Universidade de Aveiro
2009

Departamento de Electrónica,Telecomunicações e Informática

Igor Alexandre Almeida Pais

Análise de Desempenho e do Comportamento do Utilizador em redes 3G

End User Behaviour and Performance Analysis in 3G Networks

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica do Dr. Amaro Fernandes de Sousa, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

o júri

presidente

Prof. Dr. Rui Luis Andrade Aguiar
Universidade de Aveiro

Prof. Dr. Carlos Manuel da Silva Rabadão
Esc. Sup. De Tecnologia e Gestão do Inst. Politécnico de Leiria

Prof. Dr. Amaro Fernandes de Sousa
Universidade de Aveiro

agradecimentos

À minha namorada Lara e à minha família pela motivação e suporte incansável que sempre demonstraram.
Ao Professor Doutor Amaro Sousa pelo seu essencial contributo para este trabalho.
À minha empresa, Nokia Siemens Networks, pelo espiríto inovador que me foi habituando e por tornar possível este estudo.

palavras-chave Comportamento dos utilizadores, percepção da qualidade de um serviço por parte dos utilizadores, qualidade de serviço, métricas de desempenho da rede

resumo A Qualidade de Serviço (QoS) é uma preocupação para os operadores, mas devido à evolução da rede para um enorme número de serviços com requisistos diferentes, garantir uma boa QoS não é exatcamente sinónimo de utilizadores satisfeitos. A percepção da qualidade de serviço por parte dos utilizadores (QoE) garante aos operadores uma visão do grau de satisfação do utilizador final. O objectivo de uma boa QoS deve ser promover uma melhor QoE nos utilizadores. A QoE permite aos operadores saberem de que forma é que as condições da rede satisfazem as expectativas dos seus utilizadores em termos de confiança, disponibilidade, escalabilidade, velocidade, desempenho e eficiência.

O objectivo deste trabalho é o desenvolvimento de mecanismos que permitam aos operadores analisarem ao mesmo tempo o comportamento dos utilizadores e o estado da rede em termos de qualidade numa determinada região. Com este tipo de informação disponível os operadores podem adaptar os mecanismos de QoS da rede de modo a prencherem na totalidade as expectativas do utilizador final numa determinada região.

keywords End User Behaviour, Quality of Experience, Quality of Service, Network Performance, Key Performance Indicators

abstract The Quality of Service (QoS) is already a major concern for operators, but things are changing and, although in many cases better QoS results in better Quality of Experience (QoE), fulfilling the required performance parameters is not a synonym of satisfied users. QoE conditions can immediate response on the user satisfaction and thus the goal of QoS assurance should be to promote a better QoE. This will give the operator a deeper sense of the contribution of network's performance to the overall level of customer satisfaction in terms of reliability, availability, scalability, speed, accuracy and efficiency. The main goal of this work is to provide operators with mechanisms for end user behaviour analysis and at the same instant detailed network status. With this information operators know the end users behaviour in a certain region, know in detail network performance metrics and can adapt QoS mechanisms to fulfil end users expectations.

Document Index

Figures Index

Tables Index

1. Introduction

1.1. Motivation

Telecommunication networks are evolving gradually along the years and mobile networks have evolved from a voice-only oriented service to an enormous pool of multimedia data services. The increase of interest on such types of data is becoming an important component in operators' business models [1], [2]. This diversity of available services and applications with multiple and heterogeneous requirements imposes new challenges to operators for managing the performance of their networks and applications.

Current network performance tools are based on measuring and displaying Key Performance Indicators (KPI), which define network performance metric values. With these indicators, it is possible to know how a service is being provided and supported by the network. Nevertheless, such indicators do not give us enough information on how the services are being received by the end user and, most important of all, how is the end user satisfaction with each service.

The Quality of Experience (QoE) is used to describe the end user perception of the performance of a service and to provide the best possible usability in a cost-effective way. Numerous technical factors affect the end user experience and these factors are usually grouped in usage, accessibility, integrity and retainability. In order to deliver a high level of QoE, operators must understand the factors that contribute to the user perception of the target services, and apply that knowledge to adjust network parameters.

Two practical approaches can be used to measure QoE: realizing the satisfaction related with a service by using samples of population, or guessing the QoE based on the Quality of Service (QoS) parameters monitored in the network. The first approach can be done by inquiring the clients about important issues related with a service. The second approach uses values from some available Key Performance Indicators (KPI) related with QoS and tries to map them into QoE values. In this work, the end user behaviour is characterized and network performance indicators are achieved in order to understand both network and end user perspective. Monitoring the end users behaviours allows an operator to adjust the network parameters in order to fulfil different needs from different groups of users. Prediction algorithms can be created in the network to be configured accordingly to a profile of a pool of users previously classified.

Operators need to deal with end user behaviour and network performance at the same time. The main motivation of this work is to provide operators with mechanisms to analyse jointly user behaviour and network performance.

1.2. Objectives

As stated previously, the main goal of this study is to develop a methodology suitable for operators to monitor the end user behaviour at the same time that network performance and capabilities are known.

The objective is to develop mechanisms to understand user behaviour, service quality perceived by the end-user and have a detailed view of network status.

To do so, it is necessary to develop capabilities to know which services are most used and what is the end user opinion about these services. It is also important to create metrics to classify the quality of these services in an end user perspective and to know what end users think of this quality.

Detailed network status should be also monitored by the operator. So, a full set of metrics from the network side that can cause impact in service delivery performance are specified. In resume, the objectives for this work are:

- Develop mechanisms for monitoring end users behaviour (most/less used applications/services, average times of usage, …);
- Understand end users expectations and satisfaction by knowing the end users perspective of the quality of a service;
- Build network performance metrics to achieve detailed network status and understand which are the causes for the degradation of QoS (Quality of Service) when it happens

1.3. Organization of the Dissertation

Besides this Introduction chapter, this dissertation is organized in 8 chapters.

Chapter 2 introduces 3G network architecture, elements, interfaces and protocols. Also a brief comparison between present and future mobile networks characteristics is performed. After this introduction is described a 3G network performance management system and the importance of such a system to an operator. Chapter 2 is finalized with the illustration of the function and examples of KPIs (Key Performance Indicators), reports and a troubleshooting use case example.

Chapter 3 intends to introduce User Behaviour analysis, Quality of Service (QoS) and Quality of Experience (QoE). This chapter shows briefly some of the available services, describes QoS mechanisms, QoS attributes and identifies some of the challenges of QoS management in a mobile network. In chapter 3 could also be perceived how to achieve QoE, understand the relationship between QoE, QoS and End user experience and see an example of user profile classification.

In chapter 4 is presented User Behaviour monitoring tool architecture, user behaviour models, service quality reports and service based agent KPIs.

Chapter 5 describes Network reporting tool and shows network KPIs specification.

In chapter 6 is illustrated the architecture proposal to accomplish the initial objectives and the required extensions.

Chapter 7 explain implementation issues and shows the acquire results for end user behaviour and network performance metrics.

Finally, in chapter 8, conclusion of the work done and future directions are described.

2. 3G Network Performance

This chapter provides an overview of the 3G network architecture, protocols and interfaces. It also introduces some network performance concepts such as KPI (Key Performance Indicator) and network performance reports. The main goal of this chapter is to present a brief description of network performance providing the key concepts necessary to understand the following chapters.

2.1. 3G Network Description

This first section introduces 3G Network architecture, protocols and interfaces.

2.1.1. The UMTS PLMN Subsystems

UMTS implements a Public Land Mobile Network (PLMN). Fig 1 shows the subsystems of the UMTS public land mobile network and their system units.

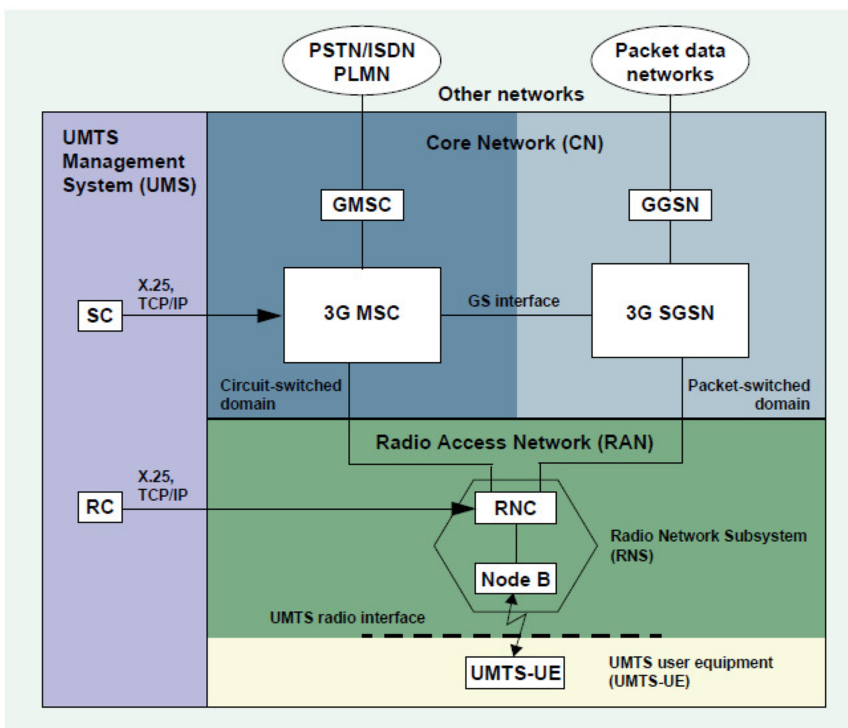


Figure 1 - Subsystems of the PLMN and their Network Elements

The following topics summarize the subsystems present in the figure and describe their functions.

A. UMTS User Equipment (UE)

The UMTS-UE provides user access to network services by supporting all the operating functions for subscribers.

B. UMTS Terrestrial Radio Access Network (UTRAN)

The UTRAN consists of one or more Radio Network Subsystems (RNS):

B.1. Radio Network Subsystem (RNS)

The RNS provides all the transmission and control functions that are necessary for radio coverage of the service area. It includes:

B.1.1. one or more base transceiver stations (Nodes B), distributed throughout the entire service area;

B.1.2. the Radio Network Controller (RNC) to which each Node B is connected.

B.2. Core Network (CN)

The CN consists of a:

B.2.1. Circuit-switched (CS) domain:

The circuit-switched domain provides all circuit-switching functions for the UMTS PLMN, those specific to mobile radio as well as those necessary for the fixed network. The CS system consists of a MSC (Mobile-services Switching Centre) and a GMSC (Gateway MSC).

The circuit-switching functions are required for:

– Independent operation of the UMTS PLMN
– Operation of the UMTS PLMN in conjunction with a fixed network (PSTN/ISDN)
– Another circuit-switched PLMN

The circuit-switched domain can also manage complex databases (subscriber data, Network data) and handle the various signalling protocols used for establishing and clearing down connections.

B.2.2. Packet-switched (PS) domain:

The packet-switched domain provides all packet-switching functions for UMTS PLMN, those specific to mobile radio as well as those necessary for the fixed network. The PS domain consists of a SGSN (Serving GPRS Support Node) and a GGSN (Gateway GPRS Support Node).

The packet-switching functions are required for:

– Independent operation of the UMTS PLMN
– Operation of the UMTS in conjunction with a fixed network (Internet)
– Another packet-switched PLMN

The packet-switched domain can also manage complex databases (network data) and handle the various signalling protocols used for establishing and clearing down connections.

C. UMTS Management System (UMS)

The UMS provides all functions relevant for remote and local operation of the UMTS network in addition of recording the information about system performance.

The UMS requires:

- processing, memory and supervision equipment at a central location

- a Switch Commander (SC) or Radio Commander (RC) or a Management framework
- other local operation and maintenance equipment, either fixed or portable

The SC or RC represents the element management level. It may be complemented by a network management level and a service management level.

2.1.2. Layers and Protocols

The UMTS network consists of two independent service layers, the Access Stratum (AS) and the Non-Access Stratum (NAS), which correspond to the logical division of functions within the network.

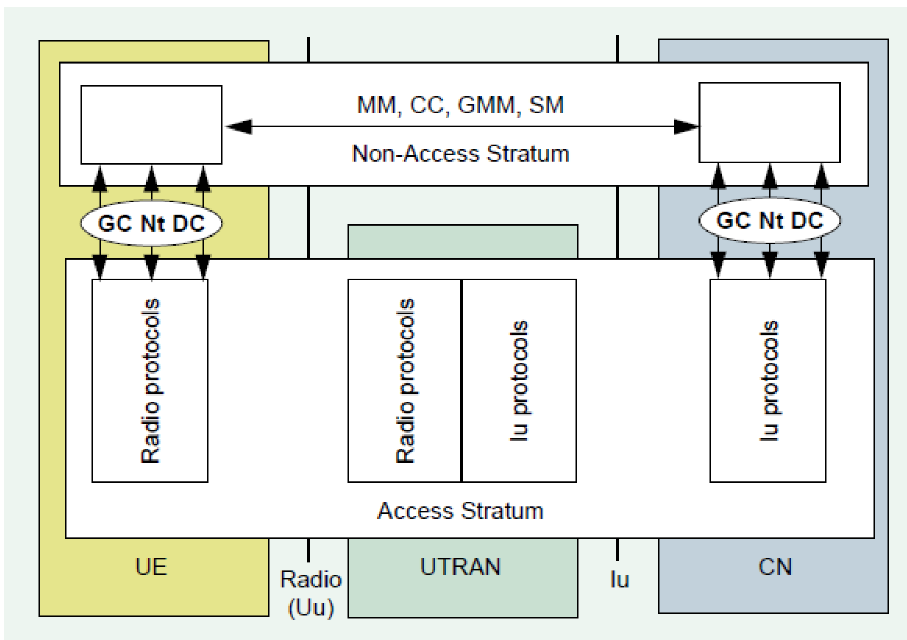


Figure 2 - Layers and protocols for AS and NAS

The Access Stratum comprises all functions relating to the access network of UMTS, which is, in compliance with the 3GPP specifications, completely integrated in this layer. As shown in Fig 2 it also comprises a part of the UE (the one, which manages the protocols of the radio Uu interface) as well as a part of the CN (the Iu interface).

The Non-Access Stratum comprises all other functions of the UMTS network that are independent of the access network, such as functions for:

– Connection establishment, which correspond to the protocol layers CC (Call Control) for CS calls and SM (Session Management) for PS services

- Mobility management in standby mode, which correspond to the layer protocols MM (Mobility Management) in CS mode and GMM (GPRS Mobility Management) in PS Mode

Table 1 shows the division of UMTS functions to the separate layers.

Table 1 - AS and NAS functions

UMTS Functions	Access Stratum	Non-Access Stratum
Call Processing		x
Authentication		x
Handover	x	
Management of additional services		x
Management of radio channels	x	
Ciphering	x	(x)
Compression	x	(x)
Billing functions		x

The UMTS functions ciphering and compression are represented in both layers. In the standard they are described as part of the Access Stratum, but they can alternatively belong to the Non-Access Stratum (in the table marked in brackets). The Access Stratum acts as a service provider for the Non-Access Stratum. A number of connections, so-called Service Access Points (SAPs), are specified between the AS layers and the NAS layers in the UE and in the CN. These SAPs allow for the classification of interactions between NAS and AS depending on the type of services offered or requested.
There are three such access points:

- GC (General Control);
- Nt (Notification);
- DC (Dedicated Control).

2.1.3. The Radio Network Subsystem

The RNS is divided into an intelligent, centralized controller (the Radio Network Controller - RNC) part and several transceiver stations (Nodes B) as it is shown in Fig 3. This structure is well adapted to both small-cell networks, as preferably used in urban areas, and large-cell rural networks. The advantage of small-cell networks is the internal handover offered by the RNCs, higher capacity and throughput, while the advantage of large-cell networks is to obtain more coverage with less Network Elements.

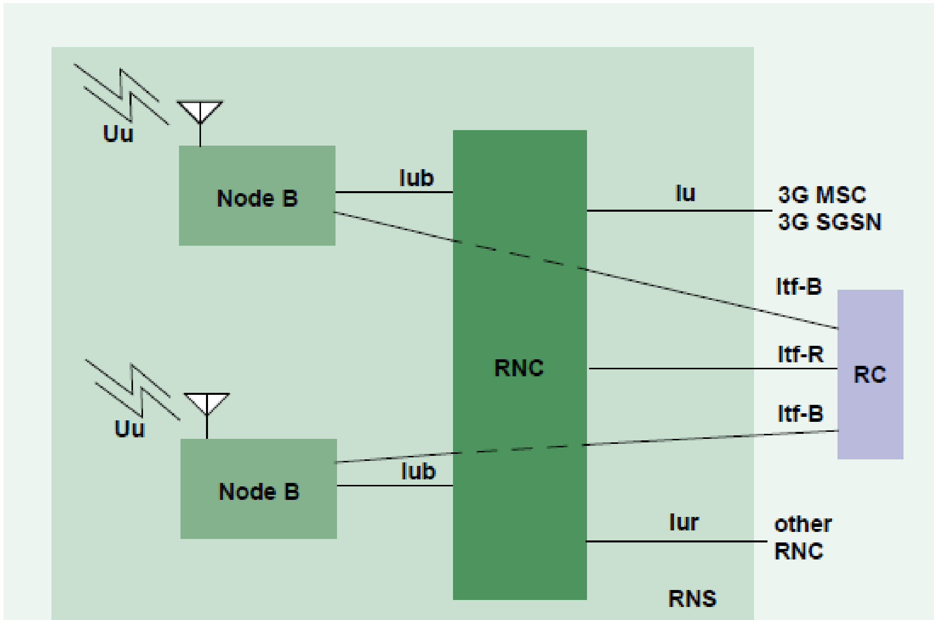


Figure 3 - Radio Network Subsystem

Node B is a logical node and responsible for the radio transmission to the User Equipment (UE) and the radio reception from the UE. Each Node B can serve a variable number of radio cells (e.g. a 2/2/2 configuration means to support 6 cells). Nodes B are distributed throughout the entire radio service area. Each RNC supports at least one Node B, though generally more Nodes B are supported. Node B terminates the Iub interface towards the RNC.
The RNC controls Node B and is responsible for core processing in the UTRAN. Up to three RNCs are linked to a 3G MSC or 3G-SGSN. The RNCs can be physically grouped together at a central point on 3G-MSC or 3G-SGSN sites or remotely in a shelter or in a confined space.

2.1.4. The Concept of Network Area

This section describes the concept of “area” in a UMTS system and some mechanisms associated. Concept of “area” is important because all performance metrics are related with a specific “area” in the network and subscriber location perception is only possible because of this “area” concept. Subscriber location and registration mechanisms are described next.

Fig. 4 shows the UE registration and connection principles within the UMTS. AS it is possible to observe in the figure, subscriber data delivered from the home location register (HLR) and mobility data delivered by Mobility Management (MM) functions are temporarily stored in the visitor location register (VLR, CS domain) or the subscriber location register (SLR, PS domain). The HLR contains data on subscription restrictions, services assigned to the mobile subscribers, and the current subscriber status including information on the current location. For both CS and PS there are three different logical states associated to UE:

- DETACHED
- IDLE

- CONNECTED

These states are performed trough an RRC connection (connection between the RRC (Radio Resource Control) entities on the UE and the RNC) and the respective RRC connection sets are saved in the SGSN/SLR and the MSC/VLR.

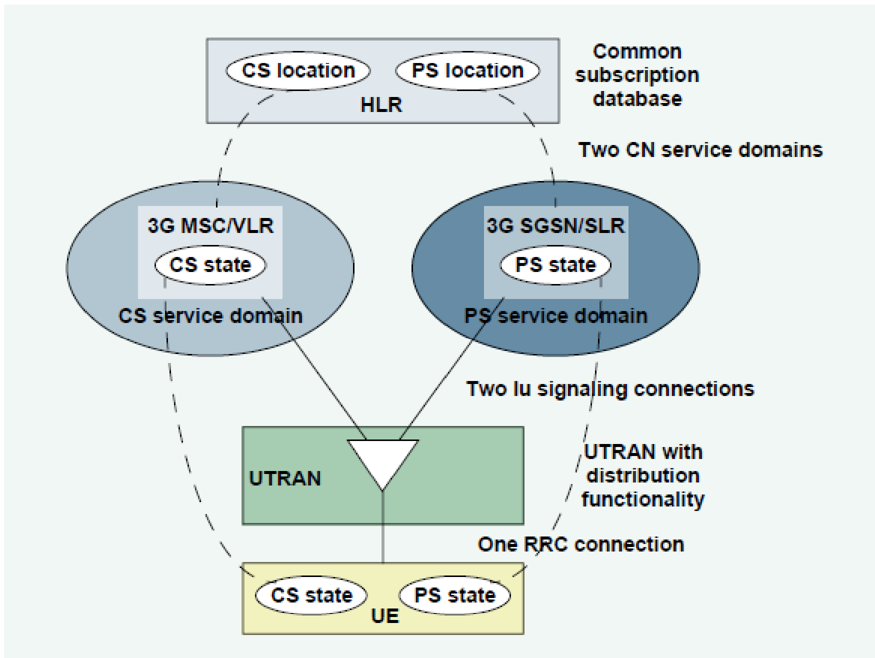


Figure 4 - UE connection and registration for MSC and SGSN

MM can be divided into the following functions:

- Update of the UE location in the CS or PS domain
- Establish or delete an MM context for a UE in the network node
- Page and search
- Update the subscriber database
- Protect against unauthorised service usage, e.g. authentication and service request validation, ciphering.

Fig. 5 shows the connection of UEs (3G) and MSs (2G) to the core network and the mobility management. This figure is useful to understand the possible flows of data between the mobile terminal and the HLR. The possible paths are:

- MS-BSS-MSC/VLR-HLR (2G CS connection);
- MS-BSS-SGSN/SLR-HLR (2G PS connection);
- UE-RNS-MSC/VLR-HLR (3G CS connection);
- UE-RNS-SGSN/SLR-HLR (3G PS connection).

Base Station Subsystem (BSS) is the equivalent of the 2G system for the RNS in the 3G system.

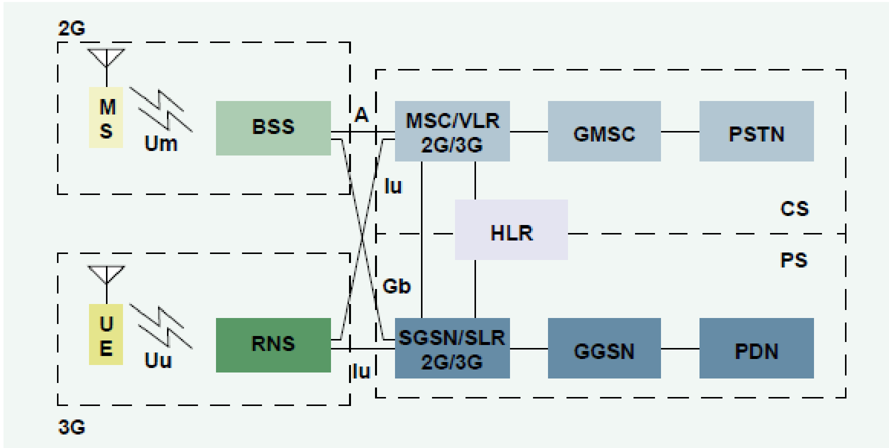


Figure 5 - Connection of User Equipment and Mobile Station to Core Network and mobility management

2.1.4.1. Location Areas and Routing Areas

For locating a subscriber, the network is divided into:

- . Location areas for CS services
- . Routing areas for PS services

The location or routing area is used, for example, during paging procedure (described bellow in 2.1.4.3), a temporary identity is assigned to the UE. This identity is unique within a location area or routing area. In Figure 6 are shown the area concepts associated to a 3G network: location area and routing area.

The mapping between location area and RNC is handled within the MSC/VLR to which the location area is assigned. The mapping between routing area and RNC is handled within the SGSN/SLR that owns this routing area. The RNC handles the mapping between location area and cells and between routing areas and cells.

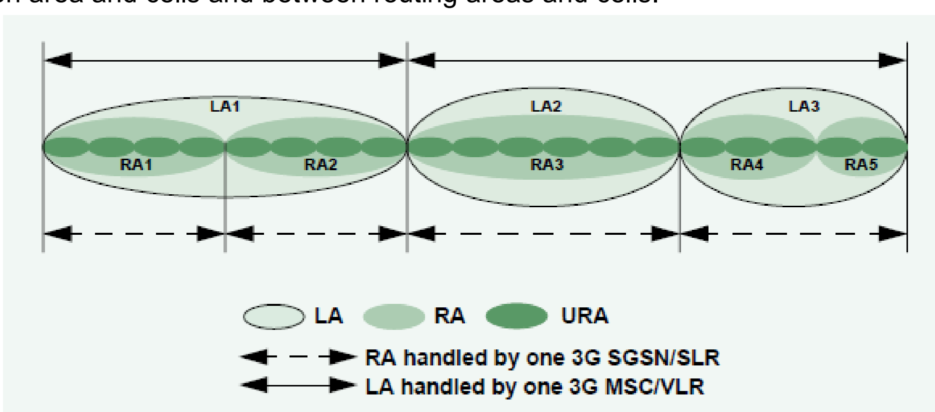


Figure 6 - Area concepts (cells are not shown)

2.1.4.2. Location Area Update and Routing Area Update

A UE invokes a location area update procedure via the 3G-MSC if it changes the location area or if a certain timer has expired. If a new location area is in an area served by another Core Network (CN), the location area update also triggers the registration of the subscriber in the new CN node and a location update for CS services towards the HLR.

2.1.4.3. Paging

Paging information is transmitted to an individual UE in Idle mode (Idle mode happens when there is no connection between the UE and the UTRAN, in Idle mode the UE is not involved in any active session) using the paging control channel (PCCH). A paging message sent by the RNC contains the paging area ID parameters, i.e. the information on the area in which the paging message is to be broadcast.
The location area or routing area is taken from a cell identifier list. If a UMTS cell is paged, the cell identifier list contains only one dummy cell from which to derive the location area. The RNC itself creates the list of cells to be paged in. Paging is completely independent for CS and PS services.

2.1.4.4. Location Services

Location services (LCS) allow the determination of the geographical location of a UE. This information may be requested by the UE itself or by a client (application running in the other user's mobiles or in the network) within or attached to the CN. Location services are the basis for new services such as:

- Localised advertising, tracking services (e.g. fleet management), navigation;
- Location-dependent billing;
- Enhanced support of emergency calls by determining the originator's location;

2.1.5. Interfaces

The Network Elements (NEs) within UMTS are connected via defined logical and physical interfaces based on WCDMA and ATM.
Fig. 7 shows the adopted interfaces between the User Equipment (UE) and the Radio Network Subsystem (RNS), within the RNS, as well as between the RNS and the Core Network (CN).
For multipoint to multipoint communication, Fig. 7 shows:

- the Intra-Mobile Communications Network Link (M-M Link)
- the Mobile Communications Network-Landline Telephone Network Link (M-L Link)

Fig. 7 is important to understand some topics in the following chapters because interfaces performance is analyzed in detail to guarantee the overall system efficiency.

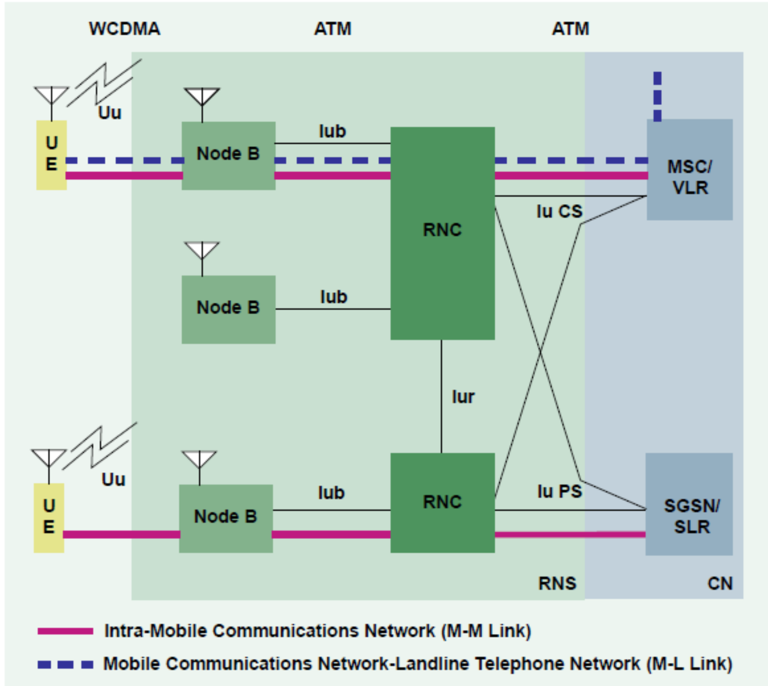


Figure 7 - Interfaces for multipoint to multipoint communication

2.1.6. Interface Protocol Structure

The interface protocol architecture consists of two horizontal layers, see Fig. 8:
Radio Network Layer – defines procedures relating to the operation of Node B. All UTRAN-related issues are visible in this layer.
Transport Network Layer – defines procedures for establishing physical connections between Node B and the RNC. This layer provides transport-related technologies.

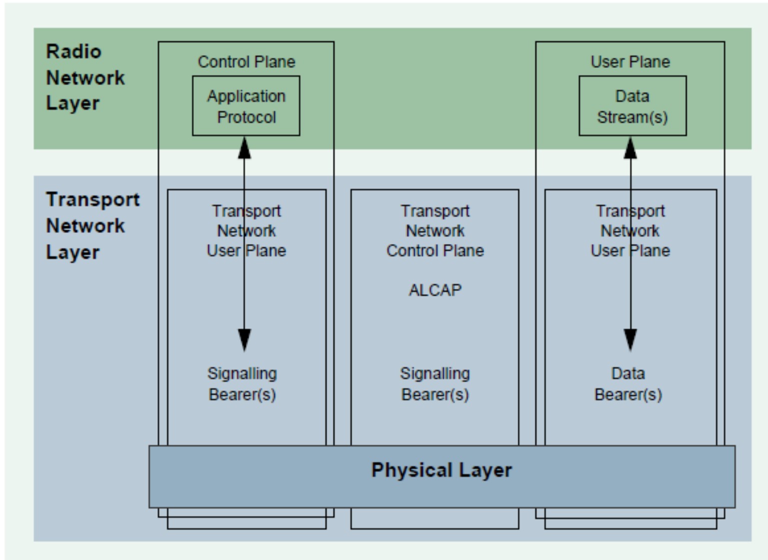


Figure 8 - Protocol stacks in the Radio Network and the Transport Network Layer

In addition, the protocol stack of an interface consists of 3 vertical layers:

- **User plane**

The user plane transports all user data including data stream(s) and data bearer(s). It supports both circuit-switched domain and packet-switched data transport protocols.

- **Control plane**

The control plane provides UMTS-specific control signalling including the application protocols, i.e. RANAP, RNSAP and NBAP, and the signalling bearer for transporting the application protocol messages. It supports: (i) control signalling protocols for circuit-switched/packet-switched service management, user management and resource management and (ii) transport signalling protocols for the allocation of the bearers between the RNC and the 3G-MSC in the case of a circuit-switched domain.

- **Transport network control plane**

The transport network control plane provides all control signalling within the transport layer. This plane mediates between the control plane and the user plane to keep the application protocol of the control plane independent of the technology selected for the data bearer in the user plane.
Several protocol stacks reflect the data transfer via individual interfaces between the UMTS network elements. These interfaces are defined by 3GPP technical specifications. For example, Fig.9 shows the voice communication protocol stack for the control plane in an M-L link (described before in section 2.1.5), while Fig.10 shows the corresponding user plane.

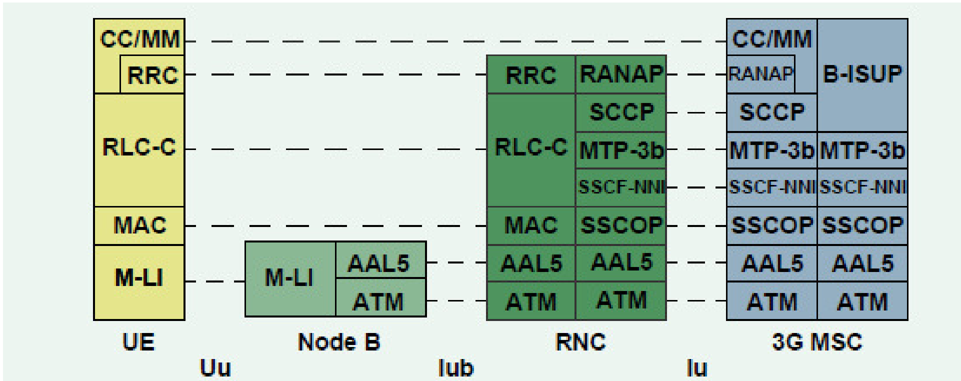


Figure 9 - Voice communication (C-plane) protocol stack in M-L Link

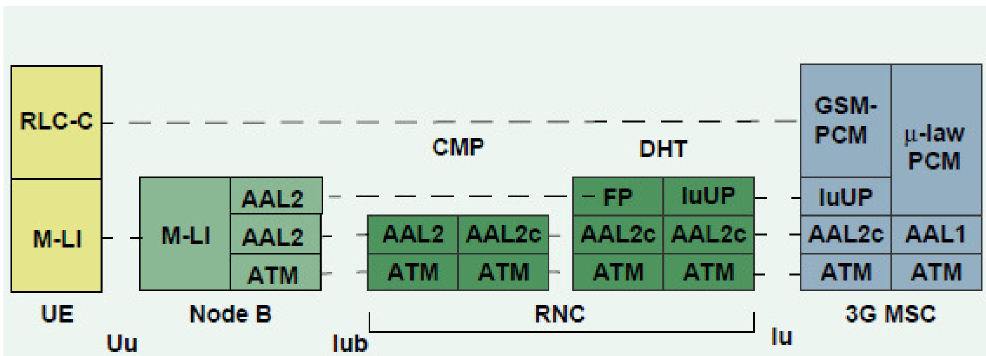


Figure 10 - Voice communication (U-plane) protocol stack in M-L Link

2.1.7. Future Mobile Networks

This section aims to provide a brief resume on how mobile networks are planned to evolve in the future, what are their future characteristics and what will change in the current architecture.

Historically, UMTS was using W-CDMA (Wideband Code Division Multiple Access) in the air interface. This technology evolved to HSDPA (High Speed Downlink Packet Access) also named 3.5G that introduces higher data rates and capacity in the downlink of the UMTS system. Afterwards, HSUPA (High Speed Uplink Packet Access) introduces higher data rates in the uplink part of the network. These two protocols HSDPA and HSUPA together are jointly designated as HSPA which consists of an enhancement to W-CDMA protocols.

A new evolution of HSPA, known as HSPA+ or i-HSPA, is a standard defined in 3GPP release 7. The main difference to the traditional HSPA (High Speed Packet Access) is the flatness architecture and the use of all-IP that provides higher data rates. I-HSPA is many times denominated as pre-LTE technology because of the similarities in the flat architecture.

LTE (Long Term Evolution) is the 4th generation or 3GPP release 8 of mobile networks and was designed to increase data rates of i-HSPA system. A huge amount of

investments into researching the deployment of LTE is now carried by network elements vendors.
On the other end, WiMAX (Worldwide Interoperability for Microwave Access) is a wireless technology with multiple transmission modes based in IEEE 802.16 and companies are evaluating this technology for the "last-mile" connectivity in very specific areas.

Table 2 presents a comparison between actual and future mobile networks technologies.

Table 2 - Peak data rate, Latency, Architecture and Services Comparison for 3G and next technologies

	WCDMA	HSPA R6	i-HSPA	WIMAX	LTE
Peak data rate (DL/UL)	384/384 Kbps	14/5.7 Mbps	28/11.5 Mbps (rel7) 43/11.5 Mbps (rel8)	40/10 Mbps	173/58 Mbps
Latency	100-200 ms	40-60 ms	25 ms	30-50 ms	10-20 ms
Architecture	RNC based	RNC based	Flat	Flat	Flat
Services	Cs and high speed PS	Broadband PS	Broadband PS and CS over HSPA	PS only, VOIP	PS only, VOIP

From the table, it is possible to understand the huge increase in the peak data rate provided by the envisaged future technologies, together with a reduction of the latency and also the tendency for a flat architecture and packet-switched (PS) only services.

Fig 11 shows the evolution in mobile network architecture by reducing the number of nodes, which permits to achieve a TTI (Time Transmission Interval) of 2 ms for i-HSPA and 1 ms for LTE.

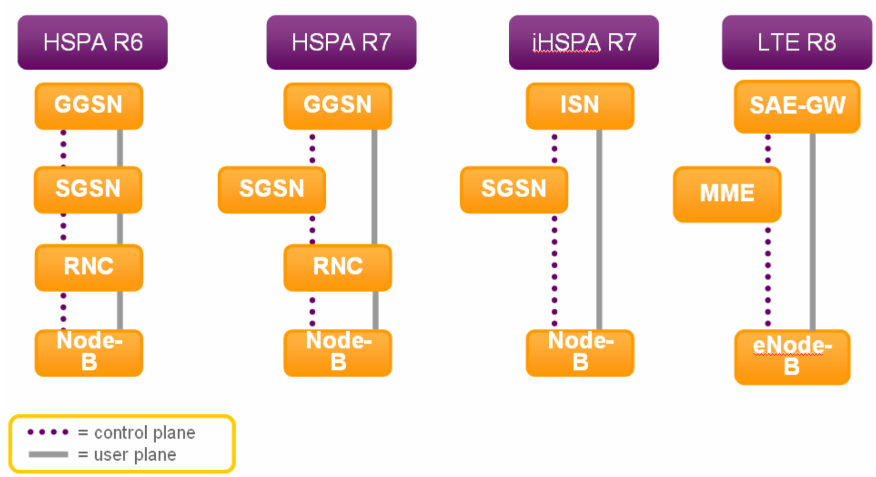


Figure 11 - Architecture evolution from HSPA to LTE

2.2. 3g Network Performance Introduction

This section describes the key concepts of performance management (PM), explains why it is needed and gives an introduction to KPI (Key Performance Indicators) and Reports. At the end, it is described a use case of 3G network performance analysis and improvement.

2.2.1. Purpose of Performance Management

The aim of any performance management activity is to collect data to support the following activities:

- Verify the physical and logical configuration of the telecommunications network;
- Monitor continuously how the network is working;
- Localize potential problems as early as possible;
- Monitor subscriber behaviour;
- Provide optimum services to mobile subscribers.

Depending on how the performance management applications use the collected data, they can be divided into two main types: performance monitoring and performance reporting applications.

Performance monitoring is online-oriented and provides almost real-time information of the network. Therefore, performance monitoring applications generally use shorter measurement output intervals because fast reaction is needed in severe problem cases, in which a 24-hour interval would be too long. The outputs of performance monitoring applications are mainly used as additional information for problem cases in which no alarm information is available. Performance monitoring rules are intelligent threshold rules and can be calculated before or after the measurements are stored in the database.

Performance reporting is offline-oriented and provides information on what happened in the network over a certain period of time. Performance reporting applications mainly rely on performance indicators and produce reports, which can be used, for example, when troubleshooting, planning and/or optimising the network.

2.2.2. Using the Performance Information

Network elements produce data to support performance evaluation, for example on the following areas:

- Traffic levels within the PLMN, including the level of both the user traffic and the signalling traffic
- Verification of the network configuration
- Resource access measurements
- Quality of Service (QoS)
- Resource availability

It is possible to use various types of performance information for monitoring and managing the network.

When monitoring a network to manage its performance, data is collected for locating potential problems as early as possible and verifying the physical and logical configuration of the network. It is also possible to use performance data to monitor subscriber behaviour by charting out the usage of different services that are available to end users. Such information provides input for business and service management decisions during the optimization and expansion of the network.

Networks continuously produce an enormous amount of performance data. Performance Management (PM) applications need to filter the information to suit the needs of various users. The following aspects can be identified for reporting needs:

- End user perception of the services about traffic, quality, or availability of the services.
- Technical considerations in the network, for example monitoring the radio access and the core network separately.
- Management and operational levels at the operator.

Fig. 12 shows these reporting needs in a combined model. Each chunk of the pyramid corresponds to a report set or report group that PM applications can provide to the various user groups.

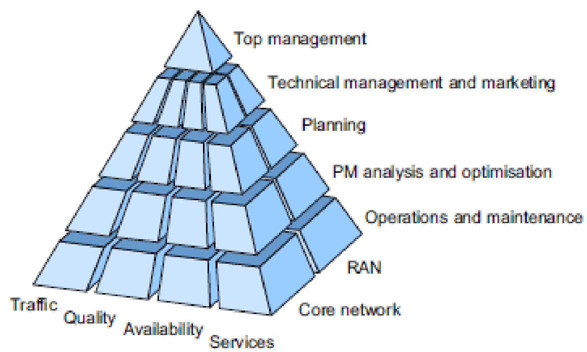


Figure 12 - Network Management level

The following table describes examples of how various user groups can use performance measurement data for managing the network and what kind of PM applications they can use.

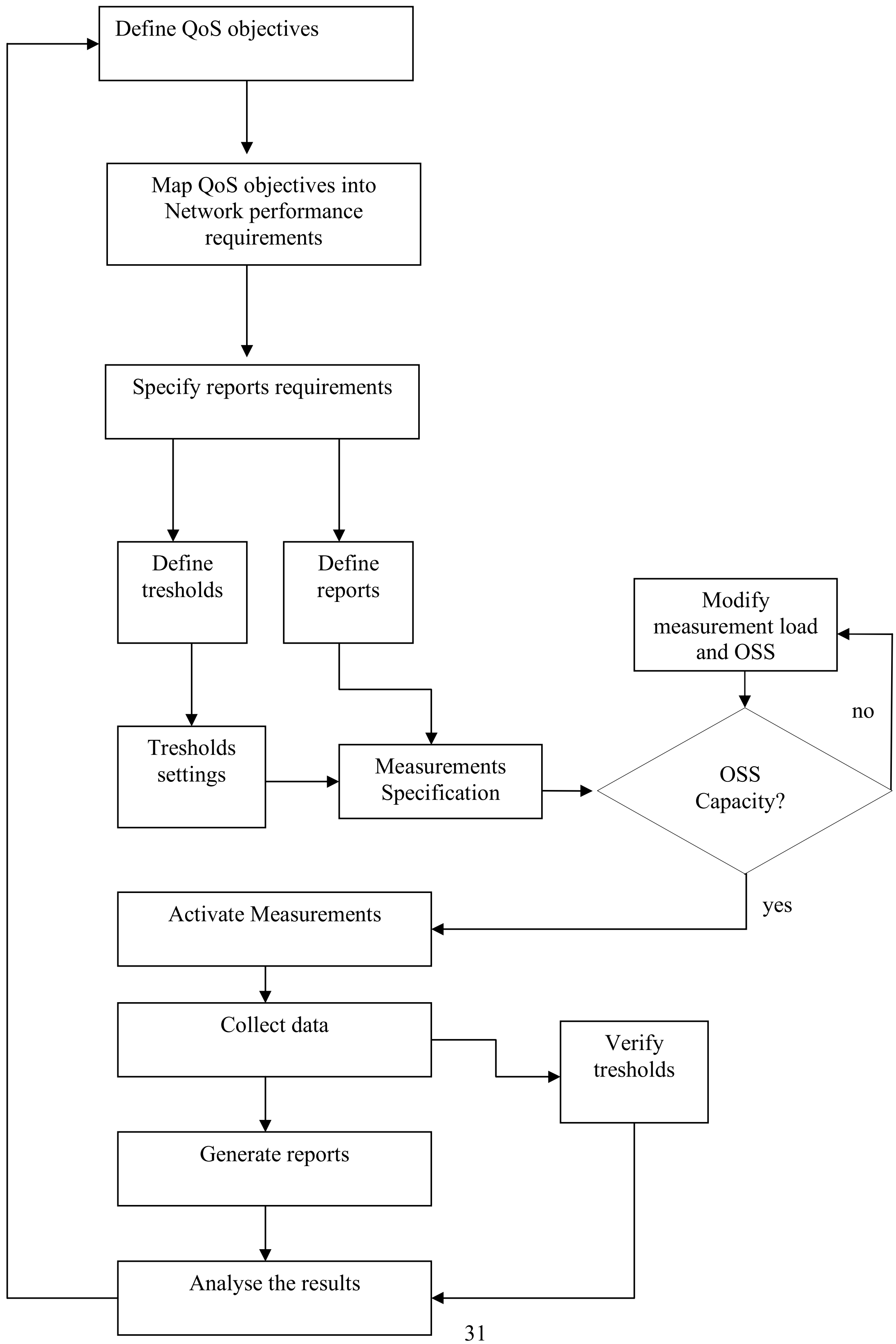
Table 3 - Management performance by groups of profiles

Group	Objective	Report Specification	Report characteristics
COO/Management	▪ Ad-hoc reports ▪ Future trends forecasting ▪ Overall status of network services	▪ PLMN level ▪ Trends ▪ Offline reports	▪ Long periods reporting ▪ Tendencies ▪ Graphical reports

Marketing	▪ End user behaviour ▪ New introduction opportunities analyzing current trends	▪ High-level information with possibility to drill down ▪ Check the most important metrics for each service	▪ Weekly and monthly reports ▪ Graphs and tables ▪ NodeB level is important ▪ Service usage statistics
O&M	▪ Fault detecting with alarm capabilities ▪ Localization of areas with problems ▪ Root-cause analysis	▪ Low-level ▪ Fast switch between different NE ▪ Fault management reports	▪ Daily reports ▪ Text and graphical
Performance, Optimization and Planning	▪ Long term reports evaluating network resources ▪ Quality of different services in detail ▪ Coverage study with PI of HO failures and geographic distribution	▪ Traffic profiles ▪ Network usage ▪ QoS monitoring for each service ▪ QoE KPI's ▪ Log-term traffic profiles ▪ Benchmark KPI's	▪ Tendency reports ▪ Raw data in some cases ▪ Text and graphical
Customer assurance	▪ Analyze customer service complaints and identify network problems	▪ Failing NE reports ▪ Daily status reports	▪ Fault management reports

2.2.3. 3G Network Performance Management Process

This section describes a view on the performance management process in order to provide a pre-defined QoS goal or to improve the existent in the network.



In the following points this diagram is explained in detail:

- **Defining or revising Quality of Service goals**

When managing a particular network, the Quality of Service (QoS) of the network means the effective maintenance of the quality for the network services according to some agreed criteria. Maintaining and improving the QoS of the network includes regular monitoring of the service performance and the evaluation of the service performance problems with respect to the targeted level of service delivery.

To improve the service, we need to define QoS goals that guide the evaluation of network and service performance. These goals can represent, for example, the point of view of the operator (efficient hardware usage, for example efficient use of network elements (NE's) considering the type and capacity of the NE's, their location and number to support the services) or the point of view of the subscriber (availability, speed, and accuracy of the service).

To develop and expand the network, it might be necessary to update and revise QoS goals. Mature operators competing for more market benefit from introducing new services to differentiate themselves from other service providers. At this point, all kinds of service and profile measurements are useful. The quality of service becomes extremely important and, for example, optimization measurements should be taken.

The following tasks belong to this stage:

  - Setting objectives for the desired level of quality
  - Measuring the QoS by monitoring real-time data and long-term trends
  - Planning corrective actions

- **Translating Quality of Service goals to network performance requirements**

To achieve the QoS goals that were defined, the detailed requirements for the network should be prepared or updated, including the requirements on network performance. For example, the quality of a voice service can be correlated with the number and ratio of dropped calls, while the accuracy of an IP service can be correlated with the number of received packets.

- **Specifying reporting requirements**

The network performance requirements serve as the basis for reporting requirements and provide information on which measurements we need to monitor with the Performance Monitoring (PM) applications.

- **Identifying other reporting needs**

Monitoring the day-to-day performance of the network is not the only reason for using performance reports. It is also needed to gather information from reports, for example, if we want to examine short-term changes when we are upgrading operator system or if we want to check long-term trends to gather information for optimizing or expanding the existing network.

If we can translate these reporting needs into measurements, we can directly define the reporting requirements for them and translate them into measurement activation criteria. However, some reporting requirements can give rise to a need for a new or an additional PM application.

We might also need a new or additional PM report when investigating uncommon situations or when upgrading or expanding the network.

Before we establish new reports, it is recommended to check whether the active measurements already provide the counter information that we need, or whether the required measurements are available but have not been activated yet, or whether it is sufficient to change the measurement settings to obtain the information that you need.

- **Defining reports**

  Reports are defined and scheduled based in previously established requirements.

- **Adapting the measurement load and the Operations Support System (OSS)**

  If a new measurement is needed, we have to ensure that the capacity of the OSS, for example, disk space and insertion time, is not exceeded when the new measurement is implemented and activated. If the OSS capacity is exceeded, we must perform one of several possible actions: we can adapt the entire measurement set by reducing the number of the measurements or increase the measurement intervals, or increase the Management system capacity (for example, with additional processors). If the capacity has not been exceeded we can proceed with initializing the additional measurements.

- **Initializing measurements**

  Activating all the necessary measurements across the network could impose a significant but unnecessary load on the database and the PM processes. To analyse this case, the following factors must be taken into consideration:

  - time period when measurements are active in the network element or when the measurement information is uploaded from the network element to the network management system
  - measurement intervals
  - number and distribution of the network elements for which the measurement is activated

- **Defining thresholds**

  Network management systems allow us to define thresholds for the measurements. When a measurement value exceeds a threshold, an alarm is raised. Therefore, we can follow the status of network performance with the fault management applications in which the alarm is visible and also with the performance management applications in which the measurement information is processed. The details of defining thresholds depend mostly on the particular OSS solution in operation.

- **Collecting data**

  The measurement information can be pre-processed in the network element, depending on the NE type and the configuration of the NE. If pre-processing is applied, the data is filtered (not all the measurement information is sent to the network management system) or some calculations are done on the measurement data before uploaded to the OSS through the NE-specific interface.

  The measurement data can also undergo further processing in a post-processing phase, for example, when busy hour information is needed for a report. The results of the post-processing are also stored in the OSS database. The data is used by the reporting applications or is exported through the open interfaces to external systems.

- **Monitoring network performance**

  We can typically find performance-related problems by checking the alarms, the measurements, and the customer complaint reports. We can also use information from drive tests and Call Detail Reports (files that contain detailed information about mobile control procedures during a call).

Regular reporting such as running a report once a day or once a week is too slow for continuous network monitoring. Furthermore, the default measurement interval for most counters, which is one hour, may create a huge delay. However, we can also analyze the traffic statistics and collect detailed information from the network for further analysis with reports when we discover a problem. Prompt reaction is sometimes required. Therefore, it is recommended to set thresholds for the performance indicators so that the OSS raises alarms based on the indicator values in unusual situations or when errors occur.

- **Generating and distributing reports**

When defining the reporting requirements, we also consider gathering information on the user groups (the envisaged types of persons to whom the reports are addressed to) and on the type and frequency of the reports that they need. We have to ensure that the users of the OSS have access to, or receive the reports that best support their work. Reporter applications allow us to turn the often overwhelming amount of measurement data into information on the performance of the resources in the network, which we can use for determining whether the network performance goals are met and whether problem-determination procedures should be initiated based on performance.

2.2.4.3G Network KPIs

Key Performance Indicators (KPIs) are the most important indicators of network performance. KPI reports allow the operator to detect the first signs of performance degradation and prevent the development of critical network problems. KPIs on the regional level can be used for analyzing performance trends, for example, on the cell level for troubleshooting specific cells.
The Performance Management database contains a very large number of counters which are all elementary performance indicators. For practical purposes, however, it can be difficult to interpret the indicators, because they often provide data on a very specific aspect of the network, and it may be hard to see what relevance the value of a counter has in terms of the network if you analyse the counter in an isolated way. Fig. 13 shows the process sequence to build a report. Normally network reports contain more than one KPI and KPIs are done based on Performance Indicators (PIs). These PIs are simple counters that belong to a measurement in the network. Measurements can include thousands of counters and, so, it is not a good approach analyze counter by counter. KPIs filter these PIs into the most important metrics from the network according to operator needs.

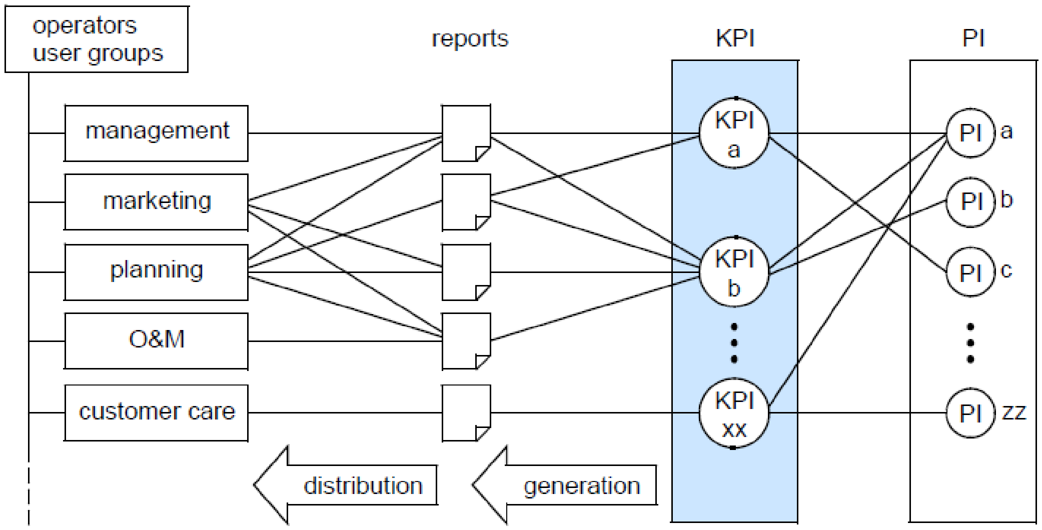


Figure 13 - Performance Indicators and Key Performance Indicators

KPIs are used for assessing the performance of the network and their formula should remain constant over time. KPIs use counters received from one or several measurements that are mapped directly to a counter or use a formula based on several counters.

To create a formula, all measurements that are input data to the formula must be activated and, in addition, these measurements should have the same output interval (in case the measurements with different intervals, they can be computed only on the basis of the longest interval).

KPIs are created and selected to belong to a group. For illustration purposes, Table 4 presents the definition of some groups that are used in this work and for each group, the set of defined KPIs. The use of groups for KPIs helps when building a report and we need to find a KPI related with a specific issue. If the user needs a KPI that is related with Accessibility of the network, for instance, he just needs to go to the Accessibility folder and see if the KPI is already available.

Table 4 - Examples of KPIs by groups

Groups of KPIs	KPIs
Accessibility	▪ Radio Resource Control (RRC) Connection setup Success Ratio ▪ Radio Access Bearer (RAB) Setup Access Success Ratio ▪ Radio link setup Success Ratio ▪ Call Setup Success Ratio
Retainability	▪ RRC Drop Ratio ▪ RAB Drop Ratio ▪ Radio Bearer Reconfiguration Failure Ratio ▪ Transport Channel Reconfiguration Complete Ratio
Integrity	▪ Share of MAC PDUs dropped due to retransmissions ▪ HSDPA congestion rate in Iub ▪ Percentage of Good Cell Quality Indicator (CQI) reports ▪ Average BLER (Block error ratio)

Mobility	▪ Intra-System Hard Handover Failure Rate ▪ Soft Handover Success Rate ▪ HSUPA Serving Cell Change Success Ratio
Usage	▪ Average of free Chanel Elements ▪ Bit rate utilization ▪ Peak number of HSDPA users in IADA ▪ HSDPA Throughput

Usually, the reporting tools use the groups presented in Table 4: Accessibility contains KPIs related with the first attempt to the network and access phase; Retainability contains KPIs that are related with the cases of calls failing after the access phase (call drops are usually in this group); Integrity contains KPIs related with transmission and retransmission errors and quality in the cells; Mobility contains KPIs covering all issues related with handovers and cell changes; finally, Usage contains KPIs showing the status of the capacity of the network.

2.2.5. 3G Network Reports

Performance Management (PM) reports are mostly used for two purposes: to gather information for troubleshooting (short-term and ad-hoc reports) and prevention, and for developing the network and the services (longer-term reports).

For example, the top management of an operator can use PM reports for obtaining high-level information on the network operation at a glance. It is possible to use a report about the length of the calls to determine whether the service is becoming more or less popular among the subscribers and in what time periods of the day subscriber's use particular services the most. The call durations plotted against the day gives information on the preferences of the subscribers and helps operators to decide about marketing campaigns, special rates for busy hours or off-peak hours.

It is also possible to monitor whether a network expansion has the expected results on the quality of services by regularly checking customized reports. With regular checks, operators can avoid significant negative impact by taking corrective actions as soon as deviations are observed in the quality of the service.

When optimization of the network is needed or when expansion is considered, operators can use various PM reports to gather information on the past. Reports are used for monitoring the network performance over a certain period of time and check how the QoS and other quality objectives are met and identify possible problem areas in the network. For optimization, for example, operators might need detailed quality information on particular measurements in a particular RNC. For expansion and planning, operators might need a less detailed report but over a longer period of time to observe trends in the subscriber behaviour or network resource usage.

A set of pre-defined reports for different network technologies cover such issues as quality of service and capacity and many others. Operators can exploit the collected information from the network from the beginning. Fig 14 shows the process that is needed to obtain a report since the data collection, database management and reports specification. Data collectors collect information from the network and store it in the performance management database. Then Reporting Tools access the database to run the mechanisms necessary to acquire the values for the KPI and reports.

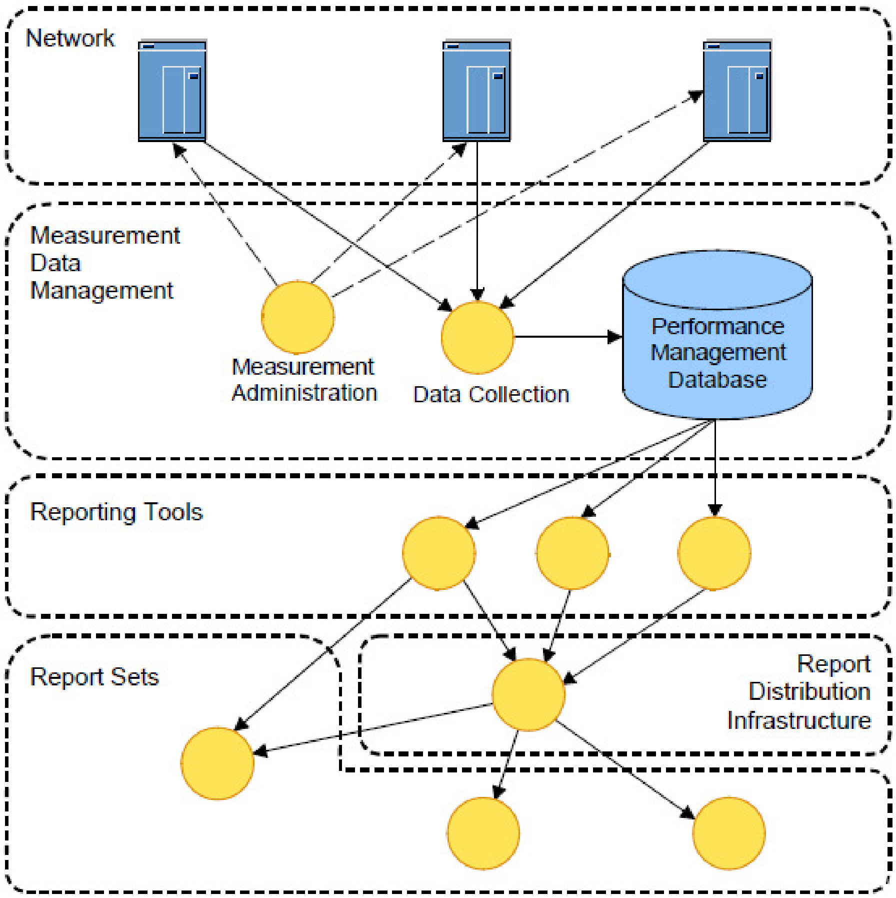


Figure 14 - Process to obtain a Report Set

Reports are specified with the purpose to give a status of specific issues in the network. In this work, some of these reports are specified and results are show based on measurements taken from a live network.

2.2.6. 3G Network Troubleshooting Use Case Example

In this subsection, a use case example for network troubleshooting is specified with the aim of showing what normally happens in a real environment for solve network problems. It covers the Call Setup Success Ratio (CSSR) because it is one of the most frequent problems in a mobile network. Fig 15 defines the phases of call setup procedure. The UE reads the System Information (available in the broadcast channel) and decides to initiate the call. First an RRC connection request message is send by UE on a CCCH (Common Control Channel) and this message contains information such us Initial UE identity and establishment cause. Then several issues are performed between NodeB and RNC (example: Downlink and Uplink synchronisation) until the reception of an RRC connection setup message. This message is sent by the RNC to UE on the CCCH with the parameters required to establish a connection on a DCH (Dedicated Channel). After the establishment of a radio link an RRC connection setup complete is sent on the DCCH

(Dedicated Control Channel). After this other parameters are negotiated until the RRC connection access is completed and then the UE haves an active RRC connection. Another phase starts with the RAB (Radio Access Bearer) establishment procedure that contains messages such as RAB assignment request, Radio bearer setup and others. After a successful RAB Setup the UE haves a Call Setup Success and is connected. If the RAB is dropped, the call is also dropped. More details of this Call setup procedure could be found in [11].

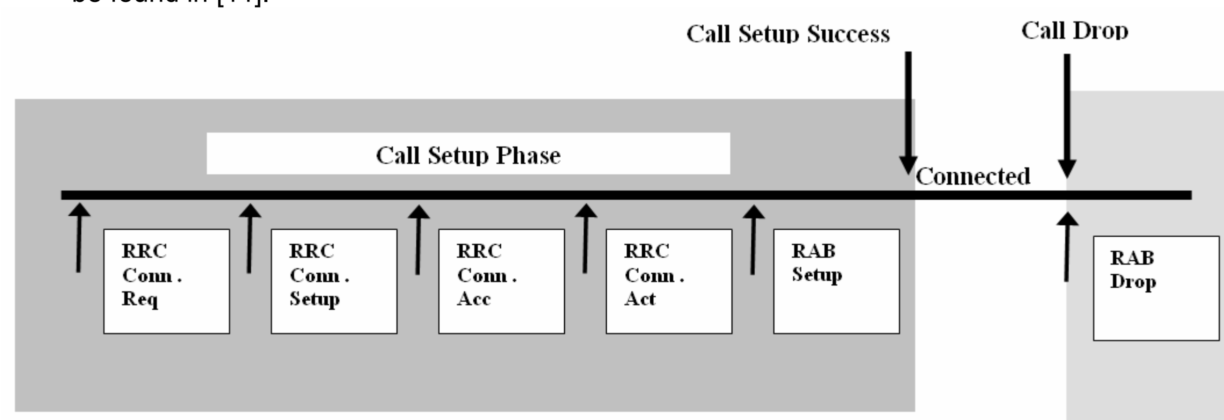
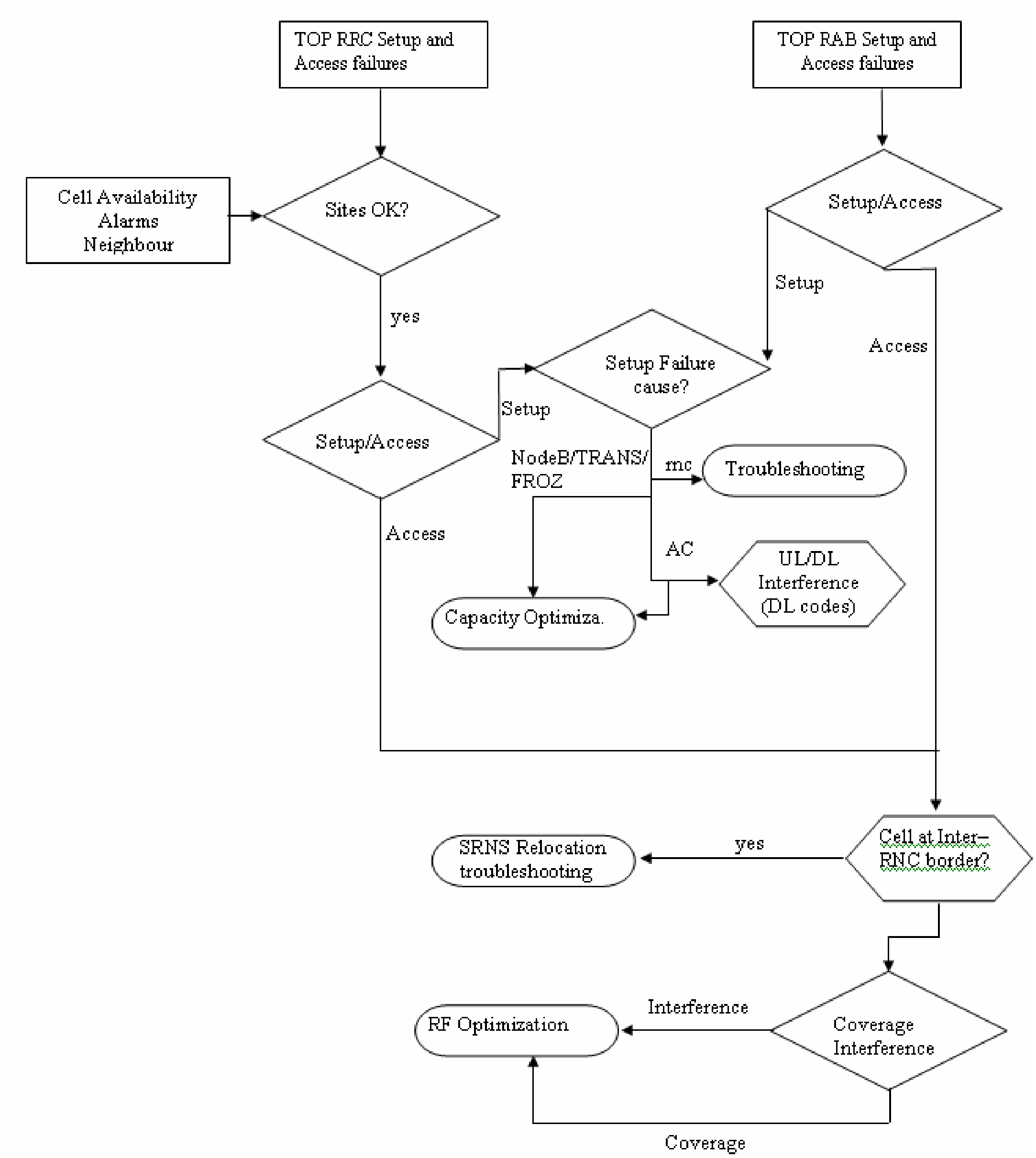


Figure 15 - Call Setup procedure phases

After analyzing the call setup phases, it is possible to understand that the CSSR will be degraded when one of the following events occurs:

- RRC (Radio Resource Control) Connection Setup Fail
- RRC Connection Access Fail
- RAB (Radio Access Bearer) Setup fail
- RAB Setup Access Fail

The next flow chart specifies the process adopted to analyze the RRC and RAB setup and access improving the CSSR. Afterwards, the adopted KPIs for this troubleshooting process are defined in more detail (chapter 5.2) and reporting values in a real network are presented (chapter 7.4).



This process (defined by the flow chart) has tree inputs: RRC Setup and access failures, RAB setup and access failures and Cell Availability Alarms Neighbour. If Alarms from cell availability or RRC failures arrive, we should first verify if Sites are correctly working because sites could be shutdown, destroyed or suffer of other problem that are not related with capacity or normal performance issues. If sites are ok, then it is necessary to verify and analyze the Setup and access problems because these points provide the connection to the network. Also, in case of RAB failures, we should analyze the Setup and Access problems for verify if problems in RAB setup are happening. In case of Setup failure, this means that the access to network has occurred, we should specify KPIs to analyze if the problem is from the Radio Network Controller, from the Node B, from Transmission, from a Frozen BTS or finally from Admission control problem because these are the possible causes for a setup problem. In case the problem is related with the RNC, we should run RNCs reports and do a deep analysis to the RNC. In case the problem is related with Admission Control (AC), we should analyze UL/DL interference, because this could happen due to allocation off codes or to another reasons related with AC mechanisms For all other cases, a capacity optimization should be performed because this is normally due to overload of users or traffic. If the problem is related with the Access phase, then it should be verified if the cell is at Inter-RNC border because one of the major problem is relate with relocation; if yes, the Serving Radio Network Subsystem (SRNS) Relocation procedure should be analyzed; if no, Coverage and Interference should be optimized in RF part because of exclusion of all the others.

3. Classifying the User Behaviour

3.1. Present and Future Mobile Services

Telecommunications are in gradual modification along the years and mobile networks have evolved from a voice-only oriented service to an enormous pool of multimedia data services. In the first generation networks, the only provided service was voice. With the second generation networks, SMS, MMS and WAP became a reality and some data transfer services were already introduced.

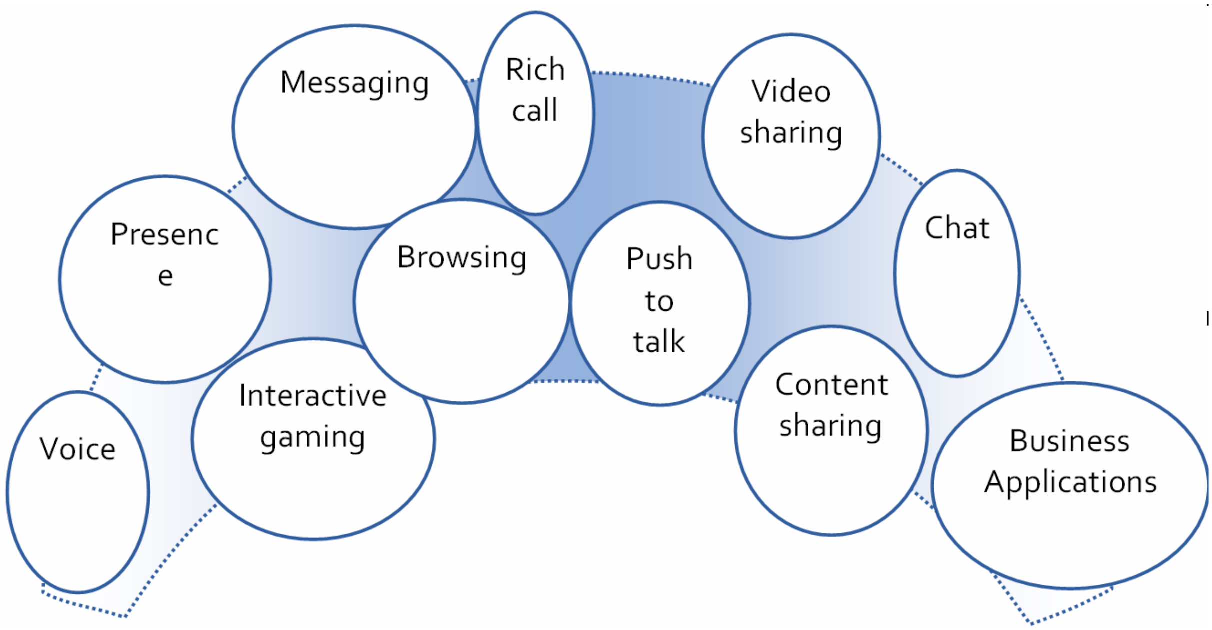


Figure 16 - Available Services in a mobile network

The increase of interest on such types of services is, thus, an important component in operators' business models [1], [2] in current days. This diversity of available services and applications with multiple requirements imposes new challenges to operators for managing the performance of networks and applications (Fig 16 highlights the services that are already available in a mobile network).

In this chapter is introduced User Behaviour analysis, Quality of Service (QoS) and Quality of Experience (QoE). This chapter shows briefly some of the available services, describes QoS mechanisms, QoS attributes and identifies some of the challenges of QoS management in a mobile network. At this chapter is also possible to perceive how to achieve QoE, understand the relationship between QoE, QoS and End user experience and see an example of user profile classification.

3.2. User Expectations and Perception of a Service

User Expectations can be understood as how several end user's think about what should be the quality of a specific service or group of services. The User Expectations are different for each user depending on several factors such as age, academic profile, demographic context, technology expertise, price of the service and others.

The perception of a service is dependent on each user sensitiveness for that service and can be defined as the opinion that the end user has about the service after the usage.

These two terms are very important because if the perception of a service matches the user expectations, then the customer is satisfied and will use that service more often. Fig 17 illustrates different examples of how different users have different expectations and different needs from the network.

Figure 17 - Different users, different expectations

3.3. QoS

This subsection intends to give an overview of Quality of Service (QoS) mechanisms in a mobile network.

3.3.1. Need of QoS

Every telecommunications operator that sells a service to a customer needs to guarantee that the service will be available with some quality. In a mobile network, it is very hard to define requirements for quality, because of hardware limitations and, since there is an enormous pool of available services, for each service the customer needs different requirements (Table 5 presents just a brief summary on the different requirements for different services).

Table 5 - Services and their requirements from the network

Services	Requirements from the network
Voice	▪ Setup Time ▪ Speech Round Trip Time ▪ Speech Quality
Video	▪ Setup time ▪ End-to-end delay ▪ Jitter ▪ Error free transmission
Browsing	▪ Setup time ▪ Page download time
Gaming	▪ Very tight delay for action games ▪ Setup time
FTP, SMS, MMS, Email	▪ Reliability ▪ Error free ▪ Delay is not very important

These quality of service (QoS) requirements for the different services can be directly mapped into network cost to the operator. For example, money is wasted if high priority and tight delay is given to advertise a user that a new email has arrived to his email server. The user will not have the perception that the advertisement was in real-time. So, differentiation of network resources per service is needed in order to ensure that resources are used in an efficient way by guarantying the correct bandwidth and other metrics for each service.

Three metrics that are always in our QoS quotidian are:

- Delay
    - Low delay is required for conversational types of services
    - Low delay variations are needed for streaming services
- Error rate

- Data services need error-free transmission (can be solved by retransmissions if delay requirements are not so tough)
- Other services may work OK with some errors – although the quality is best with few errors

- Bandwidth

- Higher bandwidth improves the response time of interactive services like web-browsing, which relies on fast file downloading

Another reason for service QoS differentiation is, for instance, the economical aspect based on how much the customer pay.

3.3.2. QoS Mechanisms in the Mobile Network

To ensure that a certain quality is given to a service, all network elements must be configured in an appropriate way to reach a common target for that service (as it is possible to see in Fig 18). If a single element fails in providing a voice call with minimum delay, the whole system will be perceived as poor in terms of delay. It is a challenge to ensure that all Network Elements, in many cases from different vendors, assure that all conditions are fulfilled.

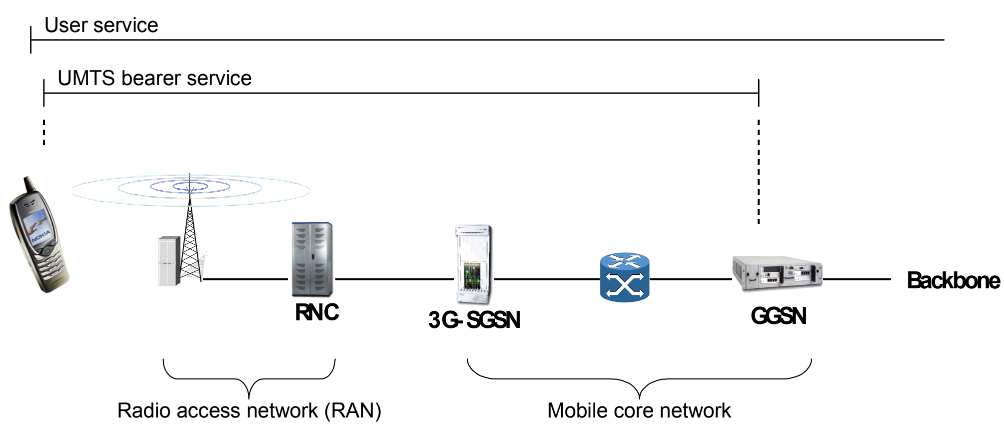


Figure 18 - Network elements involved in service provisioning should guarantee QoS target

A call is carried over the UMTS network through a UMTS bearer service. The UMTS bearer service is what a mobile terminal asks for, when it wants to transmit something over the UMTS network – a “service” offered by the UMTS network to the application layer. The UMTS bearer service is set up as a result of negotiations between the mobile terminal and the network. In the setup process, the mobile (and the network) specifies certain QoS attributes describing the call to be carried by the UMTS bearer service. These attributes can then be read by all network elements used for carrying the call (through the UMTS bearer service). The UMTS bearer service is carried by bearer services subsystem over the network as it is visible in Fig 19. However, the QoS attributes are copied or mapped from the UMTS bearer service to the subsystem bearer services, so all network elements have access to the same information.

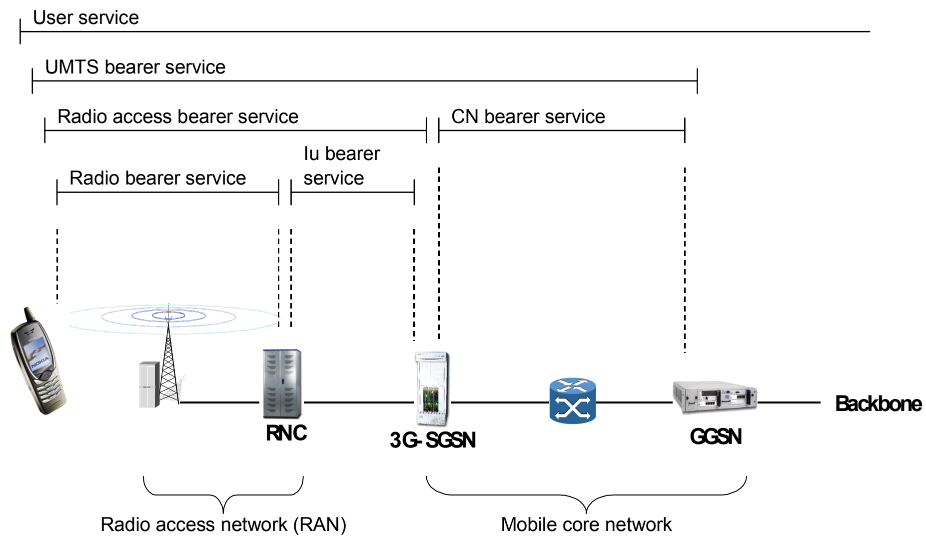


Figure 19: Architecture of the UMTS bearer service

Along the end-to-end path, the bearer services use different transport technologies such as ATM and/or IP. IP routers also support QoS differentiation, but do not support UMTS bearer service QoS attributes. So, some mapping to the IP QoS system is needed. In the next subsection, this topic will be briefly explained.

3.3.3. QOS Attributes

The main QoS attribute specified for the UMTS bearer service is Traffic Class (TC), which can have one of four values (listed in Table 6). The main factor distinguishing the traffic classes represented by the values of this attribute is the delay sensitivity. The Conversational class requires minimum delay and should be used for e.g. voice and video calls as well as for real-time games requiring the fastest possible data transport. Streaming services also require low delay but not as strict as Conversational. Interactive services do not need any guarantees on the delay although the service is typically better perceived if the delay is not too high. Calls setup with the Background value will get minimum attention when it comes to delay. These should be more or less delay-independent.

Table 6 - Values of the Traffic Class QoS attribute

Traffic Class value	Delay Sensitivity	Typical examples
Conversational	Very high	Phone calls, real-time multi-user games
Streaming	High	Video streaming
Interactive	Low	Web-browsing
Background	Very low	Emails

All network elements should operate to provide a delay dependent on the Traffic Class attribute. This means that Conversational and possibly Streaming should always be forwarded to the next network element as soon as it arrives at the element, avoiding, if possible, any time spent in queues used for scheduling. Interactive and Background data, on the other hand, should be placed in queues and wait for periods with no real-time traffic, or maybe even wait for e.g. a positive fade in its radio link before it is scheduled, so that the resources are utilized in a more resource-optimized way. At the same time, Interactive data should get priority over Background data. Because Conversational and Streaming are treated in a similar way in implementations and the same goes for Interactive and Background, it is common to classify Conversational and Streaming as "real-time" (RT) services, and Interactive and Background services as "non-real-time" (NRT) services.

In addition to the main QoS attribute, Traffic Class, several more QoS attributes should be taken into account.

Two important QoS attributes, Traffic Handling Priority (THP) and Allocation/Retention Priority (ARP), aim to further subdivide the priority of calls for network resource manager modules to do more optimal resource allocation.

The Traffic Handling Priority attribute is only valid for calls with Traffic Class=Interactive.

The reason is that Traffic Class=Interactive is expected to be used for a wide range of interactive packet services. So, the Traffic Handling Priority attribute allows these to get an additional priority to allow the system to compare them to each other. Table 7 shows the three possible values (1, 2, and 3) together with a few examples of services having those values set.

Table 7: Values of the "Traffic Handling Priority" QoS attribute

THP value	Priority	Examples
1	High	Web browsing
2	Average	Playing a game of chess over the mobile
3	Low	Email "Send/Receive Messages"

The reason behind the prioritization in the "Examples" column of the previous table is that web browsing is typically an ongoing activity where the end user clicks many links within relatively short time while searching relevant information. For such use, it would be nice to get relatively fast response from the system although no guarantees need to be given. Playing a game of chess is less demanding in terms of delay since the user can use the time for thinking about the next move. Email "Send/Receive Messages" is an action the user will conduct only once every now and then, so the response from the system needs not be so fast, and therefore it is given the lowest priority.

The Allocation/Retention Priority QoS attribute is somehow orthogonal to Traffic Class and Traffic Handling Priority. The Allocation/Retention Priority specifies how to treat the call when it requests new capacity (where Traffic Class and Traffic Handling Priority specifies how the call should be treated once it is up and running). New capacity is requested in admission control algorithms, and the system should prioritize calls with Allocation/Retention Priority=1 over calls with Allocation/Retention Priority=2 and 3. This means that if the load is "high", the system may choose to reject calls with Allocation/Retention Priority=3 to save capacity for arriving calls with higher priority. Also, in case the load grows close to outage, load control mechanisms acting to decrease the load should first deallocate calls with the highest Allocation/Retention Priority value (lowest priority). Table 8 shows the values of Allocation/Retention Priority.

Clearly, the value of the Allocation/Retention Priority attribute cannot be requested by the mobile, but has to come from the Home Location Registry (HLR) in the network, since prioritization in resource allocation is typically a function controlled by the network and not by the client application.

Table 8 - Values of the Allocation/ Retention Priority QoS attribute

ARP value	Priority
1	"Gold"
2	"Silver"
3	"Bronze"

Although theoretically all possible combinations of Traffic Class, Traffic Handling Priority, and Allocation/Retention Priority are possible, it is recommended that only the following combinations are used for Traffic Class=Interactive and Background (since this makes things more compatible with GPRS and older 3GPP releases):

Table 9 - Recommended allowed TC/THP/ARP combinations for Interactive

Traffic Class	THP	ARP
Interactive	1	1
Interactive	2	2
Interactive	3	3
Background	-	3

Table 9 shows that for Interactive, the Allocation/Retention Priority value should simply be equal to the Traffic Handling Priority value, and for Background only Allocation/Retention Priority=3 is an option. For Traffic Class=Conversational and Streaming, all values of Allocation/Retention Priority are possible.

For illustrative purposes, in Table 10 there are other QoS attributes that can be found in 3GPP specifications.

Table 10: UMTS QoS attributes

Traffic Class
Maximum Bitrate (kbps)
Delivery Order
Maximum SDU Size (octets)
SDU Format Information
Delivery of Erroneous SDUs
Residual BER
SDU Error Ratio
Transfer Delay (ms)

Guaranteed Bitrate (kbps)
Traffic Handling Priority
Allocation/Retention Priority
Source Statistic Descriptor

3.3.4. QOS Differentiation in IP Transport

In many cases of real UMTS network, the IP (Internet Protocol) is used for carrying the UMTS bearer service from one network element to another. In particular, this is common in the core network part, but also the connection between SGSN and RNC or RNC and Node-B may be carried also over IP. In the IP world, there is a QoS mechanism somewhat similar to the UMTS QoS framework for discriminating data packets based on what is carried. This mechanism is called DiffServ (Differentiation Services) and is standardized in IETF (RFC 2474).

Instead of a number of QoS attributes, each IP data packet has a DSCP (DiffServ Code Point) value set in the IP header. The DSCP field in the IP header is only six bits long (compared to UMTS bearer services, which uses several attributes). The DSCP value is read by the network node (typically a router) and used to determine which treatment to apply for the packet. The different treatments supported by a router are called per-hop behaviours (PHBs). A per-hop behaviour is one forwarding behaviour the network node applies to the served packets. Typical behaviours are Expedite Forwarding (EF), Assured Forwarding (AF), and Best Effort (BE). As these names indicate, they can be compared to UMTS traffic classes. EF has a resemblance with Conversational and Streaming, AF is similar to Interactive, and BE is more or less the same as Background. A range of DSCP values are standardized to refer to specific per-hop behaviours (like those mentioned here), whereas others can be defined to have a local meaning in an IP sub-network.

Fig 20 shows how the CN bearer service could be transported over an IP connection between the GGSN and the SGSN. The transmitted IP packets should have DSCP marked through a mapping from the QoS attributes of the UMTS bearer service. In this way, the IP routers will be able to treat the packets according to the carried user service (which is hopefully properly described by the QoS attributes of the UMTS bearer service).

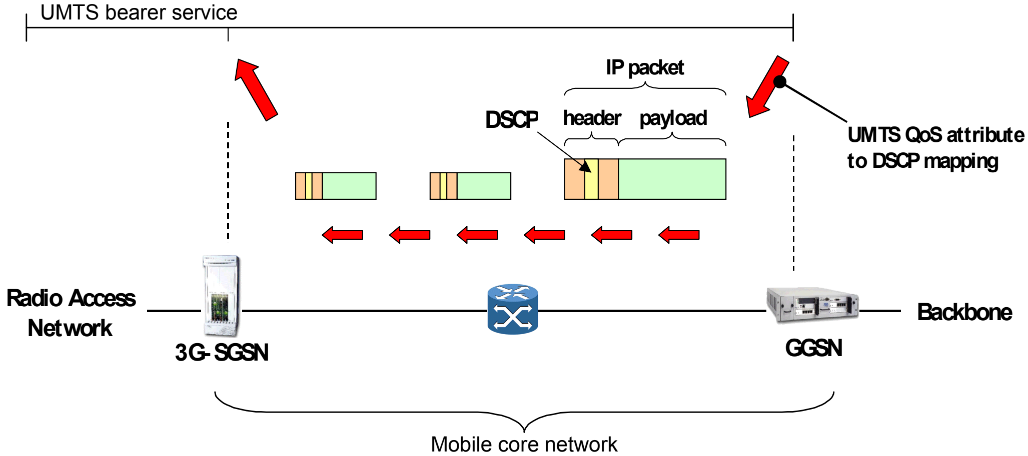


Figure 20 - UMTS bearer service over IP

Fig 20 shows a CN bearer service implemented with an IP connection from GGSN to SGSN. Fig 20 shows data sent only in one direction, but, the opposite direction is similar with the mapping function located in the SGSN.

3.3.5. QOS Management

The management of a QoS system is another challenge since although the high-level QoS requirements have been harmonized in standardization, the different network elements still have vendor-specific implementations, which will be configured through vendor-specific sets of QoS-enabled parameters. What is here called QoS-enabled parameters is a subset of all the available configuration (planning) parameters used to steer the behaviour of network elements. QoS-enabled parameters are exactly those that allow the operator to configure the differentiated treatment of calls – parameters with a different setting for different Traffic Classes, Traffic Handling Priority values, and/or Allocation/Retention Priority values. An example is a downlink power threshold, ptxThreshold, used in call admission control in the RNC; if this threshold has a separate value for each different Allocation/Retention Priority value, then it is QoS-enabled.

So, a QoS management system is an operator user interface enabling customization of the QoS mechanisms. This interface receives inputs from all network elements, as it is illustrated in figure 21.

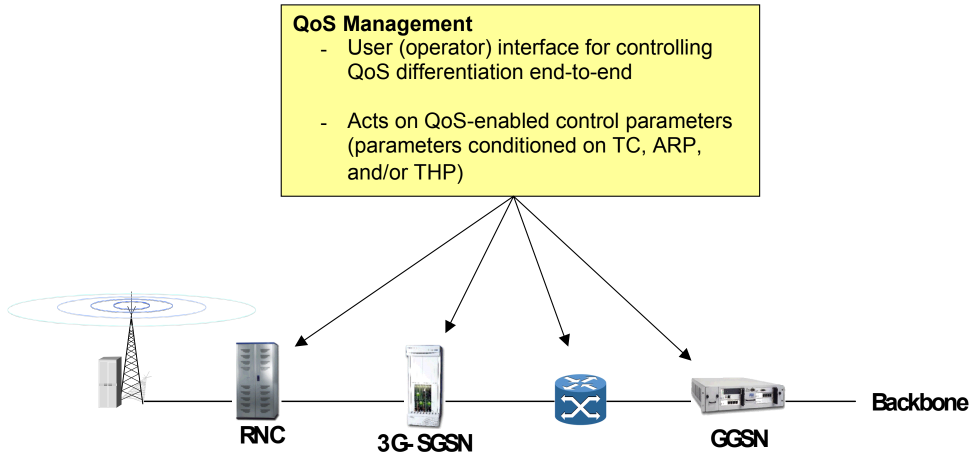


Figure 21 - User interface for controlling QoS differentiation end-to-end

One special case of a QoS management system is a policy management system. Such a system tries to abstract over the vendor-specific and network element specific QoS-enabled parameters and provide a more high-level policy to the whole system (or parts of the system) using general terms to give higher priority to one service (e.g. streaming) over other services (e.g. speech). If implemented in the right way, such a system can be very powerful but we will not go deeper into policy management since it was not a goal for this work.

3.4. Quality of Experience (QoE)

Quality of experience can be defined as the overall performance of a network from end user point of view. It is a subjective measure of end to end service performance from the user perspective and it is an indicator of how well the network accomplishes the user's needs. Quality of experience is different for different users under the same conditions because everything is dependent of the expectations of a certain person for a certain service and how this person reacts to service performance. This "human factor" is very hard to measure and is related with concepts such as social class, geographical localization, cultural aspects, how deep is the technology know how of the user, how much the user is able to pay for the service and many others.

3.4.1. End User Centric Approach to Analyze, Improve and Assure QoE

Since QoE has a subjective nature, we need to understand which factors are tangible and which are intangible. Some tangible factors are:

- Wide portfolio of reliable services
- E2E network and service quality
- Network and service accessibility
- Device usability and functionality
- Easy service activation

And some intangible factors are:

- Relevance to needs
- Emotional engagement
- Service discovery & ease of use
- Price vs. value
- Ability to share experience/content
- Quality of customer care and support

QoE is affected by many issues over the network as it is possible to perceive from Fig 22. In the figure is possible to observe some concerns for a good QoE related with service layer such as provisioning, aplication protocol, security and content packaging. But all these concerns are also dependent on the physical layer and all the problems related with netwrok elements illustrated in the botom of the figure.

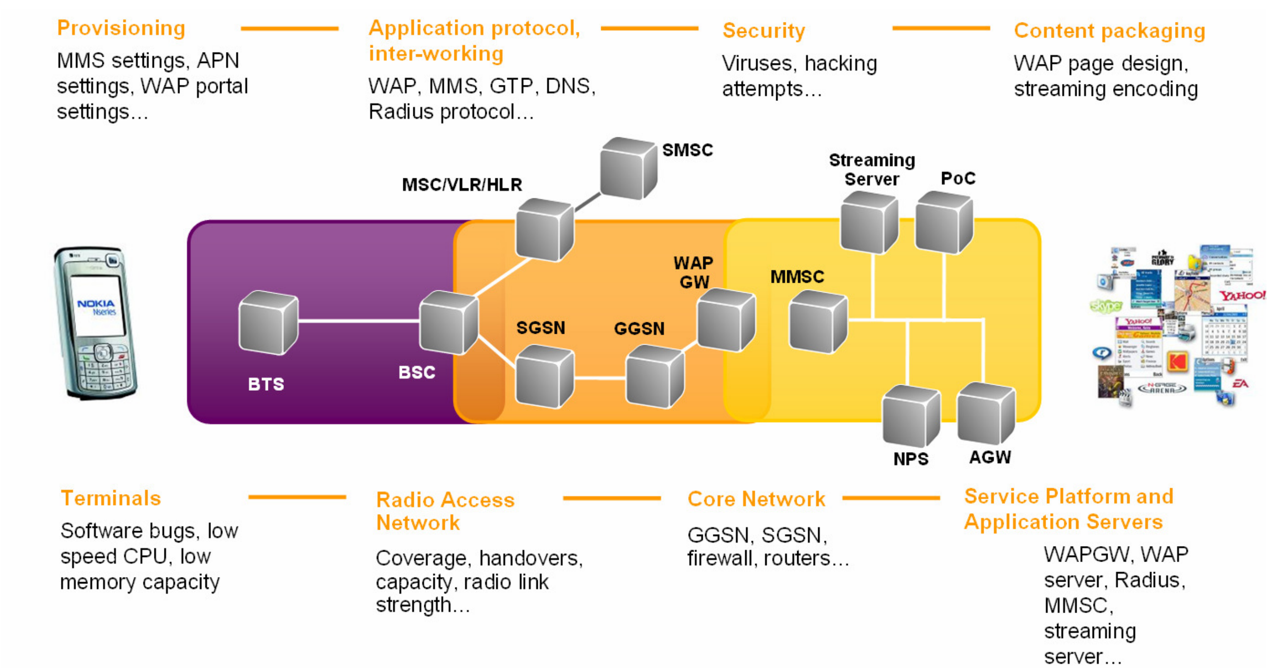


Figure 22 - Several issues impact QoE over the network

The user centric approach can be defined as a Top-Down method to relate network KPIs (describing QoS parameters) and the tangible and intangible factors of QoE.

Fig 23 illustrates the Top-Down method that is characterized by start to take into account more in-tangible QoE metrics, for instance service integrity, and go deeper and deeper into more tangible factors to perceive the root cause for a certain problem.

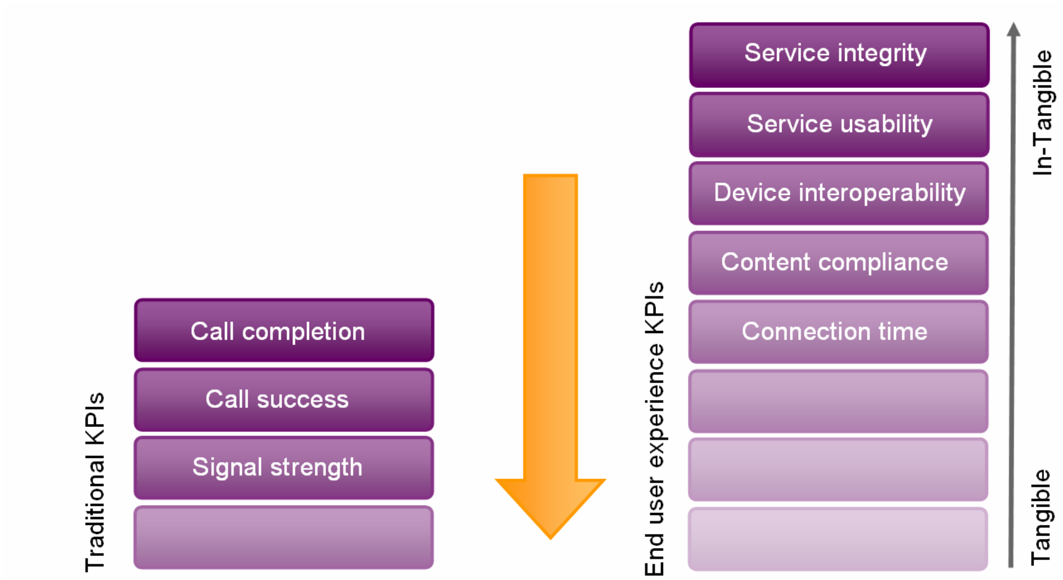


Figure 23 - Top Down method to relate network KPIs to QoE

3.4.2. QoE Achievement

There are several methods to measure QoE for the various mobile services. Some are based in people's opinion such as MOS (Mean Opinion Score) [6]. MOS is an evaluation method where a questionnaire is done to the end user after the usage of a service and the user classifies the quality of the service in a scale from 1 (very bad quality) to 5 (very good quality). Other methods are based in some algorithms such as PESQ (Perceptual Evaluation of Speech Quality) [7] that do exactly the same as described to MOS, but in this case it is only used for voice classification based on an automatic algorithm (no annoyance for the user). There are still methods based in quality degradation models such as E-model [8] and some are based in a neural approach of the network such as Rubino's PSQA (Pseudo Subjective Quality Assessment) [9].

Typically operators use two approaches:

- Statistical Samples: In this case, the available services in the network are benchmarked via questionnaires by trying to apply MOS and some questions to different types of persons in different locations.
- Optimize QoS to assure a better QoE: With this method, operators improve QoS performance metrics to ensure a good quality of the network in general and expect to affect QoE in a positive way. Some correlation models have been proposed for this case [5].

The first approach evaluates statistical samples with surveys and gives the real end user perception for a certain service or group of services. However, this method has some disadvantages such as time and money consuming and an additional degree of annoyance towards the users who need to take a survey. The second approach relies on a QoE and QoS parameters mapping. This is not very accurate for QoE analysis because it is possible to make a correlation model between QoS and QoE parameters but it is not fully true that by optimizing QoS parameters and network performance metrics the QoE will have great improvements.

3.4.3. Relationship between QoE, QoS and User behaviour

QoE is dependent of the QoS according to the network conditions which depend on the user behaviour. If we consider that customers who are more satisfied with the quality of a service will potentially make more requests and interact more often with the network, we establish a closed loop (presented in Fig 24) with an extremely wide range of variables, including not only technical KPIs but social indicators as well.

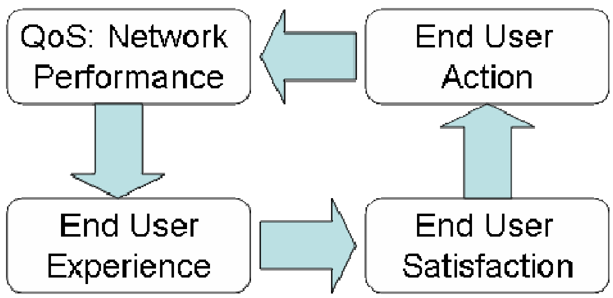


Figure 24 – QoS, QoE, End user Action and End User Satisfaction relationships

Next, a mapping is defined showing the dependency of QoE metrics from network performance metrics (QoS KPI's). To identify these metrics, I assume that the two most important principles to ensure a good end user experience are:

- Availability and accessibility of a service
- Quality and integrity of a service when reaches the end user

In Table 11, a mapping for correlating the most important QoE metrics with network performance QoS metrics is proposed (this mapping was done as part of this work). For the Accessibility QoE goal KPIs related with access phase like RRC access success, RAB setup and access, PDP context failure ratio and service access times were selected. Service access times do not exclude the user from accessing the network, but certainly degrade the quality of experience when accessing a certain service. For availability goal KPIs related with the number of channels available and interfaces availability were chosen because these are the most common limitations for this goal. For Quality and Integrity, KPIs related with packet loss and errors are specified. To map bit rate goal into QoS, the selected KPIs are channels capacity and throughput.

Table 11- Mapping between QoE and QoS metrics

QoE goals	QoS KPI's (measured in the network)
Accessibility	○ RRC Access Success ratio

QoE goals	QoS KPI's (measured in the network)
	o RAB Setup and Access Success Ratio per class service (Background, Conversational, Streaming and Interactive) o PDP context Failure ratio o Service access Times
Availability	o % of channels available for different services (AMR allocations, Background, Conversational, Streaming and Interactive) o Interfaces availability (ATM, SDH, PDH)
Quality and Integrity	o Packet loss ratio for different services o RTP layer loss ratio ingress/egress o Traffic loss at AAL2 layer o HSDPA/HSUPA congestion ratio o BLER for MAC PDUs o Frame Protocol error ratio
Bit Rate	o Bit rate o Allocated channels capacity o Throughput
End-to-end delay	o Average end-to-end delay o Jitter

All these QoS KPI's can be found in Reporting Tools used by the operators to accomplish a status of their network conditions. Although network performance in general as an "entire system" or a "global service" can be acceptable, the QoE, being more related with the experience that the end user perceive from a specific service, may be undergoing performance issues.

3.4.4. End User Behaviour Impacts the Network and the Network Impacts End User Behaviour

At this point, I want to show the close relation between end user behaviour and the network performance and vice versa. Several parameters are configured by the telecommunications' operators in the network infrastructure and some of them are directly associated to mobile oriented services.

Typically the services' configurations are done concerning the strategy of the operator in terms of network cost, new possible features and market directions. Usually, the performance of the mobile services is limited, since it is not possible to have lots of sites (BTS, NodeB, etc) everywhere due to the cost of network elements. Optimization engineers realize the hard job of optimizing a network with limited resources in order to provide the "best in class" services performance to end user. In a complex system such as an advanced mobile network, there are always some factors that can not be controlled by operators. One good example of a very important external factor that cannot be controlled or totally predicted is the end user behaviour. Next, two examples are presented that can show the influence of end user in the network performance:

- People decide to go to a festival in a wide open zone without good network conditions or a car accident causing a very large amount of people in a small area trying to communicate;
- Large group of people connected to the same radio element that decide to do browsing of different contents (with different performance requirements) or video calls/streaming at the same time.

These two examples will certainly cause a degradation of the network performance causing a negative reaction in the end user behaviour. These are typical examples of a close loop between end user behaviour, network performance and services optimization as illustrated in Fig 25.

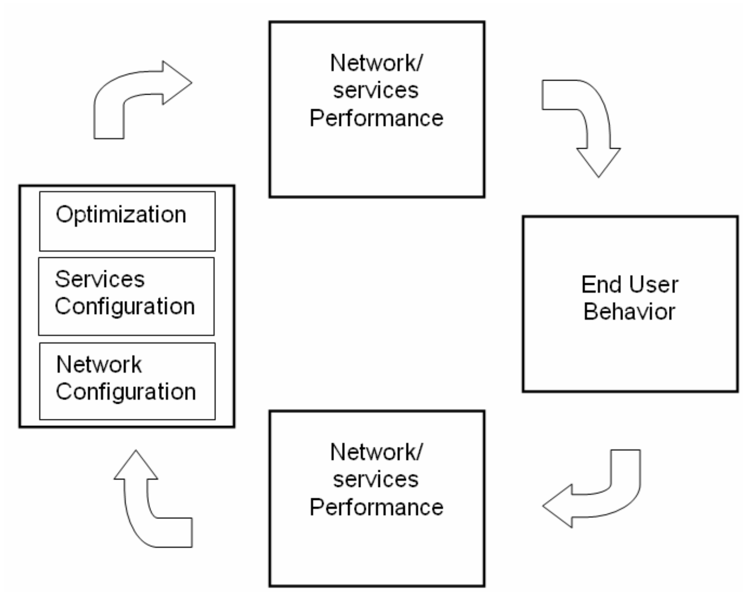


Figure 25 - Loop between user behaviour, network performance and services optimization

Figure 24 intends to illustrate the relation between user satisfaction, QoS and QoE. The main goal of figure 25 is to show the dependency of service configuration and optimization in the end user behaviour.

3.4.5. User Profile Classification

Usually, users want the lowest price for mobile services but accept to pay more money for specific services that are more important to them. So, operators, need to segment the market in order achieve all customers in an accurate way and services should be designed with price differentiation.

Free trials are a method to obtain customers for a service. Services are discovered by word of mouth, user recommendations and buddy invitations. Once again, to realize advertisement correctly for each user or group of users, operators need to know the user profile.

Understanding more about customers and their behaviour means that services can be package to meet exactly with customer lifestyle needs. When this happens, customers see the service provider as more valuable and in turn will be more loyal. Content packages must appear as if were designed for each user individually, as if designed by the customers themselves.

In resume, operators need to segment users by key customer groups (based on research about each user preferences) and after this, the users should be targeted with advertisement specific for them based on time, services in use and value.

Examples of profiles that could be created in an operator database to group users are as follows:

- 'Teenager' - Deeply into social networking and all new ways to interact with peers. Video is central to how they communicate;
- 'Top executive' - Senior business role, it's vital they stay in regular contact with colleagues, clients and suppliers while travelling and in the field;
- 'Always ON' - Content from browsing, TV and music are crucial to their total mobile experience. They need to feel 'connected' all of the time.

Fig 26 illustrates how marketing advertisement could be done by the operators for each group of profiles, defined before. In the figure is possible to perceive that Mobile TV could be advertised to a mix group of people with profile A (video calling/sharing) and people with profile C (browsing). Voice SMS service is a good service to advertise people from profile A (video calling/sharing) and profile B (Push E-mail or Push to talk over cellular). Location _Based Services or Video Blog are good services to advertise to people from profile B and C. This is just an example diagram of what operators could do to map all profiles established.

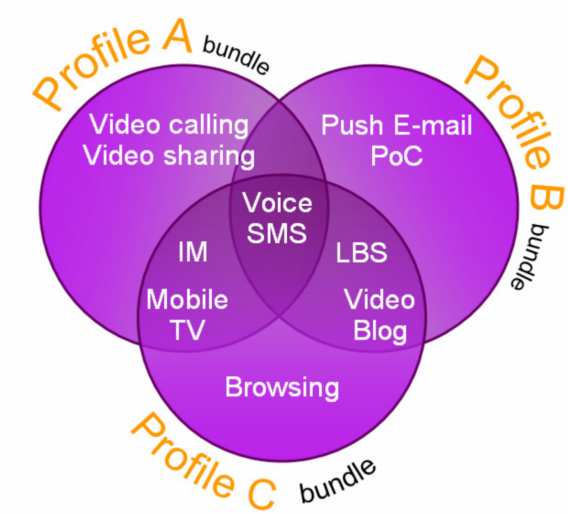


Figure 26 - User Profiles matching

4. User Behaviour Monitoring

This chapter describes the User Behaviour monitoring tool, its user profiles and metrics, whose definition was one of the aims of the present work.

4.1. Description of Monitoring Tool Architecture

Main purpose of the monitoring tool terminal software application is to collect information from mobile devices. It outputs the collected pieces of information in the log file and it sends this file to the database through the wireless network for accurate analysis.

The monitoring tool consists of the following modules:

- Client (in the mobile phone)
- SMS Gateway
- Log Reader
- Recruitment panel
- Panel Manager
- Database

The data flow between all these modules is described in Fig 27.

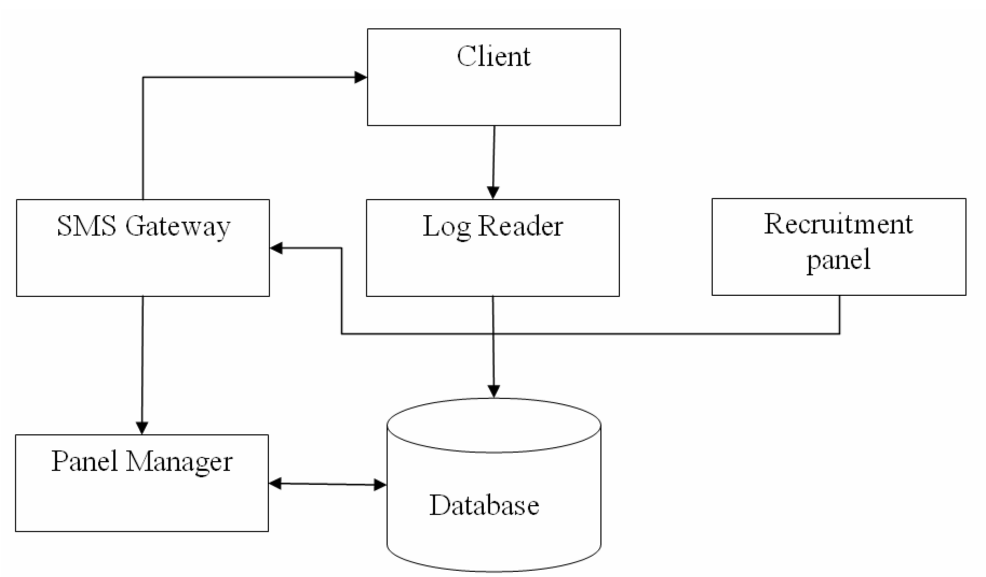


Figure 27 - Monitoring tool architecture

The process starts with the recruitment panel that sends sms with information for software download link. Then the user installs the client software in the mobile and activates it with the help of the panel manager. The Log Reader stores in a log file the metrics collected in the client and upload it over the wireless network to the database.
After the data is available in the database, it is processed and presented through a Graphical User Interface (GUI).

4.2. User Behaviour Models

User Behaviour models intend to give information of services usage and how the end user experienced that usage. The final objective is to acquire the information on the user (or group of users) most used services and on how these users experienced the services.
The user behaviour models that are collected by the tool are:

- Personal Network
    - The Personal Network report shows the distribution of voice calls, SMS and MMS among the top 5 destination users compared to all the destinations.

- Mobile Questionnaire
    - The Mobile Questionnaire is a specific Observer which asks Questions about the device user.

- CS Call 24h profile
    - The CS Call 24 Profile report shows the average number of calls as well as the call duration for each hour of the day and each day of the week. The hourly average is first taken per terminal over all days in the defined time frame that the terminal was active and then over all filtered terminals.

- Message 24h Profile
    - The Message 24h Profile report shows the average number of sent and received messages for each day of the week and per hour of the day. The hourly average is taken first per terminal over all days in the defined time frame that the terminal was active and then over all filtered terminals. Also the text length for sent and received messages can be displayed and which message types for display can be filtered: SMS, MMS, Bluetooth, infrared, E-mail.

- Application 24h Profile
    - The Application 24h Profile report shows the average number of application usage and usage time of different application categories for each hour of the day per each day of the week. The hourly average is taken first per terminal over all days in the defined time frame that the terminal was active and then over all filtered terminals.

- Camera vs Sent Attachment
    - The Camera vs. Sent Attachment report consists of several reports about how many pictures and videos have been taken and how many of them have been sent via MMS or E-mail.

- Web Browsing Usage
    - The Web Browsing report consists of the most visited URL's and how long they where being visited.

- Message and Call Stimulus
    - The Message and Call Stimulus report shows the number of different types of reactions by the terminal user to a specific action such as:
        - Received and read SMS
        - Received and read MMS
        - Answered calls
        - Missed calls

- Top Application Usage and User Activity
    - The Top Application Usage and User Activity report shows the absolute or relative application usage in a chart.

4.3. Service Quality Reports

Service Quality reports give an overview of the status of each service in the network. QoE metrics such as Integrity and Accessibility are very important in an end user perspective.

The following reports can be computed with the tool:

- Service Dashboard
    - The Service Dashboard provides a comprehensive overview about the individual services available in the mobile terminal;
- CS Call Service
    - For the CS Call service is monitored the accessibility, integrity and retainability. PIs such as call duration, voice quality and RX quality (signal quality in the receiver) are taken into account.
- PS Call Service
    - The PS Call quality report shows the accessibility of the service (Setup failure ratio, Attach failure ratio and context activation failure ratio) and the integrity of the service (PS call setup duration, PS call data traffic, PS call Throughput and PS call max Throughput);
- Browsing Service
    - The Browsing service analysis report provides an overview of the Browsing service structured in accessibility (HTTP Service Accessibility Ratio and HTTP IP-Service Non-Accessibility Ratio), integrity (Webpage download duration, traffic and throughput) and retainability (Webpage success ratio, fastest site download and slowest site download);

- Messaging Service
    - The Message service provides an accessibility report (SMS accessibility ratio and MMS accessibility ratio) and an integrity report (SMS/MMS send tries, duration and size).

4.3.1. Service Agent based KPI's

To monitor how the services are being delivery by the network, the following KPI's could be used for each service measured in nearly real time:

- Voice Telephony
    - Setup Time Telephony
    - Service Accessibility Telephony
    - Composite Call Success rate
    - Call Completion Rate Circuit Switched Telephony
    - Dropped call rate
    - Speech Quality on Call Basis PESQ
    - Speech Quality on Call Basis MOS

- SMS Transfer
    - Service Accessibility SMS MO
    - Access Delay SMS MO
    - SMS sending time
    - End-to-end Delivery Time SMS
    - Completion Rate SMS Circuit Switched

- MMS Transfer
    - MMS send failure ratio
    - MMS retrieval failure ratio
    - MMS send time
    - MMS retrieval time
    - MMS notification failure ratio
    - MMS end-to-end failure ratio
    - MMS success ratio
    - MMS end-to-end Delivery time
    - MMS size check success rate

- WAP Transfer
    - WAP Browsing Success ratio incl PDP
    - WAP Browsing Delivery time incl PDP
    - WAP Browsing Success ratio
    - WAP Browsing Delivery time

- Video Stream
    - Streaming Non-Accessibility
    - Streaming service Access time

Results to some of these KPIs are presented in chapter 7.3.3.

5. Network Reporting Tool

This chapter starts by describing the functions of the reporting tool that was used in this work. Then, it describes the work conducted in the specification of KPIs providing the appropriate status of the network, specification of the KPIs for the analysis of the troubleshooting example described in section 2.2.6 and, finally, specification of the reports, based on a scorecard, that can give the status of a service.

5.1. Description of Network Reporting Performance Tool

With a reporting tool, the operator can view and analyze performance, fault and configuration data from multiple sources across the whole network. Raw data is translated into meaningful information that is visualized in graphical KPIs.
A reporting tool is designed to help operators to analyze and optimize the entire managed network, in the presence of the latest technologies. Reporting tool provides operators with:

- A constant source of information in the network status;
- Reports in all network levels;
- Long-term data statistics;
- Multi-vendor support;
- Interfaces for measurement data export.

The network statistics and measurements are collected and stored temporarily in the network elements, or element management systems. Specified raw counters are processed into Performance Indicators (PI) and Key Performance Indicators (KPI), and transferred to the PM Database. Different performance measurements are combined into data modules. Typically, the measurements from one network technology are bundled in a data module.
Three tasks have to be executed to enable PM database to receive measurement data of a network element:

- Install the adaptation that supports the network element type in PM database.
- The physical network element has been integrated into PM database, so that performance and alarm data can be exchanged.
- The network element has been configured to send measurement data, either manually, or by using Administration of Measurements.

These preparation tasks enable a continuous PM data stream from the network element to the receiving database system. The received data is transformed in the native PM database format by using mediation components. Required conversion configurations or scripts are part of the respective adaptation. After that, the data is inserted into the database. After the insertion, the data is available for further usage, e.g. network reporting, troubleshooting or optimization.
Fig 28 presents the typical architecture of a Reporting tool. Network elements send data to PM platform with the help of mediations. After this, the adaptation for the network elements should be available in the platform and raw data is aggregated. The aggregation of data could be done in “time domain” (combine, for instance, data that arrives every 15

min into day, week or month) or in “Object level domain” (combine, for example, data from all the NodeBs at RNC level).

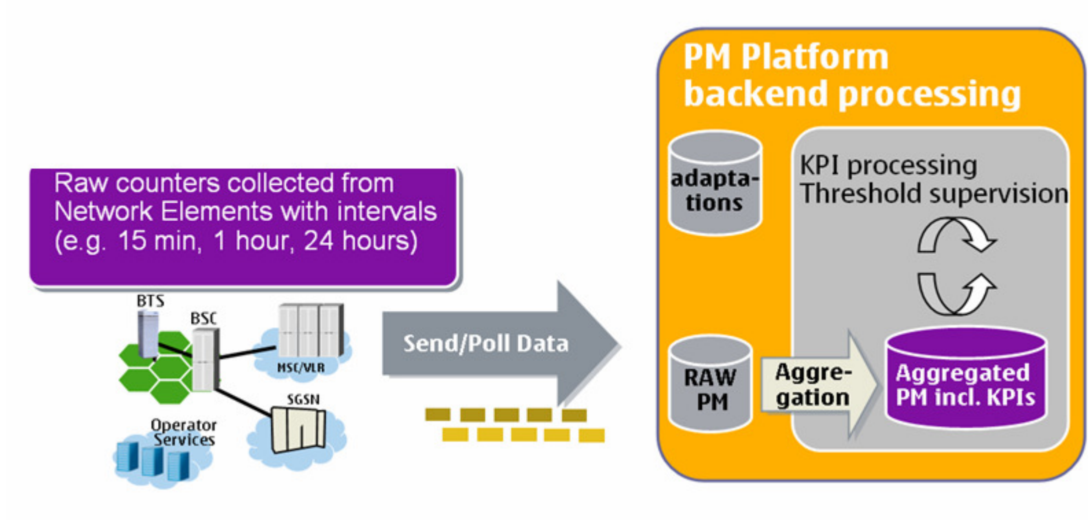


Figure 28 - Reporting tool architecture

5.2. Network KPI’s Specification

This section describes the work done in the specification of the network KPIs that provide a detailed overview of the network status. To do this specification, I had to do a detailed study of 3G technology and a verification on which counters or performance metrics were available in the network. Then, the specification of the appropriate KPIs was done for the several network elements, network interfaces and protocols where problems could occur. Results to these KPIs are presented in section 7.4.

5.2.1. WCell Quality indicator

WCell (this term is used to refer to a WCDMA 3G cell) Quality indicator KPI can be used to relate network perceived quality for a specific cell with the one reported by the end user in MOS for that cell region. The KPI formula is:

$$CQI = \frac{\sum (CQI\_1 * 1 + \mathrm{CQI2} * 2 + \ldots + \mathrm{CQI\_n} * \mathrm{n})}{\sum \mathrm{CQI\_1} + \mathrm{CQI\_2} + \ldots + \mathrm{CQI\_n}}$$

CQI PI shows the number of reported CQI values matching to a certain class and this PI is updated by the NodeB. The numerator multiplies the number of samples of a certain class by the quality of the class to make all of them equivalent in terms of samples and the de denominator is the sum of all samples reported. The result is the overall quality of a wcell during a certain period.

5.2.2. WCell Active Throughput Downlink

WCell Active Throughput shows the active throughput (Mbps), in a network perspective, to the end user. The active time refers to the scheduled TTI (Transmission Time Interval). The KPI formula is:

$$CellActiveThp = \frac{\sum\left[((\mathrm{Re}ceived\_MACd\_PDUs)-(Discarded\_MACd\_PDUs))\right]}{Schedule\_TTI\_time}$$

The numerator is subtraction of discarded packets from all packets because this is the throughput to end user. Some packets are discarded and never reach the end user. The denominator is the sample of time for this throughput measurement.

5.2.3. WCell Active Throughput Uplink

This KPI shows the total number of transmitted data (Mbps) to the end user, divided by the total amount of active simultaneous HSUPA user's allocation time.
Formula is similar from previous point (5.2.2) but for uplink direction.

5.2.4. WCell Availability

This KPI shows the total time that a cell is available.

$$Cell\_Availability = 100 * \frac{\sum Time\_Wcell\_in\_working\_state}{\sum Time\_Wcell\_exists\_in\_RNW\_Database}$$

The numerator is the total amount of time that a cell is in working state (this is the time that effectively the cell is working) divided by the total amount of time that a cell is active in RNW (Radio Network) Database. Note that a cell could be active in the RNW database but not working due to some problem.

5.2.5. Data Downlink Service Accessibility

This KPI provides the network perspective of 3G downlink data service accessibility. This factor is calculated by the division of downlink channel allocations (number of times a channel is in use) with the downlink channel selections (number of time that a channel was selected to transmit). Note that a channel can be selected but not allocated, this means that a channel is selected to transmit but the user will not transmit in that channel due to some problems.

$$Data\_Downlink\_accessibility = 100 * \frac{\sum Downlink\_channel\_allocations}{\sum Downlink\_channel\_selections}$$

5.2.6. Data Uplink Service Accessibility

This KPI provides the network perspective of 3G uplink data service accessibility. This factor is calculated by the division of uplink channel allocations with the uplink channel selections.

$$Data\_Uplink\_accessibility = 100 * \frac{\sum Uplink\_channel\_allocations}{\sum Uplink\_channel\_selections}$$

5.2.7. Call Setup Success ratio

This KPI combines RRC setups, Signalling Radio Bearers (SRB) success and RAB success ratio.

$$Call\_Setup\_SR = 100 * [Call\_RRC\_SR * RAB\_SR * Standalone\_SRB\_SR]$$

This formula is the combination of all processes that are involved in a setup phase of a call. First is necessary to establish a RRC connection and after this is needed a RAB establishment (this process is detailed explained in section 2.2.6). SRB is a radio bearer established in the DCCH for transport signalling data. SRB is used during call establishment to provide the RAB and to deliver signalling during the connection.

5.2.8. Channel Elements Utilization Ratio

This KPI describe how Channel Elements (CE) are being used in the network. It is an important KPI to HW resources analysis.
The formula selects the maximum number of CE that were used ate the same time either for downlink or for uplink and divides it for the total of CEs available.

$$CE\_utilization\_ratio = \frac{MAX(\max\_used\_CE\_DL, \max\_used\_CE\_UL)}{\sum \max\_available\_channel\_elements}$$

5.2.9. HSPA Users (Wcell level)

This KPI provides the average number of users for HSPA service. This KPI, HSPA users, is divided into two KPIs: HSDPA users and HSUPA users.

$$HSDPA = \frac{\sum sampled\_values\_for\_measuring\_simultaneous\_HSDPA\_users\_in\_the\_cell}{Number\_of\_samples}$$

$$HSUPA = \frac{\sum sampled\_values\_for\_measuring\_simultaneous\_HSUPA\_users\_in\_the\_cell}{Number\_of\_samples}$$

Formula divides the total number of sampled values for all the users in the cell (for uplink and downlink direction in different KPIs) by the total number of samples. This gives the possibility to know how many users are connected in a cell for HSDPA or HSUPA.

5.2.10. Max HSPA Users (RNC level)

This KPI calculates the maximum usage of the RNC when service type is HSDPA or HSUPA. This KPI uses the Digital Signal Processor (DSP), which is responsible for processing all data in RNC, with a filter for the specific service to calculate the usage. There is a separate KPI for Downlink and Uplink. The formulas are bellow:

$$MAX\_HSDPA\_USERS\_RNC = MAX(DSP\_allocated\_resources\_HSDPA\_services)$$

$$MAX\_HSUPA\_USERS\_RNC = MAX(DSP\_allocated\_resources\_HSUPA\_services)$$

This formula calculates the maximum value of a PI that counts the allocated resources (these are actually two different PIs, one for HSDPA and other for HSUPA, which origins two different KPIs).

5.2.11. HSPA Throughput (RNC level)

This KPI calculates the throughput at RNC level. The KPIs below, for HSDPA and for HSUPA, are similar to the ones present in sections 5.2.2 and 5.2.3 (Wcell level) but now is for RNC level. Two KPIs are specified, one for HSDPA and other for HSUPA service. The formulas are bellow:

$$HSDPA = \sum \frac{Received\_MACd\_PDUs\_for\_HSDPA\_service}{Allocated\_time}$$

$$HSUPA = \sum \frac{Received\_MACe\_PDUs\_for\_HSUPA\_service}{Allocated\_time}$$

These KPIs are the result of the total packets sent and received, for HSDPA and HSUPA, by the RNC divided by the total allocation time. Note that know the discarded packets are not subtracted (has it happens in section 5.2.2 and 5.2.3) because these KPIs show the throughput in a RNC perspective and not in an end user perspective.

5.2.12. RRC Setup and Access Call Ratio

RRC Setup and Access Complete Ratio over the reporting period show RRC Setup and Access phases from RRC Connection Request to RRC Connection Setup Complete.

$$RRC\_stp\_acc\_call = 100 * \frac{\sum \begin{bmatrix} (RRC\_ACC\_COMP) \\ +(RRC\_CONN\_STP\_COMP\_AND\_DIRECTED) \end{bmatrix}}{\sum \begin{bmatrix} (RRC\_STP\_COMP) \\ +(RRC\_CONN\_STP\_COMP\_AFTER\_DIRECTED) \\ -(RRC\_STP\_Reject\_Due\_To\_Emergency\_Call\_Redirect) \end{bmatrix}}$$

The numerator of this KPI is the sum of all RRC accesses completed (total accesses that were performed for a RRC connection) and the RRC connection setups that were directed. The denominator is the sum of all succeed RRC accesses (counted in the RRC_STP_COMP PI), plus the succeed directed RRC accesses and the subtraction of RRC setup rejected due to emergency call, because this is not a network error, so it should not count as a failure.

5.2.13. RRC Fails per Cause

This KPI calculates the RRC connection active failures related with the radio part of the network. Contributions for these failures are Radio, NodeB, Iur interface, CPICH channel (related with the common pilot from NodeB), RNC and User Equipment. Formula is:

$$RRC\_Fails\_per\_cause = \sum \begin{bmatrix} RRC\_Fails\_due\_to\_Radio \\ +RRC\_Fails\_due\_to\_NodeB \\ +RRC\_Fails\_due\_to\_Iur \\ +RRC\_Fails\_due\_to\_CPICH \\ +RRC\_Fails\_due\_to\_RNC \\ +RRC\_Fails\_due\_to\_UE \end{bmatrix}$$

5.2.14. RAB Setup and Access Call Ratio

RAB Setup and access Call ratio is a very important KPI to monitor how a user can access to the several services available in the network.

$$RAB\_STP\_ACC\_COMP\_ratio = \frac{RAB\_ACC\_COMP\_for\_Voice}{RAB\_STP\_ATT\_for\_Voice}$$

This KPI is the division of the total RAB accesses completed with success by the total attempts for a RAB establishment.

5.2.15. Packet Service RAB Drops per Cause

RAB drops per cause in radio part are calculated by the sum of all RAB drops contributions of radio network elements and interfaces. Formula is:

$$RAB\_Fails\_per\_cause = \sum \begin{bmatrix} RAB\_Fails\_due\_to\_Radio \\ + RAB\_Fails\_due\_to\_NodeB \\ + RAB\_Fails\_due\_to\_Iur \\ + RAB\_Fails\_due\_to\_Iu \\ + RAB\_Fails\_due\_to\_RNC \\ + RAB\_Fails\_due\_to\_UE \end{bmatrix}$$

5.2.16. PRACH Propagation Delay

PRACH (Physical Random Access Channel) Propagation Delay gives an estimate of the distance between the UE and the base station. This is important for network optimization in order to know if NodeBs geographical location is the most correct. When the UE send a RRC Connection Request or Cell Update, one of PRACH Propagation Delay distribution counters is updated by value 1. For PRACH propagation delay, it is not needed a KPI. It is just needed to make visible all PRACH counters (from class 0 to 20).

5.2.17. UE Power Headroom

UE power headroom (UPH) is the division of the UE maximum transmission power and the corresponding physical channel code power. UPH is updated when NodeB receives Scheduling Information PDU in serving radio link set or in a non serving radio link set. It is important to understand power distribution in the network.
This parameter does not need to be implemented in a KPI since it is possible to observe the values directly from the network counters.
UPH value is divided in classes and the range is from 0 to 31.

5.2.18. Average Traffic per Call

Here, two KPIs are specified, one that measures the average CS voice traffic per call and other one for conversational (already explained in section 3.3.3) normally associated with phone calls, real-time multi-user games. The KPI is the result of the average CS voice call duration in minutes divided by the number of active CS voice calls per minute. The counters are updated when a RAB is released.

$$Average\_Traffic\_per\_call = \frac{\text{Average CS Voice Call duration in minutes}}{\text{Number of active CS voice calls * hour duration in minutes * 1000}}$$

$$Average\_Traffic\_per\_call\_conv = \frac{\text{Average\_CS\_Voice\_Call\_duration\_in\_minutes\_for\_conv}}{\text{Number\_of\_active\_CS\_voice\_calls\_for\_conv * hour\_duration\_in\_minutes * 1000}}$$

5.2.19. Iub FP Layer Data Throughput

This KPI provides information about the Iub interface throughput in BTS Frame Protocol level. The BTS measures the FP level throughput from the payload of FP data frames (i.e. excluding the FP headers). This KPI is divided in two parts, DL and UL.

$$Iub\_FP\_Thp\_DL = \begin{bmatrix} FP\_payload\_for\_CCH \\ + FP\_payload\_for\_DCH \\ + MACd\_data\_for\_HS - DSCH \end{bmatrix}$$

$$Iub\_FP\_Thp\_UL = \begin{bmatrix} FP\_payload\_for\_CCH \\ + FP\_payload\_for\_DCH \\ + MACe\_data\_for\_E - DCH \end{bmatrix}$$

This KPI is the sum of payload in CCH (Common Channel), DCH (Dedicated Channel) and HS-DSCH (High Speed Downlink packet access Channel) for Downlink or E-DCH (Enhanced Dedicated Channel) for uplink.

5.2.20. Iub FP Layer Data Volume

This KPI reports the Iub FP Layer data volume and is divided in UL and DL direction:

- DL

The Downlink Traffic includes all Dedicated Channel (DCH) user plane, plus control plane FP payload (excluding FP header data) traffic and HSDPA shared channel (HS-DSCH) FP payload (excluding FP header data) for a LCG (Local cell group) traffic between RNC and BTS.

$$Iub\_FP\_data\_Vol = \sum \begin{array}{l}(Payload\_data\_of\_FP\_common \\ + Dedicated\_traffic \\ + HS-DSCH\_traffic\_received\_for\_Iub\_in\_the\_LCG)\end{array}$$

- UL

The Uplink Traffic includes all user plane and control plane FP payload (excluding FP header data) traffic between RNC and BTS.

$$Iub\_FP\_data\_Vol = \sum \begin{array}{l}(Payload\_data\_of\_FP\_common \\ + Dedicated\_traffic \\ + E-DCH\_traffic\_received\_for\_Iub\_in\_the\_LCG)\end{array}$$

5.2.21. Iub Congestion Rate

This KPI shows the Iub congestion rate in the Downlink and in the Uplink. For the Downlink direction, provides the ratio between the congestion events to the number of received HS-DSCH frames:

$$Iub\_congestion\_rate\_DL = \frac{\sum number\_of\_congestion\_indications\_sent}{All\_HS-DSCH\_related\_frames}$$

For Uplink direction, the formula is similar, but E-DCH frames are monitored.

$$Iub\_congestion\_rate\_UL = \frac{\sum number\_of\_congestion\_indications\_sent}{Missed\_EDCH\_frames}$$

$$\sum + Delayed\_EDCH\_frames$$

$$SUCC\_REC\_EDCH\_frames$$

This KPI is important because enables the HSPA congestion control feature to limit the radio layer throughput when Iub performance decreases.

6. Architecture Proposal

This chapter describes the architecture of the developed system, the reasons behind the adopted architecture and the required extensions.

6.1. Architecture

Usually, operators work with reporting performance tools to gather information of network status and QoS and to do troubleshooting and optimization of the network. The problem of this solution is that operators have full control over the network and know with which QoS a service is being delivered by the network but don't know what are the end users expectations and how the services are perceived by them (QoE).

In this architecture, it is proposed a simultaneous use of a monitoring and a performance tool. With such solution, we can know how a specific user or a specific group of users are using the services available in the network (thus creating a profile), evaluate how end user reacts to service performance provided by the network (for instance, a user video call, can keep the call if the performance is good, or i can shift for a voice call and send MMS if quality is bad) and visualize the network performance. The great advantage of this approach is to know how end users react to the available service performance, further know his opinion and collect detailed network metrics at the same time to quantify the performance.

Figure 29 presents the architecture capable of acquiring all the metrics stated before.

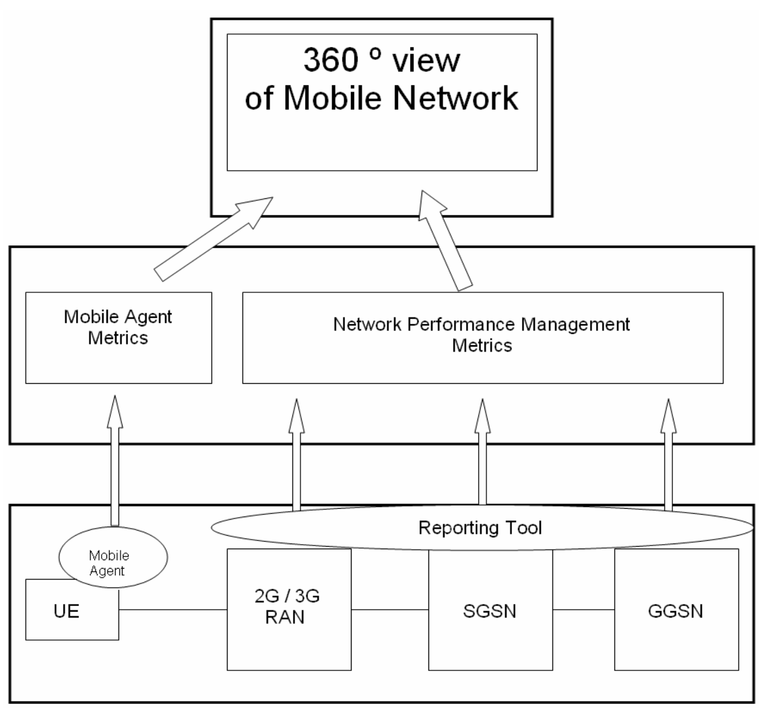


Figure 29 - Architecture of 360 º view of the Network

The architecture includes a Mobile Agent (monitoring tool) and a Reporting Tool. With the Mobile Agent, it is possible to collect QoE metrics and real opinion from end user (in this case by surveys realized with the agent and applying the MOS method). The agent collects QoE metrics on how end user perceives the quality of a service.

The agent also gives us the opportunity to collect some counters with relevant information for 2G and 3G technologies (Rx_Level, Rx_Quality, Ec/No, RSCP, SIR Target ...) that can give an idea of network conditions in an end user perspective. But, since the collection module is installed in the mobile terminal, some problems might occur such as mobile phone resources consumption or missing of detailed network measurements. So, to complete the vision of detailed network metrics, the powerful reporting tool is also included in the architecture. With the reporting tool, it is possible to collect detailed measurements from network perspective for all available mobile technologies in the market (GSM, UMTS, HSPA, etc.). With such a combination, operators have a "360º view" of the network, or in other words, have the end user perspective and the network perspective. One of the advantages of this architecture, by acquiring end user satisfaction and relating it with the network metrics, is to provide the operators with the opportunity to have OPEX costs lowered because the end users only get what they really need or expect. In Figure 30, it is shown an example, acquired from a recent study [10], which shows that it is not necessary to provide more than -81dBm of radio signal from NodeB to end user terminal because the end user feedback information already reports that the call speech quality is excellent.

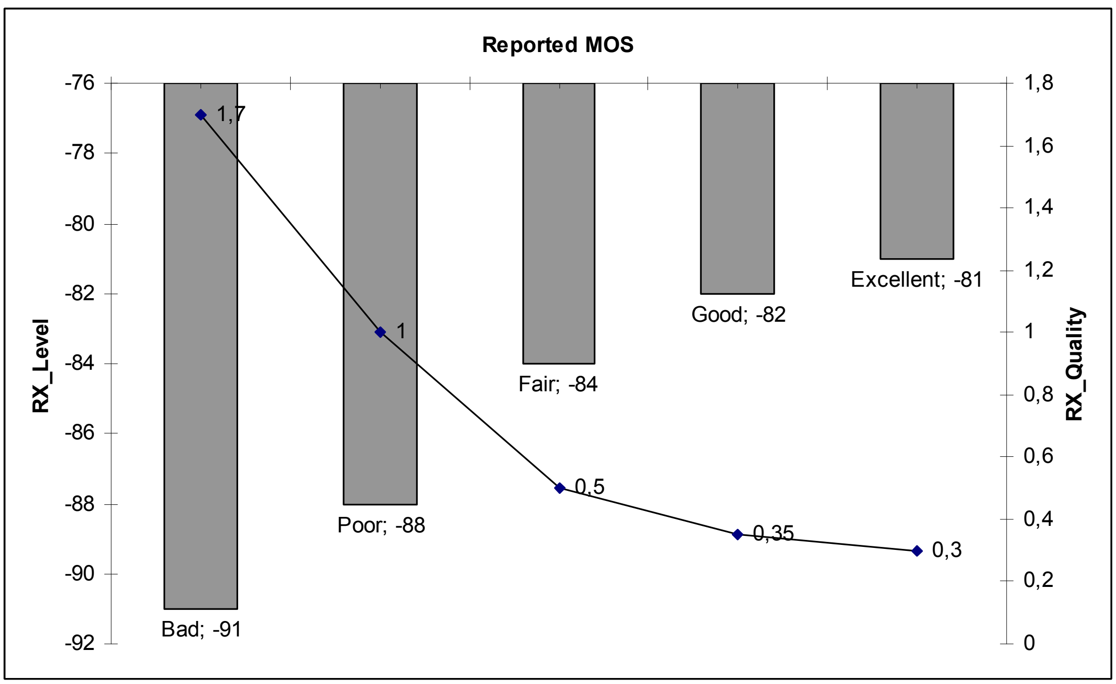


Figure 30 - Relation reported MOS, RX_Quality and RX_Level

Acquiring this information about the end-user is very important because of the "human factor" and all other conditions (demographics, price of service, etc...) stated before that change the end user expectations. By classifying a specific region in terms of end user expectations, the operators can predict how their reaction/acceptance is going to be. It is also possible to create a database with the profiling (which services a specific user uses more, how long, how is the satisfaction reported, etc.). With this profile information in a database, operators can optimize more one service instead of another, according to end users preferences in a certain region. Another possibility is to do marketing oriented approach to a specific user because operators already have a detailed knowledge of user's profile.

6.2. Required Extensions

To reach the previous described architecture with an easy to use or user friendly powerful tool, some extensions should be implemented. Two possible alternatives are: one that complete integrates both tools into one tool and another scenario where only the database is common.

The architecture proposed in the beginning of this chapter is a merge of a reporting tool and an agent based monitoring tool. If operators want to have data available from both tools, they have to buy both, install them in separate machines and the most difficult task is to relate data from each other synchronously.

So, the easiest scenario to software vendors that can simplify operator's task is to develop adaptations and mediations for both tools and to integrate data into one common database. Since data is under the common database, the network topology is exactly the same and data is completely synchronous. With this option, operators have two separate GUIs, one for the monitoring tool and another one for the reporting tool, but the hard task of relate data from two sources is solved. In Fig 31, the architecture extension proposal with a common database is illustrated.

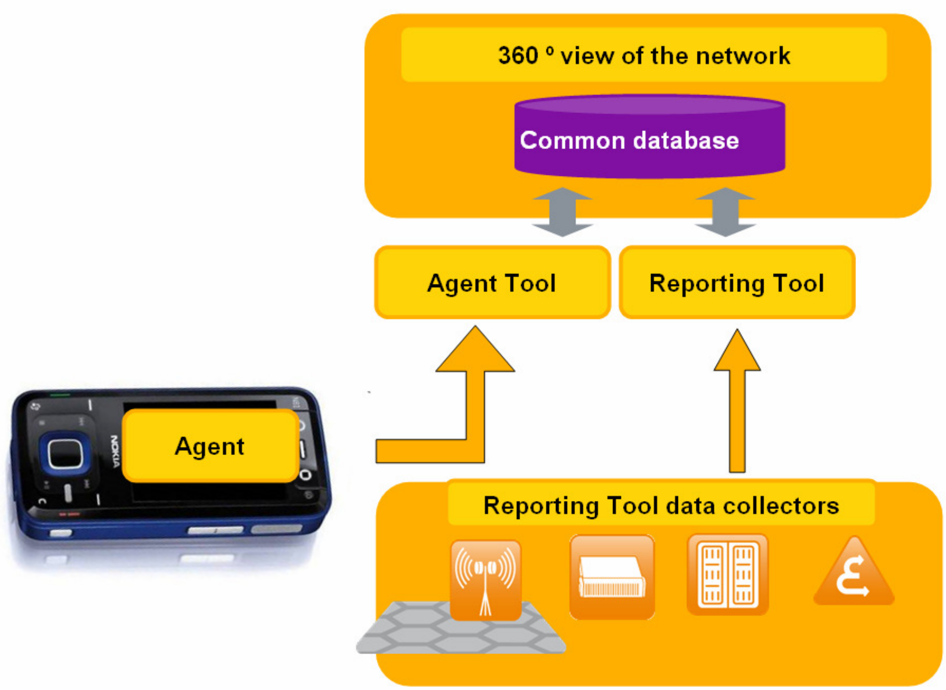


Figure 31 - Architecture extension with a common database

With this alternative, a problem still persists which is that operators need two different installations and effort to maintain both machines working correctly. They also need training in two different tools and specific know how from each one. So, the ideal scenario to operators is to have both tools integrated into one. This scenario is the most complex to software vendors but it is the one that really accomplishes operator's needs. Figure 32 illustrates the alternative of a complete merge between both tools.

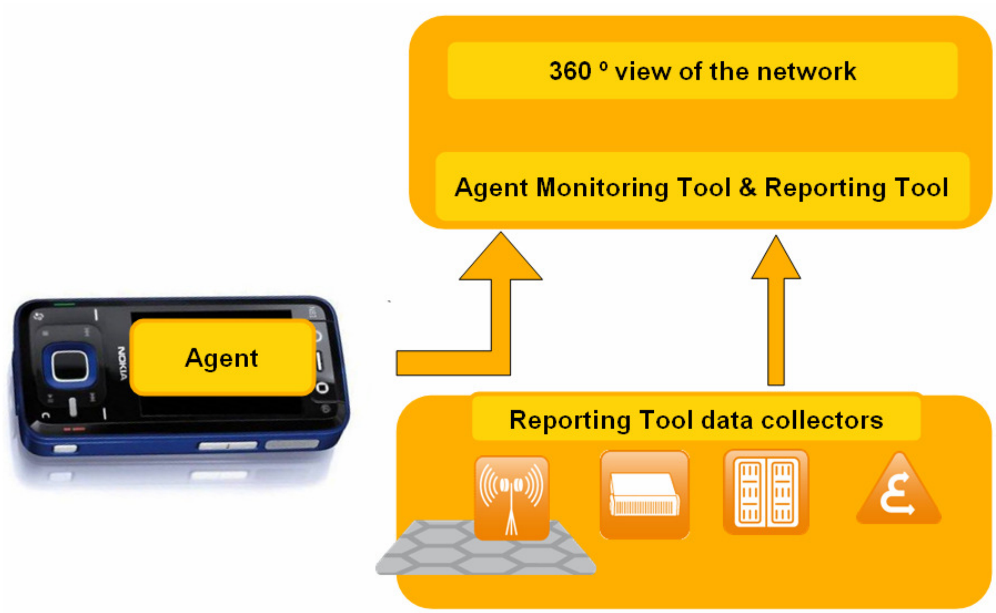


Figure 32 - Architecture extension with both tools merged

7. Implementation and Results

This chapter describes the implementation of the proposed monitoring and reporting system and the results obtained by it on real networks.

7.1. Implementation Issues

All data that is shown in this chapter is from real networks. User profiles and behaviours are from users around the world from 23 different operators. Network performance metrics were collected from a real network of a mature operator (with a stable and optimized network). Although possible, it does not make very sense to collect metrics from different operators around the world with the developed agent because in this way we are covering a huge geographical area, and very different behaviours and expectations from the users. The reason because this combination was done is to not happen the chance of data collected identify an operator (confidential issues). Data is from real operators and is confidential, so during results presentation, all details that can identify a specific operator are hidden. The idea of this work is to prove that the architecture works and operators can use it to achieve detailed network status and end users preferences. After the realization of a case study like this one in a single operator network, the network will be classified by regions according with user's needs and expectations. After this, the operator should analyze what the network is able to offer and should change QoS priorities mechanisms, giving priority to services that are more used and that are more profitable. In Fig. 33 is present a typical QoS queue in an operator network.

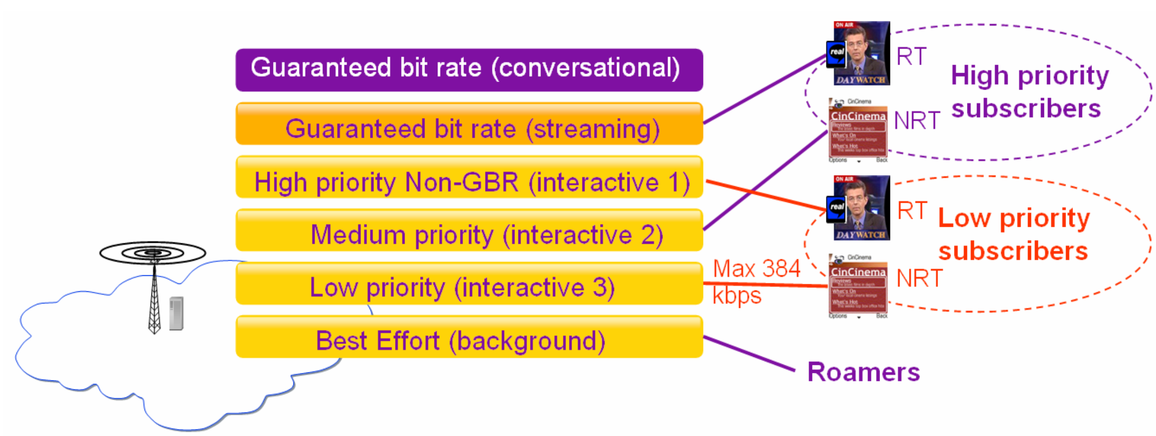


Figure 33 - Typical QoS queue in an operator

Normally, operators give the highest priority to interactive traffic instead of background. But, for instance, in a very specific area, after analysing user behaviour and needs, the operator concludes that, in that area, email service is very important and interactive services have a good classification in terms of user experience. In this case, it will be a good approach if the operator gives more priority to background traffic instead of interactive traffic.

7.2. Implementation Description

My implementation consists of several agents (in this specific case: 423) installed in users from different operators around the world from 72656 3G network cells. Data was collected for a week long (from 29 of October to 5 of November). This data was stored in the monitoring tool database and processed. After this, in order to acquire user behaviour, the previous defined user behaviour models (section 4.2) are applied.

The other part of the implementation consists of network data collectors implemented in a single operator. After collection, data is stored in the reporting database and aggregated by the aggregation engine. After this, is reported the network status graphical information based on the KPIs previously specified (section 5.2). Figure 34 shows the overall solution described before in this subsection.

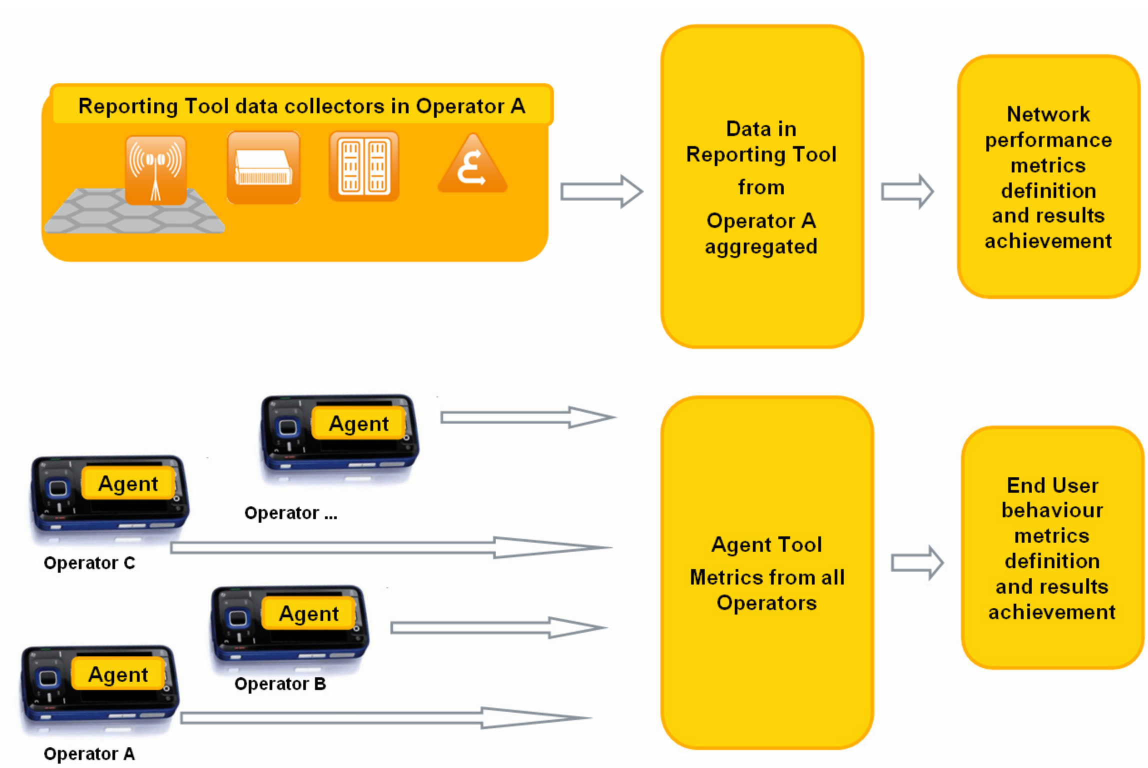


Figure 34 - Implementation description

My tasks regarding implementation issues were the following:

- Monitoring tool:
  - Tool configuration by select and activate NodeB's within a geographical area to collect measurements

    - Select terminals with agent deployed in those areas
    - Filter the amount of days to run the report
    - Dashboard configuration
- Reporting tool:
    - Guarantee that measurements needed are active
    - Deployment of adaptation for the data collected in the database
    - Export data for further processing
    - Specify network KPIs and create graphics for network status overview

7.3. User Behaviour and Service Status Metrics Collected by the Agent

n this section I present user behaviour metrics previously described in section 4.2. To show the results I adopt the same structure of chapter 4.

7.3.1. User Behaviour Metrics

- **Personal Network**

Top SMS

The Top SMS report shows that the user with more usage of SMS service sent 715 msgs in one week, followed by users with 293, 277, 232 and 194 msgs.

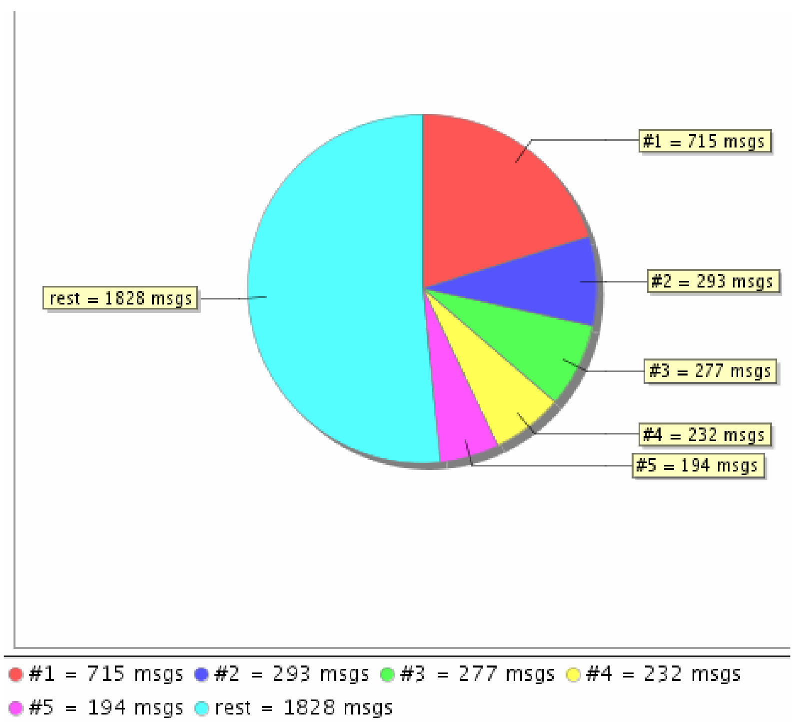


Figure 35 - Top SMS report

Top MMS

The Top MMS service shows that the top user sent only 3 mms messages. The mms service is the less used in these networks for this period.

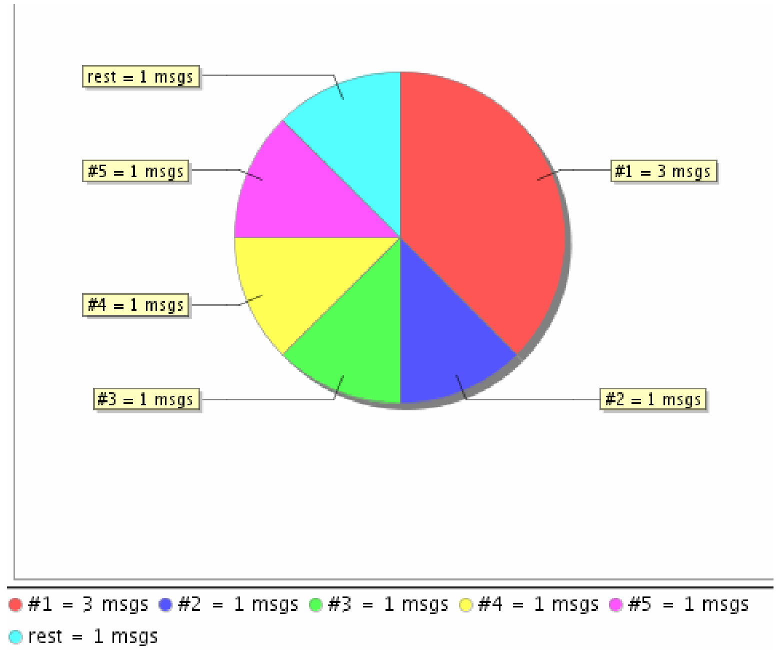


Figure 36 - Top MMS report

Top voice

The top voice service shows a top usage of 795 calls per user, followed by 330, 323, 312 and 175 calls.

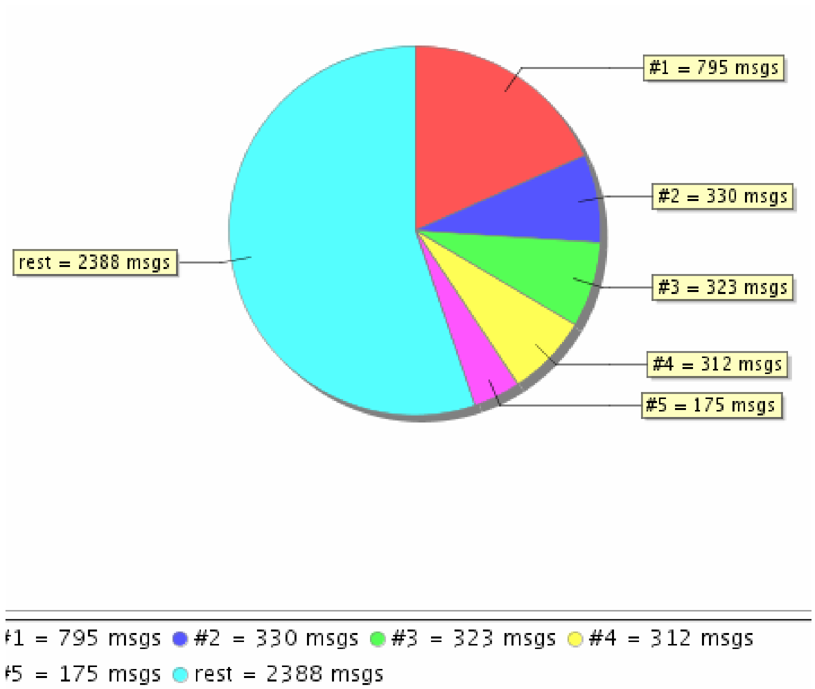


Figure 37 - Top voice

- **CS call 24 hours profile**

In CS call profile, it is possible to understand that the majority of the calls were outgoing calls with an average duration of 2 min.

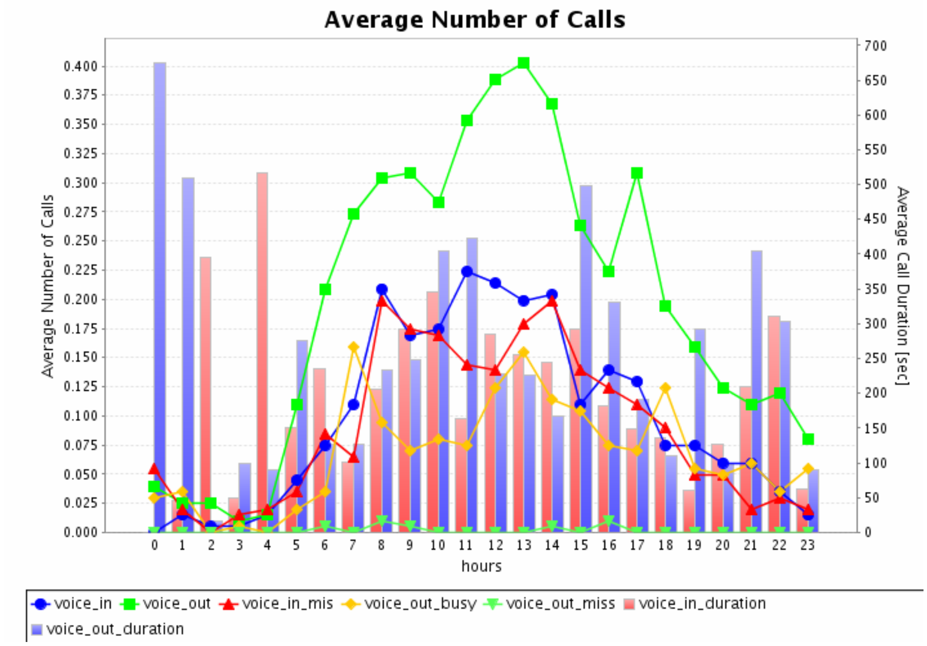


Figure 38 - CS call 24 hours profile

- **Messaging 24 h sms**

For the SMS service, we can see there are more received SMS messages than transmitted ones and the average length is 67 characters.

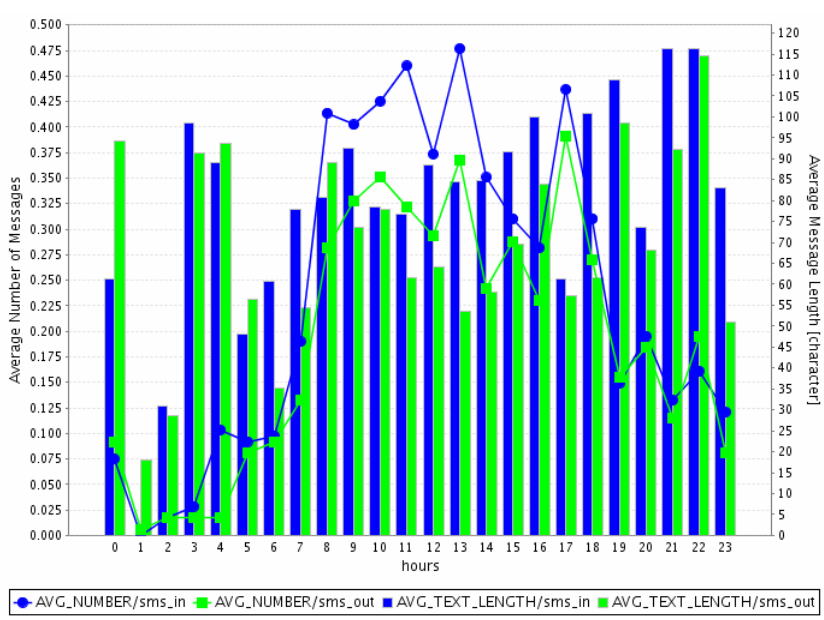


Figure 39 - Messaging 24 h sms

- **Messaging 24 h email**

For email, the graphic shows that almost all emails are received ones (this is due to the fact that the agent is installed in mobile phones, and people hardly use the mobile terminal to write emails).

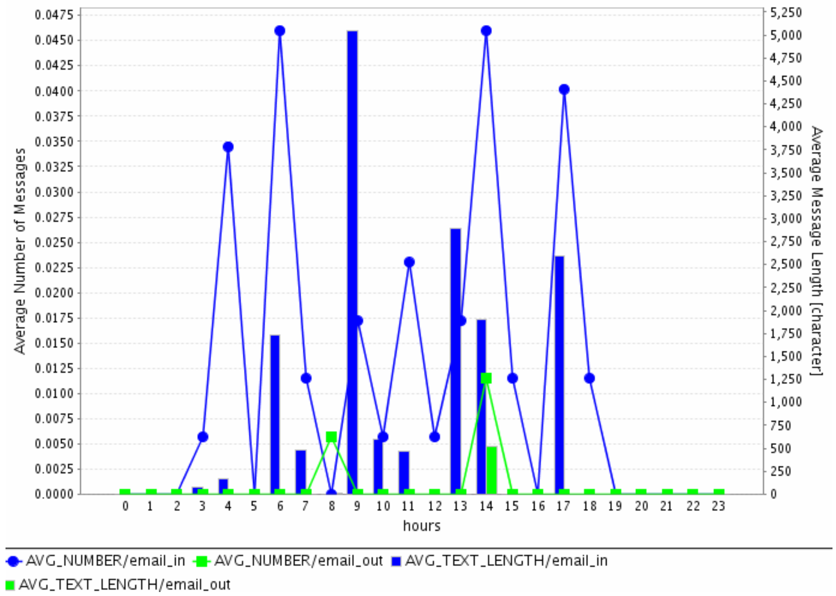


Figure 40 - Messaging 24 h email

- **Application – Browsing**

The most relevant information that can be taken from this graphic is that browsing sessions have an average duration of 5 min.

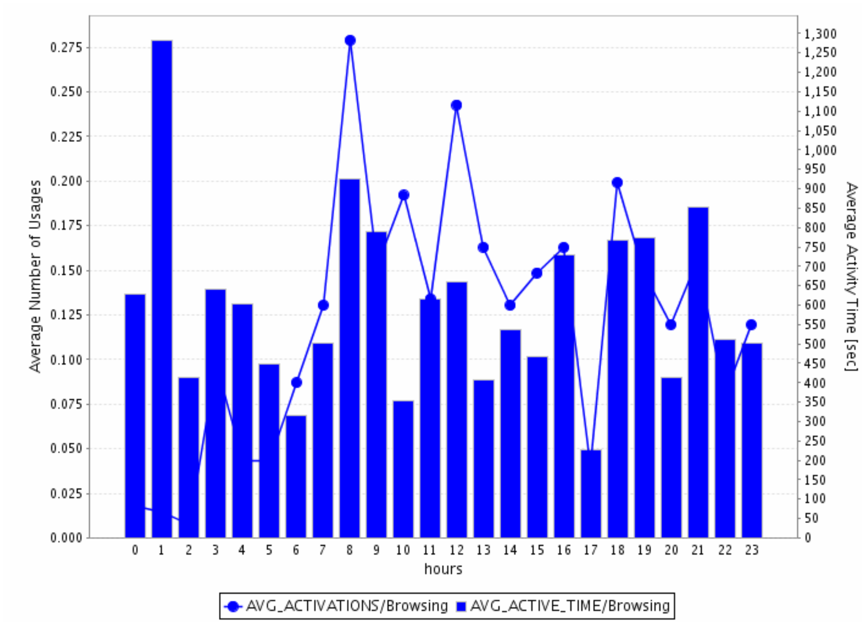


Figure 41 - Application – Browsing

- **Pictures and video**

This report shows that users take a quite amount of photos and few videos but only a very small part is used in MMS or email.

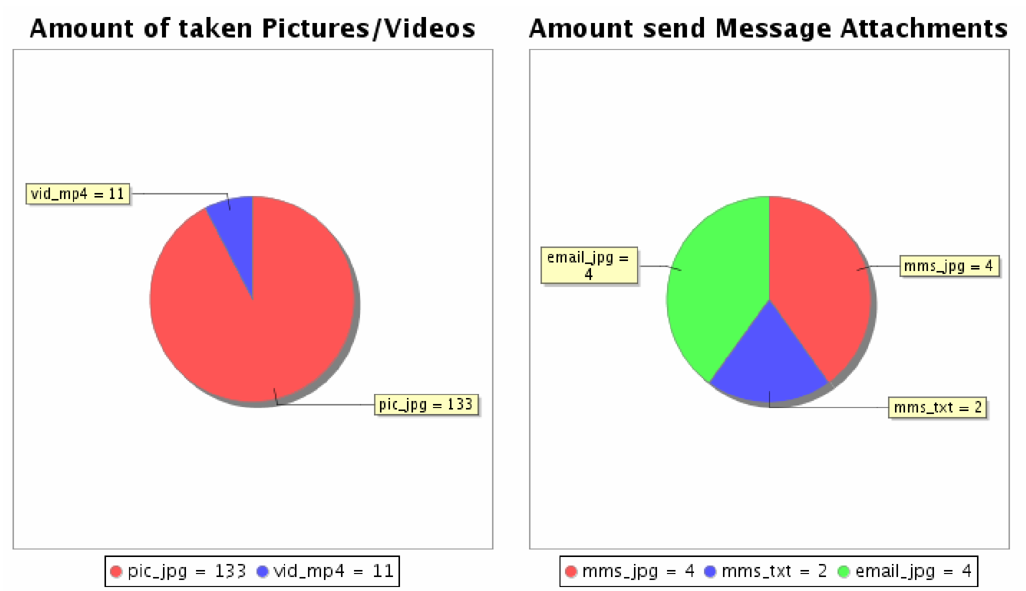


Figure 42 - Pictures and video

- **Camera usage vs Attachment sending**

As already seen from previous report, this one also shows that only a small amount of pictures are sent to other users.

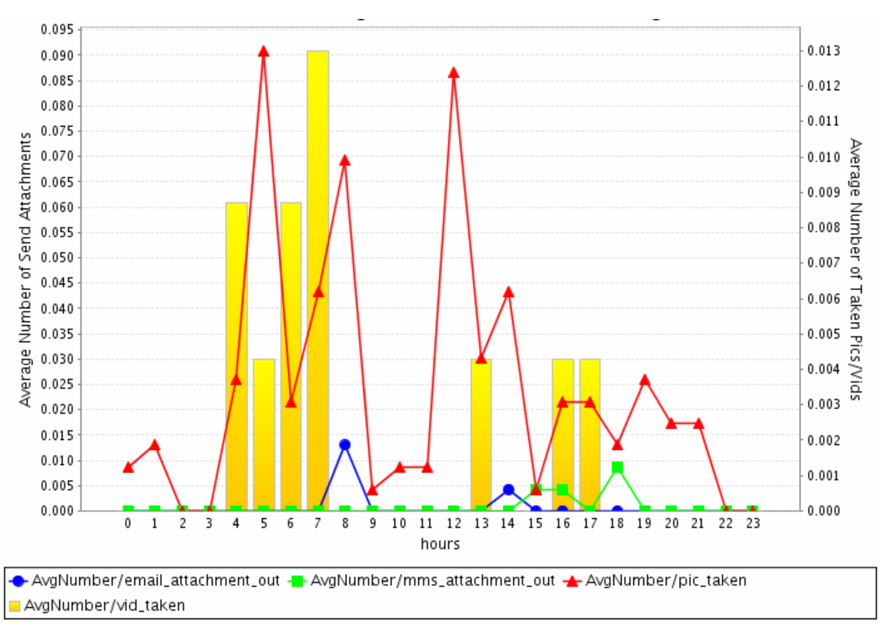


Figure 43 - Camera usage vs. Attachment sending

- **Web browsing usage**

Here is interesting to notice that "betfair" is one of the most visited url's. All the calls are done by 3G bearers, but this is because of my initial filter for 3G.

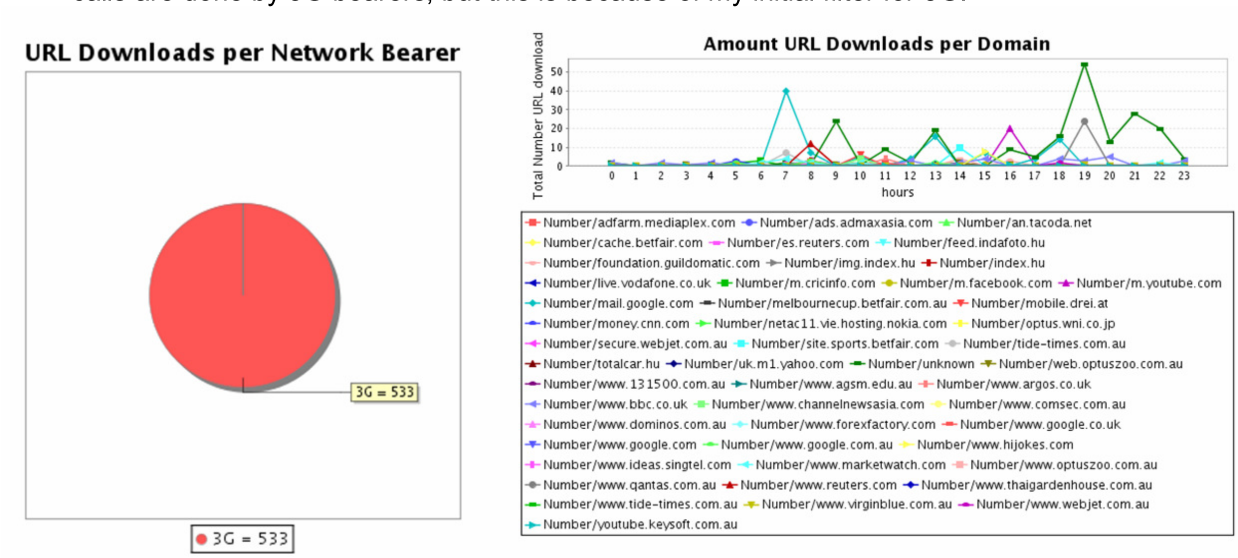


Figure 44 - Web browsing usage

- **Message and Call Stimulus**

SMS Stimulus

This report shows the amount of reactions of users after reading an SMS. Users normally react to the reception of an SMS in the next 5 min.

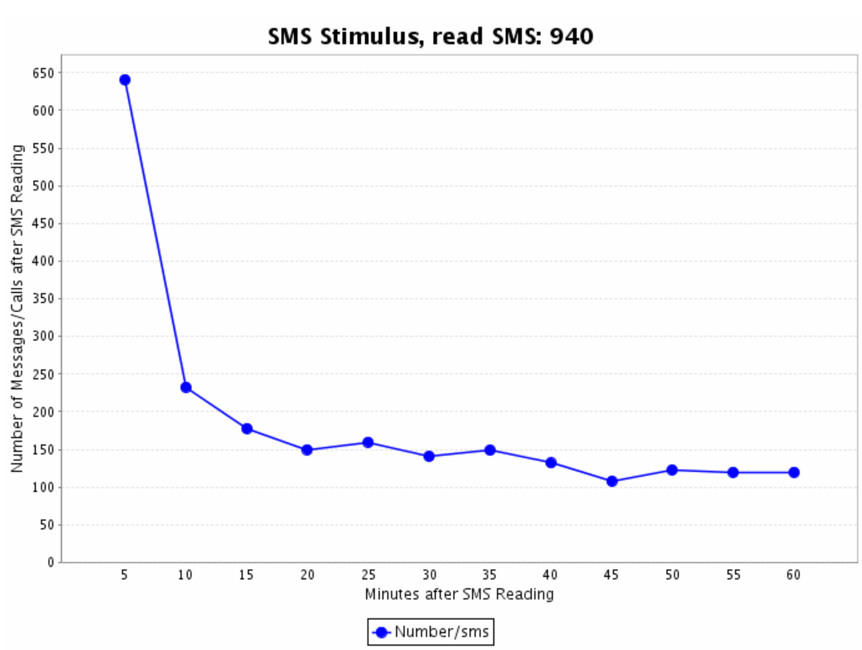


Figure 45 - SMS Stimulus

Voice call Stimulus

For voice call stimulus the reaction is different from the SMS. In this case, the reactions are more or less linear along the day. They are not focus in the next 5 min of answering a call.

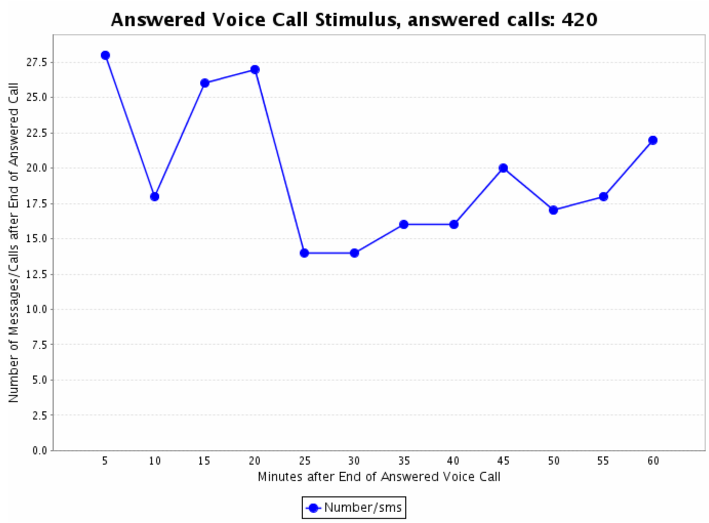


Figure 46 - Voice call Stimulus

- **Top N Report**

Top N report resumes the user's behaviour in the network. In this case the most used service is data calls because of my filter for 3G network. The second one is voice calls followed by SMS and browsing. It was also observed some usage of camera and calendar but almost none of the pictures are sent to other users. It was also important to verify that video calls and gaming have no activity.

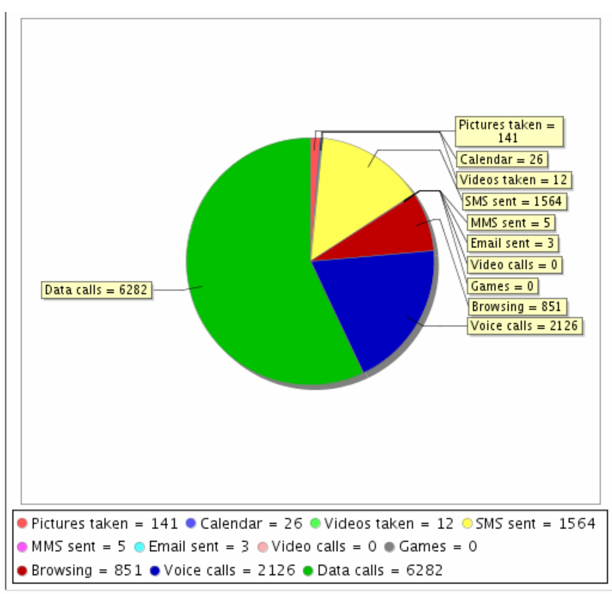


Figure 47 - Top N Report

7.3.2. Service Quality reports

- **Service Dashboard**

This report shows the dashboard of a general usage by service.

CS Call Service Summary:

This report is based on 2298 CS calls with a total duration of 6010.4 minutes and with an average duration of 2.62 minutes.
This report shows that the majority of voice calls were outgoing calls, then incoming calls, then incoming missed calls (because of no reply), then the busy calls and finally (with almost no percentage) the outgoing missed calls. Since I set the filtering for 3G cells, the report registered that almost all calls are 3G (the 2G calls are due to handovers from 2G cells). The report also shows that most of the calls are from the home network and only a few in roaming. The outgoing failures are mostly normal (no action from the called user), a small part are blocked (the called user rejects the call) and the rest were just dropped for some other reason. It is also possible to perceive that CS call setup time is between 3 and 8 seconds for almost all the calls. The empty graph for MOS (Mean opinion score) is because of the non-feedback from the users to CS call quality.

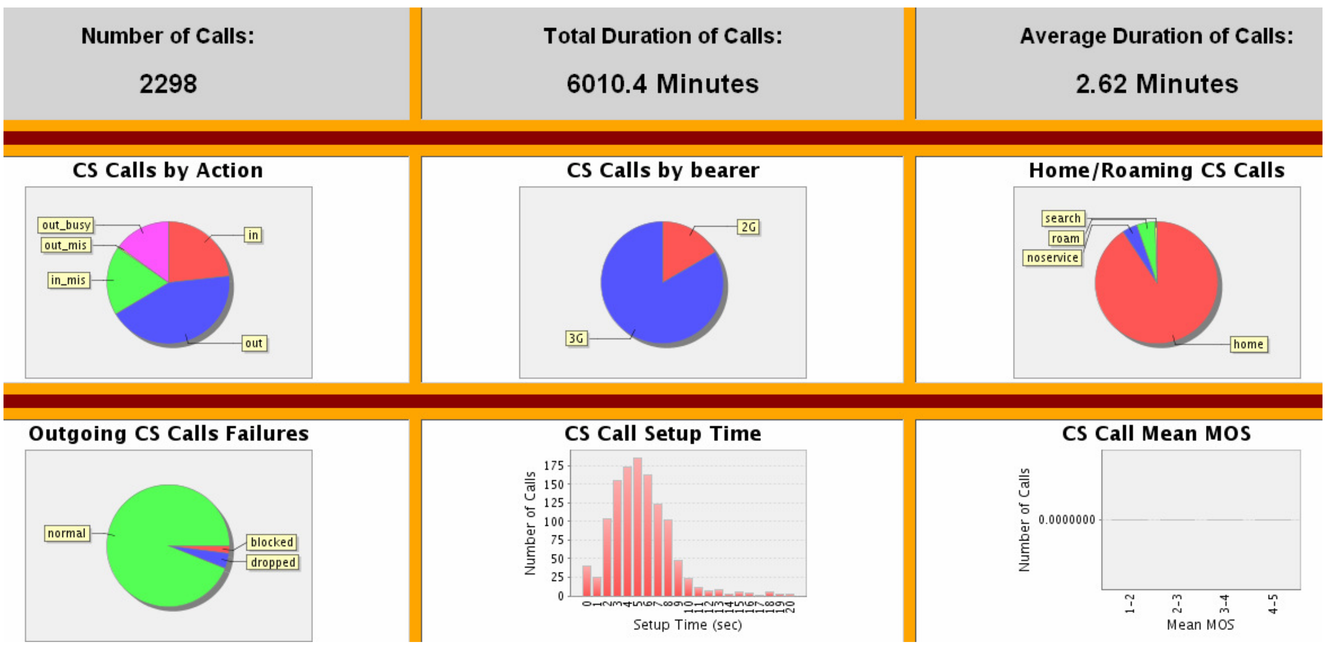


Figure 48 - CS Call Service Summary

PS Call Service Summary:

The PS call summary report has a total number of 3074 PS calls with the total duration of 11015.66 minutes and a total data traffic of 246.97 Mbytes. The average calls per user are 93 with the average duration of 3.58 minutes and with an average traffic per call of 97.91 Kbyte.

Almost all calls are in a WCDMA network and bearer because of my 3G selection. These calls are in home network and the major part of failures during the Setup and other part are related with PDP failure. The setup time is around 2 sec and the access point with more requests is the internet followed by live.vodafone. The major part of traffic is for Browsing.

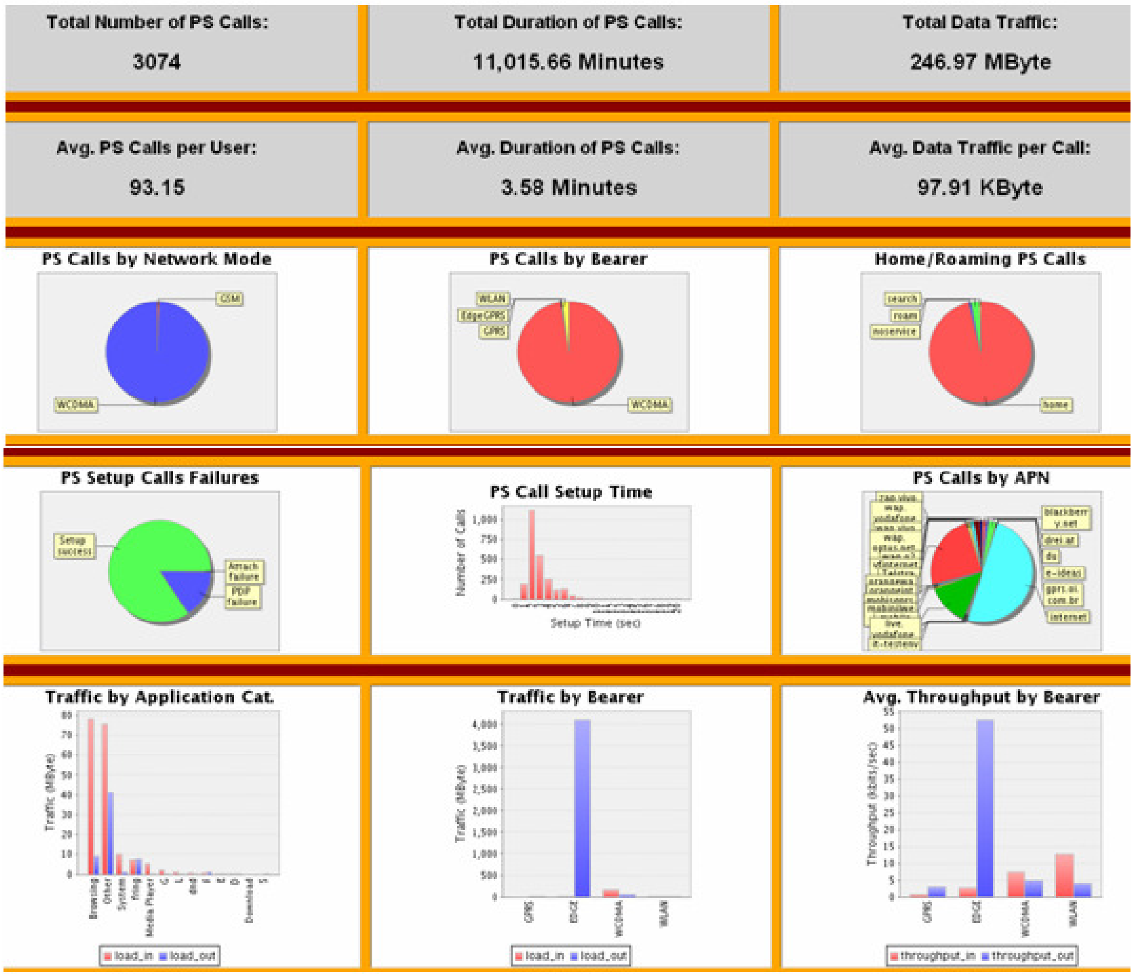


Figure 49 - PS Call Service Summary

SMS Service Summary:

The panellist of SMS service contains 887 received SMS and 723 sent SMS with the average SMS size of 72 chars. The major number of SMS is in the home network and in a 3G bearer. There were only two SMS failures in this panellist and the send duration is around 2-3 seconds.

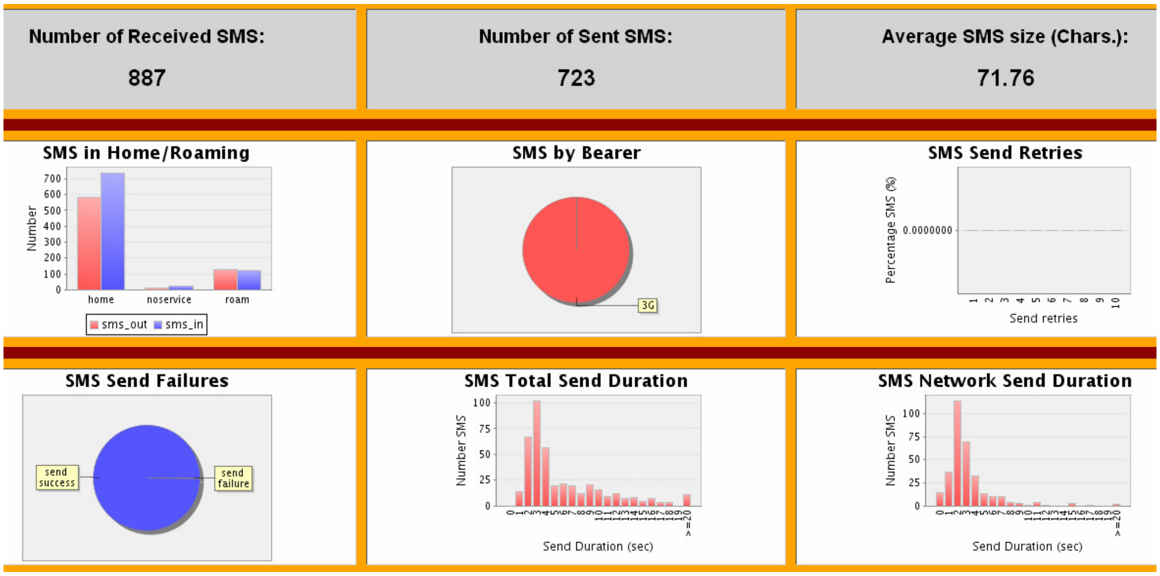


Figure 50 - SMS Service Summary

- ## CS Call Service

At this point, is possible to observe an average accessibility ratio around 98 % and call duration around 5 minutes. The day with more calls is Thursday and on Saturday is when the calls have a longer duration.

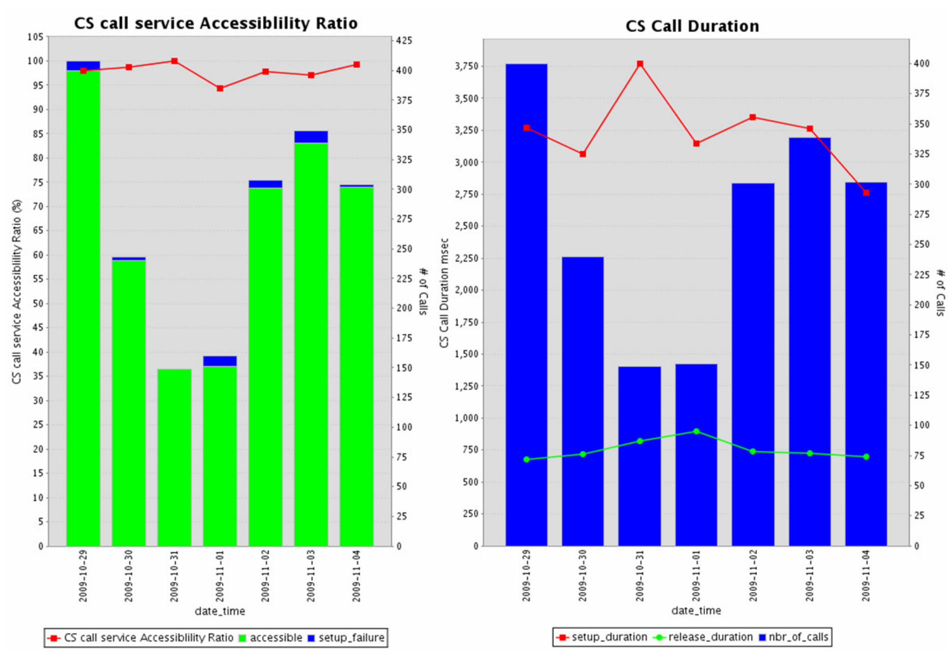


Figure 51 – CS call service Accessibility Ratio and Call Duration

The release of a call occurs in most of the times in the first 100 msec and the call success rate is 96.5 %. This call success rate is bad (good values are around 99%) and this is one of the elements that normally differ when measured in the end user side or in the network side. This happen because from the user point of view we achieve the call success rate for a specific user and from the network point of view we achieve the average call success rate of a group of users.

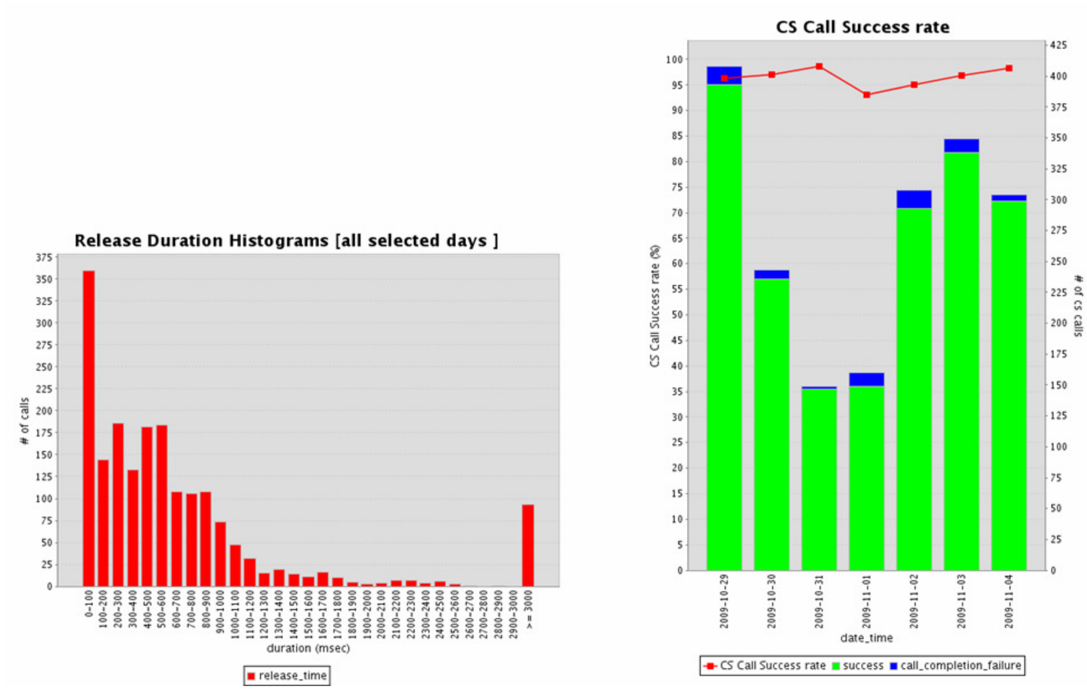


Figure 52 - Release Duration and CS Call Success rate

- ## PS Call Service

For PS service, it is possible to observe a very high setup failure ratio of 15.3 %. Possible causes for this failure ratio could be problems in the attach procedure and in the context activation procedure. Nevertheless, a deeper analysis shows that the context activation takes only 15% of the failures. From the bottom graphic, we can see that almost all the failures are from an unknown reason. In this case, it is necessary to do a deep analysis in the reporting tool, more specifically in the core part to understand why context activation is failing.

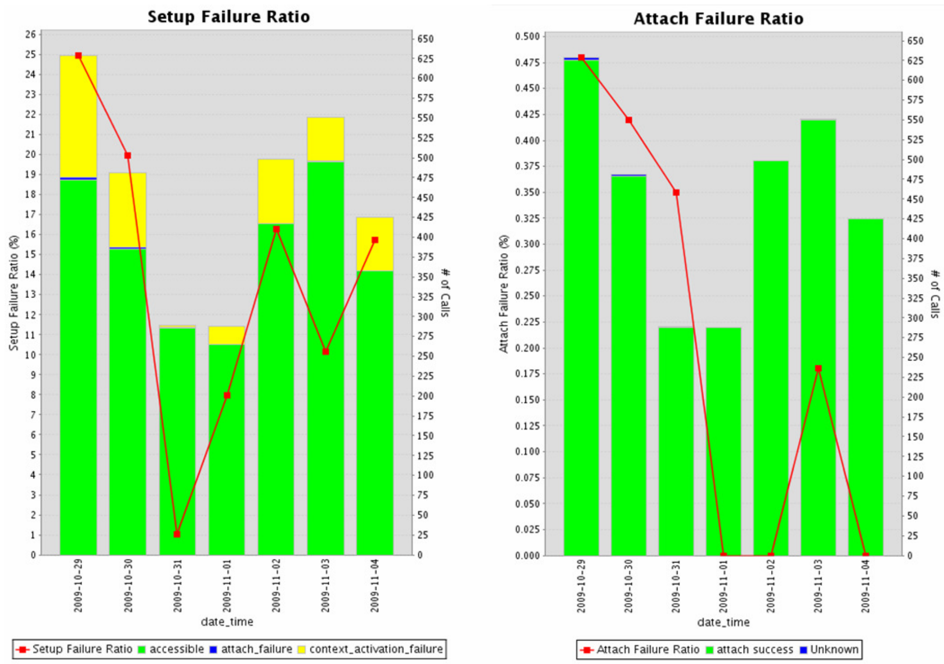


Figure 53 - PS Call Service

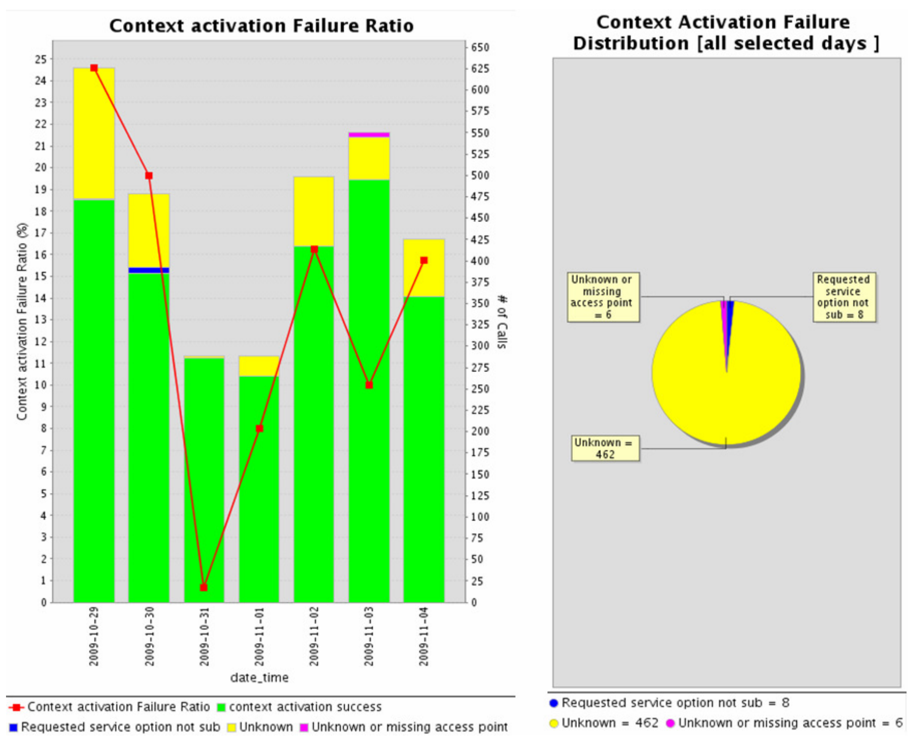


Figure 54 - Context activation Failure Ratio and Failure Distribution

The PS call duration is high not because of attach time (time when user is trying to connect to the network), that is only 2 sec, but because of context activation time (time that the user is negotiating connection parameters with the network) that as an average of 21 sec. The average data traffic is 85105 bytes.

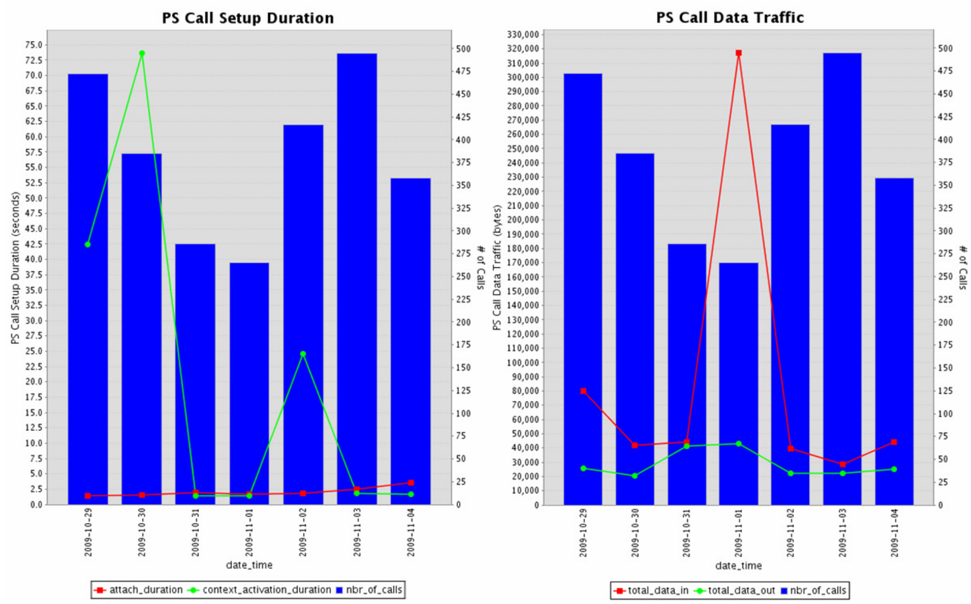


Figure 55 - PS Call Setup Duration and Data Traffic

- ## Browsing Service

For the browsing service, the metrics achieved are 99,76 for accessibility, 45 sec of webpage duration, 9 sec for page download, 17 Kbps for downlink throughput and 2,5 Kbps for uplink throughput. Although all metrics reported before are ok, the webpage success ratio is only at 79%. So, a detailed look should be taken in the network metrics to understand the source of the problem.

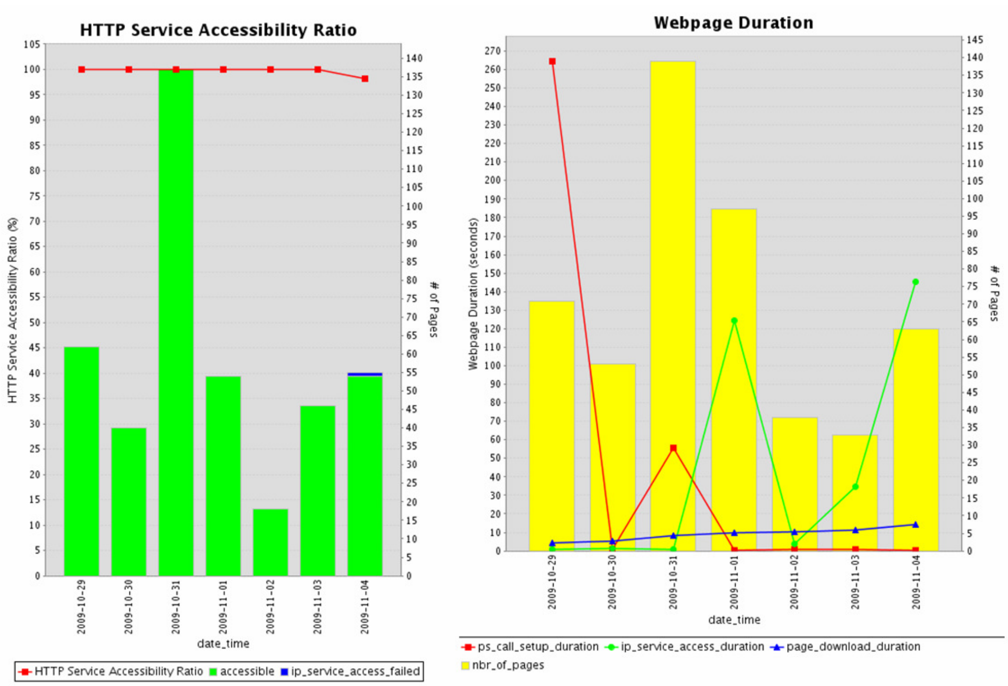


Figure 56 - Browsing service

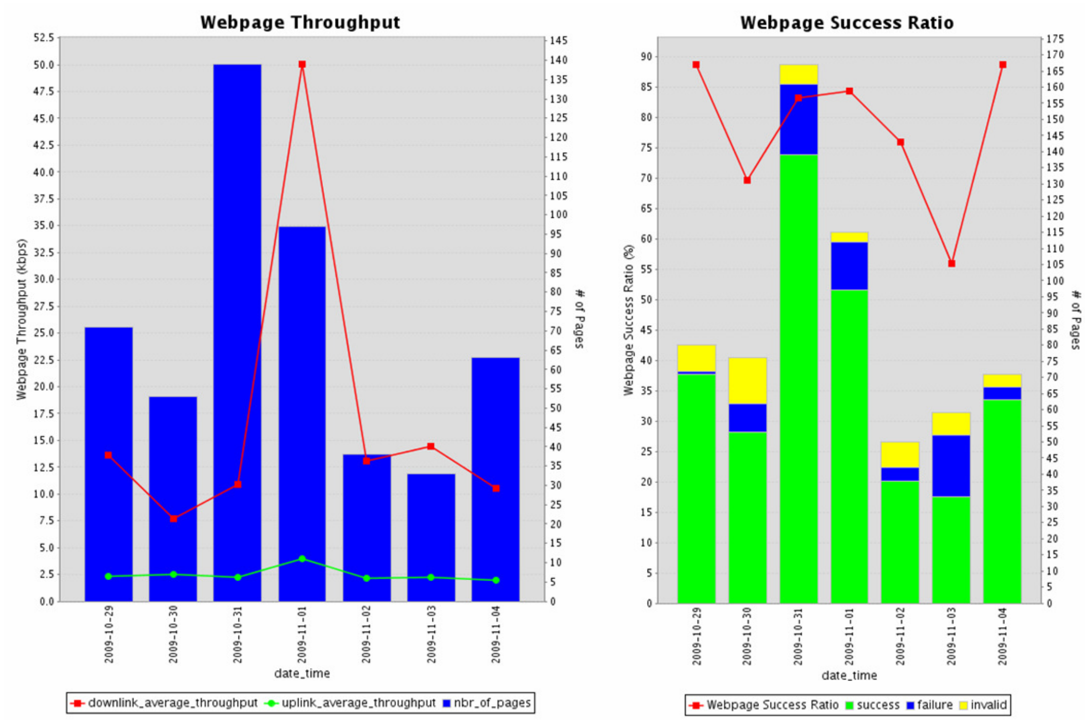


Figure 57 - Webpage Throughput and Success Ratio

- ### SMS Service

The SMS service shows an accessibility success ratio of more than 99%, an average duration of 6,5 seconds and an average message size of 67 char per message. All these metrics are OK and accomplish the requirements for SMS service.

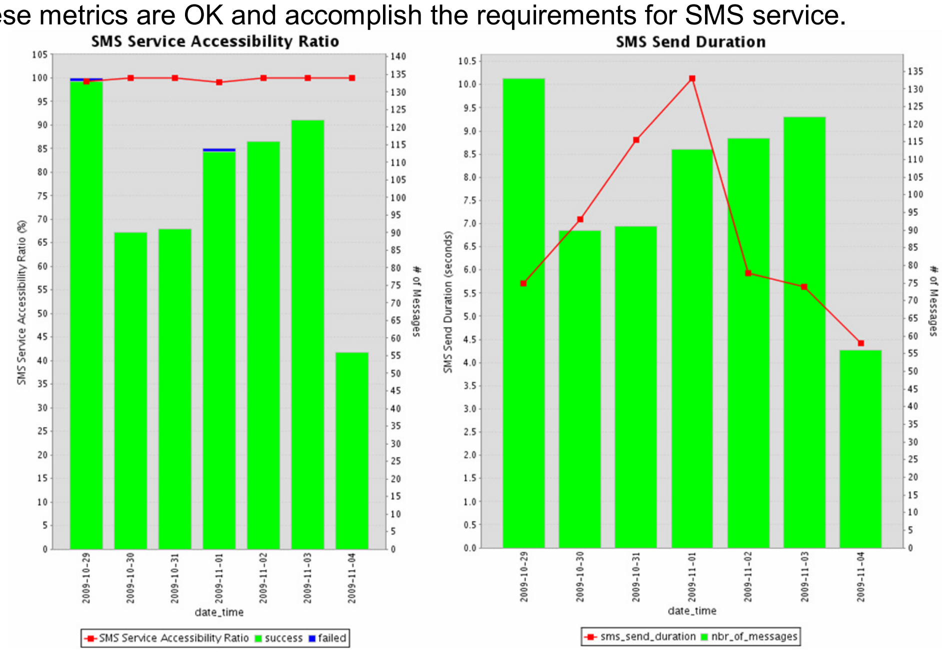


Figure 58 - SMS Service

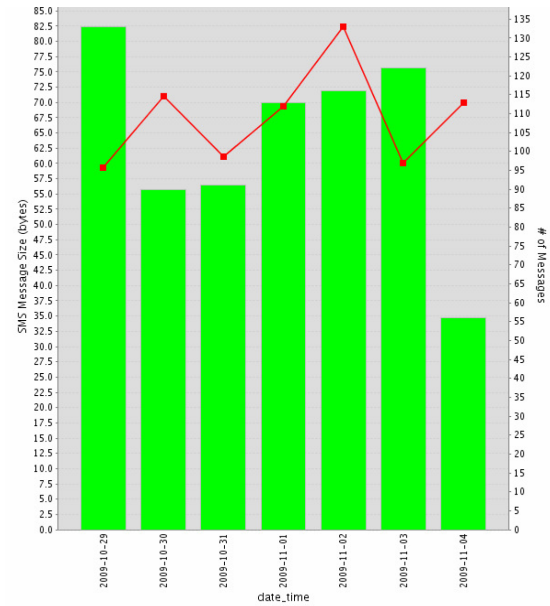


Figure 59 - SMS Message Size

7.3.3. Service Agent Based KPIs

At this section I present some KPIs based on metrics collected by the agent.

- **Voice service**

In Figures 60 and 61 is possible to observe values between 6 and 10 seconds for setup time and it is also visible the Rx_Level per cell (this is just an example of the amount of parameters that could be correlated with all the collected metrics). Since metrics are collected by the agent, we don't have detailed information on the network status. To complement this analysis, we can always go to the reporting tool and analyse some available KPIs such as the RRC access and setup, RAB access and setup, RRC and RAB attempts, service access times and many others.

The call speech quality was gathered based on two methods: PESQ and MOS (Figure 62 and 63). The MOS is reported by users classifying call quality and PESQ is an algorithm. The quality was around 3.3 (between "fair" and "good" classifications) and very similar in the two methods although it is possible to observe that MOS method is slightly more aggressive because of higher values for good and bad calls. With the reporting tool, we can analyze AMR status in the network (capacity, allocations …), analyze spreading factors, codes in use, and many others to correlate with this speech quality report and discover what is affecting the end user quality.

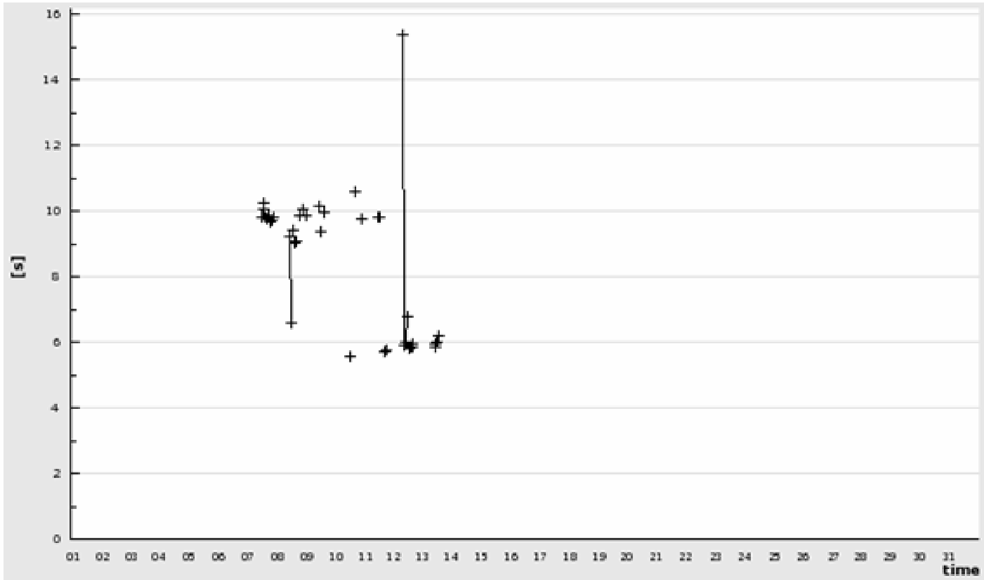


Figure 60 - Setup time for voice

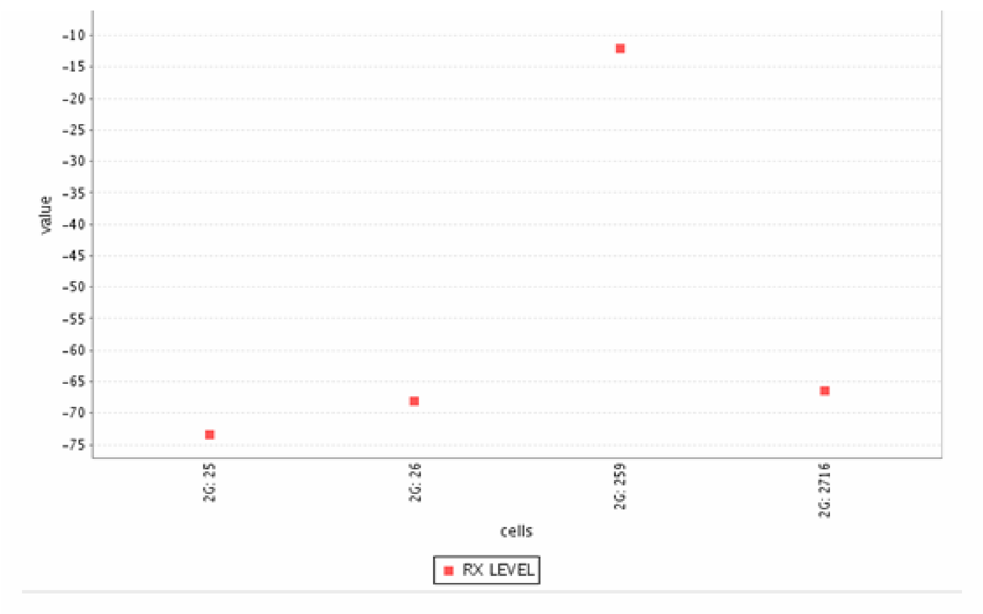


Figure 61 - Rx_Level per cell

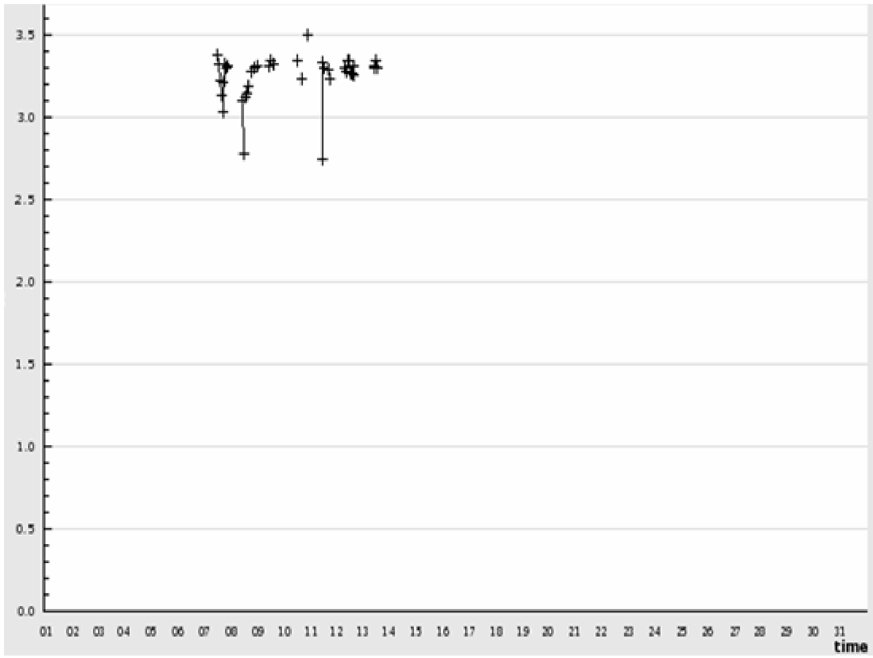


Figure 62 - Speech Quality- PESQ

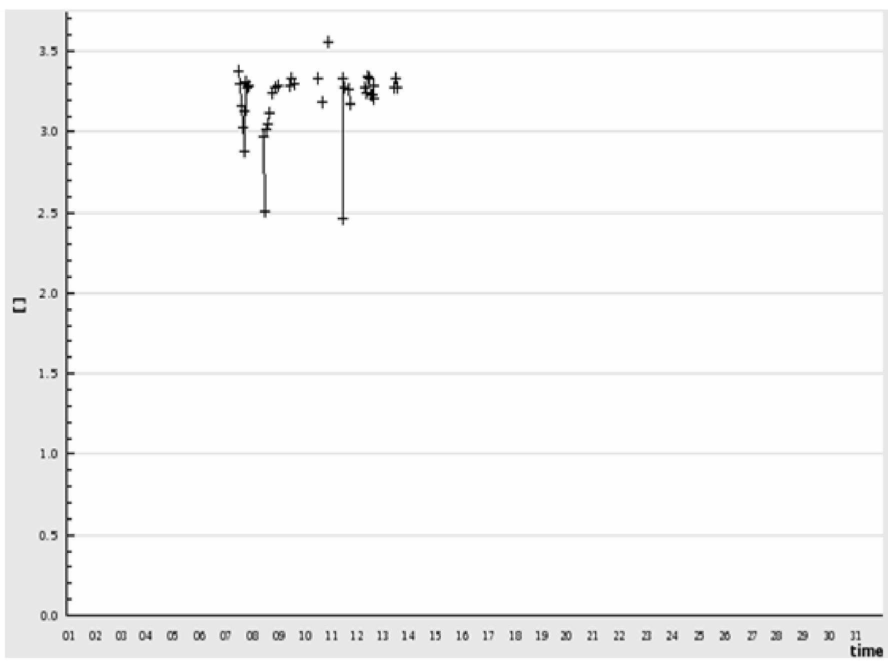


Figure 63 - Speech Quality- MOS

- **SMS service**

For the SMS service, the results show an accessibility of 100 % (Figure 64 a)), access delays and sending times around 6 sec (Figures 64 b) and 65 b)) and end-to-end delivery around 12 sec (Figure 65 a)). In this case the values are acceptable, but in case of facing some problems with this service we can get form the reporting tool detailed description about what is happening. Some KPIs like SMS attempts for users in idle mode, Rejected Gb MO SMS attempts, Success ratio of Gb CAMEL SMS delivery attempts, Iu MO SMS success ratio and others could be used to have a detailed perception of the service.

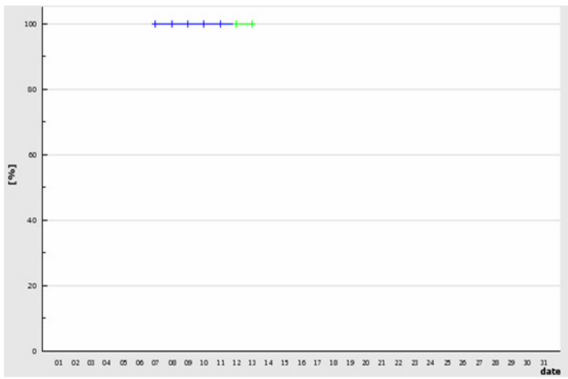
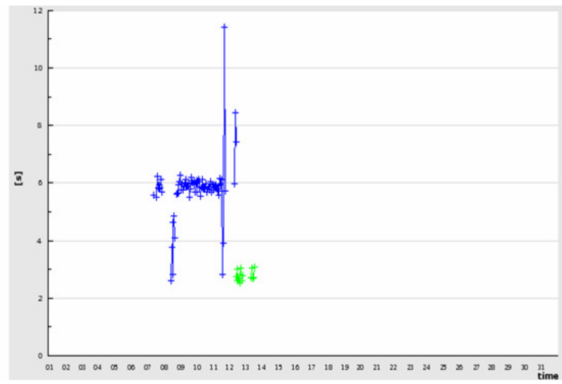


Figure 64 a) SMS accessibility



b) SMS access delay

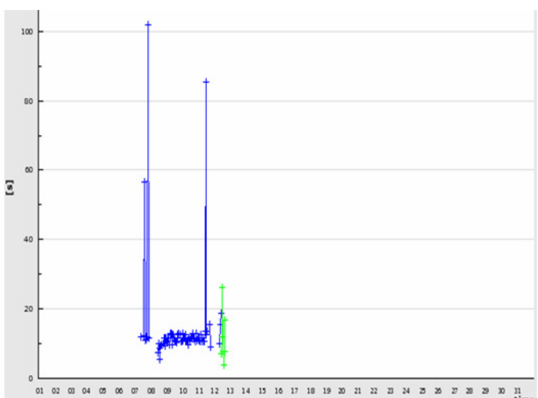
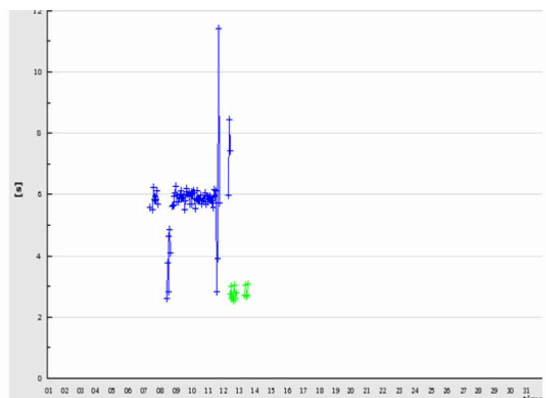


Figure 65 a) SMS end-to-end delivery time



b) SMS sending time

- **WAP service**

For WAP, service is presented the success rate including PDP and the delivery time.

This service exhibits a success rate of almost 100% all the time and the delivery time is between 30 and 70 sec. For the analysis of the browsing service in the network side, we could take into account some KPIs related with packet data such as Data dropped per session, Data Volume for several NE's, Data throughput, HSDPA schedule priority indication, Data discarded ratio and others.

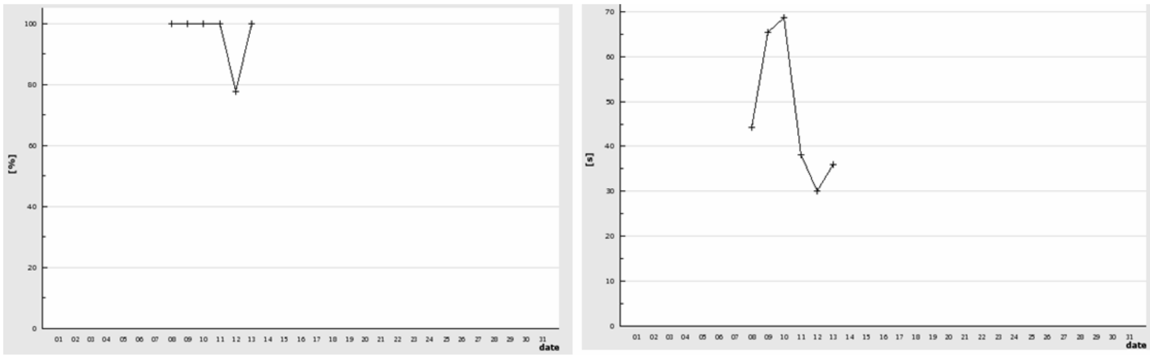


Figure 66 a) WAP success Incl PDP **b) WAP delivery time**

- ## General packet data

Here, we can observe a session time around a minute and a good round trip time of 0.3 sec. It is also observed a service setup time of 5 sec and a mean data rate of 100 Kbits/sec. For the analysis packet data in detail from the network side, we have lots of KPIs in the reporting tool like the following: Data volume per interface, HSUPA/HSDPA data, CS data, Throughput per class service, PS allocated capacity, Peak number of HSPA users and Cell availability.

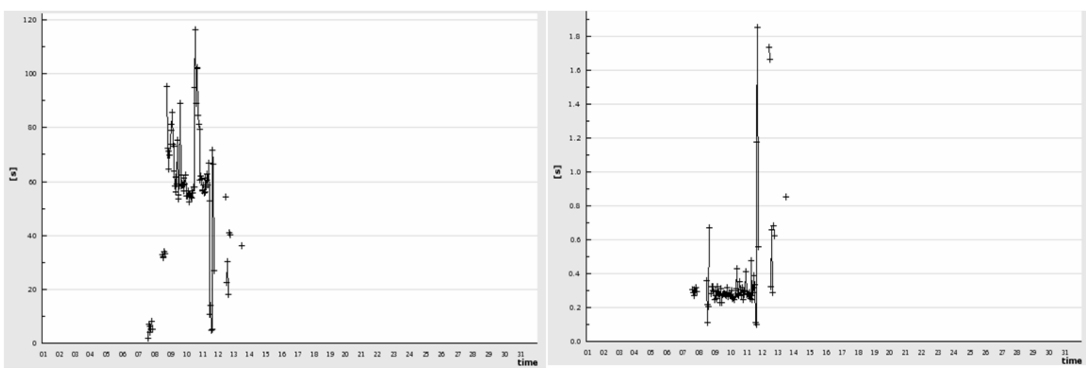


Figure 67 a) Packet session time **b) Round trip time**

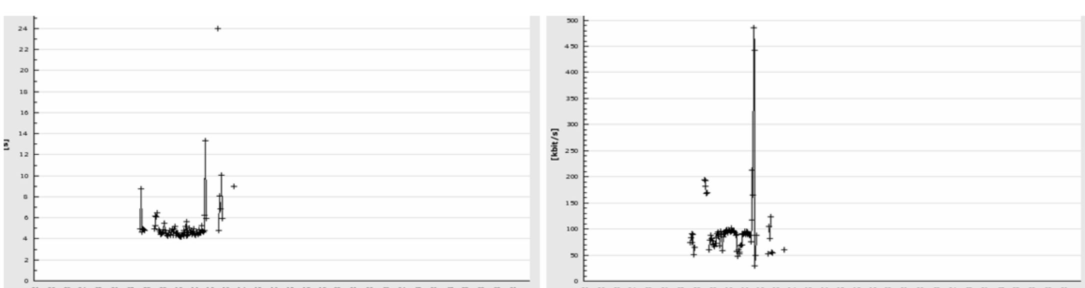


Figure 68 a) Service setup time **b) Mean data**

7.4. Network Performance Results

At this chapter, I present the results acquired from a live network with the KPIs specified in chapter 5.2.

7.4.1. Cell Quality indicator

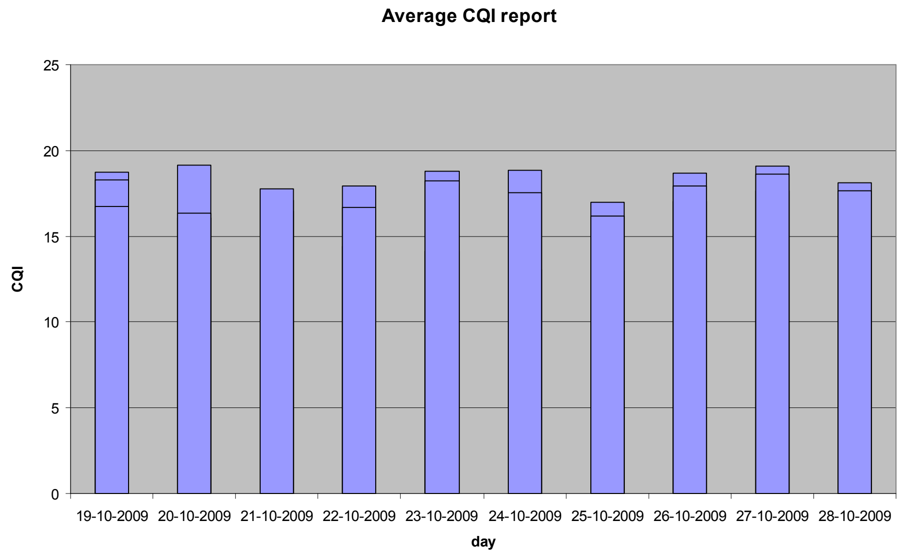


Figure 69 - Cell Quality indicator

Cell Quality indicator shows a status of the quality in a cell. Cell quality values are from 0 to 30 (where 0 is the worse and 30 is excellent). Since in this network the CQI is around 17, it is possible to classify the cell as good quality. Normally values around 13 and 14 are still acceptable, but bellow these values, the cell is very bad.

7.4.2. Cell Throughput Downlink

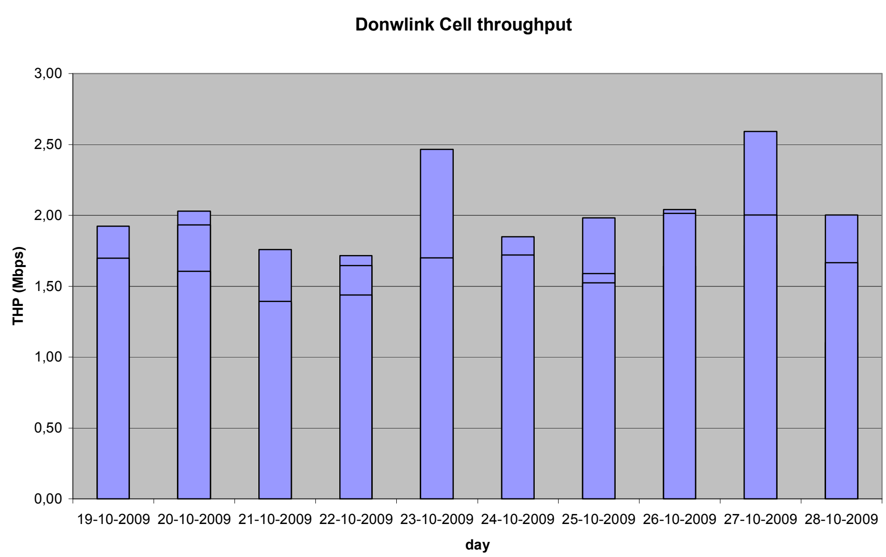


Figure 70 - Cell Throughput Downlink

This graphic presents the cell throughput for downlink direction. The achieved Average Throughput of 1,7 Mbps is very acceptable for this operator network.

7.4.3. Cell Throughput Uplink

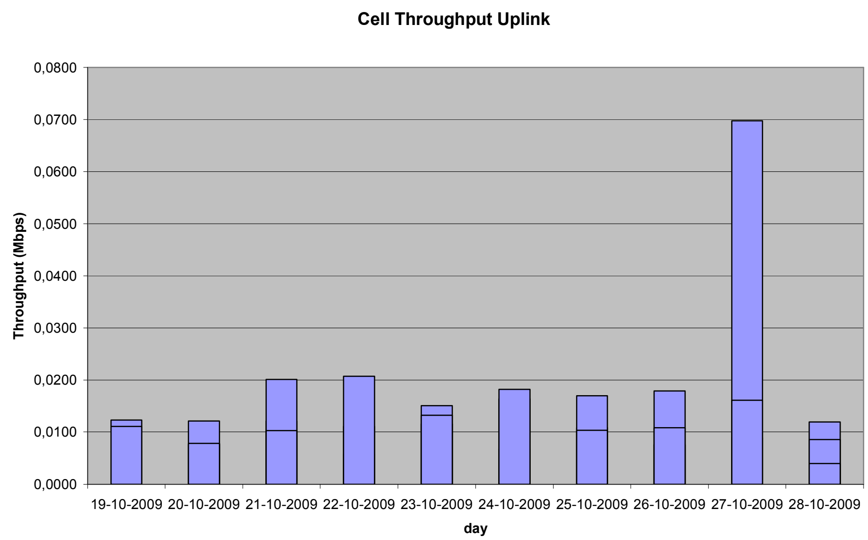


Figure 71 - Cell Throughput Uplink

This graphic presents the cell throughput for uplink direction. It is possible to observe that uplink traffic is very small in the network, and has almost no expression when compared to the downlink.

7.4.4. Cell Availability

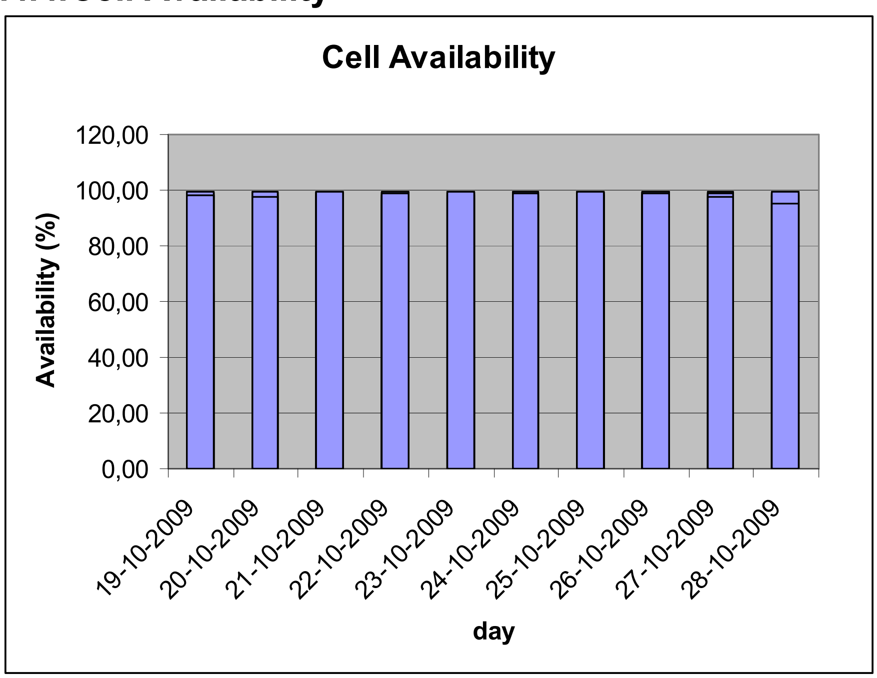


Figure 72 - Cell Availability

Cell Availability is always near 100% as it was expected. In case of cell related problems this is the first KPI that alarms the situation.

7.4.5. Data Downlink service accessibility

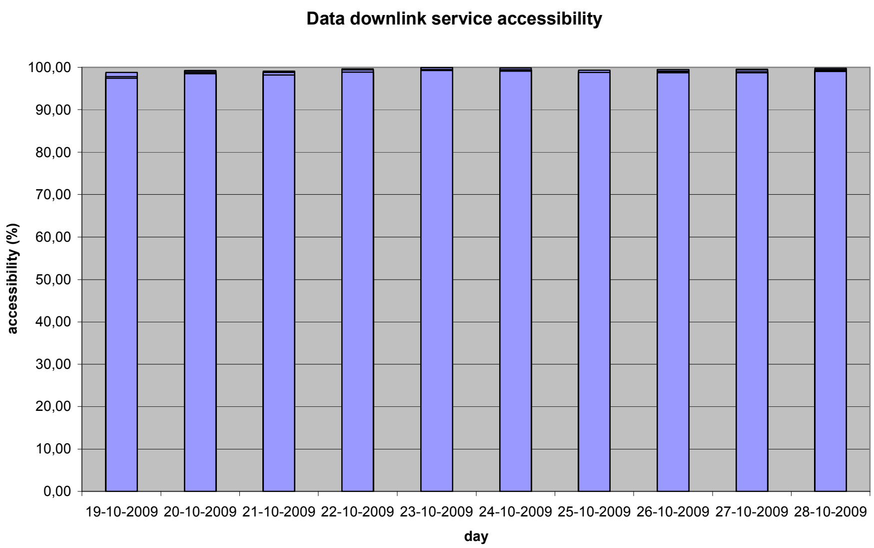


Figure 73 - Data Downlink service accessibility

Data Downlink service has a very high accessibility around 99%. As expected is possible to perceive that this accessibility good but less that cell availability observed before (section 7.4.3 around 100 %).

7.4.6. Data Uplink service accessibility

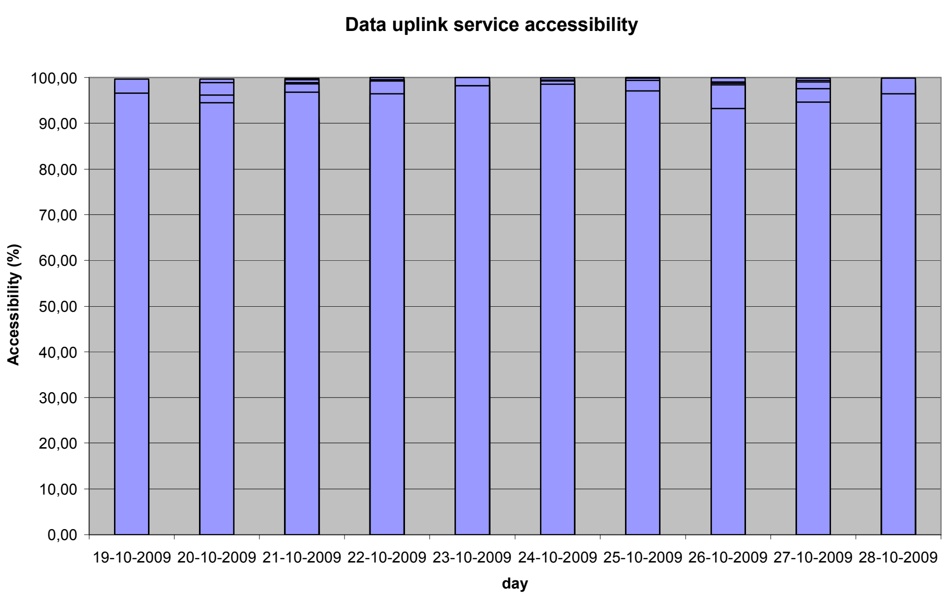


Figure 74 - Data Uplink service accessibility

Data Uplink service shows very good values of accessibility for this network.

7.4.7. Call Setup Success ratio

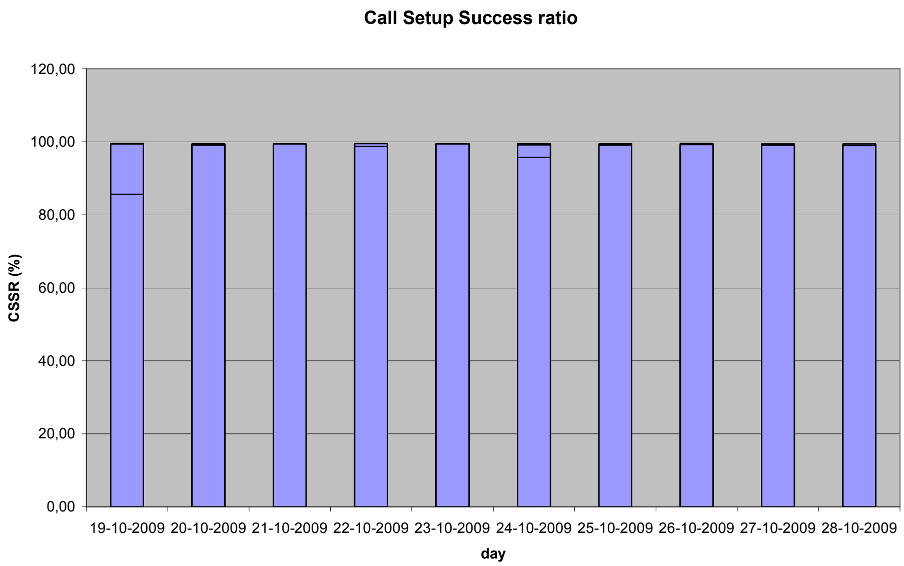


Figure 75 - Call Setup Success ratio

Call Setup Success shows normal values around 99%. For the most Europe mobile operators, values less than 96%, depending on the geographical area, are considered bad.

7.4.8. Channel elements utilization ratio

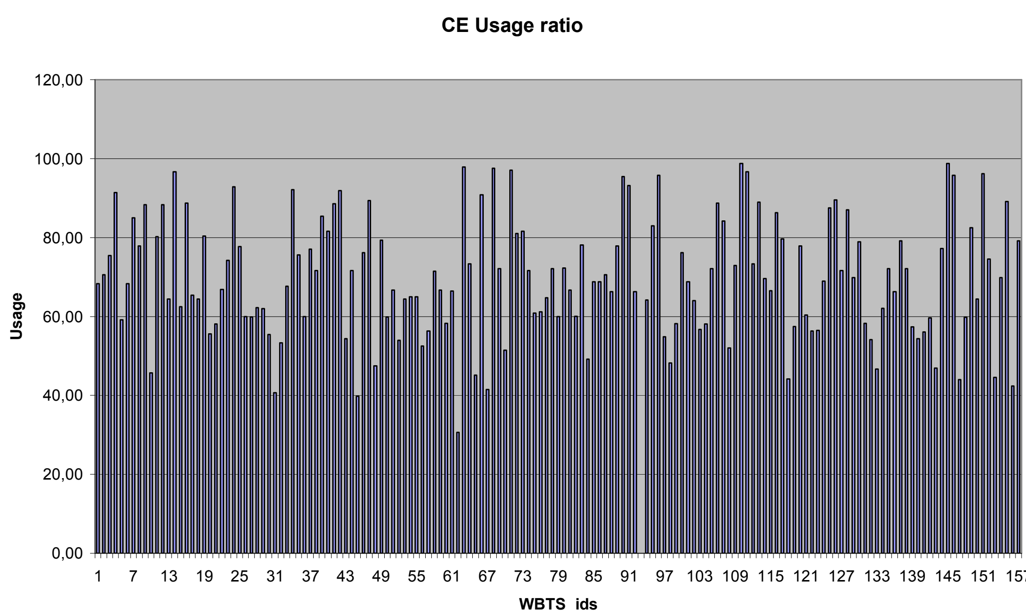


Figure 76 - CE Usage ratio

Channel elements utilization ratio around 60% is good. When values start to become around 80% operators should consider upgrading HW resources. I should refer that in the graphic I changed the WBTS ids for a sequential number to not show confidential data.

7.4.9. HSPA Users (Wcell level)

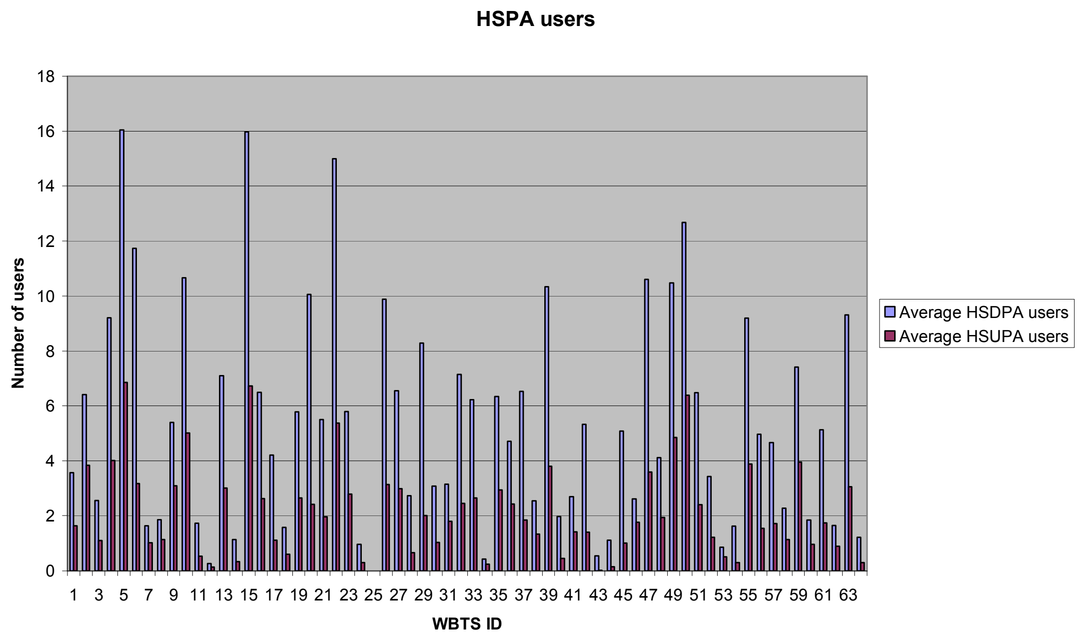


Figure 77 - HSPA users

The number of HSPA users is the contribution of the downlink users and uplink users. It is possible to observe that the number of downlink users is more than two times the number of uplink users.

7.4.10. Max HSPA Users (RNC level)

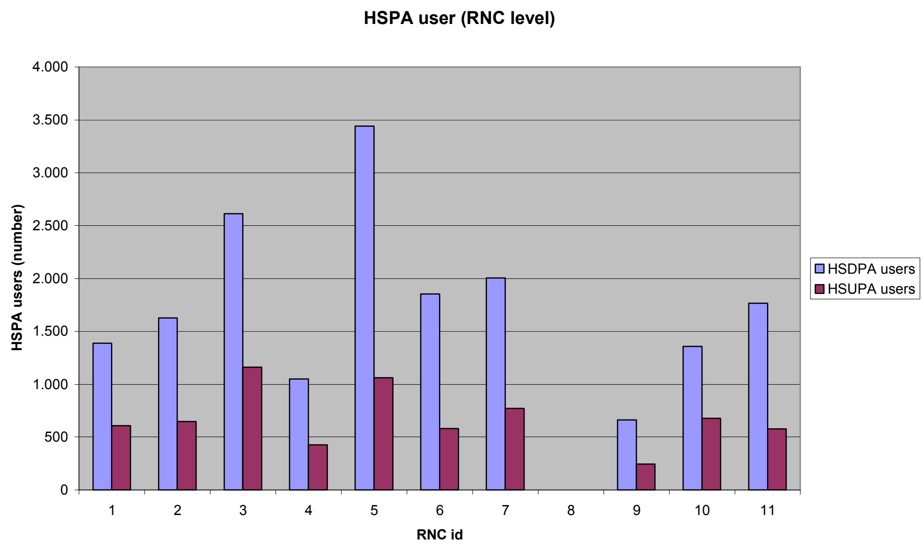


Figure 78 - Max HSPA Users (RNC level)

Here, we observe the number of HSPA users by RNC. I should refer that in the graphic I changed the RNCs ids for a sequential number to not show confidential data.

7.4.11. HSPA Throughput (RNC level)

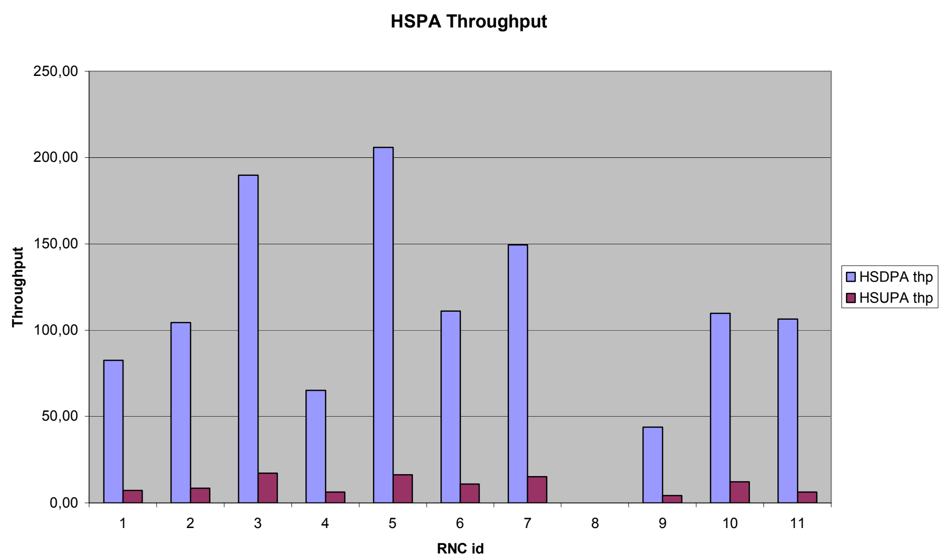


Figure 79 - HSPA Throughput (RNC level)

This KPI is useful to understand how the traffic is being handled in the several RNCs and how is used for RNCs traffic load distribution observation.

7.4.12. RRC setup and access call ratio

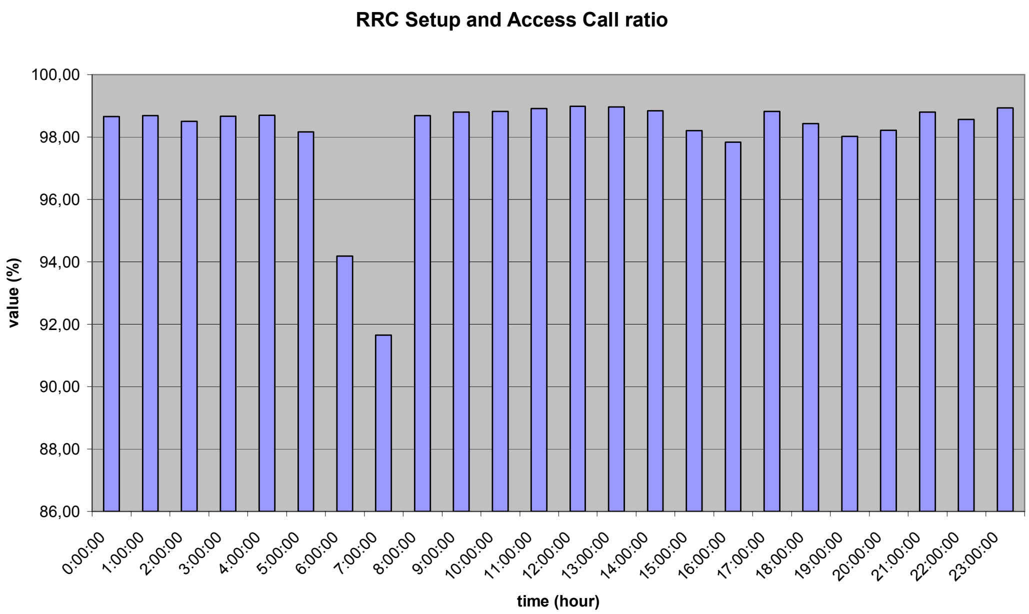


Figure 80 - RRC setup and access call ratio

RRC Setup and Access Call ratio is a very important indicator of how easy user gets access to the network. The target for this KPI should be values around 99%. In this case the KPI shows values near the target, except for 6:00 and 7:00 am. This could be related with some HW/SW upgrade issue.

7.4.13. RRC Fails per cause

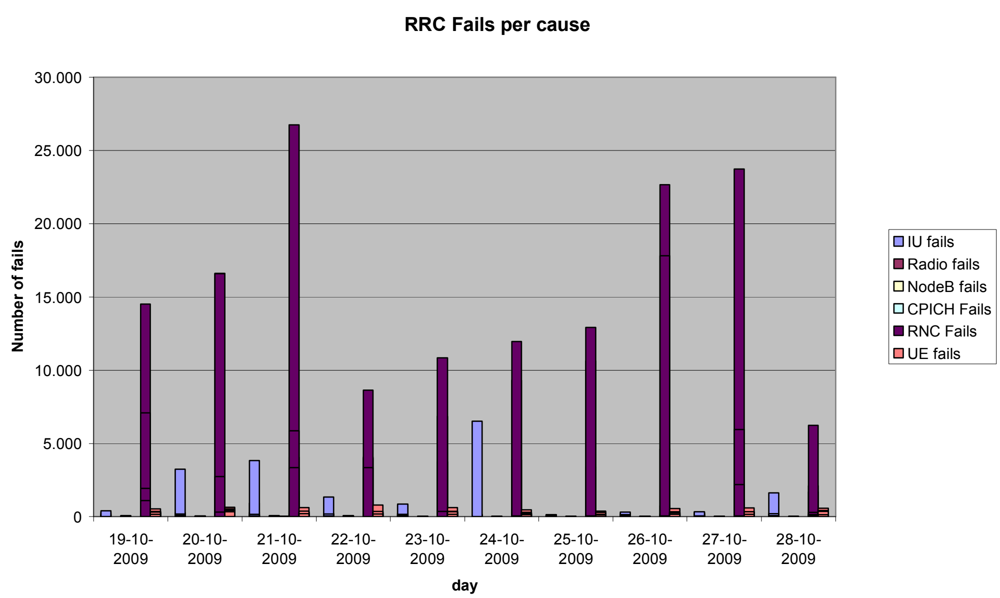


Figure 81 - RRC Fails per cause

Radio Resource Control fails is one of the most common problems in a network. Here, a detailed analysis of RRC drops, in this network, show that the bigger contributors are RNC and Iu interface fails. There is still a small amount related with the UE.

7.4.14. RAB Setup and Access ratio

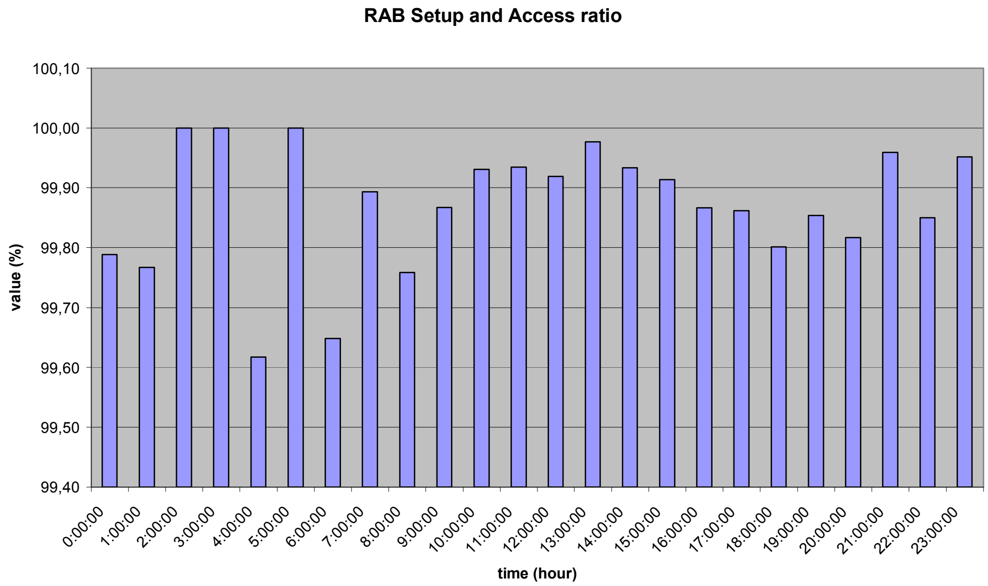


Figure 82 - RAB setup and access call ratio

Together with RRC setup and access ratio, this KPI shows user accessibility to the network. The target value is also more than 99%. In this case, the monitored network has very good values.

7.4.15. Packet service RAB drops per cause

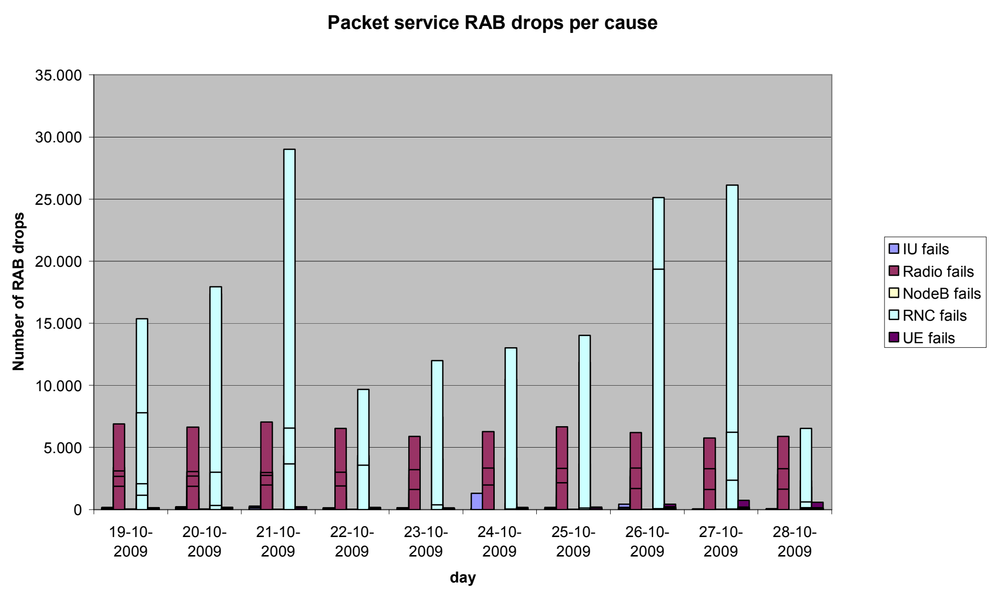


Figure 83 - Packet service RAB drops per cause

Radio Access Bearer drops together with RRC drops fulfil almost all radio related call drops in the network. This KPI shows that in this network, the highest contributors for RAB drops are RNC and radio interface fails.

7.4.16. PRACH Propagation delay

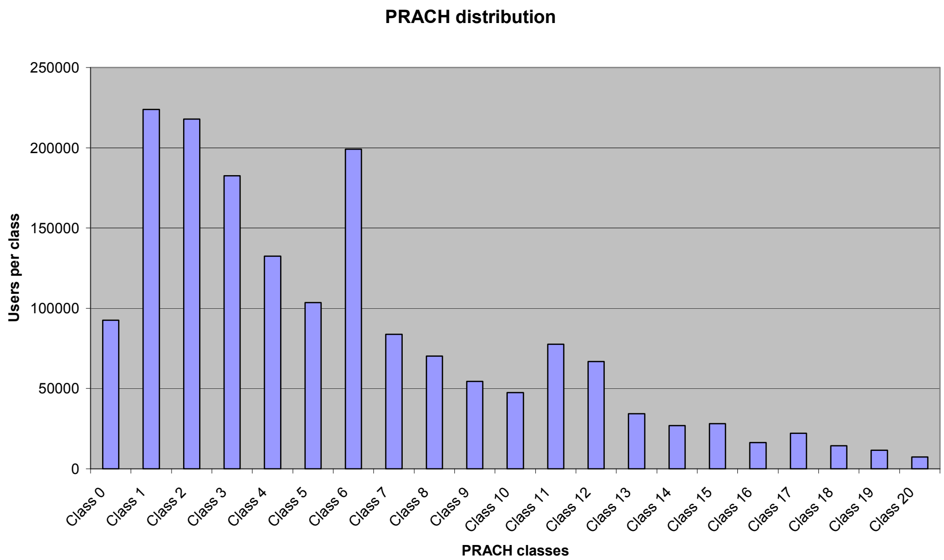


Figure 84 - PRACH propagation delay

PRACH propagation delay shows user distribution, in distance, to the NodeB. In this graphic, we could understand that we are monitoring an urban zone, because users are very close to the NodeB and not very dispersed. Note that PRACH is classified in classes from 1 o 31, were 1 is very close to NodeB and 31 is very far.

7.4.17. UE Power Headroom

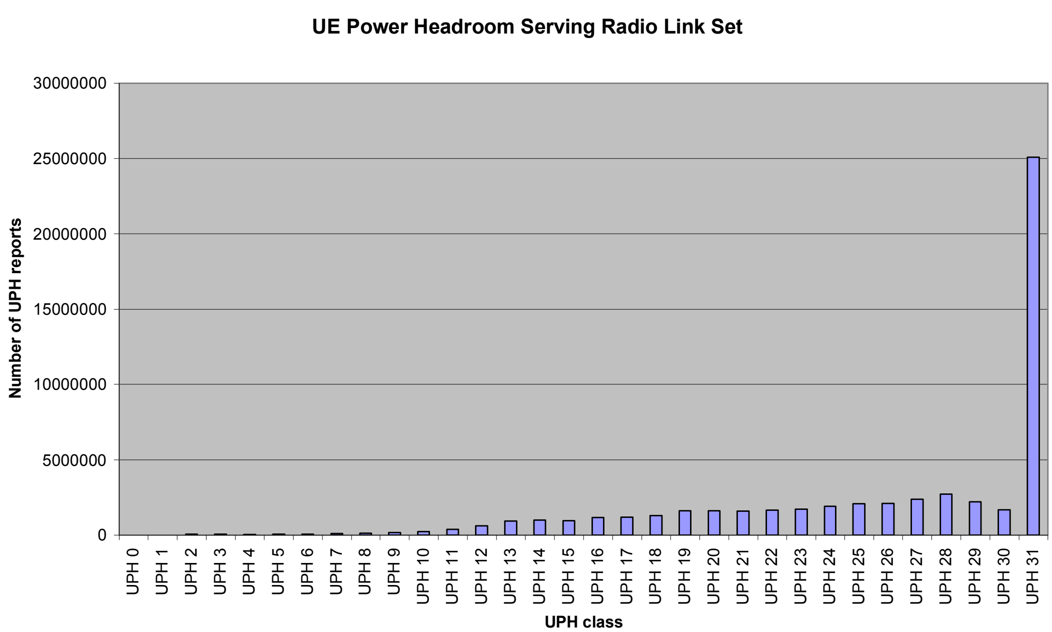


Figure 85 – UPH for serving radio link

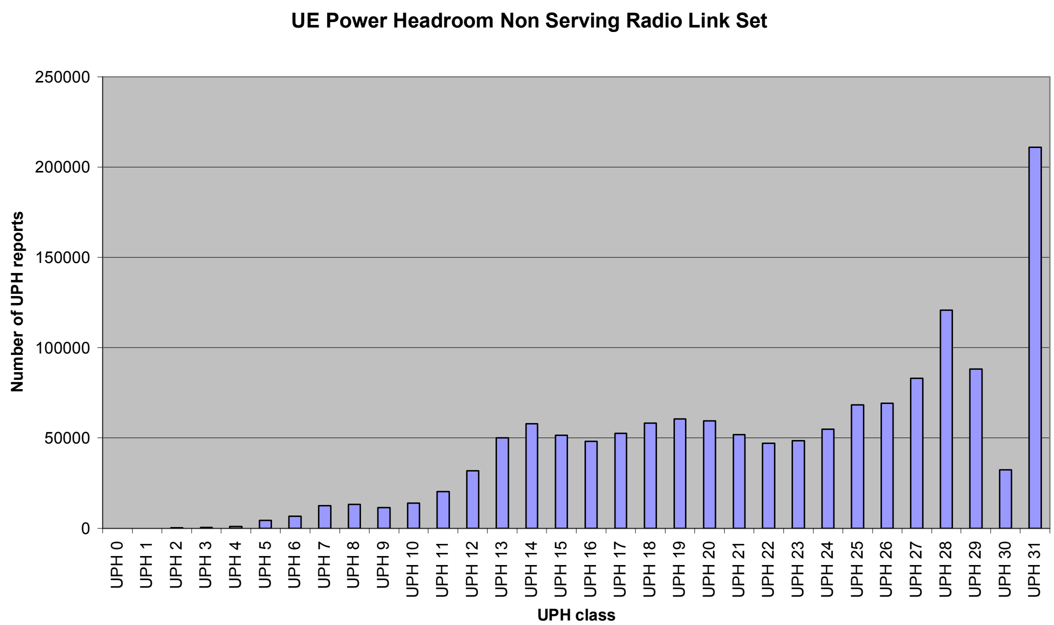


Figure 86 - UPH for non-serving radio link

Since UE power headroom is the division of the UE maximum transmission power and the corresponding physical channel code power, it is possible to understand from previous graphics, that the serving radio link (connection that actual is being used to transmit data) has better values than the non-serving radio link. This KPI is important to monitor the relation between UE power and network physical channel power.

7.4.18. Average Traffic per call

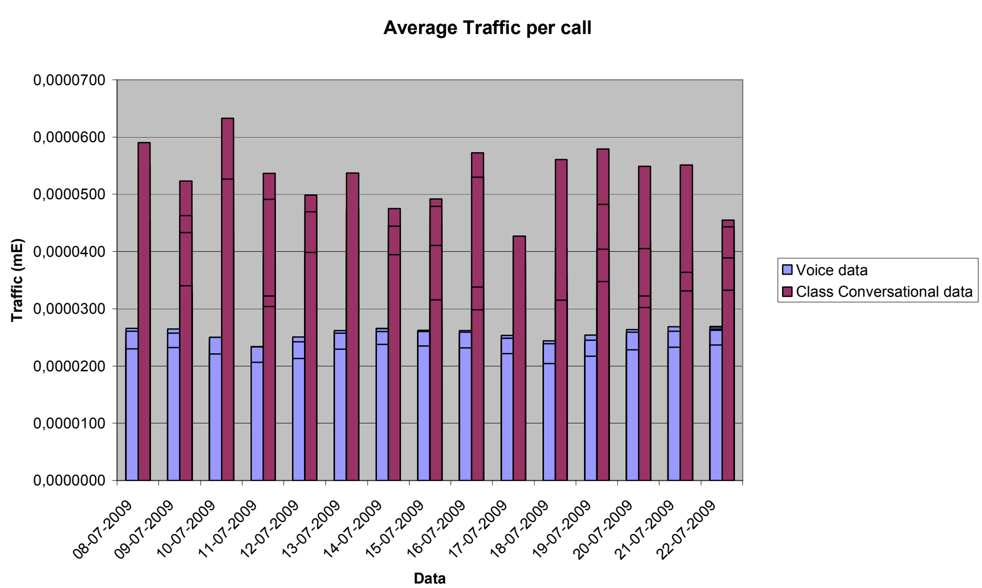


Figure 87 - Average traffic per call

This KPI shows that the average traffic per call is more for conversional class than for voice as it is expected.

7.4.19. Iub FP Layer data Throughput

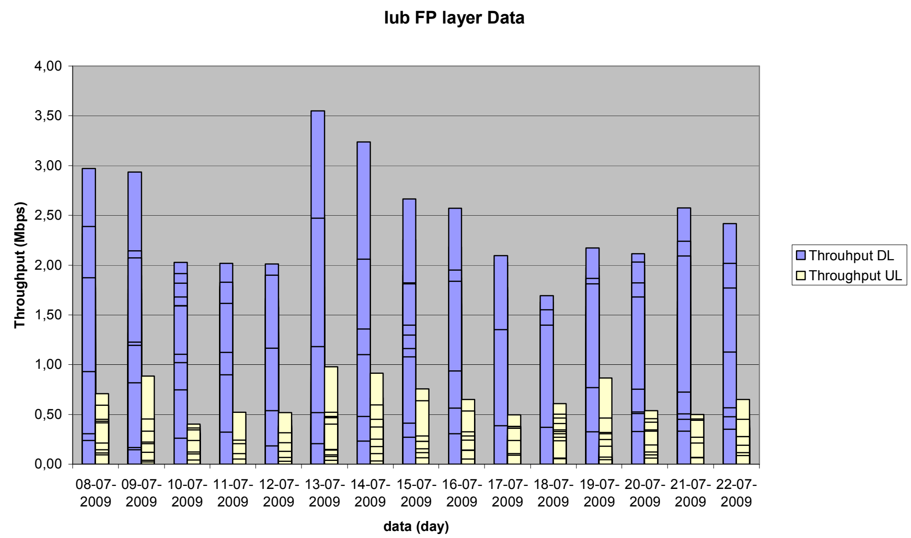


Figure 88 - Iub FP Layer data Throughput

Iub Throughput is used to monitor the throughput between the NodeB and the RNC. DL throughput should be dimensioned to be more than UL.

7.4.20. Iub FP Layer data Volume

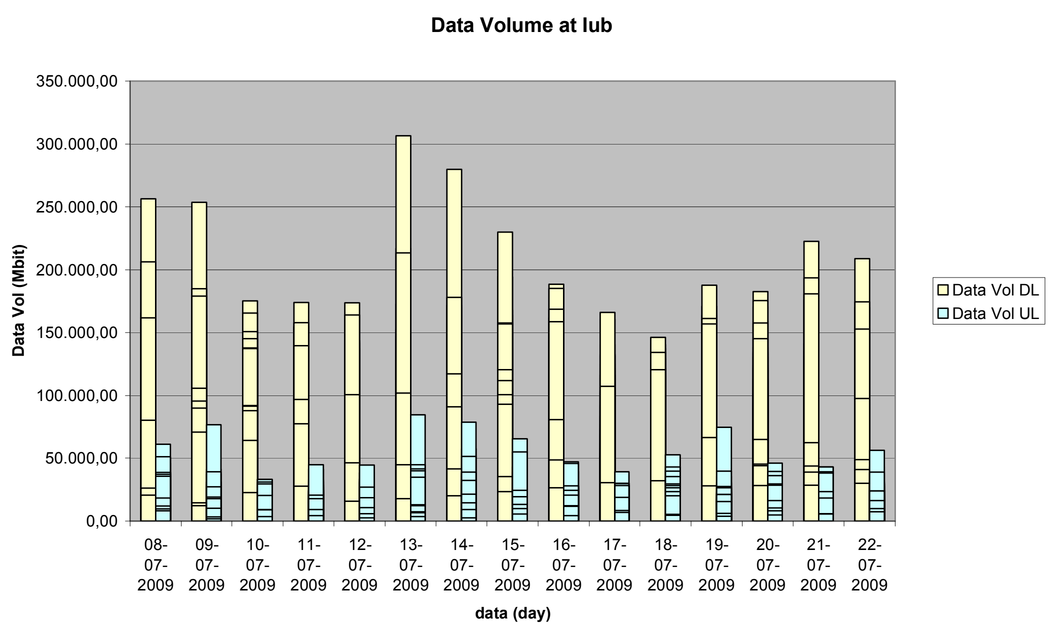


Figure 89 - Iub FP Layer data Volume

At this point, it is observed the total transmitted data between NodeB and RNC in both directions.

7.4.21. Iub congestion rate

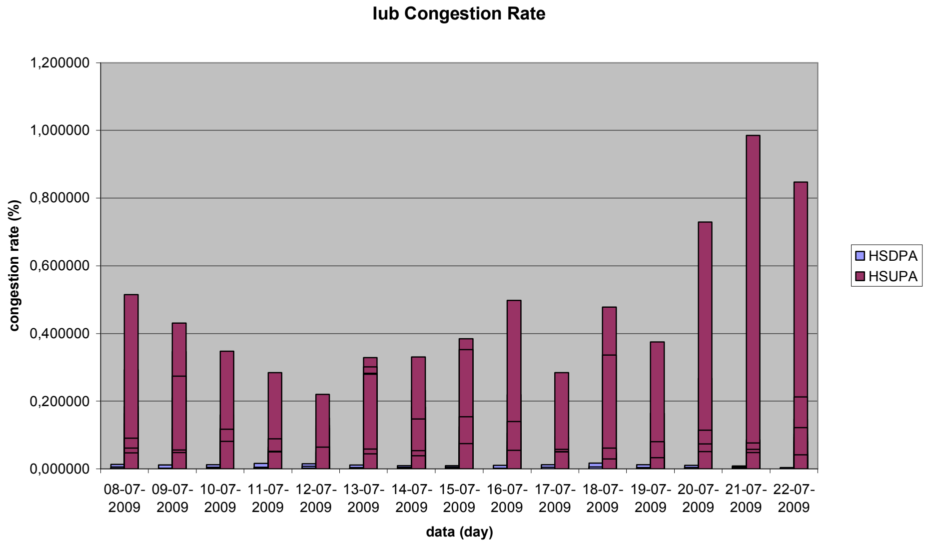


Figure 90 - Iub congestion rate

Iub congestion rate shows how traffic is being handled between NodeB and RNC. In this case, HSUPA is more congested than HSDPA, probably because HSUPA as limited resources from the network. Operator should give more resources to HSUPA traffic in this interface.

8. Conclusions

Monitoring network KPIs and End User KPIs is extremely important for OPEX reduction purposes and problem mitigation, since operators can predict issues and avoid unnecessary costs. To help operators to identify the different requirements of the different services, it is shown a table with the most important metrics for each service and defined the most important QoS KPI's per service application. Since QoE is certainly dependent of QoS is also defined a relation table between some QoE goals and QoS KPI's from network side.

With the architecture proposed in chapter 6 operators can monitor end user behaviour and network at the same time which gives them the possibility of classify a specific region/user in terms of profiles. This allows operators to configure network parameters in order to fulfil different needs of several users and predict how they would react by estimating the acceptance degree of a certain service in a given region.

The main goal of this work was the development of mechanisms that that allow telecommunication operators to monitor end user behaviour and network performance metrics. This objective is fulfilled and results are illustrated in chapter 7. To improve quality of the work an effort has been done in order to collect metrics from real networks around the world.

The work proves that is possible to collect, for the several services available in a mobile network, an amount of metrics that allows operators to understand services quality (access times, delays, bit rate, etc...), user expectations or satisfaction (MOS method), users behaviour (more used services, date and time of usage, combination of services in use, etc…). Network detailed performance metrics such as interfaces capacity, throughputs (for downlink and uplink), cells availability, protocols error ratio and many others. From the users behaviour metrics collected is possible to conclude that the most used services over a 3G network are data calls, browsing, SMS and voice calls. These services are with good quality in general and user recognises it. Network detailed status is gathered, and for this case, the overall performance of the network is excellent as it was expected because the study was performed in a mature operator with a stabilized network.

In future work, the impact of several users for different services in the network performance and vice-versa will be measured. It will also be understood for the different services how end users react to service degradation and will be performed the definition of breakpoints correlating network metrics and user reported quality for services performance.

9. References

[1] Harri Holma and Antti Toskala, "WCDMA for UMTS HSPA Evolution and LTE", John Wiley & Sons, 4st edition, 2007

[2] Kaaranen, Heikki, et. Al., "UMTS Networks Architecture Mobility and Services", John Wiley & Sons, 2nd Edition, 2005

[3] Kreher, Ralf, "UMTS Performance Measurement - A Practical Guide to KPIs for the UTRAN Environment", John Wiley & Sons, 2006

[4] A. Clark, et. Al., "RTCP HR - High Resolution VoIP Metrics Report", ietf draft, February 2008

[5] Hyun-Jong Kim, Dong-Hyeon Lee, Jong-Min Lee, Kyoung-Hee Lee,Won Lyu, Seong-Gon Choi, et Al., "The QoE Evaluation Method through the QoS-QoE Correlation Model", NCM 09, September 2008

[6] ITU-T P.800, "Mean Opinion Score(MOS) terminology," Mars 2003.

[7] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ)," February 2001.

[8] M. Almeida, R. Inácio, S. Sargento, "Cross layer design approach for performance evaluation of multimedia contents", IWCLD, Maiorca, June 2009

[9] ITU-T G.107, "The E-model, a computational model for use in transmission planning," Mar 2005.G. Rubino, "The PSQA project," INRIA Rennes-Bretagne Atlantique

[10] Nokia Siemens Networks, Managing User experience "In-Touch Study", September 2008

[11] 3GPP TS 25.303 and 3GPP TS 25.331

[12] Jaana Laiho, Achim Wacker, Tomas Novosad, "Radio Network Planning and Optimisation for UMTS", John Wiley & Sons, 2nd Edition, 2006

[13] Maciej J. Nawrocki, Mischa Dohler, A. Hamid Aghvami, "Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation ", Wiley, 2006

[14] Alex Brand, Hamid Aghvami, "Multiple Access Protocols for Mobile Communications GPRS, UMTS and Beyond", John Wiley & Sons, 2002

[15] 3GPP TS 25 series