Real-Time Systems, 21, 269–284, 2001

Quality of Service Support for Multimedia Applications in Third Generation Mobile Networks Using Adaptive Scheduling

PAUL RICHARDSON
Department of Electrical and Computer Engineering, University of Michigan (Dearborn)

LARRY SIEH
Vetronics Technology Center, U.S. Army Tank-Automotive RD&E Center

AURA GANZ
Department of Electrical and Computer Engineering, University of Massachusetts (Amherst)

Abstract. Third generation mobile network will support services such as video-telephony, video-conferencing and other multimedia applications. Therefore, this network must provide quality of service (QoS) to these applications consistent with that offered by fixed networks. However, this is a very challenging task due to the instability of the wireless channel and the diverse quality of service requirements dictated by different multimedia applications. In this paper we introduce a resource allocation algorithm for the wireless downlink that takes into account the wireless channel characteristics, the QoS required by the applications as well as a pricing value function. Our solution is based on an adaptive scheduling algorithm originally developed for scheduling real-time processes during transient surges. This algorithm tends to maximize the wireless network operator profit while satisfying the customers' quality requests.

Keywords: mobile computing, real-time communication, robust communications, adaptive scheduling, third generation mobile systems, voice transmission, quality of service support, multimedia applications and wireless networks

1. Introduction

The rapid expansion of wireless networks is making it possible for mobile users with portable computers or personal digital assistants (PDAs) to access information anywhere and at any time. The number of wireless terminals with voice, data, and video applications is growing rapidly. It is inevitable that these users should be integrated with the Internet. Therefore, it is important that the quality model in wireless networks be consistent with that of the Internet.

The next generation of services will include a price tag for premium services [1, 2]. Each application will have a service contract with the carrier (of course default values will be used as taken from the user profiles). The service contract, consistent with the Internet differential or integrated services will specify the quality of service (QoS) expected by the application and the price the user is willing to pay for this service [1, 2]. Moreover, the wireless operator needs to be ready to reimburse the users if the QoS contract cannot be met.

One lesson that the cellular network providers have learned is that effective radio resource management is essential to promote quality and efficiency in the system. This resource management will become critical in systems that include data and video in addition to voice traffic.

The admission control policy negotiates with the users on a specific QoS service contract that includes network support parameters such as bandwidth and time delays, as well as a revenue model (price that the user pays for the session). The problem of satisfying such a QoS contract in a wireless network is the instability of the wireless link that can have transient blackouts. These blackouts can overload the network, forcing packets to be late, i.e., violating the QoS contracts agreed by the admission control module. Under these conditions, conventional scheduling approaches cannot provide QoS support for real-time applications such as voice or video inter-mixed with traditional WWW traffic.

This work focuses on developing a new framework for radio resource management on the wireless downlink from the backbone to the mobile terminals. The proposed approach relies on using adaptive real-time scheduling techniques that take into account the characteristics of the wireless channel as well as the QoS contracts for each traffic stream (session). In this paper we assume that the QoS parameters for real-time streams such as voice traffic are defined in terms of maximum delay tolerated per packet as well as maximum packet loss probability. Moreover, we also assume that each session has an associated reward which is obtained when the QoS contract is satisfied and an associated penalty when this contract is violated. To capture all these parameters we will implement a wireless adaptive scheduling algorithm (WASA). WASA mitigates network overloads caused by the wireless channel interference by ordering the transmission of packets based on sessions' state and user-requested QoS. We monitor the network non-intrusively to detect these overloads. WASA uses a value-based model [3, 4] to order packet transmissions during overloads. The goal of a value model is to accurately capture the reward or penalty of meeting or missing the QoS contract (i.e., time constraints). Our simulation results show that for specific channel and traffic characteristics, the proposed adaptive scheduling approach provides more than 20% higher revenues as compared with the earliest deadline first (EDF) approach and more than 100% higher revenues when compared to FIFO.

The proposed approach will enable wireless service providers to support a wide range of applications as well as personalized service per user, significantly increasing their revenues.

The paper is organized in the following way. In the next section we provide an overview of the wireless network architecture. Section 3 describes the adaptive scheduling algorithm. Section 4 describes the simulation model we used and the numerical results we have obtained. Section 5 concludes the paper.

2. Wireless Network Architecture

The 3G system that we investigate follows an UMTS architecture [5–8] as depicted in Figure 1. The architecture consists of the mobile terminals (MTs), the base transceiver

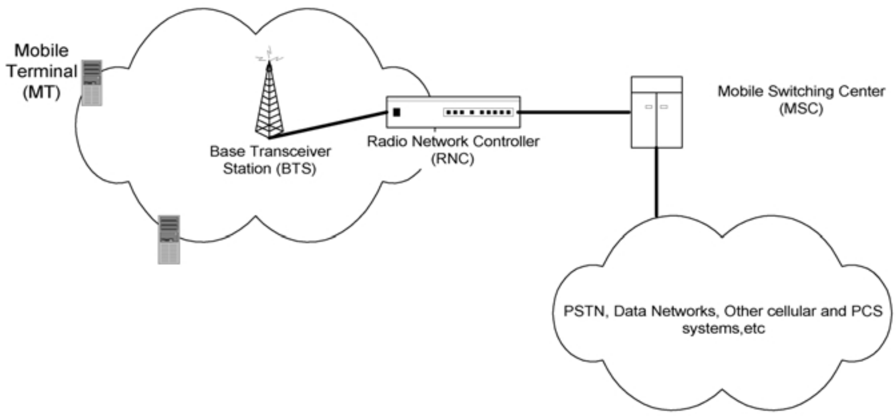


Figure 1. Depicts the UMTS 3G wireless network architecture.

station (BTS) and the control nodes such as the radio network controller (RNC) that has the responsibility of managing the radio resources. The RNC and BTS are connected through a high-speed transport infrastructure.

The main functions of the RNC:

1. Interfaces between the wired and wireless segments. It converts between the packet formats.

2. Controls the wireless uplink and downlink activities on the wireless segment as well as represents the wireless segment in the QoS end-to-end contract negotiations.

The proposed WASA will reside in the RNC (see Figure 2) and will manage the downlink between the RNC and the MTs. The packets arriving from the backbone are stored at the RNC. The module schedules the packets for transmission on the downlink based on the applications' QoS requirements and channel characteristics. Through implicit or explicit probes, WASA obtains information on the wireless channel.

3. The Proposed Wireless Adaptive Scheduling Algorithm

We consider services for both voice and data transmissions from Internet web browsing. The voice packets have firm deadlines equal to their inter-arrival time (e.g., 20 ms). Firm deadlines are those in which some late deliveries can be tolerated, but once late, an individual packet has no further value. Data packets for web browsing are assumed to be TCP/IP packets and they are given soft deadlines (e.g., 3000 ms), which reflects the estimated delay at which point a web browser will start to become irritated. Packets with soft deadlines have tolerance for late deliveries and individual packets retain a reduced value once they are late.

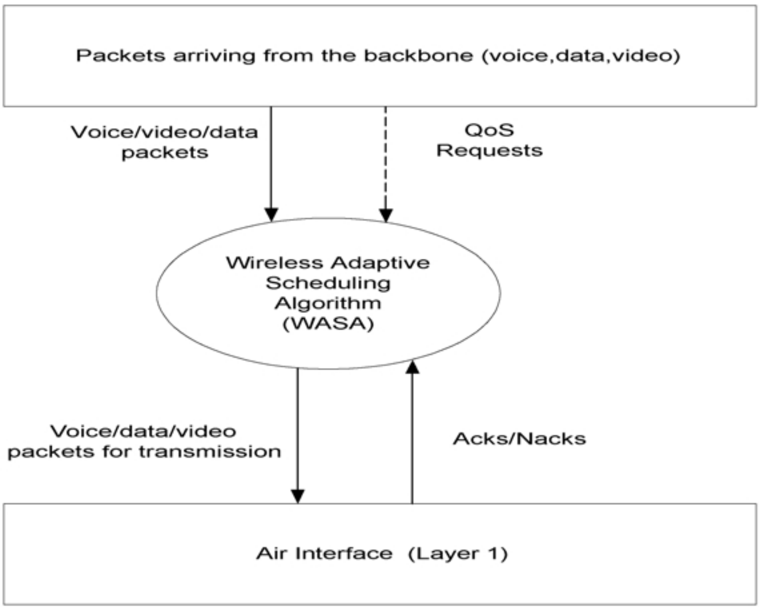


Figure 2. Block diagram of the RNC-MS downlink, showing the WASA module at the RNC.

Obviously when channel deterioration occurs, no scheduling policy can guarantee that every deadline is met. So the real question is how the scheduling policy can handle the transient channel deterioration while still providing the QoS to specific high return sessions.

Earliest deadline first (EDF) [9] is a preemptive deadline-based scheduling policy that can guarantee the time constraints for a set of real-time processes under certain conditions. EDF is optimal in the sense that if any scheduling policy can guarantee the time constraints for a set of real-time processes, then EDF can as well. Under nominal channel conditions, assuming the use of an appropriate admissions control policy, EDF can efficiently guarantee all requested QoS support. However, EDF performs poorly in case of channel overloads. This is due to its inability to dynamically adapt to deteriorating channel conditions.

Therefore, in this paper, during deteriorating channel conditions, we use value-based scheduling, which is able to order packet transmissions based on QoS support requested by users. In Jensen et al. [4] and Jensen [10], Jensen describes value-based models that more accurately represent the cost and benefit of missing or meeting time constraints. Jensen's model considers the importance of events as a function of time. In Kilkki [3], Brown modifies this approach to account for system state, and develops a fuzzy priority scheduler, to schedule events under overload conditions. Richardson and Sarkar [11] present an adaptive scheduling approach that is the predecessor to the proposed algorithm.

In the remainder of this section, we describe the WASA that is able to support QoS requests under nominal and deteriorated channel conditions. Under nominal channel conditions, WASA uses the EDF scheduling policy to order packet transmissions. Under these conditions, assuming the appropriate admissions control, all QoS requests can be guaranteed. When channel conditions deteriorate, WASA uses value-based scheduling to assign priorities to packet transmissions. The objective under these conditions is to provide quality service to high return sessions during transient channel overloads by adapting the packet priorities as a function of the session state. The session state can reflect the fact that packets are lost on the wireless channel, e.g., the percentage of lost packets over a time window, or it can capture the fact that consecutive packets are lost in a burst. Each session is assigned a value function as determined by the session QoS parameters such as reward. The changes in the value function will occur when the channel deteriorates and will be determined by the QoS contract, the current time relative to the packet deadline, and the current session state. During transient overloads, each packet's priority will be determined by its value function.

Our value-based scheduling algorithm will accept any value function defined by these parameters. Currently, there are a number of efforts in the Internet community to define pricing as a function of QoS [1, 2] which will eventually be translated to the wireless operators value function.

3.1. Value Model and Cost Assignment to Sessions

We define three levels of the value function: the system value function V, per session value function V_i and per packet, p_{ij}, value function $v_i(j)$ defined as follows:

1. $v_i(j) \equiv$ the value function per packet p_{ij}. When a scheduling event occurs at time t, a value, $v_i(j)$ is assigned to the jth packet transmission.

2. $V_i \equiv$ session S_i's value function that is composed of the mean individual packets value functions. For a session with n packet transmissions we have

$$V_i = \frac{1}{n}\sum_{j=1}^{n} v_i(j) \tag{1}$$

3. $V \equiv$ the system value function reflects the reward/penalty obtained by the system operator for (1) a given scheduling policy, (2) sessions' characteristics (per session traffic characteristics and QoS requirements), and (3) channel conditions. For m sessions we have

$$V = \sum_{i=1}^{m} V_i \tag{2}$$

If the net value function for a session is positive, then there is a reward for the session, if it is negative, there is a net penalty (refund) for the session. While the user is allowed to specify and pay for desired levels of service in the QoS contract, these specifications must be restricted to prevent a greedy user from disrupting network performance.

The value function $v_i(j)$ assigned to packet p_{ij} is determined by four parameters. The weight parameter, denoted by w_i, establishes bounds for the maximum reward and greatest penalty for each packet transmission. A packet transmission can have the greatest reward of w_i, if it is on-time and the largest penalty of $-w_i$, if it is exceedingly late. The lateness parameter, $l_i(t)$, captures the impact of tardiness on a packet transmission for values between $[+w_i, -w_i]$. The parameters dw_i and dl_i are used to alter the weight and lateness parameters, w_i, and $l_i(t)$, respectively, to capture changes in session state caused by the variations in the channel quality. The QoS contract specifies C_i and P_i, the bounds for cost and penalty (refund) associated with the quality of the session. The maximum reward provided by the user if the contract is fully met is specified by C_i. The maximum penalty assessed against the service provider in the case the contract is violated is P_i.

The initial and maximum value of w_i is C_i, which reflects the reward associated with delivering the proper QoS to the user. The minimum value of w_i is P_i which is the worst case penalty that must be refunded to the user if the contract is violated. The impact of the value function parameters on the cost or penalty (refund) of a session and their affect on performance evaluation and priority assignment are described below.

The value assigned to p_{ij}, the j-th packet in session S_i at time t assuming t_{dj} is its deadline is given by

$$\begin{aligned} \text{if } t \leq t_{dj} \text{ then } v_i(j) &= w_i \quad (p_{ij} \text{ is timely}) \\ \text{otherwise} \qquad v_i(j) &= \min(w_i, w_i(1 - l_i(t))) \quad \text{if } w_i(1 - l_i(t)) \geq 0 \\ &= \max(-w_i, w_i(1 - l_i(t))) \quad \text{if } w_i(1 - l_i(t)) < 0 \end{aligned} \tag{3}$$

The initial value of $l_i(t)$ is determined by the characteristics of the service request. For example, voice packet transmissions have firm deadlines and no value once they are late. Thus $l_i(t)$ for these packets would be ∞, indicating a step function decline in their value once they are late. Data packets for web browsing on the other hand, have soft deadlines, that reflect a users willingness to wait for services. Since these packets retain value after they are late, $l_i(t) < \infty$. For data sessions, which have soft deadlines, the QoS class chosen by the user will specify $l_i(t)$ and the deadline which in turn determines the rate the value will increase under deteriorated channel conditions.

The value function, $v_v(j)$ for a voice packet in session S_v, is shown in Figure 3(left). The weight parameter, w_v is established by the QoS contract and can differ for different line qualities. In this paper the model for voice packets specifies a firm deadline of 20 ms, hence, $l_v(t) = \infty$, which causes the value of p_{ij} to drop to $-w_v$ once it is late. Mobile customers desiring differing level of services can specify and pay for differing amounts which will be reflected by C_v, P_v, and dw_v.

The value function, $v_d(j)$ for a data packet in WWW session S_d, is shown in Figure 3(right). As with voice, w_d is established by the QoS contract and can differ for different line qualities. As described earlier, the lateness parameter, $l_d(t)$, will be determined by the QoS class chosen by the user. The deadline determines at which point the value function will begin to decline if the packet has not been successfully transmitted. The specification of $l_d(t)$ determines how fast $v_d(j)$ declines after its deadline, reaching its maximum penalty, $-w_d$, at some u_d.

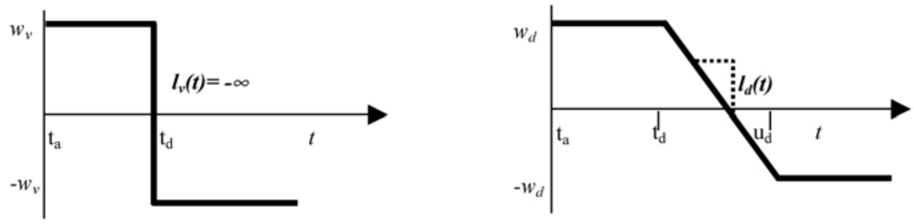


Figure 3. Value function for a voice session (left) and for a WWW session (right).

The intent of using parameters dw_i and dl_i is to capture the changes in session state caused by varying channel conditions. These changes will be determined by the QoS contract agreed with the user, i.e., the packet loss sensitivity of a specific session and its impact on the session price/reward. For example, an occasional packet transmitted late has some negative impact on session state, although it may be minimal. Whereas a large number of consecutive packets transmitted late has a correspondingly worse impact on session state, and the QoS delivered. The parameter dw_i is used to modify the largest penalty, $-w_i$, to reflect changes in session state caused by varying channel conditions.

If p_{ij} is delivered late, then the largest penalty for $p_{i,j+1}$, becomes

$$-w_i = \max(-P_i, -w_i - dw_i)$$

For each timely delivery, p_{ij}, the largest reward for p_{ij+1}, becomes

$$w_i = \min(C_i, w_i + dw_i)$$

The example depicted in Figure 4 captures the effect of modifying w_i on the value function. Initially, the parameters of the value function are w_{i0} and l_{i0}, after late transmissions, the parameter w_{i0} is modified by dw_i to reflect the impact of late deliveries on session state. In a similar manner $dl_i(t)$, is used to alter $l_{i0}(t)$, to vary the rate at which the value function declines.

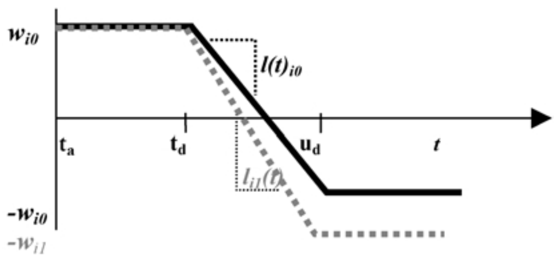


Figure 4 Capturing the effects of session states modifications on the value function. The parameters w_{i0} and $l_{i0}(t)$ are modified to w_{i1} and $l_{i1}(t)$ to reflect the impact of late arrivals to session state.

3.2. Assignment of Priorities to Packet Transmissions

The assignment of priorities to packet transmissions proceeds as follows: assume that we have the j-th packet of session S_i with deadline t_{dj}, at time t.

We differentiate between two cases:

Case 1 Ideal Channel Conditions

Under ideal channel conditions, EDF is used to schedule packet transmissions, thus the closer a packet comes to being late, the more important it becomes (smaller values are more important):

$$\text{priority } p_{ij} = t_{dj} - t \tag{4}$$

Case 2 Deteriorated Channel Conditions

During channel deterioration, the value function is used to assign priorities to packets in danger of being late. Once a packet is in danger of being late, the priority assignment is based on the weight and lateness parameters of the session's value function to capture the packets importance and session state.

At some threshold time, $t_{dj} - k_j$, when a packet is in danger of being late, if $t \leq t_{dj} - k_j$ then priorities are assigned by

$$\text{priority } p_{ij} = t_{dj} - k_j - t \quad \text{(EDF scheduling)}, \tag{5a}$$

otherwise $(t > t_{dj} - k_j)$ priority assignment corresponds to

$$\text{priority } p_{ij} = \max(-w_i, w_i(1 - l_i(t))) \tag{5b}$$

Figure 5 depicts the assignment of priorities for voice and data packet transmissions for this case. An optimal value of k_i was determined experimentally to be 2.5 times the packet transmission time.

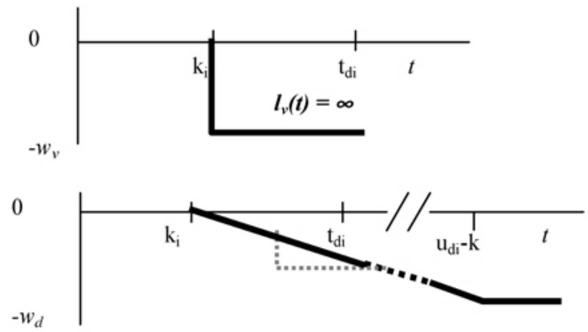


Figure 5. Packet priority assignment for voice packets (top) and WWW data packets (bottom).

The ordering of packets by their value functions results in packets being ordered by their importance and session state with respect to time. As we will see in the simulation, this offers considerable advantage over the classical EDF deadline scheduling policies during overloads.

It is worth noting that when even a single session enters WASA other sessions that are scheduled by EDF are disturbed and optimality based on EDF criteria is violated. However, we need to remember that the EDF scheduling algorithm is optimal strictly from a time-based perspective, i.e., minimizes maximum lateness. In this paper we deal with optimization based on relative value of packets during overloads, i.e., maximize reward. This optimization based on value results in a disturbance to the time-based optimality. During network overloads, EDF still tends to minimize maximum lateness and is limited in the sense that it does not consider the relative value of packet transmissions. Conversely, WASA recognizes the wide variance in the value of different packet transmissions and attempts to optimize based on value and time. Notice that under nominal conditions, EDF and WASA are both time-driven and synonymous.

4. Simulation and Numerical Results

A simulation was performed to verify the performance of the proposed WASA on the RNC-Mobile Host downlink. For comparison, the classical EDF scheduling policy and FIFO scheduling policy were simulated under identical conditions.

Two types of services were chosen to be representative of third generation mobile service requirements, encoded voice, and Internet web browsing. They embody the combined voice and data requirements of future systems.

It is assumed that for maximum usage efficiency, voice will be compressed. An encoding system used in CDMA systems [8] will be assumed. A four-state stochastic model is used to model the varying lengths (2, 5, 10, 22 octets) of the voice packets. The packet sizes depend on the nature of the speech at any given instant. One packet is transmitted every 20 ms, with an average of 8.34 octets/frame (without overhead), and a data rate of 417.3 octets/s (3339 b/s).

Web browsing, presently the most popular data service, is used to represent data service requirements. A three-state Markov arrival model [8] is used, with the states representing ''sending'', ''active'', and ''sleeping''. Sending corresponds to the burst of activity as a user downloads a web page with all of its constituent files. Active implies that a user is actively engaged in a browsing session, while sleeping corresponds to a user taking a break from browsing. Each state implies a hyper-exponential distribution of arrival times between TCP connections, with means of 0.3, 30, and 4000 s, respectively. Each TCP connection implies the transmission of one file, whose length is modeled by an Erlang distribution with a mean of 2500 bytes.

The RNC-MT link was assumed to have a bandwidth of 64 kbps, a frame overhead of 64 bits per frame, and a variable frame size with a maximum allowable frame size of 512 bits of payload data. We simulate a total of six voice streams $(S_1, \ldots, S_6)$ and four Internet web (WWW) browsing data streams $(S_7, \ldots, S_{10})$.

Value functions were assigned to each packet stream to reflect possible user QoS requests. Voice streams S_1 and S_2 are high value voice streams that reflect a need to have a high QoS under degraded wireless channel conditions. Users are charged a premium for this grade of service. Voice streams S_3 and S_4 are medium value voice streams. These voice streams have value functions whose values will be less than that of high value voice streams, but will still have significant emphasis placed on them when channel blackouts occur as the result of deteriorating channel conditions. Charges for this grade of service will be somewhat less than for high value voice streams. Voice streams S_5 and S_6 are low value voice streams that reflect a user desire to have an economical voice service that receives no consideration if the wireless channel deteriorates. Data streams S_7–S_{10} are low value data streams. The deadlines, value function parameters $(w, dw, l(t))$, mean packet lengths, and total packet arrivals for these streams are summarized in Table 1.

The duration of the simulation was for 300 s of simulated time. We have run two separate simulations for WASA, EDF, and FIFO under non-ideal channel conditions. In these runs, deteriorated channel conditions were simulated with an exponentially distributed arrival rate with a mean of 1000 ms and an exponentially distributed duration with means of 100 and 200 ms, respectively.

We have collected the following metrics:

1. Percentage of dropped packets (for voice sessions) or late packets (for WWW sessions).

2. Per session value metric for each scheduling approach, V_i.

3. System total value metric for each scheduling approach, V.

Figures 6(a) and (b) depict the per session percentage of dropped packets (for WASA, EDF, and FIFO) for blackouts with mean duration of 100 and 200 ms, respectively. We clearly observe that the WASA policy results in the lowest dropping rate of high value sessions such as S_1–S_4, with 90% or better of them arriving on time for the blackouts with

Table 1. Summary of packet data for 300 s simulation: deadlines, value function parameters, mean packet lengths, and total arrivals.

Session identifier	Type	Importance	Deadlines t_d (ms)	w	dw	$l(t)$	Mean packet length (bits)	Total packet arrivals (300 s)
S_1	Voice	High	20	30	1.5	∞	66	15,000
S_2	Voice	High	20	30	1.5	∞	66	15,000
S_3	Voice	Medium	20	15	1	∞	66	15,000
S_4	Voice	Medium	20	15	1	∞	66	15,000
S_5	Voice	Low	20	5	0.5	∞	66	15,000
S_6	Voice	Low	20	5	0.5	∞	66	15,000
S_7	Data	Low	3000	2	0.5	t/t_d	19577	20
S_8	Data	Low	3000	2	0.5	t/t_d	20217	20
S_9	Data	Low	3000	1	0.5	t/t_d	20994	22
S_{10}	Data	Low	3000	1	0.5	t/t_d	19795	18

mean duration of 100 ms and 80% or better for blackout with mean duration of 200 ms. WASA also postpones the transmittal of low value packets, resulting in a higher dropped or lost percentage. EDF is more balanced with each of the voice streams, reflecting its tendency to minimize maximum lateness and its inability to distinguish between QoS requests for each packet stream. FIFO does relatively poorly for most of the voice packet streams.

Figures 7(a)–(c) depict the per session performance for each one of the scheduling policies, WASA, EDF, and FIFO, respectively, for 100 ms mean blackout period. The shaded area represents the QoS specification agreed to in the contract. The linear graph depicts the per session value function, which in turn determines the actual reward or penalty per session. We observe that the contour, or reward obtained by WASA is the highest for high value sessions $(S_1–S_4)$ and the lowest for low value sessions $(S_9–S_{10})$. These plots correlate with the results obtained in Figure 6(a), showing the preference of WASA to scheduling of high value packets, maximizing the system value function.

From Figure 7(a) we observe that for the premium voice sessions (S_1, S_2) WASA will have the largest net reward from the users. For example, for session S_1 WASA will obtain 18 cost units out of maximum possible 30 units, which is 60% of the maximum reward. EDF (Figure 7(b)) obtains for S_1 10 cost units, while FIFO (Figure 7(c)) needs to provide a refund of 8 cost units for S_1. For the medium grade voice sessions, with a maximum reward of 15 cost units, the trend is similar. For example, for session S_3: WASA has the highest net reward of about 6.3 cost units (42% of maximum reward), EDF averages about 4 cost units (26% of maximum reward), and FIFO incurs a penalty of an average or -2.85 cost units.

As expected from the adaptive scheduling policy WASA degrades the low grade/reward voice and data sessions. For example, for data session S_7 with maximum reward of 2.0 cost units WASA and EDF each received 0.25 cost units and FIFO received 0.65 cost units.

Figures 8 and 9 provide the total value function, V, for all sessions per scheduling policy for 100 and 200 ms mean blackout times, respectively. The total value function is given as a percentage of the nominal total value function under ideal conditions. In Figure 8, for 100 ms mean blackout period the performance is consistent with the previous observations. We notice that the WASA policy will maximize the operator revenues in a wireless network with fading channels. In this example, WASA will result in a 7.5% higher value than EDF and over 100% improvement over FIFO. Figure 9 depicts V per scheduling policy for 200 ms mean blackout period. We observe that when the wireless channel deteriorates, the advantage of WASA over EDF and FIFO is more pronounced. This is due to the fact that in this case more packets will be late and consequently need to be dropped.

The reason FIFO scheduling performs much better for data channels can be viewed as ‘‘blocking’’ of voice packets by data packets. Voice packets are much smaller and arrive at a much greater rate than data packets. When a large data packet (average length 20 k bits) is scheduled using the FIFO algorithm, all voice packets that arrived later than the data packet will wait until the entire packet is transmitted (in 512 bit frames). This causes a large number of voice packets to be late. Conversely, with their significantly longer

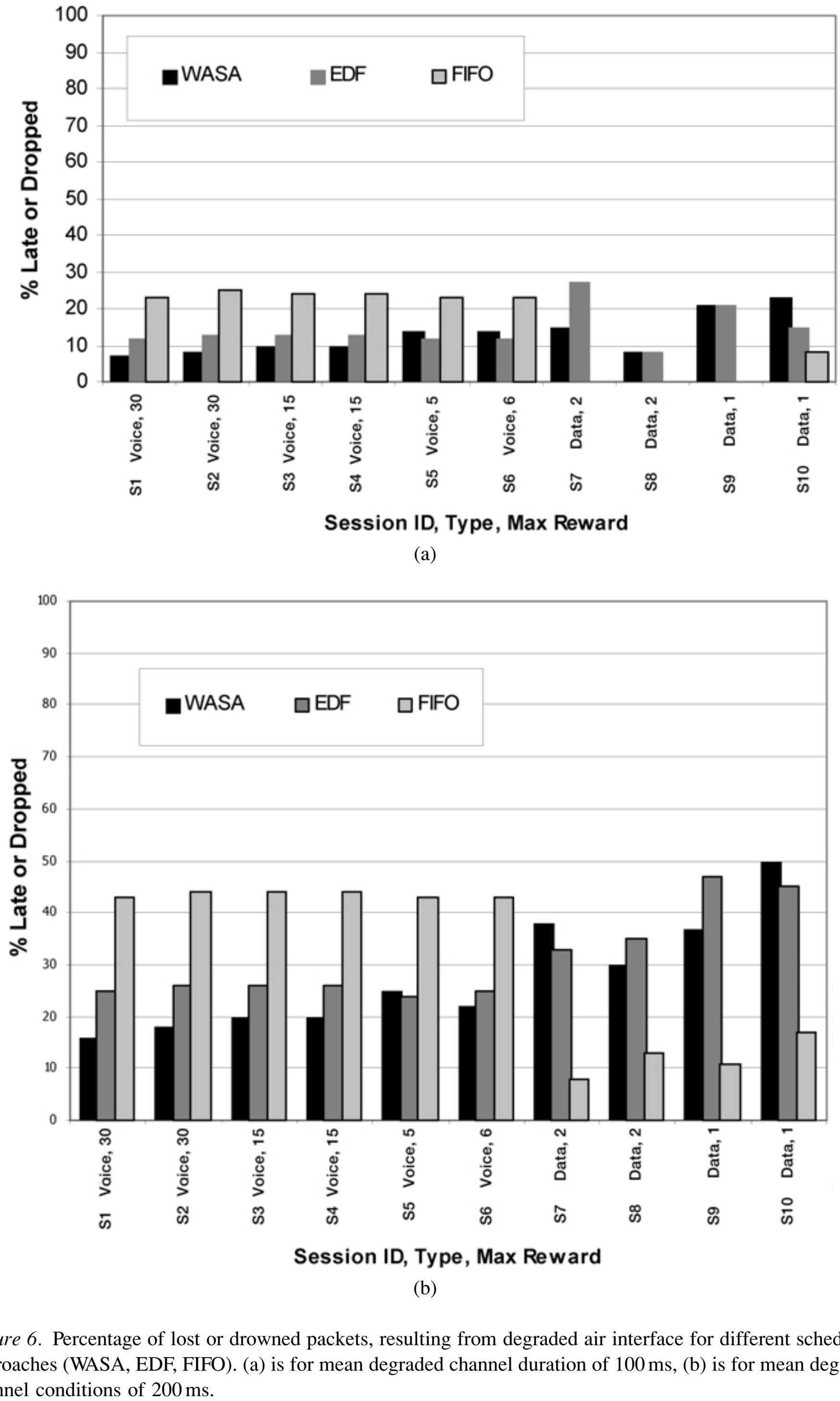
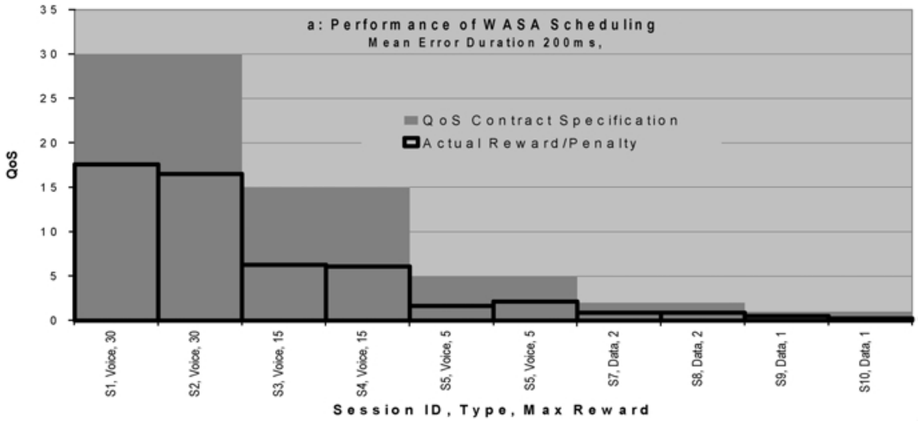
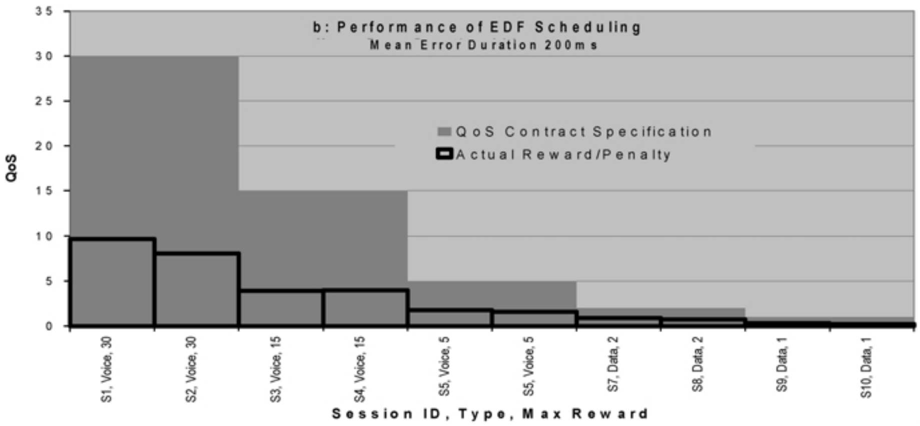


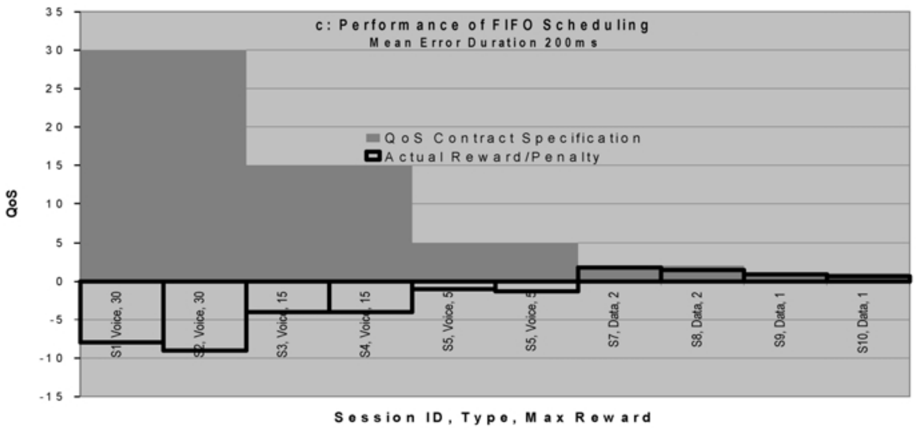
Figure 6. Percentage of lost or drowned packets, resulting from degraded air interface for different scheduling approaches (WASA, EDF, FIFO). (a) is for mean degraded channel duration of 100 ms, (b) is for mean degraded channel conditions of 200 ms.



(a)



(b)



(c)

Figure 7. Per session performance for different scheduling algorithms, channel blackouts with mean duration = 200 ms. Shaded area is QoS requested in contact. Lineargraph is per-session value functions, which also determines per session reward and penalty. The performance of WASA, EDF, and FIFO are given in (a), (b), and (c) respectively.

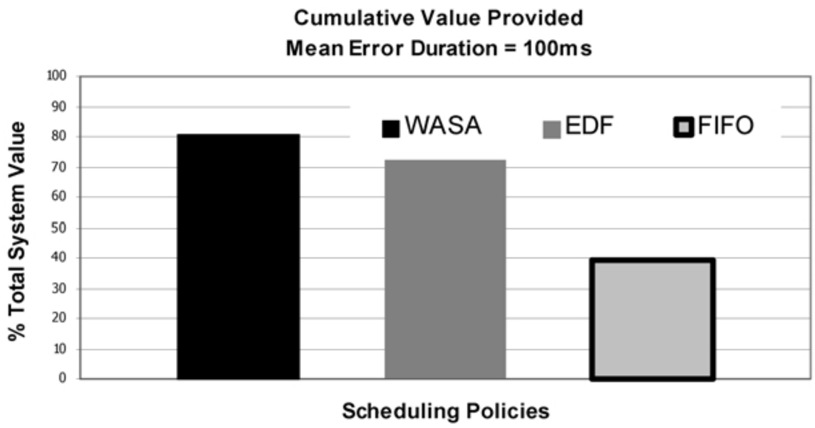


Figure 8. Total value function for all sessions with mean degraded channel duration = 100 ms, as percentage of nominal total value function, per scheduling algorithm.

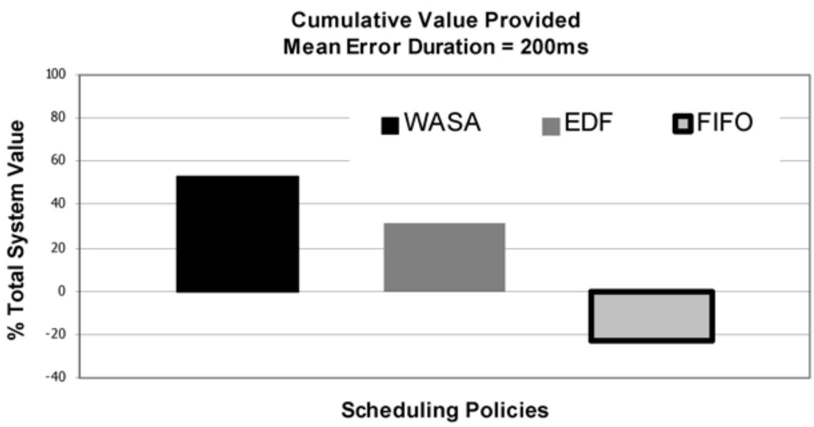


Figure 9. Total value function for all sessions with mean degraded channel duration = 200 ms, as percentage of nominal total value function, per sceduling algorithm.

deadlines and inter-arrival times, the ''blocking'' of data packets occurs much less frequently.

5. Conclusions

In this paper we have introduced an adaptive scheduling algorithm for the wireless downlink in third generation mobile networks. The module will be implemented in the RNC which becomes the intelligent gateway between the wired backbone and the mobile users on the wireless segment. The use of this algorithm at the RNC provides the wireless network the ability to complement the QoS support provided on the wired backbone. The proposed algorithm accounts for (1) the QoS requested by the application in terms of bounded delay and packet dropping probability, (2) price per session, and (3) the dynamics of the wireless channel which can cause frequent surges leading to channel overload. While accounting for all these factors, the algorithm thrives to maximize the cellular operator profit. As shown in the numerical results, the proposed algorithm provides significantly higher profits than traditional scheduling approaches.

References

1. Fishburn, P. C., and Odlyzko, A.M. 1998. Dynamic behavior of differential pricing and quality of service options for the Internet. In *Proceedings of the First International Conference on Information and Computation Economies* (*ICE*).
2. Kilkki, K. 1999. *Differentiated Services for the Internet*. Macmillan Technology Publishing.
3. Brown, F. R. 1998. Scheduling real-time processes using fuzzy logic. PhD Dissertation. Utah State University, Logan, Utah.
4. Jensen, E. D., Locke, C.D. and Tokuda, H. 1985. A time-value driven scheduling model for real-time operating systems. In *Proceedings of the Symposium on Real-Time Systems*. IEEE Computer Society.
5. Berruto, E., Columbo, G., Monogioudis, P., Napolitano, A., and Sabatakakis, K. 1997. Architectural aspects for the evolution of mobile communications toward UMTS. *IEEE Journal on Selected Areas in Communications* 15.
6. Bjorland, D. F., and Lauritzen, G. O. 1996. UMTS – The universal mobile telecommunication system. *Teletronikk*, August: 127–132.
7. Marent, B. G. 1997. Third generation mobile systems for the support of mobile multimedia based on ATM Transport. In *Proceedings of the 5th IFIP Workshop*. Ukley, UK, 1–32.
8. Valko, A. G., Racz, A., and Fodor, G. 1999. Voice QoS in third generation mobile systems. *IEEE Journal on Selected Areas in Communications* 17(1).
9. Liu, C. and Layland, J. 1973. Scheduling algorithms for multiprogramming in a hard real time environment. *ACM* 20(1): 46–61.
10. Jensen, D. 1994. Eliminating the 'hard'/'soft' real-time dichotomy. *Embedded System Conference*, 84–96.
11. Richardson, P., and Sarkar, S. 1999. Adaptive scheduling: overload scheduling for mission critical systems. *RTASS99*. Vancouver, BC.
12. Burns, A., and Wellings, A. 1997. *Real-Time Systems & Programming Languages*. Addison-Wesley.
13. Jeffay, K., Stanat, D., and Martel, C. 1991. On non-preemptive scheduling of periodic and sporadic tasks. *12th IEEE Real-Time Systems Symposium*. San Antonio, TX.
14. Kamat, S., and Zhao, W. 1996. *Real Time Performance of Two Token Ring Protocols. Advances in Real Time Systems*. Prentice Hall.
15. Richardson, P., and Sieh, L. 1999. Real-time LANs in combat vehicles: feasibility criteria for non-preemptive messages and multiple message streams originating from individual nodes. *IEEE ICCCN '99*. Boston, MA.
16. Strosnider, L., and Sha. 1988. Advanced real time scheduling using IEEE802.5 token ring. In *Proceedings of the IEEE Real-Time Symposium*.

Paul Richardson is an assistant professor in the Department of Electrical and Computer Engineering at the University of Michigan-Dearborn. His interests include *embedded real-time systems, in-vehicle networks* and *mobile networks*. Prior to joining the University, Dr. Richardson was a research associate and team leader at the U.S. Army Research Center in Warren, MI, where he studied embedded networks for military applications. The author has a B.S.E. in Computer Engineering, an M.S.E. in Computer and Electrical Engineering, and Ph.D. in Systems Engineering from Oakland University.

Larry Sieh is a research associate at the U.S. Army Research Center in Warren, MI. His research interests include Wireless Combat Networks, Multimedia Networks, Network Simulation, Combat Vehicle Simulation, and Software Engineering. The author has a B.S.E. in Computer Engineering, and M.S. Degrees in Electrical Engineering, Systems Engineering, and Computer Science from Oakland University. He is currently a Ph.D. candidate in Software Engineering at the Naval Postgraduate School.

Prof. Aura Ganz's research interests include Multimedia Networks, Wireless Networks, Security, High Speed Networks, Satellite Systems, Performance Evaluation and Algorithm Design of Computer Communication Networks. Prof. Ganz received her Ph.D. degree in Computer Science from Technion, Haifa, Israel. She has published over 120 papers in various Journals and Conferences. Prof. Ganz is also the co-founder of the UMASS spin-off, EnrichNet Inc that provides software solutions for multimedia support in wireless and wired LANs.

She is an associate professor at the Department of Electrical and Computer Engineering—Computer Systems Engineering Group at the University of Massachusetts, Amherst.