Scheduling for Efficient Telemedicine Traffic Transmission over Next Generation Cellular Networks

Theodora Kouskouli and Polychronis Koutsakis

Dept. of Electronic and Computer Engineering,
Technical University of Crete, Greece
ntora_kousk@hotmail.com, polk@telecom.tuc.gr

Abstract. Telemedicine traffic transmission has gained in importance during the past few years. Due to the fact that it carries critical information regarding the patients' condition, the expedited and errorless transmission of multimedia telemedicine traffic is of fundamental importance. The prioritized or guaranteed transmission of telemedicine traffic, however, can lead to the violation of the Quality of Service (QoS) requirements of regular traffic users and to the loss of guaranteed bandwidth in cases when it is left unused, due to the infrequent nature of telemedicine traffic. To resolve these problems, we propose a Multiple Access Control (MAC) protocol, for the integrated transmission of regular and telemedicine traffic transmission over next generation cellular networks and an adaptive bandwidth reservation scheme based on road map information and on user mobility and a fair scheduling scheme for telemedicine traffic transmission over cellular networks. The combination of the two schemes achieves high channel bandwidth utilization while offering full priority to telemedicine traffic over regular traffic. Three scheduling algorithms are evaluated, quantitatively and qualitatively, in terms of the QoS and the fairness they offer to different types of traffic.

1 Introduction

Telemedicine has been defined as using telecommunications to provide medical information and services [1], and is already being employed in many areas of healthcare.

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 [3, 4]. Thus, 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 [3, 5]. Mobile healthcare (M-health, "mobile computing, medical sensor and communication technologies for healthcare" [6]) 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.

Next generation cellular networks will be able to provide voice, data and streamed multimedia to users on an "anytime, anywhere" basis. This will be achieved after

C.X. Mavromoustakis et al. (eds.), *Resource Management in Mobile Computing Environments*, Modeling and Optimization in Science and Technologies 3,
DOI: 10.1007/978-3-319-06704-9_18, © Springer International Publishing Switzerland 2014

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.

2 Description of System Model

In our work we consider the integration of regular and telemedicine traffic transmission. We consider voice, video, e-mail and short message service (sms) as representatives of regular traffic, and electro-cardiograph (ECG), X-ray, video and high-resolution medical still images for telemedicine traffic, in order to study the practical scenario of many different types of users simultaneously attempting to access the network and hence aggravating the access for telemedicine users. We describe below the characteristics of each traffic type.

2.1 Regular Multimedia Traffic

Four types of "regular" multimedia traffic are considered in our work: MPEG-4 video-conference, voice, email and mobile text messages (sms), which are the most common traffic types in cellular networks.

- Voice: The speech codec rate is 32 kb/s, and voice terminals are equipped with a voice activity detector (VAD) [9]. Voice sources follow an alternating pattern of talkspurt 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 1). The mean talkspurt duration is 1 s and the mean silence duration is 1.35 s. The talkspurt to silence transition probability is PTS and the silence to talkspurt transition probability is PST. The talkspurt and silence periods are geometrically distributed with mean 1/PTS and 1/PST frames, respectively. Therefore, at steady state, the probability that a terminal is in talkspurt (speech activity), PT , or silence, PS, is obtained from the following equations:

$$P_T = P_{ST} / P_{ST} + P_{TS}$$
$$P_S = 1\text{-}P_T$$

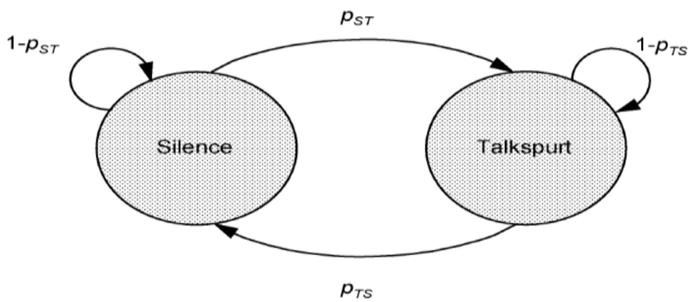

Fig. 1. The voice source activity model

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. Reserved slots are deallocated immediately. The allowed voice packet dropping probability is set to 1%, 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) [8]. 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 e-mail 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.
- SMS: Short Message Service (SMS) is a store-and-forward service that relies on a Short Message Service Center (SMSC). SMS messages are especially suitable for the transmission of small data bulks and for transmissions repeating in long time intervals (minutes to hours). The SMS payload is 140 bytes (including a header of 13 bytes) [24]. The message inter-arrival time distribution is considered exponential. In this work, in order to test our system under the strictest QoS requirements, we set an upper bound of two seconds in SMS transmission delay.
- Regular MPEG-4 Video Streams: The two MPEG-4 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 [10]. We have used the high quality version of the videos: one has a mean bit rate of 400 kb/s, a peak rate of 2 Mb/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. If a video packet does not arrive on time, the play out process will pause, which is annoying to the human eye. The allowed video packet dropping probability is set to 1% [11], as the loss of regular video packets is not of equally critical importance as that of telemedicine video packets for which the maximum allowed video packet dropping probability is 0.01%, as it will be explained in Section 2.2. The two video traces are chosen with equal probability by regular video users.

2.2 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): 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 users/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 [5] 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.
- Medical Images: Medical image files have sizes ranging from 15 to 20 Kbytes/image [2] and are Poisson distributed with mean λ_I files/frame. The upper bound for the transmission delay of a medical image is set to 5 seconds 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 [10] 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%.

3 Multimedia Integration Access Control (MI-MAC)

The Multimedia Integration Multiple Access Control (MI-MAC) protocol, introduced in [7] 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 sms 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.

3.1 Channel Frame Structure

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.

Our work focuses on the uplink (wireless terminals to Base Station) channel. The uplink channel time is divided into time frames of fixed length. The frame duration (12 ms accommodating 566 slots) is selected such that a voice terminal in talkspurt generates exactly one packet per frame. The packet size is considered to be equal to 53 bytes, 48 of which contain information. This choice was made in [7] in order to compare the protocol with other works in the literature; it is by no means restrictive, and does not influence the efficiency of our scheduling scheme. As shown in Fig. 2, which presents the channel frame structure, each frame consists of two types of intervals. These are the request interval and the information interval.

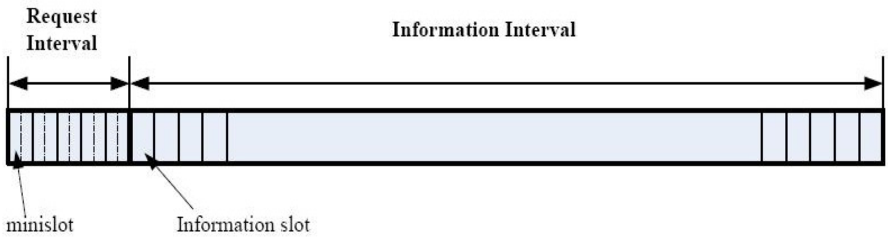

Fig. 2. Channel Frame Structure

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. We consider a 20 Mbps channel as a conservative estimation, given that next generation cellular networks are planned to have transmission rates exceeding 20 Mbps. We chose 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 those of the regular request slots.

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 non-collision (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 telemedicine traffic are transmitted first; their priority order will be explained in Section 4. The four types of regular traffic follow, with priority order: video, voice, email, and sms. 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 [12] (by email and sms terminals) to resolve contention. 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.

3.3 System State Transitions

As shown in Figure 3, 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 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 slot(s) until it exhausts its packets and returns to the silent state.

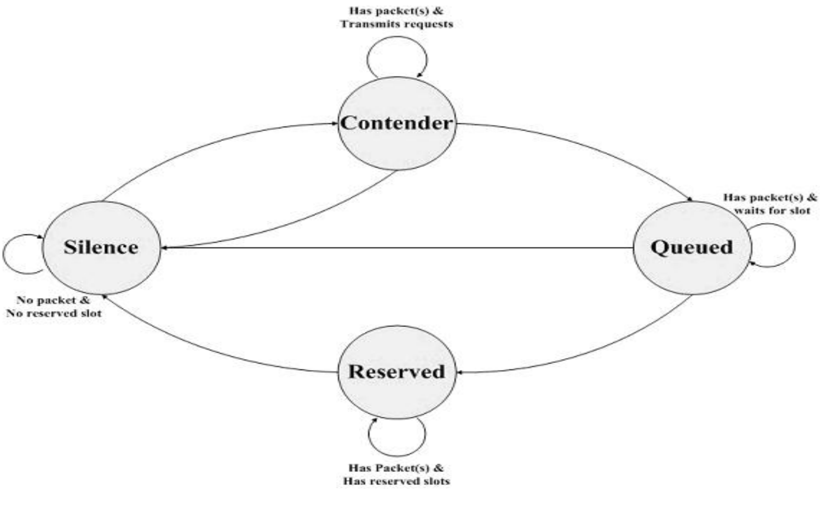

Fig. 3. State transition diagram for an active terminal

3.4 Channel Error Model

The most widely adopted wireless channel error model in the literature is the Gilbert-Elliot model [13, 14]. 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). A better choice for a more robust error model for wireless channels is the one we adopt in our study, and which was proposed in [15]. 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, as we can see in Figure 4. 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.

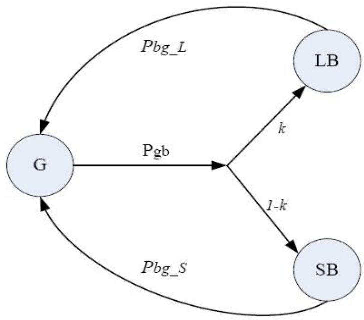

Fig. 4. Channel Error Model

4 The Proposed Scheduling Scheme

4.1 Introduction of the Scheduling Ideas

In our work the following design limitations have been adopted:

- A portion of the traffic arriving in each cell is handoff traffic from the other cells in the network. Handoff traffic is treated with full priority, with the use of the adaptive bandwidth reservation scheme which will be analyzed in Section 5.
- Video sources do not "live" permanently in the system, but have exponentially distributed sessions with a mean duration of five minutes [16]. This "relieves" a burden from the information interval of the channel, 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. A small percentage of the bandwidth (4.4% in our work, i.e., 25 slots out of the 566 slots of the channel frame) suffices to be used for requests. 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.

- 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. This new proposed mechanism aims at allocating as many of these slots as possible to other video terminals waiting for packet transmission, in order to decrease their transmission delay.

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 surpasses 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 hand-offed 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 directly 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 Scheduling Priorities and 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. 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, sms), based on the strictness of the QoS requirements for each traffic type (video and voice have the same QoS requirements of less than 1% packet dropping, but video traffic is much burstier, therefore it is granted priority over voice). Regular traffic originating from within the cell is transmitted last, with the same priority order.

We employ the two-cell stack reservation random access algorithm 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 sms terminals.

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 [7] 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:

$$B_i = M *(D_i / \Sigma_i D_i)$$

where D_i is the i^{th} user's requested bandwidth and $\Sigma_i D_i$ 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, 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, besides video, since they are allocated only one slot per frame.

4.4 FCFS – EDF – SJF Algorithms

As mentioned earlier, users of the same type of traffic are served in a FCFS order once they are admitted into the network. The First-Come-First-Served algorithm is the simplest scheduling algorithm. Processes are dispatched according to their arrival time on the ready queue. Jobs arriving are placed at the end of queue, the dispatcher selects the first job in the queue and this job runs to completion. The FCFS scheduling is fair in the human sense of fairness but it is unfair in the sense that long jobs make short jobs wait. For this reason, we wanted to compare its performance against two other well-known scheduling algorithms from the literature.

The Earliest Deadline First (EDF) algorithm operates on the logic that, among the users of one traffic class (e.g. voice), the user with the nearest deadline to transmit (a packet, message or video frame) will be accommodated first, instead of the user that arrived first (i.e. transmitted a request packet successfully in an earlier minislot). Earliest deadline first or "least time to go" is a dynamic scheduling algorithm used in real-time operating systems. It places processes in a priority queue. Whenever a scheduling event occurs (task finishes, new task released, etc.) the queue will be searched for the process closest to its deadline.

Shortest Job First (SJF), also known as Shortest Job Next (SJN), is a scheduling policy that selects the waiting process with the smallest execution time to execute next, and this process runs to completion. Shortest job next is advantageous because of its simplicity and because it maximizes process throughput (in terms of the number of processes run to completion in a given amount of time). We did not use SJF on voice users since we cannot predict a priori how long a conversation will last, while we know, e.g., the size of data messages or video frames.

4.5 Jain's Fairness Index

Aggregate performance evaluation metrics reveal very little, if any, information regarding the QoS of each wireless subscriber. Therefore, we should examine the impact of FCFS, EDF, SJF not only on commonly used performance metrics, such as throughput and delay, but also on fairness. We used Jain's Fairness Index [25] which is defined as follows:

$$J(x_1, \ldots, x_n) = \frac{\left(\sum_{k=1}^{n} x_k\right)^2}{n \sum_{k=1}^{n} x_k^2},$$

where n is the number of users and x_k is the throughput of user k. This index is continuous and bounded between 0 and 1, with 1 denoting maximal fairness. It is also very intuitive. If a ratio y of the users are treated fairly and (1 - y) are starved, then the resulting fairness index is y. We study fairness separately for telemedicine video users, regular video users and voice users. The throughput is measured in terms of the average number of allocated slots over 83 frames, which approximately corresponds to the average talkspurt duration (1 second); n is the number of users that were active during each 83-frame window (which is defined as a superframe).

5 Adaptive Bandwidth Reservation Based on Mobility and Road Information

5.1 Network and Mobility Models

It is a common assumption 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 [17,18]. 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.

We consider an architecture of seven hexagonal cells which we place in two different topologies. In the first topology, we assume the “circular” case, where after leaving the last cell (G) a user enters cell (A) again. In the second, "open" topology, mobiles can get in or out of our system only via cell A and cell G, which are placed at the two edges of our network model [23]. So, all cells are connected in a straight line, as shown in Figure 5.

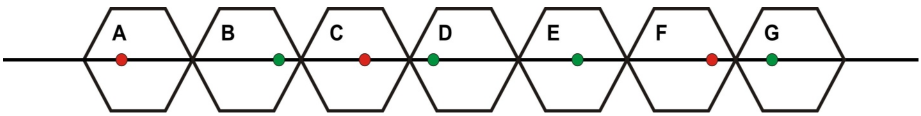

Fig. 5. Road Map and cellular network model

The cell diameter is 300 meters. One road, which is modeled by a straight line, passes through cells and connects them. 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 road. The initial location of a moving user on the 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. One traffic light is located randomly within each cell. A mobile user arriving at the traffic light of the cell might continue to go straight, or turn around with probabilities 0.9 and 0.1, respectively. If a mobile user chooses to go straight at the traffic light, it needs to stop there with probability 0.5 for a random time between 0 and 30 seconds due to a red traffic light, or else passes with probability 0.5. If the user chooses to turn around, it needs to stop there for a random time between 0 and 60 seconds due to the traffic signal [19, 20]. 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), moving direction, and speed, and can be provided with an accuracy of 1m through GPS [19, 20, 21].

5.2 Adaptive Bandwidth Reservation Scheme

In our study, we adopt the idea that mobile stations only need to update their position information to the BS of their cell when they arrive at a traffic light. If the BS of the current cell of a mobile station predicts, based on the station's location (at a traffic light) and speed that the station is going to move to another cell, 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 traffic light. 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.

6 Results and Discussion

6.1 Simulation Setup

We use computer simulations to study the performance of our scheme. 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 scenarios, where one type of telemedicine traffic might be more dominant than others in any given moment. Each simulation point presents the average result over 5 combinations which were used to create a specific traffic load, and for each combination we conducted 5 independent runs, each simulating 155000 frames (1/2 h of network operation), the first 5000 of which are used as warm-up period.

6.1.1 FCFS Results

We present below the first part of our results, in Figures 6-8. These figures include results derived with the use of the FCFS (First Come First Served) algorithm, for the "circular" topology.

In Figure 6, the improvement shown in the average video packet dropping is due to the proportionate allocation; 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.

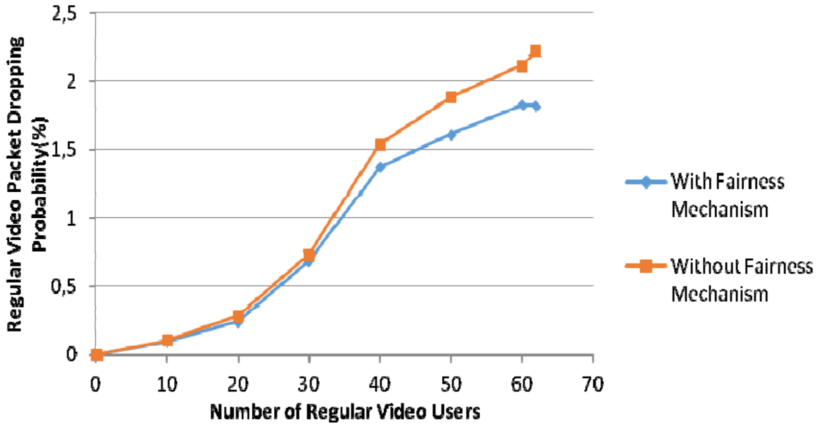

Fig. 6. Regular video packet dropping versus the number of regular video users

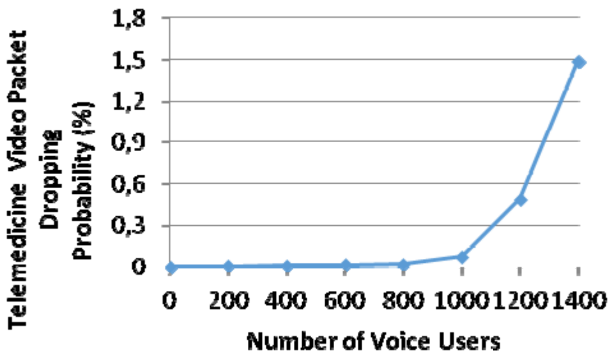

Fig. 7. Effect of regular voice traffic on telemedicine video traffic

Figure 7 shows the almost negligible effect on the QoS of telemedicine users that stems from the increase in the number of voice users. We present results on the telemedicine video packet dropping and it is clear that only in the case of very high voice loads there can be deterioration in telemedicine traffic QoS. The reason is that our combined scheduling and adaptive bandwidth reservation schemes guarantee full priority to all types of telemedicine traffic.

In Figure 8 the percentage of telemedicine video users whose QoS requirements for packet dropping (< 0.01%) are violated is much lower with the use of our fairness mechanism. The reason for this result is that our schemes are designed to offer maximum priority to telemedicine traffic; therefore, when telemedicine video packet dropping increases, the increase is almost "uniform" for all telemedicine video users.

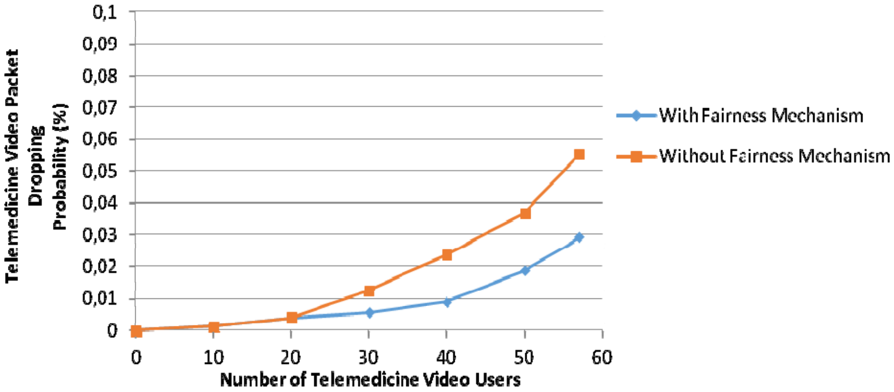

Fig. 8. Telemedicine video packet dropping versus the number of telemedicine video users

6.1.2 FCFS-EDF-SJF Results

Figures 9-10 present results derived with FCFS (First Come First Served), EDF (Earliest Deadline First) and SJF (Shortest Job First) algorithms. For these results we used the same "scenarios" for channel load as those that we used for the corresponding figures in the previous section.

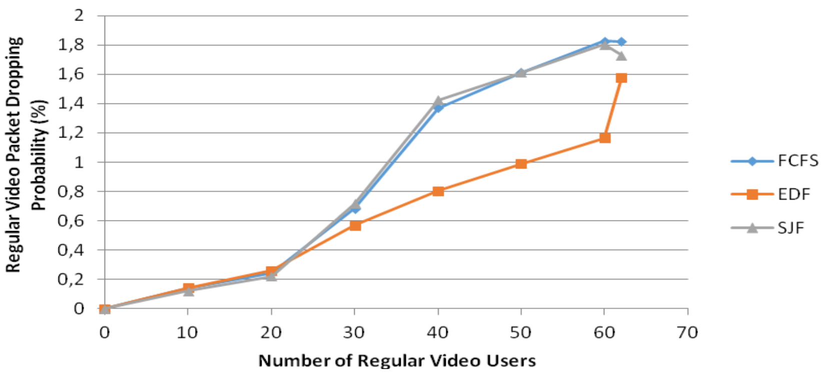

Fig. 9. Regular video packet dropping versus the number of regular video users

Figure 9 shows that the increase in the number of regular video users clearly affects that type of traffic, because the video packet dropping probability of regular video users significantly rises above the 1% acceptable upper bound. Initially, the three algorithms have almost the same efficiency. However, when the number of users in the system (i.e., over all seven cells) exceeds 20, the EDF algorithm produces significantly better video packet dropping probability results for regular video users, with FCFS being marginally better than SJF overall.

Figure 10 shows that the three queuing priority algorithms produce almost identical results for low and medium telemedicine video loads, in terms of regular and telemedicine video packet dropping. EDF excels once again in the case of high loads.

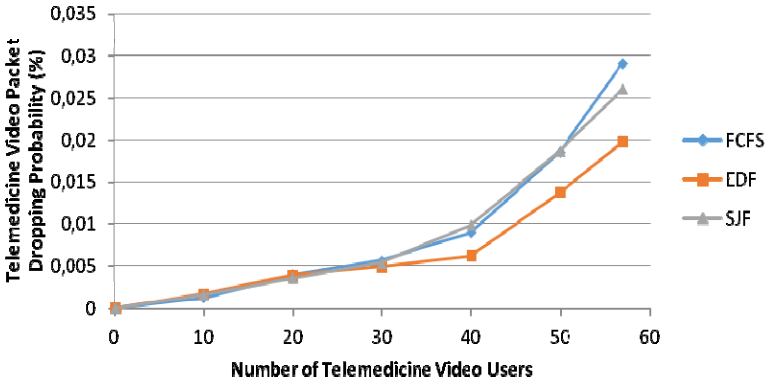

Fig. 10. Telemedicine video packet dropping versus the number of telemedicine video users

It is clear from the figures above that EDF excels in comparison to the other two algorithms. The reason is that it accommodates users based on their deadline, hence it manages to satisfy their QoS requirements in time. The FCFS and SJF algorithms base their respective policies on the arrival time and the size of new information, thereby ignoring the urgency of users' needs. The more aggressive policy implemented by EDF, of course, has, theoretically, the disadvantage that it can lead to unfairness in specific metrics; users with later deadlines can experience longer delays than they do, e.g., with the use of FCFS. For this reason, in Section 6.5 we will evaluate the fairness offered to regular and telemedicine users by each of the three scheduling algorithms. Finally, we need to note that, although FCFS and SJF have comparable results for almost all traffic loads, in very high load conditions FCFS performs always worse than SJF because it fails to offer any kind of priority based on the users' needs.

6.1.3 Open Topology Results

In this section we indicatively present results derived for the second topology that we studied. All cells are connected in a straight line and mobiles can get in or out of our system only via cell A and cell G, which are placed at the two edges of the network.

Figure 11 illustrates the increase of telemedicine video packet dropping probability, as we increase the number of telemedicine video users. It is useful to compare and contrast Figure 11 to Figure 10, where we considered the "circular" topology.

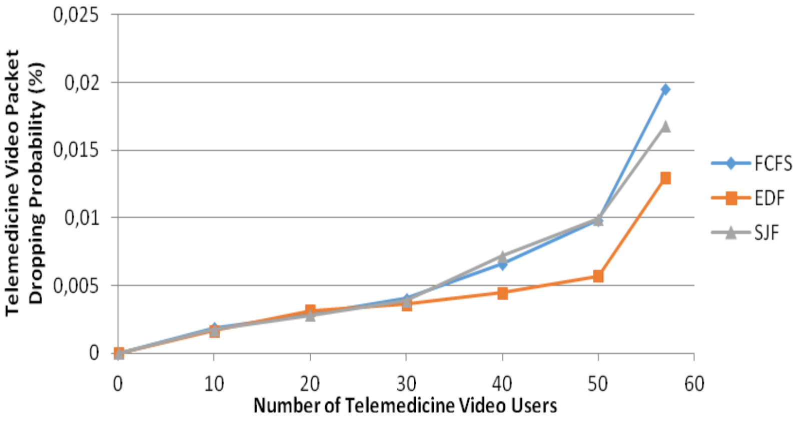

Fig. 11. Telemedicine video packet dropping versus the number of telemedicine video users

Comparing Figures 10 and 11 we note that the results are similar in nature, but the quantitative difference is large. Using the circular topology, average telemedicine video packet dropping probability was 0.03% with FCFS, 0.026% with SJF and 0.02% with EDF, while with the second map the corresponding values are 0.02%, 0.017% and 0.0128%.

Hence, in the case of the open topology, regular video packet dropping probability and telemedicine video packet dropping probability are much lower. The reason is that when the seven cells are connected in a circular fashion, users never leave the system; hence, their aggregate number steadily grows, considering that new users keep arriving in the system. On the contrary, when users can leave the system via cells A and G, the aggregate users' number fluctuates. This explains the worse QoS achieved by all algorithms in the case of the circular topology; this case actually represents a worst-case scenario for our system.

6.1.4 Jain's Fairness Results

Below, we present results derived with 3 different combinations of all loads of users, ranging from very low (9%) to very high (100%) and we present the throughput fairness results with the use of Jain's fairness index for voice, video and telemedicine video users.

Table 1. 9% traffic load

User's type/Algorithm	***FCFS***	***EDF***	***SJF***
Voice Fairness	0.9510	0.9513	-
Video Fairness	0.9978	0.9981	0.9976
Telemedicine Video Fairness	1.0000	1.0000	1.0000

Table 2. 51% traffic load

User's type/Algorithm	*FCFS*	*EDF*	*SJF*
Voice Fairness	0.9476	0.9470	-
Video Fairness	0.9868	0.9889	0.9859
Telemedicine Video Fairness	0.9963	0.9961	0.9965

Table 3. 100% traffic load

User's type/Algorithm	*FCFS*	*EDF*	*SJF*
Voice Fairness	0.9838	0.9925	-
Video Fairness	0.9876	0.9901	0.9866
Telemedicine Video Fairness	0.9920	0.9931	0.9902

The results presented in Tables 1-3 show that EDF outperforms FCFS and SJF in terms of fairness in throughput, something intuitively explained due to the nature of the algorithm (packets with imminent deadlines are transmitted first, therefore the smallest possible number of packets is dropped). FCFS outperforms SJF, which creates some unfairness by always servicing first the users with the least packets to transmit, and hence leading larger transmissions to possible packet dropping. Still, the use of throughput as a fairness metric does not suffice; delays and packet dropping are of equal importance, therefore we used them in our study and we present the results below.

We present two more cases, one with a high load (80%) and one corresponding to a traffic overload (110% of the channel capacity) and we evaluate fairness in terms of delay and packet dropping.

Table 4. 80% traffic load

	DELAY			*PDROP*		
User's type/Algorithm	*FCFS*	*EDF*	*SJF*	*FCFS*	*EDF*	*SJF*
Voice Fairness	0.9369	0.9327	-	0.9138	0.9094	-
Video Fairness	0.9799	0.9753	0.9226	0.9338	0.9203	0.8248
Telemedicine Video Fairness	0.9906	0.9893	0.9317	0.9893	0.9824	0.8465

Table 5. 110 % traffic load

	DELAY			*PDROP*		
User's type/Algorithm	***FCFS***	***EDF***	***SJF***	***FCFS***	***EDF***	***SJF***
Voice Fairness	0.8977	0.8813	-	0.8948	0.8723	-
Video Fairness	0.8858	0.8552	0.7912	0.6937	0.6779	0.6219
Telemedicine Video Fairness	0.9727	0.9788	0.8864	0.9551	0.9501	0.8065

For the results presented in Tables 4-5, we have kept constant the number of voice users and we increased the number of video and telemedicine video users. These Tables clearly show that as we increase the channel load, the fairness achieved by all three algorithms decreases significantly, especially when considering packet dropping as a metric. The traffic type that is influenced the most is regular video traffic, as voice is less demanding in bandwidth, and telemedicine video users have absolute priority in scheduling, therefore they are minimally affected. As intuitively expected and explained in Section 6.2.3, EDF does not outperform FCFS when delay and packet dropping are used as fairness metrics.

It is clear, however, from all the results presented in Tables 1-5, that our scheme achieves excellent fairness results for all traffic types and for all traffic loads that do not exceed the maximum channel capacity.

7 Conclusion and Future Work

This work has focused on the problem of scheduling integrated traffic transmissions from urgent types of traffic, like telemedicine, with regular wireless traffic over next generation cellular networks. We have extended a recent work [22] on a new MAC protocol by using three different scheduling algorithms over a simple network topology and by evaluating the algorithms' fairness based on a number of metrics. Our results have clearly shown that in terms of the provided QoS to wireless users, the EDF algorithm excels over SJF and FCFS, which has been widely used in the literature as it is the intuitively simplest choice and is marginally fairer than EDF.

In future work, we intend to experiment with more scheduling algorithms from the literature and to propose an algorithm of our own, which will incorporate the advantages of EDF and will provide increased fairness in comparison to FCFS. In order to achieve this, we believe that we will need periodic bandwidth reallocation from the BS to the wireless users, based on efficient traffic modeling.

References

1. Perednia, D.A., Allen, A.: Telemedicine Technology and Clinical Applications. The Journal of the American Medical Association (JAMA) 273(6), 483–488 (1995)
2. Chu, Y., Ganz, A.: A Mobile Teletrauma System Using 3G networks. IEEE Transactions on Information Technology in Biomedicine 8(4), 456–462 (2004)

3. Bhargava, A., Khan, M.F., Ghafoor, A.: 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)
4. Tulu, B., Chatterjee, S., Laxminarayan, S.: 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)
5. Voskarides, S.C., Pattichis, C.S., Istepanian, R., Michaelides, C., Schizas, C.N.: 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)
6. Garawi, S., Istepanian, R.S.H., Abu-Rgheff, M.A.: 3G Wireless Communications for Mobile Robotic Tele-Ultrasonography Systems. IEEE Communications Magazine 44(4), 91–96 (2006)
7. Koutsakis, P., Psychis, S., Paterakis, M.: Integrated Wireless Access for Videoconference from MPEG-4 and H.263 video coders with Voice, Email and Web traffic. IEEE Transactions on Vehicular Technology 54, 1863–1874 (2005)
8. Pang, Q., Bigloo, A., Leung, V.C.M., Scholefield, C.: Service Scheduling for General Packet Radio Service Classes. In: Proceedings of IEEE Wireless Communications and Networking Conference (1999)
9. Nanda, S., Goodman, D.J., Timor, U.: Performance of PRMA: A Packet Voice Protocol for Cellular Systems. IEEE Transactions on Vehicular Technology 40, 584–598 (1991)
10. Fitzek, F.H.P., Reisslein, M.: MPEG-4 and H.263 Video Traces for Network Performance Evaluation. IEEE Network 15(6), 40–54 (2001)
11. Tsai, C.-F., Tsang, C.-J., Ren, F.-C., Yen, C.-M.: Adaptive Radio Resource Allocation for Downlink OFDMA/SDMA systems. In: Proceedings of the IEEE International Conference on Communications, ICC (2007)
12. Cleary, A.C., Paterakis, M.: 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)
13. Gilbert, E.N.: Capacity of a Burst-noise Channel. Bell System Technical Journal 39, 1252–1265 (1960)
14. Elliot, E.O.: Estimates of Error Rates for Codes on Burst-noise Channels. Bell Labs Technical Journal 42, 1977–1997 (1963)
15. Bottigliengo, M., Casetti, C., Chiasserini, C.-F., Meo, M.: Short-term Fairness for TCP flows in 802.11b WLANs. In: Proceedings of IEEE Infocom (2004)
16. Baig, E.C.: Will Consumers Tune in to a Tiny TV in their hand (2006), `http://www.usatoday.com/tech/wireless/2006-08-17-mobile-tv_x.htm`
17. Yu, F.R., Wong, V.W.S., Leung, V.C.M.: A New QoS Provisioning Method for Adaptive Multimedia in Wireless Networks. IEEE Transactions on Vehicular Technology 57(3), 1899–1909 (2008)
18. Song, Q., Jamalipour, A.: A Negotiation-Based Network Selection Scheme for Next-Generation Mobile systems. In: Proceedings of the IEEE Globecom (2006)
19. Choi, S., Shin, K.: Adaptive Bandwidth Reservation and Admission Control in QoS-sensitive Cellular Networks. IEEE Transactions on Vehicular Technology 13, 882–897 (2002)

20. Lee, D.-S., Hsueh, Y.-H.: Bandwidth-reservation Scheme based on Road Information for Next-generation Cellular Networks. IEEE Transactions on Vehicular Technology 53, 243–252 (2004)
21. Soh, W.S., Kim, H.S.: A Predictive Bandwidth Reservation Scheme using Mobile Positioning and Road Topology Information. IEEE/ACM Transactions on Networking (TON) 14, 1078–1091 (2006)
22. Qiao, L., Koutsakis, P.: Adaptive Bandwidth Reservation and Scheduling for Efficient Wireless Telemedicine Traffic Transmission. IEEE Transactions on Vehicular Technology 60(2), 632–643 (2011)
23. Kwon, T., Choi, Y., Bisdikian, C., Naghsineh, M.: QoS Provisioning in Wireless/Mobile Multimedia Networks Using an Adaptive Framework. Wireless Networks 9(1), 51–59 (2003)
24. ETSI. Digital Cellular Telecommunications System (Phase 2+); Technical Realization of the Short Message Service (SMS); Point-to-Point (PP) (GSM 03.40)
25. Jain, R.: The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. Wiley- Interscience, New York (1991)