet type mixing.

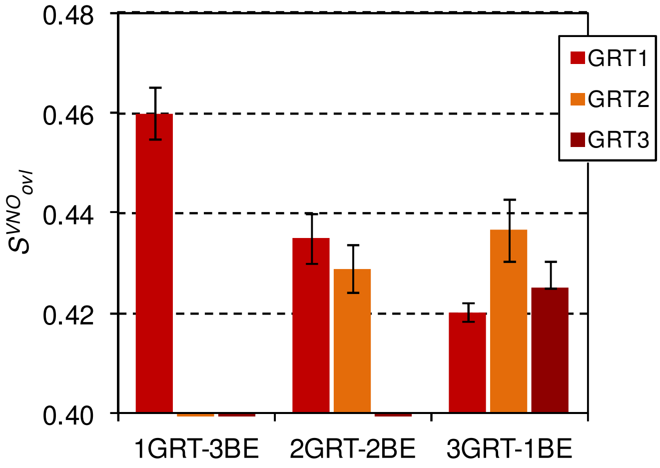



Figure 7.42. *Satisfaction Level on Extra Capacity Requested* for VNet type mixing.

The *Average Delay on Service Request InP* in GRT VBSs is in the order of a hundred of milliseconds, Figure 7.43. From the satisfaction level, it could be suggested that there is no delay in this situation, however, the delay is nonzero because the end-users not admitted on extra capacity requested may continue delayed if the VBS becomes operating under the capacity contracted and there are not radio resources available in their locations. In this case, end-users are not accounted again as not admitted and the delay is accounted as *Delay on Service Request InP*, i.e., within the contracted data rate. It can be noted that, for *1GRT-3BE*, the delay takes the minimum value, $\overline{\tau_{InP}^{VBS}} = 129$ ms, the other use cases presenting different values among them but not far from this minimum. The maximum deviation is verified between one of the VBSs in *3GRT-1BE* and another in *1GRT-3BE*, being around 20 ms.

The pattern for *Average Delay on Service Request VNO*, $\overline{\tau_{VNO}^{VBS}}$, is similar to the previous one, though values are very different, being of the order of tens of seconds, Figure 7.44. Furthermore, for multiple GRT VBSs, the VBSs having the lowest $\overline{\tau_{InP}^{VBS}}$ are the ones that have the highest $\overline{\tau_{VNO}^{VBS}}$, clearly because when the VBS is operating within the contracted data rate the delay on service request is accounted for $\overline{\tau_{InP}^{VBS}}$ and not for $\overline{\tau_{VNO}^{VBS}}$, and vice-versa.

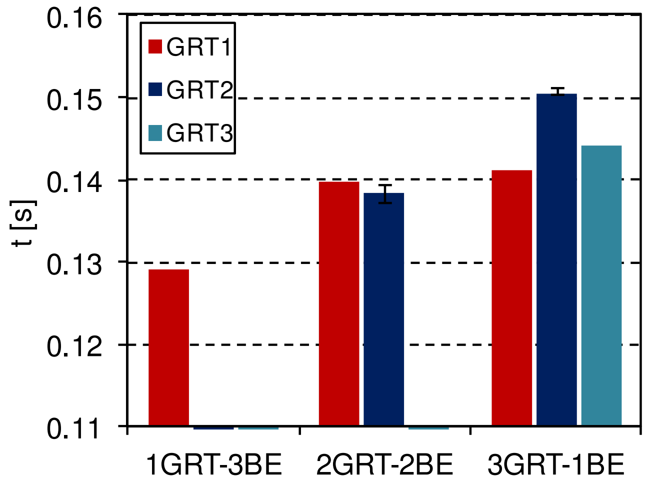



Figure 7.43. *Average Delay on Service Request InP* for VNet type mixing.

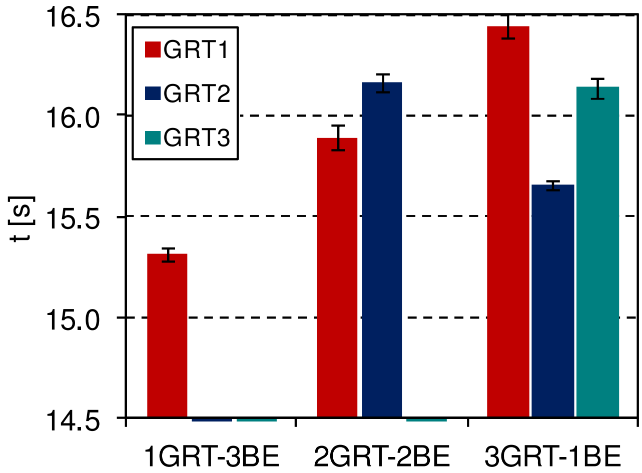
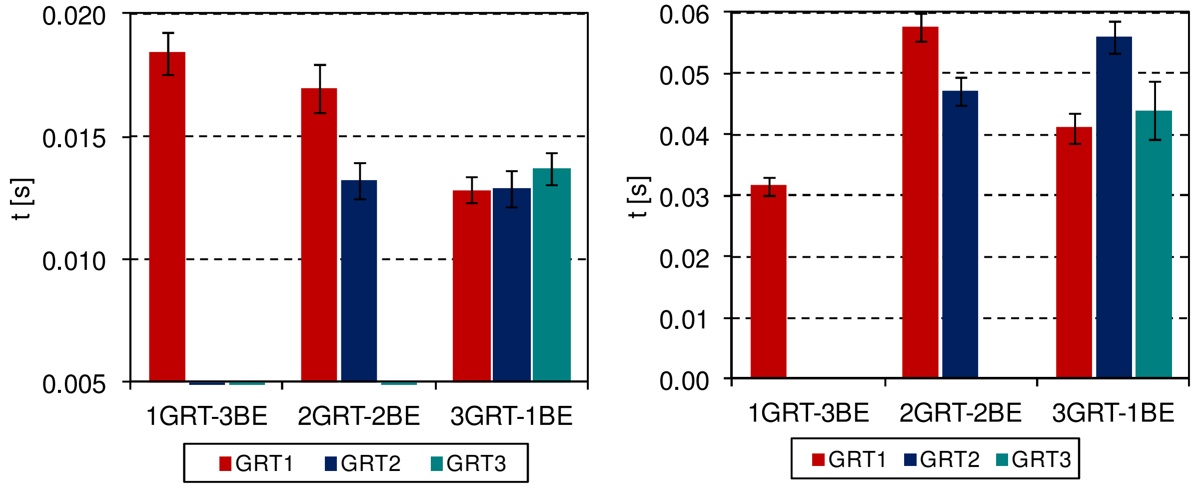



Figure 7.44. *Average Delay on Service Request VNO* for VNet type mixing.

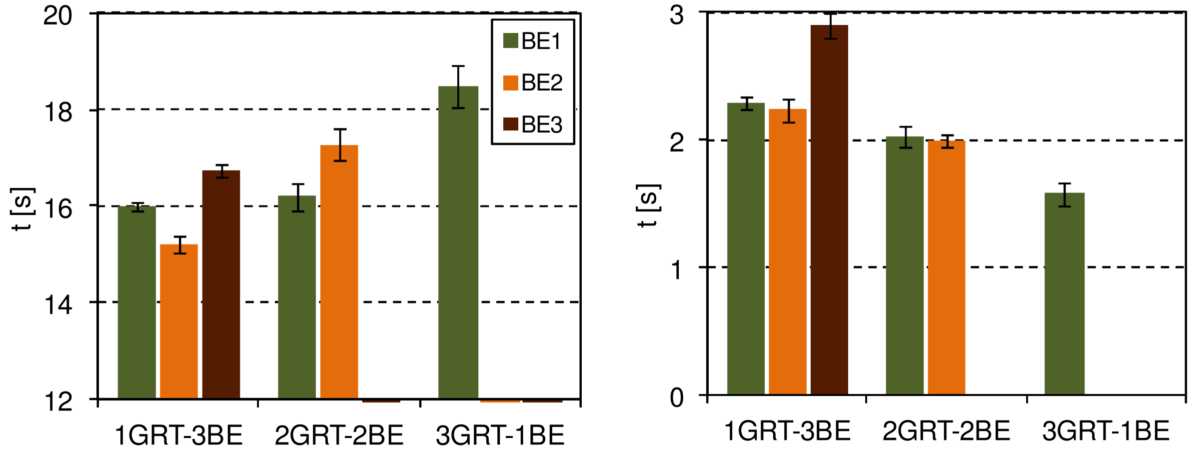
The *Average VBS Time Service Delayed* in GRT VBS is very low, in the order of tens of milliseconds, Figure 7.45. It is lower, $10 < \overline{t_{int}^{VBS}} < 20$ ms, when the VBS is operating within the contracted data rate, Figure 7.45(a), as end-users on the VBS have priority for the (re)allocation of radio resources according to their radio performance conditions. When the VBS operates on extra capacity the service delay is still low, $30 < \overline{t_{VNO}^{VBS}} < 60$ ms, because the number of end-users connected to the VBS is initially limited by the contracted data rate, since for these use cases requested data rate is high.

The *Average Time Service Delayed* in BE VBSs is much higher than for the GRT VBSs, Figure 7.46. It is of the order of tens of seconds when the VBS is operating under the contracted capacity, Figure 7.46(a), and of the order of units of seconds when $R_{req}^{VBS} > R_{ref}^{VBS}$, Figure 7.46(b). The difference of scale between them is justified by the fact that in this scenario VBSs are below the contracted data rate most of the time, what is also perceived by the obtained values for out contract.




(a) within the contracted data rate (b) on extra capacity

Figure 7.45. *Average VBS Time Service Delayed* in GRT VBS for VNet type mixing.




(a) within the contracted data rate (b) on extra capacity request

Figure 7.46. *Average VBS Time Service Delayed* in BE *VBS* for VNet type mixing.

Comparing now the several use cases, one can observe that the highest service delay is obtained for *3GRT-1BE* when the VBS is operating under the contracted capacity, getting the lowest delay when the VBS operates over capacity. For the other two use cases, with more than one BE VBS, the delay is lower per VBS, up to approximately 20%, since the services are delayed on the several VBSs just according to the performance conditions they are experienced. It is worth to note that the service delay for BE VBSs is the duration of time the end-users have not any radio resource allocated, due to the reduction strategy used to satisfy the contracted data rate of GRT VBSs or to unavailability of radio resources when the end-user initiates de service.

As a summary of the results obtained by changing the VBSs' type mixing, one can conclude the following. The total *Average VBS Serving Data Rate* of GRT VBSs is basically the same for all use cases and the individual VBS achieved value is always above the *Minimum Contracted Data Rate,* independently of the use case. It is verified that the total *Average VBS Serving Data Rate* of GRT VBSs has a maximum deviation among use cases around 0.2%, increasing slightly with the number of

GRT VBSs. To achieve this OnDemand-VRRA algorithm manages the allocation of RRUs, causing the reduction of BE end-users and preventing the BE VBSs to get their *Reference Contracted Data Rate*. The total *Average VBS Serving Data Rate* of BE VBSs is under *the Reference Contracted Data Rate* for all use cases, decreasing approximately 1.8% from *1GRT-3BE* to *3GRT-1BE.* Each GRT Average VBS *Serving Data Rate* is always above the *Minimum Contracted Data Rate* independently of the use case.

The *Average Cluster Serving Data Rate* decreases slightly with the number of GRT VBSs, reflecting the analysis done so far for the Average VBS *Serving Data Rate*. In fact, for *1GRT-3BE* use case the *Average Cluster Serving Data Rate* achieve the maximum value, $R_{serv}^{Cl} \approx 0.7\%$ above to the *3GRT-1BE*. In accordance with this behaviour, the *Average Cluster Utilisation* decreases with the number of GRT VBSs is not significantly.

The *Satisfaction Level on extra Capacity Requested* of GRT VBSs is higher when only one VBS is serving all the GRT end-users, $S_{InP}^{VNO} \approx 0.45$, decreasing up to 8.5% for the lower value.

The GRT VBSs are always operating within contract, being the *Out of Contract* value for BE VBSs maximum for the *3GRT-1BE* use case. The other use cases present some variability among the VBSs, though slightly lower than for *3GRT-1BE*.

The *Average Delay on Service Request InP* for GRT VBSs is on the order of one hundred of milliseconds, taking the minimum value for *1GRT-3BE.* The maximum deviation for the other use cases is around 20 ms. The *Average Delay on Service Request VNO* has a similar pattern to the previous one, though the values are very different being of the order of one tens of seconds. The lowest value is also achieved for the *1GRT-3BE* use case.

The *Average VBS Time Service Delayed InP* in GRT VBS is very low, in the order of tens of milliseconds. It is lower, $10 < \overline{t_{int}^{VBS}} < 20$ ms, when the VBS is operating within the contracted data rate. When the VBS operates on extra capacity requested the service delay is still low, $30 < \overline{t_{VNO}^{VBS}} < 60$ ms. The *Average Time Service Delayed VNO* for BE VBSs is much higher than for the GRT VBSs, in the order of tens of seconds when the VBS is operating under the contracted capacity, and in the order of units of seconds when the VBS is operating above the contracted data rate.

From this analysis one can say that changing the percentage of VBSs of each type instantiated in one cluster of BSs, just minor differences are perceived in the values obtained for the defined metrics, when the total number of VBSs is fixed and the relation between GRT and BE data rate contracted is maintained.

## 7.7 VNet Quantity

The main target of the following analysis is to test the algorithm when the number of VBS in the cluster increases. The Reference scenario is used for all use cases. The VBS contracted data rate is typified according to the typical data rate of the service and the selected service profile: $R_{min}^{VBS} = 10$ Mbit/s for VoIP, $R_{min}^{VBS} = 500$ Mbit/s for Video, $R_{ref}^{VBS} = 500$ Mbit/S for Web, and $R_{ref}^{VBS} = 100$ Mbit/s for File Sharing. Three situations were explored: i) the number of VBSs is increasing for all services in the same way, one VBS per service; ii) the number of GRT VBSs increases by one for each GRT service, while the number of BE VBSs maintains with two VBSs for each service; iii) the number of GRT VBSs maintains with two VBSs for each service, and the number of BE VBSs increases by one for each BE service. Table 7.5 summarises the considered use cases. Considering the analysis done so far, concerning the relation between contracted and cluster capacity, Section 7.5, use cases have been classified as in Table 7.6, allowing assessing the conclusions made before.

Table 7.5. Use cases for VNet quantity assessment.

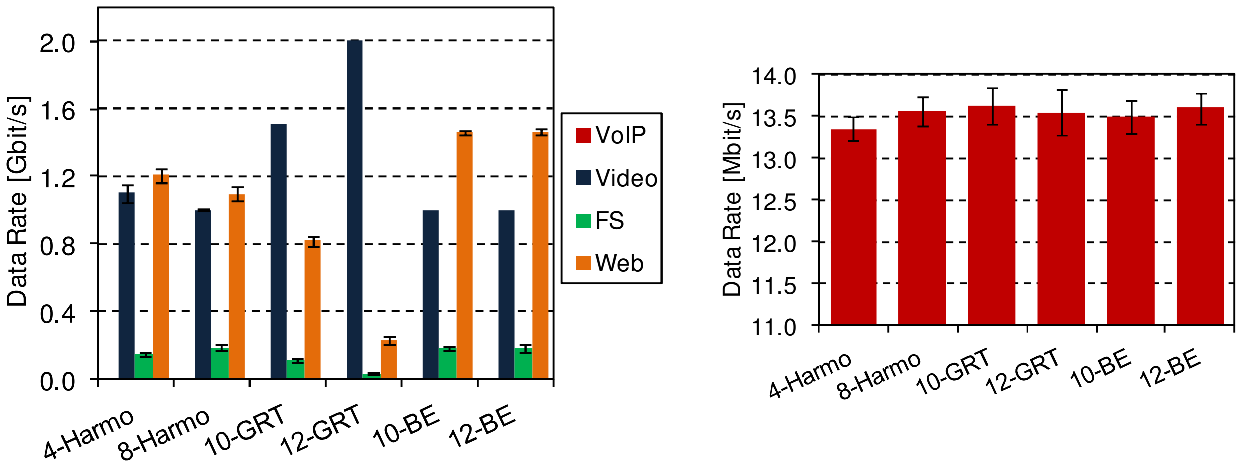
| **Parameter** | ***Harmonised*** | | ***GRT Based*** | | ***BE Based*** | |
|---|---|---|---|---|---|---|
| **Operators** | 1 NetProv, 2 VNOs, VNet scenario providing all the services | | | | | |
| **VNets** | Capacity contracted by VBS is fixed according to the service provided:<br>VoIP VBS - $R_{min}$ = 10 Mbit/s; Video VBS - $R_{min}$ = 500 Mbit/s;<br>FS VBS - $R_{ref}$ = 100 Mbit/s; Web VBS - $R_{ref}$ = 500 Mbit/s | | | | | |
| | # VNets: 4<br>1 VoIP;<br>1 Video.<br>1 FS;<br>1 Web. | # VNet: 8<br>2 VoIP;<br>2 Video.<br>2 FS;<br>2 Web. | # VNet: 10<br>3 VoIP;<br>3 Video.<br>2 FS;<br>2 Web. | # VNet: 12<br>4 VoIP;<br>4 Video.<br>2 FS;<br>2 Web. | # VNet: 10<br>2 VoIP;<br>2 Video.<br>3 FS;<br>3 Web. | # VNet: 12<br>2 VoIP;<br>2 Video.<br>4 FS;<br>4 Web. |

Table 7.6. Strategy for Data Rate Contracted for Use Case and VNet Quantity.

| **Use Case** | **Qty of VBSs** | **Total Contracted Data Rate [Mbit/s]** | **Strategy for Contracted Data Rate** | **Comment** |
|---|---|---|---|---|
| ***4-Harmo*** | 4 | GRT VBSs: 510<br>BE VBSs: 600 | Under | *Harmonised*<br>equal number of VBSs providing each service |
| ***8-Harmo*** | 8 | GRT VBSs: 1020<br>BE VBSs: 1200 | Under/Average | |
| ***10-GRT*** | 10 | GRT VBSs: 1530<br>BE VBSs: 1200 | Average | *GRT Based*<br>more VBSs providing GRT services |
| ***12-GRT*** | 12 | GRT VBSs: 2040<br>BE VBSs: 1200 | GRT Over | |
| ***10-BE*** | 10 | GRT VBSs: 1020<br>BE VBSs: 1800 | Average/BE Over | *BE Based*<br>more VBSs providing BE services |
| ***12-BE*** | 12 | GRT VBSs: 1020<br>BE VBSs: 2400 | BE Over | |

Observing the total *Average VBS Serving Data Rate* per service, i.e., the sum of $\overline{R_{serv}^{VBS}}$ for all VBSs providing a given service, Figure 7.47, one can see that it is larger than the contracted data rate for all VBSs in *4-Harmo*, since the total contracted capacity is below the average cluster capacity, Under strategy, and the requested data rate is larger than it. For *8-Harmo*, in which one VBS is added per service, the total $\overline{R_{serv}^{VBS}}$ of Video VBSs is limited by the contracted data rate as the FS VBSs are operating below the contracted data rate.

Increasing the number of GRT VBSs, *10-GRT* and *12-GRT*, $\overline{R_{serv}^{VBS}}$ of Video VBSs increases according to the total contracted capacity of the corresponding VBSs, given that the requested video traffic is larger than the contracted data rate. On the other hand, VBSs providing BE services have their $\overline{R_{serv}^{VBS}}$ reduced in order to maintain the amount of capacity contracted by GRT VBSs. When the number of BE VBSs increases, *10-BE* and *12-BE*, Video traffic is limited by the sum of the two VBSs' contracted data rate, $\overline{R_{serv}^{VBS}}$ of BE VBSs being roughly the same for both *10-BE* and *12-BE*. Regarding the latter use case, it can be said that VBSs performance is determined by the average cluster capacity, the GRT VBSs contracted data rate being ensured. $\overline{R_{serv}^{VBS}}$ of VoIP VBSs is presented in Figure 7.47(b); it is observed that VBSs are serving approximately all the data rate requested, since the requested VoIP traffic, except for *4-Harmo,* is always less than the capacity contracted by the corresponding VBSs.

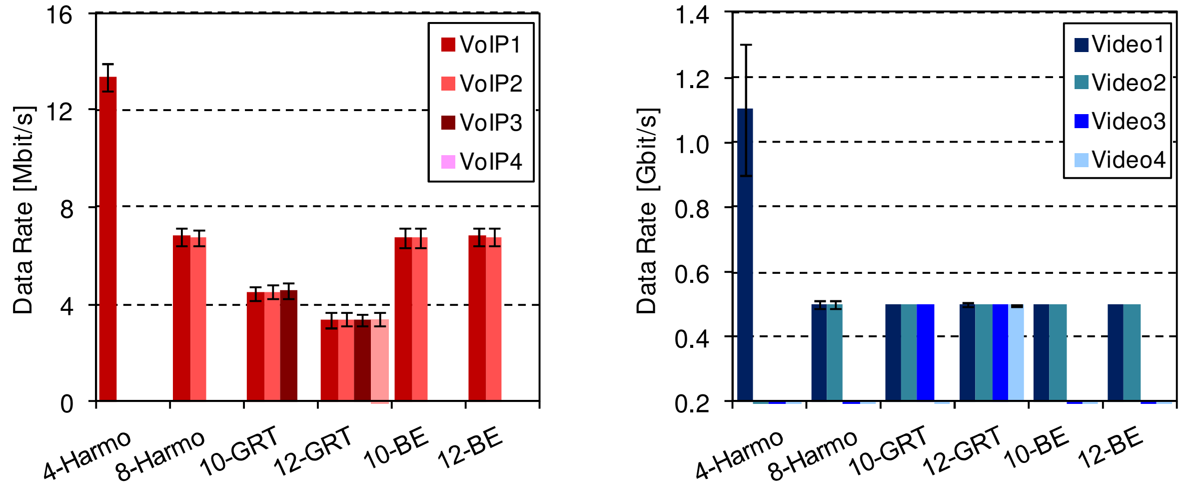



(a) All services

(b) VoIP

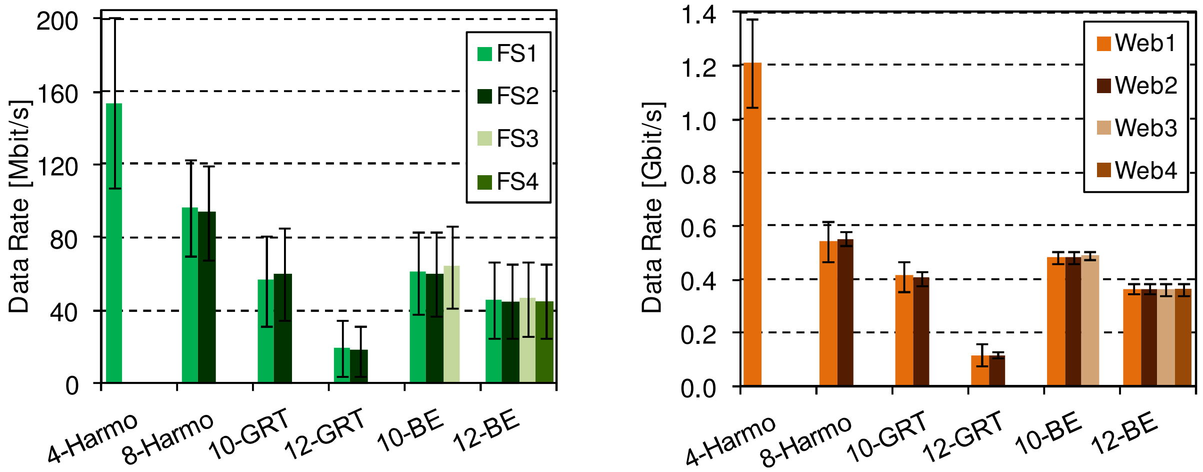
Figure 7.47. *Total Average Serving Data Rate* per type of service for VNet quantity use cases.

Concerning the individual VBSs, it can be seen that the data rate contracted by GRT VBSs is always ensured, Figure 7.48, at the expenses of reducing the *Average Serving Data Rate* of BE VBSs, Figure 7.49. When the amount of contracted data rate is essentially GRT, the *12-GRT* use case, the BE VBSs' *Average VBS Serving Data Rate* is strongly reduced, achieving around 20% of their contracted data rate, due to the need of maintaining the $R_{min}^{VBS}$ of GRT VBSs, $R_{min}^{VBS} = 2$ Gbit/s in total. If the contracted data rate is predominantly from BE VBSs, the *12-BE* use case, BE VBSs achieve an *Average VBS Serving Data Rate* closer to the contracted data rate, approximately 45% and 72% of one for FS and Web VBSs respectively. In case the total contracted data rate is below the average capacity of the cluster, *4-Harmo*, the *Average VBS Serving Data Rate* is not limited by its contracted data rate but rather by the capacity of the cluster, the extra served data rate being distributed mainly according to $R_{req}^{VBS}$ in each VBS.




(a) VoIP VBS (b) Video VBS

Figure 7.48. *Average VBS Serving Data Rate* of GRT VBSs for VNet quantity use cases.




(a) FS VBS (b) Web VBS

Figure 7.49. *Average VBS Serving Data Rate* of BE VBSs for VNet quantity use cases.

Concerning the *Average Cluster Serving Data Rate*, Figure 7.50, *BE Based* use cases achieve the highest values, $\overline{R_{serv}^{Cl}} \approx 2.65$ Gbit/s, which is in line with the analysis made in Section 7.5, since the strategy for contracted data rate in these cases is the "BE overbooking". The worst performance is obtained for *12-GRT*, $\overline{R_{serv}^{Cl}} \approx 2.2$ Gbit/s, which depicts the GRT overbooking situation. The *8-Harmo* use case presents also a bad performance for the considered amount of requested data rate, $\overline{R_{serv}^{Cl}} \approx 2.3$ Gbit/s, due to the generic rule of denying extra capacity to VBSs when there is one VBS out of contract, even if the VBS is of the BE type. In *8-Harmo*, FS VBSs are out of contract implying that there are serving data rate limitations in GRT VBSs, namely the VBSs providing Video.

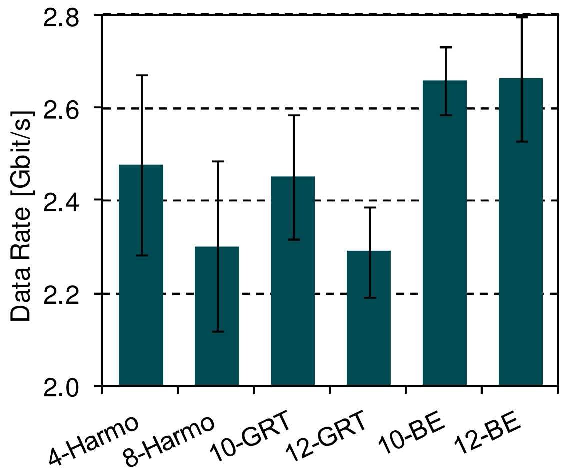



Figure 7.50. *Average Cluster Serving Data Rate* for VNet quantity use cases.

*Average Cluster Serving Data Rate* and *Average Cluster Utilisation*, Figures 7.50 and 7.51 respectively, show that the best RRU efficiency is achieved for *10-BE* and *12-BE*, which is when the strategy for the overall capacity provision is to limit the capacity contracted by GRT VNets, overbooking the capacity contracted by BE VNets. The worst RRUs efficiency is observed for *12-GRT,* in which $\overline{\eta_{CI}} = 1$, meaning that all RRUs are assigned, and $\overline{R_{serv}^{CI}} \approx 2.2$ Gbit/s. It must be stressed that these considerations are also aligned with the analysis done in Section 7.5.

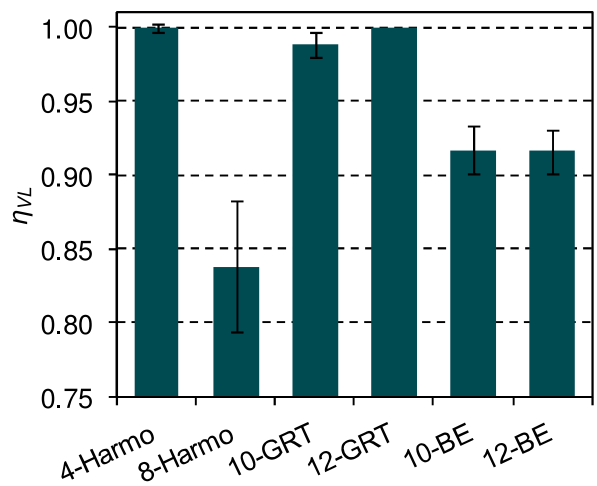
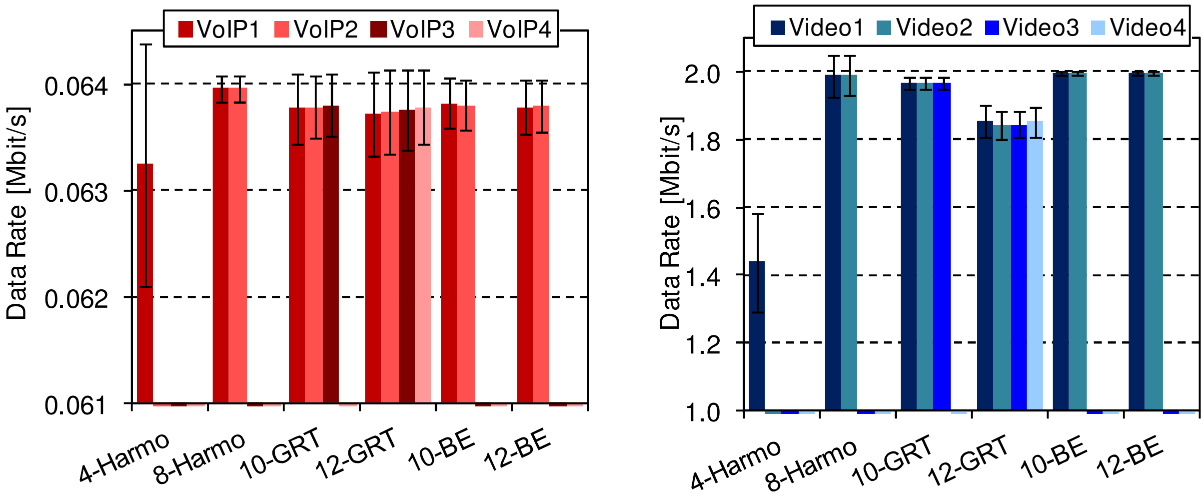



Figure 7.51. *Average Cluster Utilisation* for VNet quantity use cases.

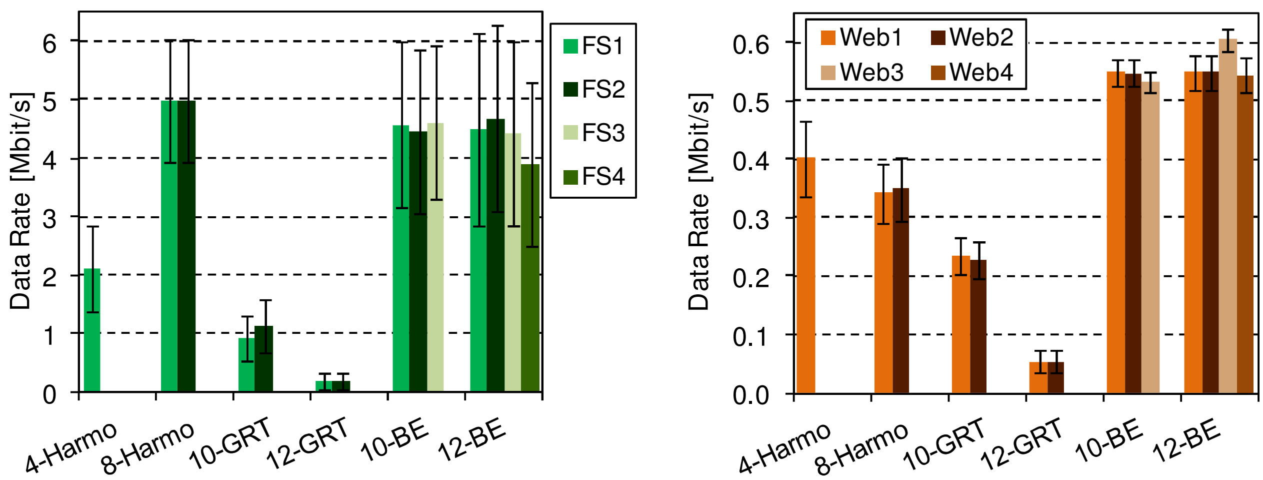
The *Average End-user Data Rate* for GRT VBSs is near the typical data rate of the service the VBSs are providing, i.e., $\overline{R_{serv}^{EU}} = 0.064$ Mbit/s for VoIP and $\overline{R_{serv}^{EU}} = 2$ Mbit/s for Video, Figure 7.52. Exceptions are registered for *4-Harmo* in both VoIP and Video VBSs, and for *12-GRT* in Video VBSs. In the former, the VBSs' contracted capacity is lower than the average cluster capacity allowing GRT end-users to receive service on extra-capacity requested, still decreasing $\overline{R_{serv}^{EU}}$ as only partially $R_{min}^{VBS}$ is ensured; in the latter, the number of end-users connected to the several GRT VBSs is high and the cluster capacity is not enough to provide the typical service data rate. However, it must be stressed that all end-users connected to GRT VBSs receive at least the minimum service data rate.




(a) VoIP VBS (b) Video VBS

Figure 7.52. GRT VBSs' *Average End-user Data Rate*.

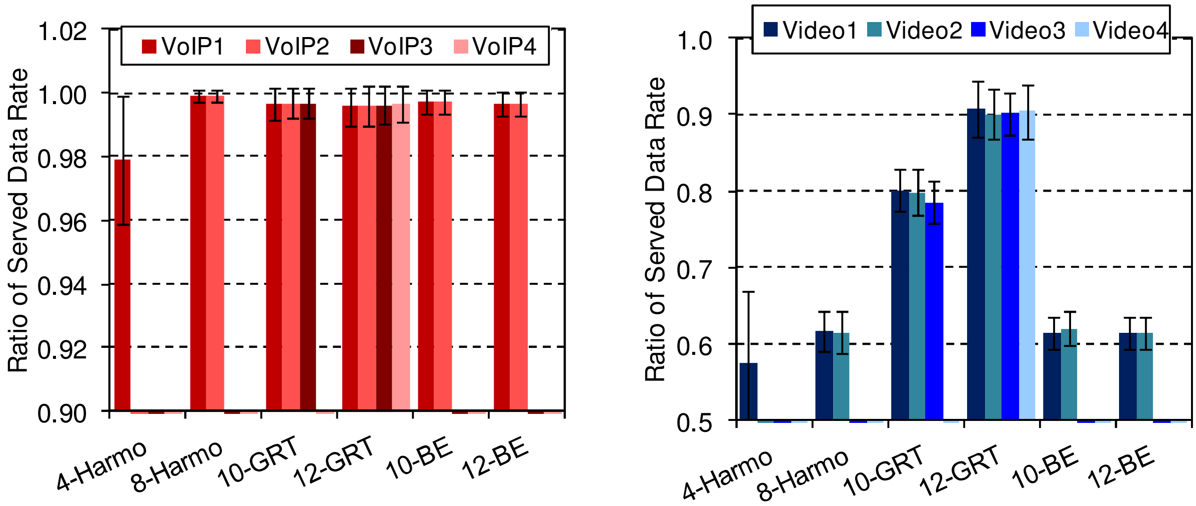
Concerning BE VBSs, Figure 7.53, for GRT based use cases $\overline{R_{serv}^{EU}}$ decreases considerably, since the contracted data rate is limited, only two VBSs per BE service, and the high number of end-users connected to the GRT VBSs forces the data rate reduction of BE end-users. The highest value of $\overline{R_{serv}^{EU}}$ is observed for BE based use cases, mainly because the total BE VBSs' contracted data rate is higher. Still, the distribution of end-users among VBSs allows achieving values around 50% of the typical service data rate, 10 Mbit/s and 1 Mbit/s for FS and Web services, respectively.




(a) FS VBS (b) Web VBS

Figure 7.53. *Average End-user data rate* of BE VBSs for VNet quantity use cases.

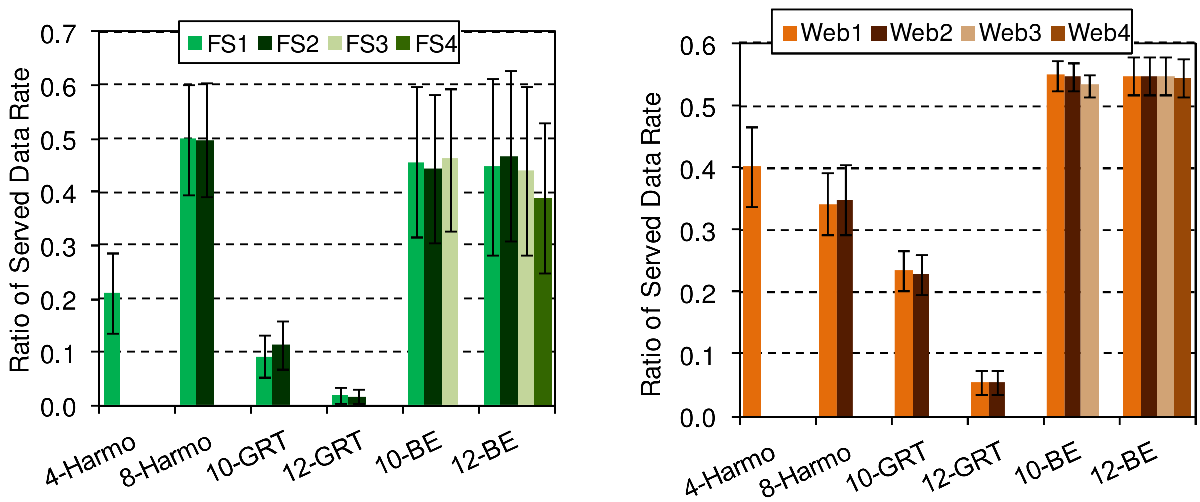
Regarding *Ratio of Served Data Rate,* $r_{serv}^{VBS}$, for GRT VBSs, Figure 7.54, it follows the amount of contracted data rate, since for GRT VBSs that amount is always ensured whenever the contract is realistic compared to the physical capacity of the cluster. In the case of VoIP, the contracted data rate is always above the data rate requested by end-users, exception made for *4-Harmo*, $r_{serv}^{VBS} \approx 1$. For VoIP, as the number of end-users is high according to the service profile used, $R_{req}^{VBS} > R_{min}^{VBS}$ for all use cases, it causes the increase of the *Ratio of Served Data Rate* with this amount. It is worth noting that this ratio is independent of the number of BE VBSs, as it can be verified for *8-Harmo*, *10-BE* and *12-BE* use cases, all with two Video VBSs, in which this ratio is maintained around 0.6.




(a) VoIP VBS (b) Video VBS

Figure 7.54. *Ratio of Served Data Rate* of GRT VBSs for VNet quantity use cases.

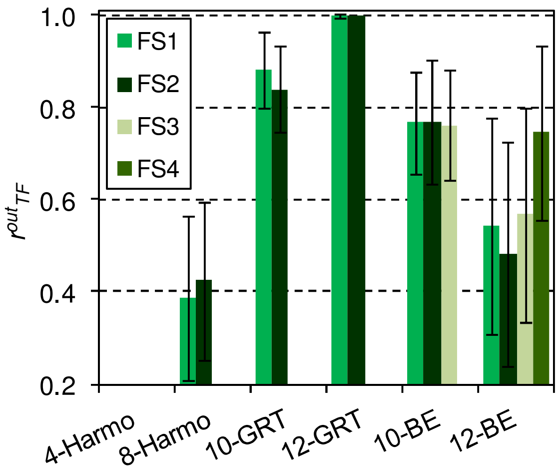
The *Ratio of Served Data Rate* for BE VBSs has a similar behaviour as the one referred for $\overline{R_{serv}^{EU}}$, Figure 7.55. In fact, since all end-users performing BE services are entering the network, they are accounted for both $R_{req}^{VBS}$ and $r_{serv}^{VBS}$, the same as the $\overline{R_{serv}^{EU}}$ divided by the typical data rate of the service, which is exclusive of each VBS.



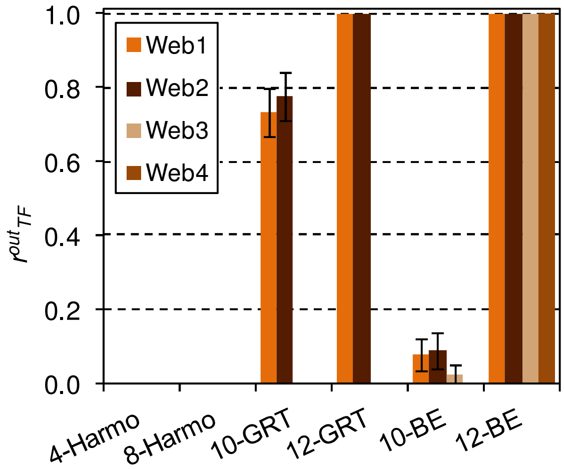

(a) FS VBS (b) Web VBS

Figure 7.55. *Ratio of Served Data Rate* for BE VBSs for VNet quantity use cases.

Results obtained for *Out of Contract* show that GRT VBSs are always running within contract, independently of the use case, but then BE VBSs are forced to run out of contract in some time frames. Figure 7.56 shows that BE VBSs are out of contract between 80% and 100% of time, $0.8 \le r_{TF}^{out} \le 1$, in GRT Based use cases. VBSs providing the FS service are more affected, $r_{TF}^{out} = 0.4$ for *8-Harmo* and $r_{TF}^{out} = 0.75$ for 10-BE, Figure 7.56(a), because the FS service is first reduced to maintain the minimum service data rate of GRT end-users in bad performance conditions. However, for *12-BE*, the out of contract percentage of time is reduced due to the low number of end-users requesting the service, only 3% of the total, which are distributed by the 4 VBSs, causing that most of the time the requested data rate is less than the contracted one.



(a) FS VBS

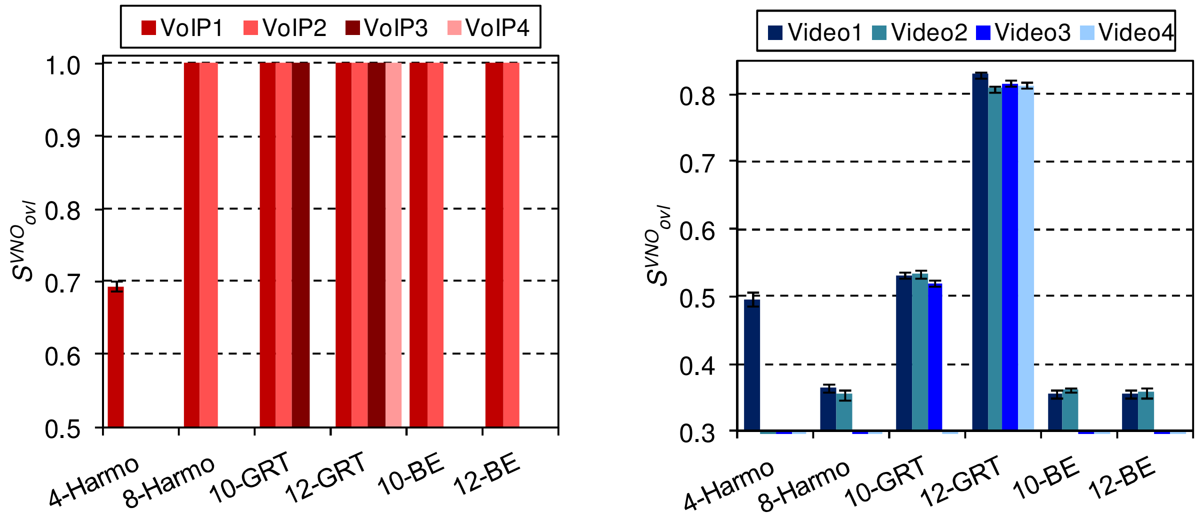


(b) Web VBS

Figure 7.56. *Out of Contract* for BE VBSs for VNet quantity use cases.

The VBSs providing Web services may run within contract most of the time for the *10-BE* use case, $0.02 \le r_{TF}^{out} \le 0.09$. However, when the number of BE VBSs increase to 4, *12-BE*, Web VBSs run always out of contract, because of the limited capacity of the physical cluster. Contracting more data rate for BE VBSs over the same physical capacity does not cause more BE VBSs' serving data rate, since the implemented strategy is to guarantee the minimum data rate of GRT VBSs, BE VBSs being just using the remaining capacity of the cluster.

The *Satisfaction Level* is only presented *on Extra Capacity Requested*, since when end-users are requesting service with the contracted capacity it is always 1, i.e., end-users are always entering the network. From Figure 7.57, the satisfaction level raises when the total VBSs contracted capacity increases, as expected. For VoIP VBSs, Figure 7.57(a), only for *4-Harmo* it is lower than 1, approximately 0.7, since for the other use cases VBSs are never running on extra capacity requested. For Video VBSs, Figure 7.57(b), the requested capacity is always above the total contracted one, and the *Satisfaction Level on Extra Capacity Requested* never reaches the maximum, $0.35 \le S_{ovl}^{VBS} \le 0.82$, increasing with the amount of contracted data rate, *10-GRT* and *12-GRT*, because the total requested data rate for all VBSs in the cluster is higher than the cluster capacity, and the allocation of radio resources is managed in order to follow the contracts established. However, if the other VBSs are running is low utilisation, the data rate requested by Video end-users would be served.




(a) VoIP VBS (b) Video VBS

Figure 7.57. *Satisfaction Level on Extra Capacity Requested* for VNet quantity use cases.

The *Average Delay on Service Request InP* is of the order of hundreds of milliseconds for Video VBSs, Figure 7.58, $\overline{\tau_{InP}^{VBS}} < 300$ ms. It increases with the number of VBSs due to the larger number of GRT end-users in the network, which cause a high utilisation of the radio resources in order to guarantee the service data rate of end-users in bad performance conditions. However, since the *12-GRT* use case depicts a GRT Based overbooking situation, it can be considered as a worst case or a limit situation for implementation.

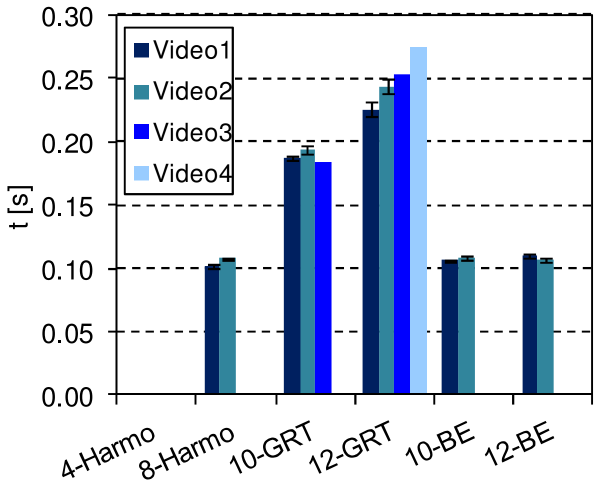



Figure 7.58. *Average Delay on Service Request InP*.

Regarding the *Average Delay on Service Request VNO*, it is of the order of ten of seconds for Video VBSs, except for *12-GRT*, Figure 7.59, in which the contracted data rate is near the requested one, thus, end-users are less delayed when they try to enter in the network. It can be also noted that delay increases to approximately 20 s when the total VBS contracted data rate is lower, i.e., on *8-Harmo*, *10-BE* and *12-BE* use cases. The *Average Delay on Service Request VNO* for VoIP VBSs is only different from zero for the *4-Harmo* use case, $\overline{t_{InP}^{VBS}} \approx 1.16$ s, since it is the only use case in which the VBS is operating on extra capacity requested, $\overline{R_{serv}^{VBS}} > R_{min}^{VBS}$.

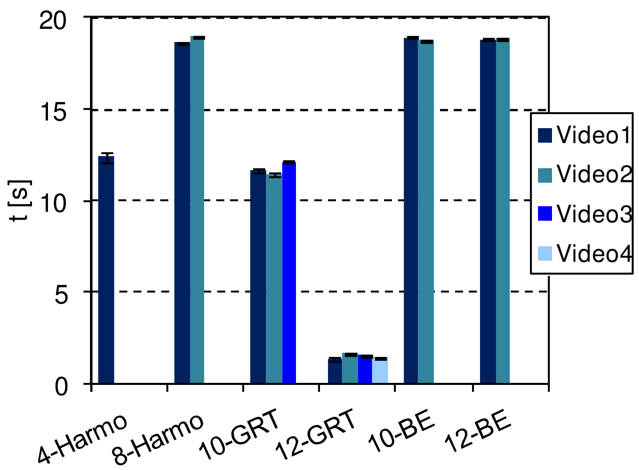
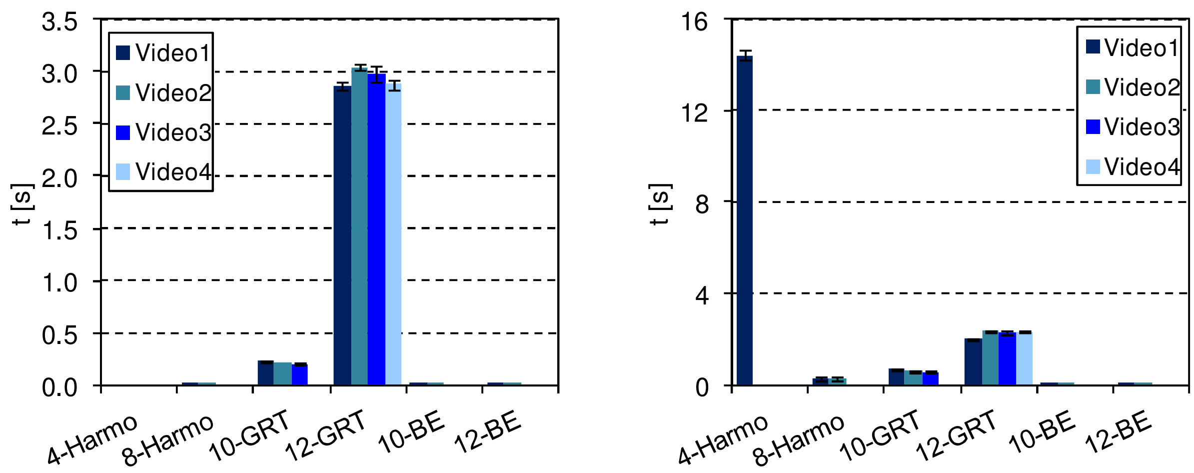



Figure 7.59. *Average Delay on Service Request VNO* for Video VBS.

Results for *Average VBS Time Service Delayed* in Video VBSs, Figure 7.60, show that the service can be strongly delayed, in the order of seconds, when the number of GRT end-users in the network is considerable, *4-Harmo* and *12-GRT* use cases.

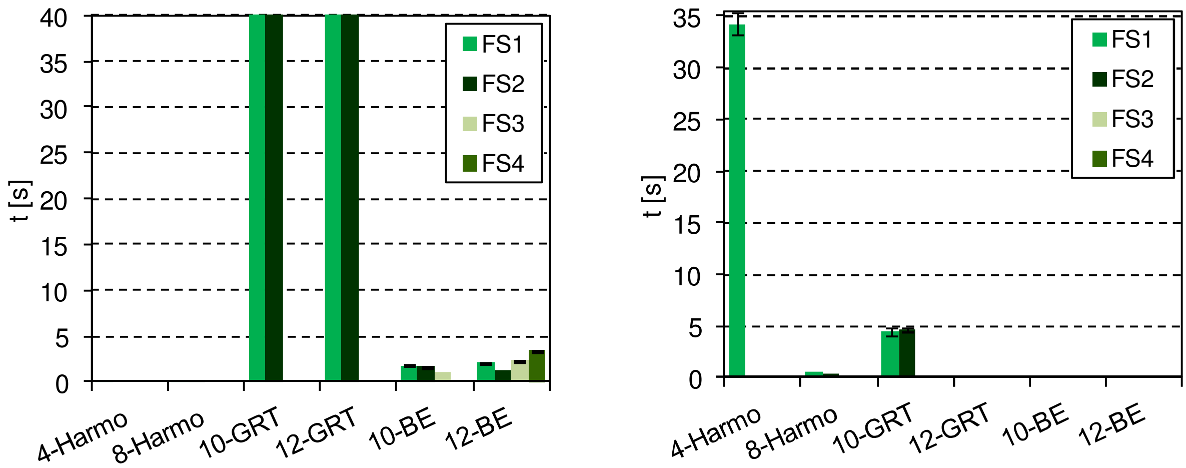



(a) within the contracted data rate (b) on extra capacity requested

Figure 7.60. Video VBSs *Average Time Service Delayed*.

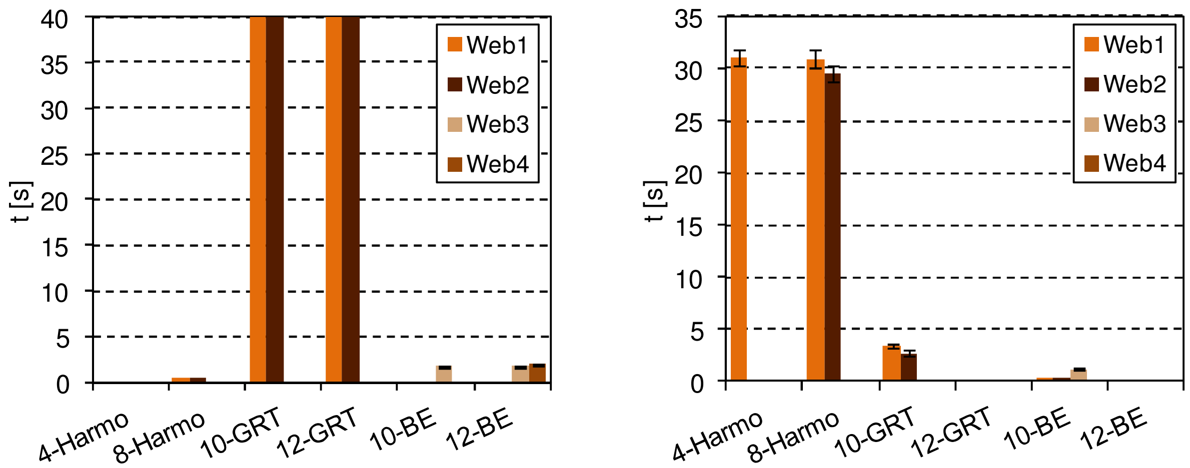
For *4-Harmo*, $\overline{\tau_{VNO}^{VBS}} \approx 14$ s, the problem is related to end-users connected on extra requested capacity, Figure 7.60(b), since no guarantees exist for them, though the service is not denied. *12-GRT* appears one more time as an extreme case of poor performance, $\overline{\tau_{int}^{VBS}} \approx 3$ s, due to the high GRT VBSs contracted data rate. In fact, in bad performance conditions the contracted data rate can be only satisfied at the expenses of delaying the service of GRT end-users, since the cluster is operating in its capacity limit. This again suggests that the strategy for reducing BE end-users should be applied considering service delay and not only the minimum service data rate.

The *Average Time Service Delayed* of BE VBSs denotes the strategy used to guarantee the contracted data rate by GRT VBSs. The following analysis considers delays both within contracted data rate and on extra requested capacity, since BE VBSs are running under or over $R_{ref}^{VBS}$ depending on the state of GRT VBSs. From Figures 7.61 and 7.62, it sees that end-users performing FS or Web services are delayed, $\overline{\tau_{int}^{VBS}} > 40$ s, Figures 7.61(a) and 7.62 (a), for *10-GRT* and *12-GRT* use cases.




(a) within the contracted data rate (b) on extra capacity requested

Figure 7.61 *Average VBS Time Service Delayed*. of FS *VBSs* for VNet quantity use cases.




(a) within the contracted data rate (b) on extra capacity requested

Figure 7.62. Web *VBSs Average Time Service Delayed*.

For *4-Harmo* and *8-Harmo*, $\overline{\tau_{VNO}^{VBS}} \approx 30$ s, Figures 7.61(b) and 7.62 (b), except for VBSs providing FS services in *8-Harmo*, due to the lower number of end-users performing this service. For *10-BE* and *12-BE, Average Time Service Delayed* is of the order of units of seconds on BE VBSs, the average delay not being as high as it is for GRT based use cases, because GRT VBSs are limited by their contracted data rate, allowing most of end-users in BE VBS to reach at least a residual data rate.

As a summary, one can say that changing the quantity of created VBSs as well as the contracted data rate in the cluster, GRT VBSs continue to achieve their minimum contracted data rate, though the *Average Cluster Serving Data Rate* can decrease if the number of GRT VBSs is higher than the number of BE VBSs. It is observed that the best RRU efficiency is achieved when the strategy for the overall capacity provision is to limit the capacity contracted by GRT VNets, overbooking the capacity contracted by BE VNets. The worst RRUs efficiency is observed when the number of GRT VBSs is the highest, the GRT overbooking situation being considered as the worst case or the limit situation for virtual wireless access implementation. Furthermore, contracting more data rate for BE VBSs over the same physical capacity does not cause the increase of their *Average VBS Serving Data Rate*, since the main target is to guarantee the minimum contracted data rate of GRT VBSs, BE VBSs being just using the remaining capacity of the cluster.

## 7.8 Service Profile

The assessment of the OnDemand-VRRA algorithm when the usage service profile, i.e., the percentage of end-users per service, changes is presented in this section. The main goal is to verify how the data rate requested by end-users performing services in each VBS, and specifically in each kind of service, affects the operation of the VNets. One takes the use cases presented in Table 6.3 as *Default*, *FSW* and *ViVo* related to service profile variations. The Reference scenario is considered. Table 7.7 summarises the use cases considered for this analysis.

Table 7.7. Service Profile related use case's parameters.

| **Parameter** | ***Default*** | ***FSW*** | ***ViVo*** |
|---|---|---|---|
| **Operators** | 2 VNOs - one provides GRT services and the other provides BE services | | |
| **VNets** | 1VNet GRT, 1VNet BE | | |
| | 2 VBSs deployed over the cluster | | |
| | VBS GRT provides VoIP and Video<br>BE VBS provides Web File Sharing | | |
| | Total capacity contracted by VBSs within cluster capacity | | |
| | Capacity contracted:<br>GRT VBS - $R_{min}$ = 1250 Mbit/s BE VBS - $R_{ref}$ = 1500 Mbit/s | | |
| **Service Profile** | VoIP(4%); Video(35%)<br>FS (3%); Web(58%) | VoIP(11%); Video(9%)<br>FS (23%); Web(57%) | VoIP(46%); Video(34%)<br>FS (6%); Web(14%) |
| **End-users** | From 1000 to 30000 end-users in the cluster | | |

The *Average VBS Serving Data Rate*, $\overline{R_{serv}^{VBS}}$, computed from (4.24), is depicted in Figure 7.63 and Figure 7.64, as a function of the cluster requested data rate, $R_{req}^{Cl}$.

For the use case *Default,* Figure 7.63, $\overline{R_{serv}^{VBS}}$ increases with the requested data rate, showing the trend to reach the contracted data rate of each VBS more or less linearly, because in both VBSs $R_{req}^{VBS}$ rise similarly relative to $R_{min}^{VBS}$ and $R_{ref}^{VBS}$ for GRT and BE VBSs, respectively.

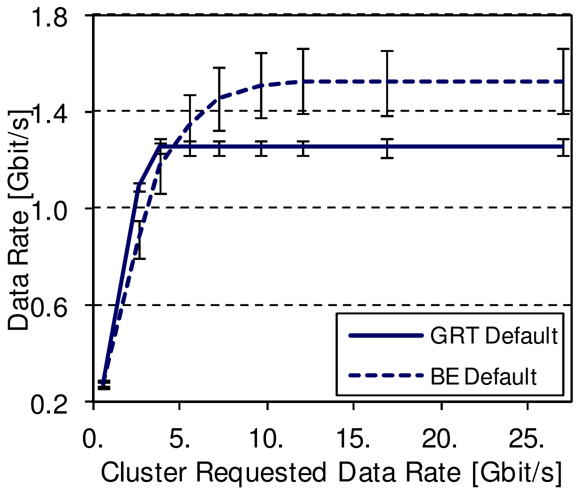
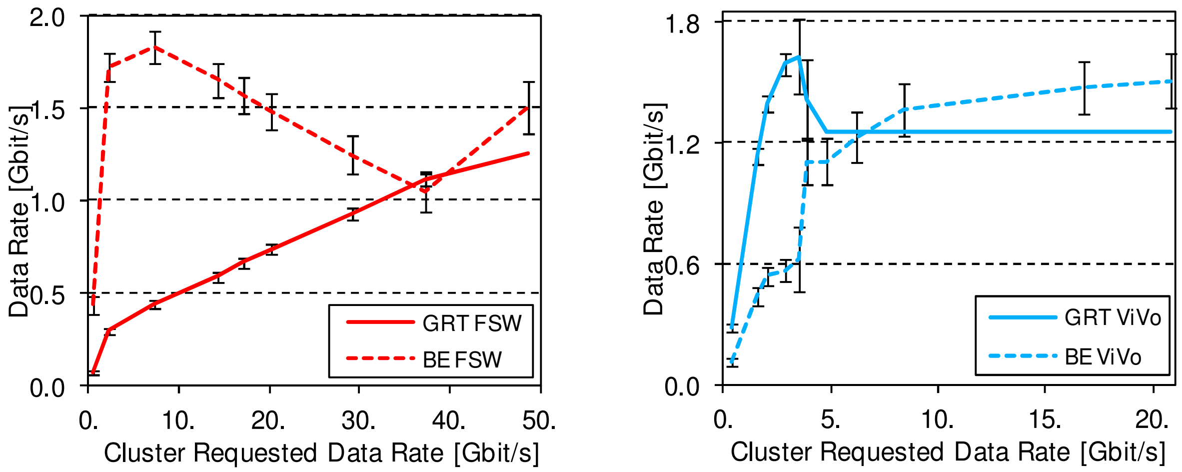



Figure 7.63. Average VBS *Serving Data Rate* for *Default* use case.

For *FSW* and *ViVo*, Figure 7.64, in which the number of end-users is asymmetrically distributed between GRT and BE VBSs, $\overline{R_{serv}^{VBS}}$ increases in the VBS with more end-users exceeding the contracted data rate, while in the other VBS one has $R_{req}^{VBS} < R_{min}^{VBS}$ or $R_{req}^{VBS} < R_{ref}^{VBS}$ according to the type of VBS. The maximum values of $\overline{R_{serv}^{VBS}}$ are bounded by the cluster total capacity, being approximately 1.6 Gbit/s in GRT VBS for $R_{req}^{Cl} = 2$ Gbit/s, Figure 7.64(a), and 1.8 Gbit/s in BE VBS for $R_{req}^{Cl} = 6.8$ Gbit/s, Figure 7.64(b). Such behaviour is similar in this initial phase whatever the type of VBS with the greatest requested data rate, confirming the allocation of the available RRUs in the cluster when the other VBS does not use them. After that, when $R_{req}^{Cl}$ increases, both VBSs decrease their serving data rate, tending to reach the contracted value. However, due to the VNet priority mechanism, the $\overline{R_{serv}^{VBS}}$ of BE VBS for *FSW*, decreases below $R_{ref}^{VBS}$ for $R_{req}^{Cl} = 36$ Gbit/s, $\overline{R_{serv}^{VBS}} \approx$ 0.7 $R_{ref}^{VBS}$) to allow the GRT VBS to reach its $R_{min}^{VBS}$.

In conclusion, while for both VBSs $R_{req}^{VBS} < R_{min}^{VBS}$ and $R_{req}^{VBS} < R_{ref}^{VBS}$, GRT end-users are served with priority ( $\overline{R_{serv}^{VBS}} \approx R_{req}^{VBS}$ ),the remaining cluster capacity being allocated to VBS BE end-users. On the other hand, if $\overline{R_{req}^{VBS}} > R_{min}^{VBS}$, the GRT VBS Serving Data Rate is limited by its $R_{min}^{VBS}$ until the BE VBS reach its $R_{ref}^{VBS}$, situation that is maintained when the number of end-users increases due to the limited cluster capacity. Finally, whenever $R_{req}^{VBS}$ for a VBS is below the contracted data rate, being above it for the other, the OnDemand algorithm manages to allocate all the remaining RRUs in the cluster to the VBS with more requested data rate, independently of the VBS type.




(a) FSW (b) ViVo

Figure 7.64. Average VBS *Serving Data Rate* for *FSW* and *ViVo* use cases.

The *Average Cluster Serving Data Rate* is presented in Figure 7.65 for the three uses cases considered. All use cases achieve slightly more than the sum of VBSs' contracted data rate, 2.75 Gbit/s, but it depends on how $R_{req}^{Cl}$ is distributed among VBSs, because the maximum *Average Cluster Serving Data Rate* is only achieved when for both VBSs the $R_{req}^{VBS}$ is larger than the contracted data rate.

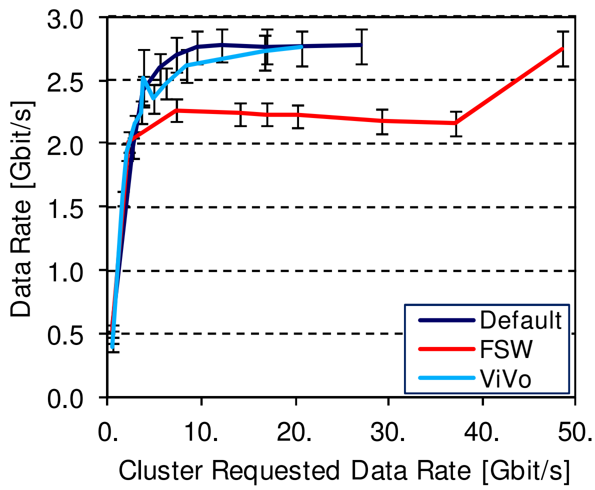



Figure 7.65. *Average Cluster Serving Data Rate* for *Default*, *FSW* and *ViVo* use cases.

Figure 7.65 shows that the maximum value $\overline{R_{serv}^{Cl}} = 2.77$ Gbit/s is achieved when $R_{req}^{Cl} = 9.6$ Gbit/s for *Default*, when $R_{req}^{Cl} = 16.7$ Gbit/s for *ViVo*, and only when $R_{req}^{Cl} = 48.5$ Gbit/s for *FSW*. These $R_{req}^{Cl}$ boundaries depend on the level of asymmetry between the percentage of end-users and the typical data rate of the provided services by each VBS. Only when in both VBSs the *VBS Requested Data*

*Rate* is larger than the contracted one, the maximum $\overline{R_{serv}^{Cl}}$ is achieved, because only in that situation GRT VBS is serving at least $R_{min}^{VBS}$, loosing the priority, and allowing the BE VBS to tend to $R_{ref}^{VBS}$.

Regarding both *Average Cluster Utilisation* and *Average Cluster Serving Data Rate*, for *Default* and *ViVo*, the increase of cluster utilisation, Figure 7.66, is according to the increase of $\overline{R_{serv}^{Cl}}$, Figure 7.65. Moreover, when the maximum $\overline{R_{serv}^{Cl}}$ is achieved, one gets $\overline{\eta_{Cl}} \approx 1$. Still, the *FSW* use case shows a high utilisation of RRUs much before the maximum $\overline{R_{serv}^{Cl}}$ is reached, because most of the RRUs are assigned to BE end-users, in large number for this use case, following the oscillations of their performance as there is no reduction process associated with BE end-users.

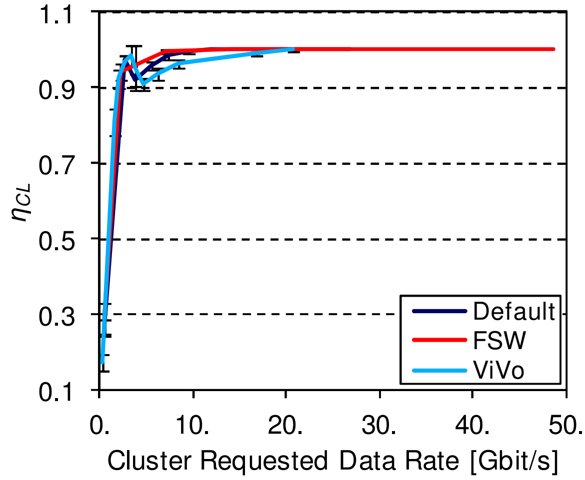



Figure 7.66. *Cluster Utilisation* for *Default*, *FSW* and *ViVo* use cases.

Results for *Average End-user Data Rate*, Figure 7.67, though with small oscillations in *ViVo*, show that end-users in GRT VBS maintain the average data rate even when the requested data rate in the cluster increases, Figure 7.67(a), however, in BE VBS it decreases considerably with the increase of $R_{req}^{Cl}$ whatever the use case, Figure 7.67(b). The explanation for this behaviour is that end-users in BE VBS are always connected, since BE services do not have a minimum service data rate, and GRT end-users are only connected if there are enough RRUs available to achieve the minimum data rate of the requested service. Still, GRT end-users served data rate is dynamically adapted for the service typical data rate according to the VNet priority scheme and data rate reduction strategy.

Results for *Out of contract*, $r_{TF}^{out}$, show that it is always zero for the GRT VBS, meaning that it is served at least with the minimum contracted data rate, whenever $R_{req}^{VBS} > R_{min}^{VBS}$. On the other hand, the BE VBS presents a peak in out of contract ($r_{TF}^{out}$ =1), Figure 7.68, corresponding to the situation in which the data rate requested for one VBS is above the minimum data rate contracted and the other is reaching this value. In this situation, the VNet priority scheme and reduction strategy enforce BE end-users to reduce their data rate in order to allow the GRT VBS to be served with its contracted data rate, occurring sooner for *Default*, $r_{TF}^{out}$=0.93 for $R_{req}^{CI} = 3.8\, Gbit/s$, and later for *FSW*, $r_{TF}^{out}$=0.99 for $R_{req}^{CI} = 37.1\, Gbit/s$. For *ViVo*, the out of contract starts increasing for $R_{req}^{CI} = 3.8\, Gbit/s$ ($r_{TF}^{out} = 0.79$), because the requested data rate in the BE VBS is close to the contracted minimum, but it is not enough to compensate for the data rate reduction imposed by the GRT VBS.

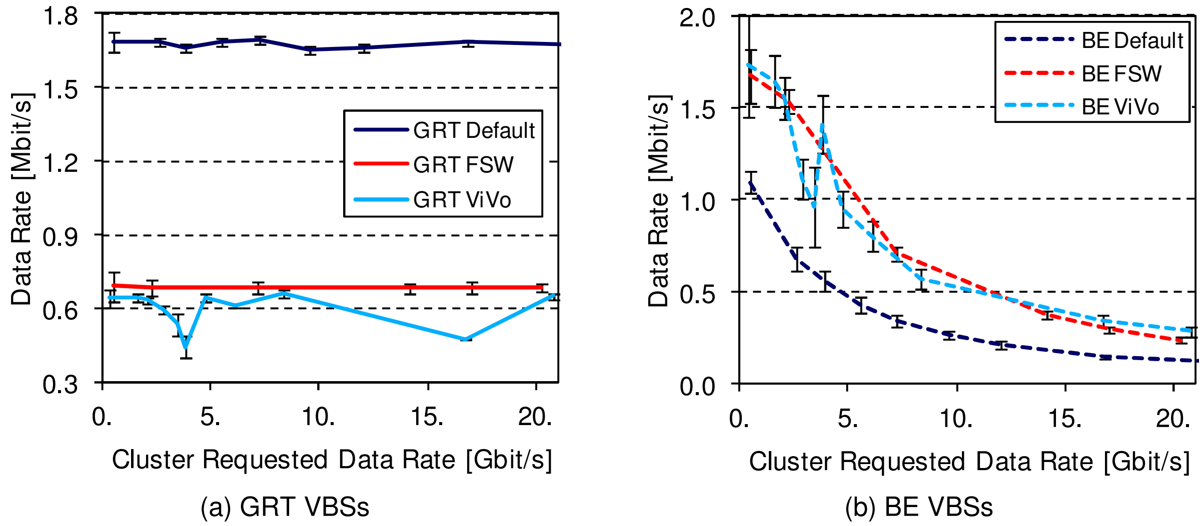



Figure 7.67. *Average End-user Data Rate* for different percentages of end-users.

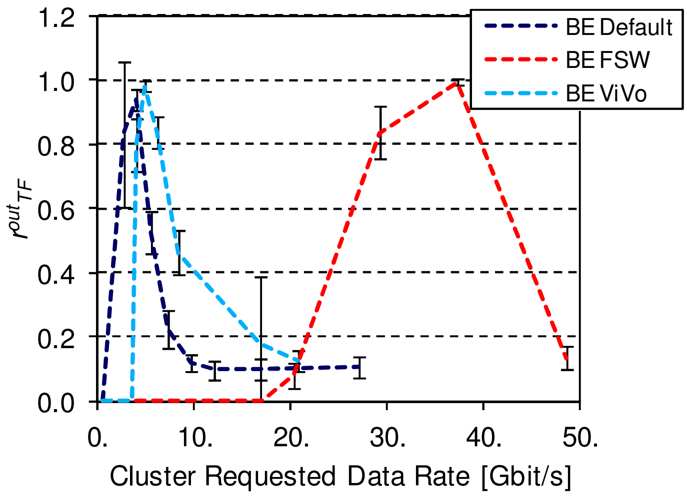



Figure 7.68. Out of Contract for *Default*, *FSW* and *ViVo* use cases.

When both VBSs are serving the contracted data rate, the network comes to a state in which the BE VBS out of contract is roughly maintained, $r_{TF}^{out}$≈0.1, which is due to the total VBSs' contracted data rate being within the average cluster capacity, which allows in most time frames to achieve the contracted data rate, since there are enough end-users in the coverage area of all BSs. This allows benefiting from all RRUs left by the imposed limit of the contracted data rate of GRT VBS.

Results for *Satisfaction Level* are only presented for GRT VBS on extra capacity requested, i.e., for situations in which $\overline{R_{serv}^{VBS}} > R_{min}^{VBS}$, Figure 7.69. In fact, due to several reasons, *Satisfaction Level on* *the infrastructure Provider* is always $S_{InP}^{VNO} = 1$: on the one hand, for BE VBS, end-users are all connected to the VNet since the minimum service data rate is considered as equal to zero; on the other hand, GRT VBS end-users who wish to connect do so until the VBS reaches $R_{min}^{VBS}$ due to VNet priority scheme and reduction strategy.

Figure 7.69 shows that, as expected, the *Satisfaction Level on extra Capacity Requested* only decreases when $\overline{R_{serv}^{VBS}} = R_{min}^{VBS}$. The use cases with more end-users in GRT VBSs, *ViVo* and *Default*, thus requesting a higher data rate, reach $R_{min}^{VBS}$ sooner, the *Satisfaction Level* being reduced also earlier. This reduction is imposed by the limited capacity of the cluster, and by the fact that GRT VBSs no longer have priority after reaching the contracted data rate, hence, GRT VBS end-users are only entering the network if the other VBSs are already operating within their contracts and there is some remaining cluster capacity.

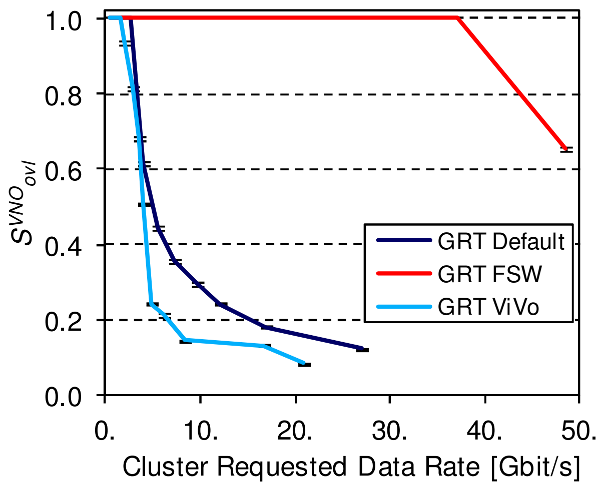
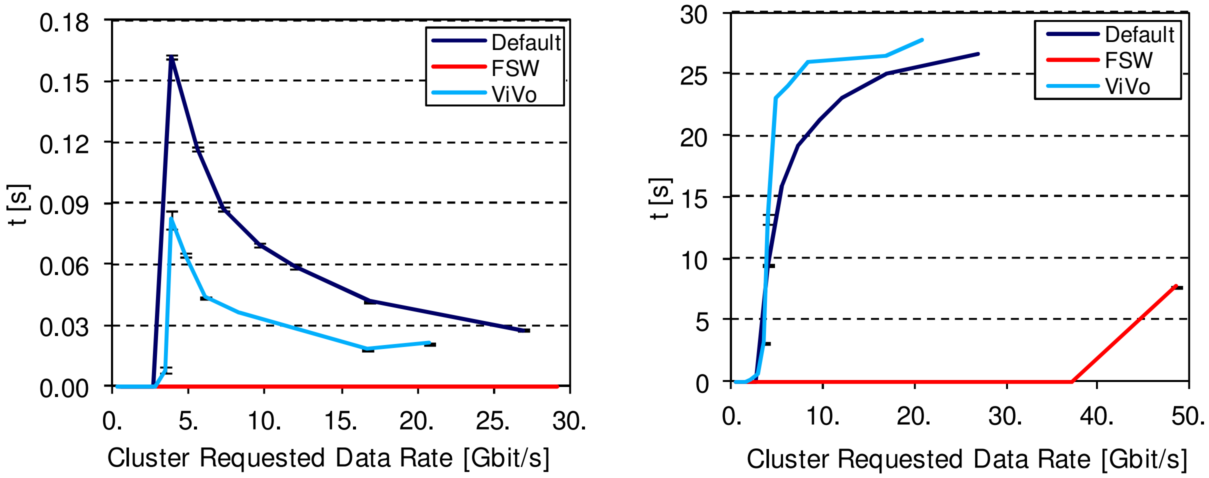



Figure 7.69. *Satisfaction Level on extra capacity requested* in GRT VNet for *Default*, *FSW* and *ViVo* use cases.

Concerning the delays end-users are subject to when trying to enter to the network, *Average Delay on Service Request*, and when performing services on the network, *Average VBS Time Service Delayed*, results show that for GRT VBS they are of the order of milliseconds when the VBS is operating within the contract, Figure 7.70(a) and Figure 7.71(a), and of the order of seconds, when the VBS is operating above the contracted data rate, Figure 7.70(b) and Figure 7.71(b). This denotes the strategy used: to guarantee the contracted data rate of GRT VBSs, but when the contract is reached, GRT end-users should compete on equal terms with other end-users.

It can be noted that $\overline{\tau_{InP}^{VBS}}$ presents a peak for the same amount of $R_{req}^{CI}$ of the peak BE VBS presents for *Out of Contract*, Figure 7.70(a), however, these peaks do not exceed 160 ms in the worst case, the *Default* use case. Still, it is for the same amount of $R_{req}^{CI}$ that $\overline{\tau_{VNO}^{VBS}}$ increases considerably, Figure 7.70(b), since the VBS is most of the time operating above the contracted data rate.

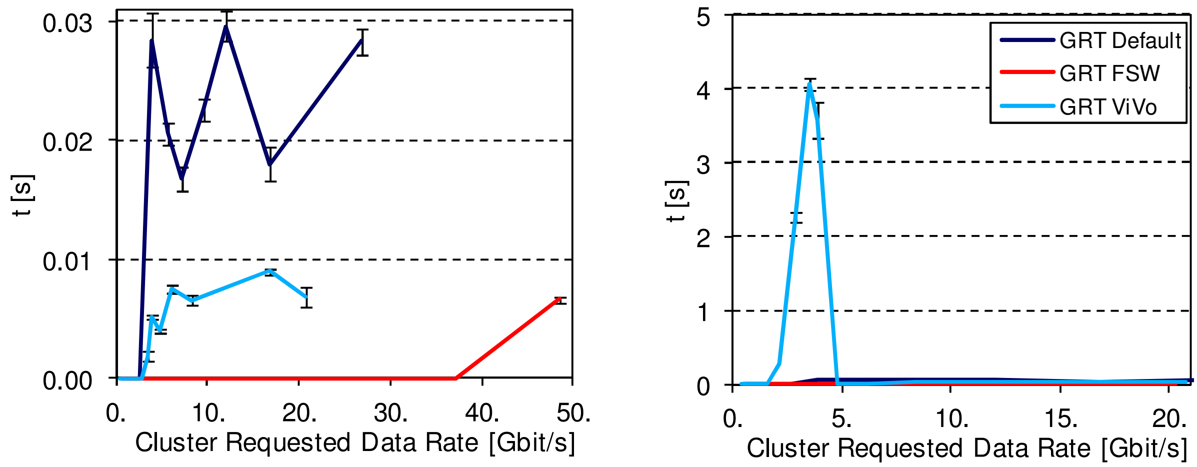



(a) within the contracted data rate ( $\overline{\tau_{InP}^{VBS}}$ ) (b) on extra capacity request ( $\overline{\tau_{VNO}^{VBS}}$ )

Figure 7.70. *Average Delay on Service Request* for *Default*, *FSW* and *ViVo* use cases.

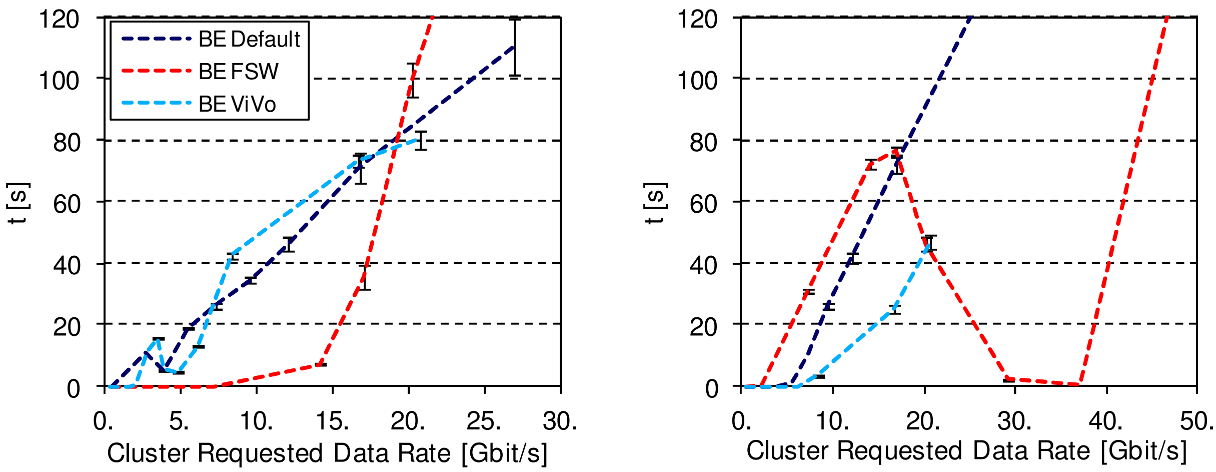
There is a similar behaviour for *Average VBS Time Service Delayed* in GRT VBS, Figure 7.71, with proper adjustment. When the VBS is operating within contract, the values are smaller than $\overline{\tau_{InP}^{VBS}}$, Figure 7.71(a), being of the order of tens of milliseconds. For the extra capacity operation of the VBS, delays are of the order of units of seconds, being only detected when the VBS is operating in this situation, *ViVo* use case for $1.6 < R_{req}^{Cl} < 4.8$ Gbit/s.

Regarding the BE VBS, delays are only accounted for *Average VBS Time Service Delayed*, as it is assumed that BE end-users are always admitted even if there are no available RRUs. For the VBS operating within contract, Figure 7.72(a) shows that only for low amounts of $R_{req}^{Cl}$, end-users are connected without any RRU assigned for $\overline{t_{int}^{VBS}} < 1$ s, when $R_{req}^{Cl} = 0.5$, 2, 7.2 Gbit/s for *Default*, *FSW* and *ViVo*, respectively. As soon as $R_{req}^{Cl}$ increases, the *Average VBS Time Service Delayed* takes values of the order of seconds, $\overline{t_{int}^{VBS}} = 30$ s for $R_{req}^{Cl} = 8.3$ Gbit/s in *ViVo,* $\overline{t_{int}^{VBS}} = 42$ s for $R_{req}^{Cl} = 9.6$ Gbit/s in *Default*, and $\overline{t_{int}^{VBS}} = 36$ s for $R_{req}^{Cl} = 17$ Gbit/s in *FSW*. For *ViVo* and *Default* the amount of $R_{req}^{Cl}$ corresponds to the situation in which the $R_{req}^{VBS} > R_{min}^{VBS}$, and the number of end-users in BE VBS increases in the way that a large quantity of them are connected but with delayed service. For *FSW*, the number of BE end-users is too high and the GRT VBS $R_{req}^{VBS}$ is forcing BE end-users to be delayed. If the VBS is operating on extra capacity, the service is delayed considerably, Figure 7.72(b), reaching $\overline{t_{VNO}^{VBS}} >$ 120 s when the number of BE end-users is considerably high.




(a) within the contracted data rate (b) on extra capacity request

Figure 7.71. *Average VBS Time Service Delayed* in GRT VBS, for Default, *FSW* and *ViVo* use cases.




(a) within the contracted data rate (b) on extra capacity request

Figure 7.72. BE *Average VBS Time Service Delayed* for *Default*, *FSW* and *ViVo* use cases.

Summarising, it can be said that by changing the service profile, the OnDemand-VRRA allows achieving isolation among the virtual resources, since the requested data rate of one VBS does not prevent the other to achieve the contracted data rate, if they are GRT. It is verified that even when for 80% of end-users are requesting service in the BE VBS, the GRT VBS reaches the contracted data rate as soon as the requested data rate is greater or equal to it.

# Chapter 8

# Conclusions

This chapter summarises the major results, drawing conclusions and pointing out aspects to be developed in future work.

## 8.1 Summary

This thesis is organised into 8 chapters. Chapter 1 gives a historical perspective of the two main topics in the work developed in the thesis, wireless and mobile networks, and network virtualisation. The motivation and main objectives, the novel aspects and the research strategy are also presented.

Chapter 2 presents the basic concepts and the topics that are the basis of the work done in the thesis, specifically, the strategies for RRM in current RATs, including the CoRRM. An introduction to VNets, the state of the art in wireless virtualisation and other approaches to network sharing are presented in Chapter 3. The overview of a novel framework for the provision and management of connectivity services, OConS, is made in this chapter.

In Chapter 4, the proposals for novel models and algorithms to manage the radio resources in virtualised environments are presented, as well as the strategies used for RRM in virtualised environments and the characterisation of the proposed Cooperative VNet RRM (CVRRM) main functions. Furthermore, the network architecture, the main assumptions and inputs, the analytical model, the strategies and algorithms for virtual radio resources allocation algorithms, the initial VNet selection and the VNet handover support and the metrics for evaluation are also presented. The modelling of the OnDemand-VRRA algorithm to be included in the OConS Architecture is also shown.

The implementation of the proposed models in a simulator is described in Chapter 5, the main assumptions taken for simulator development being presented, as well as its main features and blocks and details of implementation. Still in this chapter, the simulator assessment is presented, referring the analysis of the transitory interval and influence of the number of simulations in output values.

In Chapter 6, the scenarios for evaluation are identified and the theoretical results are compared with the ones obtained from simulations.

The analysis of simulations results for the evaluation of VRRA algorithms is done in Chapter 7. Initially, the performance of the Adaptive-VRRA algorithm is evaluated, based on the strategy for instantiation of several virtual resources in the same physical cluster. Then, the comparison of our virtualisation approach with the actual network deployment, and between the fixed allocations of radio resources is made, considering the OnDemand-VRRA algorithm. After, the exploitation of several strategies for the provision of virtual capacity, the variation on the number of virtual resources of each type, the impact of changing the quantity of virtual resources created in one physical cluster, and the variations in the service profile are analysed.

Finally, the current chapter presents a summary of the thesis, the main results and the novelty of the work. Some directions for future work are also provided.

## 8.2 Major Results

The concept of network virtualisation has been considered as the basis to address the problem of sharing the wireless infrastructure for provision of capacity to VNets. Following this approach, it is intended that the basic principles of network virtualisation, such as the isolation between virtual resources and the possibility to deploy different protocols to take the diverse service requirements into account can be applied.

The allocation of transmission resources is a challenging problem in virtualised environments, where they are shared among the different virtual resources, and there is the need to fulfil contracted capacity requirements. In wireless networks, the problem is even more challenging, due to the inherently limited resources. In fact, the available radio resources are scarce with variable performance, and there is lack of spare spectrum.

In this thesis, a reference network architecture for the virtualisation of the wireless access based on the generic network virtualisation environment is proposed. Both physical and virtual perspectives are considered, and the main stakeholders are taken into account. In terms of physical infrastructure, one considers a set of different RATs, which are abstracted by the specific RRUs of each one. This allows the management of radio resources by the coordination of a pool of RRUs, each having particular capabilities. Concerning the virtual resources, they can be defined differently by setting their type, GRT or BE, allowing to differentiate end-users handling, according to the VNet they belong to.

A new tier of RRM is proposed for inter-VNet RRM, designated by Cooperative VNet RRM, managing how radio resources are allocated to the several VNets in order to satisfy the contracted VNets' capacity. This new level of management is proposed to interact with CoRRM and RRM, which is considered to be an intra-VNet RRM, thus, under the responsibility of VNOs. Moreover, the generalisation of the inter-VNet RRM as a CooRRM problem with an additional level of abstraction, the virtual RRM level, allows following an approach of integration of the several levels of RRM, which needs to be adapted, but that actively participates in the process to achieve the main target of provision of the contracted level of service for all the VNOs operating on the common infrastructure. Naturally, the added virtual RRM level assumes the coordination role of all the underlying ones, as it is aware of VNets requirements and has the responsibility to satisfy them. Still the specific algorithms to implement the needed functionality at underlying RRM levels can evolve without overthrowing the outlined approach. The functionalities proposed for the initial VNet Selection and VNet Handover Support are essential to provide CVRRM with the set of functionalities assigned to CoRRM, thus, allowing it to be considered as a transposition of CoRRM to the virtualisation environment, though they are not further implemented. The VRRA function, being considered indispensable for the virtualisation of the wireless access, is described in more detail, two different algorithms being proposed.

The two novel VRRA algorithms proposed, according to the type of guarantees of VBSs, the amount of contracted capacity and VBSs' utilisation, take the variability of the wireless medium into account, and continuously influence RRM mechanisms, namely admission control and MAC scheduling, to be aware of the VBSs' state relative to their service level agreement. Instead of looking at the wireless virtualisation from the perspective of the instantiation of virtual machines into wireless nodes, our view is the virtualisation of the wireless access to provide a contracted capacity to the VNet, in order to serve its end-users. Our approach is then agnostic to the point where the virtual node instantiation takes place, being possible to have the virtual nodes in each physical wireless node, or somewhere in the cloud requesting virtual access over a given geographic area covered by a set of wireless nodes. It is worthwhile noting that this capacity can be modified on demand, without manually changing the configuration of the network.

Besides the innovative aspects already referred, major results related to specific achievements obtained throughout this work are summarised in the following paragraphs.

Comparing the wireless access virtualisation, supported by the proposed OnDemand-VRRA, with the standard approach, in which there are more than one network operator, each owning part of the physical infrastructure, it can be concluded that the former allows achieving a better performance from a cluster of BSs of several radio access technologies, enabling the provision of contracted capacity for GRT VNets. It is demonstrated by simulation that in virtual scenarios the Cluster Serving Data Rate may increase by approximately 46% and the utilisation by 13%. On the other hand, for Virtual scenarios the serving data rate of GRT VBSs is always greater than the minimum contracted being constrained by the defined BE VBSs reference data rate, which tends to be followed. The values achieved for the serving data rate are the lowest for BE services for the standard approach, but they can be larger for GRT services in the virtual approach, denoting, on the one hand, the limitation arising from the split of the total cluster capacity by two operators, and on the other hand, the consequences of an uncoordinated allocation of radio resources when end-users are handled independently.

Furthermore, the comparison between the OnDemand-VRRA algorithm and the Fixed allocation of RRUs to VBSs shows that, when some of the VBSs are being requested to provide service with less than the contracted data rate, *OnDemand* performs better than *Fixed*, approximately 14% higher, as it can allocate all the available RRUs to any VBS, since they are operating according to the established contracts. However, if the requested data rate is higher than the contracted one in all VBSs, the cluster performance is approximately 14% higher in *Fixed*, the main difference being experienced in BE VBSs, as *OnDemand* privileges GRT end-users, decreasing the data rate of BE end-users in order to guarantee the GRT VBS contracted data rate.

The proposed algorithm has been analysed for different strategies for capacity provision, several usage profiles, diverse combinations of VBSs from different types, and several quantities of VBSs deployed on the physical cluster.

It is concluded that a limit for the percentage of GRT VBSs' contracted data rate should be defined as a function of the average capacity of the cluster and of the BE VBSs' contracted capacity. From the comparison between the virtual access approach and the standard one, it can be said that the setting of minimum and reference values for VNet's contracted data rate allows end-users to have a better network experience, considering the data rate as the main parameter to evaluate it.

Concerning the different strategies for capacity provision, i.e., when the amount of contracted capacity by VBSs is over, on average, or under the physical capacity, it is concluded that a limit for the contracted data rate by GRT VBSs should be established in order to allow an efficient use of RRUs among all the VBSs deployed within the cluster. It is verified that the Cluster Serving Data Rate may increase by approximately 20% if the amount of contracted capacity by BE VBS is 85% of the average capacity of the cluster, compared to the use case where the contracted capacity by the GRT VBS is the one with 85% of the cluster average data rate.

When the service profile is changing, OnDemand-VRRA allows achieving isolation among the virtual resources, since the requested data rate of a VBS does not prevent the other to achieve the contracted data rate, if they are GRT. It is verified that even when 80% of end-users are requesting service in the BE VBS, the GRT VBS reaches the contracted data rate as soon as the requested data rate is greater or equal to it.

When changing the percentage of VBSs of each type instantiated in a cluster of BSs, only minor differences are perceived in the values obtained for the defined metrics, namely, VBS and Cluster Average Serving Data Rate, when the total number of VBSs is fixed and the relation between GRT and BE data rate contracted is maintained. The total *Average VBS Serving Data Rate* has a maximum deviation among the defined use cases around 0.2% for GRT VBS and approximately 1.8% for BE VBSs. According to this, also small variations are verified for the Average Cluster Serving Data Rate, increasing around 0.7% when the number of GRT VBSs increases.

By varying the quantity of created VBSs, as well as the contracted data rate in the cluster (although GRT VBSs are maintained within contract), the average cluster data rate can decrease if the number of GRT VBSs (hence, the contracted capacity) is higher than the number of BE VBSs. It is observed that the best RRU efficiency is achieved when the strategy for the overall capacity provision is to limit the capacity contracted by GRT VNets, overbooking the capacity contracted by BE VNets. The worst RRUs efficiency is observed when the number of GRT VBSs is the highest, which depicts the GRT overbooking situation as the worst case or the bound for virtual wireless access implementation. Furthermore, contracting more data rate for BE VBSs over the same physical capacity does not increase BE VBSs' serving data rate, since the implemented strategy is to guarantee the minimum data rate of GRT VBSs, BE VBSs being just using the remaining capacity of the cluster.

A brief reference to the Adaptive-VRRA algorithm, which was also implemented in the simulator, is done in what follows. It is concluded that it is not performing good enough and has limitations in terms of flexibility on the allocation of radio resources. However, some initial results were obtained in order to evaluate several strategies of instantiating more than one VBS of the same VNet in the cluster and on the allocation of RRU per BS to the VBS.

When the VRRA algorithm is based on the pre-allocation of an amount of RRUs to the VBS, Adaptive-VRRA algorithm, it is verified that the satisfaction level of the VNO can be improved by 40%. However, it must be stressed that although this VRRA algorithm may give some support to GRT VNets, the compensation can be insufficient when the services provided by VNets are more data rate demanding, since more RRUs are simultaneously affected by end-users in bad performance conditions.

One concludes that by sharing the already deployed heterogeneous wireless capacity through wireless access virtualisation, InPs can offer GRT and BE virtual capacity resources. Introducing VRRA algorithms, as OnDemand-VRRA, allows supporting the minimum bandwidth requirement for virtual access in a wireless cluster, composed of several physical BSs from different RATs, providing service over a given coverage area.

## 8.3 Novelty

This thesis claims novelty in a new approach for capacity sharing in wireless and mobile networks by the virtualisation of the wireless access, allowing extending the VNet concept to the wireless access, then giving the possibility to deploy different logical network functionality on top of it. Our proposal is to virtualise the wireless access by managing, in a common way, the radio resources available from heterogeneous wireless systems in order to provide capacity to the several virtual resources, the VBSs, created over it. A broader perspective of virtual resources, as an aggregated connectivity resource abstracted from a pool of RRUs of different RATs is adopted, allowing benefiting from Cooperative RRM strategies, overcoming the limited bandwidth availability of wireless technologies. Instead of looking at the wireless virtualisation from the perspective of the instantiation of virtual machines in the wireless nodes, our view is the virtualisation of the wireless access to provide a contracted capacity to the VNet, in order to serve its end-users. Our approach is then agnostic to the point where the virtual node instantiation takes place, being possible to have the virtual nodes in each physical wireless node, or somewhere in the cloud requesting virtual access over a given geographic area covered by a set of wireless nodes.

Additionally, the introduction of differentiation among the virtual resources is also claimed as a novelty. By handling the VBSs differently, according to their type of requirements, it supports the deployment of VNets with minimum guaranteed capacity, GRT VNets, and with a reference capacity to be provided whenever possible, BE VNets. All available RRUs in the cluster may be allocated to any VBS if they have been requested, as soon as all the other VBSs in the cluster have their contracted capacity being satisfied or a low demand.

Furthermore, the integrated approach to RRM in VNet environments, considering intra- and inter-VNet RRM levels, is also an innovative contribution. It intends to transport the main functions of CRRM to manage the interaction among VNets, maintaining the actual RRM and CRRM functionality. Hence, from the initial access selection, the vertical handover and the scheduling among different RATs of CRRM functionality, new functions have been proposed for CVRRM with a similar purpose, but one level up, at the virtualisation level. Although algorithms have been proposed for all CVRRM functions, namely, Initial VNet Selection, VRRA and VNet Handover Support, only for VRRA the algorithms have been implemented for simulation purposes. Two algorithms for VRRA have been proposed, the Adaptive-VRRA, in which radio resources are pre-allocated according to the contracted capacity, being adapted according to wireless medium conditions and network load, and the OnDemand-VRRA, which is more flexible, in which the allocation of radio resources is done when capacity is requested.

Finally, to highlight the configurability of the proposed algorithms, an Open Connectivity Service (OConS) architecture is proposed, flexible and modular in the description of connectivity resources and mechanisms, based on the identification of functional entities and their interfaces. It enables the orchestration of both legacy and enhanced connectivity mechanisms, which can be dynamically adapted and orchestrated into OConS Services offered to the network. Within this framework, it is worthwhile noting that this contracted capacity can be modified on demand, without manually changing the configuration of the network.

## 8.4 Future work

The work presented in this thesis can be followed on, by exploring several other topics that can be investigated in the future. Examples of these topics are proposed below.

One possibility it the extension of the set of possible requirements for VBSs to include not only capacity but also, e.g., delay and reliability. The definition of several CFs related to the type of service/requirements of the VBSs should be analysed, possibly allowing for the optimisation of the handling of end-users within the VBSs. Hence, VNOs may define the strategies to compare and classify the most relevant nodes in the radio network for their end-users according to the defined criterion.

The scheduling mechanisms and reduction strategies applied to the end-users of each VBS, although can be implemented at RRM level by the InP, must be enforced by the VNOs when the virtual resource is contracted. This means that InPs should provide a set of parameterised scheduling mechanisms that should be applied according to the VBS the end-user are connected. This allows VNOs to independently determine how their clients are handled, providing programmability to the network. The performance adaptation of GRT end-users should also consider service delay, depending on the specific service and the parameterisation made by the VNO.

Developing additional KPIs to reflect the need of resources for virtualisation, e.g., Central Processing Unit (CPU) and processing capacity available is also a need, as well to evaluate more strategies for resource comparison through CF.

The evaluation performed in this work takes implicitly the mobility of end-users into account by changing their SINR. However, specific aspects related to mobility, as handover, should be further analysed by considering the main scenarios defined for mobility. It can bring an additional degree of freedom, allowing to better explore the physical capacity utilisation according to statistical multiplexing.

The interaction with VNOs should be explored, since several parameters, such as satisfaction level on extra capacity requested and VBS utilisation, can be used by the VNO to adapt the contract capacity to the demand of their end-users.

One should consider historical patterns of end-users' service per VBS and estimation of demand provided by VNOs for a more accurate allocation of radio resources.

Future work should also explore new business models for delivery of new services for physical resources optimisation. For example, defining a new type of VNet, for tolerant delayed services, that are only served when the physical network is in a low load state after all the other VNets have their contracts satisfied. This is the case of utility operators that can receive the information of remote sensors or sell cheap capacity for given periods in which low traffic hours.

# Appendix A

# Data Rate Adaptation

A summary of the methods for data rate adaptation performed for the several radio access technologies taken in the thesis is presented in this appendix. The modulation and coding rates as well as the SINR and the data rate for the mobile and wireless systems considered as representative of the radio access technologies are also presented.

Tables A.1. to A.3. present a summary of the methods used for data rate adaptation and the peak data rate of TDMA, CDMA and OFDM based systems.

Table A.1. Data rate adaptation for TDMA based systems (based on [NaBK00]).

| System or Standard | Method of rate adaptation | Channel quality feedback | Peak data rate [kbit/s] |
|---|---|---|---|
| **GSM/GPRS** | Time-slot aggregation, Adaptive coding | Measurement reports in Automatic Repeat request (ARQ) Status message:<br>• Signal and interference;<br>• BER;<br>• Signal variance. | 160.0 |
| **TDMA 136+** | Time-slot aggregation, Adaptive modulation, incremental redundancy | Channel quality feedback (CQF):<br>• In UL - ARQ Status Message;<br>• In DL - packet channel feedback. | 44.4 |
| **GSM/EGPRS** | Time-slot aggregation Adaptive coding, Adaptive modulation, incremental redundancy | Measurement reports in ARQ Status message:<br>• Signal and interference;<br>• BER. | 473.6 |

Table A.2. Data rate adaptation for CDMA based systems (based on [NaBK00]).

| System or Standard | Method of rate adaptation | Channel quality feedback | Peak data rate [kbit/s] |
|---|---|---|---|
| **cdmaOne** | M supplemental code channels each at 8 or 14 kbit/s | SCRM, PSMM | 64 |
| **CDMA2000** | Variable-rate supplemental code channel - variable spreading and coding | SCRM, PSMM, power control bits (800Hz) | 2 048 |
| **UMTS** | Variable rate traffic channel - variable spreading and coding | Measurement report:<br>• Pilot strengths;<br>• SINR;<br>• BER, BLER. | 2 048 |

The modulation and coding schemes as a function of SINR and the achieved data rates are presented for the several mobile and wireless systems representative of each the RAT under study, i.e., GSM/EGPRS, UMTS/HSPA, Wi-Fi and LTE.

The data rates per RRU achieved for a given MCS *n*, $R_{MCS_n}^{RAT_r}$, were based on the BS capacity of the wireless network system, i.e., the number of RRUs and the data rate reached for the MCS:

$$R_{MCS_n}^{RAT_r} = \frac{R_{MCSn}^{BS}}{N_{RRU}^{RAT_r}} \qquad \text{(A.1)}$$

where:

- $R_{MCSn}^{BS}$ - BS data rate achieved for MCS *n*;
- $N_{RRU}^{RAT_r}$ - total number of RRUs per BS of RAT *r*.

The data rates are for DL, which is the link considered in this work.

Table A.3. Data rate adaptation for OFDM based systems.

| System or Standard | Method of rate adaptation | Channel quality feedback | Peak data rate [Mbit/s] |
|---|---|---|---|
| **Wi-Fi** [WiFi07] | Adaptive coding, Adaptive modulation | Not defined in standard. Probe packets, consecutive successes/losses, SNR, long-term statistics | 600 |
| **WiMax** [WiMA09] | | Channel Quality Indicator (CQI) | 300 |
| **LTE** [Agil09] | | | 326 |

GSM/EGPRS is the TDMA technology used in the scenarios defined in Section 6.1. The maximum data rate for GSM/EGPRS is 473.6 kbit/s per carrier. Knowing that for each carrier 8 time-slots are available, the maximum data rate per time-slot is 59.2 kbit/s. One TDMA BS is then one GSM/EGPRS BS with 3 carriers and a total maximum capacity of:

- Radio Unit (Time-slot) -> 8 × 3 = 24 RRUs
- Data Rate -> 3 × 473.6 kbit/s = 1.42 Mbit/s

In GSM/EGPRS, there are 9 different coding schemes defined, 5 coding schemes for 8PSK and four coding schemes for GSMK, Table A.4.

Table A.4. MCS, Modulation and Data Rate per Timeslot for GSM/EGPRS (extracted from [3GPP09b]).

| Modulation and Coding Scheme | Modulation | Data Rate per Timeslot [kbit/s] |
|---|---|---|
| MCS-1 | GMSK | 8.8 |
| MCS-2 | | 11.2 |
| MCS-3 | | 14.8 |
| MCS-4 | | 17.6 |
| MCS-5 | 8PSK | 22.4 |
| MCS-6 | | 29.6 |
| MCS-7 | | 44.8 |
| MCS-8 | | 54.4 |
| MCS-9 | | 59.2 |

Table A.5 is used to compute the data rate achieved by an end-user of a GSM/EGPRS system. The end-user data rate is the sum of the assigned RRUs' data rate, which are derived from the applied MCS determined by end-user's SINR value.

Table A.5. MCS, SINR and Data Rate used for GSM/EGPRS (based on [LDCQ01] and [3GPP09b]).

| MCS | SINR [dB] | Data rate per time-slot [kbit/s] |
|---|---|---|
| 1 | $0 < \gamma \leq 4$ | 0 |
| 2 | $4 < \gamma \leq 10$ | 9 |
| 3 | $10 < \gamma \leq 11$ | 15 |
| 4 | $11 < \gamma \leq 13$ | 25 |
| 5 | $13 < \gamma \leq 21$ | 35 |
| 6 | $21 \leq \gamma$ | 59 |

UMTS/HSPA evolution, 3GPP release 7 or HSPA+ (without Multiple Input Multiple Output (MIMO)), is the CDMA system considered. High Speed Downlink Packet Access (HSDPA) utilises advanced link adaptation and adaptive modulation and coding to ensure all users enjoy the highest possible data rate. This upgrade technology adapts the modulation scheme and coding to the quality of the appropriate radio link. While the spreading factor is fixed, the coding rate can vary between 1/4 and 3/4, and the HSDPA specification supports the use of five, 10 or 15 multicodes. This more robust coding, fast Hybrid ARQ (HARQ), and multi-code operation eliminates the need for variable spreading factor. This approach also allows users with good signal quality (higher coding rate) typically close to the BS, and those at the more distant edge of the cell (lower coding rate) to each receive an optimum available data rate [HoTo04].

The maximum data rate for UMTS/HSPA+ is 21.6 Mbit/s per carrier, using 64QAM and 15 codes. Knowing that for each carrier 15 codes are available, the maximum data rate per code is 1.44 Mbit/s. One CDMA BS in the context of this work is one HSPA+ BS with 3 carriers and a total maximum capacity of:

- Radio Unit (Code) -> 15 × 3 = 45 RRUs
- Data Rate -> 3 × 21.6 Mbit/s = 64.8 Mbit/s

In UMTS/HSPA+ various modulation and coding schemes are defined based on QPSK, 16QAM and 64QAM modulations, Table A.6. From all the possible Modulation and Coding Schemes presented, the thresholds and associated data rates per code presented in Table A.7 have been assumed.

IEEE802.11n is the wireless system chosen to represent the OFDM technology. The maximum data rate for IEEE802.11n with channels of 40 MHz, 400 ns guard interval and MIMO 4×4 is 600 Mbit/s. The existence of 108 sub-carriers per channel leads to consider the elementary radio unit as one sub-carrier, being its maximum data rate of 5.5 Mbit/s. Although this is a simplistic view, since sub-carriers are not assigned independently, it can be considered as an adequate approach as our concern is not the time scale of RRM scheduling function but a greater time span. For simulation purposes, an IEEE802.11n Access Point is considered as an OFDM BS, which has a maximum capacity of:

- Radio Units (sub-carriers) -> 108 RRUs
- Data Rate -> 600 Mbit/s

Various modulation schemes and coding rates are defined in IEEE802.11n standard being represented by a Modulation and Coding Scheme index value. From all the possible Modulation and Coding Schemes, the thresholds and associated data rates per sub-carrier presented in Table A.8 have been assumed.

Table A.6. Modulation, Coding Rate and required SINR for UMTS/HSPA (based on [HYFP04]).

| Modulation | Coding rate | SINR |
|---|---|---|
| No transmission | | $\gamma < 4$ |
| QPSK | 1/2 | $4 \leq \gamma < 6$ |
| | 2/3 | $6 \leq \gamma < 6.8$ |
| | 3/4 | $6.8 \leq \gamma < 10$ |
| 16QAM | 1/2 | $10 \leq \gamma < 12$ |
| | 2/3 | $12 \leq \gamma < 13$ |
| | 3/4 | $13 \leq \gamma < 17.7$ |
| 64QAM | 2/3 | $17.7 \leq \gamma < 21$ |
| | 3/4 | $21 \leq \gamma$ |

Table A.7. SINR and Data rate used for UMTS/HSPA+ (based on [HYFP04]).

| Modulation | SINR | Data rate per code [Mbit/s] |
|---|---|---|
| - | $\gamma < 4$ | 0 |
| QPSK | $4\leq \gamma < 7$ | 0.4 |
| | $7\leq \gamma < 10$ | 0.55 |
| 16QAM | $10 \leq \gamma < 13$ | 0.67 |
| | $13 \leq \gamma < 17.7$ | 0.96 |
| 64QAM | $17.7 \leq \gamma$ | 1.44 |

Table A.8. SINR and data rate used for IEEE802.11n (based on [IEEE09]).

| MCS Index | Modulation | Code Rate | SINR | Data rate [Mbit/s] | Data rate per sub-carrier [Mbit/s] |
|---|---|---|---|---|---|
| - | - | - | $\gamma \leq 4$ | 0 | 0 |
| 24 | BPSK | 1/2 | $4 < \gamma \leq 10$ | 60 | 0.56 |
| 25 | QPSK | 1/2 | $10 < \gamma \leq 12$ | 120 | 1.11 |
| 26 | | 3/4 | $12 < \gamma \leq 15$ | 180 | 1.67 |
| 27 | 16QAM | 1/2 | $15 < \gamma \leq 17.7$ | 240 | 2.22 |
| 28 | | 3/4 | $17.7 < \gamma \leq 21$ | 360 | 3.33 |
| 29 | 64QAM | 2/3 | $26 \leq \gamma$ | 480 | 4.44 |
| 30 | | 3/4 | $17.7 < \gamma \leq 21$ | 540 | 5 |
| 31 | | 5/6 | $26 \leq \gamma$ | 600 | 5.56 |

Concerning OFDMA technology the LTE is the representative system. The maximum data rate for LTE release 9 with MIMO 4×4 is 326 Mbit/s. Knowing that there are 102 sub-channels (PRBs), the maximum data rate of one resource block is 3.2 Mbit/s. The LTE BS is the example of OFDMA BS used for algorithm assessment as an OFDM BS, which has a maximum capacity of:

- Radio Units (sub-channel or PRB) -> 102 RRUs
- Data Rate -> 326 Mbit/s

Several modulation schemes and code rates are defined by 3GPP for LTE being represented by a CQI index value, Table A.9. Table A.10 shows the SINR thresholds for each modulation and code rate that have been assumed.

Table A.9. Modulation, Code Rate and Spectral efficiency used for LTE (extracted from [3GPP10]).

| CQI index | Modulation | Code rate (x 1024) | Efficiency (information bits per symbol) |
|---|---|---|---|
| 0 | out of range | | |
| 1 | QPSK | 78 | 0.1523 |
| 2 | | 120 | 0.2344 |
| 3 | | 193 | 0.3770 |
| 4 | | 308 | 0.6016 |
| 5 | | 449 | 0.8770 |
| 6 | | 602 | 1.1758 |
| 7 | 16QAM | 378 | 1.4766 |
| 8 | | 490 | 1.9141 |
| 9 | | 616 | 2.4063 |
| 10 | 64QAM | 466 | 2.7305 |
| 11 | | 567 | 3.3223 |
| 12 | | 666 | 3.9023 |
| 13 | | 772 | 4.5234 |
| 14 | | 873 | 5.1152 |
| 15 | | 948 | 5.5547 |

Table A.10. Modulation, coding rate and required SINR for LTE (based on [LHHC09]).

| MCS | Modulation | Coding rate | SINR [dB] |
|---|---|---|---|
| 0 | No transmission | | $\gamma < 0$ |
| 1 | QPSK | 1/4 | $0 \leq \gamma \leq 2.8$ |
| 2 | | 1/2 | $2.8 < \gamma \leq 8.5$ |
| 3 | 16QAM | 1/2 | $8.5 < \gamma \leq 12$ |
| 4 | | 3/4 | $12 < \gamma \leq 15.95$ |
| 5 | 64QAM | 2/3 | $15.95 < \gamma \leq 18.32$ |
| 6 | | 5/6 | $18.32 \leq \gamma$ |

The summary of the BS characteristics per wireless system considered is presented in Table A.11, identifying the RRU in each case, the number of RRUs per BS, and the maximum data rate per BS and per RRU.

Table A.11. Summary of BS characteristics from several RATs.

| | Radio Resource Unit (RRU) | Total number of RRUs per BS | Maximum data rate per BS [Mbit/s] | Maximum data rate per RRU [Mbit/s] |
|---|---|---|---|---|
| **TDMA (GSM/EGPRS)** | Time-slot | 24<br>(3 carriers x 8 time-slots) | 1.42<br>(473.6 kbit/s/carrier) | 0.059 |
| **CDMA (UMTS/HSPA+)** | Code | 45<br>(3 carriers x 15 codes) | 64.8<br>(21.6/carrier) | 1.44 |
| **OFDM (802.11n)** | Sub-carrier | 108 | 600 | 5.5 |
| **OFDMA (LTE 4×4MIMO)** | Sub-channel | 102 | 326 | 3.2 |

# Appendix B

# Traffic generation

In this appendix the traffic generated by VRRA simulator is presented. First the values obtained over time for the service inter-arrival time, and after the values obtained for service time and data volume per service.

## B.1 Inter-arrival Time per Service

The following graphs were obtained from simulations to verify the statistical behaviour of each service related parameter.  Figures B.1 to B.4 depict the inter-arrival time for each service considered.

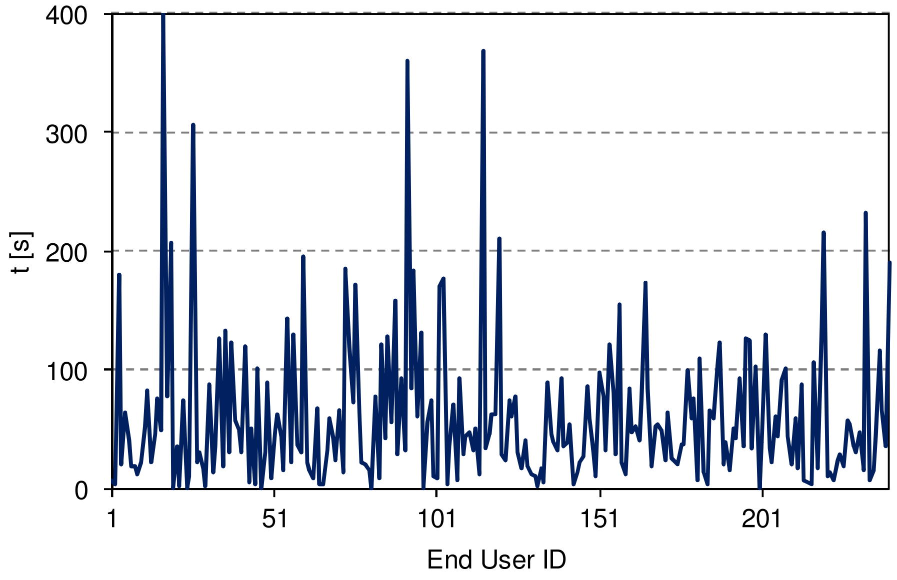



Figure B.1. Inter-arrival time for VoIP service with statistical distribution Exp[60].

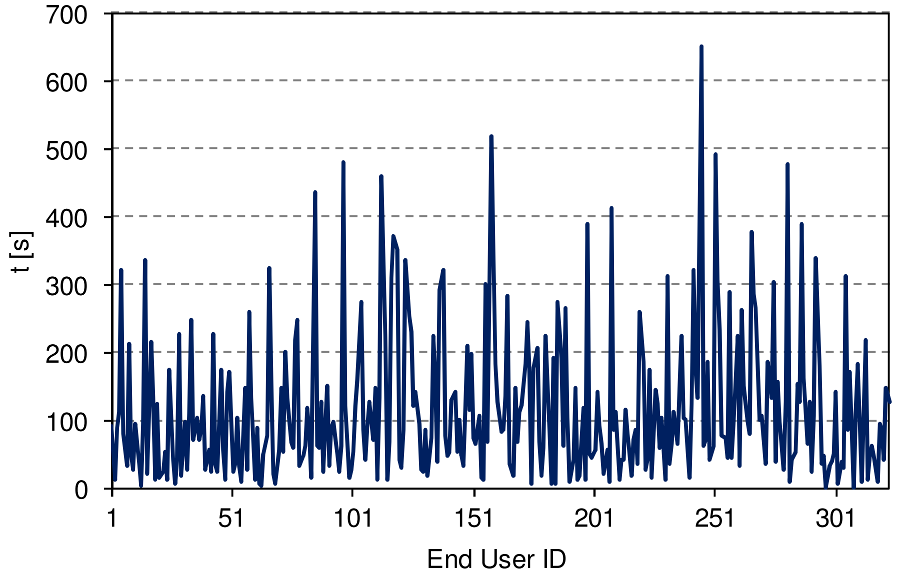



Figure B.2. Inter-arrival time for Video service with statistical distribution Exp[120] (obtained from simulator).

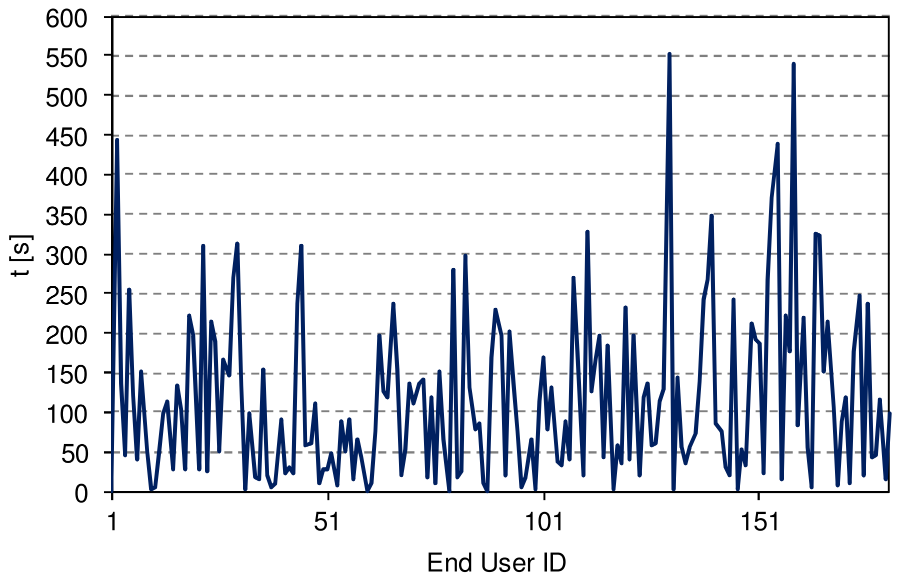



Figure B.3. Inter-arrival time for File Sharing service with statistical distribution Exp[120] (obtained from simulator).

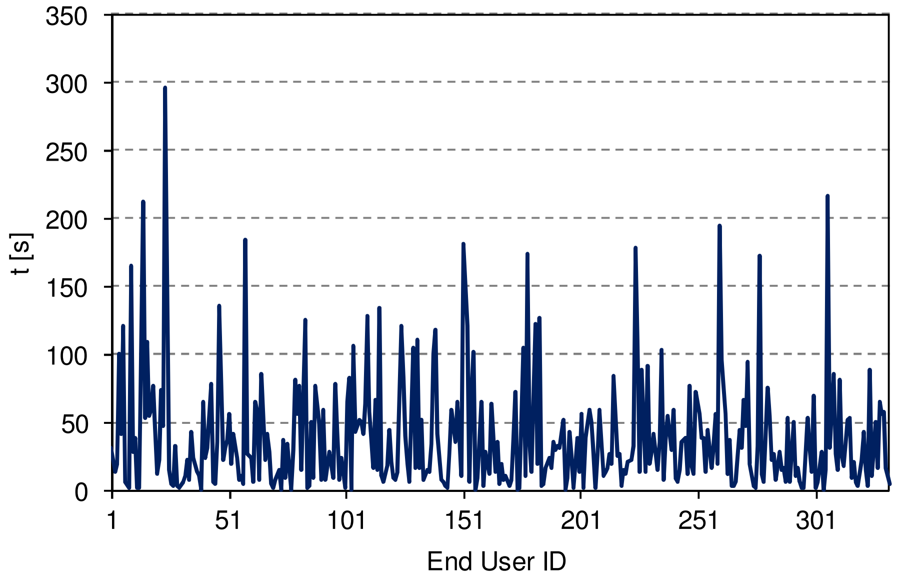



Figure B.4. Inter-arrival time for Web/Data service with statistical distribution Exp[36] (obtained from simulator).

## B.2 Service Time and Data Volume

The service time for VoIP is presented in Figure B.5, and the data volume for Video, File Sharing and Web/Data services are presented in Figures B.6 to B.8.

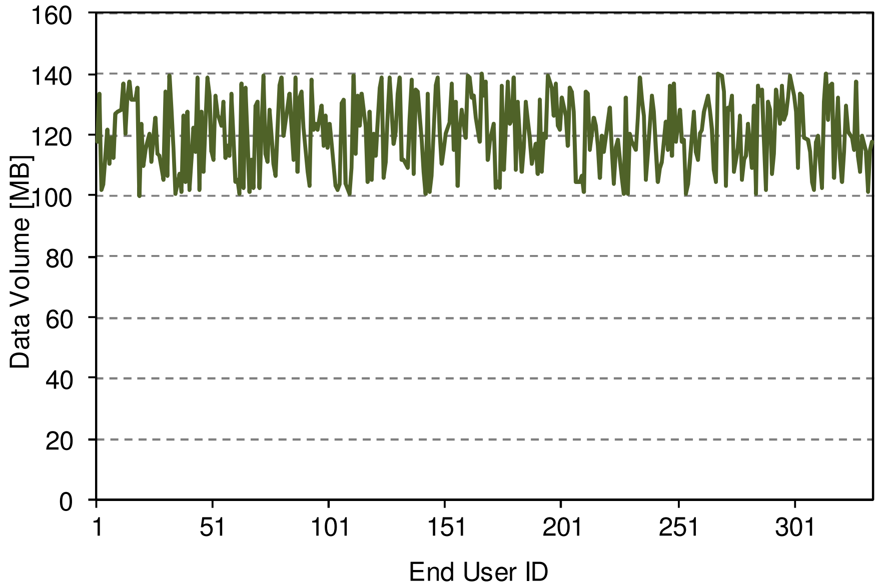



Figure B.5. Service time for VoIP service with statistical distribution Uniform[100,140]

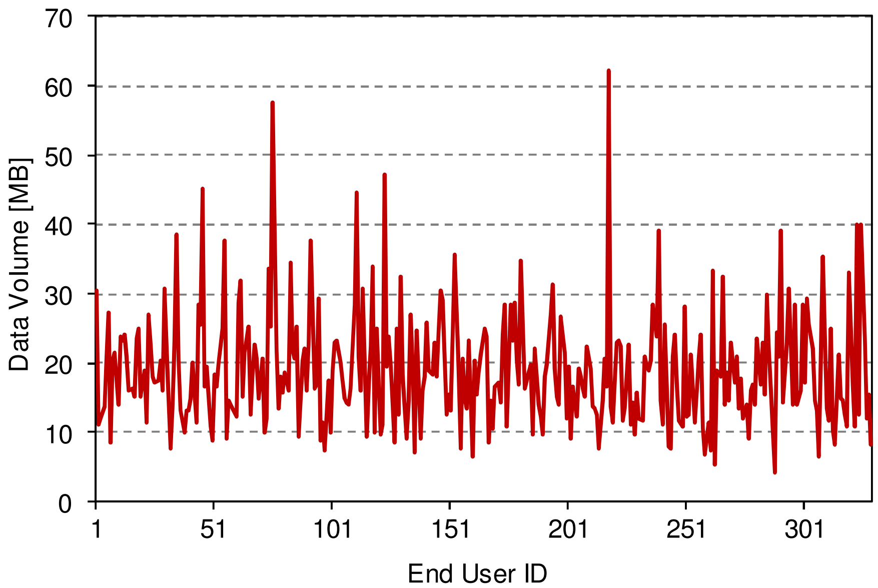



Figure B.6. Data Volume for Video service with statistical distribution Lognormal[17.5,10]

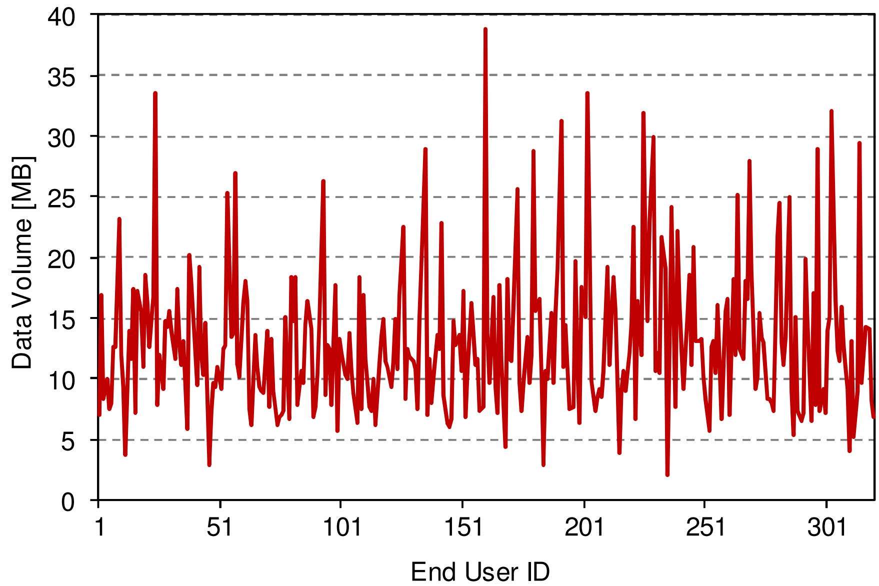



Figure B.7. Data Volume for File Sharing service with statistical distribution Lognormal[12.5,5]

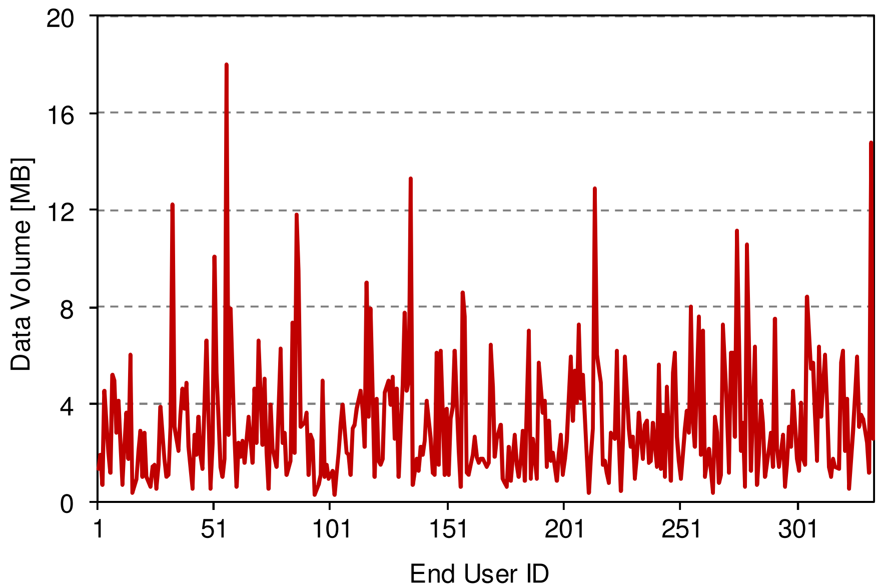



Figure B.8. Data Volume for Web/Data service with statistical distribution Lognormal[3,5].

# Appendix C

# Simulator Assessment Results

Results obtained for simulator assessment are presented in this appendix. Initially, the results related to the simulator transitory interval and after those related to the sensitivity to the number of simulations, are the results shown here.

## C.1 Simulator Transitory Interval

The values collected from simulation over time for *Average Cluster Serving Data Rate*, *VNet Out of Contract, Satisfaction Level on extra Capacity Requested* and *Average Delay on Service Request InP* are graphically represented in Figures C.1 to C.4, respectively.

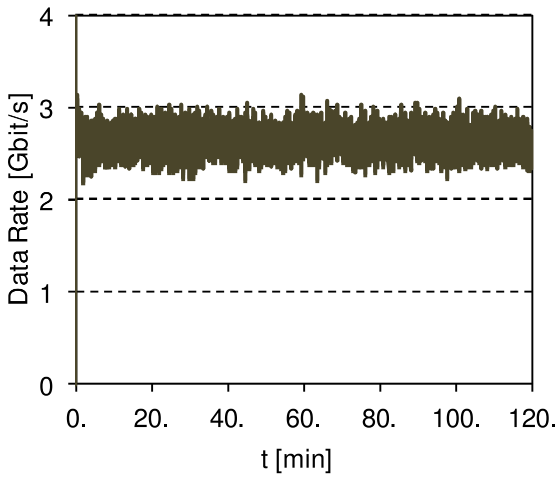



Figure C.1. *Average Cluster Serving Data Rate* over time.

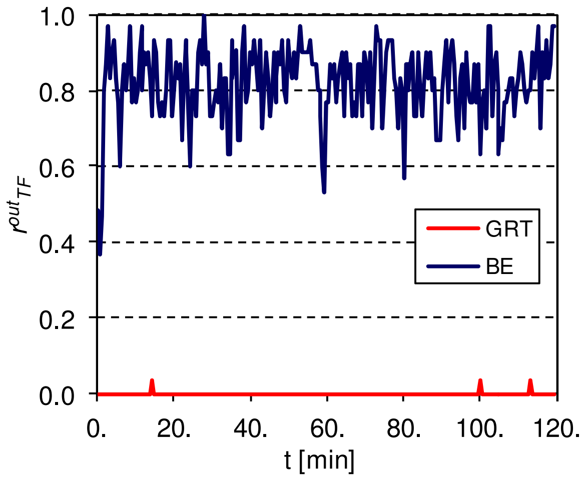



Figure C.2. *Out of Contract* over time.

The relative deviation for *Satisfaction Level on extra Capacity Requested*, *Delay on Service Request InP* and *Average VBS Time Service Delayed* are represented graphically in Figure C.5 to C.7, the numerical values being presented in Table C.1.

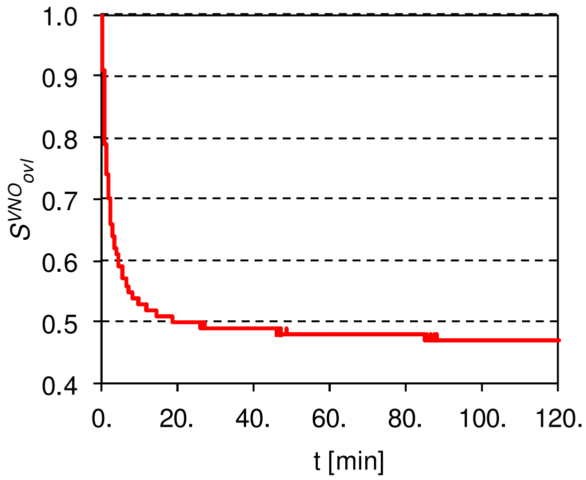



Figure C.3. *Satisfaction Level on extra Capacity Requested* over time.

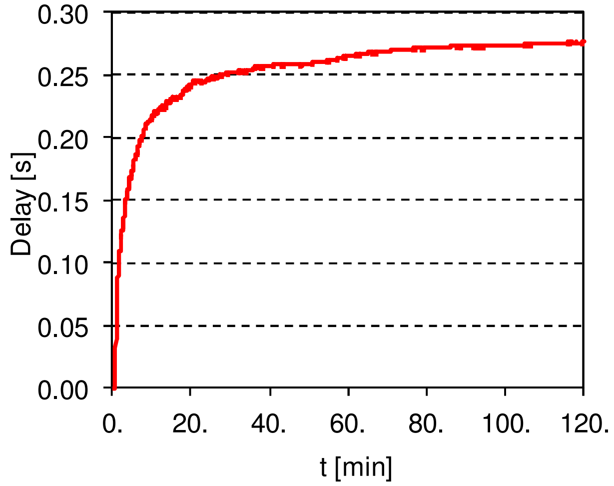



Figure C.4. *Delay on Service Request InP* over time.

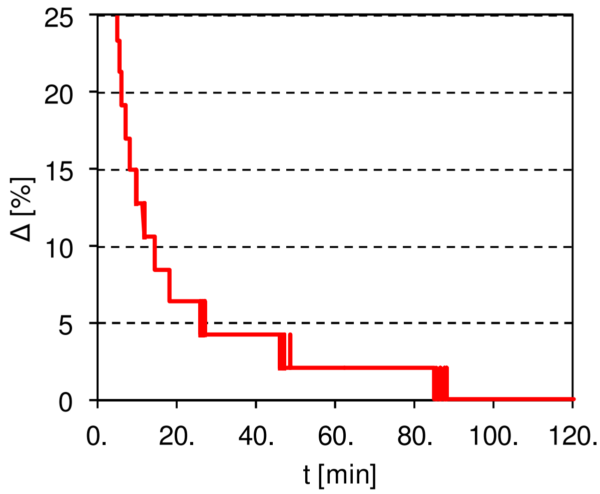



Figure C.5. Relative deviation of *Satisfaction Level on extra Capacity Requested*.

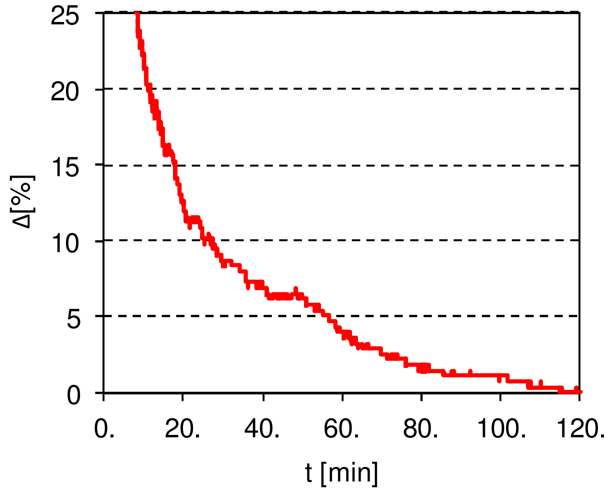
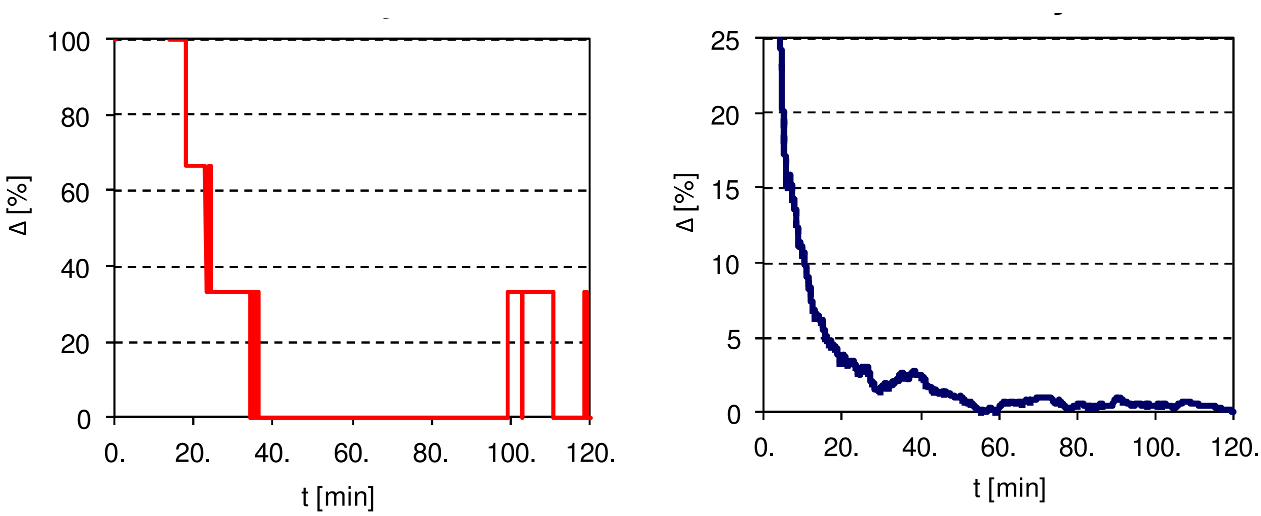



Figure C.6. Relative deviation of *Average Delay on Service Request InP*.




(a) VNet GRT (b) VNet BE

Figure C.7. Relative deviation of *Average VBS Time Service Delayed*.

Table C.1. Values and deviation for the several parameters.

| **Time [min]** | **VNet GRT** | | | | | | **VNet BE** | |
|---|---|---|---|---|---|---|---|---|
| | **Satisfaction Level** | | **Delay on Service Request** | | **Service Delayed** | | **Service Delayed** | |
| | **Value** | **Δ [%]** | **Value [ms]** | **Δ [%]** | **Value [ms]** | **Δ [%]** | **Value [s]** | **Δ [%]** |
| 10 | 0.53 | 12.8 | 216 | 21.7 | 7 | 133.3 | 20.15 | 10.6 |
| 15 | 0.51 | 8.5 | 231 | 16.3 | 6 | 100.0 | 21.19 | 6.0 |
| 20 | 0.50 | 6.4 | 242 | 12.3 | 5 | 66.7 | 21.74 | 2.1 |
| 25 | 0.50 | 6.4 | 248 | 10.1 | 4 | 33.3 | 21.86 | 3.0 |
| 30 | 0.49 | 4.3 | 252 | 8.7 | 4 | 33.3 | 22.15 | 1.7 |
| 120 | 0.47 | 0.0 | 276 | 0.0 | 3 | 0.0 | 22.53 | 0.0 |

## C.2 Sensitivity Analysis as a Function of the Number of Simulations

The results for different sets of simulations for *Satisfaction Level on extra capacity requested*, *Average VBS Time Service Delayed* and *Average Cluster Serving Data Rate*, are depicted graphically in Figures C.8 to C.11, as a function of the number of simulations performed. The *Out of Contract* for VNet GRT is not represented graphically as it is constant for the different sets of simulations, as it can be verified in Table C.2.

To quantify the variation achieved in the observations, the deviation percentage relative to the average of all simulation values, computed from (5.1), for each set of simulations, are presented in Table C.2.

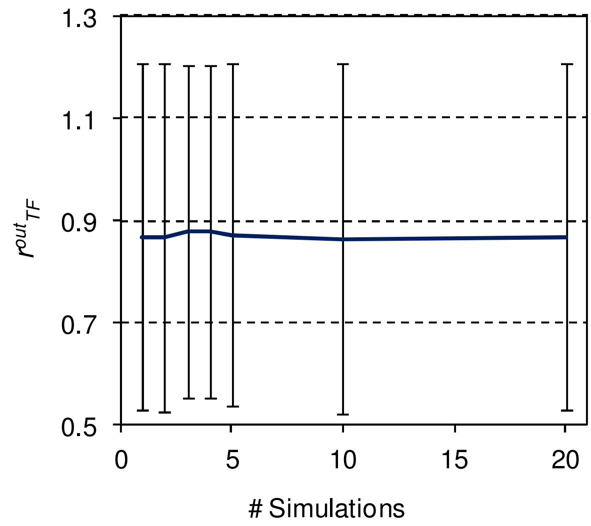



Figure C.8. *Out of Contract* (VNet BE) for different number of simulations.

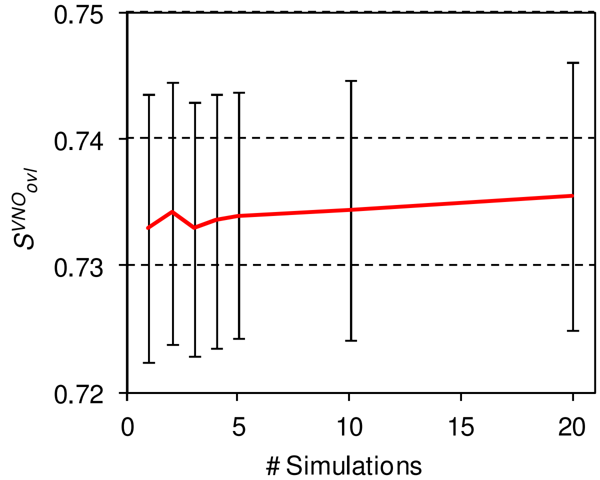
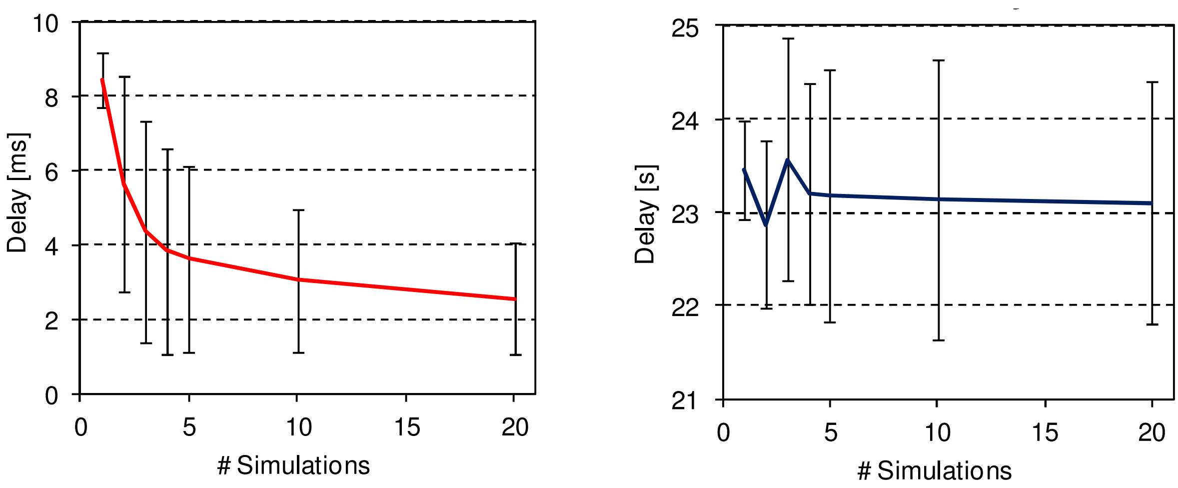



Figure C.9. *Satisfaction Level on extra Capacity Requested* for different number of simulations.




(a) VNet GRT (b) VNet BE

Figure C.10. *Average VBS Time Service Delayed* for different number of simulations.

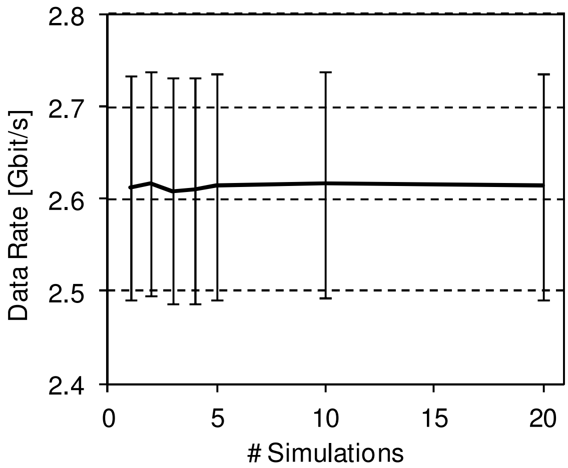



Figure C.11. *Average Cluster Serving Data Rate* for different number of simulations.

Table C.2. Average, relative and standard deviation for several number of simulations.

| VNet | # Sim. | Serving Data Rate | | | Out of Contract | | | Service Delayed | | | Satisfaction Level | | | Delay on Service Request | | | Cluster Data Rate | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\overline{X}$ [Mbit/s] | Δ [%] | σ [Mbit/s] | $\overline{X}$ | Δ [%] | σ | $\overline{X}$ [ms] | Δ [%] | σ [ms] | $\overline{X}$ | Δ [%] | σ | $\overline{X}$ [ms] | Δ [%] | σ [ms] | $\overline{X}$ [Mbit/s] | Δ [%] | σ [Mbit/s] |
| GRT | 1 | 1251.06 | | 0.64 | | | | 8 | 166.7 | 1 | 0.73 | 1.4 | | 271 | 1.1 | 6 | 2613.31 | 0.0 | 121.16 |
| | 2 | 1251.05 | | 0.64 | | | | 6 | 100.0 | 3 | 0.73 | 1.4 | | 268 | 0.0 | 8 | 2616.30 | 0.1 | 121.41 |
| | 3 | 1251.05 | | 0.63 | | | | 4 | 33.3 | 3 | 0.73 | 1.4 | | 269 | 0.4 | 8 | 2609.24 | 0.2 | 121.97 |
| | 4 | 1251.05 | 0 | 0.63 | 0 | 0 | 0.02 | 4 | 33.3 | 3 | 0.73 | 1.4 | 0.01 | 268 | 0.0 | 8 | 2609.29 | 0.2 | 121.69 |
| | 5 | 1251.05 | | 0.63 | | | | 4 | 33.3 | 2 | 0.73 | 1.4 | | 268 | 0.0 | 7 | 2613.73 | 0.0 | 122.51 |
| | 10 | 1251.05 | | 0.63 | | | | 3 | 0.0 | 2 | 0.73 | 1.4 | | 268 | 0.0 | 7 | 2616.36 | 0.1 | 122.81 |
| | 20 | 1251.06 | | 0.63 | | | | 3 | 0.0 | 2 | 0.74 | 0.0 | | 268 | 0.0 | 9 | 2614.17 | 0.0 | 122.40 |
| BE | 1 | 1362.25 | 0.1 | 121.18 | 0.87 | 0.0 | 0.34 | 23.46 | 1.6 | 0.53 | | | | | | | | | |
| | 2 | 1365.25 | 0.2 | 122.52 | 0.87 | 0.3 | 0.34 | 22.87 | 1.0 | 0.90 | | | | | | | | | |
| | 3 | 1358.18 | 0.4 | 122.81 | 0.88 | 1.2 | 0.33 | 23.57 | 2.0 | 1.31 | | | | | | | | | |
| | 4 | 1358.24 | 0.4 | 122.39 | 0.88 | 1.3 | 0.33 | 23.20 | 0.4 | 1.18 | | | | | | | | | |
| | 5 | 1362.68 | 0.0 | 122.02 | 0.87 | 0.4 | 0.33 | 23.17 | 0.3 | 1.35 | | | | | | | | | |
| | 10 | 1365.31 | 0.2 | 122.50 | 0.86 | 0.5 | 0.34 | 23.14 | 0.2 | 1.50 | | | | | | | | | |
| | 20 | 1363.11 | 0.0 | 122.46 | 0.87 | 0.0 | 0.34 | 23.10 | 0.0 | 1.30 | | | | | | | | | |

# References

[3GPP03] 3GPP, *Improvement of RRM across RNS and RNS/BSS (post-Rel-5) (Release6)*, RAN, Technical Report TR 25.891 v0.3.0, June 2003, (http://www.3gpp.org).

[3GPP08] 3GPP, *3GPP System Architecture Evolution (SAE): Report on technical options and conclusions*, Technical Report TR 23.882 v8.0.0, Sep. 2008, (http://www.3gpp.org).

[3GPP09a] 3GPP, *3GPP Technical Specification Group GSM/EDGE Radio Access Network; Physical layer on the radio path; General description (Release 8)*, Technical Specification TS 145.001 v8.0.0, Feb. 2009 (http://www.3gpp.org).

[3GPP09b] 3GPP, *3GPP Technical Specification Group GSM/EDGE Radio Access Network; Physical layer on the radio path; General description (Release 9)*, Technical Specification TS 45.001 v9.0.0, May 2009 (http://www.3gpp.org).

[3GPP10] 3GPP*, 3GPP Technical Specification Group Radio Access Network; LTE - Evolved Universal Terrestrial Radio Access; Physical Layer Procedures (Release 9),* Technical Specification TS 36.213 v9.3.0, Oct. 2010 (http://www.3gpp.org).

[3GPP12] 3GPP*, 3GPP Technical Specification Group Services and Systems Aspects; Policy and Charging Control Architecture (Release 10),* Technical Specification TS 23.203 v10.6.0, Mar. 2012 (http://www.3gpp.org).

[3GPP13a] 3GPP, *3GPP Technical Specification Universal Mobile Telecommunications System (UMTS); LTE; Network Sharing; Architecture and functional description (Release 11)*, Technical Specification TS 23.251 v11.4.0, Jan. 2013 (http://www.3gpp.org).

[3GPP13b] 3GPP, *Study on Radio Access Network (RAN) sharing enhancements (Release 12)*, Technical Report TR 22.852 v12.0.0, June 2013 (http://www.3gpp.org).

[4WAR10] *Architecture and Design for the Future Internet* (4WARD), EC FP7-ICT Project 216041, June 2010 (www.4ward-project.eu).

[Agil09] Agilent, *3GPP Long Term Evolution: System Overview, Product Development, and Test Challenges*, Agilent Technologies White Paper, June 2009 (http://cp.literature.agilent.com/litweb/pdf/5989-8139EN.pdf).

[AKAR06] *Architecture Design Project for New Generation Networks* (AKARI), NICT Project, Japan, May 2006 (http://akari-project.nict.go.jp).

[AlBa05] Al-Jarbou,Y. and Baroudi,U., "Performance Of Heterogeneous Traffic In Roaming Based Sharing Multi-Operator WCDMA Networks", in *Proc. of ISWCS2005 - 2nd International Symposium on Wireless Communication Systems*, Siena, Italy, Sep. 2005.

[Alca12] Alcatel-Lucent, *Network Sharing in LTE: Opportunities & Solutions*, Technology White Paper, July 2012 (http://alcatellucentmediaroom.files.wordpress.com/2010/07/lte_network_sharing_en_techwhitepaper1.pdf).

[AMSE06] AlQahtani,S.A., Mahmoud,A.S., Sheltami,T.R. and El-Tarhuni,M., "Adaptive Radio Resource Management for Multi-Operator WCDMA Based Cellular Wireless Networks with Heterogeneous Traffic", in *Proc. of PIMRC'06 - 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Helsinki, Finland, Sep. 2006.

[Ambi07] *Ambient Networks* (AN), EC FP6-IST Project 027662, Jan. 2007 (www.ist-world.org).

[APST05] Anderson,T., Peterson,L., Shenker,S. and Turner,J., "Overcoming the Internet impasse through Virtualization", *IEEE Computer Magazine*, Vol. 38, No. 4, Apr. 2005, pp.34-41.

[Bauc08a] Baucke,S. (ed.), *Milestone Report*, Report R-3.1, EC ICT-4WARD Project, Sep. 2008 (www.4ward-project.eu).

[Bauc08b] Baucke,S. (ed.), *Virtualization Approach: Concept (Draft)*, Deliverable D-3.1.0, EC ICT-4WARD Project, Jan. 2009 (www.4ward-project.eu).

[Bauc09a] Baucke,S. (ed.), *Virtualization Approach: Concept*, Deliverable D-3.1.1, EC ICT-4WARD Project, July 2009 (www.4ward-project.eu).

[Bauc09b] Baucke,S. (ed.), *Virtualization Approach: Evaluation and Integration (Draft)*, Deliverable D-3.2.0, EC ICT-4WARD Project, Dec. 2009 (www.4ward-project.eu).

[Bauc10a] Baucke,S. (ed.), *Virtualization Approach: Evaluation and Integration*, Deliverable D-3.2.1, EC ICT-4WARD Project, Jan. 2010 (www.4ward-project.eu).

[Bauc10b] Baucke,S. (ed.), *Virtualization Approach: Evaluation and Integration - Update*, Deliverable D-3.2.2, EC ICT-4WARD Project, June 2010 (www.4ward-project.eu).

[BBCC04] Bavier,A., Bowman,M., Chun,B., Culler,D., Karlin,S., Muir,S., Peterson,L., Roscoe,T., Spalink,T. and Wawrzoniak, M., "Operating system support for planetary-scale network services", in *Proc. of USENIX NSDI'04 - 1st Symposium on Networked Systems Design and Implementation*, San Francisco, CA, USA, Mar. 2004.

[BFHP06] Bavier,A., Feamster,N., Huang,M., Peterson,L. and Rexford,J., "In VINI Veritas: Realistic and controlled network experimentation", in *Proc. of ACM SIGCOMM 2006 - Conference of the Special Interest Group on Data Communication*, Pisa, Italy, Sep. 2006.

[Bing90] Bingham,J., "Multicarrier modulation for data transmission: an idea whose time has come", *IEEE Communications Magazine*, Vol. 28, No. 5, May 1990, pp. 5-14.

[BMMM08] Bhatia,S., Motiwala,M., Mühlbauer,W., Mundada,Y., Valancius,V., Bavier,A., Feamster,N., Peterson,L. and Rexford,J., "Trellis: A platform for building flexible, fast VNets on commodity hardware", in *Proc. of ACM CoNEXT 2008 - Conference on Emerging Network Experiment and Technology*, Madrid, Spain, Dec. 2008.

[BSMR10] Bhanage,G., Seskar,I., Mahindra,R. and Raychaudhuri,D., "Virtual Basestation: Architecture for an Open Shared WiMAX Framework", in *Proc. of VISA'10 - Virtualized infrastructure Systems and Architectures*, New Delhi, India, Sep. 2010.

[BVSR10] Bhanage,G., Vete,D., Seskar,I. and Raychaudhuri,D., "SplitAP: Leveraging Wireless Network Virtualization for Flexible Sharing of WLANs", in *Proc. of GLOBECOM 2010 - Global Communication Conference*, Miami, FL, USA, Dec. 2010.

[ChBo10] Chowdhury,N.M. and Boutaba,R., "A Survey of Network Virtualization", *Computer Networks*, Vol. 54, No 5, Apr. 2010, pp. 862-876.

[ChGu06] Chen,H. and Guizani,M., *Next Generation Wireless Systems and Networks*, John Wiley & Sons Ltd, London, UK, 2006.

[Cisc12a] Cisco Systems, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016*, Cisco Visual Networking Index (VNI), Feb. 2012 (www.cisco.com).

[Cisc12b] Cisco Systems, *Cisco Visual Networking Index: Forecast and Methodology, 2011–2016*, Cisco Visual Networking Index (VNI), May 2012 (www.cisco.com).

[Cisc13] Cisco Systems, *The Zettabyte Era – Trends and Analysis*, Cisco Visual Networking Index (VNI), May 2013 (www.cisco.com).

[Dike00] Dike,J., "A Usermode Port for the Linux Kernel", in *Proc. of USENIX ALS'00 - 4th Annual Linux Showcase & Conference*, Atlanta, GA, USA, Oct. 2000.

[ElFa09] Elliott,C. and Falk,A., "An update on the GENI project", *ACM SIGCOMM Computer Communication Review*, Vol. 39, No.3, June 2009, pp. 28-34.

[FeGR07] Feamster,N., Gao,L. and Rexford,J., "How to Lease the Internet in Your Spare Time", *ACM SIGCOMM Computer Communication Review*, Vol.37, No.1, Jan. 2007, pp.61-64.

[Fren11] Frenzel,L.E., *Wireless Companies Follow The Roadmap Past 4G And On to 5G*, Electronic Design, June 2011 (http://electronicdesign.com/print/4g/wireless-companies-follow-roadmap-past-4g-and-5g).

[GEPW07] Gross,J., Emmelmann,M., Puñal,O. and Wolisz,A., "Dynamic Single-User OFDM Adaptation for IEEE 802.11 Systems", in *Proc. of MSWiM'07 - 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, Chania, Greece. Oct. 2007.

[HaGB05] Hassan,H., Garcia,J.M. and Brun,O., "Generic modeling of multimedia traffic sources", in *Proc. of HET-NETs '05 - 3rd International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks*, Ilkley, UK, July 2005.

[HeWh06] Hew,S. and White,L.B., "Fair Resource Bargaining Solutions for Cooperative Multi-Operator Networks", in *Proc. of IZS - International Zurich Seminar on Communications*, Zurich, Switzerland, Feb. 2006.

[HJSY06] Ho,S.K., Jongkyung,K., Sung,K.T., Yong,K.Y. and Jongsoo,S., "Adaptive Admission Control Algorithm for Multiuser OFDMA Wireless Networks", in *Proc. of INFOCOM 2006 - 25th IEEE International Conference on Computer Communications*, Barcelona, Spain, Apr. 2006.

[HLPS05] Hooli,K., Lara,J., Pfletschinger,S., Sternad,M., Thilakawardana,S. and Yutao,Z., *WINNER Spectrum Aspects: Assessment Report*, Deliverable D6.3, EC IST-WINNER Project, Dec. 2005 (www.ist-winner.org).

[HLZP10] Houidi,I., Louati,W., Seghlache,D., Papadimitriou,P. and Mathy,L., "Adaptive VNet Provisioning", in *Proc. of VISA'10 - Virtualized infrastructure Systems and Architectures*, New Delhi, India, Sep. 2010.

[HoTo04] Holma,H. and Toskala,A., *WCDMA for UMTS*, John Wiley & Sons, Chichester, UK, 2004.

[HSHL97] Hamalainen,S., Slanina,P., Hartman,M., Lappetelainen,A., Holma,H. and Salonaho,O., "A novel interface between link and system level simulations", in *Proc. of ACTS Mobile Telecommunications Summit'97*, Aalborg, Denmark, Oct. 1997.

[HYFP04] Hu,H., Yanikomeroglu,H., Falconer,D. and Periyalwar,S., "Range Extension without Capacity Penalty in Cellular Networks with Digital Fixed Relays", in *Proc. of IEEE Globecom 2004 - Global Communication Conference*, Dallas, TX, USA, Dec. 2004.

[HZLL08] He,J., Zhang-Shen,R., Li,Y., Lee,C.-Y., Rexford,J. and Chiang,M., "Davinci: Dynamically Adaptive VNets for a customised internet", in *Proc. of ACM CoNEXT 2008 - 4th ACM International Conference on Emerging Networking Experiments and Technologies*, Madrid, Spain, Dec. 2008.

[IC1013] *Cooperative Radio Communications for Green Smart Environments* (IC1004), European Cooperation in the Field of Scientific and Technical Research – COST Action IC1004, Jan. 2013 (http://www.ic1004.org).

[IEEE04a] IEEE, *802.1D - IEEE Standard for Local and Metropolitan Area Networks. Media Access Control (MAC) Bridges*, 802.1 WG, IEEE 802.1D-2004, June 2004 (http://www.ieee.org).

[IEEE04b] IEEE, *802.16 - IEEE Standard for Local and Metropolitan Area Networks. Part16: Air Interface for Fixed Broadband Wireless Access Systems*, 802.16 WG, IEEE 802.16-2004, Oct. 2004 (http://www.ieee.org).

[IEEE05a] IEEE, *IEEE Standard for Information Technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*, 802.11 WG, IEEE 802.11e-2005, Sep. 2005 (http://www.ieee.org).

[IEEE05b] IEEE, *IEEE Standard for Local and metropolitan area networks: Virtual Bridged Local Area Networks*, 802.1 WG, IEEE 802.1Q-2005, Dec. 2005, (http://www.ieee.org).

[IEEE06] IEEE, 802.16 - IEEE Standard for Local and Metropolitan Area Networks. Part16: Air Interface for Fixed Broadband Wireless Access Systems - Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, 802.16 WG, IEEE 802.16-2004, Feb. 2006 (http://www.ieee.org).

[IEEE09] IEEE, *IEEE Standard for Information technology - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 5: Enhancements for Higher Throughput*, WG802.11 - Wireless LAN Working Group, IEEE 802.11n-2009, Oct. 2009 (http://www.ieee.org).

[JoKS04] Johansson,K., Kristensson,M. and Schwarz,U., “Radio Resource Management in Roaming Based Multi-Operator WCDMA Networks”, in *Proc. of VTC 2004 Spring - 59$^{th}$ IEEE Vehicular Technology Conference*, Milan, Italy, May 2004.

[KlLL01] Klemm,A., Lindemann,C. and Lohmann,M., “Traffic Modelling and Characterisation for UMTS Networks”, in *Proc. of IEEE Globecom 2001 - Internet Performance Symposium*, San Antonio, TX, USA, Nov. 2001.

[KMZR12] Kokku,R., Mahindra,R., Zhang,H. and Rangarajan,S., “NVS: A Substrate for Virtualizing Wireless Resources in Cellular Networks”, *IEEE/ACM Transactions on Networking*, Vol. 20, No. 5, Oct. 2012, pp. 1333-1346.

[KMZR13] Kokku,R., Mahindra,R., Zhang,H. and Rangarajan,S., “Cellular Wireless Resource Slicing For Active RAN Sharing”, in *Proc. of COMSNETS 2013 – 5th International Conference on Communication Systems and Networks*, Bangalore, India, Jan. 2013.

[LCBG12] Landi,G., Ciulli,N., Buysse,J., Georgakilas,K., Anastasopoulos,M., Tzanakaki,A., Develder,C., Escalona,E. and Parniewicz,D., "A Network Control Plane architecture for on-demand co-provisioning of optical network and IT services", in *Proc. of Future Network & Mobile Summit 2012*, Berlin, Germany, July 2012.

[LDCQ01] Leung,K., Driessen,P., Chawla,K. and Qiu,X., “Link Adaptation and Power Control for Streaming Services in EGPRS Wireless Networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 19, No. 10, Oct. 2001, pp.2029-2039.

[LHHC09] Lee,H., Hany,K., Hwangz,Y. and Choi S., “Opportunistic Band Sharing for Point-to-Point Link Connection of Cognitive Radios”, in *Proc. of CROWNCOM '09 - 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Hannover, Germany, June 2009.

[LiNi05] Li,L. and Niu,Z., “A Resource Allocation and Admission Control Scheme for Multimedia Services in Broadband Wireless Communication Systems”, in *Proc. of APCC 2005 - 11th Asia-Pacific Conference on Communications*, Perth, Australia, Apr. 2005.

[Lope08] Lopes,J., *Performance Analysis of UMTS/HSDPA/HSUPA at the Cellular Level*, M. Sc. Thesis, IST - Technical University of Lisbon, Lisbon, Portugal, 2008.

[LYLZ12] Lu,X., Yang,K., Liu,Y., Zhou,D. and Liu,S., “An elastic resource allocation algorithm enabling wireless network virtualization”, *Wireless Communications and Mobile Computing*, Vol.14, No 1, Dec. 2012.

[Maca91] Macario,R., *Personal and Mobile Radio Systems,* P. Peregrinus, London, UK, 1991.

[Meag02] Meago,F., *Common Radio Resource Management (CRRM)*, Temporary Document TD(02)046, COST273 Project, Espoo, Finland, May 2002.

[MIND02] Mobile IP based Network Developments (MIND), EC IST Project 28584, Nov. 2002 (www.ist-world.org).

[MiSu13] Miron,A. and Suciu,L. (eds.), *Applications for connectivity services and evaluation*, Deliverable D-4.4, EC ICT-SAIL project, Feb. 2013 (www.sail-project.eu).

[MoOb06] Moisio,M. and Oborina,A., “Comparison of effective SINR mapping with traditional AVI approach for modeling packet error rate in multi-state channel”, in *Proc. of NEW2AN 2006 - 6th International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking*, St. Petersburg, Russia, May 2006.

[Mous12] Mousa,A.M., “Prospective of Fifth Generation Mobile Communications”, *International Journal of Next-Generation Networks (IJNGN)*, Vol. 4, No. 3, Sep. 2012, pp.11-30.

[Muss12] Mussman,H., *GENI: An Introduction*, GENI Project Office at Raytheon, Feb. 2012 (http://www.geni.net).

[NaBK00] Nanda,S., Balachandran,K. and Kumar,S., “Adaptation Techniques in Wireless Packet Data Services”, *IEEE Communications Magazine*, Vol. 38, No. 1, Jan. 2000, pp. 54-64.

[NEWC11] *Network of Excellence in Wireless Communications++* (NEWCOM++), EC FP7-ICT Project, Apr. 2011.

[Orbi12] www.orbit-lab.org, Nov 2012.

[Plan10] www.planet-lab.org, March 2010.

[PMRK06] Peterson,L., Muir,S., Roscoe,T. and Klingaman,A., *PlanetLab Architecture: An Overview*, Technical Report PDN–06–031, May 2006 (https://www.planet-lab.org).

[PrCJ02] Prabhu,B.J., Chockalingam,A. and Jayaram,D., “Performance of Code Allocation Algorithms on UMTS Uplink with Mixed Voice/Data Traffic”, in *Proc. of ICPWC’02 - IEEE International Conference on Personal Wireless Communications*, New Delhi, India, Dec. 2002.

[RaWa07] Schoenen,R. and Walke,B., “On PHY and MAC performance of 3G-LTE in a multi-hop cellular environment”, in *Proc. of WiCom 2007 - International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China, Sep. 2007.

[Rome10] Romero,J.P. (ed.), “Final report of the JRRM and ASM activities”, Deliverable DR9.3, NEWCOM++ Project, Nov. 2010 (www.newcom-project.eu).

[RRSS05] Ryu,S., Ryu,B., Seo,H., Shin,M. and Park,S., “Wireless Packet Scheduling Algorithm for OFDMA System Based on Time-Utility and Channel State”, *ETRI Journal*, Vol. 27, No. 6, Dec. 2005., pp. 777-787.

[RSAD05] Perez-Romero,J., Sallent,O., Agusti,R. and Diaz-Guerra,M., *Radio Resource Management Strategies in UMTS*, John Wiley & Sons, Chichester, UK, 2005.

[RSOG05] Raychaudhuri,D., Seskar,I., Ott,M., Ganu,S., Ramachandran,K., Kremo,H., Siracusa,R., Liu,H. and Singh,M., “Overview of the Orbit Radio Grid Testbed for Evaluation of Next-Generation Wireless Network Protocols”, in *Proc. of WCNC 2005 - IEEE Wireless Communications and Networking Conference*, New Orleans, LA, USA, Mar. 2005.

[SaBa08] Sachs,J. and Baucke,S., “Virtual radio: a framework for configurable radio networks”, in

*Proc. of WICON '08 - 4th Annual International Conference on Wireless Internet*, Maui, HI, USA, Nov. 2008.

[SABG06] Sachs,J., Aguero,R., Berg,M., Gebert,J., Jorguseski,L., Karla,I., Karlsson,P., Koudouridis,G.P., Lundsjo,J., Prytz,M. and Strandberg,O., "Migration of Existing Access Networks Towards Multi-Radio Access", in *Proc. of VTC'06 Fall - 64th Vehicular Technology Conference*, Montreal, Canada, Sep. 2006.

[SAIL13] Scalable and Adaptive Internet Solutions (SAIL), EC FP7-ICT Project 257448, Jan. 2013 (www.sail-project.eu).

[SaSr06] Sanjoy,P. and Srini,S., *Technical Document on Wireless Virtualization*, Document GDD-06-17, GENI Project, USA, Sep. 2006. (http://www.geni.net).

[Scou97] Scourias,J., *Overview of the Global System for Mobile Communications*, Internal Report, University of Waterloo, Waterloo, Canada, Oct. 1997, (http://ccnga.uwaterloo.ca/~jscouria/GSM/gsmreport.html).

[Seba07] Sebastião,D., *Algorithms for Quality of Service in a Wi-Fi Network*, M.Sc. Thesis, IS - Technical University of Lisbon, Lisbon, Portugal, Dec. 2007.

[SeCo07] Serrador,A. and Correia,L.M., "Policies for a Cost Function for Heterogeneous Networks Performance Evaluation", in *Proc. of PIMRC 2007 - IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, Athens, Greece, Sep. 2007.

[SHHS02] Sumaryo,S., Hepworth,E., Higgins,D. and Siebert,M., "A Radio Resource Management Architecture for a Beyond-3G Network", in *Proc. of International Workshop IST-MIND*, Budapeste, Hungary, Nov. 2002.

[SiSD12] Singh,D., Srinivas,K and Das,D., "A Dynamic Channel Assignment in GSM Telecommunication network using Modified Genetic Algorithm", in *Proc. of EATIS 2012 - 6th Euro American Conference on Telematics and Information Systems*, Valencia, Spain, May 2012.

[Suci11] Suciu,L. (ed.), *Architectural Concepts for Connectivity Services*, Deliverable D-4.1, EC ICT-SAIL Project, Jul. 2011 (www.sail-project.eu).

[SuTi12] Suciu,L. and Timm-Giel,A. (eds.), *Architectural Concepts for Connectivity Services - Addendum*, Deliverable D-4.1-A, EC ICT-SAIL Project, Jan. 2012 (www.sail-project.eu).

[SWLM04] Sachs,J., Wiemann,H., Lundsjo,J. and Magnusson,P., "Integration of multi-radio access in a beyond 3G network", in *Proc. of PIMRC 2004 - 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Barcelona, Spain, Sep. 2004.

[SWPF09] Schaffrath,G., Werle,C., Papadimitriou,P., Feldmann,A., Bless,R., Greenhalgh,A., Wundsam,A., Kind,M., Maennel,O. and Mathy,L., “Network Virtualization Architecture: Proposal and Initial Prototype”, in *Proc. of VISA'09 - ACM SIGCOMM Workshop on Virtualized infrastructure Systems and Architectures*, Barcelona, Spain, Aug. 2009.

[TCST04] Takeda,D., Chow,Y.C., Strauch,P. and Tsurumi, H., “Threshold Controlling Scheme For Adaptive Modulation And Coding System”, in *Proc. of PIMRC 2004 - 15$^{th}$ IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Barcelona, Spain, Sep. 2004.

[TiSu13] Timm-Giel,A. and Suciu,L. (eds.), *Architecture and Mechanisms for Connectivity Services*, Deliverable D-4.2, EC ICT-SAIL Project, Feb. 2013 (www.sail-project.eu).

[TsSo03] Tsai,S. and Soong,A., *Effective-SNR Mapping for Modelling Frame Error Rates in Multiple-state Channels*, Technical report 3GPP2-C30-20030429-010, Apr. 2003 (ftp://ftp.3gpp2.org).

[TuWa05] Tuomaala,E. and Wang,H., “Effective SINR Approach Of Link To System Mapping In OFDM/Multi-Carrier Mobile Network”, in *Proc. of IEEE Mobility Conference 2005 - 2$^{nd}$ International Conference on Mobile Technology, Applications and Systems*, Guangzhou, China, Nov. 2005.

[WiFi07] Wi-Fi Alliance, *Wi-Fi CERTIFIED™ 802.11n draft 2.0: Longer-Range, Faster-Throughput, Multimedia-Grade Wi-Fi®Networks*, Internal Report, 2007 (http://www.wi-fi.org).

[Will07] Williams,D.E., *Virtualization with Xen: Including XenEnterprise, XenServer, and XenExpress*, Syngress Publishing, ACM eBooks, 2007.

[WiMA09] WiMAX Forum, *WiMAX, HSPA+, and LTE: A Comparative Analysis*, Internal Report, Nov. 2009 (http://resources.wimaxforum.org).

[XKYG11] Xia,L., Kumar,S., Yang,X., Gopalakrishnan,P., Liu,Y., Schoenberg,S. and Guo,X., “Virtual Wi-Fi: Bring Virtualization from Wired to Wireless”, in *Proc. of VEE’11 – International Conference on Virtual Execution Environments*, Newport Beach, CA, USA, Mar. 2011.

[ZhRe05] Zheng,J. and Regentova,E., “QoS-Based Dynamic Channel Allocation for GSM/GPRS Networks”, in *Proc. of NPC 2005 – IFIP International Conference on Network and Parallel Computing*, Beijing, China, Dec. 2005.

[ZhVi05] Zheng,H. and Viswanathan,H., “Optimizing the ARQ Performance in Downlink Packet Data Systems with Scheduling”, *IEEE Transactions On Wireless Communications*, Vol. 4, No. 2, Mar. 2005., pp. 495-506.

[ZLGT10] Zaki,Y., Liang,Z., Goerg,C. and Timm-Giel,A., “LTE wireless virtualization and spectrum management", in *Proc. of WMNC'10 - Wireless and Mobile Networking Conference*, Budapest, Hungary, Oct. 2010.

[ZZRR08] Zhu,Y., Zhang-Shen,R., Rangarajan,S. and Rexford,J., "Cabernet: Connectivity Architecture for Better Network Services", in *Proc. of ReArch'08 - Re-Architecting the Internet ACM Workshop*, Madrid, Spain, Dec. 2008.