DUBLIN CITY UNIVERSITY

SCHOOL OF ELECTRONIC ENGINEERING

Resource Allocation for Multimedia Messaging Services over EGPRS

Edward Nowicki

September 2003

MASTER OF ENGINEERING

IN

ELECTRONIC ENGINEERING

Supervised by Dr. J. Murphy

Acknowledgements

I would like to thank my supervisor Dr John Murphy for his guidance, enthusiasm and commitment to this project. Thanks are also due to Dr Philip Perry for supporting this work to date. Finally I would like to express my deep appreciation to the whole Performance Laboratory for the nice atmosphere and help during my research work at Dublin City University.

Declaration

I hereby declare that, except where otherwise indicated, this document is entirely my own work and has not been submitted in whole or in part to any other university.

Signed: [illegible] Date: 17.08.2003

Abstract

The General Packet Radio Service (GPRS) is a new bearer service for GSM that greatly simplifies wireless access to packet data networks, such as the Internet, corporate LANs or to mobile portals. It applies a packet radio standard to transfer user data packets in well-organized way between Mobile Stations (MS) and external packet data networks.

The Enhanced General Packet Radio Service (EGPRS) is an extension of GPRS, offering much greater capacity. These enhancements have allowed the introduction of new services like Multimedia Messaging Services (MMS). MMS enables messaging with full content versatility, including images, audio, video, data and text, from terminal to terminal or from terminal to e-mail.

The Wireless Application Protocol (WAP) is the WAP Forum standard for the presentation and delivery of wireless information and telephony services on mobile phones and other wireless terminals.

In this thesis it is indicated that efficient radio resource allocation is necessary for managing different types of traffic in order to maintain the quality demands for different types of services. A theoretical model of MMS and WAP traffic is developed, and based on this model a simulator is implemented in Java programming language.

This thesis proposes two techniques to improve the radio resource allocation algorithm performance called "radio link condition diversification" and "interactive traffic class prioritization". The radio link condition diversification technique defines minimum radio link quality that allows the user to receive their packets. The interactive traffic class prioritization technique defines different priorities for WAP packets and for MMS packets. Both techniques give good results in increasing user's perception of services and increasing network efficiency.

This thesis indicates also that the prioritization mechanism successfully improves the response time of the interactive service by up to 80% with a setting of priority for interactive traffic class and decreasing the performance of the background traffic. This decrease is within a range acceptable by the end-user and that the link conditions limit mechanism has an advantage in terms of resource utilization.

Table of Contents

List of Figures

List of Tables

Chapter 1

1. Introduction

1.1 Introduction to the Thesis

Mobile and personal communication is recognised as major driving force for European industrial competitiveness and for sustained economic growth. On a personal level, telecommunications change the lifestyles of most people making them more mobile. For instance, as an amount of public or private transportation to work and leisure activities is increasing, people spend more and more time in transit. Time spend commuting is not work time or free time, but "mobile time". This has been recognised by developers of electronic technology which have made mobile electronics part of the mass consumer market.

Since the introducing of the Global System for Mobile communications (GSM) in 1992, mobile phones has become devices owned by more than half of the population in many European countries. The market is expecting not only voice services but also more and more data services like Internet and multimedia. Principally, the intention is to bring together the GSM solution for wireless access with the Internet solution for information sharing, forming a solution for personal information access. These new solutions as well as demand for mass-market mobile data services opening up for new business opportunities for operators and content providers.

The General Packet Radio Service (GPRS) is a new bearer service for GSM that greatly simplifies wireless access to packet data networks, such as the Internet, corporate LANs or to mobile portals. It applies a packet radio standard to transfer user data packets in well-organized way between Mobile Stations (MS) and external packet data networks.

The Enhanced General Packet Radio Service (EGPRS) is an extension of GPRS, offering much greater capacity. These enhancements have allowed the introduction of new services like Multimedia Messaging Services (MMS). MMS enables messaging with full content versatility, including images, audio, video, data and text, from terminal to terminal or from terminal to e-mail.

The Wireless Application Protocol (WAP) is the WAP Forum standard for the presentation and delivery of wireless information and telephony services on mobile phones and other wireless terminals.

WAP and MMS have different traffic characteristics. The MMS is a typical background messaging service that requires no user interaction. Conversely, WAP is an interactive service, which is delay sensitive and introduces very bursty traffic.

1.2 Purpose of the Thesis

This thesis addresses problems appearing in the introduction of new multimedia services such a Multimedia Messaging in the EGPRS radio communication system. The focus is on downlink transmission scheduling over wireless links. Flexibility of the radio resource allocation is the most important means for achieving high bandwidth efficiency in the wireless link, and also satisfying the required communication quality for the served applications.

The essence of this statement is that our communication system must be able to assign the scarce wireless resources where they are needed, and where they are expected to provide the most benefit. It is important not to waste the bandwidth on things such as repetitions of already transmitted information. The statement also implies that the applications should be able to adjust their requirements such a service priority on the requested services from the communication system. This thesis suggests an approach where traffic class prioritisation and link condition diversification is used, obtaining a bandwidth efficient predictive resource allocation.

1.3 Related Work

The allocation of radio resources in GPRS and EGPRS is an active research area and there are many papers that address this problem from different perspectives. Researchers have been studying the problem of radio resource allocation for many years.

In [22], [7], [16], [8], [11], [10], [24] and [25] the authors use simulations to investigate the algorithms used for resource allocation in GPRS or EGPRS networks.

In [22] Pang et al presents an evaluation of several traffic scheduling methods, including FIFO, static priority scheduling and earliest deadline first by simulations, with the objective of meeting the delay profiles defined for a number of GPRS classes. Traffic sources representative of GPRS applications, including e-mail, fleet management and World Wide Web applications are employed. This work is focussed on the downlink link, which represents the bottleneck of a typical GPRS data connection.

In [7] Ajib et al propose several service disciplines to provide best possible performance for best-effort WWW traffic, describe their mechanisms, similarities and differences. These policies performance is studied in terms of throughput, delay and loss rate by simulating the WWW traffic application on uplink direction in GPRS system. In [16] authors investigate the weighted fair queuing and priority queuing algorithms for resource allocation, including the influence that TCP/IP has on the users' performance in EDGE system.

In [8], the authors investigate how the adaptation capability of "elastic" multimedia applications diminishes the differences between user perceived performance in the cases with/without reservation or prioritisation. Bajaj et al in [8] incorporates two new elements, into the analysis: the utility function of the applications (how application performance depends on network service), and the adaptive nature of applications (how applications react, to changing network service).

In [10] the authors explore different scheduling algorithms at the radio access network, where the radio link quality dominates the overall performance, for supporting differentiated service offerings.

In [11] Chang et al. explore radio resource mechanisms based on allocating more radio resources, e.g. multiple time slots, to premium users and to see if these are effective in supporting differentiated services in the context of the EGPRS system.

In [24], [25] and [26] Todinca et al. investigate the efficiency of different algorithms used for resource allocation in data transfer over EGPRS networks. A number of algorithms are presented for the resource allocation, and some generic mathematical results are presented. . The simulation presented in [24], [25] and [26] is focussed primarily on the resource allocation algorithms, rather than system details like the interaction with TCP/IP or network signalling.

In [15] Graja et al. investigate the influence of different probability distributions that can be applied in user data source models and compare their effect on the user perceived performance of a simulated GPRS system. In this paper Graja et al. investigate four different distributions of the time between two consecutive web page requests and consider non-probabilistic models of the worst-case and best-case data source models.

This thesis proposes a new approach in radio resource allocation using *radio link quality diversification* and *interactive traffic class priority mechanism*. The similar approach where presented in [10], but in [10] Chang et al consider QoS class in terms of service offering, such as premium and basic, rather than traffic class. Chang et al. considers two classes of background data service: premium and basic with the same traffic characteristics. This thesis considers two traffic classes with different traffic characteristics.

In [10] Chang et al propose a simple yet intelligent scheduling technique that takes into account the radio link quality and investigate the resulting performance in turns of system capacity, delay performance over the ordinary scheduling technique.

The radio link quality mechanism investigated in this thesis for background messaging services allows the operator to improve the overall system utilisation and the user perceiving of interactive services such a WAP browsing.

1.4 Structure of the Thesis

The second chapter of the Master's Thesis presents activities in the resource allocation research area and papers that address similar problem from different perspectives.

The third chapter of the Master's Thesis explains the technical background of the EGPRS system, Multimedia Messaging and radio resource allocation principles.

The fourth chapter explains the structure of the EGPRS model and simulation methodology, shows the input parameters and simulation results and finally in the fifth chapter conclusions are drawn. Last you also find appendices with abbreviations and a glossary which can be convenient to use during the reading of this thesis.

Chapter 2

2. Technical Background

2.1 Evolution of cellular networks

There are three different generations as far as mobile communication is concerned. The first generation, 1G, is the name for the analogue or semi-analogue (analogue radio path, but digital switched) mobile networks established after the mid- 1980s, such as NMT 450 and NMT 900 (Nordic Mobile Telephone system), TACS (Total Access Communication System) in United Kingdom and AMPS (Advanced Mobile Phone Service) in the United States and Canada. These networks offered basic services for the users, and the emphasis was on speech.1G networks were mainly national efforts and very often they were standardised after the networks were established. Due to this, the 1G networks were incompatible to each other. All these systems are based on analogue speech transmission using analogue modulation and Frequency Division Multiple Access (FDMA).

As the need for mobile communication increase, also the need for a more global mobile communication system increased. The international specification bodies started to specify what the second generation, 2G, mobile communication system should look like. The emphasis on 2G is/was on compatibility and international transparency; the system should be a global one and the users of the system should be able to access it basically anywhere the service exists. Due to some political reasons, the concept of globalisation did not succeed completely and there were some 2G systems available on the market like GSM (Global System for Mobile Communication) in Europe, D-AMPS (Digital-Advanced Mobile Phone Service) in the America region and PDC (Personal Digital Cellular) in Japan. Out of these, the commercial success story is GSM and its adaptations. The GSM services were launched in 1991 and employs digital transmission and Time Division Multiple Access (TDMA).

The third generation, 3G, is expected to complete the globalisation process of the mobile communication. Again there are national interests involved. Also some difficulties can be foreseen. The 3G will mostly be based on GSM technical solution due to reasons that , the GSM as technology dominates the market, and that great investments made to GSM should be utilised as much as possible. Based on this, the specification bodies created a vision about how mobile telecommunication will develop within the next decade. The main requirements for the 3G system are that it shall use more efficient available spectrum, provide higher bit rates to mobile devices and access to same services as fixed network delivers. Multimedia and all of its components must be supported through the system. The radio access to the 3G must be generic and the services for the end users must be independent i.e. radio access and the network infrastructure must not limit the services to be generated.

Because there are some political desires involved, the issue is not as simple as described. Global system means global business and this is why there has been a lot of pressure to select or emphasis certain solution more than others. This political debate actually delayed the specification work remarkably, and finally an organisation was established to take care of the harmonisation issues. This organisation, OHG (Operator Harmonisation Group) aims to find a common understanding concerning the global issues. The results of this organisation are used as inputs in 3rd Generation Partnership Project (3GPP) work, and also in 3G future implementation. The 3GPP is a separate specification body, which takes care of the specification work in co-operation with four international standardisation bodies acting as "generators" for 3G specification work:

- ITU-T (International telecommunication union). This organisation has finished its development process for IMT2000, International Mobile telephone – 2000 and the specification work is transferred to the 3GPP.
- ETSI (European Telecommunication Standard Institute). This organisation is in dominant role in the 3G-specification work. In Europe the ETSI-driven 3G variant is called UMTS (Universal Mobile Telecommunications System)
- ARIB (Alliance of Radio Industries and Business). This organisation provides commercially oriented contribution for the specification process from the Australia-Asia area.
- ANSI (American National Standard Institute). This organisation is the American specification body that is mainly concentrating on a competing 3G Air interface technology selection called CDMA2000. ANSI strongly motivated 3GPP to work on CDMA2000 standards.

The aim of the OHG work is to affect the specification so that all radio access variants are compatible with all the variants meant for switching, which will ensure true globalisation for 3G. All the above-presented variants have "marketing names" used globally, which makes 3G a bit confusing.

- IMT-2000 is the ITU-T name for the 3rd generation cellular system. The Japanese view of 3G is based on the IMT-2000. The switching part of this variant is quite an open issue, but is expected to be based on existing GSM technology. The radio access is almost similar to the European variant.
- CDMA2000 is the name for American 3G variant. In the American variant, the TDMA scenario, which is partially based on the EDGE (Enhanced Data rates for GSM Evolution) is also considered as an alternative.

- UMTS (Universal Mobile Telecommunication System) is the name for the European, ETSI-driven 3G variant. It emphasis the interoperability and backward compatibility between the 3G implementation and GSM.

The evolution towards UMTS takes place in several steps: (as shown in the figure 1)

- GPRS (General Packet Radio Service)
- EDGE (Enhanced Data rates for GSM Evolution)
- UMTS (Universal Mobile Telecommunications System)

GPRS and EGPRS are an intermediate stage that is added to the evolution. This generation is labelled 2+ (or even 2.5G) this transition-gained momentum during the year 2000 and beyond.

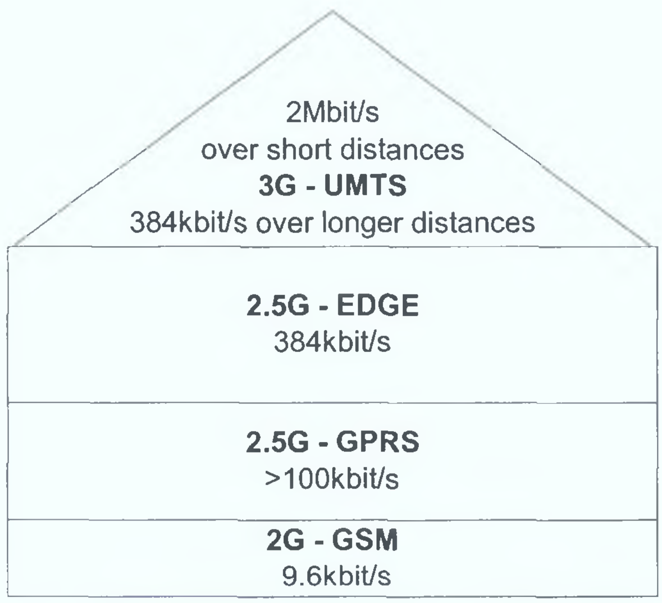


FIGURE 1: EVOLUTION FROM GSM TOWARDS UMTS

The first step in the evolution process towards UMTS is GPRS, which is able to offer a data rate of over 100kbit/s. GPRS uses a packet-based technique, which will enhance GSM data services significantly, especially for bursty Internet traffic. New GPRS network elements are required to be installed parallel with the current GSM core network elements to provide direct IP (Internet Protocol) access from the mobile station to data networks such as the Internet.

Following GPRS will be EDGE (Enhanced Data rates for GSM Evolution), which will increase the data rate to maximum of 384kbit/s. EDGE (also refereed as EGPRS – Enhanced General Packet Radio Service) achieves this by using a more efficient modulation schemes across the air interface. This technology is backward compatible with the existing GSM standards. This step will probably be the end point for several operators due to the licensing policy (country-specific regulations). On the other hand, some operators may skip this phase and move on to the next step in this development path. The EDGE utilises everything built in the GSM, including the multiple access method used in the air interface (TDMA).

UMTS will use a new air interface employing WCDMA (Wideband Code Division Multiple Access) which is capable of carrying 384kbit/s over relatively long distances. UMTS comes into its own over short distances where it is capable of transmitting and receiving 2Mbit/s. UMTS will provide wideband wireless multimedia capabilities over mobile communication networks.

2.2 General Packet Radio Service

The General Packet Radio Service (GPRS) is an overlay to the circuit switched GSM-network. GPRS increases the possible bandwidth for data transmission by introducing packet switching. The radio interface of GPRS system is defined in [12] and an overview is given in [9].

Table 1 presents protocol sub-layers in GPRS/GSM networks. The highest sub-layer is the SubNetwork Dependent Convergence Protocol (SNDCP), which maps network level characteristic onto one of the underlying network. Under the SNDCP is the Logical Link Control (LLC) sub-layer, which provides a logical link between two entities. Under LLC sub-layer is the Radio Link Control (RLC) sub-layer, which performs the segmentation and reassemble of LLC Protocol Data Unit (PDU) into RLC data blocks. Under the RLC sub-layer is the Medium Access Control (MAC) sub-layer, which enables Mobile Stations (MS) to share one or several Physical Channels (PhCH). Under the MAC sub-layer is the Physical layer that provides services for information transfer over physical channels between Mobile Stations and the GSM/GPRS network, as is shown in Figure 2.

SubNetwork Dependent Convergence Protocol (SNDCP)	Network Layer
Logical Link Control (LLC)	Data Link Layer
Radio Link Control (RLC)	
Medium Access Control (MAC)	
Physical Layer PhCH	Physical Layer

FIGURE 2: LAYERS IN GPRS/GSM NETWORKS

On the physical layer, GSM uses a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). Two frequency bands 45MHz apart have been reserved for GSM/GPRS, namely 890 to 915 MHz for the uplink and 935 to 960 MHz for the downlink. Each of these 25 MHz bandwidths is divided into 124 single carrier channels of 200 kHz. A certain number of these frequency channels are allocated to a Base Transceiver Station (BTS) i.e. to a cell. Each frequency channels carries eight TDMA channels, called time slots and these slots form a TDMA frame. Each time slot of TDMA frame has duration of 576.9 μs, so a TDMA frame has duration of 4.613 ms and every time slot defines a physical channel. In conventional GSM, a channel is permanently allocated for a particular user during the entire call period (whether or not data is transmitted). In GPRS channels are only allocated when data packets are sent or received, and they are released after the transmission.

2.3 Enhanced General Packet Radio Service

The objective of EDGE/EGPRS (Enhanced Data rates for GSM Evolution/Enhanced GPRS) is to increase user bit rates without making major changes to previously installed GSM networks. Just as for GSM/GPRS, eight timeslots and 200KHz carrier spacing is used. The aim is to obtain maximum data rates up to about 384kbps (perceived by a single user on eight timeslots [14].

EDGE uses 8PSK (Phase Shift Keying) modulation in addition to GMSK. This means that three bits are sent for every symbol instead of only one single bit. Therefore, three times higher bit rates are theoretically possible. EDGE supports transmission rates ranging from 8.8 kbps to 59.2 kbps per channel. The 8PSK modulation is much more sensitive to noise and interference. Therefore, GMSK modulation can still be used when a more robust modulation is needed. One idea is to use 8PSK/GMSK for downlink communication, where the demand for high bit rates is larger, and use only GMSK for uplink communication, which results in less complex transmitters in the terminals.

To optimise the throughput for all radio conditions Link Quality Control (LQC) is used to dynamically adapt the modulation and code rate with respect to the current link quality. EGPRS has nine different Modulation and Coding Schemes, MCS-1 to MCS-9 as show in the Table 1 [1]. The first four use traditional GMSK and the last five use the new 8PSK modulation scheme. The MCS's also have different code rates and are therefore suitable for different radio conditions. If the conditions are very good, no coding is necessary and high bit rates can be achieved. Bad conditions require more robust coding, which results in less payload and lower bit rate.

Scheme	Modulation	Raw Data within one Radio Block (octets)	Data rate kb/s
MCS-9	8PSK	148	59.2
MCS-8		136	54.4
MCS-7		112	44.8
MCS-6		74	29.6
MCS-5		56	22.4
MCS-4	GMSK	44	17.6
MCS-3		37	14.8
MCS-2		28	11.2
MCS-1		22	8.8

TABLE 1: *CODING PARAMETERS FOR THE EGPRS CODING SCHEMES*

Figure 3 illustrates the EGPRS architecture. The EGPRS network consists of two parts: a Radio Access Network portion (the GSM/EDGE Radio Access Network (GERAN)) and a Core Network (CN) portion. The GERAN consists of base transceiver stations and radio network controllers. The Radio Network Controller (RNC) is responsible for the radio resource management, including handover decisions, for all base transceiver stations that it serves. The Core Network consists of two types of entities, Enhanced Serving GPRS Supporting Node (ESGSN) and Enhanced Gateway GPRS Supporting Node (EGGSN). RNCs are connected to the serving ESGSN, which serves as an access point to the EGPRS core network. The EGGSN serves as the gateway router connecting the EGPRS core network to external packet data networks.

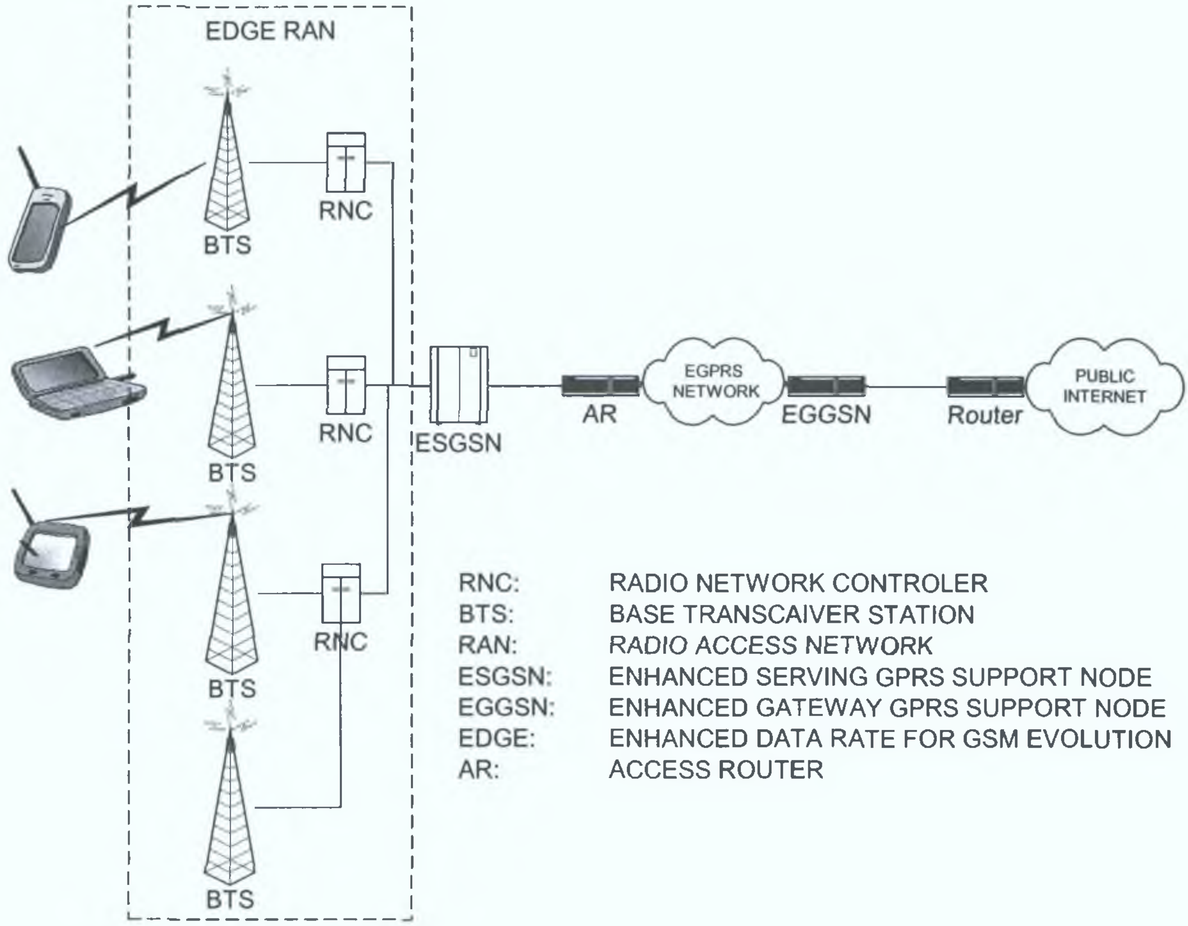


FIGURE 3: THE ARCHITECTURE OF EGPRS

Any data transfer in GPRS or EGPRS requires an active Packet Data Protocol (PDP) context. A PDP context, for example an IP context, can be initiated by the MS or the EGGSN using a "PDP context create" signalling message. This set up a connection using the packet data protocol specified between an external packet data network and the MS with specified QoS requirements. The contents of the signalling message include the desired QoS profile for the session. Thus, the PDP context creation process not only enables the MS to communicate with the external data network but also provides QoS management in the EGPRS core network.

2.4 Resource allocation principles

A cell supporting GPRS or EGPRS may allocate physical channels for GPRS traffic. Such a physical channel is denoted as Packed Data Channel (PDCH). The PDCH are taken from the common pool of all channels available in the cell. The allocation of physical channels to Circuit Switched (CS) services in GPRS is done dynamically according to the "capacity on demand" principle. It means that GPRS does not require permanently allocated PhCHs. When the PDCHs are congested but more resources are available in the same cell, then the network can allocate more channels for PDCHs. Dynamic allocation of PDCHs can be done by the allocation of unused channels as PDCHs to increase the overall GPRS QoS and by the de-allocation of PDCHs upon resource demand for other services with higher priority. Physical channels available in the cells could be shared between different traffic types, either in a fixed boundary or in a movable boundary way as shown in Figure 4 and Figure 5 [23].

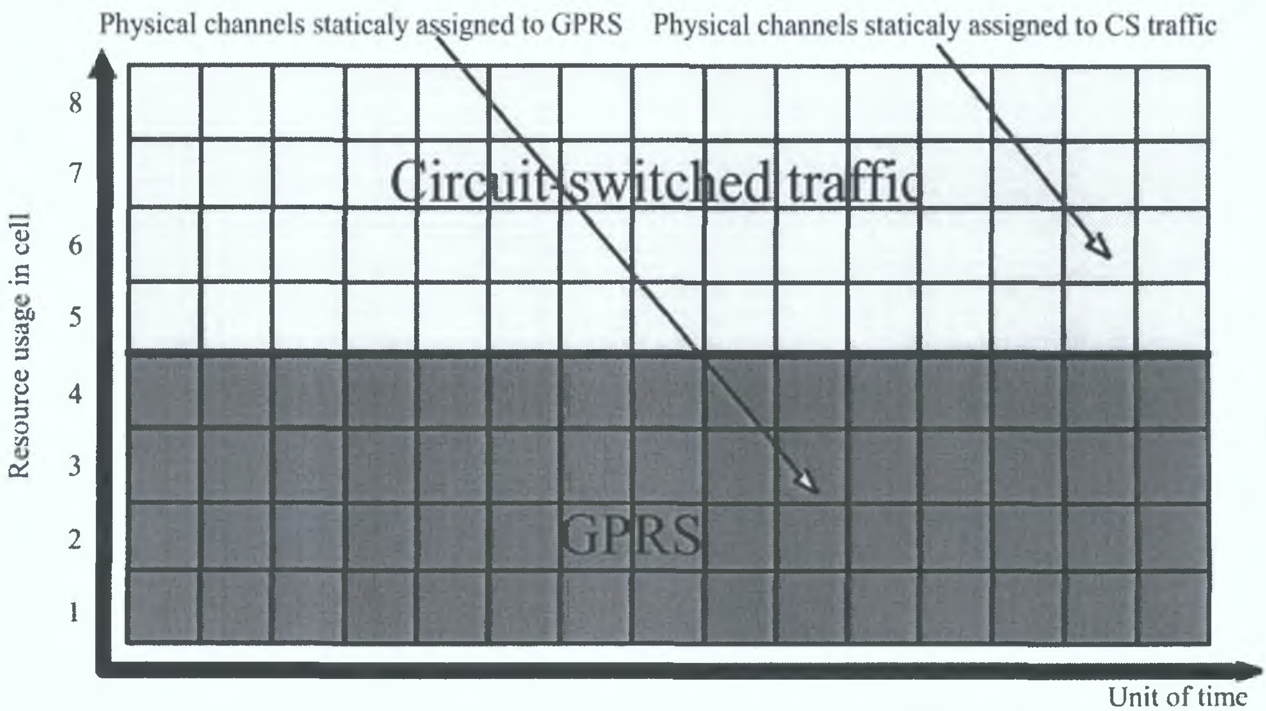


FIGURE 4: FIXED BOUNDARY ALLOCATION POLICY.

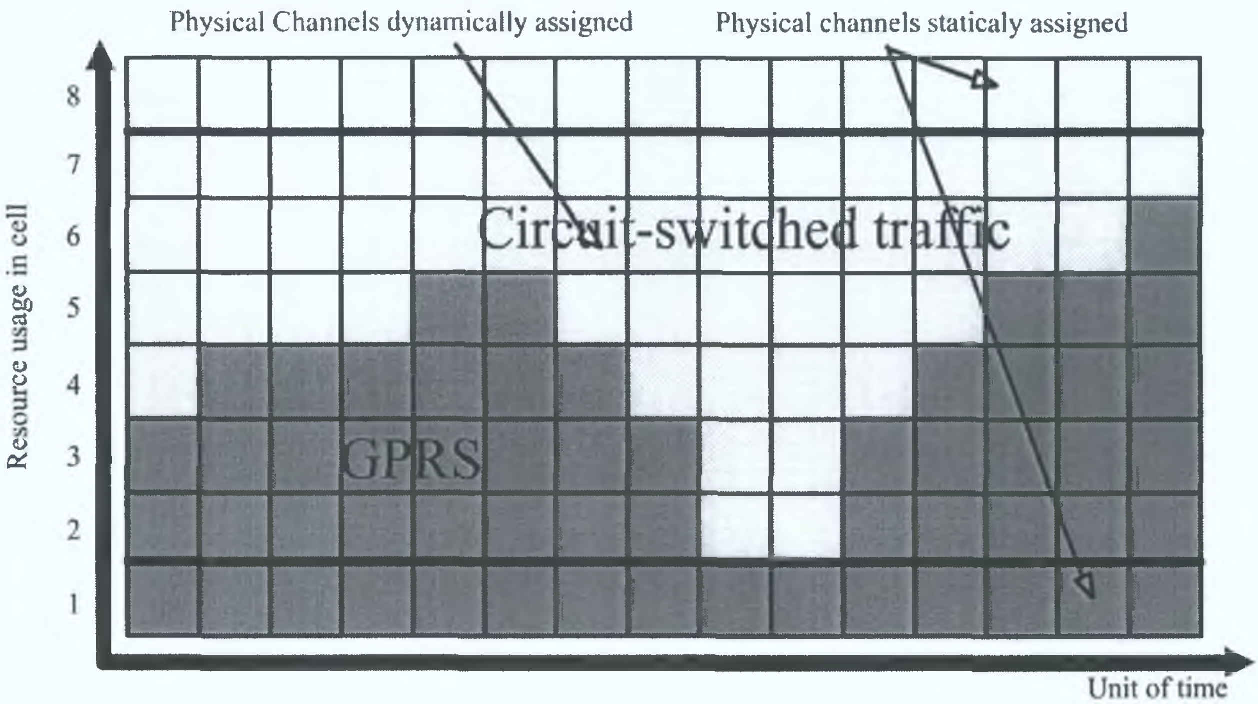


FIGURE 5: DOUBLE MOVABLE BOUNDARY ALLOCATION POLICY

The fixed boundary policy is not efficient, because resources are not fully utilised. The one type of traffic wastes the resources, while the other one is suffering from congestion. The movable boundary policy overcomes this disadvantage by allowing a limited sharing of resources. The double movable boundary policy considers that the total of resources is separated into three parts:

1. The first part is reserved for the Circuit Switched traffic
2. The second part is reserved for Packet Switched traffic
3. The third part contains a number of channels that are dynamically assigned, either to Circuit Switched traffic or to the Packet Switched traffic.

The number of fixed Circuit Switched traffic channels depends on the input load of Circuit Switched calls in the cell while the QoS experienced by the packet switched services is dependent on the traffic profile. [23].

In this thesis the focus is on Physical Channels assignment to GPRS Mobile Stations, which are shared between Circuit Switched and Packed Switched transmissions. The basic radio packet in GPRS is the RLC/MAC block. It uses a sequence of four time slots on a PDCH – see Figure 4. A Temporary Block Flow (TBF) is a physical connection used to support the transfer of a number of blocks and identified by Temporary Flow Identifier (TFI). The TFI is included in every transmitted block, so that multiplexing of blocks originated from different Mobile Stations on the same PDCH is possible.

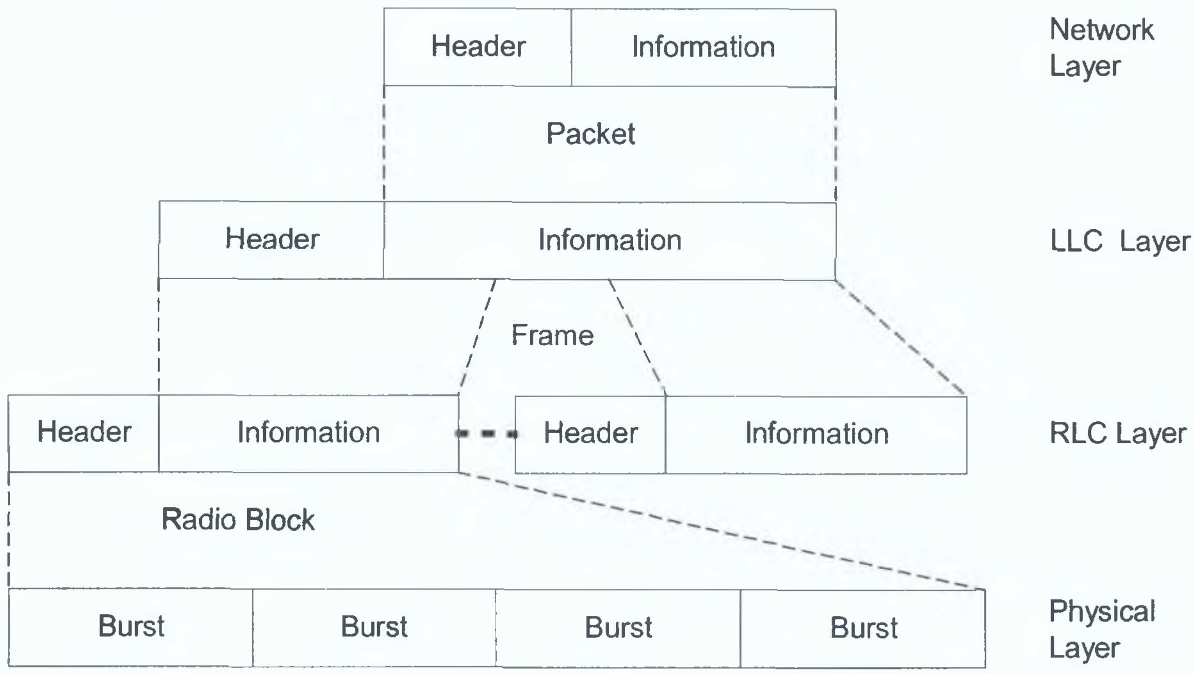


FIGURE 6 PACKET TRANSFORMATION IN GPRS

2.5 Quality of Service

2.5.1 QoS in the Internet

Transmission of IP-packets in the Internet has traditionally been driven by a "Best Effort" mentality, meaning that the links only provide a single level of service. In this case, there are no guarantees that the packets won't be delayed or that they are delivered at all. Users are increasingly demanding a differentiation of services to meet new standards of applications.

The level of service quality can be measured by four main components:

- Delay is the elapsed time for a packed to be passed from the sender, through the entire network to the receiver. This is also referred to as the latency of the network. An interactive voice and video application with a total roundtrip-delay of about 400ms is considered to be unresponsive.
- Jitter is a variation of delay at the end-points of a connection. High levels of jitter causes the signals to be distorted. Jitter is counteracked by a reassembly playback-buffer, but this again introduces additional delay to the system.
- Bandwidth is the averagebit rate between two end-points. The bandwidth is limited by the physical infrastructure as well as other flows sharing the same path.
- Reliability is often thought of as the average error rate of a medium, but for an end-to-end system it also includes the performance of the switching systems, e.g. the routers. Unreliable networks causes frequent retransmission of packets, which triggers congestion avoidance algorithms.

2.5.2 Implementation of Quality of Service

The user needs to be able to request a certain level of QoS in some way. In packet switched network, such as ATM, network capabilities are established and reserved exclusively for the user. This is provided by taking advantage of time slots in the transmission (TDMA). In this way an ATM-network is guaranteed a certain bandwidth throughout the connection. ATM actually provides a Variable Bit Rate (VBR) for bursty traffic such as HTTP and a Constant Bit-Rate (CBR) for delay-sensitive-traffic. In a link-layer technology such as Ethernet that carries IP-packets (using the CSMA/CD-mechanism), many users share the high-speed medium and can send whenever they want. If two hosts transmit at the same time on an Ethernet, a collision occurs and they retransmit after a random delay. It is difficult to allocate resources in the network unless the architecture is modified, because there is no control over who accesses the medium. Wireless systems are another area where QoS needs to be implemented. This is important due to scarcity of wireless resources. For some of the new services that will be available in 2.5G and 3G there is a need to be able to guarantee a certain quality of transmit delay, bit rate and/or bit error rate [3]. Therefore 3GPP has defined the concept of Quality of Service (QoS). All services have specific needs and have therefore been sorted into four traffic classes. The four classes are:

- Conversational: Telephony speech, voice over IP and video conferencing (real time services).
- Streaming: Video/audio streaming (semi-real time services).
- Interactive: Web/WAP browsing, data base retrieval and server access.
- Background: Delivery of E-mails, MMS and downloads from databases.

The main difference between these classes is how delay sensitive the traffic is. The conversational class has the highest demands for low delay due to the human perception.

Traffic Class	**Examples**	**Time sensitive?**	**Packet loss sensitive?**
Streaming	Video/audio streaming, Video on demand	Semi-real-time services – large jitter buffer - few seconds of delay acceptable	Packet loss acceptable, but data shall be received in the same order as it was sent.
Conversational	Telephony speech, voice over IP and video conferencing	real time services – up to 300ms delay	Packet loss acceptable, but data shall be received in the same order as it was sent.
Background	Delivery of E-mails, MMS and downloads from databases.	No – large delays allowed	Yes – packet lost is not acceptable, retransmission needed
Interactive	Web/WAP browsing, data base retrieval and server access.	Semi-time sensitive services – respond in an "acceptable" time	Yes – packet lost is not acceptable, retransmission needed

TABLE 2: QUALITY OF SERVICE CLASSES

Streaming can allow a greater delay because of buffering. The interactive class expects respond in an "acceptable" time and the background class does not have any specific needs at all considering the delay. Conversational and streaming classes require that the data be received in the same order as it was sent. Interactive and Background classes typically expect error free transmission, which makes re-transmission necessary.

Applications that run over wireless networks will be affected differently by the varying conditions of the radio-interface. Clearly, conditions of high delay and variations in delay (jitter) on the radio-interface will degrade the user's experience to a greater degree with moving image transmission than with file transfers. Similarly, it could be argued that web browsing and WAP browsing is somewhat sensitive to delays, whereas e-mail and MMS is more insensitive to delays but needs strict requirements on reliability. The ultimate goal of

Quality of Service (QoS) is to effectively differentiate these different types of applications in terms of capacity and service handling.

Several levels of QoS functionality on the end-to-end connection must interact in order to make this possible. In essence, the concept of QoS must be accepted and supported by every element in the non-voice mobile value chain, i.e. the infrastructure and terminal developers, the mobile network operators, the application developers and the end-users.

2.6 Multimedia Messaging Services

Multimedia Messaging Service enables messaging with full content versatility, including images, audio, video, data and text, from MS to MS or from MS to e-mail. MMS delivers a location-independent, multimedia communication experience. Despite the full versatility of content the service is, from the user point of view, a simple, logical extension of Text Messaging (SMS) and Picture Messaging as shown in Figure 7. MMS content can include one or several of the following content types, with minimal restrictions to message size or format:

- Picture
- Data
- Text
- Audio
- Video

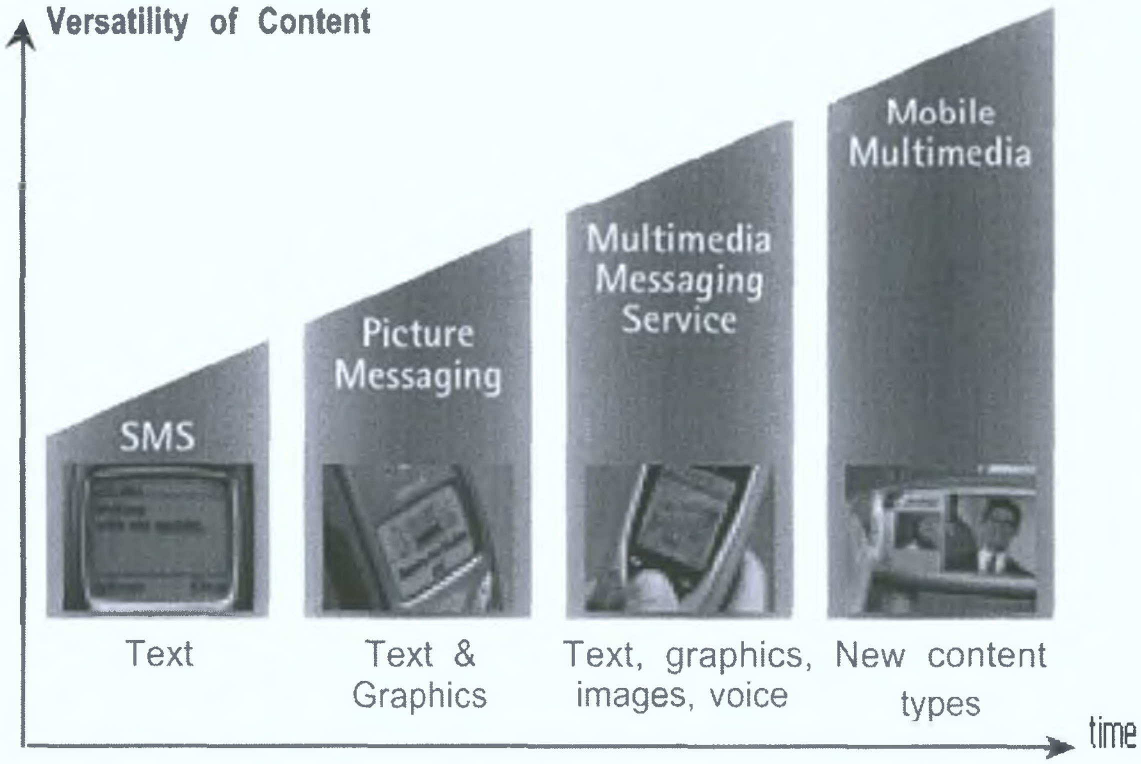


FIGURE 7: MULTIMEDIA MESSAGING MIGRATION

To create a multimedia message, a terminal with an integrated or connected camera is used for the initial digital image input. If a multimedia message is sent to a mobile phone without the MMS application, this so-called legacy terminal will receive an SMS stating that at the operator web site there is a new message along with the Internet address and personal password to fetch it. This same web site can also be used for storing favourite messages and pictures as well as for creating messages and forwarding them to the other MMS users.

From the network perspective, the Multimedia Messaging Service Centre (MMSC) is needed to perform the required store and forward operations of multimedia messages. The message flow in the GPRS network is presented below:

1. WAP POST

A multimedia message (MM) is sent from the MMS terminal to the MMS Centre using WAP POST operation.

2. MM delivery from MMS Center to recipient

When the MMS Center receives a MM from the MMS terminal it may be targeted at another MMS terminal (numbers 3 and 4 in the Figure 8), at e-mail systems, at external applications or another MMS Center.

3. WAP PUSH (over SMS) => MMS notification (over SMS)

A notification of a new MM is sent from the MMS Center to the MMS terminal using WAP PUSH operation. The bearer for WAP PUSH MMS notification is SMS.

4. WAP GET

After receiving the notification the MMS terminal sends a WAP GET request to the MMS Center. In the response an MM is delivered from the MMS Center to the MMS terminal. The downloading process is invisible to the end user and the user is notified when the whole message is downloaded and put in the MMS inbox.

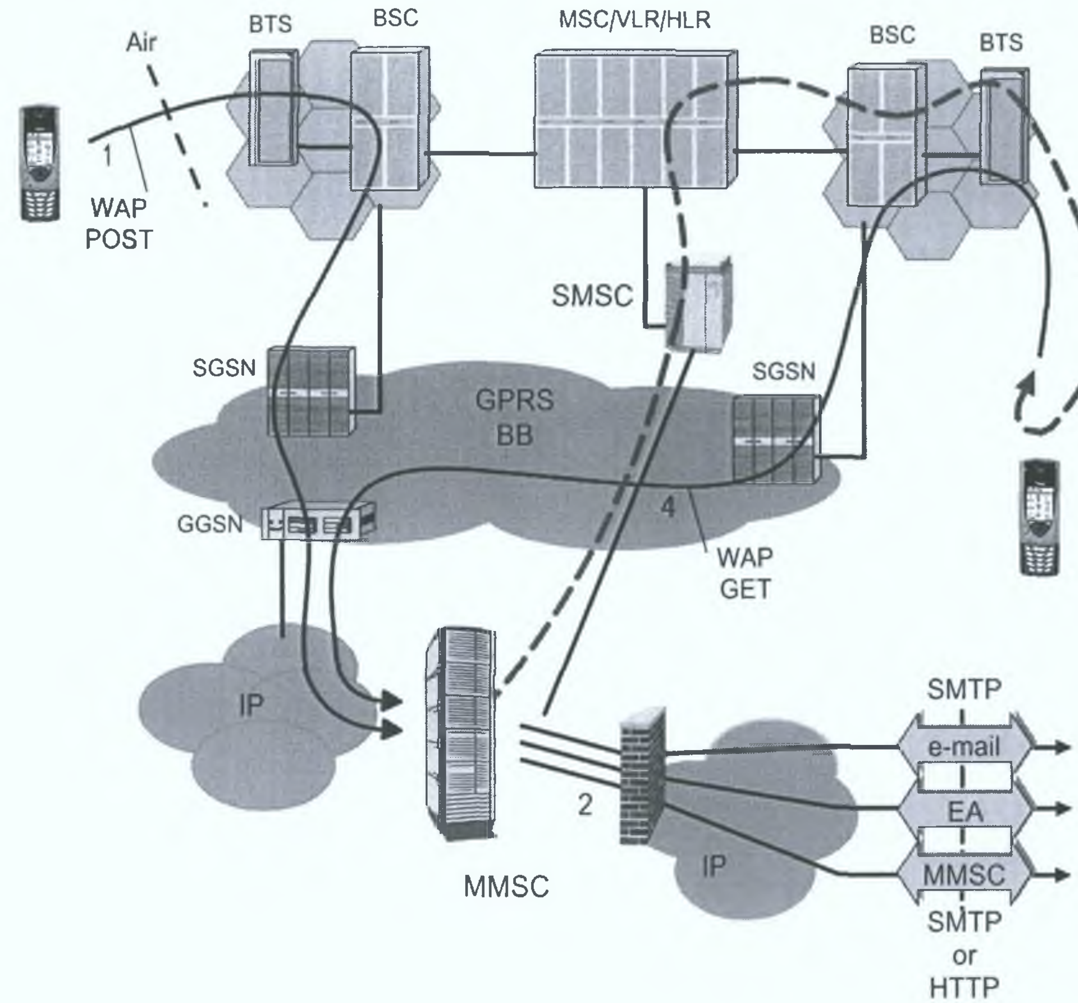


FIGURE 8: THE MMS MESSAGE FLOW IN THE GPRS NETWORK

2.7 Wireless Application Protocol

The Wireless Application Protocol (WAP) [29] is the de-facto WAP Forum standard for the presentation and delivery of wireless information and telephony services on mobile phones and other wireless terminals. WAP is positioned at the convergence of three rapidly evolving network technologies: wireless data, telephony and the Internet. The key elements of the WAP specification include:

- Definition of the WAP Programming Model as seen in Figure 9, which is based heavily on the existing WWW Programming Model.

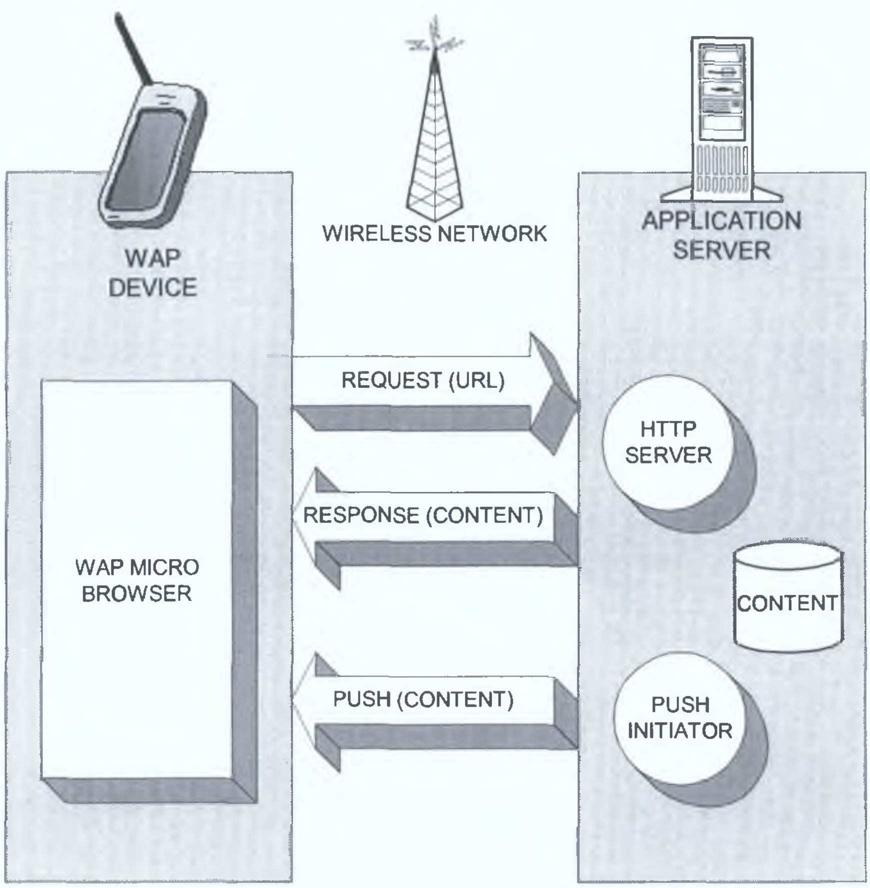


FIGURE 9: WAP PROGRAMMING MODEL

- The Wireless Mark-up Language (WML) - a mark-up language adhering to XML standards that is designed to enable powerful applications within the constraints of handheld devices. WML documents are divided into a set of well-defined units of user interactions. One unit of interaction is called a card, and services are created by letting the user navigate back and forth between cards from one or several WML documents.
- A specification for a microbrowser in the wireless terminal that controls the user interfaces and is analogous to a standard Web browser. This specification defines how the WML and WML Scripts should be interpreted in the handset and presented to the user.
- A lightweight protocol stack is used to minimise bandwidth requirements and to guarantee a variety of wireless networks can run WAP applications.

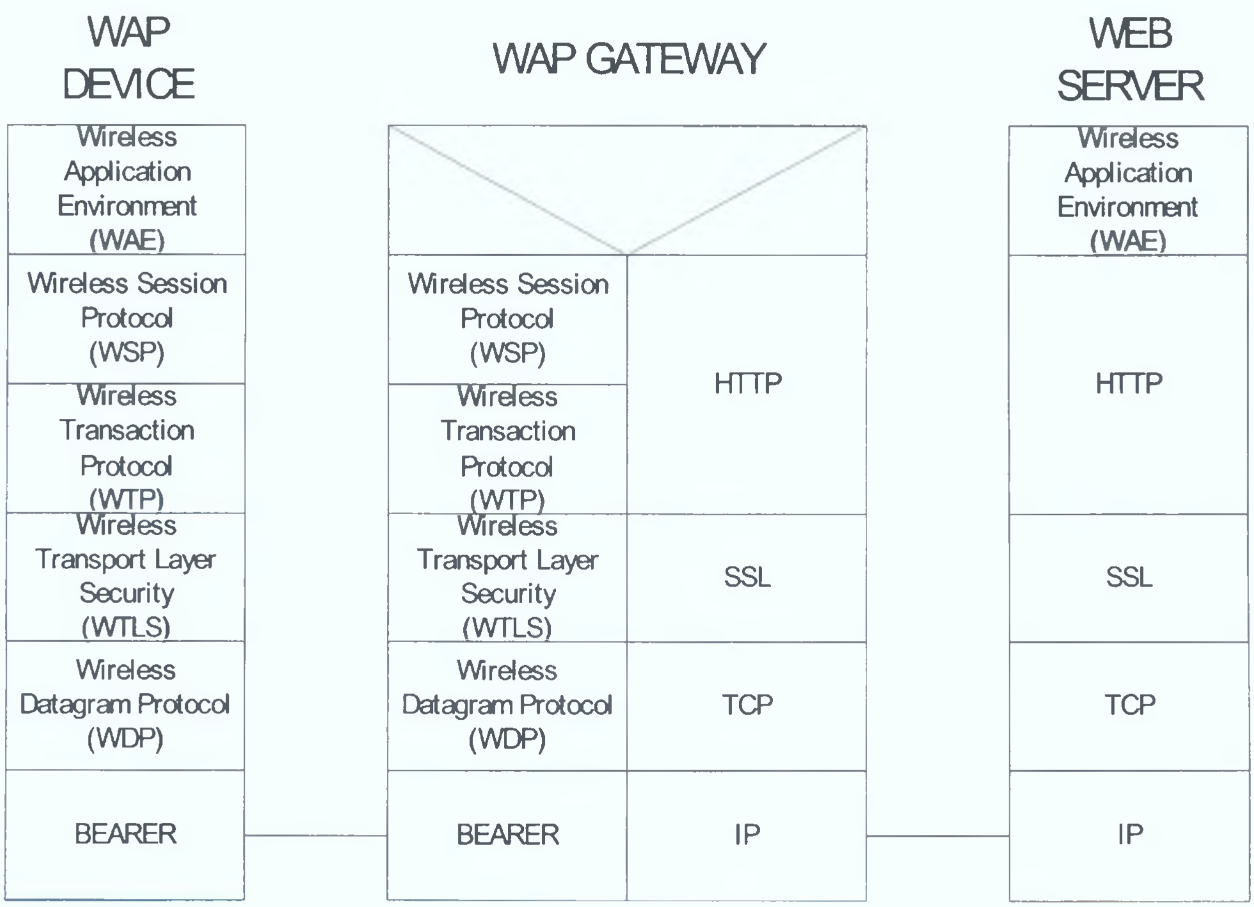


FIGURE 10: WAP PROTOCOL STACK

The WAP protocol stack is shown in Figure 10. This protocol stack has been optimised for low-bandwidth bearer networks with relatively long latency.

Wireless Session Protocol (WSP) provides HTTP/1.1 functionality and incorporates new features, such as long-lived sessions and session suspend/resume. WSP provides the upper-level application layer of WAP with a consistent interface for two session services. The first is a connection-mode service that operates above the transaction layer protocol, and the second is a connectionless service that operates above a secure or non-secure datagram transport service.

Wireless Transaction Protocol (WTP) has been defined as a light weight transaction oriented protocol that is suitable for implementation in "thin" clients (mobile stations) and operates efficiently over wireless datagram networks.

Wireless Transport Layer Security (WTLS) is designed to provide privacy, data integrity and authentication between two communicating applications. It provides the upper-level layer of WAP with a secure transport service interface that preserves the transport service interface below it.

Wireless Datagram Protocol (WDP) is a general datagram service, offering a consistent service to the upper layer protocols and communicating transparently over one of the available underlying bearer services. This consistency is provided by a set of adaptations to specific features of these bearers. This thus provides a common interface to the upper layers that are then able to function independently of the services of the wireless network.

The WAP Gateway interconnects the services offered by WSP to the HTTP protocol to permit access to data on the wired Internet.

Chapter 3

3. Simulation Models

3.1 Model produced

In this thesis the model of EGPRS system is limited to a single cell with N users and a centralised Packet Control Unit (PCU). The PCU is the part of the Base Station Sub-system (BSS) which performs the arbitration mechanism to share the radio resources between users. The users can use nine different channel coding schemes, each one allowing a certain bit rate. There are two types of users: voice users and data users. To simulate the real GSM environment, the voice users in the model are divided into two groups: high mobility users and low mobility users. The average TRX usage for high mobility users is much shorter than for low mobility users due to more frequent handover to adjacent cells. In the model it is assumed that data users are divided in to two groups: WAP data users and MMS data users. The WAP and MMS data users represent two different traffic classes: an interactive traffic class and a background traffic class. WAP data users browses WAP content – new WML page every few seconds. MMS data users receive in background MMS messages, which contains pictures of about 50KB.

The definition of data used in the thesis include all headers and error checking for WAP protocol stack i.e. Wireless Datagram Protocol, Wireless transaction Protocol, Wireless Session Protocol headers and error checking.

The user movement model assumes that all users are moving across the cell in an urban environment. In the real life not all users move in the cell. The users which are moving in the cell observes high fluctuation of interference originating from another users in the network and fluctuation of the signal level due to position in the cell and obstacles (for example buildings). The user, which is staying in one place, observes only fluctuation of interference originating from another users in the network. The signal level is quite constant (only fluctuations from interference). I used in the model two groups of users: one with strong mobility (C/I changes very quickly) and one with weak mobility (C/I changes slowly). The model does not include users, which are staying in one place and are very close to the base station or close to the edge of the cell. This affect of course presented results, which corresponds to the users, which moves (slower or faster) in the cell and observes the whole spectrum of the signal and interference level.

3.2 Simulation methodology

The simulation methodology is shown in Figure 13. The input of the simulator consists of two tables and a set of traffic parameters such as the number of carriers, the number of data users, the multislot classes, and the traffic loads of voice, WAP and MMS.

The first input table is a Block Error Ratio versus Carrier to Interference Ratio for every coding scheme from MCS-1 to MCS-9 as shown in Figure 11 [30]. This is used to calculate the error ratio for transmitted data blocks. The second table is the distribution of the link condition level during user movement in the cell as shown in Figure 12. This data is gathered from the real network environment used TEMS Investigation GSM 3.2 software and Ericsson R520m hardware. This table is used to generate user movement with specified link condition distribution.

The outputs of the simulation are the performance parameters such a delay calculated as Time to First Bit (TTFB), average throughput, bandwidth utilisation and number of retransmissions.

The radio channel transmission simulation uses multiple coding transmission from MCS-1 to MCS-8. Due to constant radio link monitoring current coding scheme is adopted to the actual radio link conditions. The lower the interference is, the higher coding scheme is used. The simulation uses multiple coding schemes switching points e.g. 2:4:6:9:11:13:20 – the switching point is the limit in dB. The value of 2 means that the MCS-1 is used from minus infinity to 2dB and the MCS-2 is used from 2dB to 4dB etc.

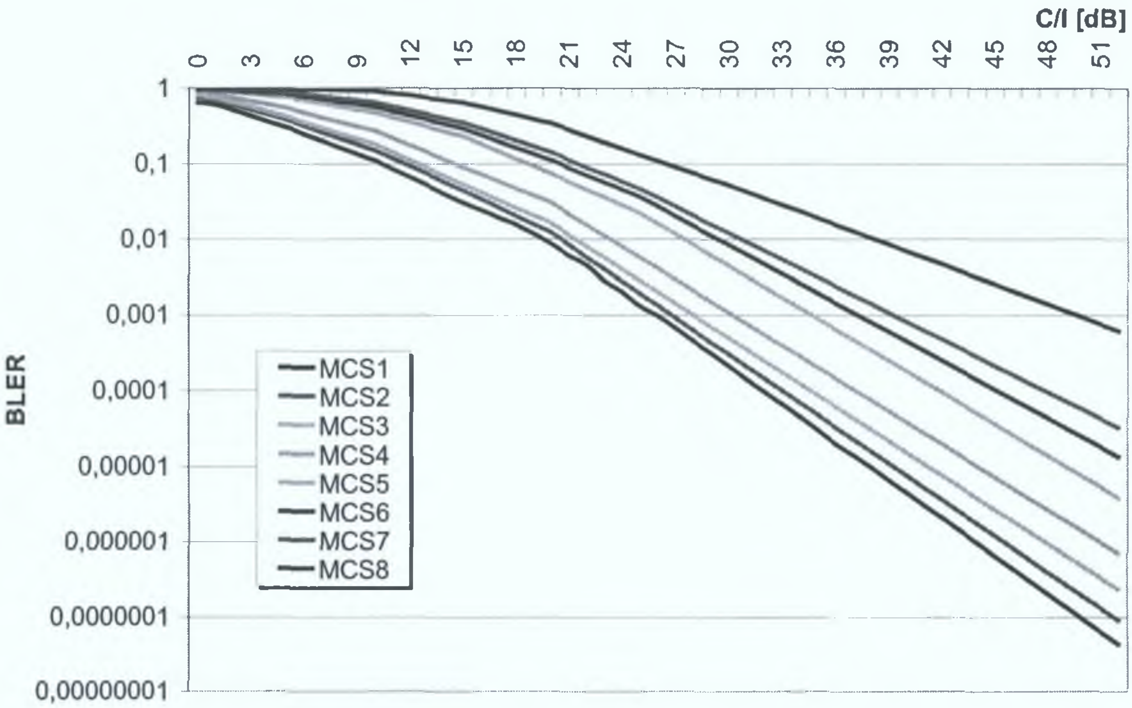


FIGURE 11: BLOCK ERROR RATIO VERSUS CARRIER TO INTERFERENCE RATIO FOR DIFFERENT CODING SCHEMES

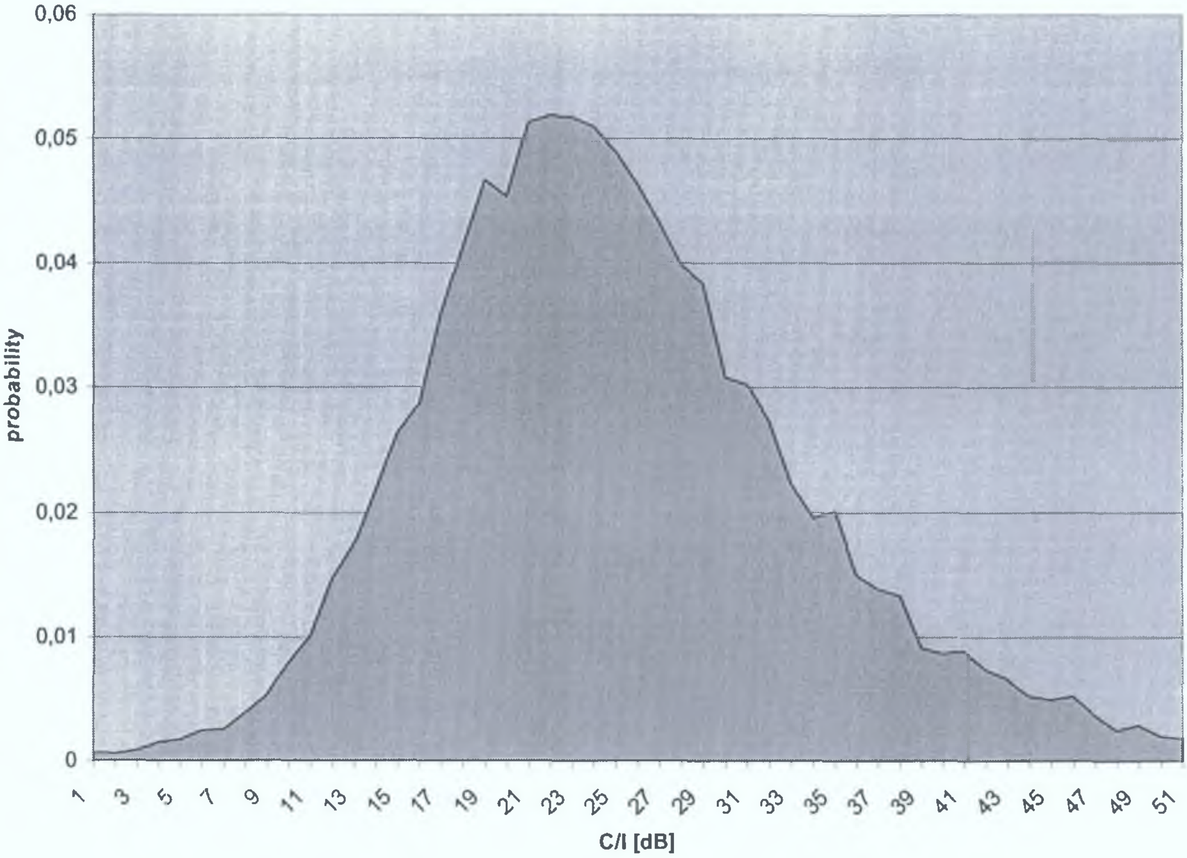


FIGURE 12: DISTRIBUTION OF THE CARRIER TO INTERFERENCE RATIO IN THE GSM CELL – EXPERIMENTAL RESULTS

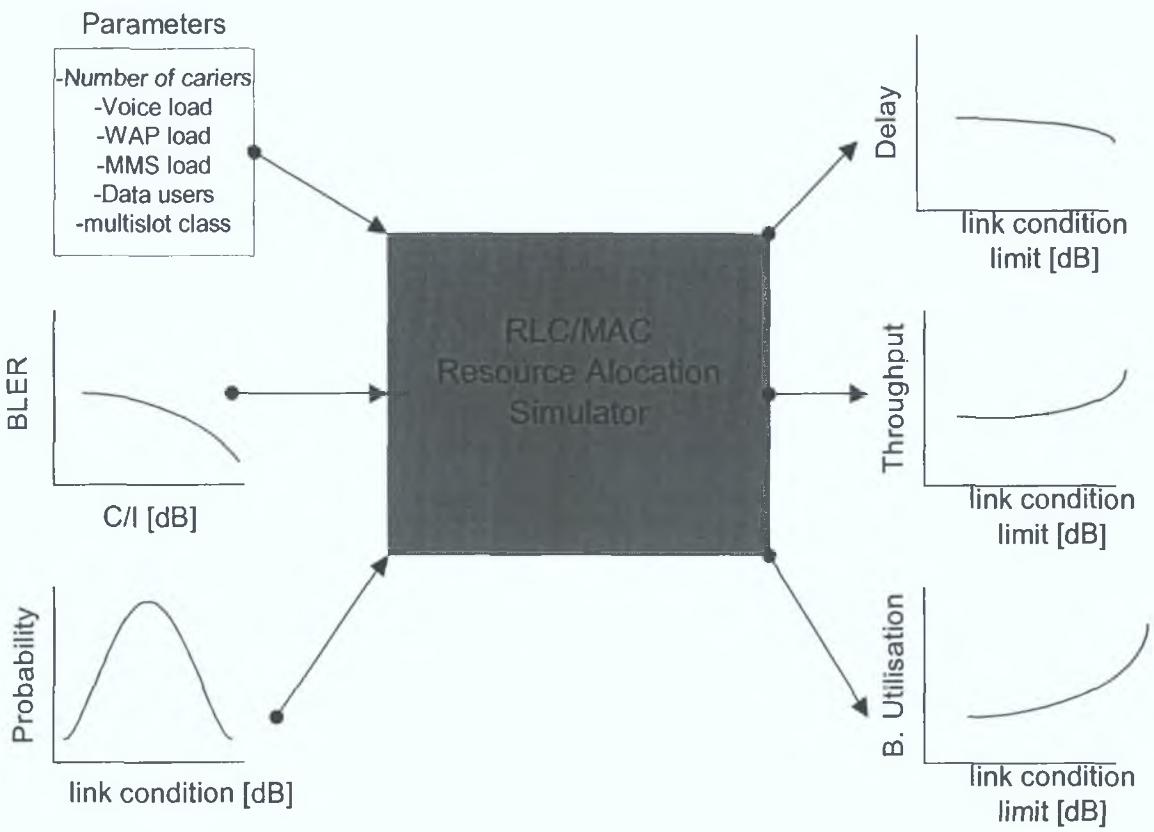


FIGURE 13: SIMULATION METHODOLOGY

3.3 Overall Model structure

The Figure 14 shows the structure of the RLC/MAC resource allocation model. The model consists of the voice traffic source, WAP traffic source, MMS traffic source, radio resource allocation algorithm at connection establishment, radio resource allocation algorithm during connection, voice traffic allocation algorithm, radio link simulator, link condition node, sink node and the system timer that controls all of the above elements.

3.4 Traffic Sources

To simulate the real GSM environment, the voice traffic sources in the model are divided into two groups: voice sources with strong mobility and voice sources with weak mobility.

The average TRX usage for high mobility users is much shorter than for low mobility users due to more frequent handover to adjacent cells. The voice source generates calls according to a Poisson process. The mean value of active time is 120 seconds for mobile stations, which remain in one cell (weak mobility), and is 45 seconds for mobile stations which move to another cell before finishing the call (strong mobility).

The data source in this model simulates the behaviour of WAP interactive sessions and also of the MMS background transmission. This can be modelled as a non real-time service, as proposed in [20] and [23]. The proposed model simulates downlink transmission. It is assumed that the uplink transmission uses one time slot and the transmission delay is equal 20 ms.

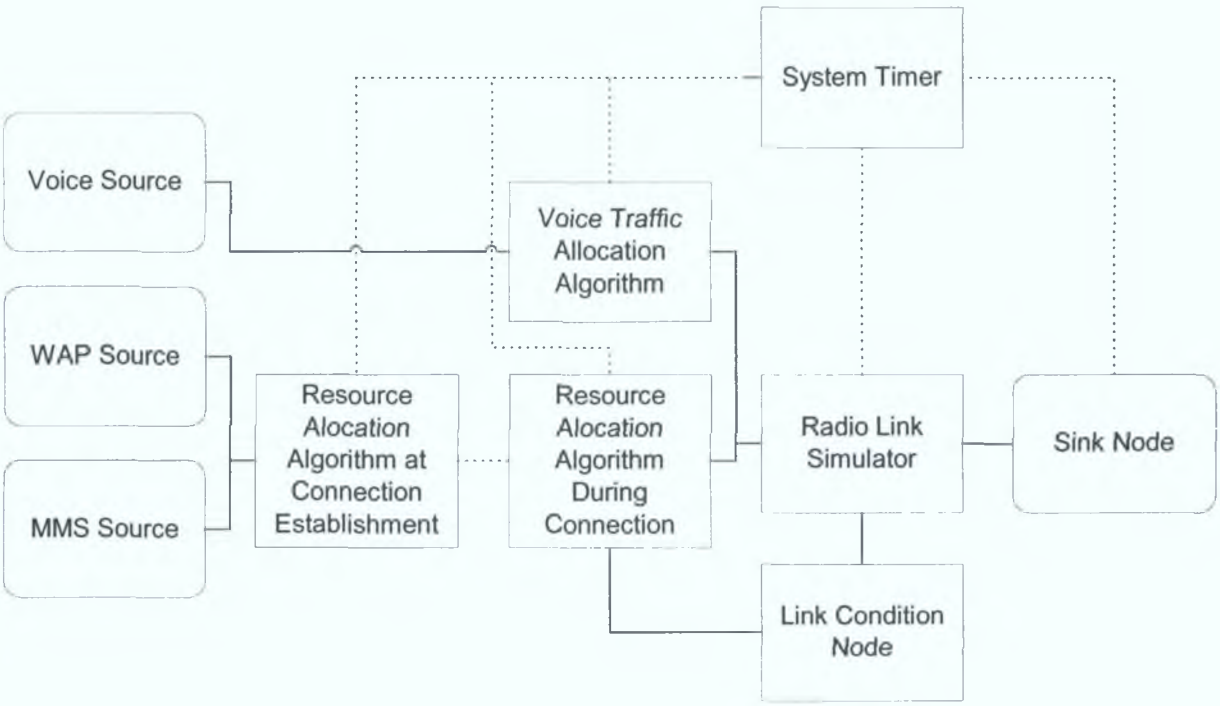


FIGURE 14: RLC/MAC RESOURCE ALLOCATION MODEL

Interactive WAP traffic sources are modelled by a bursty traffic with ON-OFF characteristics, where the OFF time represents the time during which a user is reading or browsing a WAP page. The ON time represents the time for the download of the files that belongs to a particular WML page as shown in the Figure 15. The T_d time represents download time, T_r represents reading time and D_w represents averages inter-arrival time between two WML pages. In the network model, the WAP server and backbone network are ideal except for a propagation delay.

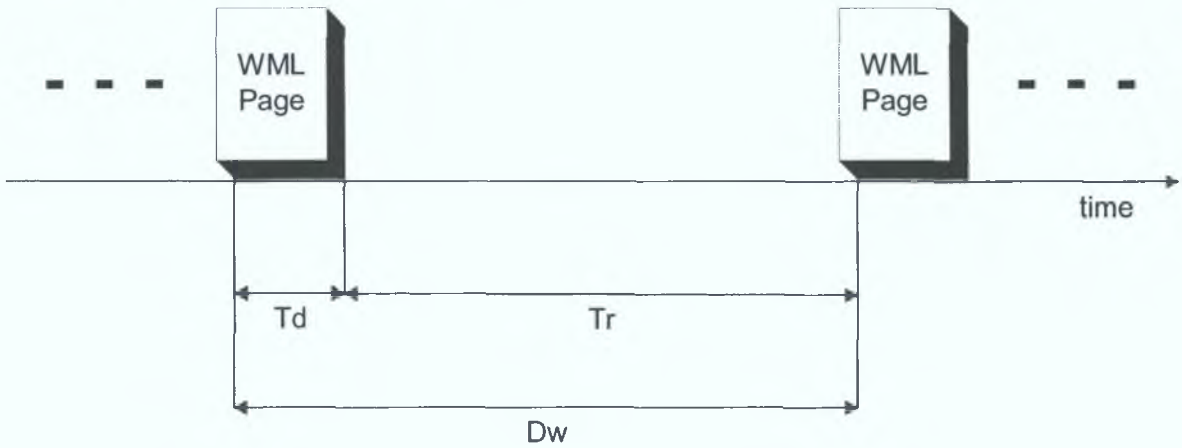


FIGURE 15: ON-OFF WAP TRAFFIC SOURCE CHARACTERISTICS

Background MMS data sources are modelled by a bursty traffic with ON-OFF characteristics as shown in Figure 16. When user receive a MMS message indication using PUSH SMS, it connects to the MMS Centre and downloads a multimedia message, which consists of several multimedia parts: pictures; sounds; and texts. The average size of MMS message is 50KB [27]. Typically the download process of MMS messages is invisible to the user.

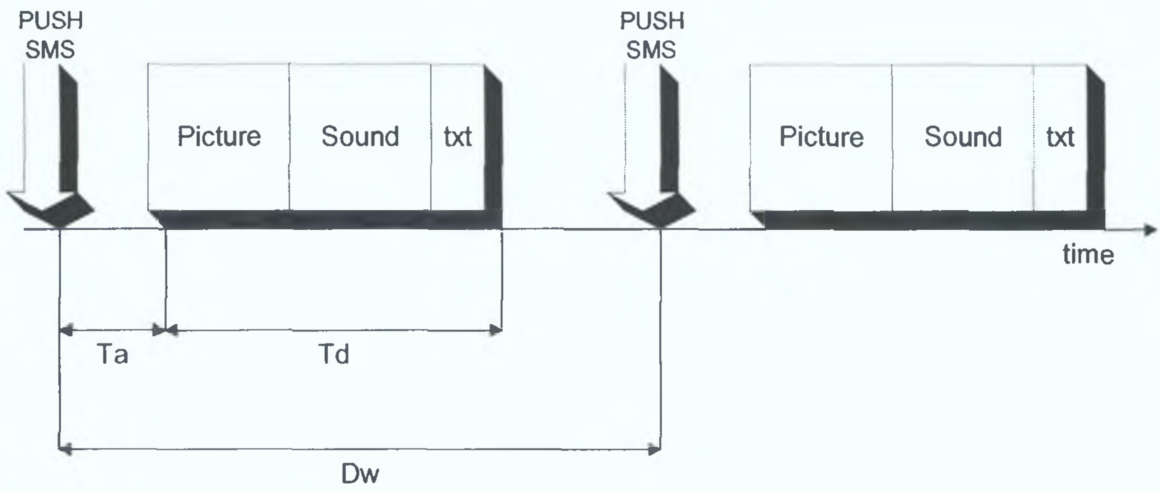


FIGURE 16: ON-OFF MMS TRAFFIC SOURCE CHARACTERISTICS

3.5 Algorithms

In the simulation it is assumed that the voice calls have priority and that there is no voice call loss due to GPRS traffic. It means that if the voice traffic increases and more and more users would like to set-up a call, then PDCHs which are used for GPRS traffic are automatically reconfigured to TCH (traffic channel used for voice).

When a user would like to transmit a file the resource allocation algorithm allocates to the user some PDCHs. If the CS traffic increases rapidly and there is no more PDCHs available to the data user, then the Temporary Block Flow will be lost and the user have to retransmit the whole data block. It is more favourable from the end user perspective to retransmit a background traffic class transmission, which is not visible to the end user than interactive session that will introduce visible delay to the user. To avoid retransmission of TBF for interactive traffic class, the radio resource allocation algorithm assigns to interactive traffic class only traffic channels, which have very small probability to reallocate it for voice traffic. The concept of safe resource allocation algorithms is described in [21].

The algorithm at connection establishment [21] allocates PDCHs to Mobile Stations within one carrier frequency. It divides the potential number of PDCHs Y into two groups. One part is (Y-[Y/2][1]) risky physical channels (high index) and the other part is ([Y/2]) safe physical channels (small index). The probability that the risky physical (high index) channel is occupied by a Circuit Switched calls is higher than the one of a safe (small index) physical channels. When a Mobile Station, whose multislot class is x(the number of time slots, which MS is able to use at the same moment), wants to establish a Temporary Block Flow (data transmission), then the network assigns for it to the ([x/2]) least loaded safe physical channels and (x-[x/2]) risky least loaded physical channels.

The second algorithm used in the model is executed during transmission. It concerns the distribution of block periods belonging to the same PDCH among Mobile Stations assigned to this PDCH [21][23]. This algorithm divides the block periods of PDCH between all the users of this PDCH according to the priority associated to the transmitted traffic class. The resource assignment is updated at each new user arrival and at each block period. The algorithm gives also a transmission priority to the set of blocks depending on the number of block retransmissions. This is like a set of FIFO queues with priority and with a Round Robin server whose allocation cycle is one block period.

The voice traffic allocation algorithm allocates incoming voice calls to the carrier frequencies, which are dedicated for Circuit Switched calls. If there are no more time slots available in the dedicated carrier frequencies, the algorithm allocates the Circuit Switched calls to the carrier frequency, which were previously allocated to the Packet Switched data.

---

[1] footnote [ ] means the integer part

3.6 Movement Model

The main role of the radio link simulator is to emulate the movement of the users in the cell and to calculate the radio block error probability associated with these movement based on the distribution of the link condition level as shown in the Figure 17.

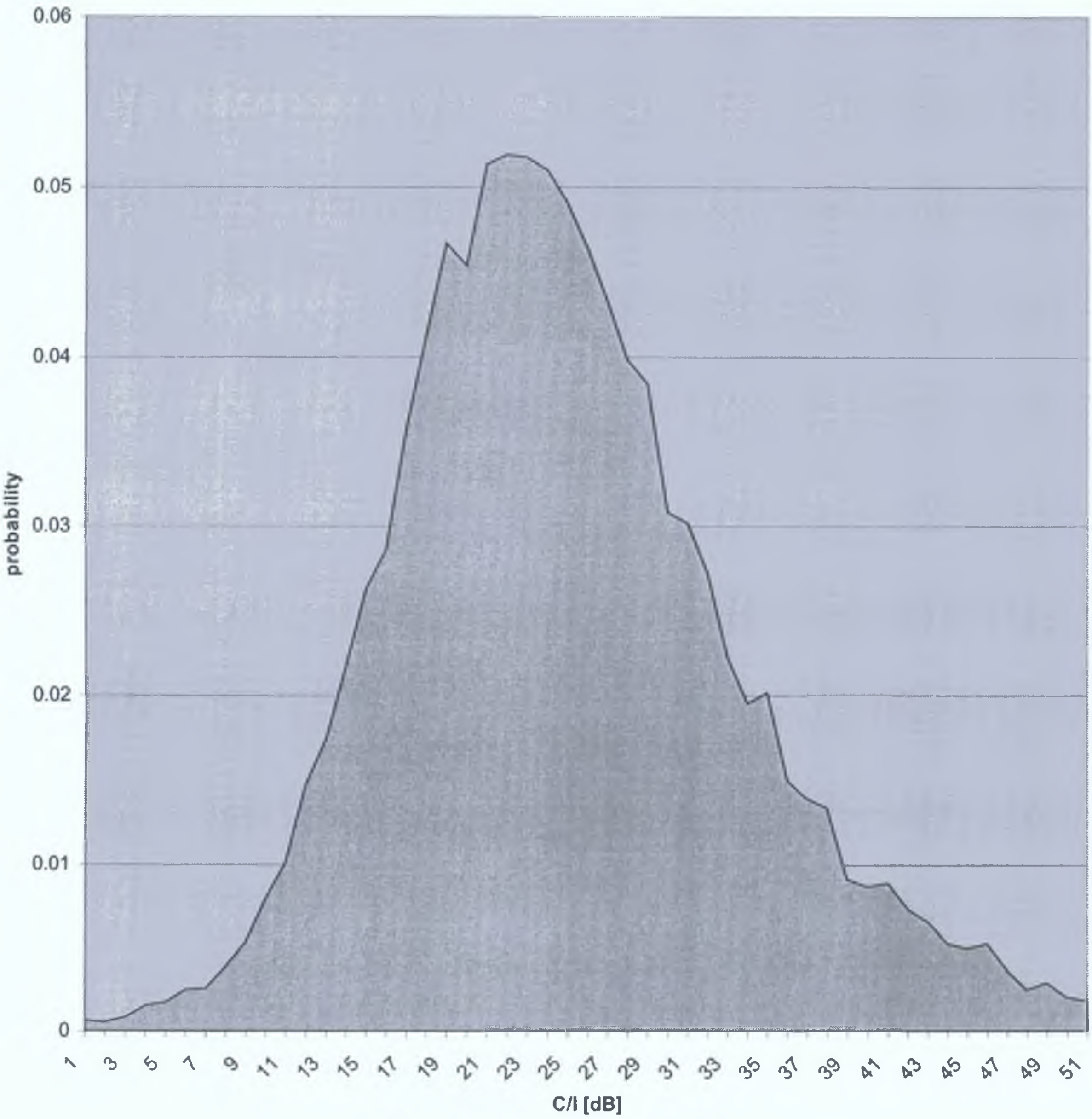


FIGURE 17: DISTRIBUTION OF THE CARRIER TO INTERFERENCE RATIO DURING USER MOVEMENT – EXPERIMENTAL RESULTS

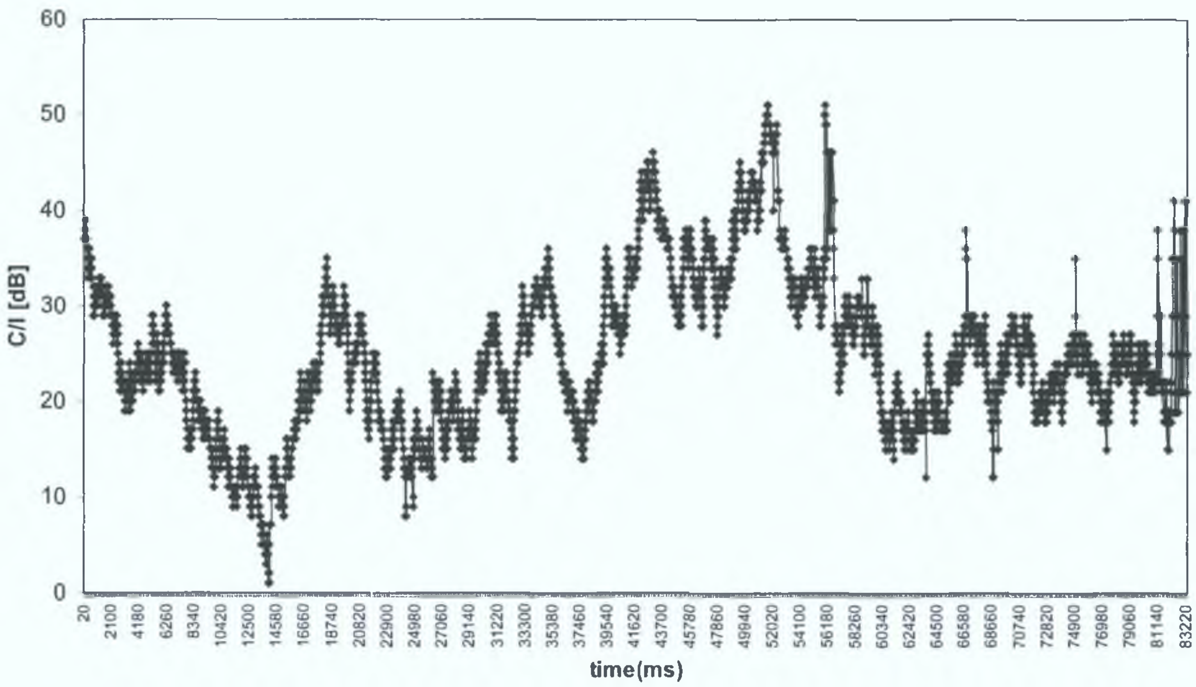


FIGURE 18: THE CARRIER TO INTERFERENCE RATIO VERSUS USER MOVEMENT IN THE CELL

Figure 18 shows how the carrier to interference ratio changes according to the user movements in the cell. The model is based on an assumption that the calculation of *C/I* level is done every block period and that the step is 1dB.

This model is based on experimental results from TEMS software and R520m Ericsson hardware. This experimental results show that the C/I level changes very fast (the changes could be in range of 4dB in the 150ms time period) and these fast changes don't depend on the user movement. The parameters, which depend on the velocity of the user, are included in the two-state user movement model and are responsible for the long-term C/I changes. The C/I values used in the simulation are results from the simulation of Markov movement state model as shown in Figure 18 with distribution of the C/I as shown in Figure 17.

A user increases or decreases its carrier to interference ratio with probability (1-p) and changes the trend with probability p. The p value used in the model was as follow:

p=0.1+random[0,1]*0.2+abs(26-(C/I))*0.005 where C/I is the actual value of C/I in the range of (0-52dB) ,random[0,1] is a random value from 0 to 1 and introduces random variation of p value, the p value differs from 0.1 (C/I equal 26 dB) to 0.43 (C/I equal 0 or 52dB).

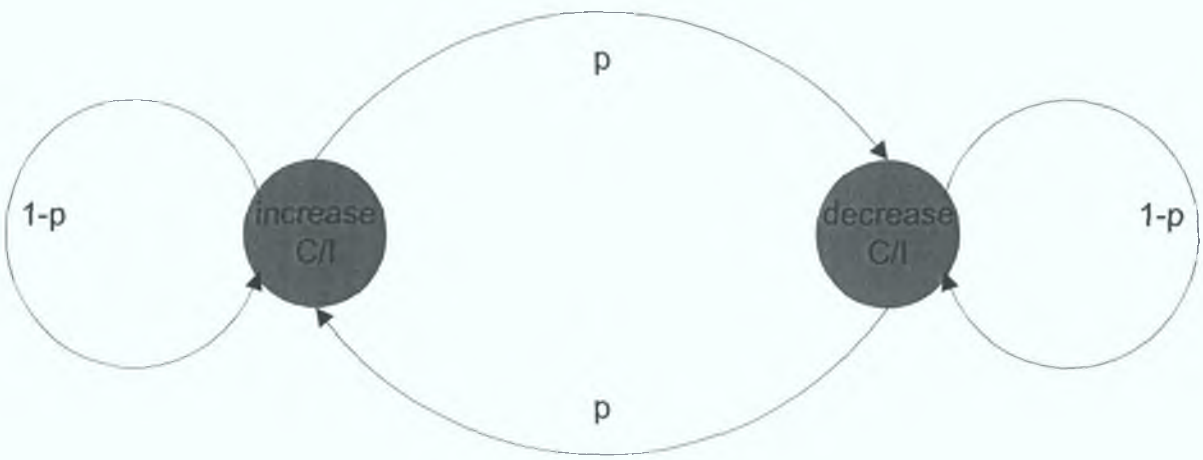


FIGURE 19: THE USER MOVEMENT DISCRETE STATE MODEL

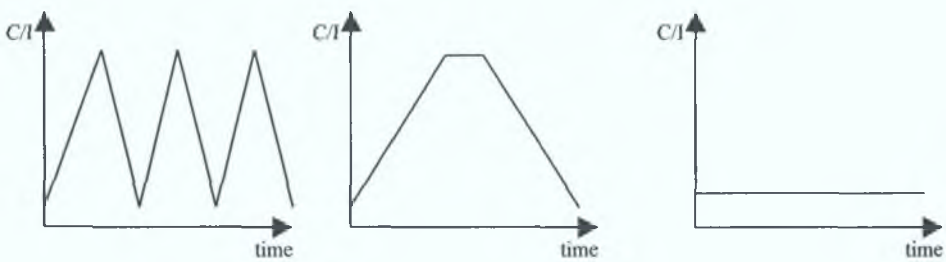
Figures 20-23 present theoretical model for the user movement in the sell. Lets assume that we have 12 units of time in which we are measuring the level of C/I for the user. In this time the fast user will move around the cell as shown in the Figure 20a) the slow user will move in the cell with half of the distance of fast user and static user will stay in the same place for the whole time. Each of the users observes interference from another users in the cell as shown in the Figure 21. The Figure 22 shows the C/I level for each of the user depending of the position in the cell. Because we are measuring C/I level in the discrete moments (each unit of time), so the user will observe the total C/I level as shown in the Figure 23. If we regard this process as a Markov model then we can observe that for fast users the probability that the next delta of measured value compared to the previous value of C/I will be with different sign, is low as the fluctuations originated from the user movement are dominating.

The slow users often change the trend of their curves due to fluctuations of interference originating from other users in the cell. On the other hand static users could be modelled with the Markov model with the probability of change the trend equal 1.

a) Fast users b) Slow users c) Static users

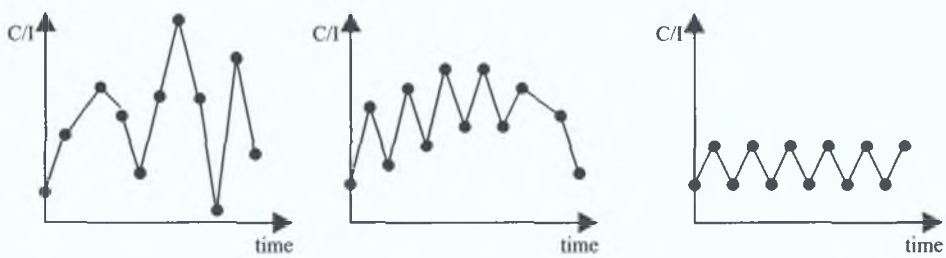
FIGURE 20: DIAGRAM OF USER MOVEMENT IN THE CELL

FIGURE 21: FLUCTUATION OF THE C/I CAUSED BY INTERFERENCE FROM OTHER USERS IN THE NETWORK



a) Fast users b) Slow users c) Static users

FIGURE 22: FLUCTUATION OF THE C/I CAUSED BY THE CHANGE OF SIGNAL LEVEL



a) Fast users b) Slow users c) Static users

FIGURE 23: SUMMATION OF FLUCTUATION OF THE C/I CAUSED BY THE CHANGE OF SIGNAL LEVEL AND BY INTERFERENCE FROM OTHER USERS IN THE NETWORK

The role of the sink node is to calculate the performance parameters, for example the time to first bit (TTFB) and the average throughput.

3.7 The interactive traffic class priority mechanism

This master thesis introduces priority mechanism based on periodical allocation of data blocks to the interactive traffic class. The prioritisation process take place in the Packet Control Unit. The PCU allocates available radio resources between all active users and different packet classes. In the 52 packet multi-frame there is 12 data blocks (4 data packet each). Without prioritisation mechanism data blocks are allocated to the users and different class of packets according to the first come first out policy. There is a separate queue for each of active user in the cell and all users have the same priority in the round robin server without any diversification related to the type of transmitted traffic class. In the interactive traffic class priority mechanism there is a diversification between different traffic classes. If the priority is equal to X, where X is between 1 and 11, then the resource allocation algorithm allocates X more data blocks to the interactive traffic class then to the background traffic class. For example if the priority is 1, then the interactive traffic class gets 1 block period more than background traffic i.e. the interactive traffic class gets 2 blocks and background traffic class gets 1 block. If the priority is High then the resource allocation algorithm gives the precedence to the interactive traffic and if there is no more interactive traffic blocks to transmit, then the resource allocation algorithm allocates remaining radio resource to background traffic class.

3.8 The radio link condition diversification mechanism

This master thesis introduces radio link diversification mechanism, which could be a part of Packet Control Unit. In the PCU the radio resource allocation algorithm allocates data packets for a particular time slot only if the approximation of the carrier to interference ratio is bigger than the link condition limit.

When a user downloads a data block and if there are bad link conditions in the network, then the transmission stops for a few moments and only resumes if the conditions are better in the network. In fact the transmission is broken into number of smaller transmissions, but it uses higher coding schemes. Another user could reuse the common transmission medium, when the radio link conditions are not sufficient for a high-speed transmission. This could reduce the unnecessary retransmissions in the EGPRS radio access network. In fact the radio resources are used in more efficient way because first the data blocks are not transmitted to the users witch are in the high interference areas of the cell and second the data blocks are transmitted with higher coding scheme (more data is transmitted in one data block).

The drawback of the radio link condition diversification mechanism is that it could introduce a delay if the user is staying in the high interference area for a long period of time. This could suggest using the radio link condition diversification mechanism only for background traffic.

Chapter 4

4. Results

4.1 Performance metrics

The first performance parameter used in these simulation is the average throughput of a Mobile Station. This is defined to be the quantity of data received correctly a Mobile Station in a unit of time. Packets failed during transmission are not counted for the average throughput and these packets are retransmitted again. The second performance parameter is the delay, and this is defined to be the time to the first bit being received. This is the exact latency introduced by EGPRS radio access network and does not include the external network and server delay. The delay is the average time which packets are waiting in the queue before they are transmitted. The delay is calculated as the time to first bit, so the delay does not depend on the length of transmitted data. The third performance parameter is the bandwidth utilisation, which is calculated as the ratio of the number of bytes sent to the total radio link capacity available for data users. The fourth performance parameter is the number of retransmitted data blocks.

The throughput, delay and bandwidth utilisation parameters consider only the TBFs correctly received. When a TBF is not correctly transmitted all blocks related to this TBF are considered failed.

The results are averaged over an entire simulation run. The simulation time has been set-up to 10000 sec.

4.2 Simulation parameters

The simulation parameters for the model are as follows:

- SNDCP header length equals 2 bytes – is the header of the SNDCP layer.
- LLC header length equals 7 bytes – is the header of the LLC layer.
- Length of LLC frames equals 1520 bytes.
-
- Coding schemes MCS-1 - MCS-9 used.
-
- multislot class equals 4 – is the maximum number of simultaneous used downlink time slots.
- number of TRXs used equals 3
- PDCH assigned for Circuit Switched equals 2
- PDCH assigned for Packet Switched equals 1
- Number of PDCHs dynamically assigned equals 20
- Control channel configuration:
    - Common control channels and dedicated control channels are transmitted on TSL0 of BCCH TRX
    - 2 voice TRX with 2 % blocking on Air Interface + 1 TRX for data PDCHs
    -
    - Voice traffic capacity equals 9 Erl
    - Number. of subscribers per cell equals 450

4.3 The influence of different interactive traffic class priority parameter

Figure 26 and 25 show the average time to first bit delay and average throughput versus different interactive class priority.

The loads of WAP and MMS presented in results equal 70 and 140kbps, this load is generated by 450 users in the cell, so each user generates an average payload of 0.15 and 0.31kbps respectively. The WAP users average WML page size is (data size to be transmitted in one block flow) is 1KB and for MMS users respectively 50KB. So each user establishes temporary block flow of 1KB and 50KB respectively.

In the first scenario if there is no priority defined, then the radio resource allocation algorithm allocates data blocks equally to the interactive and background traffic class. If the priority is equal to X, where X is between 1 and 11, then the resource allocation algorithm allocates X more data blocks to the interactive traffic class then to the background traffic class. If the priority is High then the resource allocation algorithm gives the precedence to the interactive traffic and if there is no more interactive traffic blocks to transmit, then the resource allocation algorithm allocates remaining radio resource to background traffic class.

The simulation results show that while increasing interactive traffic class priority the time to first bit delay decreases over 81% for interactive traffic class and increase only slightly – less than 11% for the background traffic class. Using the priority for interactive traffic class it is possible to increase over 65% the average throughput of the interactive traffic class and to not decrease significantly – less than 5% the average throughput of the background traffic class.

The average system throughput increases from 102kbps without any priority mechanism to 114kbps with the high priority mechanism. It gives about 10% increase in the total system throughput. This happen because MMS traffic consists of large data bloks (about 50KB) and WAP is a transmission of large number of small data blocks (about 1KB). The average throughput of small data blocks could decrease if the common transmission medium is mostly used by the transmission of large data blocks. The delay of small data blocks is large and the average throughput decreases significantly. The results show that using interactive traffic class priority the operator may offer much better connection quality for users, which uses interactive traffic class services like WAP browsing and increase the total system throughput.

It is important to note that the quality of background services like multimedia messaging, does not decrease significantly and is not visible to the end user. The interactive allocates more bandwidth to the traffic class, which is more sensitive to delay and has a more bursty traffic characteristic. This technique is easy to implement in the wireless network and gives good results.

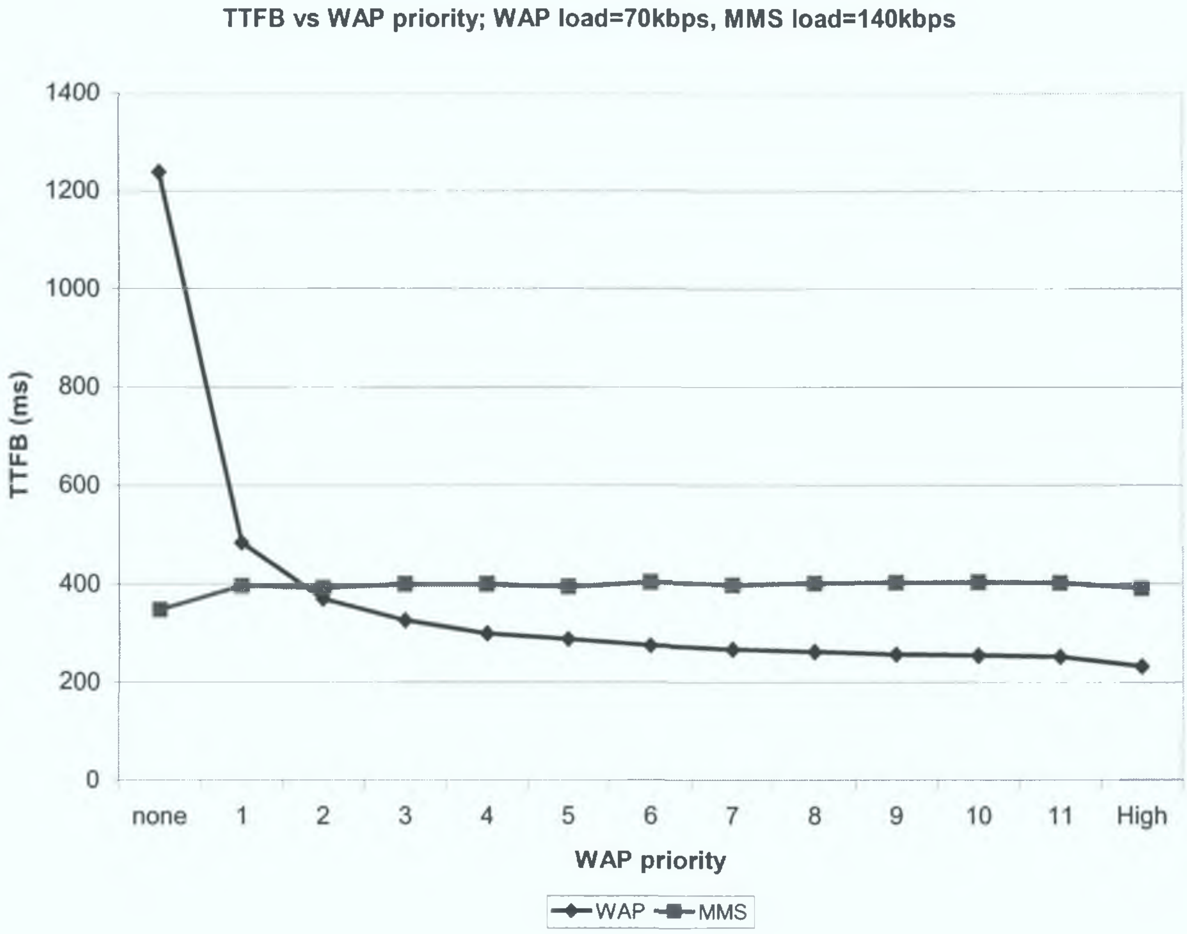


FIGURE 24: THE INFLUENCE OF THE INTERACTIVE TRAFFIC CLASS PRIORITY LEVEL ON THE TTFB

Figure 26 shows the total average system throughput versus different interactive class priority. The priority parameter defines the number of data blocks in the multiframe that is allocated to the interactive traffic class. Each data block consists of 4 consecutive time slots as defined in [12].

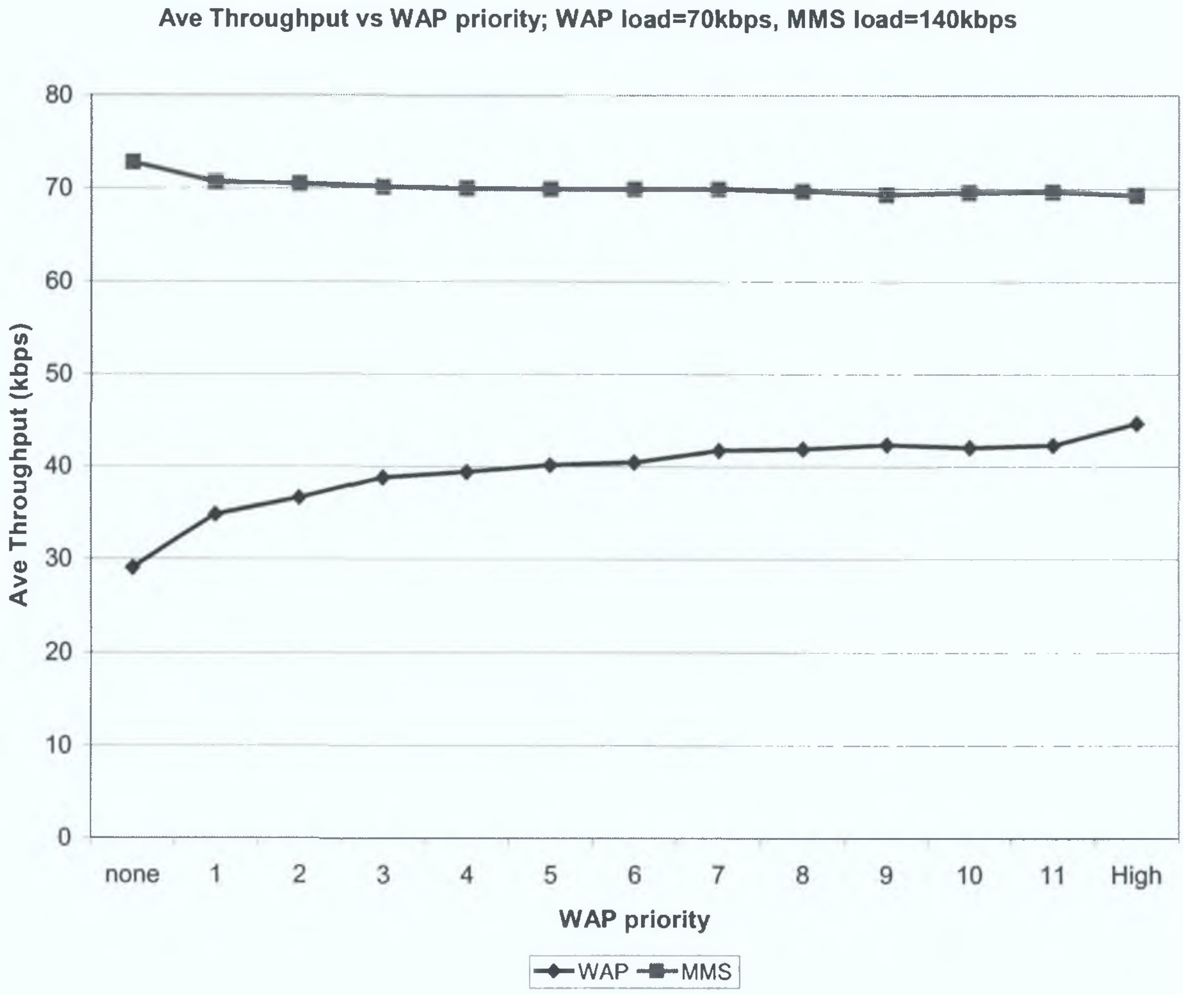


FIGURE 25: THE INFLUENCE OF THE INTERACTIVE TRAFFIC CLASS PRIORITY LEVEL ON THE AVERAGE THROUGHPUT

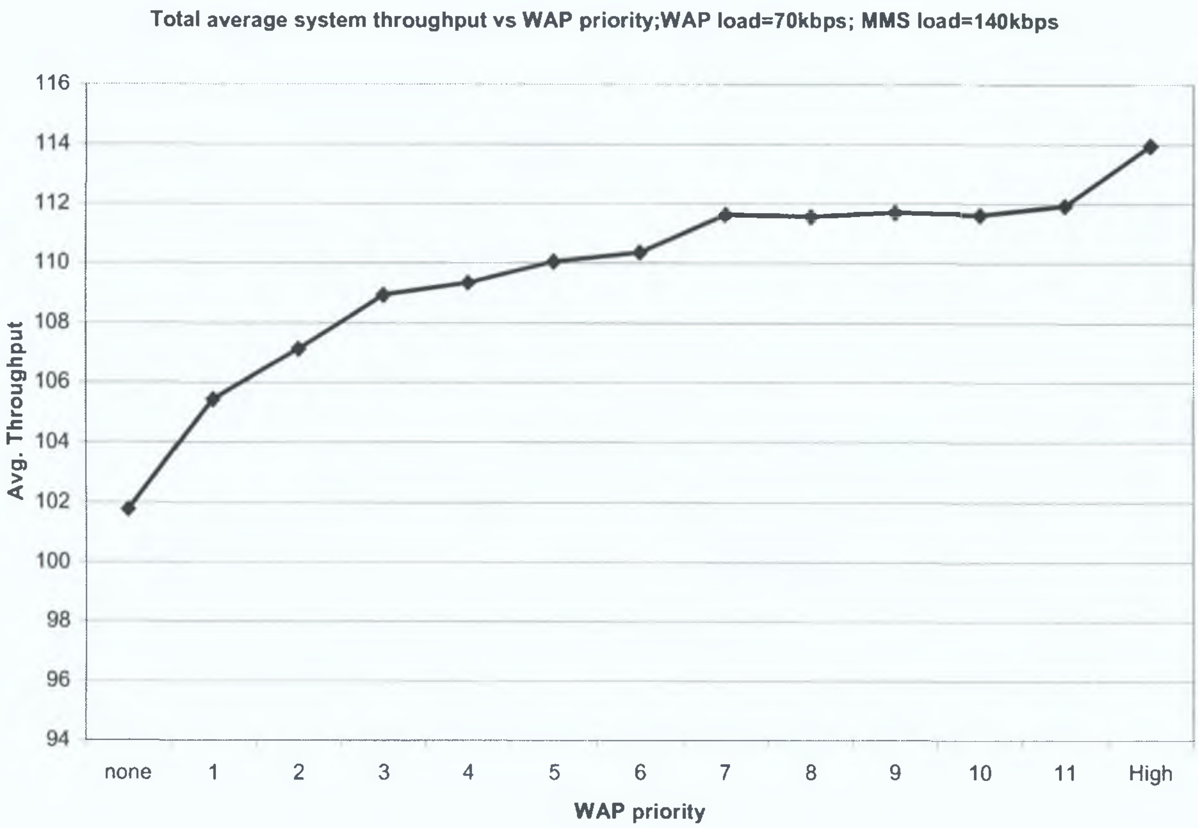


FIGURE 26: THE INFLUENCE OF THE INTERACTIVE TRAFFIC CLASS PRIORITY LEVEL ON THE TOTAL AVERAGE SYSTEM THROUGHPUT

4.4 The influence of link condition limit on background traffic class

Figures 27,28,29 and 30 show the average time delay to first bit, the average throughput and the number of retransmissions versus different link condition limit scenarios for a background traffic class. The influence of interactive traffic class is not taken in to account in this simulation. The mechanism for differentiating packet transmission based on radio link condition is a part of Packet Control Unit. The radio resource allocation algorithm allocates data packets for a particular time slot only if the approximation of the carrier to interference ratio is bigger than the link condition limit.

Figures 27 to 30 are calculated using "mean" value of C/I equal 16dB. The unit values of C/I were from 0 to 52 dB and were distributed according to the C/I distribution shown on Figure 12 but with mean C/I equal 16dB. The simulator calculates apriori from the C/I distribution the number of C/I values to be used in the simulation, so if the simulation is running the Markov movement model takes the C/I values from the list generated from the C/I distribution.

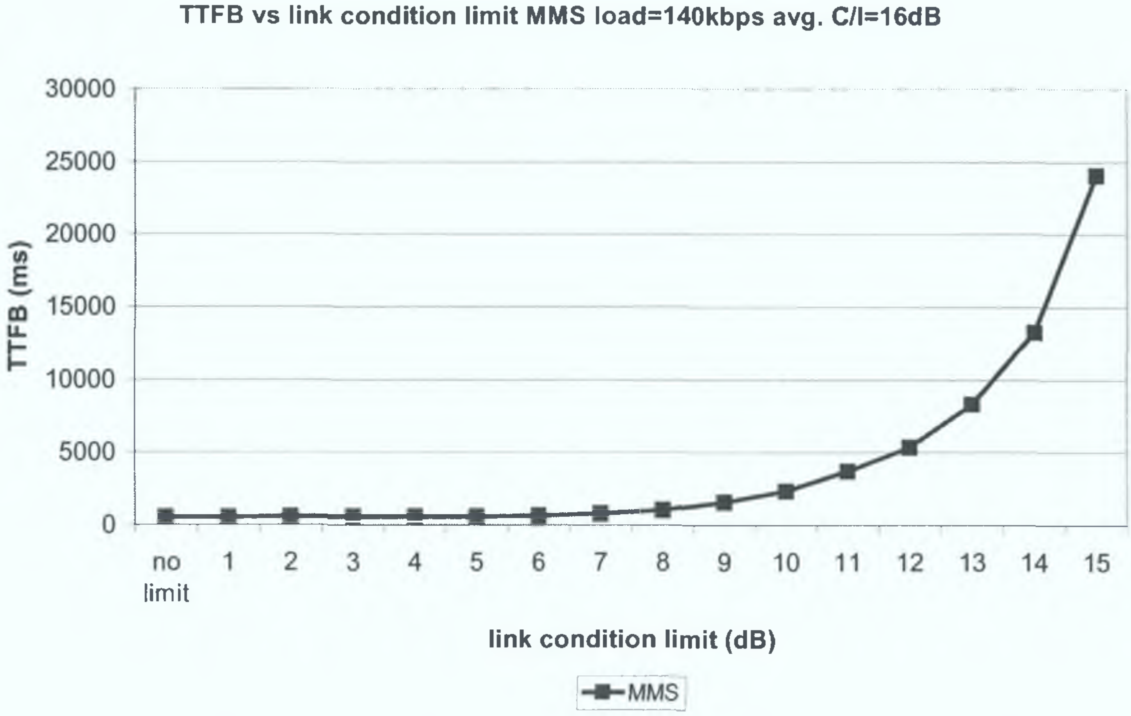


FIGURE 27: THE TIME TO FIRST BIT DELAY VERSUS LINK CONDITION LIMIT

The average time to first bit delay increases while link condition limit increases. When a user gets the push indication and tries to download a MMS message, they have to wait until the link conditions are reasonable. This produces higher average time to first bit delay. This is not critical for the end user because the downloading process is invisible to the users.

The average throughput has a very similar situation. When a user downloads the MMS message and if there are bad link conditions in the network, then the transmission stops for a few moments and only resumes if the conditions are better in the network. In fact the transmission is broken into number of smaller transmissions, but it uses higher coding schemes. The transmission with good link condition and higher coding schemes causes in increasing the bandwidth utilisation and reduction of retransmissions. The bandwidth utilisation is shown in the Figure 29.

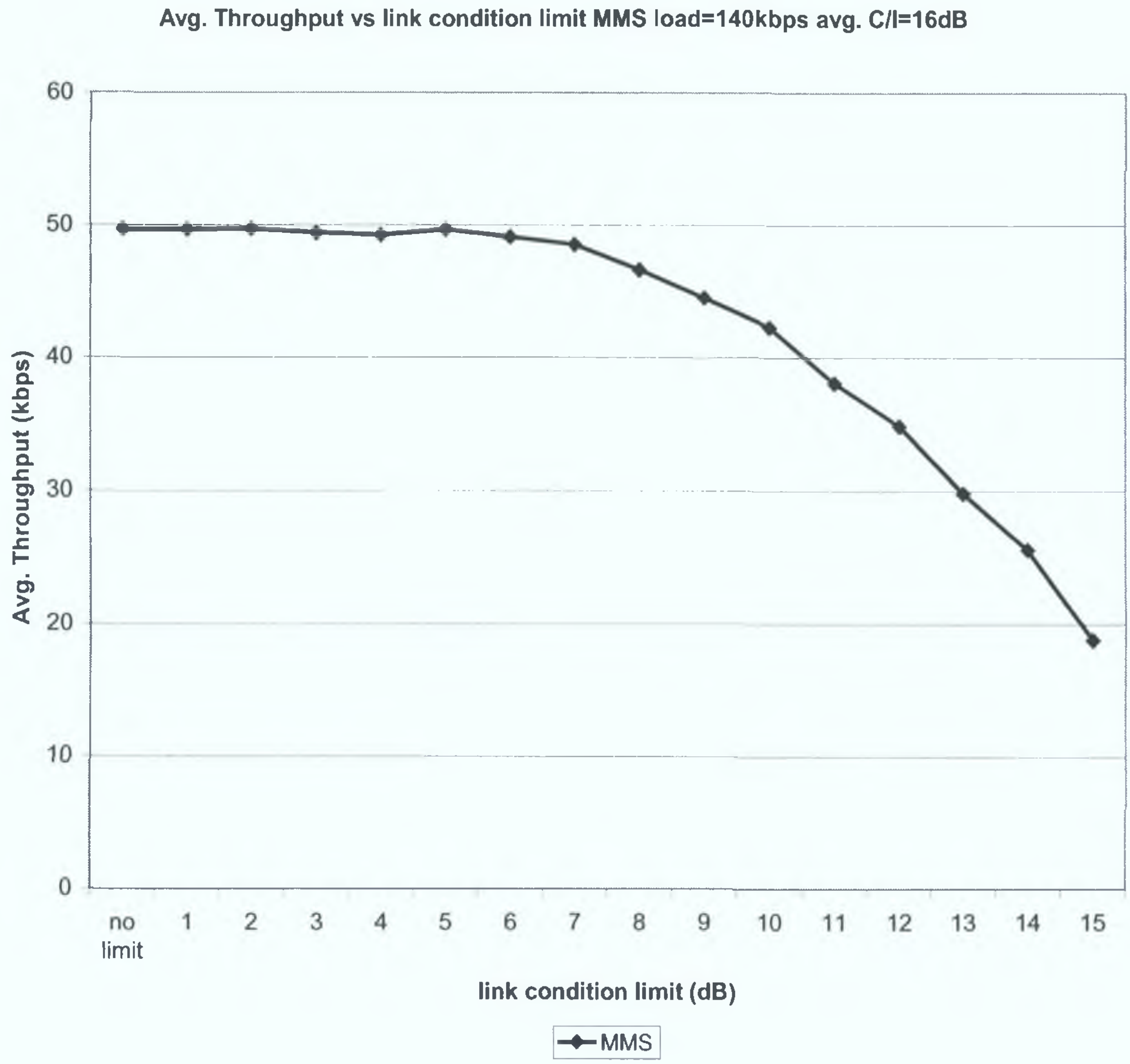


FIGURE 28: THE AVERAGE THROUGHPUT VERSUS LINK CONDITION LIMIT

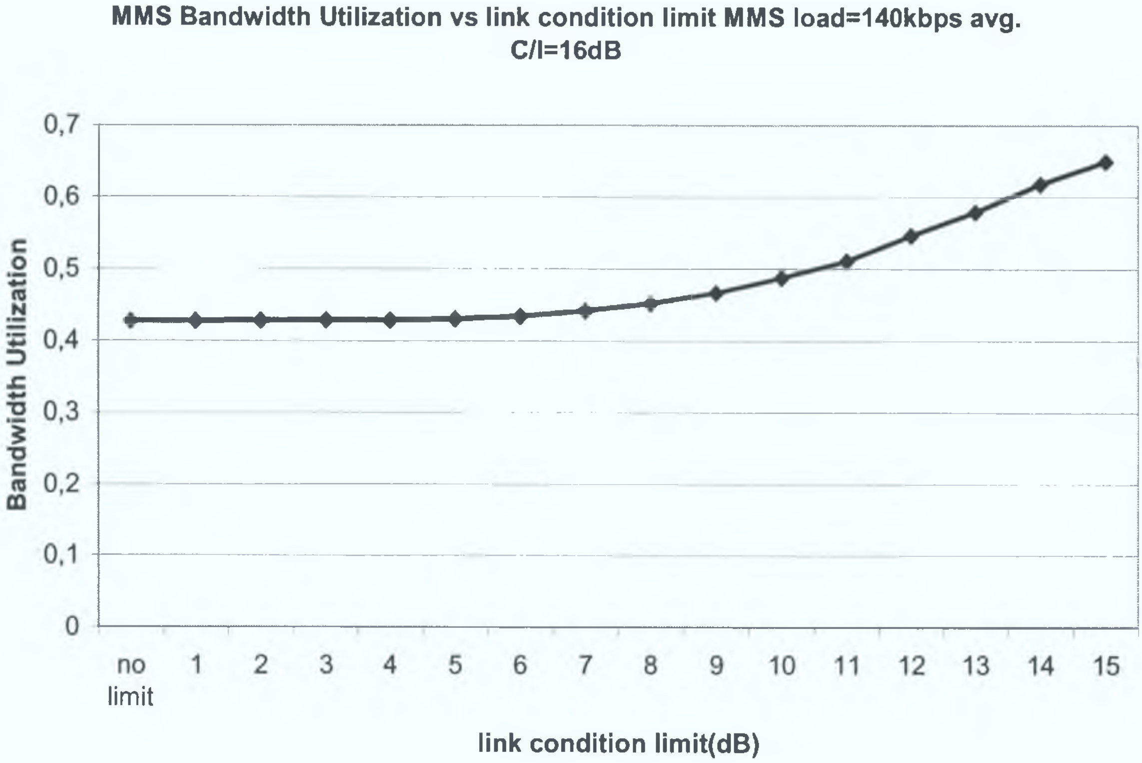


FIGURE 29: THE BANDWIDTH UTILISATION VERSUS LINK CONDITION LIMIT

The introduction of the link condition limit mechanism causes a smaller number of retransmissions in the network, as is shown in Figure 30. For example the number of retransmissions of the data packet using link condition limit equal 15dB is 58% smaller than the number of retransmissions without the link condition limit mechanism. The link condition limit mechanism saves the radio resources in the network, so that these resources could be used for other services with higher precedence.

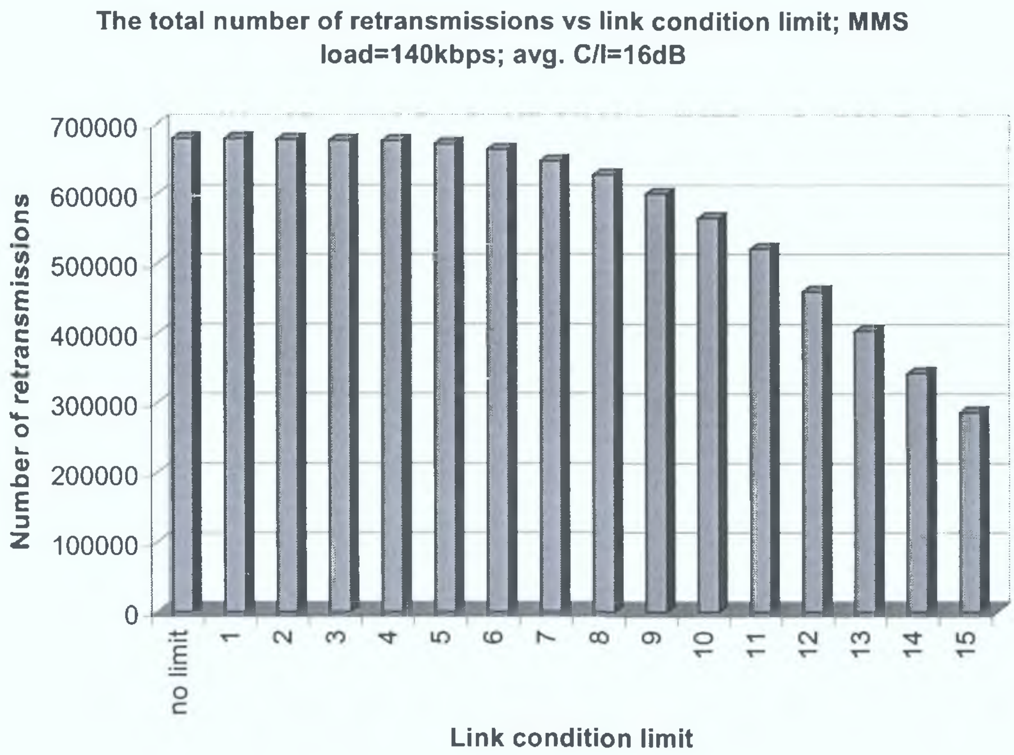


FIGURE 30: THE TOTAL NUMBER OF RETRANSMISSIONS VERSUS LINK CONDITION LIMIT

4.5 The influence of MMS link condition limit on overall performance

Figure 31 shows how the link conditions limit implementation for the background traffic class influences the average time to first bit delay of the interactive traffic class. While the link condition limit increases, the average time to first bit delay of the WAP traffic decreases. That means that the WAP users will download WAP sites much faster.

The TTFB delay could decrease from 1785 ms to 220 ms with the link condition limit equal 15dB. The reduction is in the range of 88%. This is significant improvement for the performance of the WAP traffic.

The TTFB delay for Multimedia Messaging Services traffic increases slightly and is not visible to the end users if the link condition limit used in the algorithm is small and increases significant if the link condition limit is grater than 13dB as shown in Figure 31.

The average throughput for interactive and background traffic class is not degraded significantly if the link condition limit used in the algorithm is less than 13dB. However it does degrade significantly if the link condition limit is greater than 13dB, as is shown in Figure 32.

The optimal results for this particular simulation input data are obtained when the link condition limit equals 12dB. The result of 12dB breakpoint is valid only for this model and could be slightly different for another settings and real network environment. There is a break point where there is no benefit of using radio resource diversification algorithm, but it is still worth to use such an algorithm for background services where the delay performance metrics is not so important. In the real system we don't know the average C/I a priori, so we have to use a prediction of C/I value in the future. Because such a prediction introduces some errors, so we need to setup the system to be operated at a point somewhere to the left to the optimum point in order to guarantee small improvement of the performance and save radio resources, which are wasting due to the fact of unnecessary retransmissions and band quality link conditions. Figure 33 shows that the average system throughput reaches its maximum value for link condition limit equal to 12dB, and increases by 1.4%. In the same time the bandwidth utilisation increases by 1.2%, the average WAP throughput increases by 4.3% and the TTFB delay for the WAP traffic decreases by 84% as shown in Figures 31, 32 and 34. Introducing a simple link condition limit mechanism it is possible to reduce the TTFB WAP traffic delay by more than 80% while there is only a 2 seconds TTFB delay for MMS interactive traffic. The delay is calculated per 1KB packet for WAP and MMS traffic.

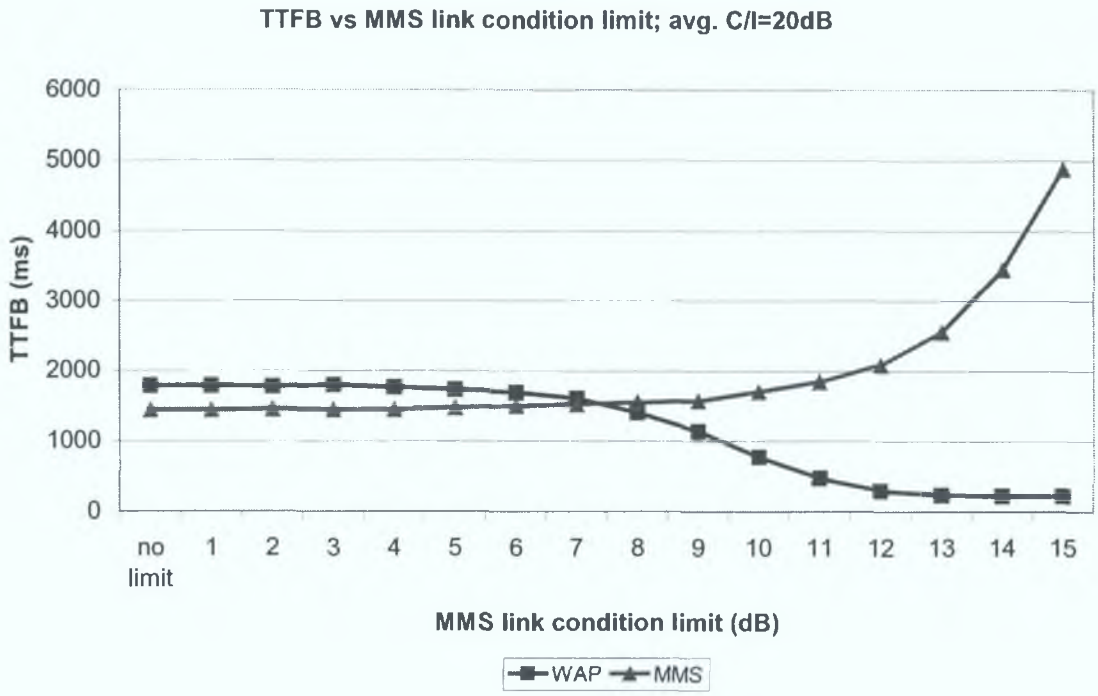


FIGURE 31: THE AVERAGE TIME TO FIRST BIT DELAY VERSUS LINK CONDITION LIMIT

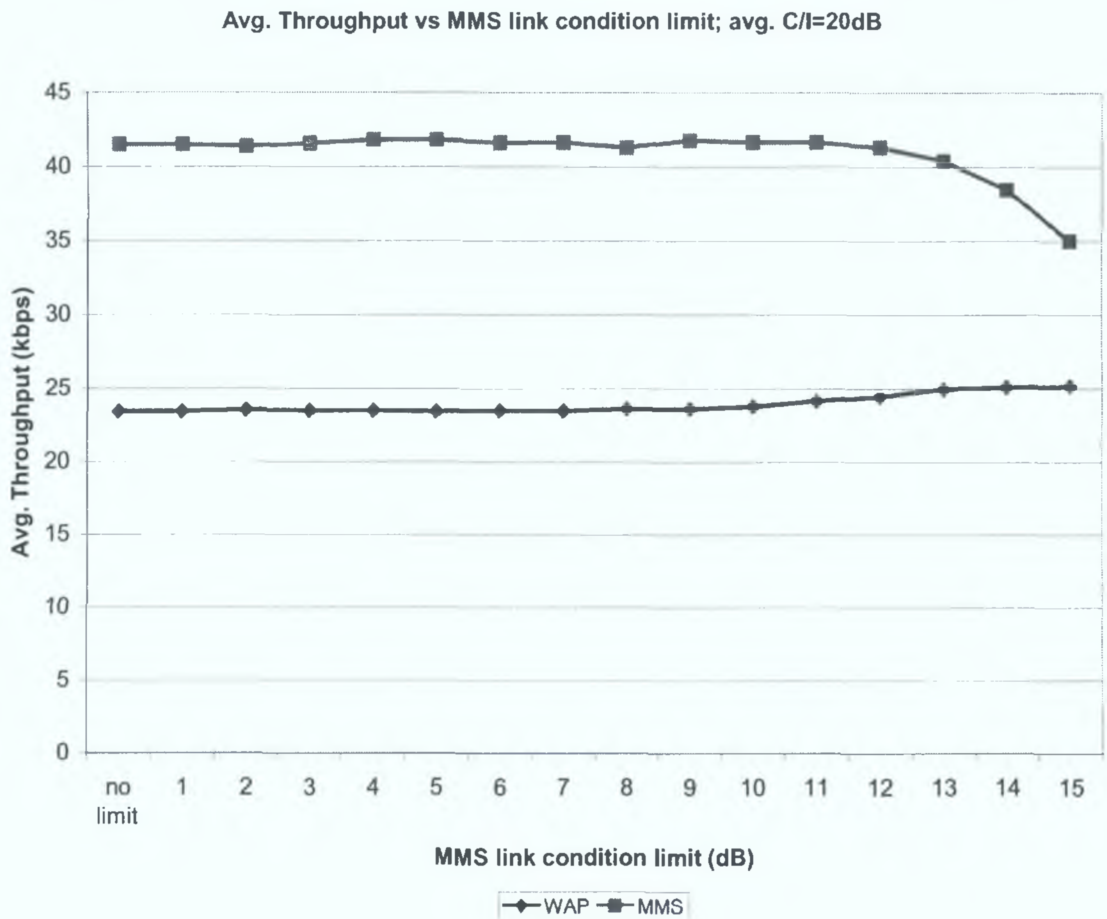


FIGURE 32: THE AVERAGE THROUGHPUT VERSUS LINK CONDITION LIMIT

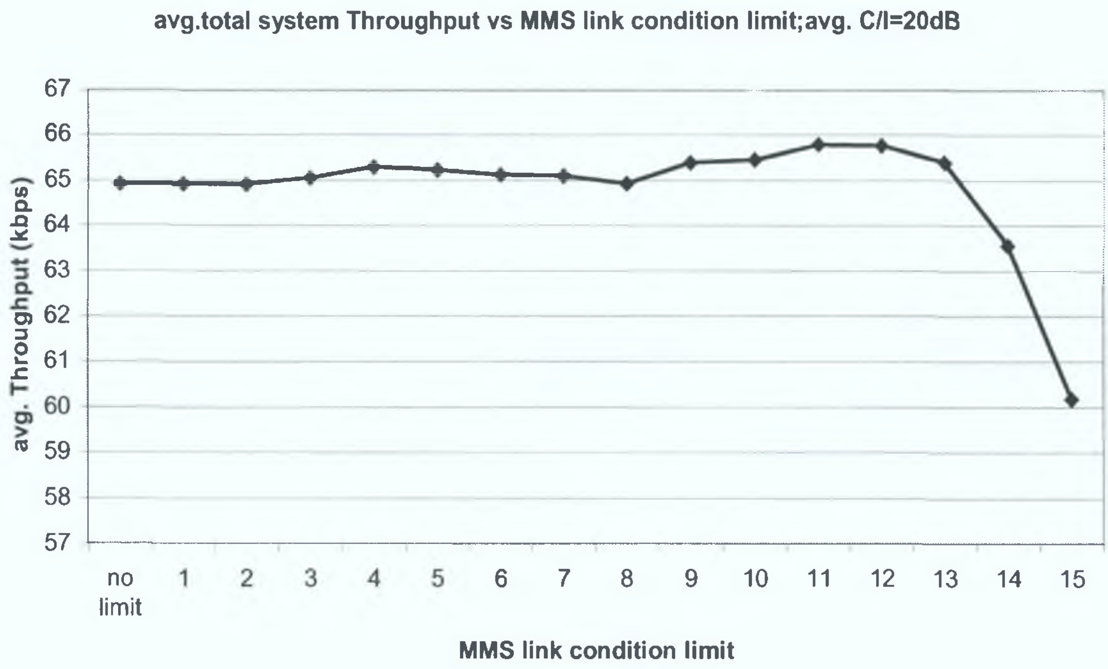


FIGURE 33: THE AVERAGE SYSTEM THROUGHPUT VERSUS LINK CONDITION LIMIT

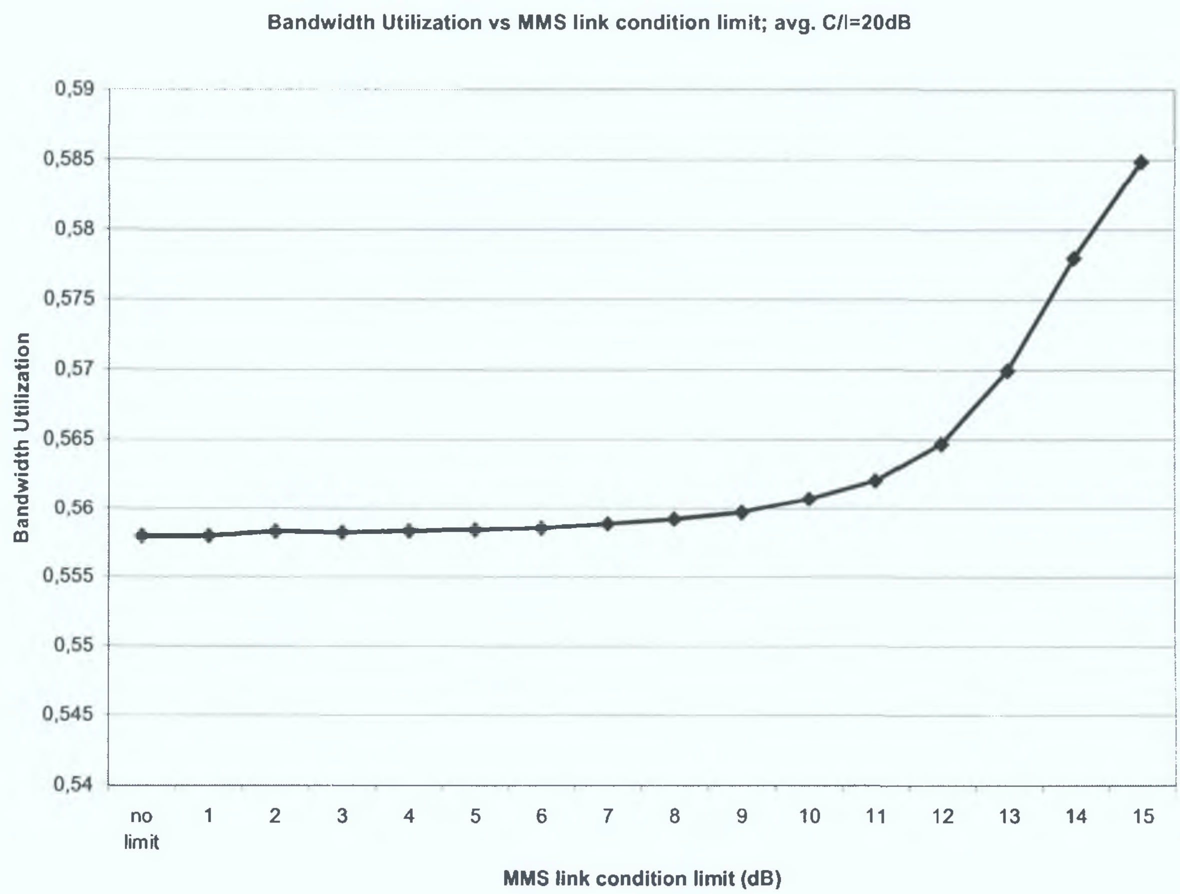


FIGURE 34: THE BANDWIDTH UTILISATION VERSUS LINK CONDITION LIMIT

This technique is much harder to implement in wireless networks than the interactive traffic class prioritisation technique because it needs to constantly monitor the link quality inside the Packet Control Unit. However this technique can give significant gains to the network operator and to the end user. The end user gets smaller delay while browsing interactive services and the operator utilises the scare bandwidth in the network in the more efficient way.

Link condition limit for background traffic class makes the wireless network more intelligent while allocating the scare bandwidth to different traffic classes. The multimedia messages are transmitted only if the link conditions are good enough to guarantee better throughput and better bandwidth utilisation. The interactive traffic is transmitted continuously not paying attention to the link conditions on the wireless link, which guarantee very low delay but not good throughput. In fact the guarantee for a high throughput is not needed for interactive WAP browsing, because the interactive WAP traffic is bursty and contains a large number of small files. The opposite is true of MMS traffic, which contains a smaller number of larger files, and needs high throughput to transmit these files in more efficient way.

Chapter 5

5. Conclusions

The academic overview that is provided in this masters thesis reveals several aspects of the Enhanced GPRS system. This thesis presents two techniques to improve the overall performance, especially for typical mobile Internet applications like WAP browsing and Multimedia Messaging Services.

The MMS messages have a bright future to share in the big success that was achieved by SMS messaging. The MMS messages introduce to the network large background traffic. This may affect performance of the WAP - the traditional mobile Internet browsing.

The presented techniques like interactive traffic class prioritisation and link condition limit for background traffic class allows the network operator to improve the performance of interactive mobile internet browsing without affecting user perception of MMS services.

The first proposed method does not affect the performance (TTFB delay and average throughput) of the Multimedia Messaging Services. Introduction of the priority mechanism increases the user perception of interactive services like WAP sessions. The improvement of the performance is done by more efficient allocation of the available resources to the interactive traffic, but it does not the bandwidth utilisation to increase and does not decrease the number of retransmitted data blocks.

The second proposed method slightly affect the performance of the Multimedia Messaging Services, but it is not critical because the transmission of MMS messages is invisible to the user. Introduction of the link condition limit increases the user perception of interactive services. This method do not give as big throughput improvement as the first method give, but the operator can profit with the bandwidth utilisation improvement and reduction of the number of retransmitted data blocks.

Both of presented methods allow the end user to improve significantly the perception of the interactive services (the TTFB delay reduction by more than 80 %). The first method is focus on the end user perception of the offered services and the second presented method is more focused on the operator perception of the network resources usage. The drawback of the second method is that the setting to high link condition limit may cause in huge delay of the MMS messages transmission and significant reduction of the average system throughput. The first method does not have this drawback and setting the highest priority for the interactive traffic class does not cause in reduction of the average MMS and WAP throughput. Both of presented method does not impact on voice call blocking probability. The implementation of the second presented method is much harder because it need the estimation of the link condition for every of the active user in the cell.

Presented simulation shows that using sophisticated resource allocation the network operator may reduce the unnecessary delay of the interactive WAP traffic and to make the most of the network resources. Introduction of new multimedia services to the network must not affect the performance of existing services, but it should be added some intelligence to the resource allocation algorithms in the network.

Previous study of radio resource allocation have investigated different algorithms and resource allocation policies in terms of service offering, such as premium and basic, rather than traffic class. This thesis presents a new approach to intelligent radio resource allocation. The resource allocation policy uses different traffic class characteristics and allocates the scare bandwidth in different way to each traffic class.

The proposed enhancements of radio resource allocation algorithms like interactive traffic class prioritisation and background traffic class link condition limit improves the end user perception of interactive services quite significant and increases the bandwidth utilisation of the operator's network. However the presented techniques were studied in the context of EGPRS network, but they could be used in any wireless packed data network including 3G, CDMA2000 and UMTS.

References

[1] 3rd Generation Partnership Project, Technical Specification Group GSM/EDGE Radio Access Network; General Packet Radio Service (GPRS); Overall description of the GPRS radio interface; Stage 2 (Release 1999), 3GPP TS 03.64, pp. 33-34.

[2] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, General Packet Radio Service (GPRS), Service description, Stage 1 (Release 1999), 3GPP TS 22.060, pp. 17-18

[3] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, QoS Concept and Architecture (Release 1999), 3G TS 23.107, pp. 9.

[4] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; General Packet Radio Service (GPRS); Mobile Station (MS) Base Station System (BSS) interface; Radio Link Control/ Medium Access Control (RLC/MAC) protocol (Release 1999) 3GPP TS 04.60, pp. 18-19

[5] 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Overall description - Stage 2; (Release 5) 3GPP TS 43.051, pp. 50-52

[6] 3rd Generation Partnership Project; Technical Specification Group Terminals; Multimedia Messaging Service (MMS); Functional description; Stage 2 (Release 1999) 3G TS 23.140, pp. 97-100

[7] W. Ajib and P. Godlewski, "Service Disciplines Performance for WWW Traffic in GPRS Systems", Proc. of IEE 3G 2000 Mobile Communications Technologies, London, U.K., March 2000, pp. 3-4

[8] S. Bajaj, L. Breslau, and S. Shenker, "Is service priority useful in networks?" SIGMETRICS '98/PERFORMANCE'98, Performance Evaluation Review, vol. 26, pp. 66-77, June 1998

[9] C. Bettstetter, H.J. Vgel, J. Eberspcher, "GSM Phase 2+, General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface", IEEE Communications Surveys Vol. 2, No. 3, 1999, pp. 15-18.

[10] L.F. Chang et. al., "Providing differentiated services in EGPRS through packet scheduling", IEEE Globecom, Vol.3, pp. 1515-1521, 2000.

[11] L.F. Chang et. al., "Providing differentiated services in EGPRS through radio resource management", ICC '01. Volume: 7, pp. 2296 -2301, 2001.

[12] ETSI, “Overall description of GPRS radio interface”, GSM 3.64, ver. 6.2.0, 5-1999, pp. 9-54

[13] C.H. Foh, B. Meini, B. Wydrowski and M. Zuckerman, “Modelling and Performance Evaluation of GPRS”, Proc. IEEE VTC 2001, Rhodes, Greece, pp. 2108-2112, 2001.

[14] A. Furuskär, J. Näslund, H. Olofsson, “EDGE Enhanced Data Rates for GSM and TDMA/136 Evolution”, Ericsson Review Vol.76 (1999): 1, pp. 28-37, 1999.

[15] H. Graja, P. Perry and J. Murphy, "Development of a Data Source Model for a GPRS Network Simulator", IEI/IEE Postgraduate Symposium Telecommunications Systems Research, Dublin, Nov. 2001.

[16] Z. Jiang, J. Chang and N.K. Shankaranarayanan, “Providing Multiple Service Classes For Bursty Data Traffic in Cellular Networks”, Electronic Edition, in Proc. of IEEE INFOCOM 2000, Vol. 3, pp. 1087-1096, 2000.

[17] C. Lindheimer, S. Mazur, J. Molnö, M. Waleij, “Third-generation TDMA.” Ericsson Review Vol.76 (2000): 2, pp. 68-79, 2000.

[18] S. Lu, V. Bharghavan, and R. Srikant, “Fair scheduling in wireless packet networks”, Proc. of ACM/SIGCOMM '97, pp. 63-74, 1997.

[19] A. Mehrotra, “GSM system engineering”, Artech House, 1997, pp. 25-120

[20] "Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS", ETSI UMTS 30.03 version 3.1.0, 1999, pp. 33-36.

[21] E. Nowicki and J. Murphy, "Resource allocation for interactive traffic class over GPRS network", Proc. of the IEE 17th UK Teletraffic Symposium, pp. 31/1-31/7, Dublin, Ireland, May 16-18, 2001.

[22] Q. Pang, A. Bigloo, V.C.M. Leung and C. Scholefield, “Service scheduling for General Packet Radio Service classes”, Proc. of IEEE Wireless Communications and Networking Conference WCNC’99, New Orleans, LA, 1999, pp. 2-5

[23] G. Stuber & B. Jabbari, “Multi-access, mobility & teletraffic for wireless communications”, Kluwer Academic, vol. 5, pp. 85-96, 2000.

[24] D. Todinca, P. Perry and J. Murphy, "Novel Prioritised EGPRS Medium Access Regime For Reduced File Transfer Delay During Congested Periods", Proc. of the IEE 3rd International Conference on 3G Mobile Communications Technologies, May 8-10, 2002, pp. 2-4.

[25] D. Todinca, P. Perry and J. Murphy, "Algorithms for Resource Allocation in Data Transfer over EGPRS Networks", 2nd European Conference on Universal Multiservice Networks (ECUMN'2002), CREF, Colmar, France, April 8-10, 2002, pp. 1-3.

[26] D. Todinca, P. Perry and J. Murphy, "Resource Allocation Algorithms for GPRS Network Performance Improvement During Congested Periods", IEI/IEE Postgraduate Symposium Telecommunications Systems Research, Dublin, Nov. 2001, pp. 2-4.

[27] Wireless Application Protocol Multimedia Messaging Service Architecture Overview Specification WAP-205-MMSArchOverview-20010425-a WAP Forum.

[28] Wireless Application Protocol Multimedia Messaging Service Client Transactions Specification WAP-206-MMSCTR-20020115-a WAP Forum.

[29] Wireless Application Protocol Architecture SpecificationWAP-210-WAPArch-20010712 WAP Forum, pp. 9-10.

[30] Xiaoxin Qiu, Justin Chuang, Kapil Chawla "Link Layer Techniques for Performance Enhancement for EDGE" AT&T Labs - Research Red Bank, NJ, USA, GPRS SEMINAR February 22, 1999 Core Auditorium, Busch Campus Rutgers University

Appendix A: Abbreviations and terminology

3G	Third Generation mobile telecommunication system
3GPP	3rd generation Partnership Project
8PSK	Eight Phase Shift Keying A new modulation technique that is used for EDGE enabling increased data rates.
AR	Access Router
BSC	Base Station Controller A high capacity switch that performs radio functionality and mobility management in the GPRS network.
BSS	Base Station Subsystem Subsystem that consists of BSC and BTS.
BTS	Base Transceiver System The radio equipment part of the BSS which includes the antenna.
C/I	Carrier to Interference ratio A measure of signal strength relative to interference level.
CN	Core Network
CS	Circuit Switched. Established circuit from one end-point to the other
CSD	Circuit Switched Data Circuit Switched data call between and connections.
EDGE	Enhanced Data rates for GSM Evolution An enhanced modulation technique of the GPRS-network that uses 8PSK.
EGGSN	Enhanced Gateway GPRS Support Node A node which is one of the new key functional elements of EGPRS.
ESGSN	Enhanced Serving GPRS Support Node A new functional element of the EGPRS network.
EGPRS	Enhanced General Packet Radio Service An enhanced modulation technique of the GPRS-network that uses 8PSK.
ETSI	European Telecommunications Standard Institute

	An Institute that brings together players from the telecommunication industry with the intention of putting forward new and improved wireless standards.
FDMA	Frequency Division Multiple Access Technique that divides the frequency spectrum into fixed channels.
GERAN	GSM/EDGE Radio Access Network A radio access network in EGPRS, which consists of base transceiver stations, and radio network controllers.
GGSN	Gateway GPRS Support Node A node which is one of the new key functional elements of GPRS.
GMSK	Gaussian Minimum Shift Keying Conventional modulation scheme used by GSM.
GSM	Global System for Mobile communication A digital mobile telephone system that is widely used in Europe and other parts of the world.
GSN	GPRS Support Node Two new components of the GPRS network, both SGSN and GGSN. Provides routing in the GPRS network.
HLR	Home Location Register A database that holds subscription information about every person who has bought a subscription from GSM/GPRS operator.
Hz	Hertz A unit for frequency.
IP	Internet Protocol A protocol used on the network layer.
LA	Link Adaptation A technique that is used for LQC
LLC	Logical Link Control A protocol layer between MS and SGSN.
LQC	Link Quality Control A mechanism that is used to dynamically adapt the modulation and code rate with respect to the current link quality
MAC	Medium Access Control Protocol in the radio level that is used to allocate the radio channel.

MCS	Modulation and Coding Scheme A certain level of data integrity check (error correction overhead) applied to data before it is transmitted over the radio-interface.
MM	Multimedia Message
MMS	Multimedia Messaging Services Messaging services used to send and receive multimedia messages contains graphics, sounds and texts in the GPRS network.
MMSC	Multimedia Messaging Services Centre A network element that is used for MMS messages exchanging.
MS	Mobile Station The phone or terminal of the end-user that consists of a MT and a TE.
MSC	Mobile Switching Centre A node that performs the switching functions of GSM CS system.
PCU	Packet Control Unit Functional element in BSS that handles upper level GPRS control in radio and resource allocation.
PhCH	Physical Channel
PDCH	Packet Data Channel Physical channel allocated for GPRS traffic.
PDN	Packet Data Network A network that carries user data in packets (e.g. Internet and X.25).
PDP	Packet Data Protocol The network protocol, e.g. IP or X.25 that is used by user.
PDU	Packet Data Unit One data packet on the protocol layer.
PS	Packet Switched Data sent as packets, which are routed in the network by address-headers. An alternative to circuit switched communication.
PUSH	A WAP 1.2 notification service used to start mobile originating connection based on SMS. Used for MMS notification.
QoS	Quality of Service A definition of service class on the connection between entities.
RAN	Radio Access Network

	A radio access network in EGPRS, which consists of base transceiver stations, and radio network controllers.
RLC	Radio Link Control
	A protocol between MS and BSS that handles retransmission and other radio related issues.
RNC	Radio Network Controller
	A controller in EGPRS network, which is responsible for the radio resource management, including handover decisions.
SGSN	Serving GPRS Support Node
	A new functional element of the GPRS network.
SMS	Short Message Service
	A service that make it possible to send or receive messages.
SMSC	Short Message Service Centre
	A central computer that handles short messages.
SNDCP	SubNetwork Dependence Convergence Protocol
	A protocol used between MS and SGSN.
TBF	Temporary Block Flow
	A physical connection used to support the transfer of a number of blocks and identified by TFI.
TCP	Transmission Control Protocol
	Protocol layer on top of conventional IP protocol used to control the end-to-end connection.
TDMA	Time Division Multiple Access
	A technique that divides each channel into time slots in order to increase the amount of data that can be carried.
TFI	Traffic Flow Identifier
	Identifier that uniquely identifies the traffic flow associated to each PDP context.
TRX	Transceiver Unit
	An antenna unit in the BTS.
TTFB	Time to First Bit
	A transmission delay calculated as the time that is needed to receive the first bit of data.

VLR	Visitor Location Register A database that contains information about all the MSs currently located in the MSC location area or the SGSN routing areas.
WAP	Wireless Application Protocol A specification for a set of communications protocols that standardise the way wireless devices can be used for Internet access.
WML	Wireless Mark-up Language A language used to write the WAP services.
WWW	World Wide Web All the resources and users on the Internet that are using the Hypertext Transfer Protocol.