Channel Efficiency Aware Scheduling Algorithm for Real-Time Services in Wireless Networks

Gengfa Fang, Jinglin Shi
Institute of Computing Technology
Chinese Academy of Sciences
Beijing, China
{gengfa.fang, sjl}@ict.ac.cn

Eryk Dutkiewicz
Wireless Technologies Laboratory
University of Wollongong
Wollongong, Australia
eryk@uow.edu.au

Abstract

In this paper, we consider the problem of scheduling real time services over time-varying wireless links in broadband wireless networks where an Adaptive Modulation and Coding (AMC) scheme is applied in the physical layer in order to decrease the packet error rate. It is well known that a properly chosen modulation and coding scheme can increase error robustness in the physical layer. However, this is at the expense of higher system complexity and decreased channel efficiency. We present a novel Near Maximum Weighted Bipartite Matching (NMWBM) scheduling algorithm, which schedules real time services in accordance with delay bounds and physical layer modulation and coding modes. Numerical results set in the context of IEEE 802.16 networks show that NMWBM can improve system packet throughput and provide higher channel efficiency compared to the existing Earliest Deadline First scheduling algorithm. NMWBM provides this improved performance while meeting delay bound and packet loss rate requirements of real time services in broadband wireless networks.

1. Introduction

Recently, there has been great interest in extending the wireline networks to wireless domain. Compared to wireline counterparts, wireless network resources are more scarce and unstable due to the multipath, fading and Doppler effects[1]. In order to deal with the error-prone channel conditions of wireless links, Adaptive Modulation and Coding (AMC) scheme is used to match transmission parameters in the physical layer [2][3][4]. When transmission conditions deteriorate, AMC may choose, for example, a more robust FEC scheme and lower modulation mode to increase the probability of successfully transmitting packets. AMC reverts to a simple FEC scheme and higher modulation mode when channel conditions improve. This new technology also makes it necessary to combine packet scheduling with the AMC mechanism in the physical layer, so as to use the wireless resources more efficiently.

Packet scheduling has been studied extensively and many packet scheduling algorithms have been proposed. Techniques in the literature for scheduling real-time services in wireline networks rely on a combination of admission control, resource reservation and packet scheduling to guarantee the deadlines of packets. It was shown in [4] that Earliest Deadline First (EDF) is optimal for scheduling such services. EDF is extensively studied in [5-8]. In [9], the author presents a Feasible Earliest Deadline First scheduling (FEDD) algorithm, which chooses to schedule the packet whose deadline has the earliest time to expire and whose channel state is in good state. It was proved that FEDD is optimal for some certain restricted arrival processes. FEDD assumes that the wireless channel has only two states: either “good” or “bad”. If the wireless channel is in the good state, a packet can be transmitted without errors. On the other hand, if the wireless channel is in the bad state, no packet can be transmitted. This assumption is not applicable, especially in wireless networks, where the wireless channel is erroneous and time varying because of deep fading and adaptation of AMC.

To the best of our knowledge, this is the first paper in the literature concerning scheduling real-time services over time varying wireless links in wireless networks, which combines the packet scheduling with AMC implemented in the physical layer. In this paper, we employ Maximum Matching Theory to schedule real time services in accordance with delay bounds and physical layer modulation and coding modes. ANear Maximum Weighted Bipartite Matching (NMWBM) scheduling algorithm is presented. Numerical results show that NMWBM can improve system packet throughput and provide higher channel efficiency compared to the existing Earliest Deadline First scheduling algorithm.

The rest of the paper is organized as follows. In the next section we introduce the system model, which includes the network model, the model of AMC. In section III, we present the NMWBM algorithm and analyze its performance. In section IV, we present numerical results and in section V, we present our conclusions.

2. System Model

2.1. Network Model

We consider the cell system, such as that in IEEE 802.16 networks, where there is one base station (BS) controlling N mobile stations (MS) in the cell. The MSs share the wireless link resources under the control of the Media Access Control (MAC) sub-layer in data link layer in the BS. TDMA is used in the physical layer where bandwidth is calculated in time slots. We assume that there is only one real-time service flow between each MS_i (i = 1, 2, …N) and BS. In this paper we consider the uplink case, where user data is sent from MS_i to BS, although our results are also applicable to the downlink. Next we describe the bandwidth request and grant process in the uplink scenario.

As depicted in Figure 1, the BS has N queues for the N MSs. Each MS has one queue in BS to store the bandwidth requests. We assume that the queues are infinite in size, so that all the bandwidth requests can be put into the corresponding queue without dropping because of overflow. This assumption is reasonable because all the bandwidth requests have deadlines beyond which the bandwidth requests will be dropped, so the bandwidth requests will not increase infinitely. The average number of bandwidth requests in the queue is given by $\lambda * D_i$, where λ is the average throughput of the application and D_i is the deadline for the corresponding service. The queue for bandwidth requests operates in the First-in First-out (FIFO) mode resulting in all the bandwidth requests in the queue being stored in a non decreasing order according to their deadlines. The queue feeds the packet scheduler in the BS one bandwidth request followed by another. The packet scheduler in the BS schedules the bandwidth requests according to their QoS and channel states. This way the uplink wireless link is shared by the N services.

2.2. Model of AMC

For M-ary QAM (MQAM) and M-ary phase shift keying (MPSK), bandwidth efficiency and robustness are inversely proportional. When the channel deteriorates, bandwidth efficiency is sacrificed and a more robust, lower order modulation scheme is chosen. High level FEC code words are selected to increase robustness at the expense of bandwidth efficiency. Under excellent channel conditions, a lower level or even no FEC is chosen in order to achieve high channel efficiency. Packets from the MAC layer will be coded using the FEC code words de-

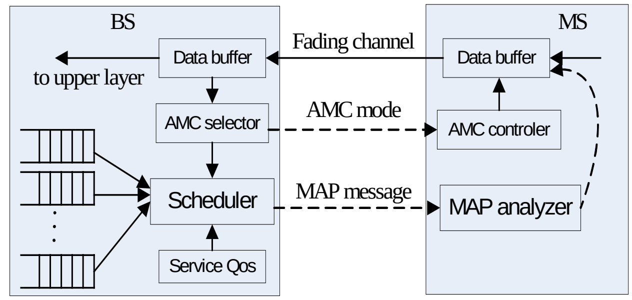

Figure 1. Uplink packet scheduling with AMC

fined in the AMC mode, where part of redundant bits will be added and the packets' size becomes P_b/ CR_n. These are then be modulated into symbols using the appropriate modulation types. Symbols are grouped into time slots. So in the physical layer, a P_b bits packet will occupy P_b/R_n symbols. Here, R_n stands for bits per symbol and it decides channel efficiency CE_n. We use CE_n to measure channel efficiency.

3. Channel Efficient Scheduling Scheme Based on Maximum Matching Theory

We consider the uplink scheduling process as described above, where the BS schedules the bandwidth requests for the next frame. For convenience, *packets* will be used to stand for bandwidth requests in the uplink scheduling process in the rest of the paper and *scheduling interval* is the time duration in a physical frame used for service data transmission.

3.1. Scheduling Model Analysis

The scheduler in BS receives all the packets before it starts to do scheduling for each *scheduling interval* in the next frame and processes all these requests together. In this scheduling model, we find that the scheduling process becomes a static problem in the following sense. Packets during the scheduling interval in the queues are stable, i.e. no new packets arrive during the scheduling process and the scheduler has full knowledge of channel conditions information before it starts scheduling. We use T_r and T_d to denote a packet's release time and deadline time accordingly. All the packets are designed to be scheduled between time T_r and T_d.

As depicted in Figure 2, service S_1 has a packet to be scheduled during $T_1 \sim T_4$. We use P_{s1} to denote the current packet in S_1, i.e. a current packet in a service is the packet with the least deadline among all the packets backlogged. Recall that all the packets in a single service are pushed in the queue in the order of packets' deadlines, so the current packet of a service should be the Head of Line

	T1	T2	T3	T4	T5	T6	T7	T8
S1	$P_{1,1}$	$P_{1,2}$	$P_{1,3}$	$P_{1,4}$				
S2	$P_{2,1}$	$P_{2,2}$	$P_{2,3}$	$P_{2,4}$	$P_{2,5}$	$P_{2,6}$	$P_{2,7}$	
S3	$P_{3,1}$	$P_{3,2}$						
S4	$P_{4,1}$	$P_{4,2}$	$P_{4,3}$	$P_{4,4}$	$P_{4,5}$			

(time →; services ↓)

Figure 2. Packets and their release time and deadlines

packet. Because of the time varying channel fading, P_{s1} may experience different fading if it is transmitted in a different time slot, which results in a different transmission profit in a different time slot, i.e. a profit of $P_{1,1}$ will be awarded if P_{s1} were scheduled in time slot T_1, and a profit of $P_{1,2}$ will be awarded if scheduled in time slot T_2. T_1 is the earliest time slot, when packet P_{s1} can be scheduled and T_4 is is the last time packet P_{s1} can be scheduled. So P_{s1} must be scheduled during $T_1 \sim T_4$, otherwise P_{s1} will be dropped by the scheduler.

We define the scheduler as a function $F(T_j)$ from time slot T_j to service S_i:

$$F(T_j) = \begin{cases} 0\ ; & \text{if no service is selected} \\ S_i\ ; & \text{if } S_i \text{ is selected for transmission} \end{cases} \quad (1)$$

A graph is said to be *bipartite graph* if all the vertices are divided into two kinds and edges are only allowed between these different kinds of vertices. From the analysis above, if we use S_i and T_i as the two kind vertices and represent $P_{i,j}$ as the edges, we can find that S_i, T_j and $P_{i,j}$ make up a weighted bipartite graph, and the scheduling process can be reduced to a Maximum Weighted Bipartite Matching problem. Given the backlogged services S = (S_1, S_2, S_3, ….., S_N), their deadlines $TD_j = (TD_1, TD_2, TD_3, ..., TD_j)$, the transmission profits in each time slots and the time slots $T = (T_1, T_2, T_3, ..., T_k)$ belonging to a frame an undirected weighted bipartite graph G = (V, E), can be constructed as follows:

$$V = S \cup T \quad (2)$$

$$E = \left\{ e_{i,j} = P_{i,j} \right\} \quad (3)$$

We use the following to calculate $P_{i,j}$:

$$P_{i,j} = CE_{i,j} \quad (4)$$

where $CE_{i,j}$ is decided by the modulation and coding mode. Each edge $e_{i,j}$ (i = 1, 2, 3, …..N; j = 1, 2, 3, ….., k) corresponds to a weighted link between S_i and T_j, whose weight is equal to the profit $P_{i,j}$.

Then our scheduling process includes the following two steps:

Step 1: Creation of the profit matrix. Before starting a scheduling cycle, the BS should create a profit matrix M[i, j] to indicate the profit M[i, j] that will be awarded if the current packet of service I, S_i, were scheduled in time slot T_j. M[i, j] is created by the BS based on the packets' release time, their deadlines and channel conditions.

Step 2: Time slot allocation. After getting the profit matrix M[i, j], the scheduler decides which service's packet should be transmitted in which times lot, while at the same time taking the System Profit into account. The purpose of the time slot allocation process is to schedule all the packets before their deadlines or reduce the number of packets because of violation of deadline constraints to least and maximize the system profit.

Next, we will apply a Near Maximum Weighted Bipartite Matching scheme (NMWBM) to implement the time slot allocation process, which guarantees packet loss rate and delay bounds for real time services.

3.2. Scheduling Based on Maximum Matching Theory

Before starting to analyze the maximum weighted bipartite matching, we first present some definitions on the graph and matching used in this section.

A weighted bipartite graph is a graph $G = (S \cup T, E)$, where row vertices set S is for the services involved, column vertices set T is for time slots within a frame and edge set E is for the link between vertexes in S and T. Each link has a weight value (or profit in this paper), which can be awarded if this link is matched.

A matching M is a subset of edges in which no vertex is adjacent to more than one edge. Matched vertex of a graph G is a vertex in G, which is adjacent to only one edge in M. A matched edge is an edge belonging to M, otherwise an edge is unmatched.

The Maximum Weighted Bipartite Matching (MWBM) works as follows:

Step 1: Start with an empty matching

Step 2: Repeat the following steps while there is an augmenting path left:

Step 2.1: Find an augmenting path with maximum score

Step 2.2: If the score > 0, then flip the edges else stop and report the maximum weighted matching M.

Eventually the above steps result in the calculation of F(t) for t from T_1, T_2, T_3, ….., T_k, so that $\sum_{T_1}^{T_k} M[F(t), t]$ is maximized and optimal. By optimal we mean that if there exits any other matching M', we have $\sum_{T_1}^{T_k} M'[F(t), t] < \sum_{T_1}^{T_k} M[F(t), t]$. The problem for traditional MWBM is that it does not take the real time services' deadlines into account, and it buys the system profit at the expense of real time services' packet loss rate. In order to improve the system profit, MWBM is willing to allocate the time slots

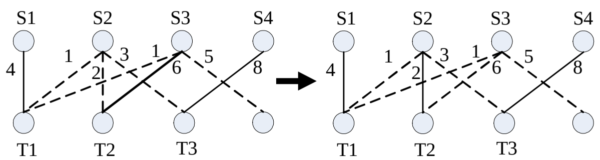

Figure 3. Scheduling Example of MWBM

to those services, whose channel conditions are in better states. So when the system becomes more overloaded, the services with bad channel conditions will suffer from packet loss because the scheduler prefers to allocate the time slots to those good channel conditions. This shortcut of MWBM will lead to large packet loss when some services' channel conditions are worse than others, which reduces the performance dramatically. We illustrate it with the following example.

As shown in Figure 3, there are four services S_1~S_4 contending to be scheduled at time T_1, which is the start time of the scheduling cycle. The edges between S_i and T_j indicate that S_i can be scheduled in time slot T_j. We can see that S_1 has a deadline of T_1 and it can only be scheduled in T_1, S_2 has the deadline of T_3 and can be scheduled in T_1, T_2 and T_3. S_3 has a deadline of T_4 and can be scheduled in time slots T_1, T_2, T_3 and T_4. The left part of Figure 3 is the result of MWBM using the steps above. T_1~T_3 are allocated to S_1, S_3 and S_4.

The overall system profit from T1 to T3 is calculated as follows:

System Profit = 4 + 6 + 8 = 18 ;

The packet in S_2 is dropped in this scheduling cycle because its deadline expires in T3. But if we use EDF to schedule the packets, the time slots will be allocated to S1, S2 and S3 as the right part of Figure 4. and the overall system profit becomes:

System Profit = 4 + 2 + 8 = 14 ;

In this case, S2 is scheduled successfully and the system profit is reduced by 4. From this example we can see that there is a tradeoff between the system profit and the packet loss of real time services. According to the mechanism in MWBM, a vertex is more likely to be matched if it has a higher edge weight than others. For instance, if a vertex has an edge, which has the highest weight among all the edges then after using MWBM, this vertex will be definitely matched.

In order to balance the system profit and packet loss rate of real time service, especially when the system becomes overloaded and the delay bounds for real time is not long enough to tolerate the time varying channel conditions, we propose a scheme, which we call Near Maximum Weighted Bipartite Matching (NMWBM). NMWBM can improve the system profit and channel efficiency and at the same time guarantee the packet loss rate and delay bounds of real time services according to their deadlines. Before we present NMWBM, we first provide a claim that will be used next.

Claim 1: If a vertex V in a weighted bipartite graph is unmatched in a maximum matching M, then all the vertices adjacent to V's edges are matched and the matched edges' weights are larger than those of edges in V.

Proof: We use reduction to absurdity to prove the above claim. Suppose there is a vertex W adjacent to one of V's edge, which is unmatched. Then because both V and W are unmatched, we can add an edge between V and W to M and set up a new M'. It's easy to see that |M'| > |M|, which is inconsistent with the fact that M is a maximum matching. Secondly, we suppose that Z is a vertex adjacent to one of V's edges E_{vz} and E_{zk} is the matched edge and $W(E_{vz}) > W(E_{zk})$. Then if we flip E_{vz} and E_{zk}, we can get a matching M' and |M'| > |M|, which is also inconsistent with the fact that M is a maximum matching.

Services in MWBM will suffer from a high packet loss rate because of bad channel conditions, even though they have more urgent deadlines than others. In order to pay more attention to packets' deadlines when allocating time slots, so that packet loss rates are bounded, we present a mechanism to adjust the time slot allocations between services according to their deadlines after MWBM process is used.

We use a packet loss indicator (PLI_i) to monitor the services' packet loss rate. When the packet loss rate of a service is larger than a threshold, we will activate its PLI to indicate its quality reduction. When the PLI of a service is active it indicates that the service has been experiencing worse channel conditions. The scheduler should attempt to adjust the result of MWBM, if possible, so as to guarantee the service packet loss rate in long deep fading and distribute packet loss among different services. Our adjustment process work as follows. When a vertex V is unmatched after MWBM and its PLI indicates that adjustment is needed then, according to Claim 1, the weights of edges belonging to this V are less than those of the adjacent edges. We choose the edge E_{vw} of V, whose weight is closest to its matched adjacent edge in vertex W, and then flip these two edges. According to Claim 1, the new graph is still a match with a $W(E_{vw}) - W(E_{vk})$ reduction of system weight. When creating the profit matrix, we sort the services in a deadlines ascending order, i.e. S_0 has lowest deadline an S_k has highest deadline. We start the adjustment from S_0 to S_k, so that an unmatched packet with the earliest deadline has the highest priority to get matched again.

Definition 1: The *least decrementing edge* of a vertex V is the edge of V, whose weight is closest to its adjacent and matched edge.

The NMWBM works frame by frame and its main steps are as follows. For each scheduling cycle perform the following steps:

Step 1: Create the profit matrix M_p
Step 2: Sort M_p in a deadline ascending order
Step 3: Perform the MWBM algorithm for M_p
Step 4: For all the service vertices in M_p do following steps:
Step 4.1: If a vertex is unmatched, find the *least decrementing edge* and flip it with its adjacent and matched edge.

4. Numerical Results

In this section, we report the results of our simulation experiments. These experiments are designed to investigate properties of NMWBM in a practical IEEE 802.16 situation and to compare NMWBM with other scheduling disciplines designed for real time services, such as EDF.

A cell consists of a BS and N MSs. One MS creates a G.711 VOIP service, which sends fixed length packets to the BS at fixed intervals. The packet size of G.711 VOIP is 32 bytes or 256 bits. The normal packet transmission interval is 0.004 second, so the application level bit rate is 64 kbps. The delay bound of VOIP in the wireless access network is restricted to 10ms, which means that the packets exceeding 10ms will be dropped. Because G.711 VOIP is a rate-constant real time service, we define it as an Unsolicited Grant Service (UGS) as supported in IEEE 802.16. With UGS, there is no need for the MS to send bandwidth requests every time the MS has data to send. The MS is guaranteed to receive grants from the BS within a fixed interval. In each simulation, the system is run for 60 seconds.

4.1. The Robustness to Channel Deterioration

The main idea of our scheme is to avoid packet transmissions when the corresponding channel is in a fading state and to transmit packets when the channel is in a better state in order to reduce the packet loss rate and make efficient use of the channel bandwidth. In this section, we use system simulations to estimate how robust our NMWBM scheduling scheme is when the channel deteriorates. We denote Pbg as the transmission probability from state S_0 to all the other states where correct packet transmission can happen. By decreasing Pbg, we can simulate the channel deterioration process.

Since there are no scheduling algorithms designed for real time services in multi state wireless channels, we compare the result of NMWBM with that of EDF, where the scheduler chooses the packet with the earliest deadline to send. Figure 4 shows the packet loss rate when the channel becomes more fading. When Pbg is less than 0.9, the packet loss rate of NMWBM is almost zero. This means that NMWBM can skip the short burst errors in the wireless link and improve bandwidth efficiency by decreasing the number of lost packets. From the results in Figure 4 we can also see that NMWBM has a relatively lower packet loss rate than EDF, especially when the network is not overloaded. There are many empty time slots that can be used and the NMWBM scheduler tries to schedule each packet in the best time slots where the packet can be transmitted correctly and improve the channel efficiency as much as possible.

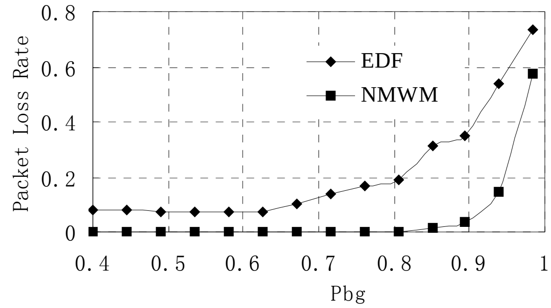

Figure 4. Packet Loss Rate vs Channel Quality

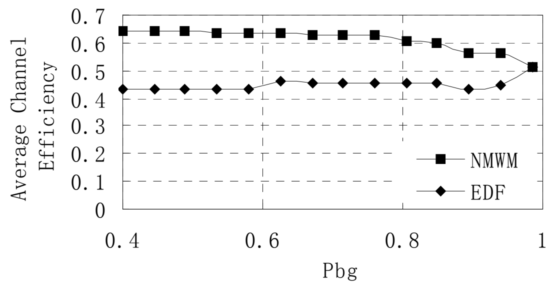

Figure 5. Average Channel Efficiency vs Channel

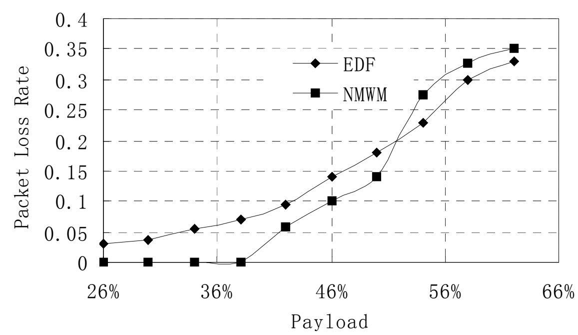

Figure 6. Packet Loss Rate vs System Payload

Figure 5 shows the average channel efficiency of NMWBM and EDF as a function of channel quality. NMWBM achieves higher channel efficiency than EDF by choosing the best time slots to send the packets, which are filled based on the maximum matching theory. When Pbg is less than 0.8, the average channel efficiency is around 0.65 for NMWBM and only 0.45 for EDF. At that level of channel deterioration EDF has also a much higher packet loss rate compared to NMWBM as shown in Figure 4. Figure 4 and 5 indicate that NMWBM can provide higher channel efficiency and lower packet loss rate when the system is not overloaded.

4.2. The Robustness to System Payload

In this section, we attempt to show the performance of NMWBM when the system payload becomes higher than that used in Figure 4 and Figure 5. We continue to compare NMWBM with EDF and we increase the system load by adding more MSs in the simulation scenario of the previous section. We choose Pbg = 0.4 in this case, where the burst errors are not very bad and the packets' deadline is larger than the longest error burst. The results in Figure 6 indicate that when the system load is below 38% in this scenario, NMWBM can guarantee that no packet is lost because of the deadline expiration. However, EDF still displays small packet loss at low system load because EDF has no mechanism to avoid link errors in the physical layer. When the payload is larger than 50%, NMWBM has a higher packet loss rate than that EDF. This is because NMWBM tries to improve channel efficiency at the expanse of the packet loss rate. At this load most of the packets loss happen because of the deadline expiration.

5. Conclusions

This paper addressed the problem of scheduling real-time services over error-prone broadband wireless networks where Adaptive Modulation and Coding is used in the physical layer. We introduced a Near Maximum Weighted Bipartite Matching (NMWBM) scheduling algorithm to guarantee QoS of real time services in a statistical manner and to improve channel efficiency and system throughput by a combination of packet scheduling in the MAC layer and an Adaptive Modulation and Coding mechanism in the physical layer. Numerical results based on an IEEE 802.16 network scenario show that the NMWBM scheme outperforms the EDF scheme in terms of packet loss rate and channel efficiency Our future work will focus on further optimization of the NMWBM scheduling algorithm and its performance evaluation in broadband wireless networks required to support mixes of real-time multimedia applications.

6. References

[1] S. Lu, V. Bharghavan and R. Srikant, "Fair scheduling in wireless packet networks," in: Proc. ACM Sigcomm'97 (September 1997).

[2] M. Elauod and P. Ramanathan, "Adaptive Use of Error-correction Codes for Real-time Communication in Wireless Networks," Proc. INFOCOM'98, pp. 548-555, Mar. 1998.
[3] Qingwen Liu; Shengli Zhou; Giannakis, G.B, "Cross-Layer Scheduling With Prescribed QoS Guarantees in Adaptive Wireless Networks," IEEE Journal on Selected Areas in Communications, Volume 23(5), pp:1056 – 1066, May 2005.

[3] X. Liu, E. K. P. Chong, and N. B. Shroff. "Optimal opportunistic scheduling in wireless networks". In Proceedings of IEEE INFOCOM, 2000.

[4] S. S. Kulkarni and C. Rosenberg, "Opportunistic scheduling policies for wireless systems with short term fairness constraints," In Proceedings of IEEE GLOBECOM, December 2003.

[5] Andrews, M, "Probabilistic End-to-End Delay Bounds for Earliest Deadline First Scheduling," INFOCOM 2000.

[6] Sivaraman, V.; Chiussi, F, "Providing End-to-End Statistical Delay Guarantees with Earliest Deadline First Scheduling and Per-Hop Traffic Shaping," INFOCOM 2000.

[7] SANJAY SHAKKOTTAI, "Scheduling Real-Time Traffic With Deadlines over a Wireless Channel," Wireless Networks 8, 13–26, 2002.

[8] F. Babich, G. Lombardi, "A Markov Model for the Mobile Propagation Channel," IEEE Trans. Vehicular Tech., pp. 63-73, Jan. 2000.

[9] Qinqing Zhang; Kassam, S.A , "Finite-State Markov Model for Rayleigh Fading Channels" IEEE Transactions on communications, Volume 47(11), pp:1688 – 1692, Nov. 1999.

[10 N. Figueira and J. Pasquale, "A schedulability condition for deadlinebased service disciplines," IEEE/ACMTransactions on Networking 5(2) , April 1997.

[11] Y. Cao and V. O. K. Li, "Scheduling algorithms for broadband wireless networks," Proc. of the IEEE, Vol. 89, No. 1, pp. 76 - 87, Jan. 2001.

[12] K. Kumaran, G. E. Margrave. D. Mitra, and K.R. Stanley, "Novel techniques for the design and control of Generalized Processor Sharing schedulers for multiple QoS classes," In Proceedings of IEEE INFOCOM 2000. March 2000.