SAPHYRE

Contract No. FP7-ICT-248001

SAPHYRE Reference Scenario Parameters and Novel Interference Models (final)
D3.3b

Contractual date:	M24
Actual date:	M24
Authors:	David Gesbert, Zuleita Ka-Ming Ho, Jacek Kibiłda, Martin Haardt, Dominik Hamera, Eduard Jorswieck, Eleftherios Karipidis, Jan Sýkora, Jianhui Li, Radosław Piesiewicz, Jianshu Zhang
Participants:	TUD, CTU, ECM, LiU, TUIL, WRC
Work package:	WP3
Security:	Public
Nature:	Report
Version:	1.0
Number of pages:	63

Abstract

This document is the final version of our specification of relevant scenarios and performance metrics within SAPHYRE. Three main steps to perform the evaluation of the proposed sharing solutions have been identified: reference topologies definition, performance metrics definition and description of the sharing models. At first, physical layer sharing solutions have been generalized to a set of reference topologies. Next, the proposed performance metrics have been mapped onto system level Key Performance Indicators. Finally, an overview of passive and active infrastructure sharing models has been given.

Keywords

Backhaul sharing, gain evaluation, KPI, infrastructure sharing, performance metrics, QoS, relay sharing, scenarios, sharing models.

Executive Summary

The performance of the resource sharing schemes which are being developed within the SAPHYRE project has to be evaluated. Previously, the Deliverable D3.3a addressed this issue by providing an overview of the three main evaluation steps that we have identified: reference topologies definition, performance metrics definition, and description of the sharing models. In this document (D3.3b), we retain most of the initial document structure so that D3.3b can be read as a self-contained deliverable specifying our assumptions and scenarios. However, updates are provided in the final version which are explicitly listed in the next section.

In the first step the deliverable provides a detailed description of the four reference topologies, envisioned for SAPHYRE physical layer solutions. The first topology shows a typical interference channel model, which consists of nodes of different operators, sharing the same spectrum. The second topology enhances the scenario by collocating nodes, where for example joint coding, decoding or processing is possible. The last two topologies present a scenario which includes a relay node, thus enabling in the non-collocated case the introduction of various different signal forwarding strategies, like: amplify-and-forward, decode-and-forward, among others. The relay-collocated topology additionally allows for usage of joint signal processing techniques.

According to the second step of the proposed evaluation methodology, we propose a set of physical layer metrics, that evaluate the performance of the developed solutions. The representative metrics consist of single user rate, sum rate, outage probability, SINR or error rate. The metrics are being used within the corresponding solutions to quantitatively describe the SAPHYRE gain. Here, the SAPHYRE gain, according to the deliverable, can be understood as the system utility in the sharing scenario compared to the exclusive use of the spectrum and infrastructure by a single operator (typically TDMA). Alternatively, the fractional SAPHYRE gain is the ratio between the total utility received by users in the sharing scenario to the average utility received in single user scenarios. Furthermore we also provide a link between physical layer solutions and system level aspects such as Quality of Service. We achieve this through the definition of Key Performance Indicators that are used to quantify operators' Quality of Service levels. Binding the two aspects, we propose a mapping between the performance metrics and Key Performance Indicators. Additionally, we give a derivation of utility metrics to assess the users' gain in the case of multiple antenna systems.

The last section of the deliverable deals with the description of the infrastructure sharing models. There, we give a broad analysis of different aspects of the infrastruc-

ture sharing, starting from the top level division between active and passive sharing. Furthermore, we expand and detail the division to five new categories: passive RAN sharing (sharing involves only passive elements such as site, masts), passive RAN sharing with Access Transmission sharing (passive sharing with shared backhauling links), Active RAN sharing with MORAN (sharing of active resources with static virtual resource division), Active RAN sharing MOCN and GWCN (sharing of RAN active resources with dynamic resource assignment), and finally Roaming-based sharing (full sharing based on inter-operator agreements). Among the listed sharing paradigms, two elements of shared infrastructure are especially important, as they might become system bottlenecks. The two elements are backhaul link and relay node. Due to the availability of different technologies, we give a deep analysis of the technologies behind backhaul link sharing. We put special attention to the QoS provisioning problem, where we propose a solution based on IP QoS service model and LAN virtualization using Ethernet Virtual Circuits. The analysis of the relay node sharing presents initial assumptions on the interference relay channel and the applicable forwarding scheme, such as amplify-and-forward, which require the least signaling as well as minimum knowledge on the transmitted signal (modulation, coding).

Eventually, we conclude the deliverable providing a summary of research efforts undertaken as well as highlighting the next steps to realize a complete vision of performance measures for the resource sharing systems.

Contents

Abbreviations

3-GPP	3rd Generation Partnership Project
AAA	Authentication Authorization and Accounting
AF	Amplify and Forward
aGW	access Gateway
AMC	Adaptive Modulation and Coding
AUC	Authentication Center
BER	Bit Error Rate
BTS	Base Transceiver Station
CAPEX	Capital Expenditures
CBS	Commit Burst Size
CF	Compress and Forward
CIR	Commit Information Rate
CSI	Channel State Information
DF	Decode and Forward
DWDM	Dense Wavelength Division Multiplex
EBS	Excessive Burst Size
EIR	Excessive Information Rate
eNodeB	evolved NodeB
EPC	Evolved Packet Core
EVC	Ethernet Virtual Circuit
GPRS	General Packet Radio Service
GTP	GPRS Tunneling Protocol
HDF	Hierarchical Decode and Forward
IC	Interference Channel
ICIC	Inter-Cell Interference Coordination
IMT	International Mobile Telecommunications
IN	Interference Neutralization
IP	Internet Protocol
IPSec	Secured IP
IRC	Interference Relay Channel
ITU-T	International Telecommunications Union - Telecommunications
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LAN	Local Area Network
LMDS	Local Multipoint Distribution System
LTE	Long Term Evolution
MAC	Media Access Control

MEF	Metro Ethernet Forum
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MME	Mobility Management Entity
MPLS-TP	Multi Protocol Label Switching - Transport Profile
NAS	Non-Access Stratum
NGMN	Next Generation Mobile Networks Alliance
NMS	Network Management System
OAM	Operation, Administration and Maintenance
OPEX	Operational Expenditures
PBB-TE	Provider Backbone Bridging with Traffic Engineering
PLMN	Public Line Mobile Network
PHY	Physical Layer
PON	Passive Optical Networks
PWE3	PseudoWires
QCI	QoS Class Identifiers
QoS	Quality of Service
RAB	Radio Access Bearer
RAN	Radio Access Network
RED	Random Early Detection
RF	Radio Frequency
RTT	Round Trip Time
SCTP	Stream Control Transmission Protocol
SDH	Synchronous Digital Hierarchy
SGW	Serving Gateway
SIMO	Single Input Multiple Output
SINR	Signal to Interference plus Noise Ratio
SISO	Single Input Single Output
SLA	Service Level Agreement
SNR	Signal to Noise Ratio
TCO	Total Cost of Ownership
TCP	Transmission Control Protocol
TDM	Time Division Multiplex
TDMA	Time Division Multiple Access
UDP	User Datagram Protocol
VDSL	Very high bitrate Digital Subscriber Line
VLAN	Virtual LAN
VoIP	Voice over IP
WAN	Wide Area Network
WiMAX	Worldwide Interoperability for Microwave Access

1 Introduction

1.1 Update on Initial Deliverable

In the initial version of this document (D3.3a) we have introduced the basic reference scenarios and key performance metrics for the various resource sharing frameworks envisioned in SAPHYRE. The purpose of this new deliverable (D3.3b) is multi-fold. First, we recall the basic concepts related to the SAPHYRE vision and the definition of the Sharing Gain, which is the central differentiator of SAPHYRE and a common feature behind all contributions produced under this project, for all sharing types (spectrum sharing, relay sharing, backhaul sharing, etc.). Note that we preserve much of the initial document structure: reference topologies, performance metrics for resource sharing, finishing with infrastructure sharing models, such that this document can be read in a self contained way without the need to go back to the D3.3a version. Nevertheless, we propose a number of updates in the D3.3b version.

First, we establish the relation between the proposed classification of reference scenarios, and their actual use in the recent technical contributions of SAPHYRE. We also illustrate the concept of the SAPHYRE gain with an example drawn from the case of two operators sharing a relay. There, the gain is shown in terms of reduction of the required transmit power. When it comes to the models on infrastructure sharing, significant efforts were made with the aim to reflect recent evolution of wireless standards (notably 3GPP) in our understanding of sharing scenarios.

1.2 The SAPHYRE Vision and the Sharing Gain Concept

The vision of SAPHYRE is to demonstrate how the paradigm of exclusive resource allocation shifts towards cost, spectrum and energy efficient voluntary physical resource sharing which is realized through innovative use of radio spectrum and network infrastructure under economic and regulatory constraints. SAPHYRE realizes the vision by focusing on resource sharing aspects between wireless network operators, where by resource sharing one should understand passive or active pooling of available resources (spectrum, network elements, physical links or site hardware) for the joint purpose of cost savings, performance enhancement and overall greater efficiency. Thus, SAPHYRE intends to develop a number of techniques that enable introduction of resource sharing schemes that lead to increased utility. However, in order to ensure the feasibility of sharing schemes, in all objectives, SAPHYRE solutions must be evaluated with common measures, which allow to emphasize the sharing gain.

The sharing gain, namely SAPHYRE gain, can be defined as the performance comparison in terms of various metrics (e.g. system sum-rate, achievable rate region). Each of the proposed resource sharing schemes shall clearly highlight the gain arising from implementation of the sharing scheme, in question. The metrics, which ought to show the SAPHYRE gain shall also be used to provide the notion of fairness and cover Quality of Service aspects. The notion of fairness between users is normally not explicitly built into the scheduling criterion, since the operator only seeks to maximize a capacity metric under possible Quality of Service (QoS) constraints. Therefore a new approach towards fairness in resource sharing schemes should be developed. Regardless from fairness aspects in the inter-operator domain still QoS levels must be maintained, therefore a two way approach is necessary, where the QoS is divided into: guaranteed level (denotes minimum required service) and excessive level (denotes the maximum possible service, where the notion of fairness is applied). It is worth pointing out that QoS is especially important in the case of QoS-sensitive IP-based applications (e.g. VoIP), which are typically provided by 4G technologies and are envisioned also for future systems such as IMT-Advanced. On the business level, QoS is ensured through contracts, called SLAs (Service Level Agreements), which oblige mutually operators to obey the policy rules. Typically policy rules describe the amount of network resources required to realize QoS services [39]. The quantitative measures for policy rules are given by KPIs (Key Performance Indicators). KPIs are set up to evaluate the system as well as monitor the current progress of the solutions in respect to set goals. However, due to the general nature of KPIs, a mapping between link level and system level performance evaluation, KPIs need to be unambiguously tied with physical layer performance metrics. Proposed mapping, based on [3], will be presented in Section 3.3.

Apart from KPIs, each of the proposed performance metrics need to be applied also to the specific reference topology, which would enable comparison and classification of the developed resource sharing schemes. In Chapter 2, the key topologies for sharing are presented. The enhancement in the sharing context, exposes novel physical layer processing techniques related to the sharing of a relay node. From the functional point of view reference topologies are used to provide an insight into the assumptions of the particular solution, for example duplex mode, element collocation, available demodulation and coding scheme.

Apart from the performance metrics and reference topologies typically available in the methodology, also infrastructure sharing models need to be provided. The infrastructure sharing models in SAPHYRE, can be divided into passive and active, which then can be further divided according to different technical solutions based on the type of the network element or site equipment that is being shared [19]. This creates a multidimensional sharing problem, therefore a detailed analysis of different degrees of infrastructure sharing will be given in the document. Furthermore the

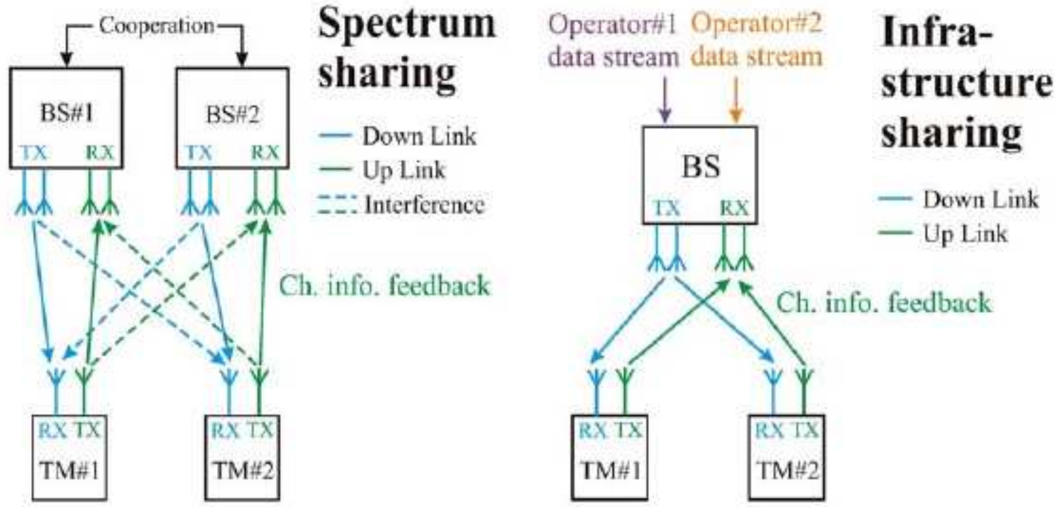


Figure 1.1: Two reference scenarios of SAPHYRE: a) spectrum sharing, b) infrastructure sharing.

document will provide a deeper insight into the problems related to backhaul link sharing as well as relay node sharing, which will further be used in deliverable D3.1b to aid the development of resource sharing schemes.

The aim of this deliverable is to provide a general understanding of the assumptions underlying the performance evaluation and potential benefits of the physical layer techniques developed within SAPHYRE project. Therefore, in Chapter 2, we start the document with unified reference topologies. In Chapter 3 we propose a set of performance metrics to describe SAPHYRE sharing solutions with enhancement of the metrics to system level and example detailed analysis of the utility achieved in multiple antenna scenario. Leading deeper with the discussions on sharing, in Chapter 4, we provide analysis of infrastructure sharing models with a closer look into the backhaul link and relay node sharing.

2 Description of Reference Topologies

2.1 Motivation

Each WP, layer or viewpoint perspective (technical, business, etc.) requires a different level of abstraction when defining what is called by the term “scenario”. Different aspects of the “scenario” are defined here for classifying solutions in different contexts.

2.2 Scenario Classification

Here, a multi-dimensional “scenario coordinate map” is proposed.

• Each coordinate (dimension) introduces the classification relevant to a given layer/viewpoint. This creates the required flexibility and avoids confusion around the term scenario later on.

• The scenario description is intentionally kept in a very generic form and defines only fundamental characteristics. All finer details (e.g. various quantitative parameters, finely defined subclasses) should be only specified as a parameter of that scenario.

• Note that we use the term “scenario” to express a classification of business related contexts. Technically-oriented WPs, such as WP2 and WP3 for instance, will use the term e.g. “topology” in order to mark the difference.

2.3 Scenario Coordinates

2.3.1 Scenario (S) Coordinate Classes

The Scenario (S) coordinate describes operator’s viewpoint. This is the first (top level) coordinate.

Values:

• S1 = shared RAN & shared spectrum

• S2 = shared RAN only

• S3 = shared spectrum only

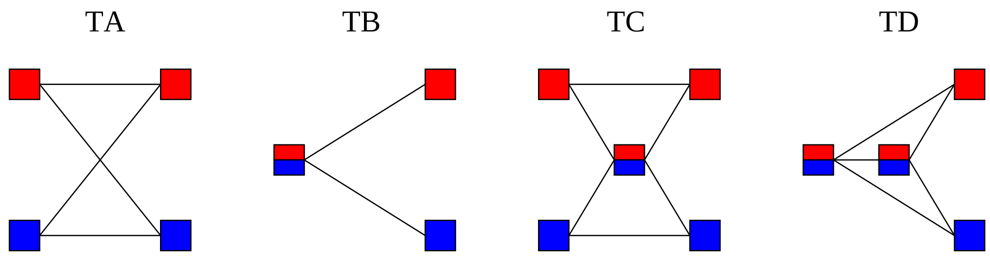


Figure 2.1: Topology (T) coordinate classes.

Notes:

- RAN sharing refers to sharing whatever type of HW (including e.g. the relay)

2.3.2 Topology (T) Coordinate Classes

The Topology (T) coordinate describes characteristics important mainly from the perspective of communication system design and algorithms. In particular, it indicates:

- The presence or absence of relays.
- The physical collocation of base stations belonging to different operators.

The following key topologies are considered in SAPHYRE (see Fig. 2.1):

- TA = no relay, no base station collocation, spectrum sharing (interference channel)
- TB = no relay, base station collocation (with perhaps joint modulation/coding, joint demodulation/decoding at terminal)
- TC = relay present and shared, no base station collocation (all possible relaying strategies AF, Joint DF, CF, HDF, etc)
- TD = shared relay, share base station (all relaying strategies with join modulation/coding, joint demodulation/decoding at terminal)

Notes:

- Presence of various link types (bidirectional/unidirectional, presence of direct link, presence of side-information link, etc.) is only reflected through the parameters of the topology. The same holds for all other attributes like SNR, channel type, synchronization assumptions, etc.

2.3.3 Other Coordinate Classes

A number of other classes, not relevant to WP3, are defined. Their particular description is in the respective WPs. Namely, there are

- Business (B) coordinate classes
- Regulatory (R) coordinate classes

2.4 Use of Reference Topologies in SAPHYRE Contributions

We have above proposed a flexible model of reference scenarios envisioned for SAPHYRE physical layer solutions. This model of reference scenarios covers different perspectives required from either non-technical (e.g. business, regulatory, etc.) or technical viewpoint. We have identified key independent parameters (scenario-coordinates) of the reference model: *a) degree of sharing (S)* (S1: sharing RAN and spectrum, S2: sharing RAN only, S3: sharing spectrum only) and *b) spatial topology (T)* which determines configuration in the space including presence of relays and physical co-location of sources (base stations). The four key proposed reference topologies are TA: interference channel, TB: interference channel with co-located sources, TC: interference relay channel and TD: interference relay channel with co-located sources. Related to the above mentioned topologies are additional parameters further describing model assumptions in more detail such as: channel models, link SNR, signal processing capability of nodes, presence of line-of-side/side-information links, synchronization assumptions, etc. This flexible structure of reference topologies fully covers all cases needed for SAPHYRE resource sharing approaches. Therefore, we do not need to extend or update the initially proposed model.

Especial importance possess following main topologies: interference channel with shared spectrum (TA+S3) and interference relay channel with shared relay (TC+S1 or S2), since TB and TD strongly correspond (via co-located sources) to TA and TC. These topologies have appeared in the following deliverables: deliverable [SAPHYRE D2.1a] considers: 1) an interference channel with MISO IC and two operators sharing the spectrum assuming single/multiple user decoding capabilities which are covered by TA+S3 scenario and 2) a 2-source relay network with presence of complementary-side information links (also denoted as "butterfly" network) which are covered by TC+S2. Deliverable [SAPHYRE D2.2a] considers TA+S3 scenario as MISO IC further distinguishing between non-cooperative and cooperative case and assuming priority users. Deliverable [SAPHYRE D2.3a] assumes MIMO IC as TA+S3 topology and "butterfly" network and multi-operator two-way relay channel as TC+S2 topology. In deliverable [SAPHYRE D3.1a] we have employed K-user MISO IC and MIMO IC as TA+S3 and TC+S2 topology which occurs also in in deliverable [SAPHYRE D3.2a]. The reference topologies are naturally discussed in deliverable [SAPHYRE D5.1a,b].

3 Performance Metrics for Resource Sharing Schemes

3.1 SAPHYRE Gain

As mentioned in SAPHYRE deliverable D3.1a, in order to evaluate the benefits of novel signal processing algorithms to exploit the additional degrees of freedom brought by sharing in multi-user and multi-cellular environments, it is important to:

- define a performance metric,
- show the gain (loss) with respect to the chosen performance metric as compared to a non-sharing scenario,
- point out conditions when a significant gain can be achieved for the chosen scenario (topology), and
- illustrate the order of magnitude of this gain.

We denote this sharing gain as the SAPHYRE gain. Formally it can be defined as the performance comparison in terms of various performance metrics (e.g., the system sum-rate, the achievable rate region, etc.). In this deliverable, we define two types of SAPHYRE gain in terms of system utility function of the sharing scenario compared to the exclusive use of the spectrum and infrastructure by a single operator (in this case, the users are served via TDMA). The absolute SAPHYRE gain is defined as

$$\Xi_{\mathrm{A}} = \sum_{k=1}^{K} U_k - \frac{1}{K} \sum_{k=1}^{K} U_k^{\mathrm{SU}}, \tag{3.1}$$

and the fractional SAPHYRE gain is defined as

$$\Xi_{\mathrm{F}} = \frac{\sum_{k=1}^{K} U_k}{\frac{1}{K} \sum_{k=1}^{K} U_k^{\mathrm{SU}}}, \tag{3.2}$$

where $k \in \{1, 2, \cdots, K\}$ is the index of the users. The utility function of the kth user in the sharing scenario and the time division case are denoted by U_k and U_k^{SU}, respectively, where SU stands for Single-User.

3.1.1 SAPHYRE Gain in Power

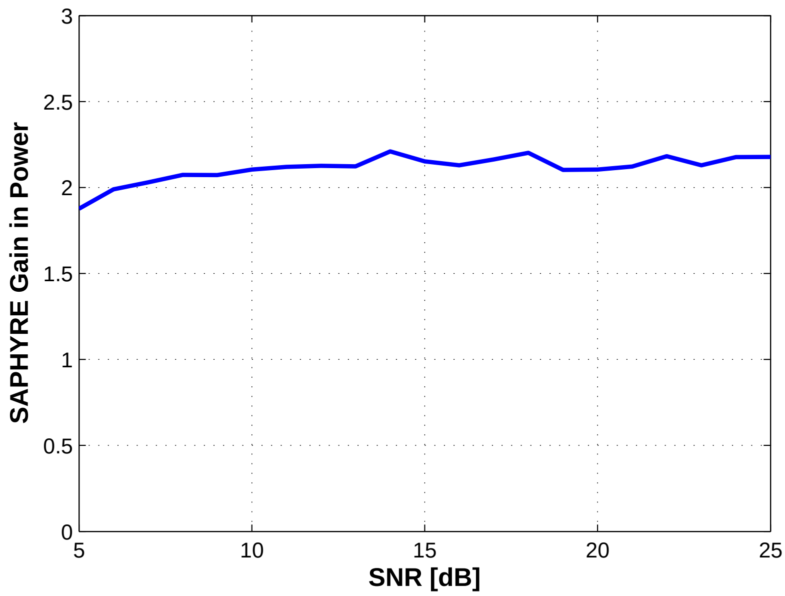


Figure 3.1: SAPHYRE gain in terms of power obtained using EReSh-PM [22] for topology TC with $K = 3$, $M_{\mathrm{R}} = 8$

In addition to the SAPHYRE gain defined in the previous section, we can also interpret the SAPHYRE sharing gain in terms of the consumed transmit power. That is, the transmit power consumed in the sharing scenario is compared to the corresponding one when there is exclusive use of the spectrum and infrastructure by a single operator (TDMA access). The fractional SAPHYRE gain in terms of transmit power is defined as

$$\Xi_{\mathrm{F}} = \frac{\frac{1}{K}\sum_{k=1}^{K} P_k^{\mathrm{SU}}}{\sum_{k=1}^{K} P_k}. \tag{3.3}$$

where the numerator denotes the average required transmit power for achieving certain QoS metrics (e.g., minimum required total data rate of the network, minimum required SNR per user) in the non-sharing case and the factor $\frac{1}{K}$ is due to the use of K resources. The denominator denotes the required transmit power for achieving the same performance metrics in the sharing case.

In this section, we give an example to illustrate the SAPHYRE gain with respect to the power. The scheme considered here is the multiple-operator one-way amplify-and-forward (AF) relaying, which belongs to the topology TC defined in Chapter 2, where multiple base stations communicate pairwise with their target users via a shared AF relay and the direct link between the base stations and users is so weak that can be neglected. As an example, we consider the case of $K = 3$ users, each served by a different operator. Both the base stations and the user terminals are equipped with single antennas while relay employs $M_{\mathrm{R}} = 8$ antennas. For this SAPHYRE scenario, the relay is designed using the algorithm named efficient resource sharing power minimization (EReSh-PM) [22] in order to minimize the relay transmit power while the SINR constraint at each user has to be satisfied. To obtain the SAPHYRE gain, we define a TDMA scenario as a benchmark, where each operator accesses the relay and the spectrum in a round-robin manner. For each operator, the optimum relay amplification matrix is designed to minimize the relay transmit power subject to the SINR constraint at each user. The SAPHYRE gain in terms of relay transmit power is plotted in Fig. 3.1, which is calculated according to equation (3.3). As seen from Fig 3.1, the SAPHYRE gain is significant; more than double power is consumed observed with orthogonal use of the resources than when the spectrum and infrastructure (relay) are shared.

3.2 Definitions of Performance Metrics for Resource Sharing Schemes

3.2.1 Definition of Physical Layer Metrics

SAPHYRE results can only be assessed with use of appropriate performance metrics. The assessment is performed by means of QoS realization, which in highly loaded packet-switched networks needs to be preserved by all means [12]. The following physical layer metrics influence the resource allocation process (performed in the Radio Resource Management layer), in a way that they reflect the possible QoS levels that can be provided to a single user for a specific service.

In order to properly assess the results of SAPHYRE WP2 and WP3 solutions we need to define the set of performance (utility) metrics for resource sharing schemes proposed in the physical layer:

- **Sum-rate** - maximum system achievable throughput, regardless of fairness, in resource allocation between the users. The maximum system achievable throughput is in fact the maximum sum of the rates taken over all rate vectors in the rate region [20]:

$$C_{SR} = \max_{(R_1,..,R_k)\in C} \sum_{k=1}^{K} R_k \tag{3.4}$$

- **Quality-of-Service related metrics**:

- *Single-user rate* (in bits/s) - the user rate can have two values, the first is the guaranteed rate and the second is the maximum theoretical rate. The guaranteed value represents a rate that can be offered and supported to a single user of one operator/technology in case of spectrum/infrastructure sharing scenarios, typically when operator/technology owns an exclusive band or its traffic has higher priority over others (this rate will not be violated by unfavorable sharing conditions - high load of other operator(s) traffic). The maximum rate represents the theoretical upper bound that can be achieved by a single user in a spectrum/infrastructure sharing scenario. The maximum rate cannot be guaranteed in any case and it is dependent on the current load situation. The relation between the values is described by inequality: *Guaranteed user rate* $\leq$ *Maximum user rate*
- *Outage probability* - the probability that the target bit error rate performance of the users can not be met, that is due to the fact that the power at the receiver is below the minimum reception threshold (which can be seen as outage in transmission). The metric can be used to represent the combination of transmission/reception methods (MIMO, MISO), transmission schemes (Network Coding), level of noise and interference on the reception of the signal. Additionally it can be also used to evaluate the system deployment of relay nodes in the scenarios with relays. The typical models for outage probability under combined path loss and shadowing, and fading channel are given below [20]:

$$\begin{aligned} p_{out}(P_{min}, d) &= p(P_r(d) \leq P_{min}) \\ P_{out}(\gamma \leq \gamma_0) &= \int_0^{\gamma_0} p_{\gamma_s}(\gamma) d\gamma \end{aligned} \tag{3.5}$$

  where P_{min} denotes minimum required power, $P_r(d)$ received power at given distance from the transmitter (which is log-normally distributed), γ_s received SNR, which is a random variable with distribution $p_{\gamma_s}(\gamma)$, γ_0 specifies the minimum SNR required for acceptable performance.
- *Service latency/delay* - represents the time needed to transmit the information over the radio interface between the transmitter and receiver (sometimes can be expressed also as in term of RTT), factors which influence latency: channel type, number and complexity of processing operations (both at transmitter and receiver), size of the information burst, etc. The latency constraint can be used to describe the feasibility of the scenario as in respect to specific services offered by operators, e.g. low latency is crucial for operators providing real-time applications such as video streaming, voice. In the most general form can be expressed as combination of various delays (measured in the units of seconds):

$$Latency = transmit\_time + propagation\_delay + processing\_time \tag{3.6}$$

- *SINR* (Signal-to-Interference-plus-Noise-Ratio) - describes channel quality (especially crucial in multiuser systems where it characterizes interference environment) and therefore directly influences the rate achievable by each user. The idea is to keep it as constant as possible so to maintain channel quality for the user and provide stable QoS value. The metric can be used as a good descriptor of the interference level (important factor in spectrum sharing) produced in the system and QoS provisioning, e.g. if SINR fluctuates severely it is unlikely to provide high QoS to the users. The most general model for SINR at the receiver (measured in dB) is given by:

$$SINR = \frac{P}{N_0 + I} \tag{3.7}$$

  where P denotes average received power of the signal, N_0 average received noise power and I average received interference power.
- *Bit Error Rate (BER)* - Empirical metric defined as the number of erroneous bits to the total number of received bits, which can be used to represent the reliability of the received data information. Factors which influence the error rate: noise, interference, distortion, fading, etc. Generally, the BER can be improved by providing better SINR (stronger transmitted signal power or lower interference or noise leves), lower modulation schemes, etc. The metric can be used to describe the reliability of the transmission in respect to the modulation and coding schemes proposed, if the schemes are less prone to interference, then the BER shall be smaller.

$$BER = \frac{Number\_of\_bit\_errors}{Total\_number\_of\_bits\_transmitted} \tag{3.8}$$

- **Amount of additional side-information**:
  - *exchanged outside PHY* (inter/intra operator) - depending on the sharing scenario, operators (or technologies in the intra-operator scenario) may need to exchange information regarding the interference conditions, used (free) time slots, power allocations, CSI (Channel State Information). The metric describes the signaling overhead required to distribute the information, as well as additional interfaces and network components.

- *required inside PHY* - typically Side Information is the amount of information on other transmitted information streams (send to other destinations) in multi-user system, that is used to decode the own information stream. It can be split into either full or partial side information knowledge, depending on the coding technique used. The side information can either be sensed from the environment or it can be supplied via the pilot channel (see also cognitive pilot channel) or cable link. The metric can be used to represent limitations to the planning of the network and additional links that need to be taken into account when deploying the scenario.
- *overhead penalization* (total rate reduction) - used as a weighting factor the utility function, the metric describes single user rate/sum-rate reduction due to additional overhead related with the exchange of side information such as supplying the CSI to the transmitter.

3.3 Quality of Service Aspects in SAPHYRE Solutions

Typically operators when deploying new networks seek to maximize their available capacity under specific QoS constraints. In particular QoS is an important measure that allows the operators to differentiate from each other [24]. It is important to consider it when designing different sharing schemes for SAPHYRE. In principle, resource sharing solutions for cellular networks shall provide QoS awareness. This imposes the same requirement on the design of performance metrics which will evaluate the sharing solutions, as in the ideal case they should also show the compromise between sharing and individual QoS achievements.

3.3.1 Mapping of Performance Metrics on System Level KPIs

Key Performance Indicators (KPIs) are primary metrics to define the success rate of an enterprise. In principle they are set up to evaluate the system as well as monitor the current progress of the solutions in respect to set goals. According to 3GPP, KPIs describe strategic goals of the enterprise, and cascade through the entire organization. KPIs are specified through definition and measurement of key parameters of input/output of internal network system and/or maintenance & operation progress of an enterprise [10]. The term strictly connected with KPIs is Service Level Agreement (SLA), which can be seen as a contract which describes common understanding of the service as well responsibilities between parties involved and performance objectives. In order to provide information on how the SLA agreements are realized, KPIs (indication of service resource performance) and Key Quality Indicators (service element performance) are measured and compared towards objective targets included in the SLA. KPIs are proved by aggregation of network performance data from network elements [10].

3GPP has proposed such a classification of KPIs for 2G and 3G systems, with definitions from ITU-T recommendation [3]:

- *Serveability* - The ability of a service to be obtained (with specific tolerance and other conditions) when requested by the user and continue to be provided without excessive impairment for a requested duration.
  - *Accessibility* - The ability of a service to be obtained (with specific tolerances and other given conditions) when requested by the user.
  - *Retainability* - The ability of a service (once obtained) to continue to be provided under given conditions for a requested duration.
  - *Integrity* - The degree to which a service (once obtained) is provided without excessive impairments.
- *Availability* - The ability of an item to be in a state to perform a required function at a given instant of time or at any instant of time within a given time interval, assuming that the external resources, if required, are provided.
  - *Reliability* - The ability of an item to perform a required function under given conditions for a given time interval.
  - *Maintainability* - The ability of an item under stated conditions of use, to be retained in or restored to, a state in which it can perform a required function, when maintenance is performed under given conditions and using stated procedures and resources.
  - *Utilization* - Indication of the network resource utilization, such as throughput on specific interface.
  - *Mobility* - The description of abilities to perform handovers.

SAPHYRE Key Performance Indicators

Based on the 3GPP proposed KPIs and their classification, SAPHYRE uses its own set of KPIs to describe the SAPHYRE gain. The proposed KPIs are used for business level modeling as well as for scenario benchmarking:

- *KPI*:
  - *Accessibility* - see the definition from [3].
  - *Retainability* - see the definition from [3].
- *KQI*:
  - *Capacity* - The resources that can be provided to perform the service, the resources can be represented as link capacity, number of users, etc.
  - *Coverage* - The area, which is serviced by the proposed technical solutions.
  - *Latency* - The introduced time constraints to the performed service.

– *Fairness* - The distribution of the user rates around the mean. The fairness can be measured by the Jain's index which characterizes the variance of the rates with respect to the mean computed across all users.

What is also an important issue is the impact of PHY features on the CAPEX (Capital Expenditure) and OPEX (Operational Expenditure):

- *Implementation cost/profit* - The cost/profit generated from the introduction of specific solution. This cost/profit occurs one time only at the installation or implementation phase.
- *Operational cost/profit* - The cost/profit generated periodically (e.g. each month) due to the availability/unavailability of specific solution. This cost/profit occurs due to rental fees, maintenance fees, person months, etc.

Performance metric	KPI/KQI				
	Accessibility	Retainability	Capacity	Coverage	Latency
Sum-rate	X	-	X	-	-
Single user rate	X	-	X	-	-
Outage probability	-	X	X	X	-
Latency/delay	-	X	-	-	X
SINR	X	X	X	-	X
Error-rate	-	X	-	X	X
Inf exchanged in PHY	X	-	X	-	X
Inf exchanged out PHY	X	-	X	-	-
Overhead penalization	X	-	X	-	-

Table 3.1: Map of PHY metrics on to system level Performance and Quality Indicators.

3.3.2 Examples of QoS Provisioning Agreements in Resource Sharing Scenarios

QoS and SLA in Shared Backhaul. The shared backhaul link transports multiple operator flows, which can be further divided into flows corresponding to different types of services. It is thus obvious that a QoS delivery shall be maintained on two levels: per operator flow and per service flow. All the constraints regarding the minimum and excessive parameters available for each of the flows are described in Service Level Agreements. The most popular solution that support provisioning of SLAs at the transport links is to use IP QoS service model [13]. The IP QoS model (DiffServ model[1]) has been standardized to provide a variety of quality classes for various services traversing an IP network. The standard defines different service

[1]Differentiated Service model

classes, QoS provisioning mechanisms along with architecture that can be applied to different network elements. 3GPP has also recognized this method as a standardized solution for traffic classification for IP Radio Access Bearer services [9]. The mapping between 3GPP QoS classes and DiffServ code points shall be defined by the operators. The packet classification is used by aggregation nodes to perform link scheduling (in fact rate limitation and packet marking), with respect to predefined thresholds derived from SLA [6]:

- Committed Information Rate (CIR) - guaranteed minimum throughput, this value is dedicated (it can not overlap with the throughput of other flows/operators) to specific flow. In principle the highest the priority of the flow, the highest the committed value.

- Excessive Information Rate (EIR) - an excessive throughput, which can be defined as maximum link rate that the service can be assigned. High value of EIR is typically assigned to “best effort” services. The excessive throughput is taken from the common pool of resource for all the flows/operators. In case of leased lines, EIR is usually connected with additional fees.

- Committed Burst Size (CBS) - maximum size of Ethernet frame burst expressed in bytes with guarantees on performance.

- Excessive Burst Size (EBS) - maximum size of Ethernet frame burst with no guarantees on performance.

All the above mentioned mechanisms and limitations constitute for backhaul QoS support tools, which are used at aggregation nodes to actually serve the desired rates of traffic.

3.4 Performance Metrics in SAPHYRE Topologies/Solutions

In a SISO Interference Relay Channel (IRC), we denote the sources as S_i and destinations D_i, $i = 1, \ldots, K$. An example of two-sources two destination IRC is illustrated in Fig. 3.2. The multi-antenna relay node is denoted as R. Denote the complex channel from S_i to D_j as h_{ji} and the complex channel vector from S_i to R as $\mathbf{g}_{ri}$ and from R to D_j as $\mathbf{g}_{jr}$. All channels are assumed to be independent identically distributed complex Gaussian variables, $\mathbf{g}_{ri}, \mathbf{g}_{jr} \in \mathbb{C}^{M\times 1}$, where M is the number of antennas at the relay. We assume linear processing at the source and destination nodes: no multi-user user encoding and decoding possible.

In the following, we introduce different relay channel models, the corresponding constraints and assumptions, namely half-duplex relays, full-duplex relays, non-potent relays, instantaneous relays and causal relays.

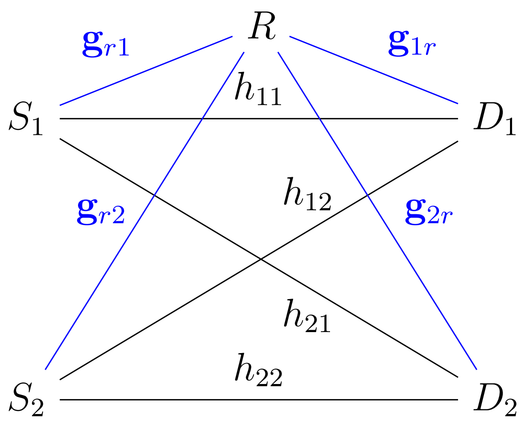


Figure 3.2: The channel model of a two sources two destinations SISO interference relay channel with a multi-antenna relay.

3.4.1 Relay Topologies and Performance Metrics

In this section, we summarize different relay topologies, the corresponding constraints, assumptions on the system and the resulting performance metrics.

- **A half-duplex relay:** the transmission of signals is considered to last for a duration of two time slots. In the first time slot (also known as the *first hop* and *broadcast phase* in the networking area), the signals travel from the source nodes to the relay node whereas in the second time slot (*second hop* or *multiple access phase*), the signals travel from the relay node to the destination node.
- **A full-duplex relay:** the relay is able to transmit and receive at the same time slot. This means that in the first time slot, the destination nodes receive signals from both the relay and source nodes.
- **A potent relay:** is a relay node that has access of power much larger than the remaining nodes in the network, e.g. a base station. A non-potent relay refers to a relay that has limited supply of power which is a much more practical assumption.
- **An instantaneous relay and a causal relay:** refer to the assumption of the knowledge of the received signal at relay: an instantaneous relay has no memory and its forwarding message only consists of what it just received; a causal relay has memory but no information about the receiver signal in the future and thus it forwards function of signals that consist of information from the beginning of the frame till present time.

In the following, we give some examples of performance metrics with a combination of the aforementioned assumptions. In particular, we give the signal model and SINR metric in a full-duplex instantaneous non-potent relay system.

Instantaneous full-duplex non-potent relay

We assume linear processing at the relay and the linear processing matrix is given by $\mathbf{R} \in \mathbb{C}^{M\times M}$. Also, we assume that the relay employs an amplify-and-forward strategy and the forwarding message consists of what it has just received. Note that, the relay is assumed to know all CSI in the system, but *not* the payload. This is because the relay only amplifies and forwards the messages and therefore has no reason to know the payload or codebooks beforehand. As the CSI for each channel state is required at the relay for interference management purposes, a fast exchange of CSI is required at both the relay and destination nodes. The source nodes, having only single antenna, do not have any information about the channel and transmit the data directly. The destination nodes are assumed to have local channel state information - all channels reaching themselves- and use this information for decoding the desired message. The signal received at R is given by:

$$\mathbf{y}_r = \sum_{t=1}^{K} \mathbf{g}_{rt}\, x_t + \mathbf{n}_r \tag{3.9}$$

where x_i are the transmit symbols from S_i which is assumed to be proper and has power constraint P, $\mathcal{E}|x_i|^2 = P_i \leq P, i = 1, \ldots, K$. The noise at the relay is denoted as n_r which is assumed to be zero mean unit variance white noise. The received signals at destination $D_j, j = 1, \ldots, K$, is

$$y_j = \sum_{l=1}^{K} \left(h_{jl} + \mathbf{g}_{jr}^H \mathbf{R}\, \mathbf{g}_{rl}\right) x_l + \mathbf{g}_{jr}^H \mathbf{R} \mathbf{n}_r + n_j \tag{3.10}$$

For brevity, denote $\mathbf{p} = [P_1, \ldots, P_K]^T \in \mathbb{R}_+^{K\times 1}$. The Signal-to-Noise ratio at destination j is

$$\mathrm{SINR}_j(\mathbf{R}, \mathbf{p}) = \frac{|h_{jj} + \mathbf{g}_{jr}^H \mathbf{R}\, \mathbf{g}_{rj}\,|^2 P_j}{\sum_{l=1, l\neq j}^{K} |h_{jl} + \mathbf{g}_{jr}^H \mathbf{R}\, \mathbf{g}_{rl}\,|^2 P_l + \|\, \mathbf{g}_{jr}^H \mathbf{R}\|^2 + 1} \tag{3.11}$$

where $\|\, \mathbf{g}_{jr}^H \mathbf{R}\|^2$ is the amplified noise from relay to destination j. The power constraint at the relay is

$$\mathbb{E}_{\mathbf{y}_r}(\mathrm{tr}\left(\mathbf{R}\mathbf{y}_r\mathbf{y}_r^H\mathbf{R}^H\right)) \leq P_r. \tag{3.12}$$

Assume that the transmit signals of the sources are proper and white noise at relay, we have $\mathbb{E}_{x_t,n_r}\left\{\mathbf{y}_r\mathbf{y}_r^H\right\} = \sum_{l=1}^{K} \mathbf{g}_{rl}\, \mathbf{g}_{rl}^H\, P_l + \mathbf{I}$. The power constraint is therefore rewritten as the following:

$$\mathrm{tr}\left(\mathbf{R}\left(\sum_{l=1}^{K} \mathbf{g}_{rl}\, \mathbf{g}_{rl}^H\, P_l + \mathbf{I}\right)\mathbf{R}^H\right) \leq P_r. \tag{3.13}$$

In the following, we introduce the idea of interference neutralization (IN) and the SINR which is achievable in this case.

Interference neutralization and its impact on performance metrics Interference neutralization is a novel technique in which the relay strategy, in this case an AF matrix, is chosen carefully such that the interference signals at each receiver are canceled out *in the air*. In order to neutralize interference, the following $K(K-1)$ equations have to be satisfied at the same time:

$$h_{dt} + \mathbf{g}_{dr}^H \mathbf{R}\, \mathbf{g}_{rt} = 0, \qquad d, t = 1, \ldots, K, d \neq t. \tag{3.14}$$

If IN is feasible, we can choose a relay matrix $\mathbf{R}$ that achieves the following SINR,

$$\mathrm{SINR}_j^{\mathrm{IN}}(\mathbf{R}, P_j) = \frac{|h_{jj} + \mathbf{g}_{jr}^H \mathbf{R}\, \mathbf{g}_{rj}\,|^2 P_j}{\|\, \mathbf{g}_{jr}^H \mathbf{R}\|^2 + 1} \tag{3.15}$$

The advantages and disadvantages of IN can be summarized as follows:

- **Advantages**:
  - The implementation of IN requires *only* relay processing. The source and destination nodes are not involved or even unaware of the existence of the relay and the processing within. This increases the practicality of the deployment as the system performance can be dramatically improved by simply introducing the interface between the intelligent relay and the simple source and destination nodes.
  - If IN is feasible, each interference signal at each receiver is completely canceled out *in the air*. Despite the system operating as the maximum degrees of freedom, the source and destination nodes enjoy this increment of rate without providing any extra processing efforts and therefore preserving their battery life.
- **Disadvantages**:
  - IN is not always feasible. With a potent relay, a limited power available to the relay, the fading nature of the wireless channel may decide on strong interference channels among the source and destination nodes and therefore making cancellation of the interference signals difficult. When the relay does not have enough power for IN, the interference signals are not completely canceled out and the system then operates on suboptimal degrees of freedom regime.
  - IN requires the relay to know all channel state information in the system in order to exactly cancel out each signal arriving at each receiver.
  - Depending on the channel parameters, IN may not be the sum rate optimal strategy. An analogy to this is the zero-forcing beamforming strategy. When the system is noise limited, complete cancellation of interference signals is a suboptimal strategy.

3.5 Utility Metrics Based on Multi-Antenna Channel Gains

We consider T transmitters and K receivers sharing the same spectral band. Define the set of transmitters as $\mathcal{T} := \{1, ..., T\}$ and receivers as $\mathcal{K} := \{1, ..., K\}$. Each transmitter sends useful information to at least one receiver. For transmitter $k, k \in \mathcal{T}$, let $\overline{\mathcal{K}}(k) \subseteq \mathcal{K}$ denote the set of its intended receivers for which useful information is sent to, and let $\underline{\mathcal{K}}(k) = \mathcal{K} \backslash \overline{\mathcal{K}}(k)$ be the set of its unintended receivers. Each transmitter k is equipped with N_k antennas, and each receiver with a single antenna. The quasi-static block flat-fading instantaneous channel vector from transmitter $k, k \in \mathcal{T}$, to receiver $\ell, \ell \in \mathcal{K}$, is denoted by $\boldsymbol{h}_{k\ell} \in \mathbb{C}^{N_k \times 1}$. The transmit covariance matrix of transmitter k is given as $\boldsymbol{Q}_k \in \mathbb{C}^{N_k \times N_k}$, $\boldsymbol{Q}_k \succeq 0$. We do not make any assumptions on the number of data streams applied at the transmitters. The basic model for the matched-filtered, symbol-sampled complex baseband data received at receiver ℓ is

$$y_\ell = \sum_{k=1}^{T} \boldsymbol{h}_{k\ell}^H \boldsymbol{Q}_k^{\frac{1}{2}} \boldsymbol{s}_k + n_\ell, \tag{3.16}$$

where $\boldsymbol{s}_k$ is the symbols vector transmitted by transmitter k and n_ℓ are the noise terms which we model as independent and identically distributed (i.i.d.) complex Gaussian with zero mean and variance σ^2. Each transmitter has a total power constraint of $P := 1$ which leads to the constraint $\operatorname{tr}(\boldsymbol{Q}_k) \leq 1$, $k \in \mathcal{T}$. Throughout, we define the signal to noise ratio (SNR) as $1/\sigma^2$. The feasible set of covariance matrices for transmitter k is defined as

$$\mathcal{S}_k := \left\{ \boldsymbol{Q}_k \in \mathbb{C}^{N_k \times N_k} : \boldsymbol{Q}_k \succeq 0, \operatorname{tr}(\boldsymbol{Q}_k) \leq 1 \right\}. \tag{3.17}$$

The performance measure of a system in an interference network is usually described by a utility function. The utility function associated with a receiver depends on the power gains originating from the transmitters in the network. Define the power gain achieved by transmitter k at a receiver ℓ as

$$x_{k,\ell}(\boldsymbol{Q}_k) = \boldsymbol{h}_{k\ell}^H \boldsymbol{Q}_{\boldsymbol{k}} \boldsymbol{h}_{k\ell}, \tag{3.18}$$

where $x_\ell(\boldsymbol{Q}_k) \in \mathbb{R}_+$ since $\boldsymbol{Q}_k$ is positive semidefinite. The utility function associated with a receiver ℓ is defined as $u_\ell : \mathbb{R}_+^T \to \mathbb{R}_+$, where T is the number of transmitters in the network.

3.1 Assumption. The utility function $u_\ell, \ell \in \mathcal{K}$, has the following properties:

A. If $\ell \in \overline{\mathcal{K}}(k)$, then u_ℓ is monotonically increasing in the power gain from transmitter k, i.e.,

$$u_\ell(x_{1,\ell}(\boldsymbol{Q}_1), ..., x_{T,\ell}(\boldsymbol{Q}_T)) \leq u_\ell\Big(x_{1,\ell}(\boldsymbol{Q}_1), ..., x_{k,\ell}(\widehat{\boldsymbol{Q}}_k), ..., x_{T,\ell}(\boldsymbol{Q}_T)\Big), \tag{3.19}$$

for $x_{k,\ell}(\boldsymbol{Q}_1) \leq x_{k,\ell}(\widehat{\boldsymbol{Q}}_k)$.

B. If $\ell \in \underline{\mathcal{K}}(k)$, then u_ℓ is monotonically decreasing in the power gain from transmitter k, i.e.,

$$u_\ell(x_{1,\ell}(\boldsymbol{Q}_1), ..., x_{T,\ell}(\boldsymbol{Q}_T)) \geq u_\ell\Big(x_{1,\ell}(\boldsymbol{Q}_1), ..., x_{k,\ell}(\widehat{\boldsymbol{Q}}_k), ..., x_{T,\ell}(\boldsymbol{Q}_T)\Big), \quad (3.20)$$

for $x_{k,\ell}(\boldsymbol{Q}_k) \leq x_{k,\ell}(\widehat{\boldsymbol{Q}}_k)$. □

Assumption 3.1 describes the settings where the performance measure at a receiver increases monotonically with increased power gain from intended transmitters and decreases monotonically with increased power gain from unintended transmitters. An example utility function which satisfies Assumption 3.1 is the signal to interference plus noise ratio (SINR).

The *utility region* is the set of all achievable utility tuples defined as:

$$\begin{aligned}\mathcal{U} \;:=\; &\{(u_1(x_{1,1}(\boldsymbol{Q}_1), \dots, x_{T,1}(\boldsymbol{Q}_T)), \dots, u_K(x_{1,K}(\boldsymbol{Q}_1), \dots, x_{T,K}(\boldsymbol{Q}_T))) : \\ &\boldsymbol{Q}_k \in \mathcal{S}_k, k \in \mathcal{T}\} \subset \mathbb{R}_+^K. \end{aligned} \quad (3.21)$$

The efficient operating points in the utility region correspond to those in which it is impossible to improve the performance of one system without simultaneously degrading the performance of at least one other system. Such operating points are called Pareto optimal and are defined formally as follows.

3.2 Definition. A tuple $(u_1, ..., u_K) \in \mathcal{U}$ is Pareto optimal if there is no other tuple $(u'_1, ..., u'_K) \in \mathcal{U}$ such that $(u'_1, ..., u'_K) \geq (u_1, ..., u_K)$, where the inequality is component-wise and strict for at least one component. The set of all Pareto optimal operating points constitutes the *Pareto boundary* ($\mathcal{PB}$) of $\mathcal{U}$. □

Next, we give an example setting where the systems' utility functions satisfy Assumption 3.1.

Example Setting

Consider two transmitters each using three transmit antennas, and three single antenna receivers as depicted in Figure 3.3. The operation of the systems is as follows:

- Broadcast Channel (BC): Transmitter 1 transmits different useful data to receivers 1 and 2 simultaneously. We assume transmitter 1 chooses the transmit covariance matrices $\boldsymbol{Q}_{11}$ with $\mathrm{tr}(\boldsymbol{Q}_{11}) = p_{11}$ for receiver 1 and $\boldsymbol{Q}_{12}$ with $\mathrm{tr}(\boldsymbol{Q}_{12}) = p_{12}$ for receiver 2. Hence, transmitter 1 can be considered as two virtual transmitters, 11 and 12, coupled by the total power constraint, $p_{11} + p_{12} \leq 1$. The receivers are identified in the following receiver sets: $1 \in \overline{\mathcal{K}}(11), 1 \in \underline{\mathcal{K}}(12)$, $2 \in \overline{\mathcal{K}}(12), 2 \in \underline{\mathcal{K}}(11)$.
- Multiple Access Channel (MAC): Transmitters 12 and 2 send distinct useful information to receiver 2. Receiver 2 decodes the data from transmitter 12 and 2 successively. Thus, $2 \in \overline{\mathcal{K}}(12), 2 \in \overline{\mathcal{K}}(2)$.

- Multicast: Transmitter 2 sends common useful data in a multicast to receivers 2 and 3. The receivers are identified in the following receiver sets: $2 \in \overline{\mathcal{K}}(2), 3 \in \overline{\mathcal{K}}(2)$.
- Interference Channel (IC): Transmitter 2 induces interference on receiver 1, while transmitter 1 induces interference on receiver 3.

The receiver sets are summarized in Figure 3.3, and the solid and dashed arrows refer to useful and not useful signal directions, respectively. The achievable rate at receiver 1 is

$$u_1(x_{11,1}(\boldsymbol{Q}_{11}), x_{12,1}(\boldsymbol{Q}_{12}), x_{2,1}(\boldsymbol{Q}_2)) = \log_2\left(1 + \frac{\boldsymbol{h}_{11}^H \boldsymbol{Q}_{11} \boldsymbol{h}_{11}}{\sigma^2 + \boldsymbol{h}_{11}^H \boldsymbol{Q}_{12} \boldsymbol{h}_{11} + \boldsymbol{h}_{21}^H \boldsymbol{Q}_2 \boldsymbol{h}_{21}}\right), \tag{3.22}$$

which is monotonically increasing in $x_{11,1}(\boldsymbol{Q}_{11})$ and monotonically decreasing in the power gains from transmitters 12 and 2. The utility at receiver 2 is its sum capacity,

$$u_2(x_{11,2}(\boldsymbol{Q}_{11}), x_{12,2}(\boldsymbol{Q}_{12}), x_{2,2}(\boldsymbol{Q}_2)) = \log_2\left(1 + \frac{\boldsymbol{h}_{12}^H \boldsymbol{Q}_{12} \boldsymbol{h}_{12} + \boldsymbol{h}_{22}^H \boldsymbol{Q}_2 \boldsymbol{h}_{22}}{\sigma^2 + \boldsymbol{h}_{12}^H \boldsymbol{Q}_{11} \boldsymbol{h}_{12}}\right), \tag{3.23}$$

which is monotonically increasing in $x_{12,2}(\boldsymbol{Q}_{12})$ and $x_{2,2}(\boldsymbol{Q}_2)$. The utility function at receiver 3 is the achievable rate,

$$u_3(x_{11,3}(\boldsymbol{Q}_{11}), x_{12,3}(\boldsymbol{Q}_{12}), x_{2,3}(\boldsymbol{Q}_2)) = \log_2\left(1 + \frac{\boldsymbol{h}_{23}^H \boldsymbol{Q}_2 \boldsymbol{h}_{23}}{\sigma^2 + \boldsymbol{h}_{13}^H \boldsymbol{Q}_{11} \boldsymbol{h}_{13} + \boldsymbol{h}_{13}^H \boldsymbol{Q}_{12} \boldsymbol{h}_{13}}\right). \tag{3.24}$$

which is monotonically increasing in $x_{2,3}(\boldsymbol{Q}_2)$. Note that the transmission rate at transmitter 2 has to be chosen such that both receiver 2 and 3 can decode the data successfully. We do not consider this requirement in (3.23) and (3.24). These rates can be achieved using rateless coding. The utility functions in (3.22)-(3.24) satisfy properties A and B in Assumption 3.1.

In this example, it can be observed that the optimization of these three utility functions is in general a multi-criteria optimization problem. The corresponding utility regions are illustrated in Figure 3.4. The region is not convex and finding efficient operating points on the boundary of the region is difficult.

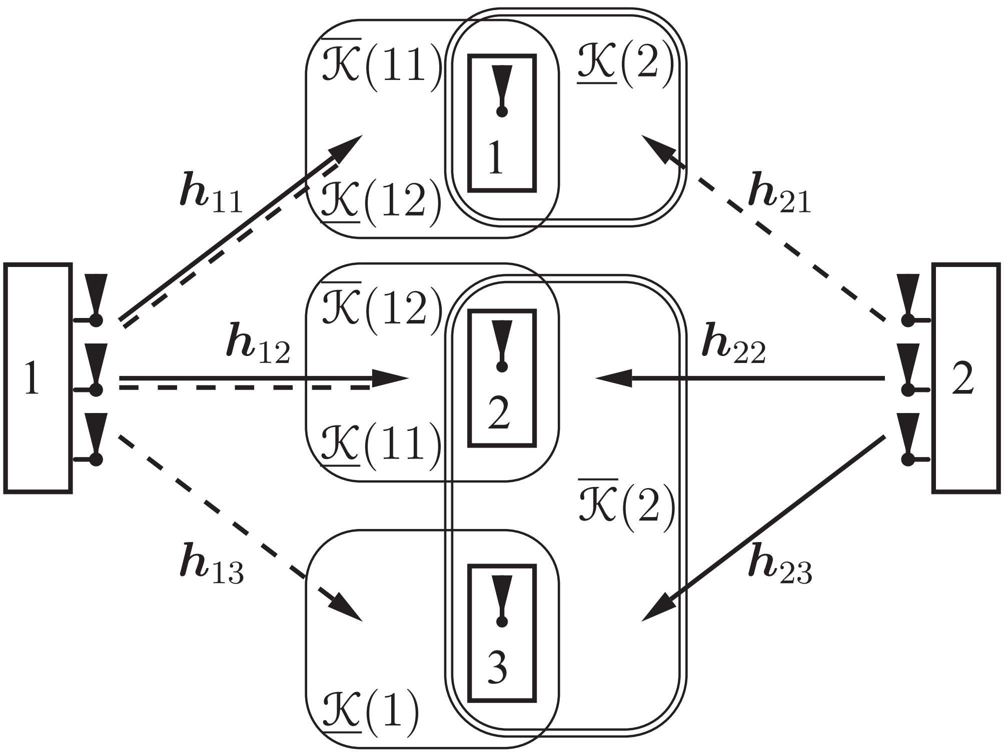


Figure 3.3: An example setting for the described system model. There exist two transmitters, each equipped with three antennas, and three single antenna receivers. The solid arrows refer to the intended receivers of a transmitter, while the dashed arrows refer to interference directions.

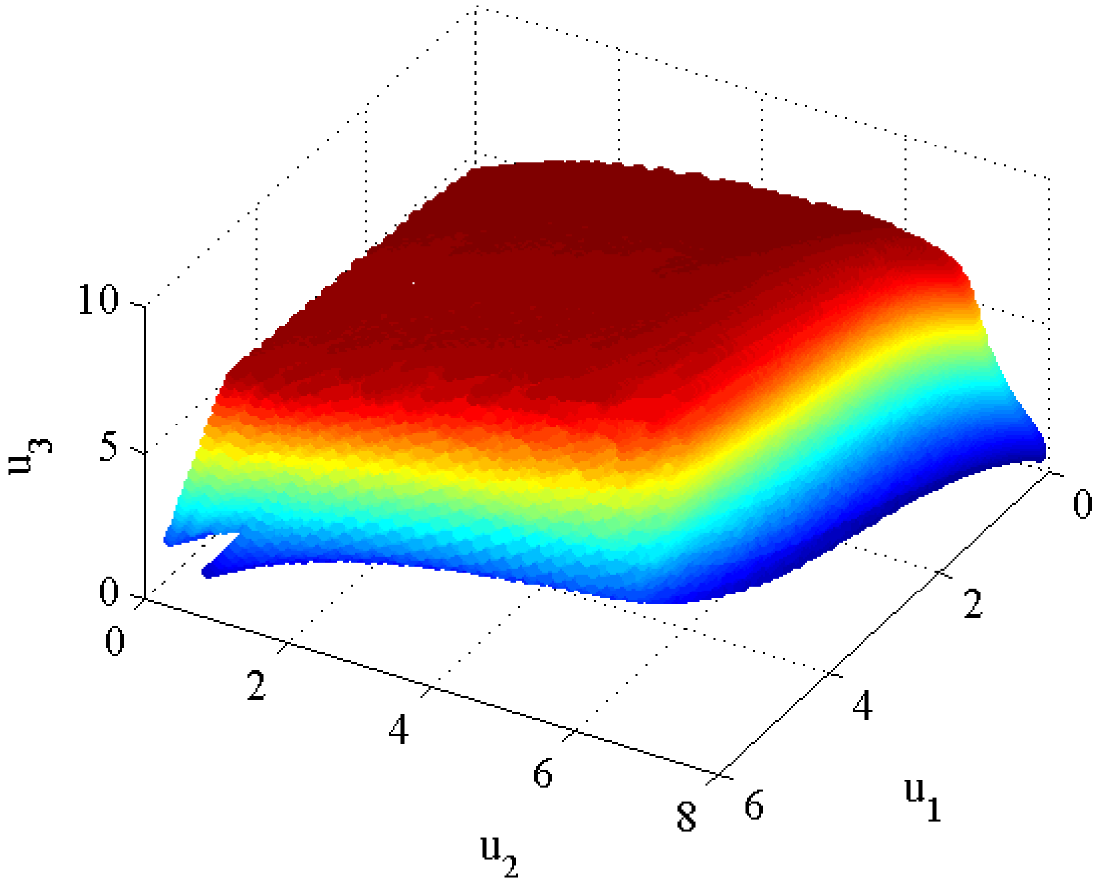


Figure 3.4: Pareto boundary of the utility region of the setting described with SNR=15 dB and $N = 3$.

4 Infrastructure sharing models

4.1 Infrastructure Sharing State-of-the-Art

One of the most challenging tasks in nowadays radio network planning is to provide the solutions that would allow the operators to significantly reduce the ever-growing OPEX [34]. It is especially important in deployment phase of new wireless technologies, as although the new wireless technologies lead to the exponential traffic growth, the increased demand for service is not balanced with increased revenues for Mobile Network Operators (see also Fig. 4.1). Temporary solution to the problem is to be the first to introduce novel technologies (in the present tense it would be LTE and LTE-Advanced), which allows to benefit from the market monopoly (increased service pricing). However introduction of new technologies is limited by the cost of familiarizing with the technology, spectrum and site acquisition and pressures from regulatory bodies ([31]) to minimize number of sites in dense urban areas[1]. Thus, current interests of Mobile Network Operators (MNO) are moved from the full ownership[2] to shared infrastructure networks. Some aspects of network infrastructure sharing are already widely popular due to lower CAPEX and OPEX imposed, e.g. co-location [31] or full network sharing [28]. Furthermore, the importance of network infrastructure sharing was noticed by mobile standards standardization bodies, such as 3GPP [35]. Application of infrastructure sharing might be especially important (can provide high cost-efficiency) in any new roll-out, consolidation (old technology replacement) or in coverage-driven network expansions[3]. In the case of capacity-driven network expansions[4] sharing might not be the most efficient solution, as capacity enhancements are usually connected with additional CAPEX related to software licenses or baseband card extension, which although providing non-negligible costs, can be compensated over time.

Network infrastructure sharing can be characterized in terms of [19]:

- *Business model* - describes parties involved in the sharing and relations be-

[1] Regulatory bodies would not only be interested in increased radio network capacity but also in providing coverage in the rural areas which are not attractive from business perspective, as a social responsibility of the MNOs.

[2] MNO plans the network, acquires and builds the sites and eventually implements, operates and maintains the network (where some of the parts might be done by subcontractors). Furthermore, it is the responsibility of MNO to obtain the required network equipment as well as software and spectral licenses.

[3] Rural area deployments.

[4] Addition of new channel elements, driven by a growing number of users (typically dense urban environments).

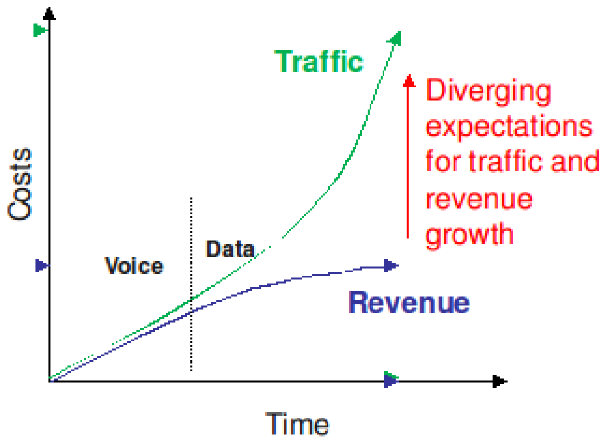


Figure 4.1: Growth in traffic data costs in comparison with operators revenues [16].

tween them.

- *Geographic model* - describes physical footprint of each of the parties involved in infrastructure sharing; it involves ownership of the network elements and the depth of sharing solutions.
- *Technology model* - describes technological solutions that are implemented to provide sharing, e.g. aggregation nodes, schedulers.

While the issues related to business and geographic models of network sharing are discussed in detail in SAPHYRE deliverables D5.1a/b and D5.2a/b, herein we concentrate on describing the technological models for network infrastructure sharing.

4.1.1 Types of Infrastructure Sharing

Depending on the solution for infrastructure sharing the cost savings can vary and most of the time they are achieved by a sacrifice in the domain of standalone network control. This aspect is very crucial especially when active resources (such as baseband processing powers or backhaul capacity) are shared as loss of standalone control might lead to QoS degradation and jeopardization of confidential operator's traffic information. Fig. 4.2 describes and positions different infrastructure sharing technological solutions against the level of network control, level of sharing and cost savings.

The technical solutions for infrastructure sharing are typically differentiated based on the type of network element or site equipment that is being shared [2, 19]:

1. *Passive RAN*[5] sharing - it is often referred to as site sharing or co-location. It has already become a solution for the operators to reduce capital (e.g.

[5] *Radio Access Network*

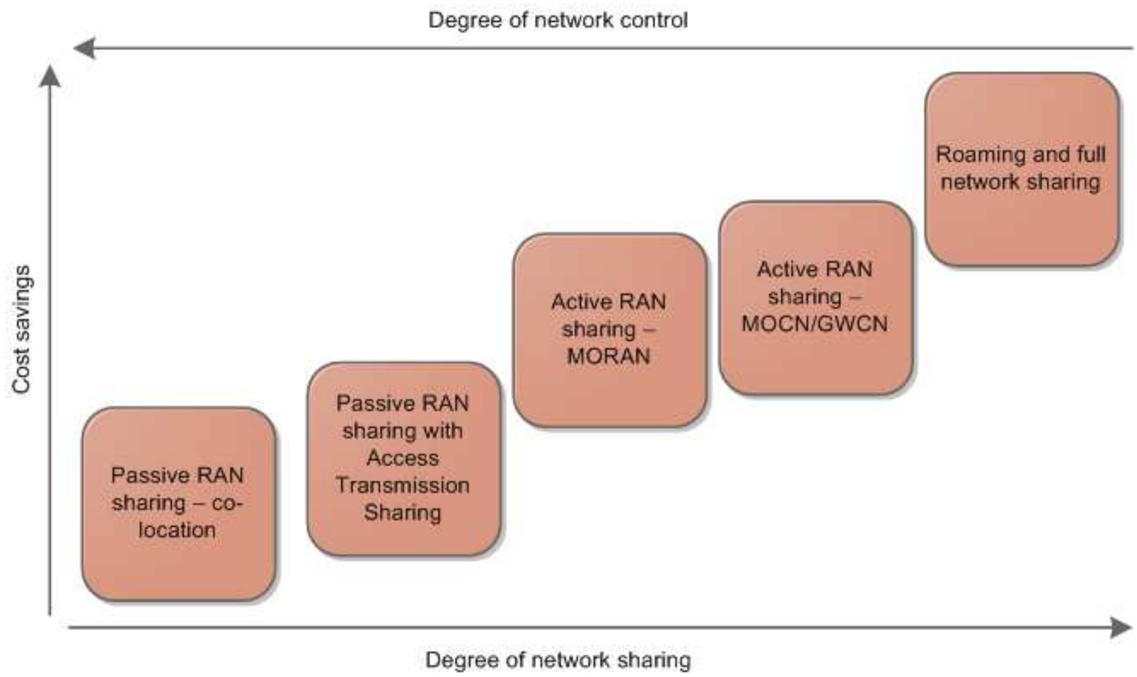


Figure 4.2: Different types of infrastructure sharing in the domains of network control, sharing and cost savings.

acquisition, civil work, mast) and operational expenditures (e.g. site rental fee, site maintenance fee) [31]. Possible solutions include also third party owners, which can be specialized rental companies that provide operation of e.g. telecommunication masts. Regulators are encouraging site sharing as it leads to the reduction of total number of sites. Site sharing may involve sharing of equipment such as site itself, mast, shelters, cabinets, electric power supply, air conditioning, diesel or biofuel generators, ducts and antennas[6]. Drawbacks of this type of infrastructure sharing are:

- The need for coordination of operational and planning aspects with sharing partners.
- In the case of antenna sharing, combined loss (combining) and no RX diversity (chaining).

2. *Passive RAN sharing with Access Transmission Sharing* - additionally to passive RAN sharing also transport links (backhauls) between a base station and a controller can be shared, e.g. in 2G between BTS and BSC (Base Station Controller), in 3G between NodeB and RNC (Radio Network Controller), in 4G between eNodeB and MME/SGW (Mobility Management Entity/Serving Gateway). It should be explained that typically backhaul sharing is realized

[6]Antenna systems can be shared in multiple ways: shared antenna radomes, shared antenna system by combining (combiners are used to connect antenna systems or receivers) or chaining (antennas are linked in chains)

by a physical separation of the two operators, where either physical links are separated but the aggregation/multiplexing equipment is common or operators use separate carriers to transmit their data. Both of the solutions provide traffic separation and moderate cost savings. Nonetheless backhaul can be shared also as an active element, where the scheduled resource is available capacity. Such type of sharing can be realized also via third party, which provides logical links with specific classes of service offered to different operators. Physically backhaul link can be realized via leased lines[7], microwave communications (e.g. WiMAX), or fiber communications.

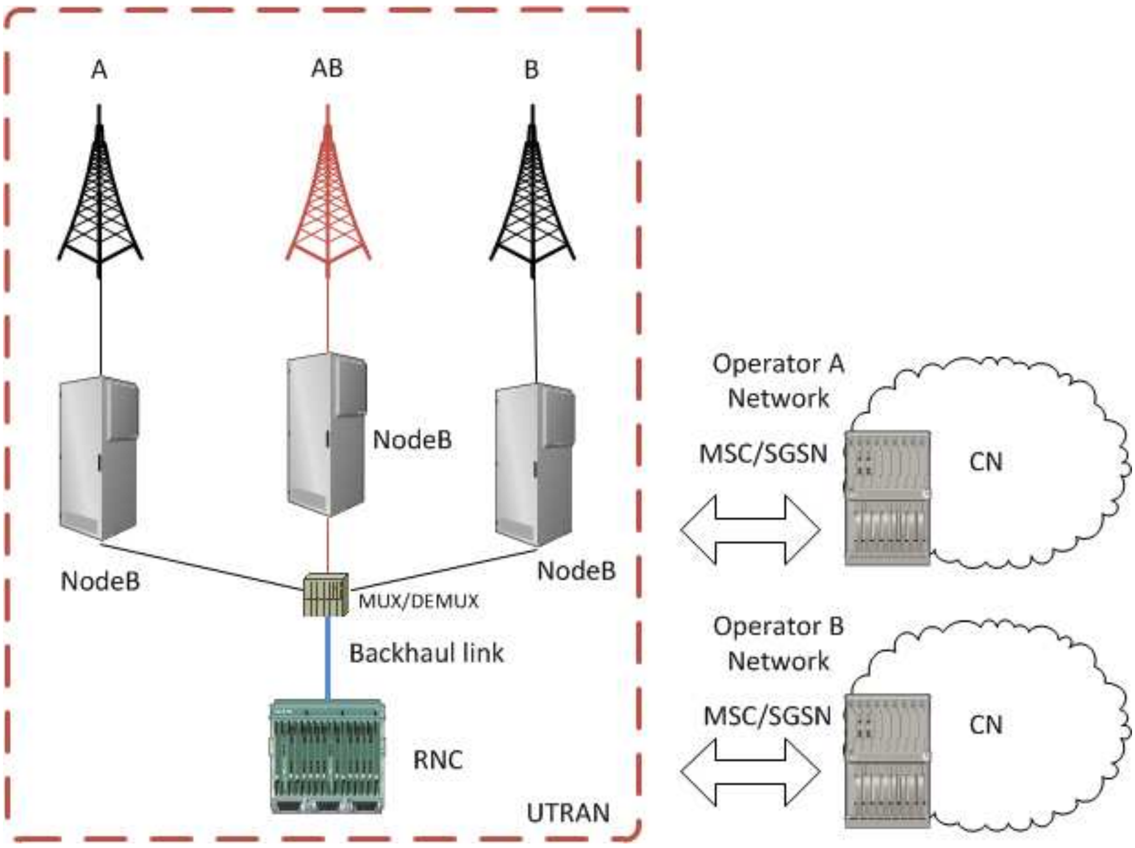


Figure 4.3: Passive RAN sharing with shared backhaul link in 3G networks.

3. *Active RAN sharing Multi-Operator RAN (MORAN)* - the approach describes another degree of sharing, where also active (in the sense of changes in the equipments software) elements (e.g. base stations) of mobile networks are shared. In this approach even though shared element is active, operator maintains control over traffic flow as well as quality aspects (coverage, capacity, link parameters). Control is maintained by virtual (logical) and static division

[7]The name is typically used to describe circuit-switched WANs, which allow permanent connection between two points set up by a telecommunications common carrier, referred to as private lines or dedicated lines [31].

of shared network elements. Possible network elements that can be shared: Base Stations (baseband cards, power amplifiers), BSC, RNC or Relay Node[8]. Such virtual access networks are then connected to the respective operator core network. This technical solution allows operators to additionally reduce costs due to lower number of network elements[9], maintain total independence in their roaming agreements and keep the sharing not visible to the users. Unfortunately Active RAN sharing solution has also many drawbacks [31]:

- operators have to use adjacent bands, e.g. due to technical limitations of power amplifiers,
- all the optional features of network elements have to be the same for both operators,
- capacity (CEs - channel elements) is pooled between the operators (one operator can exhaust available CEs),
- due to static division, elements may stay underutilized if operators have asymmetrical traffic volumes.
- O&M architecture to manage shared systems is very complex.

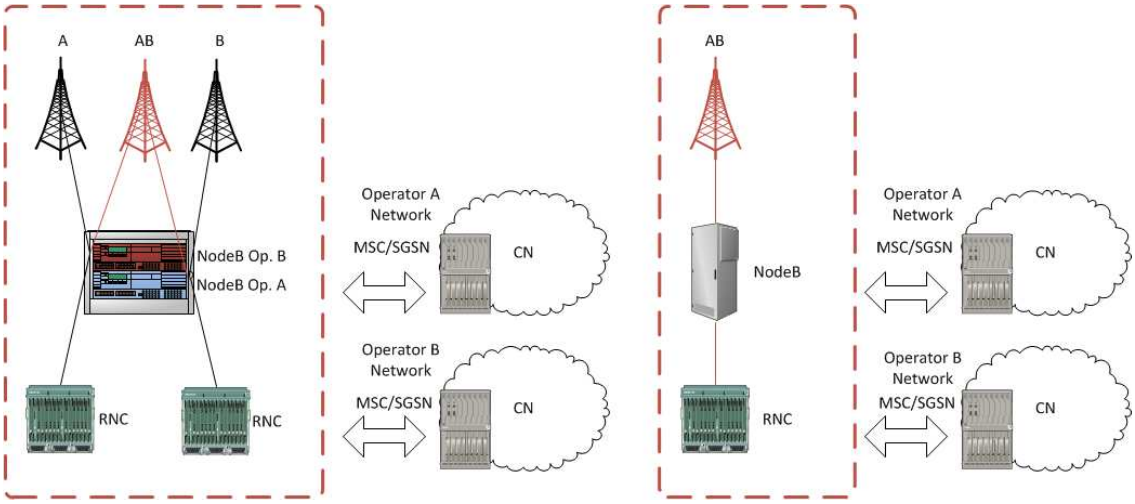


Figure 4.4: A) Passive RAN sharing (co-located site and shared cabinet). B) Active RAN sharing (shared UTRAN).

4. *Active RAN Sharing 3G Multi-Operator Core Network (MOCN) and (GWCN))* [35] - multiple operators share UTRAN[10], eUTRAN network elements and common frequency pool. GWCN approach introduces additionally partial sharing of the Core Network: in 3G MSC (Mobile Switching Center), SGSN

[8] Elements of mobile network infrastructure without a wired backhaul connection, that relay messages between the base station and mobile stations through multi-hop communication (to improve coverage, especially on the cell edges) [32].

[9] This leads to less power consumption, split of planning, optimization and maintenance costs [19].

[10] UMTS Terrestrial Radio Access Network

(Serving GPRS Support Node) and in 4G MME (Mobility Management Entity). In case of shared accesses each cell in shared radio access network broadcasts (in the system information blocks) information on available core network operators PLMN-ids. This information is used by the users during attachment, handover and cell re-selection procedures. The available core network operators shall be the same for all cells of Location Area (3G) or Tracking Area (4G). In this solution network sharing is transparent to the users. As a result of this approach operators are loosing much of their control over traffic capacity and quality. Therefore this solution is mostly acceptable for low traffic rural areas. However with proper inter-operator agreements SLA (Service Level Agreements) and resource sharing algorithms QoS regime and fairness can be provided to satisfy network sharing operators. The saved costs are comparable to the ones achieved with MORAN approach.

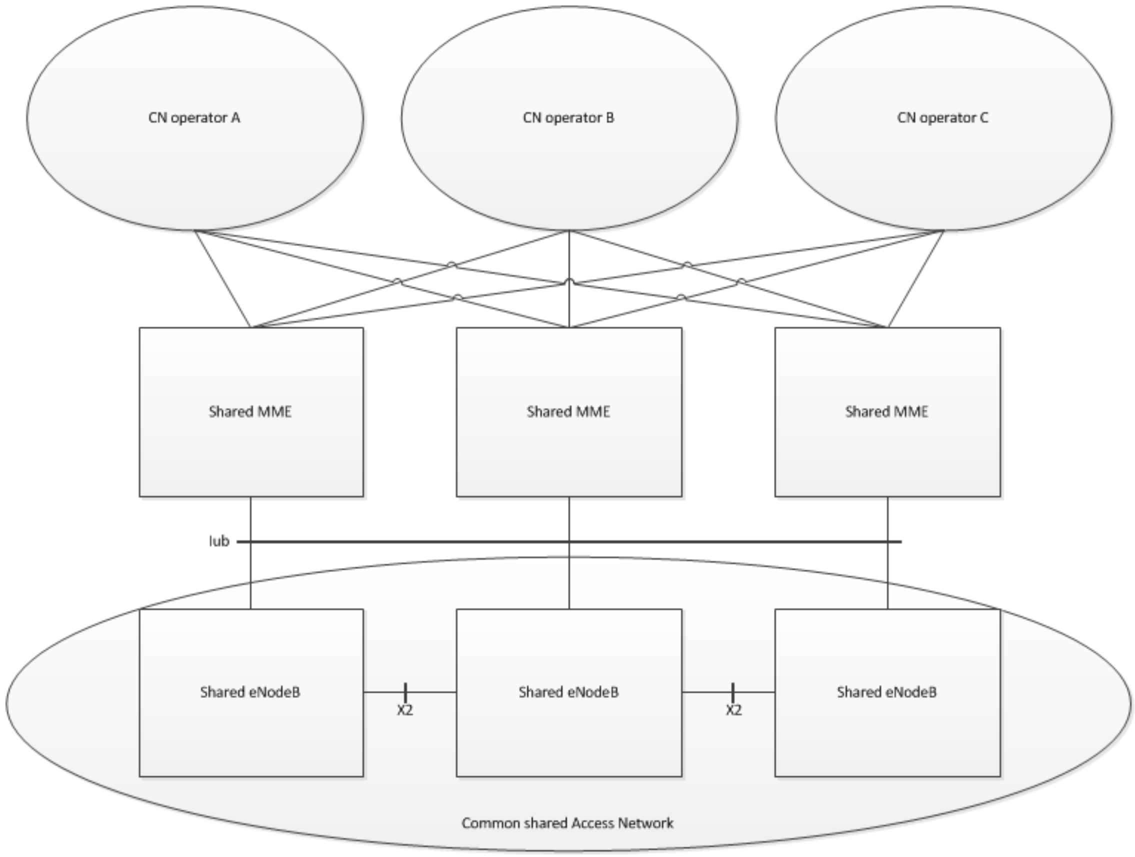


Figure 4.5: 3GPP vision of infrastructure sharing - Gateway Core Network sharing.

5. *Roaming-based sharing and full network sharing* [31] - this type of sharing relies on the fact that each operator deploys its own network with a dedicated frequency band in specific region of the world/continent/country. In order to enable roaming-based sharing each of the operator's networks need to support inter-PLMN mobility procedures. From legal perspective this means that op-

erators require national roaming agreements, Service Level Agreements and and agreement from the regulator to provide world-wide coverage. In fact up till today roaming is a typical solution for provisioning of international coverage (international roaming). National roaming can also be used by greenfield operators to provide nation-wide coverage before developing their own network. In the full sharing case operators only retain portion of the core network separated: HLR (Home Location Register), authentication (AUC and AAA server) and billing system (online and offline charging). Such a separation can be met in the deployment of MVNOs (Mobile Virtual Network Operators).

4.1.2 Network Sharing in Future Mobile Networks

Network infrastructure sharing in contemporary mobile networks is regarded as an additional feature that may decrease CAPEX and OPEX for Mobile Network Operators (MNO). However, while the mobile networks are envisioned to evolve into Cellular Network Clouds [18] or Wireless Network Clouds [26], network sharing will become a paradigm underlying the design of any of the future mobile networks.

According to the vision presented in [18] mobile networks will virtualize based on computational and storage capabilities of data centers and vast deployment of submissive network components (i.e. components that do not have inherent features that make them biased towards any particular technology or model of usage). The potential Cellular Network Cloud would consist of highly concentrated sites, each equipped with submissive base station, lattice of microwave links and optical fibre links, which shall inter-connect to the core network. The core network and its functionalities, such as RAN control, network optimization, billing or authentication, could be virtualized and physically situated in any part of the network, depending on the availability of processing units. At the same time radio access networks and baseband processing could be deployed on general purpose processors (GPU) in submissive base station equipment which would be able to switch between radio access technologies based on firmware updates. In [18] authors highlight also an important aspect of network virtualization which is connected with introduction of femtocell and user deployed accesses in general, which could provide a radio service to subscribers of any potential operators. Sharing of non-3GPP access hot spots has been indicate also in [15] as an option to alleviate the traffic burden from macrocells to user deployed accesses.

In somewhat similar vision populated by IBM [26] Wireless Network Clouds rely on wide deployment of RRH (Remote Radio Head) units, which are deployed to decouple processing units from signal up/down conversion. RRH units are connected using high speed optical fiber links to the network which enables the signals to be passed to baseband processing units implemented on general purpose processors in data centers. The inherent part of the vision is that base station functionalities are developed completely in software and along with the core network functionalities are implemented and run on servers in data centers. Pure software implementation of networking functionalities would allow for multi-threading and baseband unit pooling, thus network sharing based on data centers computational powers could be an interesting opportunity for operators seeking to decrease their OPEX as well as contribution to global energy consumption. In such a solution resources could be managed dynamically among the operators where any kind of amendments to sharing agreements could be employed during run-time.

In Fig. 4.6 we present a potential vision of the shared future mobile network where the network processing units (core and access network processing) are virtually allocated to the data centers. The baseband signals are transmitted over high speed low latency optical fiber links, which are shared between the operators (or are leased from third parties), to radio front-ends which perform up/down conversion of the signals. Virtualization of the optical backhaul network (for better inter-operator separation) allows for better resiliency, i.e. it enables fast re-routing in case of link failures or congestion, and with controlled flow parameters (e.g. latency) it allows for operation of Coordinated Multipoint (CoMP) techniques.

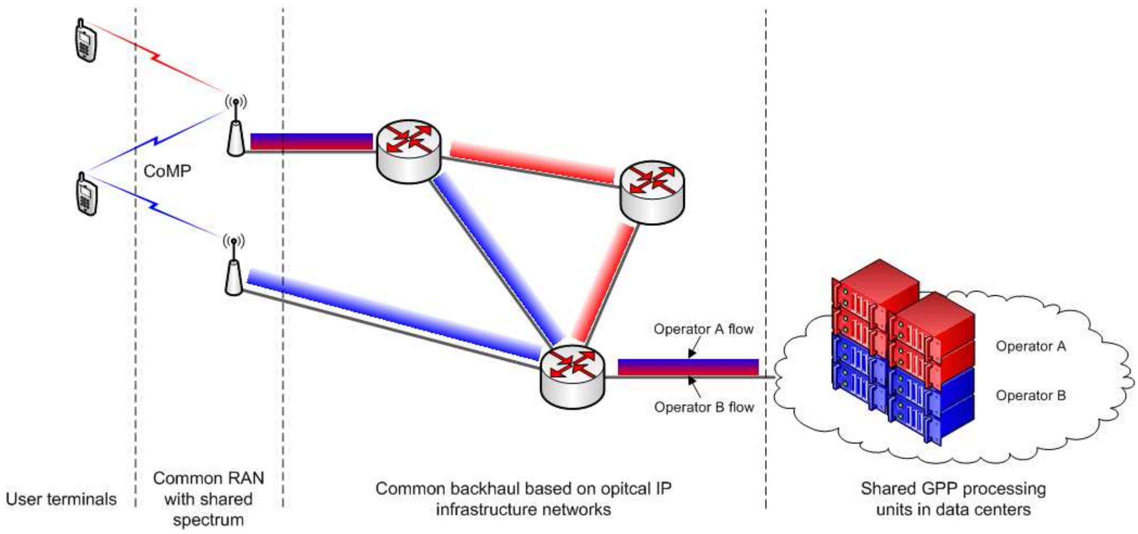


Figure 4.6: Vision of the future mobile networks with sharing based on virtualization and cloud computing.

What is inherent to each of the presented visions is that in an essence it is not necessary to deploy network elements by potential mobile operators as the radio front-end elements would be setup and connected to the home fiber line by the end-users, and all the functionality related to typical network functions would already be available as a part of the data center computational powers.

The presented visions of shared cloud-based mobile networks would require a change of architectural paradigms from the contemporary mobile networks. The requirements for infrastructure sharing in future cloud-based mobile networks can be enlisted as follows (loosely based on [23]):

- **Service level operation** - The most important part of any infrastructure sharing solution is to provide high (agreed) reliability and quality of service to all the sharing parties. It is important that sharing operators are able to set QoS and provide the service independently from each other at any times and up to levels guaranteed by the service level agreements. Virtualization of the network imposes also higher demand for hardware redundancy and resiliency, which has to be followed by any potential recovery solutions that will prevent propagation of errors or failures in network settings from one operator to another.
- **Full infrastructure sharing** - Although, roaming-based agreements and full sharing are already part of contemporary mobile networks, it is inevitable that with vast introduction of cloud servers operators would ultimately sharing their resources, i.e. even crucial core network elements would be shared between the involved parties. In such a scenario there is a need for operators to share computational and storage space in a way that does not cause disruption to the other operator's operation and does not allow the operator to access the vulnerable data of the other operator. Furthermore, such an approach shall enable simplified deployment and implementation of inter-connection points between operators, which could potentially lead to lower signalling overhead in case of inter-operator mobility.
- **Control and data plane virtualization** - while it is relatively easy to isolate data plane among operators (e.g. by means of different service bearers) it is rather inconvenient to do the same with control and management plane. The sharing operators have access to the same pool of control messages which control the shared network resources, this means that any faulty setting by any of the operators may mean also a fault to the other operator. Furthermore, the parameter settings utilized by one operator might be monitored and used by the other operator to disrupt the other operator network (e.g. decrease coverage performance) or track the behaviour of end-users to better target competitive contract offers. In such a case it is important that the operators working on shared resources are able to independently control their portion of the resources without disruption to other's operator network.
- **Network configuration and optimization** - The operators operating on future cloud-based mobile networks shall be able to configure physical parameters of the network independently from each other, taking into account their business goals (e.g. energy efficiency or coverage support) as well as also capabilities of their end-users (e.g. end terminals or services they are using). The equipment used in the mobile network infrastructure shall be as flexible as possible, especially at radio front-end where reconfigurable RF equipment shall be utilized, e.g. wideband Power Amplifiers and wideband antennas. Furthermore, the equipment shall enable the operators to define individually protocol stacks and communication protocols of their network architecture, in order to increase service dedication (e.g. protocol simplification for latency

vulnerable services such as video) and facilitate service level differentiation among operators. Network configuration and optimization flexibility shall enable also for technical differentiation among sharing operators, as one could envision also possibility for operators to provide legacy services to its subscribers based on, e.g. 2G radio access technology, or trial deployments of new entrant technologies. Interesting result of this approach might be such that while different countries and regions may deploy different radio technologies in different radio bands (e.g. cdma2000 in US and UMTS in Europe), the end-users would be able to roam world-wide maintaining connectivity using their original terminal equipment and radio technology.

4.2 Wired Backhaul Sharing Model

The broad introduction of 'all-IP' concept along with Long Term Evolution (LTE) networks [25], meant that transport networks solutions have to be definitely shifted from TDMA (Time Division Multiple Access) techniques towards Ethernet links, which are more flexible in handling IP datagrams. When deploying high capacity Ethernet links operators may initially experience high underutilization of the available link resources, i.e. users subscribe mainly for voice service or peak to average ratio of traffic volume is very high. In such a case operators are seeking the ways to reduce total cost of ownership (TCO) for backhaul links to the decrease the cost per bit of transmitted data. Ideal solution to maximize the utilization and reduce TCO is to share backhaul links with other operators. Backhaul sharing itself is one of the most interesting but very challenging aspects of Radio Access Network sharing. The biggest threat lies in the congestion (and therefore also dropping probability), that can occur when one or more operators traffic is excessing its maximum share or backhaul capacity in general. This is highly unwanted situation against which a proper prevention mechanism shall be proposed. Furthermore it is not only about possible congestion but also fairness in the available resource distribution. In fact the key problem is to fairly distribute capacity resources with guaranteed Quality of Service levels for different services and still maintain high utilization of the available capacity. In principle there are at least three possible ways to deal with the problem:

1. *Backhaul dimensioning* based on maximum traffic ratio of each of the operators. The solution provides good support for different QoS levels, but it has many drawbacks that discard it as a potential candidate for backhaul sharing model:

   - high underutilization as the maximum rates usually occur only during small parts of the day[11],
   - increased expenditures due to över-planning;

[11] In fact what is important here is the ratio between peak and average traffic volume.

- small flexibility towards future growth of demand and expansion of operator's network.

2. *Load (Congestion) Control* algorithms, which are used to avoid bottleneck problem in the limited capacity links. The idea is to reduce the transmission rate of flowing packets. The possible methods for rate limitation include algorithms such as Random Early Detection (RED) [17] or token bucket class of algorithms [38] where the incoming packets are dropped (blocked) depending on the link conditions, service priority and specific fairness rules. Such a solution can be applied to the backhaul link on different abstraction layers:
   - Radio Resource level where backhaul capacity influences radio link admission of new users (Radio Access Bearers).
   - transport layer where aggregation node or end router performs traffic classification and traffic scheduling to the link.
3. *Flow Control* (FC) algorithms, which are used to adapt the sending rate to the receiving rate at the final or intermediate node. The algorithm for flow control algorithms typically relies on end-to-end buffer state message exchange and RTT (Round Trip Time) measurements. Number of flow control algorithms has been proposed for Iub link between NodeB and RNC, as both entities can exchange their buffers state [29, 37]. The mechanism is however not fully applicable to LTE backhaul link as the communication with the Core Network may happen via external links, where routing paths might not be static, which could lead to high variations in RTT measured values. Nevertheless in [33] authors propose an algorithm for LTE radio link flow control that includes also the performance of the S1 interface.

There are also two additional aspects of backhaul sharing that are important in the process of shared backhaul link design[12]:

- Underlying physical medium transport solution.
- Economical model.

Specifically physical layer solution for backhaul link need to address requirements on:

1. capacity supported at the served cell(s),
2. number of users,
3. traffic models (including day-night traffic patterns),
4. site location.

Self-evidently, possible solutions will vary depending on the value of the above mentioned parameters. The typical solutions recognized for backhaul links are [25, 31]:

[12] Which are however not stressed in the document as they refer to the business and financial part.

- **Leased lines (dedicated lines)** - circuit-switched WANs (Wide Area Networks), which allow permanent connection between two end points of communication system via common carrier. Dedicated lines are typically used to connect geographically separated locations or to provide high capacity connection with the Internet. There are two basic types of leased lines realizations: p2p (point-to-point) and multipoint LMDS[13]. p2p lines are used to connect two locations directly for full-time and full capacity communication. LMDS lines are used to connect multiple locations to central facility over number of common transmission channels. Technologically leased lines can be realized as Very-high-bitrate Digital Subscriber Line (VDSL), NG SDH/SONET (Next Generation Synchronous Digital Hierarchy/Synchronous Optical Networking), Ethernet over fiber (Dense Wavelength Division Multiplex - DWDM) or PON (Passive Optical Networking).

- **Radio links** - p2p or multipoint connections realized over wireless media, where possible solutions include: licensed spectrum transmission paths (e.g. WiMAX), unlicensed spectrum transmission (e.g. WiFi) or even free space optics systems.

- **Self-backhauling** [21] - it is an approach proposed for LTE-Advanced to use eNodeB (which becomes anchor point) radio links as a backhaul for another eNodeB to communicate with the Core Network. It is IP layer multi-hop solution that reuses existing base stations and requires minimum system level adaptations.

The economical model in the state of the art solutions [7] for Ethernet link sharing can be realized in one of the two ways:

1. By rental of the links from fixed-line providers - where backhaul is realized as a combination of links: from base stations to the edge routers, between edge routers and operator's PLMNs. This may in fact mean that the link can be realized through the Internet.

2. By consolidation of the transport network infrastructure with other operators - where traffic separation and link scheduling is realized by traffic aggregation nodes.

The following section provides a complete model of a shared backhaul solution for SAPHYRE. The description of the model includes: LTE backhaul link architecture, transport mechanisms as well as possible solutions for flow separation between the operators.

[13]Local Multipoint Distribution System.

4.2.1 LTE backhaul link architecture

3GPP LTE offers significantly higher data rates in comparison to 2G and 3G systems. LTE is most likely to offer 100/50 Mbps in uplink/downlink (in 1 sector, with 20 MHz bandwidth) and even up to 1 Gbps for the whole site that consists of three sectors where downlink/uplink data rates are of 300/150 Mbps [36]. This in comparison with 3G peak data rates of 30 Mbps means that existing backhaul capacities need to be increased significantly by at least 10x the same capacity[14]. Another important factor that greatly differentiates 4G backhaul links from legacy systems, is the shift in architectural paradigm from hierarchical TDM (Time Division Multiplex) based architecture to flat IP-based. Flat architecture consists of lesser number of elements, which in turns forces mesh topology, and expects purely IP packet transmission between network entities. Fig. 4.7 presents a flat architecture of backhaul link connectivity, where RANs are connected with different Core Network entities as well as with each other, constituting for meshed topology.

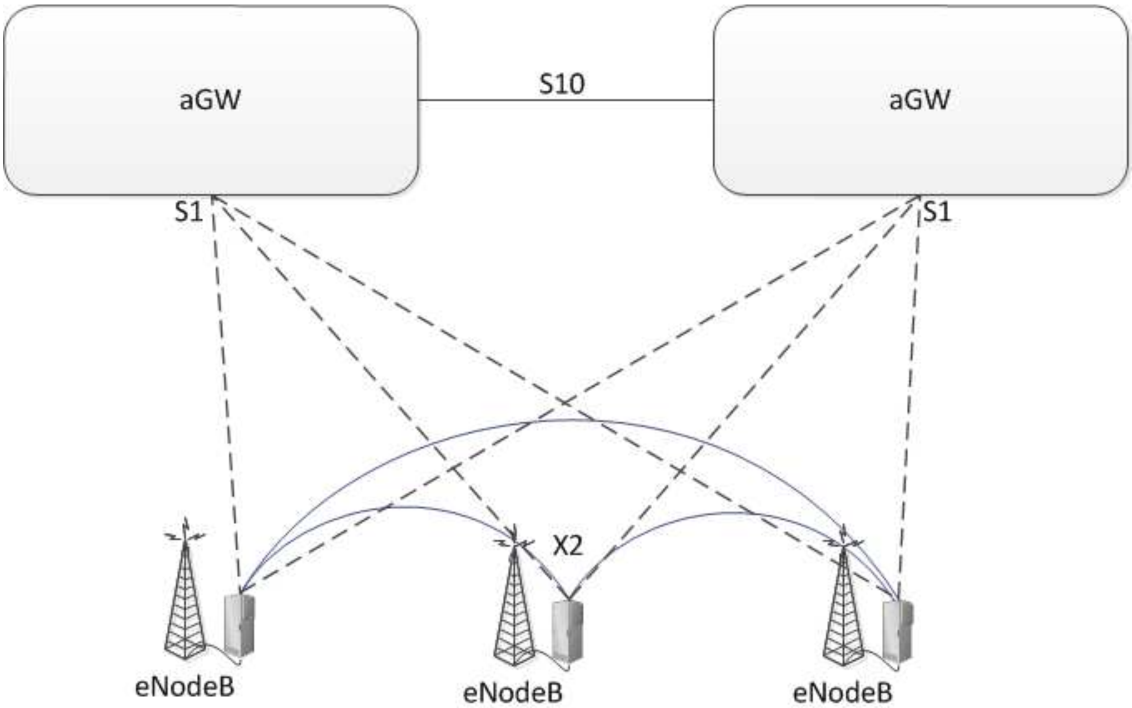


Figure 4.7: Flat architecture backhaul connectivity with inter-eNodeB communication.

Due to mesh network topology (inter-eNodeB connectivity), radio network controllers were replaced with evolved base stations (eNodeB), which perform dynamic resource allocation, radio admission control, connection mobility control, measurement configuration and intercell Radio Resource Management [25]. The introduc-

[14]It is expected that due to enhanced bandwidth and extra features proposed, LTE-Advanced will have even up to 1 Gpbs in downlink [4].

tion of new functionalities directly in Radio Access Networks lead to changes in the system architecture and appearance of new logical interfaces which constitute for evolved backhaul links (see also Fig. 4.8):

S1-MME. - which is a point-to-point[15] link that carries control plane data between eNodeB and MME. Functions of the interface include: handling of RAB (Radio Access Bearer) procedures, handover procedures, NAS (Non Access Stratum) signaling and paging. The interface is required to provide high level of reliability in order to avoid message retransmissions and unnecessary delay in control plane procedure executions [25]. Due these requirements the interface uses reliable end-to-end transport layer communication via SCTP (Stream Control Transmission Protocol)[16].

S1-U. - which is either point-to-point or point-to-multipoint (S1-flex[17]) link, that connects eNodeB with Serving Gateway(s) to transport user data packets. There is no need for flow control nor error control, nor any mechanism to guarantee data delivery over the S1-U interface, therefore in transport layer GTP (GPRS Tunneling Protocol) protocol over UDP (User Datagram Protocol) is used, which provides only data encapsulation [25].

X2. - which is a point-to-multipoint interface inter-connecting eNodeBs. It is used to transport:

- User data packets (via GTP protocol), user context information and signaling (via SCTP) in the case of handover procedures.
- Load indicators - to support load balancing management and to optimize handover decisions.
- Intercell interference coordination (ICIC) - information required to support ICIC, e.g.: allocated carriers, specific information is yet to be specified.

Altogether the three logical interfaces bring an astonishing increase in the amount of transported data and signaling, traffic prioritization and new connectivity solutions, which poses a number of requirements on the backhaul link transport architecture.

The detailed requirements for next generation backhaul transport networks have been specified by NGMN (Next Generation Mobile Networks) [30]:

[15] Although it is possible to have connection towards multiple MMEs (Mobility Management Entity), terminal can be associated with only one MME at a time.

[16] SCTP implements path selection and monitoring, flow control, validation and selective acknowledgements and order preservation [25].

[17] S1-flex means that eNodeB can be connected to multiple Serving GWs and MMEs. Where each logical control connection between UE and MME is marked with using different S1-AP Id [25]. This solution gives robustness towards Core Network node failures, more flexibility in network architecture and limitation of inter-Core Network handover procedures.

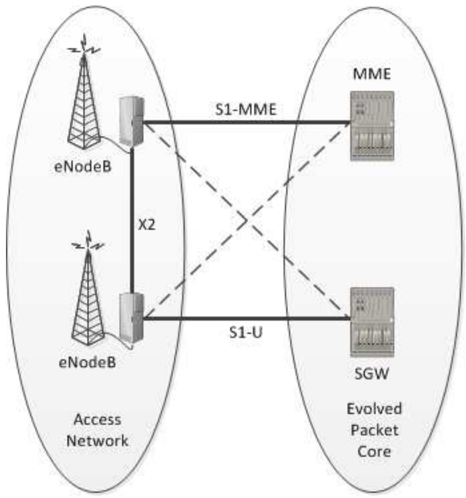


Figure 4.8: Evolved Packet Core backhaul link architecture.

- *High bandwidth* - current radio solutions require at least 450/150 Mbps in DL/UL. The demand for bandwidth (especially with introduction of LTE-Advanced) will increase over time, therefore transport network solution should be scalable to fit requirements of new radio interface solutions, different user environments (rural and urban sites) and rising interest in bandwidth consuming applications.
- *Flat architecture* - eNodeBs and Access Gateways (aGWs) shall be connected in a mesh topology type to provide many-to-many connectivity.
- *Support for QoS mechanism* - radio QoS Class Identifiers (QCI) shall be mapped on transport QoS markings, so that packet classification can be performed in the transport layer. Transport equipment is required to implement packet scheduling algorithms to guarantee requested QoS over backhaul link (in case of congestion high priority packets are sent first).
- *Low latency* - expected two-way delay shall not be higher than 10ms and it shall be possible for operators to achieve even 5ms if required.
- *Synchronization* - NGMN backhaul solutions shall support clock reference distribution over packet network. The synchronization mechanism shall support distribution of frequency, phase and time source to enable alignment of eNodeBs.
- *Link availability and fault restoration* - the availability of backhaul shall be tunable according to operators needs (e.g. in case of microwave links it can be done via Adaptive Modulation and Coding). The expected availability shall be 99,99% of time and expected link outages (before path resiliency works) shall be in the range of 50ms - 250ms.
- *Fault Management* - backhaul network elements shall have OAM (Operation, Administration and Maintenance) protocols to reactively and pro-actively re-

spond to link failures to support required end-user experience and Quality of Service. The OAM protocols shall enable backhaul link management (e.g. paths upgrade) via NMS (Network Management Systems).

- *Service continuity* - most of the already deployed sites in 2G and 3G technologies will still need to be maintained. It is highly likely that new LTE sites will be co-located with the legacy ones, therefore there is a need for next generation backhaul technologies to support also emulation of TDM services over Ethernet links or hybrid architecture were both technologies are supported using different carriers or separate physical links.

LTE Backhaul Transport Protocols

The protocols used in LTE transport network need to answer the NGMN requirements, and enable the implementation of aggregation nodes in the Layer 2, which allows for faster and less expensive traffic switching and aggregation. Fig. 4.9 presents an abstraction of possible LTE backhaul protocol stack for the consolidation backhaul link, where the main focus is on transport provisioning, realized via Ethernet protocol (possibly also Ethernet over SDH), which is further incorporated with PBB-TE (Ethernet tunnelling - Provider Backbone Bridging with Traffic Engineering) [1] or MPLS-TP (Multilabel Path Switching Transport Profile) [11] labelling to enable Virtual LAN creation. The specified protocols are used to transport either IP datagrams or legacy TDM services using circuit emulation (PWE3).

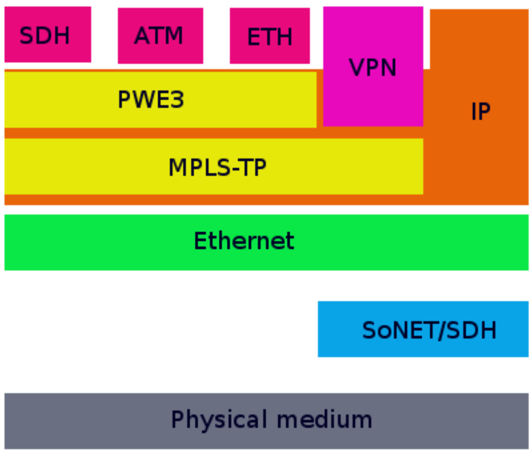


Figure 4.9: Protocol stack for consolidated backhaul link [14]

In order to efficiently realize logical connectivity in LTE backhaul networks each interface needs to be mapped to an appropriate Ethernet service[18] configuration and Ethernet Virtual Connections (EVCs), an example mapping [7]:

[18]For more information on Ethernet services, see [5].

- S1 interface can be realized as either E-Line (number of statistically multiplexed point-to-point connections) or E-Tree (point-to-multipoint connections between leaves and roots[19] as well as roots to a number of leaves) service.
- X2 interface can be realized as E-LAN service to facilitate direct point-to-multipoint interface towards other eNodeBs.

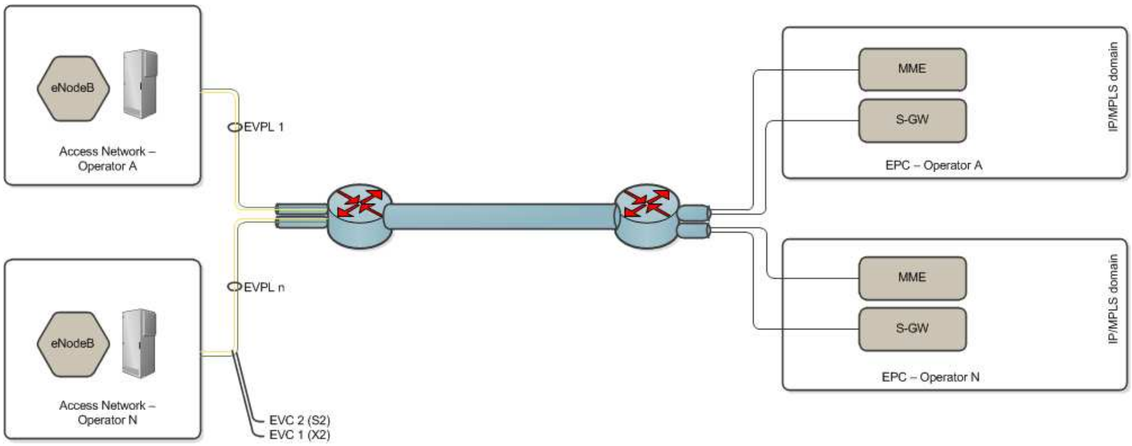


Figure 4.10: Realization of Ethernet Virtual Private Lines services in shared backhaul.

4.2.2 Backhaul Sharing Solution

In order to enable coexistence of multiple operators among same transport links at least three aspects need to be covered:

1. *Quality of Service delivery.* It shall be possible to classify packets based on the service they are carrying. The classification can be used then by traffic marking and limiting mechanisms to provide service differentiation and to guarantee minimum QoS classes described in Service Level Agreements.
2. *Traffic separation.* The second important aspect in designing shared backhaul solution is the traffic separation between the operators. Appropriate solution should enable application of different QoS policies to different traffic flows independently for each operator. The solution that enables logical separation of physical links is VLAN marking [8]. In VLAN approach flows of different operators are assigned to different VLANs and packets of each flow are marked with corresponding VLAN id. Each operator has at least one VLAN id, a number of VLANs can be used by one operator to separate also different logical interfaces.
3. *Synchronization.* Complete solution for shared backhaul shall address also the problem of synchronization provisioning (the operators typically prefer to

[19] In case of S1-flex multi-rooted E-Tree can be used.

maintain their own source of synchronization [7]). The distribution of synchronization signals (time, phase, frequency) is a challenging task in natively asynchronous packet networks. Due to the existence of multiple operators on one link, an end-to-end packet-based methods are required to provide proper separation of operators, instead of typical incorporation of reference signal into the physical layer, e.g. SDH or Sync-Ethernet.

Operators' Flows Separation

Traffic separation in a backhaul operated by multiple operators is realized via division of the available physical links into separate network domains, called Virtual LANs. Based on the general idea, there are two possible solutions to divide the resources [8]:

- Each operator maintains one VLAN network to transport its traffic.
- Each operator receives a pool of available VLAN networks to separate its traffic from other operators and to separate also different logical interfaces (S1 and X2). It is possible also to separate S1-U and S1-MME interfaces from each other, such solution however is not necessary as Core Network edge entity (aGW) routes the C-plane and U-plane traffic to destined entity.

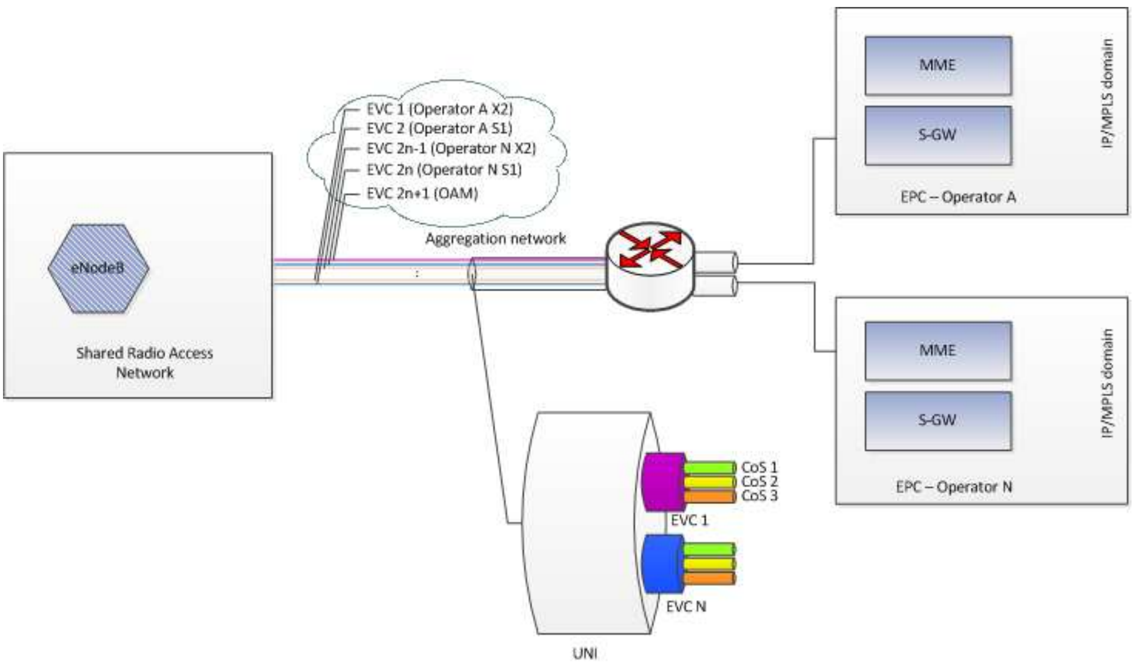


Figure 4.11: Solution for shared backhaul with EVCs for different interfaces.

Fig. 4.11 shows VLANs separation between different interfaces. Each packet of a flow assigned to specific VLAN is marked with a label called Virtual Path Label

Switched (VPLS)[20]. Typically the label is assigned by a customer edge router[21]. In the aggregation network all packets marked with the corresponding VLAN are assigned so called tunnel label that allows routing of the packets between aggregation nodes. Such solution is highly appreciated by operators as usually routers operating on lower layers are much cheaper in comparison to the ones operating on IP layer. Another degree of traffic separation can be achieved from usage of security protocols (i.e. S1-MME can be encapsulated with IPSec protocol) to minimize the possibility of jeopardizing the operators confidential data.

Each VLAN can be abstracted as Ethernet service that needs to be served with the appropriate QoS class. The provisioning of QoS agreements is done by the aggregation node. The QoS provisioning is a two-level structure that comprises both demands of services and particular operators. The realization is done through the following mechanisms[22]:

1. *Packet classification mechanism* - which is responsible for mapping of DiffServ QoS classes marked in IP header to corresponding CoS (Class of Service) provided by the transport layer protocol.

2. *Per service flow metering (rate limitation) and scheduling* - which is responsible for assignment of traffic rates (according to SLA) and pre-allocation of transport resources. The typical solution for rate limitation (and congestion control) in transport network is to use token bucket algorithms to mark the packets with specific color. Marked packets are sent to the queue of corresponding color and afterwards scheduled to the link. The queues are managed by Random Early Detection (RED) algorithm [17] to identify a priori possible congestion situations and provide early dropping of packets.

3. *Per operator flow rate limitation and scheduling* - once the requirements for services are fulfilled it is time for the aggregation nodes to confront the incoming rates against the link capacity resources and apply inter-operator resource sharing agreements. The SLA adaptation can be done again via rate limitation algorithm (this time without coloring), however it is important to note, that the proposed algorithm need to maintain minimum required rate that needs to be available to the operator's traffic. The traffic excessing link capacity is either dropped or blocked depending on the queue management solution.

Such a three step mechanism allows proper application of strict QoS parameters (e.g. throughput) to most demanding services and controlled distribution of resources among operators.

[20] MPLS or PBB-TE multi-protocol encapsulation of Ethernet frames.

[21] In the case of shared backhaul links labels can be assigned also by the base station's Transport Module

[22] The presented solution is a generalization of the solution for Carrier Ethernet transport networks presented in [6].

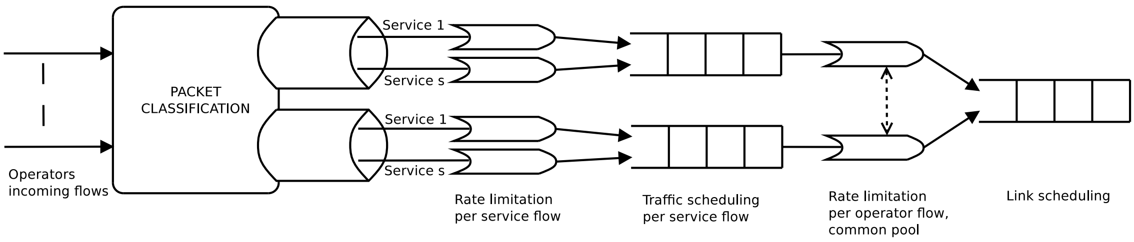


Figure 4.12: QoS provisioning system at aggregation node with multiple operators sharing backhaul link.

4.2.3 New Trends in Shared Backhaul Link

While current approaches to backhaul sharing include techniques based on VLANs, the newest trend is to deploy middleware that enables transparent (hardware independent) network virtualization and simplified flow management. One of the potential approaches to virtualize 3GPP transport link architecture is to utilize OpenFlow [23]. OpenFlow provides an open protocol to program the flow-table in different Ethernet network elements (e.g. switches, routers) [27]. It is then up to network administrator to control the OpenFlow and define different traffic flows within the network. Then each of the flows can be independently configured and managed (e.g. by means of setting up routing paths and flow control) by the flow owner. Each of the flow owners may also define different security, addressing schemes as well as apply different prioritization patterns.

In this way, OpenFlow enables dynamic sharing of link between the operators by means of controllers, which enable independent OAM operation as well as definitions of resource sharing policies if any are applied. This allows the operators to dynamically share the available network capacity and dynamically react to congestion or link failures in order to provide resiliency options. In such cases the OpenFlow controller upon detection of the demand is able to re-route the traffic from one operator backhaul link to another. Fig. 4.13 presents a potential architecture based on OpenFlow, where the whole physical network infrastructure is virtualized by means of OpenFlow. The flows are defined by the OpenFlow controller, while their characteristics are defined and controlled by each operator, in this way operators can set separate routes for their traffic as well as apply traffic prioritization policies that would allow to simultaneously serve their users (by means of transport link sharing) from a single access point (base station) that implements OpenFlow.

All these leads us to the conclusion that OpenFlow is an enabler of network and service virtualization, thus it is inevitable that the connectivity between shared radio accesses and the core networks might be efficiently realized with the aid of OpenFlow. In that sense OpenFlow might be an interesting candidate to aid deployment of future cloud-based mobile networks presented in the previous subsection.

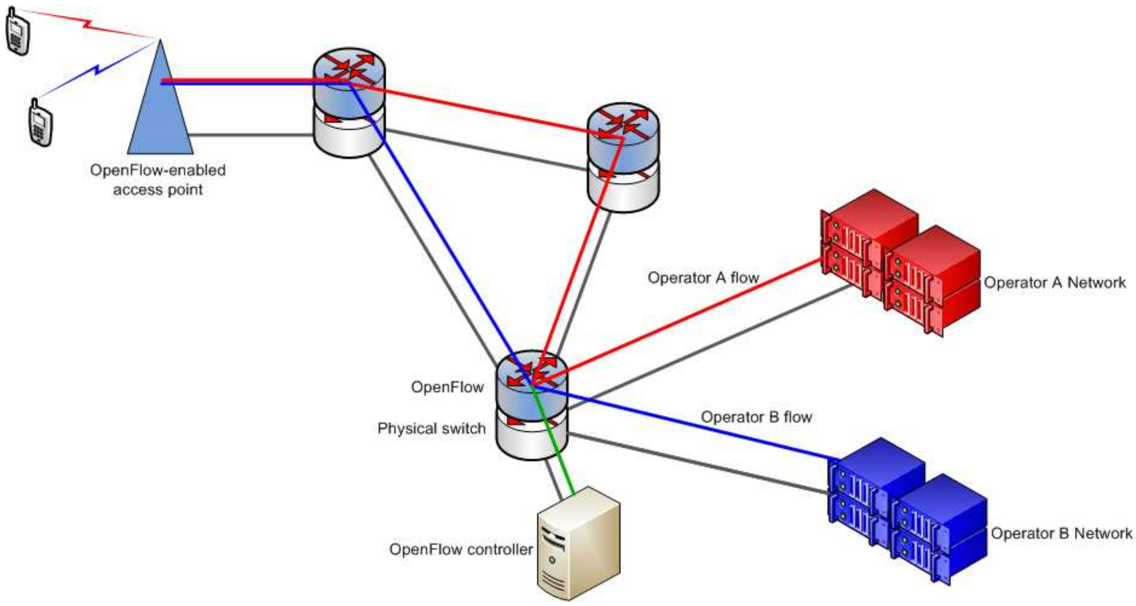


Figure 4.13: Future backhaul sharing solution based on OpenFlow virtualization.

Summary

Backhaul link sharing is an important aspect of a shared infrastructure, especially due to the fact that the backhaul performance can affect the call admission process at the radio link. Therefore, in the contemporary mobile networks it is very important to start shared backhaul deployment process from careful dimensioning and planning of an appropriate transport layer mechanisms for flow separation and service differentiation between the operators. Additionally, depending on the network architecture also aggregation nodes have to be designed properly in order to implement QoS provisioning service to the each of the operators involved in sharing (via rate limitation and scheduling).

In the future mobile networks it is envisioned that such functions will already be embedded into the network middleware. Hence, any mechanisms for scheduling, flow control, route planning and resiliency will be dynamically adjusted either by manual configuration or via cognitive mechanisms that will allow for runtime adaptation of the sharing parameters in order to maintain target (desired) network performance (defined by means of Service Level Agreements).

4.3 Wireless Backhaul Sharing Model

Recently, relays have been studied as wireless backhaul to provide efficient coverage extension and capacity increment. They can be employed with little or no incremental backhaul expense and applied in various scenarios where fixed line backhaul is

Relay technology	Analog AF	Digital AF	DF
latency	no	low	high
possible processing	real-valued scalar amplification	multiplication with a complex matrix (beamforming)	decoding re-encoding
noise forwarding	yes	yes	no
baseband processing	no	simple	complex

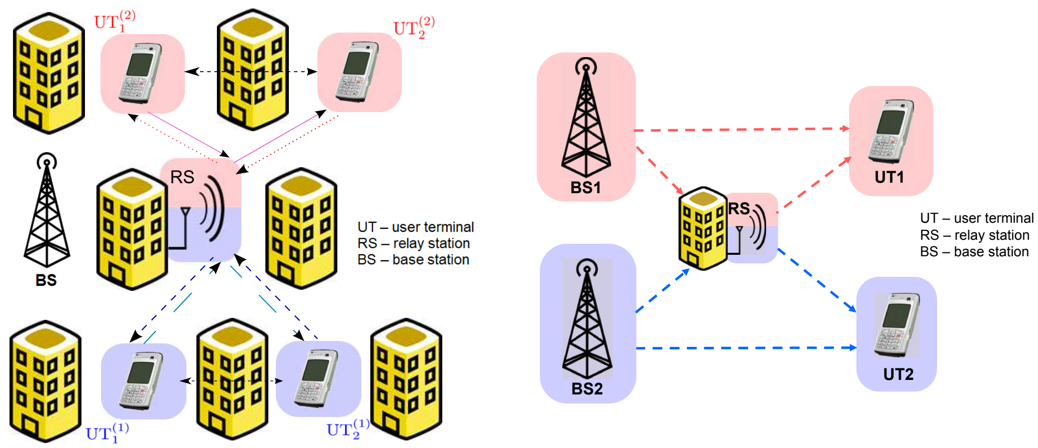
Table 4.1: Comparison of analog AF, digital AF and DF

difficult to deploy. For example, relays expand the coverage to moutainous regions or sparsely polupated areas and enhance the throughput for cell edge users.

The relay technology has experienced many years' development. The traditional relay is an analog amplify-and-forward (AF) type. The radio signals received on the downlink from the base stations or that on the uplink from the mobile users is simply amplified without further processing. Thereby, it is also called repeater or booster. The advantage of this kind of relay is that it is quite simple and no processing delay is caused. However, the interference as well as noise at the relay is amplified simutaneously together with the desired signal. Differently from the analog AF, the present AF relay usually down converts the radio frequency into the baseband and incorporates additional baseband processing such as sampling, spatial filtering, etc. After processing in the digital domain, the signal will be amplified and up converted for forward transmission. To distinguish these two types of AF relays, we call the latter the digital AF. Unlike analog AF, the digital store facilitates processing in digital domain and operation in half duplex mode. Furthermore, the digital AF does not require decoding and re-encoding of the data as decode-and-forward (DF) relay, where more complicated baseband processing causes a much higher latency. Furthermore, the rate region of the DF relay might be lower because it is restricted by the rate region obtained during the multiple access (MAC) phase or that obtained during the broadcasting (BC) phase. Thereby, it can be concluded that the digital AF is a good trade-off between performance and complexity. The comparison of these three types of relay is shown in Table 4.1.

By incorporating the digital AF into the wireless network, it is interesting to exploit the benifit of the wireless backhaul (relay) sharing for further system throughput improvement and expenditure saving. We will introduce two kinds of wireless backhaul sharing models that have been investigated under the scope of SAPHYRE.

One relay sharing model is called multiple operator two-way relaying, which corresponds to the metropolitan scenario as shown in Fig. 4.14(a). In this case, strong shadowing effects will cause many coverage holes and thereby dense networks are required to guarantee the QoS at the user terminals (UT). Moreover, considering the geometric constraints and network deployment costs, relays would be more



(a) multiple operator two-way relaying

(b) relay enhanced communications

Figure 4.14: Two concrete relay deployment scenarios. Left: metropolitan area. Right: suburban area.

suitable. Also taking into account that more than one operator or service provider would operate in the same area, if we share the relays as well as the spectrum, at the first glance it means lower capital expenditures and operating expenditures for all the operators. This model can also be applied to a disaster scenario where the base station cannot provide services any more. Then the relays can be deployed to temporarily maintain the communication among the local residents.

The other model incorporates BSs into the system, shown in Fig. 4.14(b). In order to guarantee the QoS of the cell-edge users, a relay is usually employed to assist the BS in addition to a direct transmission (might be quite weak), which is modeled as the relay channel (RC). We are motivated to extend this model to the two transceiver pair case belonging to two different operators to further improve the spectral efficiency, where the relay is shared and accessed by both BSs at the same time instead of an exclusive use of the relay for each operator in a TDMA mode. We refer to this model as an interference relay channel (IRC). More specifically, it describes the channel model where two independent transceiver pairs with multiple antennas communicate with the assistance of one relay, which operates in half-duplex mode.

5 Conclusions

SAPHYRE focuses on two forms of resource sharing, namely spectrum sharing and infrastructure sharing. The spectrum sharing introduces interference which can be resolved by the use of MIMO (multi-antenna) algorithms. Infrastructure sharing is a complex multi-level concept which comprises a number of different scenarios among which one distinguish especially the concepts of backhaul sharing and fixed wireless relay sharing. Several backhaul sharing models are provided and methodologies used to evaluate the sharing performance are outlined. A SAPHYRE gain is defined aimed at grasping the benefits of the proposed methods in a single performance number.

Bibliography

[1] *802.1Qay-2009, Virtual Bridged Local Area Networks Amendment 10: Provider Backbone Bridge Traffic Engineering*. IEEE.

[2] *Mobile infrastructure sharing*. GSM Association, www.gsmworld.com/documents/gsma.pdf.

[3] *Terms and definitions related to quality of service and network performance including dependability*. ITU-T Recommendation E.800, 1995.

[4] *LTE-Advanced - Evolving LTE towards IMT-Advanced*, October 2008. http://dx.doi.org/10.1109/VETECF.2008.313.

[5] *Carrier ethernet for mobile backhaul implementation agreement whitepaper*. Metro Ethernet Forum, February 2009.

[6] *Carrier ethernet: Service level agreement support tools*. RAD Data Communications, February 2009.

[7] *Lte backhaul: Meeting operator requirements, how transport providers can realize the business opportunity of lte with sla assurance*. RAD Data Communications, December 2009.

[8] *Network sharing in lte: Opportunity & solutions*. Alcatel Lucent Technology White Paper, May 2010.

[9] *Ts 23.107 quality of service (qos) concept and architecture (version 9.1.0 release 9)*. 3GPP, June 2010.

[10] *Ts 32.410 telecommunication management; key performance indicators (kpi) for umts and gsm (version 9.0.0 release 9)*. 3GPP, January 2010.

[11] Beller, Dieter and Rolf Sperber: *MPLS-TP The New Technology for Packet Transport Networks*. In *2 DFN Forum Munchen*, 2009.

[12] Bennis, M. and J. Lilleberg: *Inter-Operator Resource Sharing for 3G Systems and Beyond*. In *IEEE Ninth International Symposium on Spread Spectrum Techniques and Applications*, 2006.

[13] Blake, S., D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss: *RFC2475 - An Architecture for Differentiated Service*, December 1998. \url{http://www.faqs.org/rfcs/rfc2475.html}.

[14] Chundury, R.: *Mobile broadband backhaul: Adressing the challenge*. Ericsson Review, 3:4–9, 2008.

[15] Corici, Marius, Ancuta Onofrei, Dragos Vingarzan, Julius Mueller, and Thomas Magedanz: *Massive deployment of small coverage area cells in the*

all-ip wireless system - opportunities, issues and solutions. In *Proc. Technical Symp Telecom World (ITU WT) at ITU*, pages 147–152, 2011.

[16] Fletcher, S.: *Green Radio: Sustainable Wireless Networks*, February 2009. www.mobilevce.com/dloads-publ/mtg284Item_1503.ppt.

[17] Floyd, Sally and Van Jacobson: *Random Early Detection (RED) gateways for Congestion Avoidance*. IEEE/ACM Transactions on Networking, 1(4):397–413, 1993.

[18] Forde, T. K., I. Macaluso, and L. E. Doyle: *Exclusive sharing & virtualization of the cellular network*. In *Proc. IEEE Symp. New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pages 337–348, 2011.

[19] Frisanco, T., P. Tafertshofer, P. Lurin, and R. Ang: *Infrastructure Sharing and Shared Operations for Mobile Network Operators: From a Deployment and Operations View Communications*. In *ICC'08*, pages 2193–2200, 2008.

[20] Goldsmith, Andrea: *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005, ISBN 0521837162.

[21] Hoymann, C., A. Racz, N. Johansson, and J. Lundsjo: *A Self- backhauling Solution for LTE-Advanced*. In *WWRF*, 2008.

[22] Jianhui Li, Florian Roemer and Martin Haardt: *Efficient relay sharing (eresh) between multiple operators in amplify-and-forward relaying systems*. In *Proc. IEEE 4th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP2011)*, San Juan, Puerto Rico, Dec. 2011.

[23] Khan, A., W. Kellerer, K. Kozu, and M. Yabusaki: *Network sharing in the next mobile network: Tco reduction, management flexibility, and operational independence*. 49(10):134–142, 2011.

[24] Lempiainen, Jukka and Matti Manninen: *Umts Radio Network Planning, Optimization and Qos Management: For Practical Engineering Tasks*. Kluwer Academic Publishers, Norwell, MA, USA, 2004, ISBN 1402076401.

[25] Lescuyer, Pierre and Thierry Lucidarme: *Evolved Packet System (EPS): The LTE and SAE Evolution of 3G UMTS*. Wiley, March 2008, ISBN 0470059761. http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0470059761.

[26] Lin, Y., L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi: *Wireless network cloud: Architecture and system requirements*. IBM Journal of Research and Development, 54(1), 2010.

[27] McKeown, Nick, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner: *Openflow: enabling innovation in campus networks*. SIGCOMM Comput. Commun. Rev., 38:69–74, March 2008, ISSN 0146-4833. http://doi.acm.org/10.1145/1355734.1355746.

[28] Molleryd, Bengt G.: *Mobile network sharing - Swedish experiences.* PTS, May 2010.

[29] Nadas, S., S. Racz, Z. Nagy, and S. Molnar: *Providing Congestion Control in the Iub Transport Network for HSDPA.* In *Proc. IEEE Global Telecommunications GLOBECOM*, pages 5293–5297, 2007.

[30] Naveh, T.: *Next-Generation Mobile Backhaul Requirements as outlined by the NGMN Alliance, White Paper.* Ceragon Networks, May 2009.

[31] Nawrocki, M. J., M. Dohler, and H. Aghvami: *Understanding UMTS Radio Network Modelling, Planning and Automated Optimization, theory and practice.* John Wiley & Sons, 2006.

[32] Peters, S. W., A. Y. Panah, K. T. Truong, and R. W. Heath: *Relay Architectures for 3GPP LTE-Advanced.* EURASIP Journal on Wireless Communications and Networking, 2009:14, 2009.

[33] Qiu, Qinlong, Chen Jian, Ping Lingdi, and Pan Xue-zeng: *Hierarchy Virtual Queue Based Flow Control in LTE/SAE.* In *Proc. Second Int. Conf. Future Networks ICFN '10*, pages 78–82, 2010.

[34] Research Pyramid: *Demystifying Opex & Capex Budgets - Feedback from Operator Network Managers.* `http://www.researchandmarkets.com/reports/448691/demystifying_opex_and_capex_budgets_feedback`, February 2007.

[35] TSG SA WG2 Architecture: *TS 23.251 V9.2.0; Universal Mobile Telecommunications System (UMTS); LTE; Network sharing; Architecture and functional description.* Technical report, 3GPP, 2010.

[36] Venkatesan, G.K. Kulkarni, K.: *Wireless backhaul for LTE - requirements, challenges and options.* 2nd International Symposium on Advanced Networks and Telecommunication Systems. ANTS '08., pages 1–3, 2008.

[37] Weerawardane, T., A. Timm-Giel, G. C. Malafronte, D. Gianluca, S. Hauth, and C. Gorg: *Preventive and Reactive Based TNL Congestion Control Impact on the HSDPA Performance.* In *Proc. IEEE Vehicular Technology Conf. VTC Spring 2008*, pages 2296–2300, 2008.

[38] Wong, William K., Helen Y. Tang, and Victor C. M. Leung: *Token bank fair queuing: a new scheduling algorithm for wireless multimedia services.* Int. J. Commun. Syst., 17:591–614, August 2004, ISSN 1074-5351. `http://portal.acm.org/citation.cfm?id=1067899.1067905`.

[39] Zhuang, Wei, Yung Sze Gan, Kok Jeng Loh, and Kee Chaing Chua: *Policy-based qos-management architecture in an integrated umts and wlan environment.* 41(11):118–125, 2003.