

Adaptive Bandwidth Reservation and Scheduling
for Efficient Telemedicine Traffic Transmission over
Wireless Cellular Networks

Adaptive Bandwidth Reservation and Scheduling
for Efficient Telemedicine Traffic Transmission over
Wireless Cellular Networks

BY

LU QIAO, B. ENG.

BEIJING JIAOTONG UNIVERSITY

AUGUST 2008

A THESIS

SUBMITTED TO THE SCHOOL OF ELECTRICAL AND COMPUTER

ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILLMENT TO THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

©Copyright 2008 by Lu Qiao, B. Eng.

Beijing Jiaotong University

All Rights Reserved

MASTER OF APPLIED SCIENCE (2008)
(Electrical and Computer Engineering)

McMaster University
Hamilton, Ontario

TITLE: Adaptive Bandwidth Reservation and Scheduling for Efficient Telemedicine Traffic Transmission over Wireless Cellular Networks

AUTHOR: Lu Qiao, B. Eng. (Beijing Jiaotong University)

SUPERVISOR: Dr. Polychronis Koutsakis, Assistant Professor, ECE Dept., McMaster University

NUMBER OF PAGES: ix, 53

Abstract

Telemedicine traffic transmission over wireless cellular networks has gained in importance during the last few years. Most of the current research in the field has focused on software and hardware implementations for telemedicine transmission, without discussing the case of simultaneous transmission of both urgent telemedicine traffic and regular multimedia traffic over the network.

Due to the fact that telemedicine traffic carries critical information regarding the patients' condition, it is vitally important that this traffic has highest transmission priority in comparison to all other types of traffic in the cellular network. However, the need for expedited and correct transmission of telemedicine traffic calls for a guaranteed bandwidth to telemedicine users. This creates a tradeoff between the satisfaction of the very strict Quality of Service (QoS) requirements of telemedicine traffic and the loss of the guaranteed bandwidth in the numerous cases when it is left unused, due to the infrequent nature of telemedicine traffic. This waste of the bandwidth may lead to a lack of sufficient bandwidth for regular traffic, hence degrading its QoS.

To resolve this complex problem, in this thesis, we propose a) an adaptive bandwidth reservation scheme based on road map information and on users' mobility, and b) a fair scheduling scheme for video traffic transmission over wireless cellular networks. The proposed combination of the two schemes, which is evaluated over a hexagonal cellular structure, is shown to achieve high channel bandwidth utilization while offering full priority to telemedicine traffic.

Acknowledgement

In the first place I wish to express my deep appreciation to my supervisor, Dr. Polychronis Koutsakis, for his expert supervision, guidance and encouragement throughout the course of this thesis, which have been inspiring to my growth as a student and have been illuminating the way of my independent research.

I would also like to express my sincere thanks to Dr. Terence D. Todd and Dr. Dongmei Zhao, for their insightful advice on the improvement of this work. I wish to thank all my colleagues here in the Wireless Networking Group at McMaster University, especially Jun Zou, Bin Wang, Amir Sayegh, Mohammad Smadi and Ahmed Kholaiif, for their friendship and support during occasional stressful times.

I am also deeply indebted to my parents and all other family members for their understanding, love and unceasing faith in me along the way.

Contents

Abstract	iii
Acknowledgement	iv
List of Figures	vii
List of Tables	viii
List of Abbreviations	ix
1 Introduction	1
1.1 Overview	1
1.2 Motivation and Contribution of this work	2
1.3 Organization of the Thesis	4
2 Traffic Types and Models	5
2.1 Telemedicine Traffic	5
2.2 Regular Multimedia Traffic	6
3 Multimedia Integration Multiple Access Control (MI-MAC)	11
3.1 Channel Frame Structure	11
3.2 Base Station Scheduling and Actions of Terminals	13
3.3 Two-Cell Stack algorithms	15
3.3.1 Two-cell stack reservation random access algorithm	15

3.3.2	Two-cell stack blocked access collision resolution algorithm . . .	16
3.4	System state transitions	16
3.5	Channel Error Model	17
4	The Proposed Scheduling Scheme	20
4.1	Introduction of the new scheduling ideas	20
4.2	Contention resolution	23
4.3	Fair Scheduling	25
5	Adaptive Bandwidth Reservation based on Mobility and Road in-	
	formation	28
5.1	Overview	28
5.2	Network and Mobility Models	29
5.3	Adaptive Bandwidth Reservation Scheme	31
6	Results and Discussion	36
6.1	Simulation Setup	36
6.2	Results	37
6.3	Maximum Achievable Throughput	43
7	Conclusions and Future work	46
	Bibliography	48

List of Figures

2.1	The voice source activity model	7
3.1	Channel Frame Structure	12
3.2	State transition diagram for an active terminal	17
3.3	Channel Error Model	18
5.1	Road Map and cellular network model	30
6.1	Effect of regular video traffic on X-Ray traffic.	37
6.2	Percentage of regular video users who experience packet loss larger than 1%, vs. the number of regular video users.	38
6.3	Regular video packet dropping vs. the number of regular video users.	39
6.4	Effect of regular voice traffic on telemedicine video traffic.	40
6.5	Effect of regular voice traffic on telemedicine image traffic.	41
6.6	Effect of telemedicine Video traffic on regular Video traffic, with 10% handoff traffic	42
6.7	Improvement on voice packet dropping probability with the use of adaptive bandwidth reservation.	43

List of Tables

3.1	Channel Error Parameters	18
6.1	Maximum Achievable Throughput	44

List of Abbreviations

3G	Third Generation Communication Systems
4G	Fourth Generation Communication Systems
BS	Base Station
CBR	Constant Bit Rate
CRP	Collision Resolution Period
ECG	Electro-Cardiogram
FCFS	First Come First Served
GSM	Global System for Mobile communications
GPS	Global Positioning System
LAN	Local-area Network
MI-MAC	Multimedia Integrated - Medium Access Control Protocol
MSC	Mobile Switching Center
QoS	Quality of Services
SNR	Signal-to-noise ratio
TDMA-FDD	Time Division Multiple Access with Frequency Division Duplex
VAD	Voice Activity Detector
VBR	Variable Bit Rate
VF	Video Frame
WCDMA	Wideband Code Division Multiple Access

Chapter 1

Introduction

1.1 Overview

The term of telemedicine has been defined as using telecommunications to provide medical information and services [1]. Telemedicine is already being employed in many areas of healthcare, such as intensive neonatology, critical surgery, pharmacy, public health and patient education. The ultimate goal for all telemedicine applications is to improve the well-being of patients and bring medical expertise fast and at low cost to people in need [2, 3]. Currently, in the cases of motor vehicle accidents, which are the number one cause of violent death in the United States [4], prehospital teams provide on-scene initial assessment and resuscitation and transmit this information to a physician mainly via voice communication; therefore the physician can make an assessment based on what is described and continuously monitor an injured victim through visual communication (e.g., video) while at the same time receiving data of major importance regarding the victim's vital signs [5, 6]. In addition to ambulance vehicles, it is also of critical importance for the provision of health care services at understaffed areas like rural health centers, ships, trains, airplanes, as well as home monitoring [2, 7].

Mobile networks have brought about new possibilities in the field of telemedicine

due to the wide coverage provided by cellular networks and their capability of providing service to moving vehicles. Some of the related research projects include the data query mechanism proposed in [8], which takes advantage of the low cost mobile sensor networks and 3G cellular networks, the mobile telemedicine system designed in [9] over a wireless LAN, and the schemes designed in [10] for emergency QoS support over WLAN and Body Sensor Networks. In addition, various applications have been introduced, for example: the mobile tele-echography robotic application over WCDMA [11], the mobile Tele-Ultrasonography in M-health over 3G networks [12], and the Mobile Teletrauma System presented in [5] using CDMA over 3G.

As pointed out in various telemedicine research efforts during the last few years [5, 6, 7, 12, 13, 14, 15, 16], beyond 3G wireless networks can be a sufficient test bed for the development of efficient telemedicine traffic transmission mechanisms, due to the much higher channel rate they are expected to provide in comparison to previous generation's cellular networks.

4G, an acronym for Fourth-Generation Communications System, is a term used to describe the next step in wireless communications. A 4G system will be able to provide voice, data and streamed multimedia to users on an "Anytime, Anywhere" basis. Although there is no formal definition for what 4G is, its commonly assumed objective is that it will be a fully IP-based integrated system. This will be achieved after wired and wireless technologies converge and will be capable of providing very high data rates both indoors and outdoors, with premium quality and high security. 4G will offer all types of services at an affordable cost.

1.2 Motivation and Contribution of this work

Mobile healthcare (M-health, "mobile computing, medical sensor and communication technologies for healthcare" [12]) is a new paradigm that brings together the evolution of emerging wireless communications and network technologies with the concept of

“connected healthcare” anytime and anywhere. Various M-health studies have been conducted within the last few years, on very significant aspects of public health [5, 6, 7, 12, 13, 14, 15, 16]. In many of these studies, the efficient use of the cellular network resources was of paramount importance for the correct and rapid transmission of all types of telemedicine traffic (video, audio and data).

One common characteristic of all the above referenced studies is that they focus solely on the transmission of telemedicine traffic over the cellular network, without taking into account the fact that regular traffic has strict Quality of Service (QoS) requirements as well. Also, in many of these studies, despite the importance of the systems used, the accuracy in information transmission was low. For example, in [15] a high percentage (27%) of electrocardiogram (ECG) data transmission interruptions took place due to GSM channel congestion. For the above reasons, one of the main tasks of current research on the subject is the design of system hardware and software to implement a mobile telemedicine system capable of transmitting, with high QoS, significant loads of multiplexed information in next generation wireless cellular networks [5].

The work presented in this thesis is, to the best of our knowledge, the first in the literature to focus on the efficient integration and transmission scheduling of all major types of telemedicine applications with other types of wireless traffic over a high capacity cellular network.

Previous work [17] has introduced MI-MAC (Multimedia Integration Multiple Access Control Protocol), which was shown to achieve superior performance in comparison to other TDMA and WCDMA-based protocols of the literature when integrating various types of multimedia traffic (video, voice, email and web data) over next generation cellular networks.

In MI-MAC, as well as in most relevant works on MAC protocols in the literature, within each priority class the queuing discipline is assumed to be First Come First Served (FCFS). Hence the average performance evaluation metrics will give no insight

on the QoS of each individual wireless subscriber; therefore, it could be the case that certain users have their QoS severely violated while others get exceptional QoS, which would give an acceptable average QoS over the total number of users. This approach is unfair to users who arrive later in the network and hence are placed at the bottom of the Base Station (BS) service queue; the problem is especially significant in the case of video users, where early arriving users may dominate the channel by being allocated large numbers of slots, allowing just a small number of resources to be available for users arriving later. For this reason, we introduce a fair bandwidth allocation scheme in MI-MAC, as well as a number of scheduling ideas in order to improve the protocol's performance.

Furthermore, we propose a new adaptive bandwidth reservation scheme, based on road map information and on user mobility in a hexagonal cellular architecture, in order to guarantee the required QoS to all types of wireless traffic, with highest priority offered to mobile telemedicine traffic over regular traffic (individual priorities are also set among the various types of mobile telemedicine traffic, based on their current importance for medical health care).

1.3 Organization of the Thesis

The rest of this thesis is organized as follows. Chapter 2 gives a basic description of the telemedicine and regular traffic types and models used in our work. In Chapter 3, we provide a brief overview of MI-MAC protocol which is used as the basis of our work. Chapter 4 presents our proposed improved and fair scheduling scheme which we add on MI-MAC. In Chapter 5 we discuss our proposal for an adaptive bandwidth reservation scheme, which enables the system to exploit its knowledge of the users' mobility patterns. Chapter 6 includes our simulation results and a discussion on them, and in Chapter 7 we present our conclusions and ideas for future work.

Chapter 2

Traffic Types and Models

2.1 Telemedicine Traffic

Four types of telemedicine traffic are considered in our work: Electro-Cardiograph (ECG), X-ray, video and high-resolution medical still images.

- *Electro-Cardiograph (ECG)*: Similarly to [5, 16], which use data from the MIT-BIH arrhythmia database, we consider that ECG data is sampled at 360 Hz with 11 bits/sample precision. The arrival rate of ECG users is set to be λ_E user/frame following a Poisson distribution. The transmission of ECG traffic should be rapid and lossless, due to the critical nature of the data; additionally, we have set a strict upper bound of just 1 channel frame (12 ms) for the transmission delay of an ECG packet.
- *X-Ray*: We consider that a typical X-ray file size is 200 Kbytes [7] and that the aggregate X-Ray file arrivals are poisson distributed with mean λ_X files/frame. The upper bound for the transmission delay of an X-Ray file, which again needs to be lossless, is set to 1 minute. A discussion on the strictness of this bound will be made in Chapter 4.
- *Medical Images*: Medical image files have sizes ranging from 15 to 20 Kbytes/image

[5] and are Poisson distributed with mean λ_l files/frame. The upper bound for the transmission delay of an medical image is set to 5 seconds (this bound is much stricter than those used in [5, 7, 14, 16]), and the transmission needs to be lossless.

- *Telemedicine Video*: Since H.263 is the most widely used video encoding scheme for telemedicine video today, we use in our simulations real H.263 video-conference traces from [18] with mean bit rate of 91 Kbps, peak rate of 500 Kbps and standard deviation of 32.7 Kbps. The video frames arrive with constant rate (every 80 ms) with variable frame sizes. We have set the maximum transmission delay for video packets to 80 ms, with packets being dropped when this deadline is reached; i.e., all packets of a video frame must be delivered before the next video frame arrives. Due to the need for very high-quality telemedicine video, the maximum allowed video packet dropping probability is set to 0.01%.

It needs to be mentioned that existing work in the field assumes much looser QoS requirements in order for the network to be able to meet them. For example, in [6] the upper delay bound for ECG packet is 1 second, and the upper bound on voice dropping is set to 3%.

2.2 Regular Multimedia Traffic

Four types of “regular” multimedia traffic were integrated in the work in [17]: MPEG-4 video-conference, voice, email and web traffic, i.e., the most common traffic types in the 3G/B3G/4G wireless cellular networks. The same traffic types are taken into account in this work in order to study the integration of regular traffic with urgent telemedicine traffic.

- *Voice*: The speech codec rate is 32 kb/s, and voice terminals are equipped with a voice activity detector (VAD). Voice sources follow an alternating pattern of

talkspurts and silence periods (on and off), and the output of the voice activity detector is modeled by a two-state discrete time Markov chain (Figure 2.1). The mean talkspurt duration is 1 s and the mean silence duration is 1.35 s. The talkspurt to silence transition probability is p_{TS} , and the silence to talkspurt transition probability is p_{ST} . The talkspurt and silence periods are geometrically distributed with mean $1/p_{TS}$ and $1/p_{ST}$ frames, respectively. Therefore, at steady state, the probability that a terminal is in talkspurt (speech activity), p_T , or silence, p_S , is obtained from the following equations:

$$p_T = \frac{p_{ST}}{p_{ST} + p_{TS}} \quad (2.1)$$

$$p_S = 1 - p_T \quad (2.2)$$

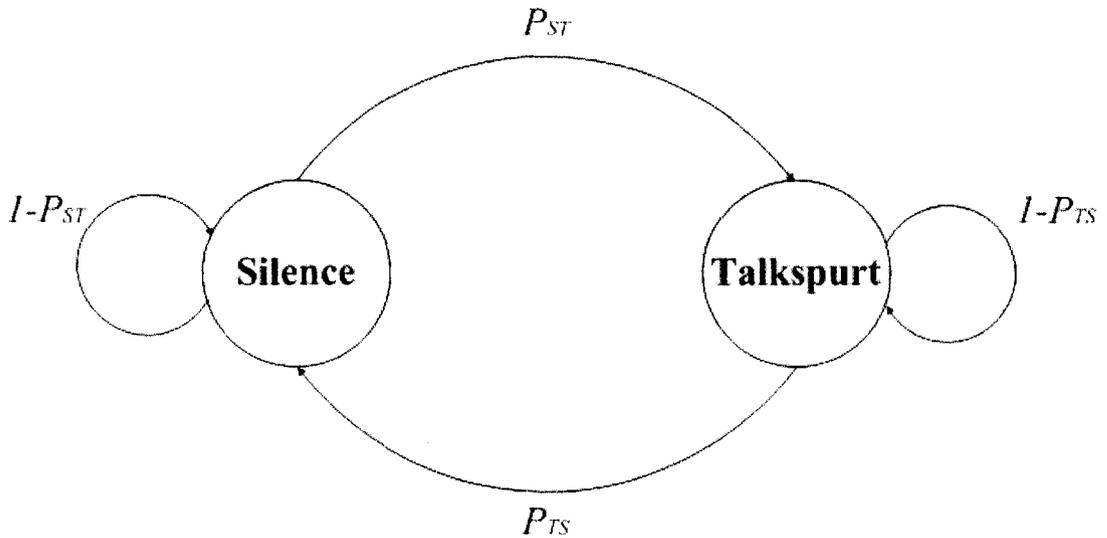


Figure 2.1: The voice source activity model

The number of active voice terminals N in the system is assumed to be constant over the period of interest. All of the voice source transitions (e.g., talk to silence) occur at the frame boundaries. This assumption is reasonably accurate, taking into consideration that the duration of a frame is equal to 12 ms here,

while the average duration of the talkspurt and silence periods exceeds 1 s. Reserved slots are deallocated immediately. This implies that a voice terminal holding a reservation signals the BS upon the completion of its talkspurt (the same assumption is made for slots reserved by data and video terminals). The allowed voice packet dropping probability is set to 0.01, and the maximum transmission delay for voice packets is set to 40 ms.

- *Email*: We adopt the data traffic model based on statistics collected on e-mail usage from the Finnish University and Research Network (FUNET) [19]. The probability distribution function $f(x)$ for the length of the e-mail data messages of this model was found to be well approximated by the Cauchy (0.8, 1) distribution. The packet interarrival time distribution for the FUNET model is exponential, and the average e-mail data message length is 80 packets. A quite strict (considering the nature of this type of traffic) upper bound is set on the average email transmission delay, equal to 5 s. The reason for this strict bound is that mobile users sending emails will be quite demanding in their QoS requirements, as they will expect service times similar to those of short message service traffic.
- *Web*: We adopt the http traffic model from [20], according to which the distributions of the random variables concerning the composition of web requests are the following:
 1. size of web request: lognormal (5.84, 0.29), with mean = 360 bytes and standard deviation = 106.5 bytes.
 2. number of web requests per www session: lognormal (1.8, 1.68), with mean = 25 pages and standard deviation = 100 pages.
 3. web request viewing time: weibull (α, β), truncated at a maximum of 15 min (if the viewing time is longer than 15 min, a new session will follow), with mean = 39.5 s and standard deviation = 92.6 s.

Other relevant random variables include the size of the main object of the requested http session, the number of inline objects, and the size of the inline objects. However, as we study the uplink channel in this work (i.e., only the web requests sent from the mobile terminals to the BS), only the distributions of the random variables presented above are needed for our model.

The arrival process of www sessions is chosen to be Poisson with rate λ_{web} sessions per second, with an upper bound on the average web request transmission delay equal to 3 s. Given that the average size of a web request is 360 bytes, i.e., 7.5 packets (less than a tenth of the average e-mail message size), it is clear that we consider web traffic to be the most delay-tolerant. Still, this upper bound is again strict, as in the case of e-mail traffic, and the reason for this choice is to test system performance when incorporating users with very demanding QoS.

- *Regular MPEG-4 Video Streams:* The MPEG initiated the new MPEG-4 standards in 1993 with the goal of developing algorithms and tools for high efficiency coding and representation of audio and video data to meet the challenges of video conferencing applications. The standards were initially restricted to low bit rate applications but were subsequently expanded to include a wider range of multimedia applications and bit rates. The most important addition to the standards was the ability to represent a scene as a set of audiovisual objects. The MPEG-4 standards differ from the MPEG-1 and MPEG-2 standards in that they are not optimized for a particular application but integrate the encoding, multiplexing, and presentation tools required to support a wide range of multimedia information and applications. In addition to providing efficient audio and video encoding, the MPEG-4 standards include such features as the ability to represent audio, video, images, graphics, text, etc. as separate objects, and the ability to multiplex and synchronize these objects to form scenes. Support is also included for error resilience over wireless links, the coding of arbitrarily shaped video objects, and content-based interactivity such as the

ability to randomly access and manipulate objects in a video scene. In our study, we use the trace statistics of actual MPEG-4 streams from the publicly available library of frame size traces of long MPEG-4 and H.263 encoded videos provided by the Telecommunication Networks Group at the Technical University of Berlin[18]. The two video streams used in our study have been extracted and analyzed from a camera showing the events happening within an office and a camera showing a lecture, respectively. We have used the high quality version of the videos: one has a mean bit rate of 400 kb/s, a peak rate of 2Mb/s, and a standard deviation of 434 kb/s, and the other one has a mean rate of 210 kbps, peak rate of 1.5 Mbps and standard derivation of 182 kbps. New video frames (VFs) arrive every 40 ms. We have set the maximum transmission delay for video packets to 40 ms, with packets being dropped when this deadline is reached. That is, all video packets of a VF must be delivered before the next VF arrives. The allowed video packet dropping probability is set to 1% [21], as the loss of regular video packets is not of equally critical importance as that of telemedicine video packets.

Chapter 3

Multimedia Integration Multiple Access Control (MI-MAC)

The Multimedia Integration Multiple Access Control (MI-MAC) protocol, introduced in [17] and based on Time Division Multiple Access with Frequency Division Duplex (TDMA-FDD), is one of the first works in the relevant literature for wireless picocellular networks that efficiently integrates voice (Constant Bit Rate, CBR, On/Off Traffic), bursty email, and web traffic with either MPEG-4 or H.263 video streams (Variable Bit Rate, VBR) in high capacity picocellular systems with burst-error characteristics. The protocol was shown to be a good candidate for next generation cellular networks, as it outperformed (in simulation results and conceptually) other TDMA and Wideband Code Division Multiple Access (WCDMA)-based protocols when evaluated over a wireless channel with burst-error characteristics [17].

3.1 Channel Frame Structure

The uplink channel time is divided into time frames of fixed length. The frame duration is selected such that a voice terminal in talkspurt generates exactly one packet per frame (packet size is considered to be equal to 53 bytes, 48 of which contain

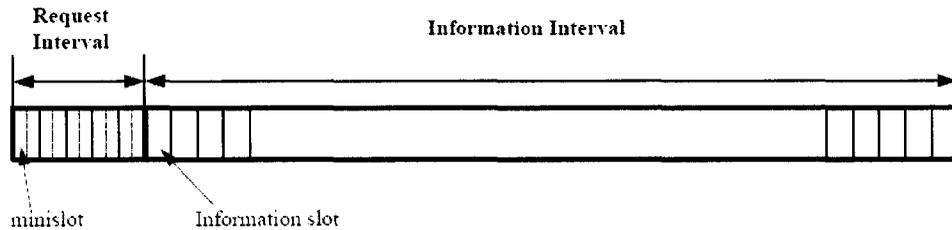


Figure 3.1: Channel Frame Structure

information; i.e., the packet size is equivalent to the ATM cell size. This choice was made for reasons of results comparison with other protocols of the literature). As shown in Fig. 3.1, which presents the channel frame structure, each frame consists of two types of intervals. These are the request interval and the information interval.

By using more than one minislot per request slot, a more efficient usage of the available request bandwidth (in which users contend for channel access) is possible. In [17], which considered a lower capacity channel, a request slot was subdivided into 4 minislots. In the present work, which considers a 20 Mbps channel, we choose the number of minislots per request slot to be equal to 2, to allow for guard time and synchronization overheads, for the transmission of a generic request packet and for the propagation delay within the picocell. Each minislot accommodates exactly one fixed-length request packet. Within an information interval, each slot accommodates exactly one fixed-length packet that contains voice, video, or data information and a header. Any free information slot of the current channel frame can be temporarily used as an extra request (ER) slot to resolve the contention between requesting users. ER slots are again subdivided into two minislots. The function and operation of ER slots are exactly the same as of the request slots in the request interval.

3.2 Base Station Scheduling and Actions of Terminals

Terminals with packets, and no reservation, contend for channel resources using a random access protocol to transmit their request packets only during the request intervals. The Base Station broadcasts a short binary feedback packet at the end of each minislot, indicating only the presence or absence of a collision within the minislot [collision (C) versus noncollision (NC)]. Upon successfully transmitting a request packet the terminal waits until the end of the corresponding request interval to learn of its reservation slot (or slots). If unsuccessful within the request intervals of the current frame, the terminal attempts again in the request intervals of the next frame. A terminal with a reservation transmits freely within its reserved slot.

To resolve contention among all requesting users, different priorities were assigned to different types of users. The four types of regular traffic studied in [17] follow the priority: video, voice, email, and web. The above prioritization by isolating each type of traffic and letting it contend only with traffic of the same type is feasible due to the use of the two-cell stack reservation random access algorithm (by video and voice terminals) and the two-cell stack blocked access collision resolution algorithm [22] (by email and web terminals) to resolve contention (this algorithm is of window type, with FCFS-like service, and will be discussed in Section 3.3). Apart from their operational simplicity, stability and relatively high throughput when compared to the PRMA (Aloha-based) [23] and PRMA-like algorithms [24]), the stack algorithms have the additional advantage of offering a clear indication of when contention has ended (for two-cell stack this happens when two consecutive non-collision signals are transmitted by the BS in the downlink).

To allocate channel resources, the Base Station maintains a dynamic table of the active terminals within the picocell. Upon successful receipt of a voice or data request packet, the Base Station provides an acknowledgment and queues the request. The

BS allocates channel resources at the end of the corresponding request interval.

Specifically, for a video terminal, if a full allocation is possible, which means that the number of idle information slots is larger than the number of requested slots, the BS assigns this user all its requested slots. If a full allocation is not possible, the BS grants to the video user as many of the requested slots as possible (partial allocation).

For email and web users, the BS allocates one slot per frame for each user.

Voice terminals that have successfully transmitted their request packets do not acquire all the available information slots in the frame. If this happened, voice terminals would keep their dedicated slots for the whole duration of their talkspurt (on average, more than 80 channel frames), and thus video terminals would not find enough slots to transmit in; hence, the particularly strict video QoS requirements would be violated. Consequently, the BS allocates a slot to each requesting voice terminal with a probability p^* . The probability p^* for the allocation of slots to voice users varied according to the video load. In this study, a near-optimal value of p^* (0.09) has been found through extensive simulations, which works well for all video loads examined. The requests of voice terminals which “fail” to acquire a slot, based on the above BS slot allocation policy, remain queued. The same holds for the case when the resources needed to satisfy a voice request are unavailable. Within each priority class, the queuing discipline is assumed to be FCFS.

In addition, the BS “preempts” email and web reservations in order to service video and voice requests. Thus, whenever new video or voice requests are received and every slot within the frame is reserved, the BS attempts to service them by canceling the appropriate number of reservations belonging to data (email/web) terminals (if any). When data reservations are canceled, the BS notifies the affected data terminal and places an appropriate request at the front of the corresponding request queue.

3.3 Two-Cell Stack algorithms

As mentioned in Section 3.2, the contention among terminals within each type of traffic is resolved by the 2-cell stack algorithm [22]. This section provides a brief discussion of the two-cell stack reservation random access algorithm (used by video and voice terminals) and the two-cell stack blocked access collision resolution algorithm (used by email and web terminals).

3.3.1 Two-cell stack reservation random access algorithm

Each terminal uses a counter, r , as follows.

1. At the start of every request interval the contending terminals initialize their counter, r , to 0 or 1 with probability $1/2$.
2. Contending terminals with $r = 0$ transmit into the first request slot. With x being the feedback for that transmission, the transitions depending on r are as follows:
 - if $x = \text{non-collision}$:
 - if $r = 0$, the request packet was transmitted successfully.
 - if $r = 1$, then $r = 0$.
 - if $x = \text{collision}$:
 - if $r = 0$, then reinitialize r to 0 or 1 each with probability $1/2$.
 - if $r = 1$, then $r = 1$.
3. Repeat step 2, until either two consecutive feedbacks indicating non-collision occur or the request interval ends.

The operation of this protocol can be depicted by a two-cell stack, where in a given request mini-slot the bottom cell contains the transmitting terminals (those with r

= 0), and the top cell contains the withholding terminals (those with $r = 1$). The especially attractive feature of this algorithm is that two consecutive non-collisions indicate that the stack is empty.

3.3.2 Two-cell stack blocked access collision resolution algorithm

To transmit data request packets, the data terminals follow the two-cell stack random blocked access collision resolution algorithm during consecutive data request intervals, due to its operational simplicity, stability and relatively high throughput.

A blocked access mechanism is established by the following first time transmission rule for newly generated data messages. Terminals with new message arrivals may not transmit during a collision resolution period (CRP). A CRP is defined as the interval of time that begins with an initial collision (if any) and ends with the successful transmission of all data request packets involved in that collision (or, if no collision occurred, ends with that mini-slot). In the first mini-slot following a CRP, all of the terminals whose message arrived within a prescribed allocation interval, of maximum length D , transmit with probability one. Terminals involved in a collision follow rules 2 ~ 3 in section 3.3.1 and the conclusion of the CRP is identified by two consecutive feedbacks indicating non-collision.

3.4 System state transitions

As shown in figure 3.2, an active terminal is described as being in one of four states: silent, contender, queued, or reserved. A silent terminal has no packet to transmit and does not require channel resources. Once the terminal has information to transmit, it enters the contender state and remains there until it either successfully transmits a request packet or drops all of its packets (in the case of video and voice terminals). Since the requests are queued at the BS, the terminal enters the queued state and

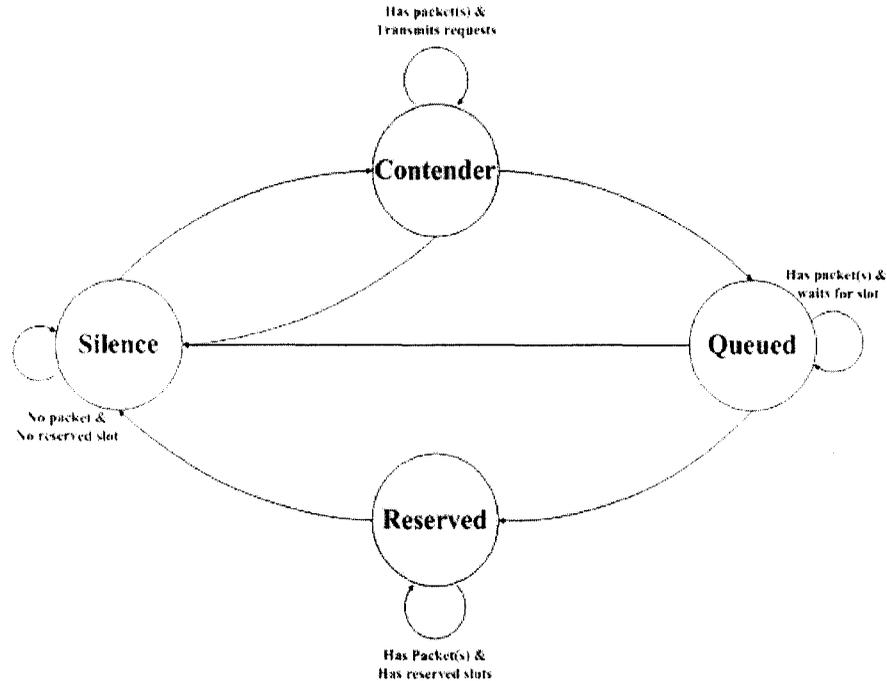


Figure 3.2: State transition diagram for an active terminal

remains there until it either receives a reservation or exits talkspurt. After receiving a reservation, the terminal enters the reserved state and transmits one (or more, in the case of video terminals) packet(s) per frame into its allotted slot(s) until it exhausts its packets and returns to the silent state.

3.5 Channel Error Model

The most widely adopted wireless channel error model in the literature is the Gilbert-Elliot model [25, 26]. The Gilbert-Elliot model is a two-state Markov model where the channel switches between a “good state” (always error-free) and a “bad state” (error-prone). However, many recent studies have shown that the Gilbert-Elliot model fails to predict performance measures depending on longer-term correlation of errors [27], minimizes channel capacity [28], and leads to a highly conservative allocation

strategy [29].

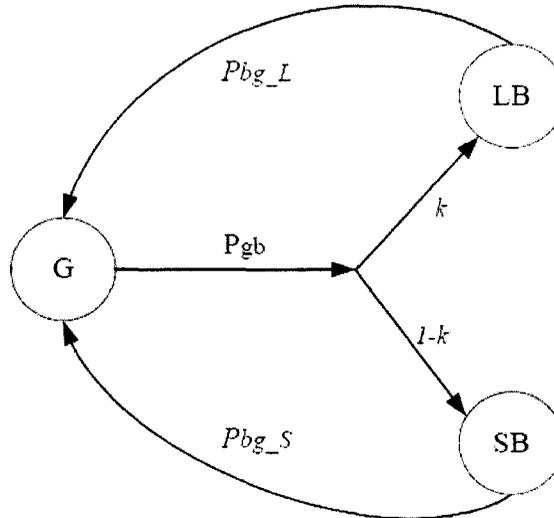


Figure 3.3: Channel Error Model

Table 3.1: Channel Error Parameters

$P_{good} = 0.99992$
$B_G = 1/P_{gb} = 65160$ slots
$B_{SB} = 1/P_{bg_S} = 2.38$ slots
$B_{LB} = 1/P_{bg_L} = 59.53$ slots
$k = 0.05$

A better choice for a more robust error model for wireless channels is the model presented in [30], which was adopted in [17] and which we also adopt in our study. This model, with the use of the short and long error bursts, makes more accurate predictions of the long-term correlation of wireless channel errors than the Gilbert-Elliot model. The error model consists of a three-state discrete-time Markov chain, where one state is the “good state” (error-free) and the other two states are the “bad states” the long bad and the short bad state (the Markov chain is shown in Fig 3.3).

A transmission is successful only if the channel is in the “good state” (G); otherwise, it fails. The difference between the long bad (LB) and short bad (SB) states is the time correlation of errors: LB corresponds to long bursts of errors, SB to short ones.

The parameters of the error model are presented in Table 3.1. The average number of error bursts, in slots, experienced when the states LB and SB are entered, are, respectively, given by $B_{LB} = 1/P_{bg,L}$ and $B_{SB} = 1/P_{bg,S}$, where $P_{bg,S}$ is the transition probability from state SB to G, and $P_{bg,L}$ is the transition probability from state LB to G. Similarly, the average number of consecutive error-free slots is given by $B_G = 1/P_{gb}$, where P_{gb} is the probability to leave state G. The parameter k is the probability that the Markov chain moves to state LB, given that it leaves state G; k also represents the probability that an error burst is long (i.e., the fraction of long bursts over the total number of error bursts).

Similarly to [17], we have chosen in our study the value of the probability P_{bad} , i.e., the steady-state probability that the channel is in a bad state, to be equal to $8 * 10^{-5}$; this value has been chosen in order to test an “almost worst” case scenario for our system, as the telemedicine video packet dropping probability is set to 0.01% and, by choosing a value of bad state probability larger than the upper bound on telemedicine video packet dropping, the strict QoS requirement of video users will certainly be violated. The values for P_{gb} and for the parameter k have been taken from [30], as well as the ratio between $P_{bg,S}$ and $P_{bg,L}$. The value for $P_{bg,L}$ is derived from the steady-state behavior of the Markov chain, for the bad state probability chosen.

Chapter 4

The Proposed Scheduling Scheme

4.1 Introduction of the new scheduling ideas

As explained in Chapter 3, we use the work on the MI-MAC protocol [17] as a basis for our scheduling scheme, and we significantly extend it to focus on the efficient scheduling for transmissions from mobile telemedicine users.

Certain design limitations had been adopted in the protocol's study in order to facilitate its comparison with other protocols in the literature:

1. Since the protocol was evaluated over one cell of the network, no traffic was considered to be arriving from other cells (handoff traffic).
2. Since video sources were assumed to “live” permanently in the system they did not have to contend for channel resources.
3. Since a picocellular wireless cellular architecture was assumed, the assumption was made that all users perceived the same uplink channel condition.

In order to evaluate the protocol's performance when integrating telemedicine traffic with “regular” traffic in a realistic wireless cellular network scenario, these assumptions need to be waived. More specifically, in our work the following respective additions/changes have been made to the wireless scenario which was studied in [17]:

1. A portion of the traffic arriving in each cell is handoff traffic from the other cells in our network model. Handoff traffic is treated with full priority, with the use of the adaptive bandwidth reservation scheme which will be presented in Chapter 5.
2. Video sources do not “live” permanently in the system, but have exponentially distributed sessions with a mean duration of five minutes [31]. This “relieves” a burden from the information interval of the channel, as video terminals occasionally leave the information interval, but adds a significant burden to the request interval, which has to compensate for the increase in contention as video users attempt to regain channel access. Similarly to previous works in the literature (e.g., [32]) we have found that a small percentage of the bandwidth suffices to be used for requests. This percentage is 4.4% in our work (25 slots used for requests out of the 566 slots of the channel frame); this value has been found via extensive simulations to provide a good tradeoff between allowing sufficient bandwidth for terminals to transmit their requests and allowing a large enough number of slots for terminals with a reservation to transmit their information packets.
3. In reality, however small the picocell radius, the channel fading experienced by each user is different, since users are moving independently of each other; therefore, in the present work *fading per user channel* is considered. As explained in Chapter 3, we adopt the robust three-state (good state, short bad state, long bad state) error model for wireless channels presented in [30] (Fig. 3.3) and we introduce the idea that the system *should take advantage of* the “problem” created when a regular video user experiences a “long bad” channel state (error burst) and is unable to transmit in its allocated uplink slots; this would normally lead to the dropping of the video packets scheduled to be transmitted in these slots, and consequently to higher average video packet dropping prob-

ability and the system's failure to satisfy the very strict QoS requirements of real-time videoconference traffic. Our new proposed mechanism aims at allocating as many of these slots as possible to other video terminals awaiting for packet transmission, in order to decrease their transmission delay. Although conceptually simple, the above approach is not equally simple to implement. The quality of each user's channel can be indicated by the signal-to-noise ratio (SNR) function; as shown in [33], in a FDD system (such as ours, which is a TDMA-FDD one) using pilot symbols that are inserted in the downlink with a certain time-frequency pattern, the mobile terminals can effectively estimate their SNR function and send it to the BS, which can then make its scheduling decisions based on all the collected cross-layer information from the terminals. This process, however, introduces both errors and delays in the estimates. Due to the random nature of the channel, it is impossible for the BS to precisely determine the state of the channel. The best estimate a BS can provide is a probability distribution over the possible channel states [34], which is our assumption in this work, i.e., that the probabilities of the Markov-chain error model have been derived with the above procedure.

Still, the BS cannot know with certainty the type of channel state transition that takes place for a mobile terminal when it leaves the good state, i.e., if the terminal's channel has entered the SB state or LB state. Therefore, the BS can only *make an estimation of* each mobile video terminal's channel conditions, by monitoring the slots allocated to the terminal and checking whether the terminal is transmitting in them or not. If the total number of a terminal's failed transmissions within its allocated slots surpass a given threshold, the BS in our scheme deduces that the terminal is in LB state, as the probability that it is in SB is very small given the high number of corrupted transmissions. Based on the channel error model it is easy to confirm by both analysis and simulation that the probability that a mobile terminal's channel is in SB when

more than 6 slots have been wasted is 6.55%; hence we have set the threshold to be 6 subsequent transmission failures (choosing a higher threshold would result in a more accurate prediction of the channel condition, as the probability of a mistake in the prediction would be significantly lower; however, it would also lead to a higher number of lost slots while the BS is awaiting to make that prediction). When the BS determines that a mobile video terminal is in LB state, if that terminal has more reserved slots in the current channel frame, the BS deallocates these slots. Full priority for these slots is given to handoff telemedicine video terminals, followed by telemedicine video users originating from within the cell, then by handoffed regular video users and finally by regular video users originating from within the cell; the allocation of the abandoned slots within each priority type is FCFS.

When the channel of the mobile terminal to which the slots were originally allocated returns to the good state, the terminal needs to inform the BS of this change, if it still has packets to transmit. This is done by transmitting a request packet. The terminal has to follow this procedure also in the case of a wrong estimation by the BS (i.e., if it was in SB state despite the long error burst). Therefore, in the (unlikely but not improbable) case of a wrong estimation, this does not influence the throughput achieved by our protocol in heavy traffic loads (slots are simply allocated to other telemedicine video and regular video users) but it results in an unnecessary increase of contention.

4.2 Contention resolution

When resolving the contention among all requesting users, the BS needs to service the telemedicine traffic first, due to its urgency. To achieve this objective, we need to guarantee highest priority to telemedicine traffic. The priority order used by the BS in our proposed scheme is the following: ECG, X-Ray, telemedicine Image, telemedicine

video. The choice of priorities has been made based on the importance that each of these traffic types currently has for medical care [2, 5, 6, 7, 12, 13, 14, 15, 16]. Highest priority is given to handoff telemedicine traffic, with telemedicine traffic originating from within the cell following in priority, in the same order (provided, of course, that the telemedicine users from within the cell have successfully transmitted their requests at the beginning of the frame request interval). Handoff regular traffic is transmitted next, with priority {video, voice, email, web}, based on the strictness of the QoS requirements for each traffic type (video and voice have the same QoS requirement of less than 1% packet dropping, but video traffic is much burstier, therefore it is granted priority over voice). Finally, regular traffic originating from within the cell is transmitted with the same priority order.

It needs to be mentioned again that, similarly to [17], we are able to ensure the priority of telemedicine traffic with the use of the two-cell stack protocol [22] for contention resolution. By exploiting the two-cell stack's advantage of clearly defining the end of contention among users of the same priority class, users of lower priority classes cannot affect the QoS of users of higher priority classes in our system. The use of the two-cell stack protocol also enables users who are moving from cell X to cell Y without having been able to access the channel in cell X, to transmit their request packets to the BS of cell Y with higher priority than new users originating from within cell Y; i.e., only when contention among the request packets of handoff telemedicine users (who were not yet transmitting in cell X) has ended, will regular users originating from within cell Y be able to transmit their request packets.

We employ the two-cell stack reservation random access algorithm (section 3.3) for telemedicine video, regular video and voice terminals, and the two-cell stack blocked access collision resolution algorithm to resolve the contention of ECG, X-ray, medical image, email and web terminals.

The major scheduling problem in serving telemedicine traffic is that not only does it have very strict QoS requirements, but it is also very bursty; hence it is necessary

for it to be transmitted immediately when it arrives, but on the other hand its arrivals are bursty, which means that the choice of constantly dedicating request bandwidth to it will often result in the loss of that bandwidth. This problem can be solved with the use of Call Admission Control (CAC) module at the entrance of the system; the BS is hence notified of which *types* of telemedicine traffic are active in the cell, and can decide how many request slots should be dedicated to telemedicine users. For example, in the extreme case when all four types of telemedicine traffic are present, both from handoffs and from traffic within the cell, at least 8 of the 25 request slots will be needed (i.e., one slot per type of handoff telemedicine traffic and one slot per type of telemedicine traffic from within the cell) to be dedicated to telemedicine users (the two-cell stack protocol needs a minimum of 2 minislots to resolve contention or to denote the absence of contention in a specific channel frame).

On the other hand, if more than 8 slots are needed to resolve the contention among users of each telemedicine traffic type, then contention will continue until all collisions have been resolved; only then will users of regular traffic (both handoff and from within the cell) get the opportunity to transmit their request packets. Since the case of 8 or more request slots being needed to resolve telemedicine traffic contention is quite infrequent, because of the nature of telemedicine traffic, it will be clear from our results that our scheme can satisfy the strict QoS requirements and the urgency of telemedicine traffic by devoting most of the time less than $8/566=1.4\%$ of the total bandwidth for telemedicine request packets.

4.3 Fair Scheduling

If users of the same type of traffic are served in a FCFS order once they are admitted into the network (as in [17] and in most relevant works on MAC protocols in the literature), the average performance evaluation metrics will give no insight on the QoS of each individual wireless subscriber; therefore, it could be the case that certain users

have their QoS severely violated while others get exceptional QoS, which would give a seemingly acceptable average QoS over the total number of users. This approach is, however, unfair to users who arrive later in the network and hence are placed at the bottom of the BS service queue; the problem is especially significant in the case of telemedicine video and regular video users, where early arriving users may dominate the channel by being allocated large numbers of slots, allowing just a small number of resources to be available for users arriving later.

For this reason, we introduce the following Fair Scheduling scheme for telemedicine video and regular video users (the scheme is enforced separately among users of each of the two types of traffic, since telemedicine video users have higher priority). The BS allocates bandwidth by comparing the channel resources to the total requested bandwidth, currently, from all active video users. If the available bandwidth is larger than the total requested bandwidth, all users will be assigned as many slots as they have requested. If, however, the available bandwidth is smaller than the total requested bandwidth, then the available bandwidth will be shared among video users proportionally. More specifically, let M be the number of currently idle information slots in the frame and B_i the amount of bandwidth that will be assigned to video terminal i in every channel frame. B_i is given by:

$$B_i = M * \frac{D_i}{\sum_i D_i} \quad (4.1)$$

Where D_i is the i^{th} user requested bandwidth and $\sum_i D_i$ is the total bandwidth requested by all of the video terminals at that moment.

It is intuitively clear, and it will also be shown from our results in Section 6, that with the use of this formula the number of telemedicine and regular, respectively, video users whose QoS is violated significantly decreases. The above scheme does not need to be implemented on any of the other types of traffic considered in our work, since they are allocated only one slot per frame.

However, an additional scheduling policy is needed for X-Ray and medical image traffic, as the upper bounds for their transmission delays are equally strict with those

for telemedicine video and ECG traffic. This strictness can be explained by the fact that X-Ray and medical image terminals are allocated only one slot per frame (close to 35 Kbps, similarly to regular voice, email and web traffic), to allow for the significantly larger numbers of slots needed by telemedicine and regular video users. Therefore, a typical X-Ray file of 200 Kbytes needs 50 seconds to be transmitted (while the upper bound for its transmission delay is set to 1 minute) and an average-sized medical image file of 17.5 Kbytes needs 4.4 seconds to be transmitted (while the upper bound for its transmission delay is set to 5 seconds). The allocation of only one slot per frame to these types of traffic, although defensive enough to prevent cases where newly arriving telemedicine video traffic cannot find enough resources to transmit, is not the most efficient in terms of bandwidth utilization. On the other hand, if these types of traffic are constantly granted more than one slot per frame, this could lead to the existence of too few idle slots for the bursty telemedicine video users. Therefore, in order to maximize system bandwidth utilization we use the following scheduling policy. After the end of the request interval at the beginning of each frame, the BS is aware of whether there are information slots during the current frame which will be left idle. These slots are allocated, *only for the current frame*, to X-Ray and medical image users who have already entered the network (with priority to X-Ray users), as additional slots to their guaranteed single slot per frame. Hence, the telemedicine traffic transmission is expedited and channel throughput is increased. This policy does not need to be extended to ECG traffic, as ECG users need only one slot per frame.

Chapter 5

Adaptive Bandwidth Reservation based on Mobility and Road information

5.1 Overview

Within a picocell, spatially dispersed source terminals share a radio channel that connects them to a fixed base station (BS). The BS allocates channel resources, delivers feedback information, and serves as an interface to the mobile switching center (MSC). The MSC provides access to the fixed network infrastructure.

It is a common assumption in the literature that the dissatisfaction of a wireless cellular subscriber who experiences forced call termination while moving between picocells is higher than that of a subscriber who attempts to access the network for the first time and experiences call blocking [35, 36]. For this reason, it is important that the system is able at any point in time to accommodate newly arriving handoff calls in any cell of the network. On the other hand, the policy of reserving a significant amount of bandwidth for possible handoff calls may lead to a portion of the bandwidth being left unused, due to small volumes of handoff traffic, while at the same time the

remaining available resources for newly generated traffic from within the cell may not suffice. The work that follows, in this Chapter, is focused on this problem.

One of the most common methods used in the literature to handle handoff traffic is to calculate either sojourn time [37, 38] or handoff probability/rate [39, 40, 41, 42] based on some updated location information. Besides the basic location information gained through the Global Positioning System (GPS), the mobility models in some of these schemes also used road map information to facilitate the prediction [38, 39, 42] while others did not [37, 40, 41, 43].

The work in [39] shows the advantages of using road map information. The proposed approach, using mobility predictions, was shown to outperform the schemes without road map information (e.g., in [43]) which lead to an unnecessary waste of reserved bandwidth. Our proposed scheme also uses road map information as a fundamental component of its function.

5.2 Network and Mobility Models

We consider a hexagonal cell architecture, as shown in Figure 5.1. We consider in all cells the uplink (wireless terminals to BS) wireless channel. Fourth generation mobile data transmission rates are planned to be up to 20 Mbps, therefore we consider this to be the uplink rate. The following mobility assumptions are adopted from [38, 39].

Each cell has six neighbors. The cell diameter is 300 meters. Roads are modeled by straight lines. Each road is assigned a weight (ω_j for road j), which represents the traffic volume. Each new call is generated with a probability of 50% to be moving on the road and 50% to be stationary. Moving users are assumed to be traveling only on the roads, and are placed on each road i with probability

$$\frac{\omega_i}{\sum_{j=1}^N \omega_j} \quad (5.1)$$

where N is the total number of roads.

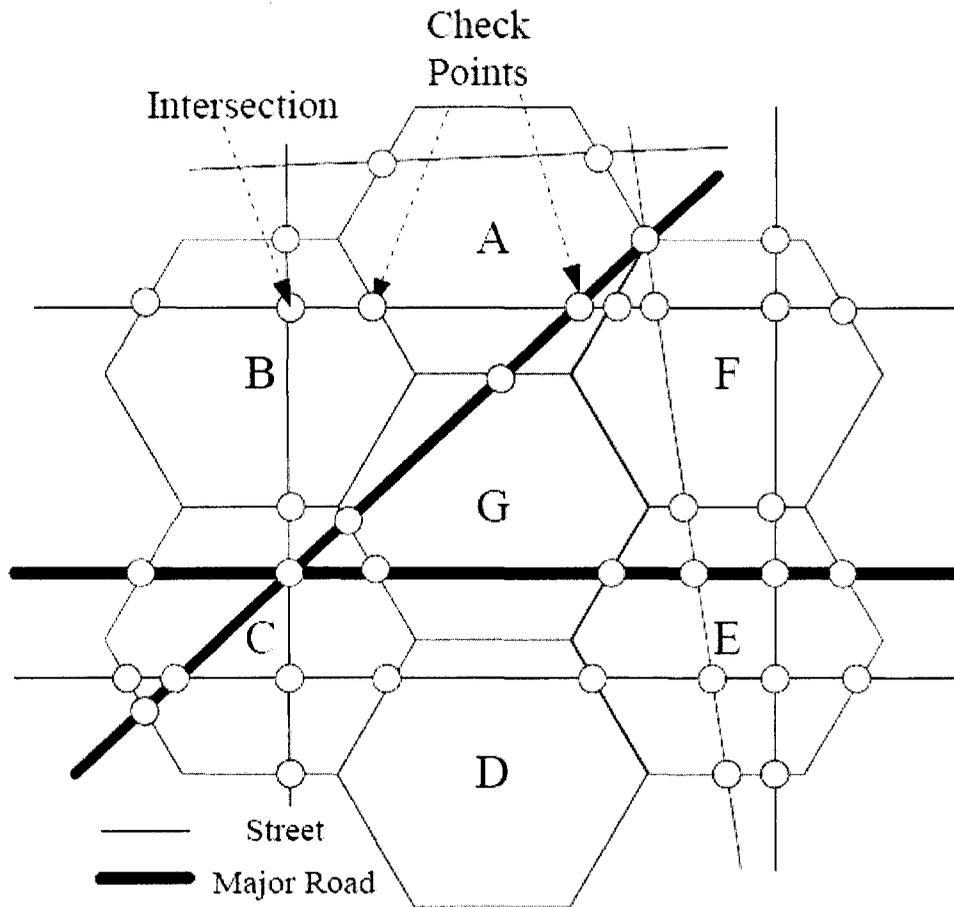


Figure 5.1: Road Map and cellular network model

The initial location of a moving user on a particular road is a uniform random variable between zero and the length of that road. During their call, stationary callers remain stationary and mobile users travel at a constant speed. Mobile users can travel in either of the two directions of a road with an equal probability, and with a speed chosen randomly in the range of $[36, 90]$ Km/h. At the intersection of two roads, a mobile user might continue to go straight, or turn left, right, or around with probabilities 0.55, 0.2, 0.2 and 0.05, respectively. If a mobile user chooses to go straight or turn right at the intersection, it needs to stop there with probability 0.5

for a random time between 0 and 30 seconds due to a red traffic light. If the user chooses to turn left or around, it needs to stop there for a random time between 0 and 60 seconds due to the traffic signal. Each base station is loaded with the road map of its coverage area and its neighboring cells. Mobile stations report their position to the BS of their cell through a control channel. The position information includes the mobile user's exact location (cell and road), moving direction, and speed, and can be provided with an accuracy of 1m through GPS [38, 39, 42].

5.3 Adaptive Bandwidth Reservation Scheme

The bandwidth reserved schemes in [38, 39, 42] require mobile stations to report their location information to the BS every T seconds (1 second in [42], 10~45 seconds in [38] and 60 seconds in [39]). For this period of time if there is a probability based on the mobile's trajectory that it will move to a new cell, then a certain amount of bandwidth is reserved in all possible future cells that the mobile may move to.

In [38] the mobile's recorded moving history is also used for the prediction. In [39] a prediction is made by the system based on each updated position information and the road map. In [35], the authors do not use a standard road map for their study, but generate a random map for each simulation. The common disadvantage of these approaches is that the length of the report period yields a tradeoff between the prediction accuracy and the computational load imposed on the system. If mobile stations report their position frequently, the computational load of processing this information and using it for bandwidth reservation will be too high; on the other hand, if mobile users' position reports are too infrequent, the prediction on the mobile's trajectory can be untrustworthy and lead to an unnecessary waste of the reserved bandwidth in neighboring cells. Because of this tradeoff, the authors in [38, 42] use additional dynamic mechanisms in their schemes in order to adjust the amount of reserved bandwidth for users in neighboring cells, based on the quality of the system

performance; we will explain in the following why this “corrective” approach, which imposes a further computational load on the system, is not needed in our scheme.

In order to eliminate the problems introduced by the time-based location information reports of the mobile station to the BS, we propose the use of *distance-based* information reports.

More specifically, we set the road intersections and cell boundaries to be the “check-points” of our system, as shown in Figure 5.1. Each cell boundary represents a unique check-point, while around each road intersection four check points are set, one in each possible direction that the user may choose after reaching the intersection. Each of the check points is assumed to be placed at a distance of 10m from the intersection, which is a distance that even the slowest moving vehicles considered in this work (with a speed of 36 Km/h) will cover within 1s after passing the intersection (naturally, if there are less than four possible directions for a mobile to follow after reaching an intersection, the number of check points needed is smaller than four). Mobile stations only need to update their position information to the BS of their cell when they arrive at a check-point.

If the BS of the current cell of a mobile station predicts, based on the station’s location (at a check point) and speed that the station is going to move to another cell (i.e., that the next check point for the station will be a cell boundary), it sends a notification to the BS of that cell, including the current bandwidth used by the station and the estimated arrival time at the next check point. Hence, the proper amount of bandwidth is reserved for the station. For telemedicine video and regular video terminals, the bandwidth that is reserved in the next cell is equal to the remaining bandwidth that the terminal will need to complete its transmission (this bandwidth is declared in each video user’s initial request to the BS). For all other types of users, the bandwidth that is reserved in the next cell is equal to their current bandwidth, so that they will seamlessly continue its transmission. Our proposed approach not only guarantees the existence of adequate resources in the new cell for handoff users

but also reduces the information update frequency.

The prediction and adaptive reservation process executed by the BS side can be summarized as algorithm 1.

Algorithm 1 Distance-based Adaptive Bandwidth Reservation Scheme

for Each new user in the system **do**

if The mobile station is not stationary **then**

When the mobile station arrives at a check point

 Update the position information

 Estimate the next arrival check-point and calculate the arrival time at that check point

if The next check point is a cell boundary **then**

 Reserve the proper amount of bandwidth for the station in the next cell

end if

end if

end for

Our proposed adaptive bandwidth reservation scheme based on *distance-based location updates* has two major advantages:

1. The position update duration is unique for every mobile user, based on their different initial position and speed. This is impossible to “capture” with the time-based location updates proposed in [38, 39, 42]. Additionally, it creates less computational load than the time-based updates which are in most proposals in the literature synchronized and therefore require simultaneous information transmission from the terminals to the BS.
2. An important parameter used to evaluate a bandwidth reservation scheme is bandwidth efficiency [39], which is calculated by $f = N_r/N_q$, where N_r is the reserved bandwidth and N_q is the actual bandwidth utilized by hand-offed users. The closer f is to 1, the higher is the efficiency achieved by the bandwidth

reservation scheme, since this would mean that there is neither lack nor waste of bandwidth in the reservation procedure. The scheme in [39] outperformed other schemes with which it was compared, achieving at best a $f = 1.047379$ (the schemes with which [39] was compared achieved a bandwidth efficiency in the range of 1.25).

We argue that with our distance-based approach the bandwidth efficiency of our scheme can asymptotically reach 1. The reasons are:

- Our scheme guarantees that there is no waste of bandwidth. For all users ready to handoff, the exact amount of bandwidth they need is reserved in the new cell and this cell is known with precision, with the use of the four check points around each intersection. The only case when bandwidth can be wasted, with the use of our adaptive bandwidth reservation approach, is when a mobile station which is predicted to move to another cell makes a stop before entering this cell (this case is not included in our adopted mobility model). Given the fact that the distance between a check point set close to an intersection and a cell boundary is in all cases much smaller than the cell diameter of 300 meters, this case can be considered a rare exception.
- Our scheme also guarantees that there will be no lack of bandwidth for hand-offed users. As explained in Chapter 4, the bandwidth needed for all types of telemedicine and regular mobile stations which do not transmit video is close to 35 Kbps (just one slot per channel frame). Therefore, this bandwidth is very small compared to the total channel capacity of 20 Mbps and can generally be reserved in the next cell with the rare exception of cases when the channel is overloaded with traffic. In the case of hand-offed telemedicine users (transmitting any type of telemedicine traffic), if the needed bandwidth is not available, then email and web users are preempted in the new cell in order for the system to grant this bandwidth to the high-priority telemedicine traffic. When email

and web reservations are canceled, the BS notifies the affected data terminals and places them at the front of the respective (email or web) queue of terminals awaiting bandwidth allocation. The priorities among telemedicine traffic types in bandwidth reservation are set in the same way as the scheduling priorities which were discussed in Chapter 4.

Chapter 6

Results and Discussion

6.1 Simulation Setup

We use computer simulations to study the performance of our scheme. The simulator is written in C programming language. Each simulation point is the result of an average of 10 independent runs (Monte-Carlo simulation), each simulating 305000 frames (the first 5000 of which are used as warm-up period).

The following weight values ω_j were assigned, without loss of generality, to the 8 roads in the road map: 1, 1, 5, 1, 6, 1, 1, 1. The two larger weights correspond to the two main roads (shown in Figure 5.1 in bold black). We have experimented with other weight values, as well, without this change having any effect on the nature of our results.

In our results, we use different traffic “combinations” from all types of traffic considered in our work, in order to test the system’s performance in a large variety of cases. In this way, we try to produce results representative of different practical cases, where one type of telemedicine traffic might be more dominant than others in any given moment. Each simulation point presents the average result over 10 combinations which were used to create a specific traffic load.

6.2 Results

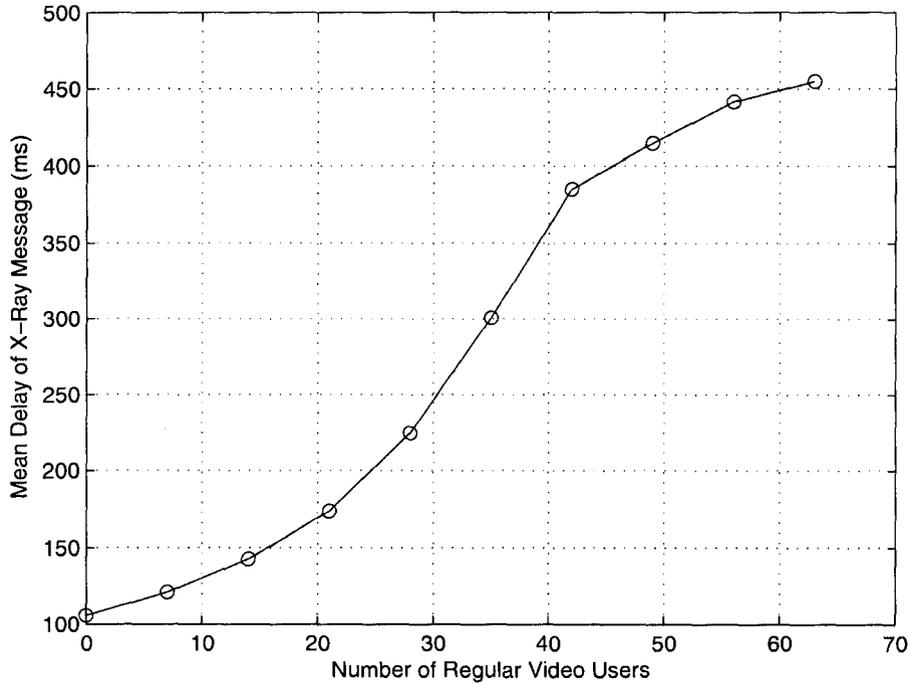


Figure 6.1: Effect of regular video traffic on X-Ray traffic.

Our results are presented in Figures 6.1 - 6.7. In these results, we keep the telemedicine traffic equal to 10% of the total channel capacity (we use 10 different traffic “combinations” of the four types of telemedicine traffic, in order to create this load; in this way, our results can be representative of different real-world cases, where one type of telemedicine traffic might be more dominant than others in any given moment).

Figure 6.1 shows the effect that the increase in the number of regular video traffic users has on the QoS of X-Ray traffic. The maximum number of video users (63) corresponds to 90% of the total traffic being generated by regular traffic. The delay in the transmission of X-Ray files increases very slowly and does not become larger

than 500 ms even for very high numbers of regular video users. Regarding the QoS of the other types of telemedicine traffic (for the same range of regular video users), telemedicine image files are transmitted on average within less than 100 ms, ECG packets are transmitted, on average, within half a frame (6 ms) and the average packet dropping probability of telemedicine video ranges between 0.003% and 0.0045% as it is affected only by the wireless channel errors.

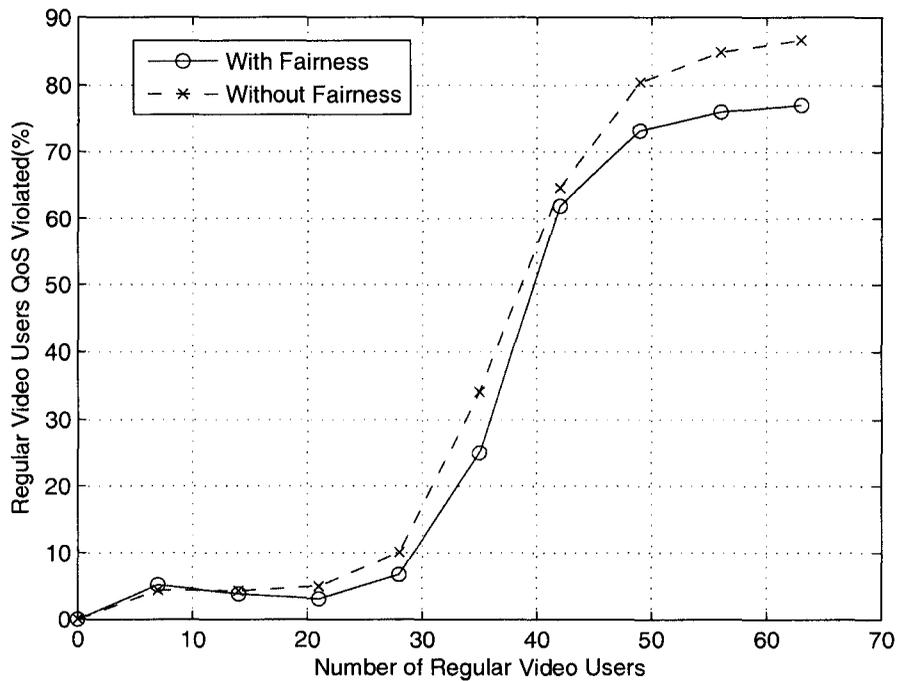


Figure 6.2: Percentage of regular video users who experience packet loss larger than 1%, vs. the number of regular video users.

All these values are far below the acceptable upper bounds. These results, combined with the fact (shown in Figures 6.2 - 6.3) that the increase in the number of regular video traffic is clearly affecting only that type of traffic (the video packet dropping probability of regular video users rises significantly above the 1% acceptable upper bound) show that our scheme succeeds in offering absolute priority to

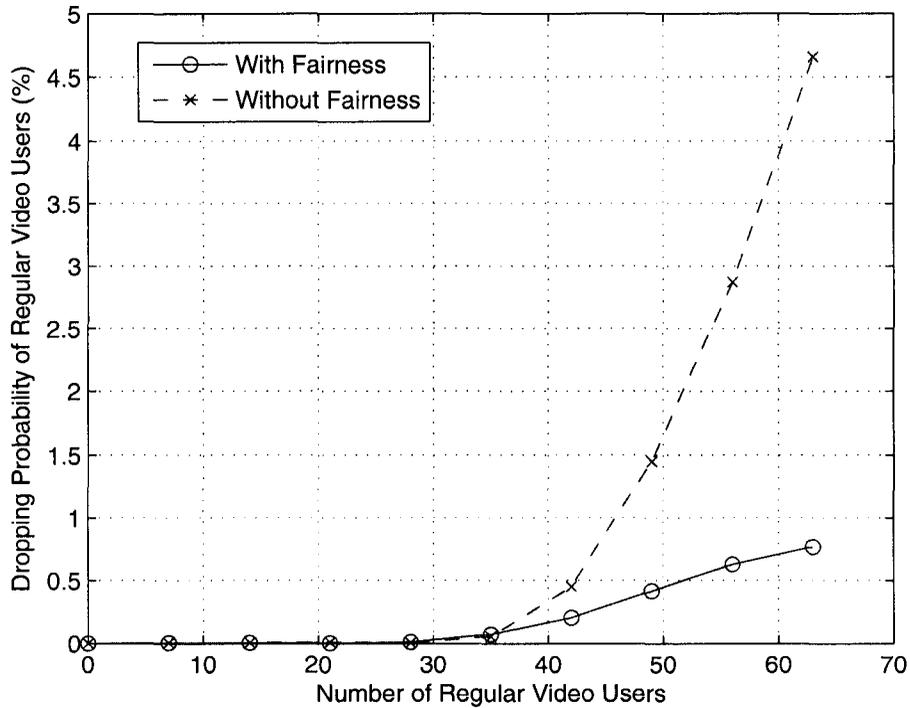


Figure 6.3: Regular video packet dropping vs. the number of regular video users.

telemedicine traffic. Figures 6.2 and 6.3 also show that the use of our fair scheduling scheme helps to substantially improve both the average video packet dropping probability over all regular video users, and the individual QoS of each regular video user. The improvement in the individual QoS (much smaller percentage of users who experience packet loss greater than the upper bound, as shown in Figure 6.2) is due to the allocation of available resources in each channel frame proportionally to each user's requested bandwidth. The result shown in Figure 6.3, regarding the improvement in the average video packet dropping is again owed to the proportionate allocation; the reason for this improvement is that our scheme prevents the case where a user whose transmission deadline is not imminent may dominate the channel, hence not allowing users with imminent transmission deadlines to transmit.

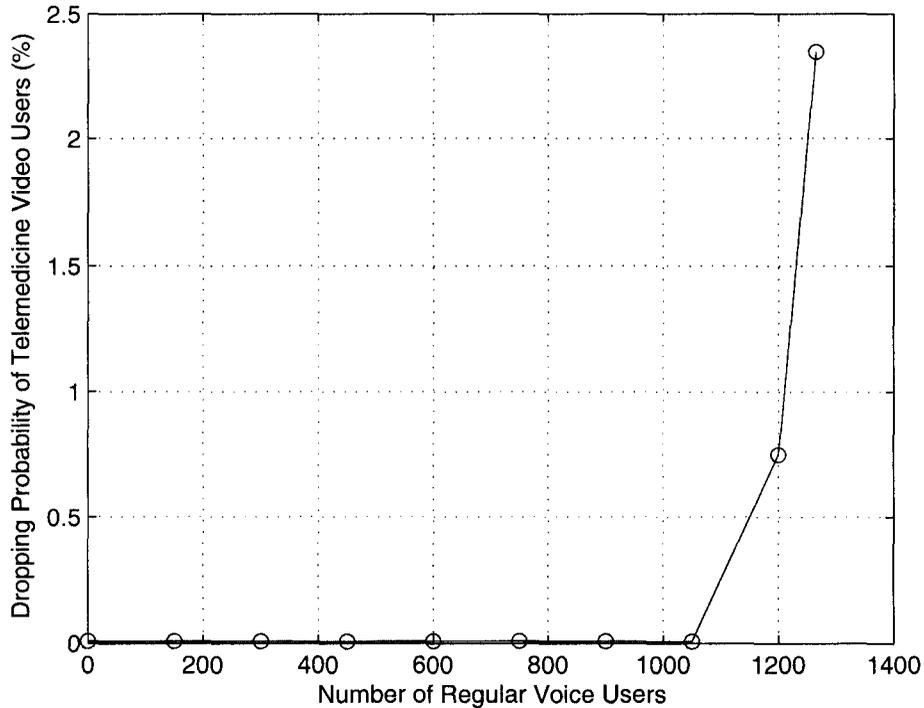


Figure 6.4: Effect of regular voice traffic on telemedicine video traffic.

Also, Figures 6.4 and 6.5 show that the increase in the number of voice users has minimal effect on the QoS of telemedicine users, such as packet dropping and delay. The reason, once again, is that our combined scheduling and adaptive bandwidth reservation schemes guarantee full priority to all types of telemedicine traffic. The average telemedicine image traffic transmission delay does not exceed 4.5 seconds even when the number of voice users exceeds 1250, which corresponds to 94% of the total channel capacity being utilized by voice users. For the same channel utilization by voice users, we found that the average X-Ray traffic transmission delay does not exceed 30 seconds. Similarly, the telemedicine video dropping probability is shown in Figure 6.4 to remain below the strict upper bound of 0.01% until the number of voice users exceeds 1070, which corresponds to 80% of the total channel capacity.

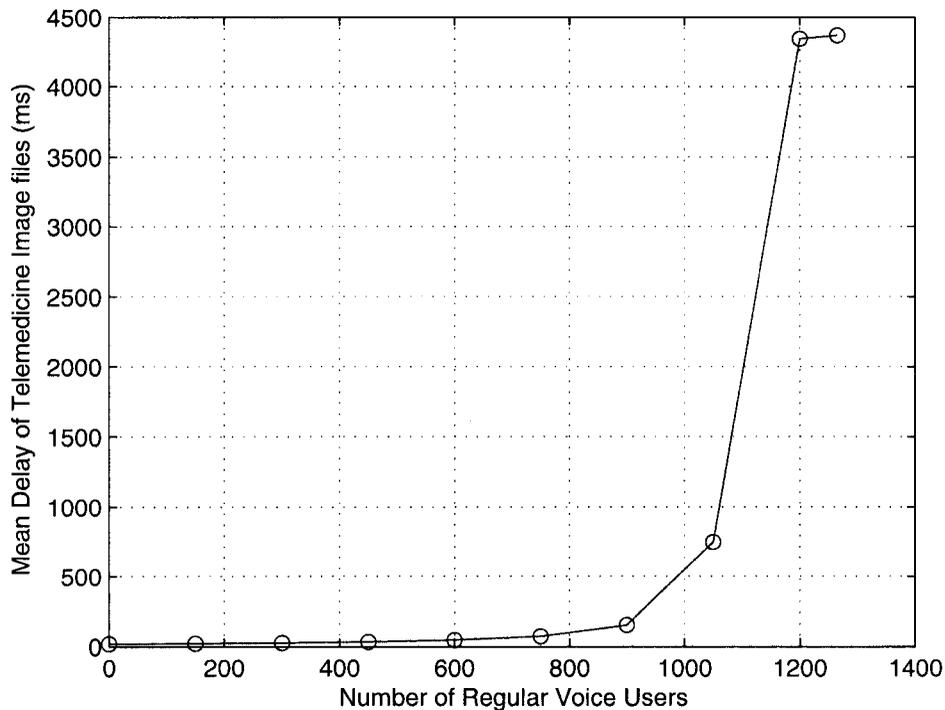


Figure 6.5: Effect of regular voice traffic on telemedicine image traffic.

Hence, only in the case of a very heavily loaded channel with voice traffic, does the telemedicine traffic experience some deterioration in its QoS.

On the other hand, Figure 6.6 shows the effect that the increase in telemedicine load has on regular video traffic, which is the most bursty of all regular traffic types. The increase in the number of telemedicine video users results in a very significant increase in regular video packet dropping, showing once again that telemedicine traffic is treated with absolute priority in our combined adaptive bandwidth reservation/scheduling schemes. The results for all other types and combinations of telemedicine and regular traffic confirm that telemedicine traffic is negligibly affected by an increase in regular traffic, while regular traffic is severely affected by increased loads of telemedicine traffic.

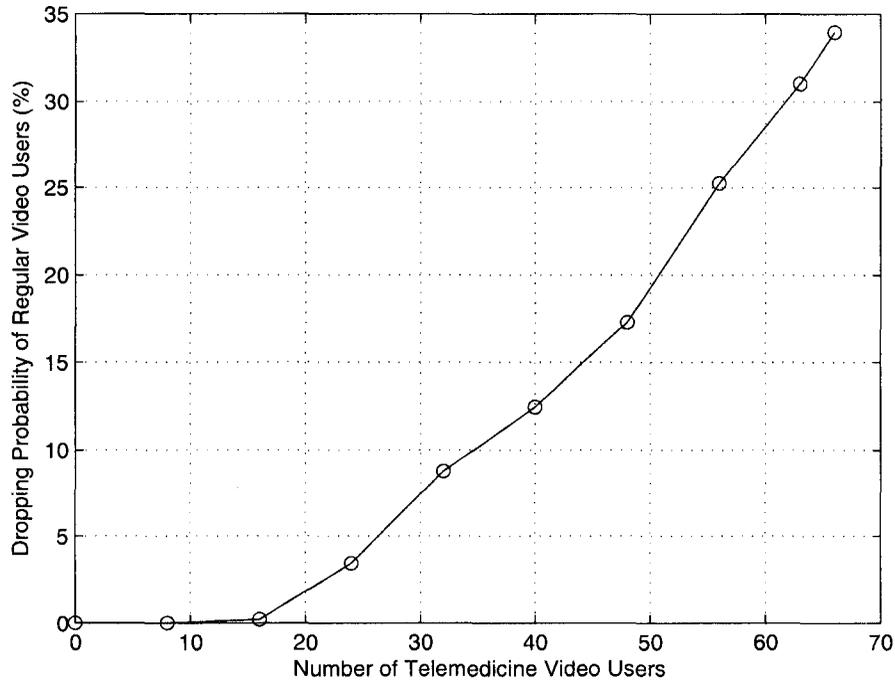


Figure 6.6: Effect of telemedicine Video traffic on regular Video traffic, with 10% handoff traffic

Finally, Figure 6.7 shows the significant improvement achieved by the use of our adaptive bandwidth reservation scheme on the QoS of the most widely used cellular application, i.e., voice. The voice packet dropping probability of handoff voice users is smaller by 16.73% on average, in comparison to the case when the bandwidth reservation scheme is deactivated. The reason is that, by intelligently reserving bandwidth in adjacent cells with bandwidth efficiency f almost equal to 1, our scheme helps to significantly decrease contention for channel resources. Therefore, once again our proposal is shown to improve the QoS of regular traffic, while always offering highest priority to telemedicine users. The use of our adaptive bandwidth reservation scheme is of equal benefit to telemedicine users as to regular users, but given the space limi-

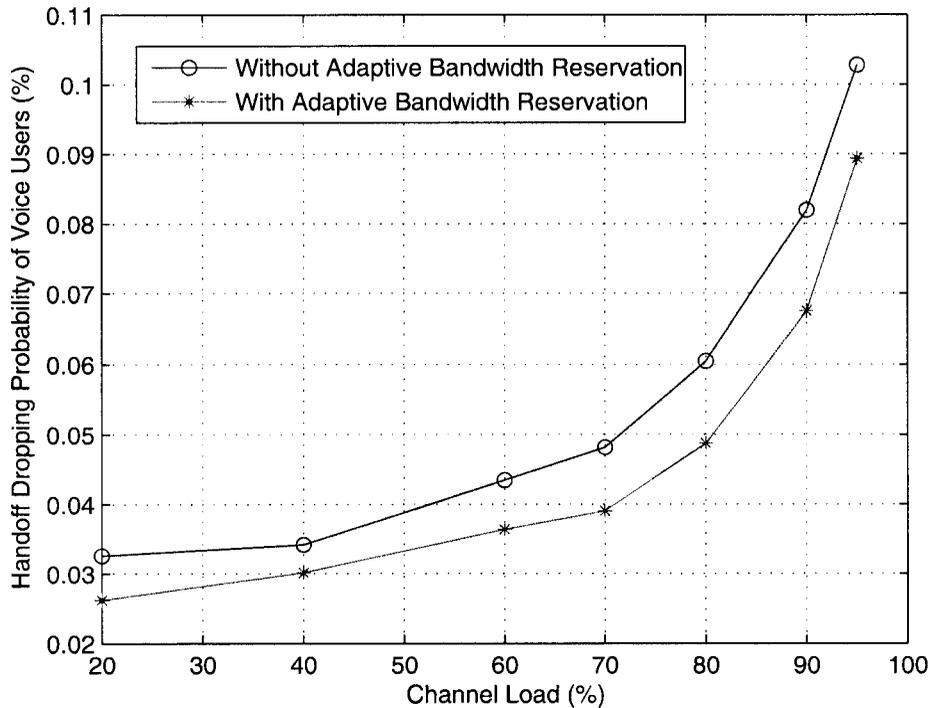


Figure 6.7: Improvement on voice packet dropping probability with the use of adaptive bandwidth reservation.

tations we chose to show its effect on voice traffic, which is certain to occupy a much larger load than telemedicine traffic in the network, and therefore comprise a much larger load of handoff traffic that the cellular network will have to accommodate.

6.3 Maximum Achievable Throughput

In this section, we present our results when varying the telemedicine traffic load in the network and studying what is the maximum channel throughput that the system can achieve while satisfying the QoS requirements of all traffic types. Given that telemedicine traffic is unlikely to ever be the dominant traffic type in the network, we

have studied three cases:

- Telemedicine traffic is 5% and regular multimedia traffic is 95% of the channel load.
- Telemedicine traffic is 10% and regular multimedia traffic is 90% of the channel load.
- Telemedicine traffic is 15% and regular multimedia traffic is 85% of the channel load.

The respective loads are created with various traffic “combinations”, both for telemedicine and for regular traffic. Our results have shown that it is the QoS requirements of regular video traffic that are violated first, in all studied cases. The reason is the strictness of their requirements, combined with the very bursty nature of MPEG-4 video conference traffic and with the fact that we offer full priority to telemedicine traffic. Table 6.1 shows that the maximum channel throughput achieved in all cases in the presence of regular video traffic is substantially smaller than that achieved when no regular video users are active in the network (by 10 ~ 15%).

Table 6.1: Maximum Achievable Throughput

Traffic Model	Maximum Throughput	
	With Regular Video	Without Regular Video
5%Telemedicine - 95%Regular	68.02%	83.30%
10%Telemedicine -90%Regular	74.29%	88.13%
15%Telemedicine - 85%Regular	78.12%	87.43%

It is also clear from Table 6.1 that the maximum achievable throughput increases with the increase of the telemedicine traffic load, in the presence of regular video. The reason is that the burstiness of regular video traffic does not allow the system capacity to be fully utilized, as in that case (e.g., if the channel is filled with voice

traffic) the regular video QoS requirements will be quickly violated upon arrival of new video users; hence, when the regular traffic load decreases, the system capacity is better utilized and the throughput increases.

In the absence of regular video traffic, we observe from the Table that the maximum achievable throughput increases when the telemedicine load moves from 5% to 10% and then slightly decreases when the telemedicine load becomes 15%. The reason for this increase and decrease is that, when no regular video traffic is present, the dominant “regular” type of traffic is voice; when the “regular” traffic load is 95% of the total load, the maximum throughput is lower, because the large number of voice users leads to higher contention. For this reason, the throughput increases when the regular traffic load decreases from 95% to 90%. However, when the regular traffic load decreases further (to 85%), the number of voice users (who are non-bursty, and for this reason easily accommodated) in the system decreases and the respective decreased voice load is substituted by telemedicine load, which is bursty and needs to be urgently serviced; this leads to a decrease in the maximum achievable throughput.

Our results have also shown that, in the absence of telemedicine traffic, the system can again achieve throughput larger than 80%, and when only voice and regular data traffic (web, email) is present, system throughput surpasses 90%.

Chapter 7

Conclusions and Future work

The importance of errorless and expedited telemedicine traffic transmission over wireless cellular networks is quickly becoming a very important issue, due to the current systems' inability to provide high QoS to transmissions from mobile telemedicine terminals. Recent studies have focused solely on the transmission of telemedicine traffic over the cellular network, without taking into account the critical fact that regular traffic has strict QoS requirements as well, therefore a practical solution should focus on the efficient integration of the two traffic categories.

In this work we propose, for the first time in the relevant literature to the best of our knowledge, the combination of a fair and efficient scheduling scheme and an adaptive bandwidth reservation scheme which enable the integration of highest-priority telemedicine traffic transmission with regular wireless traffic over cellular networks. We include in our extensive simulation study all major types of current telemedicine applications (Electro-Cardiograph, X-Ray, Video and high-resolution medical still Images), as well as the most popular "regular" applications (voice, video, email, web). With the use of new scheduling ideas and *distance-based* information reports, in order to achieve seamless handoff, our proposal, which is evaluated over a hexagonal cellular structure, is able to provide full priority and satisfy the very strict QoS requirements of telemedicine traffic without violating the QoS of regular traffic, even in the case of

high traffic loads.

Our future work will focus on the design and implementation of an intelligent Call Admission Control mechanism which will be able to make decisions on the possible degradation of the QoS of existing regular mobile users in order to accommodate new telemedicine traffic users, while at the same time attempting to maximize wireless bandwidth utilization.

Bibliography

- [1] D. A. Perednia and A. Allen, "Telemedicine Technology and Clinical Applications," *the Journal of the American Medical Association (JAMA)*, vol. 273, pp. 483–488, Feb 1995.
- [2] A. Bhargava, M. F. Khan, and A. Ghafoor, "QoS management in Multimedia Networking for Telemedicine Applications," in *Proceedings of the IEEE Workshop on Software Technologies for Future Embedded Systems*, pp. 39–42, 2003.
- [3] B. Tulu, S. Chatterjee, and S. Laxminarayan, "A Taxonomy of Telemedicine Efforts with Respect to Applications, Infrastructure, Delivery Tools, Type of Setting and Purpose," in *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS)*, 2005.
- [4] Y. B. Choi, J. S. Krause, H. Seo, K. Capitan, and K. Chung, "Telemedicine in the USA: Standardization through Information Management and Technical Applications," *IEEE Communications Magazine*, vol. 44, pp. 41–48, April 2006.
- [5] Y. Chu and A. Ganz, "A Mobile Teletrauma System Using 3G networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 456–462, 2004.
- [6] J. R. Gallego, A. Hernandez-Solana, M. Canales, J. Lafuente, A. Valdovinos, and J. Fernandez-Navajas, "Performance Analysis of Multiplexed Medical Data

- Transmission for Mobile Emergency Care over the UMTS channel,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 1, pp. 13–22, 2005.
- [7] S. C. Voskarides, C. S. Pattichis, R. Istepanian, C. Michaelides, and C. N. Schizas, “Practical Evaluation of GPRS use in a Telemedicine System in Cyprus,” in *Proceedings of the 4th IEEE International EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 39–42, 2003.
- [8] F. Hu, Y. Wang, and H. Wu, “Mobile Telemedicine Sensor Networks with Low-energy Data Query and Network Lifetime Considerations,” *IEEE Transactions on Mobile Computing*, vol. 5, pp. 404–417, April 2006.
- [9] C. Kugean, S. M. Krishnan, O. Chutatape, S. Swaminathan, N. Srinivasan, and P. Wang., “Design of a Mobile Telemedicine System with Wireless LAN,” in *Proceedings of the Asia-Pacific Conference on Circuits and Systems*, vol. 1, pp. 313–316, 2002.
- [10] C. Chigan and V. Oberoi, “Providing QoS in Ubiquitous Telemedicine Networks,” in *Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW06)*, March 2006.
- [11] W. Yao, R. Istepanian, A. Salem, H. Zisimopoulos, and P. Gosset, “Throughput Performance of a WCDMA system Supporting Tele-echography Services,” in *Proceedings of 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 310–313, April 2003.
- [12] S. Garawi, R. S. H. Istepanian, and M. A. Abu-Rgheff, “3G Wireless Communications for Mobile Robotic Tele-Ultrasonography Systems,” *IEEE Communications Magazine*, vol. 44, no. 4, pp. 91–96, 2006.

- [13] E. A. V. Navarro, J. F. N. J. R. Mas, and C. P. Alcega, "Performance of a 3G-Based Mobile Telemedicine System," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 1023–1027, 2006.
- [14] C. H. Salvador, M. P. Carrasco, M. A. G. de Mingo, A. M. Carrero, J. M. Montes, L. S. Martin, M. A. Caverio, I. F. Lozano, and J. L. Monteagudo, "Airmed-Cardio: A GSM and Internet Services-Based system for Out-of-Hospital Follow-up of Cardiac Patients," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 1, pp. 73–85, 2005.
- [15] S. Pavlopoulos, E. Kyriacou, A. Berler, S. Dembeyiotis, and D. Koutsouris, "A Novel Emergency Telemedicine System Based on Wireless Communication Technology - AMBULANCE," *IEEE Transactions on Information Technology in Biomedicine*, vol. 2, no. 4, pp. 261–267, 1998.
- [16] C.-Y. Huang and S.-G. Miaou, "Transmitting SPIHT Compressed ECG Data over a Next-Generation Mobile Telecardiology Testbed," in *Proceedings of the 23rd IEEE International Conference of the Engineering in Medicine and Biology Society*, 2001.
- [17] P. Koutsakis, S. Psychis, and M. Paterakis, "Integrated Wireless Access for Videoconference from MPEG-4 and H.263 video coders with Voice, Email and Web traffic," *IEEE Transactions on Vehicular Technology*, vol. 54, pp. 1863–1874, 2005.
- [18] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40–54, 2001.
- [19] Q. Pang, A. Bigloo, V. C. M. Leung, and C. Scholefield, "Service Scheduling for General Packet Radio Service Classes," in *Proceedings of IEEE Wireless Communications and Networking Conference*, 1999.

-
- [20] P. Tran-Gia, D. Staehle, and K. Leibnitz, "Source Traffic Modeling of Wireless Applications," *International Journal of Electronics and Communications*, vol. 55, no. 1, pp. 27–37, 2001.
- [21] C.-F. Tsai, C.-J. Tsang, F.-C. Ren, and C.-M. Yen, "Adaptive Radio Resource Allocation for Downlink OFDMA/SDMA systems," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2007.
- [22] A. C. Cleary and M. Paterakis, "An investigation of Stack Based Algorithms for Voice Packet Transmission in Microcellular Wireless Environments," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 1995.
- [23] S. Nanda, D. J. Goodman, and U. Timor, "Performance of PRMA: A Packet Voice Protocol for Cellular Systems," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 3, pp. 584–598, 1991.
- [24] D. Dyson and Z. J. Haas, "A Dynamic Packet Reservation Multiple Access Scheme for Wireless ATM," *Mobile Networks and Applications (MONET) Journal*, vol. 4, no. 2, pp. 87–99, 1999.
- [25] E. N. Gilbert, "Capacity of a Burst-noise Channel," *Bell System Technical Journal*, vol. 39, pp. 1252–1265, 1960.
- [26] E. O. Elliot, "Estimates of Error Rates for Codes on Burst-noise Channels," *Bell Labs Technical Journal*, vol. 42, pp. 1977–1997, 1963.
- [27] A. Willig, "A New Class of Packet- and Bit-level Models for Wireless Channels," in *Proceeding of the 13th IEEE International Symposium on Personal Indoor and Mobile Radio Communication (PIMRC)*, 2002.
- [28] H. S. Wang and N. Moayeri, "Finite State Markov channel - A useful Model for Radio Communications Channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.

- [29] M. Hassan, M. M. Krunz, and I. Matta, "Markov-based Channel Characterization for Tractable Performance Analysis in Wireless Packet Networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 821–831, 2004.
- [30] M. Bottigliengo, C. Casetti, C.-F. Chiasserini, and M. Meo, "Short-term Fairness for TCP flows in 802.11b WLANs," in *Proceedings of IEEE Infocom*, 2004.
- [31] E. C. Baig, "Will Consumers Tune in to a Tiny TV in their hand?," in http://www.usatoday.com/tech/wireless/2006-08-17-mobile-tv_x.htm, July 2006.
- [32] N. M. Mitrou, G. L. Lyberopoulos, and A. D. Panagopoulou, "Voice and Data Integration in the Air-interface of a Microcellular Mobile Communication System," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 1, pp. 1–13, 1993.
- [33] Y. Li, "Pilot-symbol Aided Channel Estimation for OFDM in Wireless Systems," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 4, pp. 1207–1215, 2000.
- [34] T. Holliday, A. Goldsmith, and P. Glynn, "Wireless Link Adaptation Policies: QoS for Deadline Constrained Traffic with Imperfect Channel Estimates," in *Proceedings of the International Conference on Communications (ICC)*, vol. 5, pp. 3366–3371, 2002.
- [35] F. R. Yu, V. W. S. Wong, and V. C. M. Leung, "A New QoS Provisioning Method for Adaptive Multimedia in Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1899–1909, 2008.
- [36] Q. Song and A. Jamalipour, "A Negotiation-Based Network Selection Scheme for Next-Generation Mobile systems," in *Proceedings of the IEEE Globecom*, 2006.
- [37] K. L. Yeung and S. Nanda, "Channel Management in Microcell/Macrocell Cellular Radio Systems," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 601–612, Nov 1996.

- [38] S. Choi and K. Shin, "Adaptive Bandwidth Reservation and Admission Control in QoS-sensitive Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 13, pp. 882–897, Sep 2002.
- [39] D.-S. Lee and Y.-H. Hsueh, "Bandwidth-reservation Scheme based on Road Information for Next-generation Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 243–252, Jan 2004.
- [40] R. Zander and J. M. Karlsson, "A Rate-Based Bandwidth Borrowing and Reservation Scheme for Cellular Networks," in *Proceedings of the 60th IEEE Vehicular Technology Conference*, vol. 2, pp. 1123–1128, Sep 2004.
- [41] N. Nasser, "Enhanced Blocking Probability in Adaptive Multimedia Wireless Networks," in *Proceedings of the 25th IEEE International Conference on Performance, Computing, and Communications*, 2006.
- [42] W. S. Soh and H. S. Kim, "A Predictive Bandwidth Reservation Scheme using Mobile Positioning and Road Topology Information," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, pp. 1078–1091, Oct 2006.
- [43] M. H. Chiu and M. Bassiouni, "Predictive Schemes for Handoff Prioritization in Cellular Networks based on Mobile Positioning," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 510–522, Mar 2000.