Throughput and Temporal Fairness Optimization in a Multi-Rate TDMA Wireless Network

Teerawat Issariyakul and Ekram Hossain
University of Manitoba and TR*Labs*
Winnipeg, MB, Canada.
Email: {ekram, teerawat}@ee.umanitoba.ca.

Abstract

This paper presents an optimization-based approach to solve the wireless fair-queuing problem under a TDMA (Time Division Multiple Access)-based MAC (Medium Access Control) framework. By formulating the fair scheduling problem as an assignment problem, we propose ORCA-MRT (Optimal Radio Channel Allocation for Multi-Rate Transmission) for fair bandwidth allocation in wireless data networks which support multi-rate transmission at the radio link level. The key feature of ORCA-MRT is that while allocating transmission rate to each flow fairly it keeps the inter-packet transmission delay bounded under a certain limit. We investigate the performance of the proposed ORCA-MRT scheduler in comparison to another recently proposed multi-rate fair scheduling algorithm. We also propose two channel prediction models and perform extensive simulation to investigate the performance of ORCA-MRT in terms of different system parameters such as channel state correlation, number of flows, etc.

Keywords- Optimization, wireless fair scheduling, adaptive transmission rate, finite state channel model.

I. Introduction

The problem of fair bandwidth allocation in TDMA-based wireless environment has been studied quite extensively in recent literatures. [1] summarizes most of the proposed heuristic-based approaches for fair bandwidth allocation (e.g., *Wireless Packet Scheduling* (WPS), *Channel Independent Fair Queuing* (CIF-Q), *Wireless Fair Service Algorithm* (WFS)). Recently, an optimization-based approach, *Optimal Radio Channel Allocation* (ORCA), has been shown to provide improved performance over the above heuristic-based approaches [2].

*This work was supported in part by a scholarship from the TR*Labs*, Winnipeg, Canada and in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

However, all the above scheduling algorithms are based on the assumption that only one flow can transmit at an instant and at one transmission rate only.

It is well known that in a wireless network the spectrum efficiency of the radio channels can be substantially increased by using dynamic rate adaptation based on the channel interference and fading conditions [3]. The dynamic transmission rate can be achieved, for example, through adaptive modulation (e.g., M-QAM (*M-ary Quadrature Amplitude Modulation*) [4]) and coding in TDMA-based systems and through variable spreading gain and/or multi-code transmission in CDMA (*Code Division Multiple Access*)-based systems. Analysis of the fair-queuing problem under multi-rate transmission, therefore, reveals the interesting inter-relationship among the physical level transmission parameters and the radio link level performance measures.

II. Related Work and Motivation

A. Optimal Radio Channel Allocation (ORCA)

[2]

ORCA is a frame-based scheduling algorithm where the allocations for several time slots are calculated simultaneously. The wireless channel is modeled by the two-state Gilbert-Elliott model with the perfect channel knowledge. Only one flow perceiving good channel are allowed to transmit in a time slot.

To ensure fairness and collision-free condition, ORCA imposes two more constraints. Subject to these constraints, ORCA formulates a fair scheduling problem as an assignment problem and minimizes the following total transmission cost (Ω):

$$minimize \quad \Omega = \sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij} \cdot x_{ij} \tag{1}$$

$$s.t. \qquad x_{ij} = \begin{cases} 1, & \text{flow } i \text{ is assigned with time slot } j \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

$$\sum_{i=1}^{k} x_{ij} = 1 \tag{3}$$

$$\sum_{j=1}^{k} x_{ij} = 1 \tag{4}$$

where k is matrix dimension. After formulating the cost matrix ($\mathbf{C}$) whose elements are $c_{ij} \in \mathbb{R}^{k\times k}$ from (5) below,

$$c_{ij} = \begin{cases} 0, & \text{slot } j \text{ is clean for flow } i \\ 1, & \text{slot } j \text{ is dirty for flow } i. \end{cases} \tag{5}$$

a unique set of solution (x_{ij}) can be obtained by using the following *Hungarian method*:

- **Step1:** Find the minimum element min_i in each row of the cost matrix $\mathbf{C}$ and subtract each element in the row by this value (i.e. $c_{ij} = c_{ij} - min_i$).
- **Step2:** Find the minimum element min_j in each column and subtract each element in the column by this value (i.e. $c_{ij} = c_{ij} - min_j$).
- **Step 3:** Use minimum straight lines to draw through all zeros in $\mathbf{C}$. If the number of lines is equal to the matrix dimension (i.e. k), go to step 5. Otherwise proceed to step 4.
- **Step 4:** Find the minimum min_{ij} among the elements that are not drawn through by any line. Subtract min_{ij} from each element which does not have any line drawn through and add min_{ij} to each element which has two lines drawn through. Go back to step 3.
- **Step 5:** Select a set of zeros on the lines such that there is only one zero in each row and column. Set $x_{ij} = 1$ for the selected zeros and $x_{ij} = 0$ elsewhere.

Each row in the cost matrix $\mathbf{C}$ represents the transmission cost of each flow, while each column represents each time slot. For a flow h, w_h (the weight of flow h) identical rows, $\mathbf{R_h}= \{R_1, R_2, \cdots, R_{w_h}\}$, are inserted into the cost matrix ($\mathbf{R_h}$ is the set of rows belonging to flow h). The flow h is allowed to transmit for w_h time slots (in column j where $x_{ij} = 1$; $i \in \mathbf{R_h}$) in a scheduling frame.

Since each flow can experience a *dirty* channel for a long time and might not find the channel *clean* for an entire scheduling frame, ORCA utilizes the same lead-lag compensation as in WPS [1].

B. Fairness Issues in Multi-Rate Environment

A typical objective function for a fair scheduling algorithm is to maximize the throughput under the resource constraints without deteriorating fairness. In a CDMA system, *multi-channel fair scheduler* (MFS) maximizes system throughput under the power budget constraint [5]. The rate allocation is selected by using a greedy method. In order to maintain fairness, the preference list in the greedy method is calculated by using a stochastic approximation iterative algorithm. It has been shown that the tighter the fairness constraint, the slower the convergence rate of MFS.

DEFINITION *1:* (TEMPORAL-FAIRNESS) is the property of a scheduler to fairly allocate time slots among all the flows so that they will experience similar inter-packet transmission delay[1].

DEFINITION *2:* (THROUGHPUT-FAIRNESS) is the property of a scheduler to fairly allocate transmission rates so that all the flows will transmit similar number of data packets over a certain period of time.

When all the mobiles are fully connected, IEEE 802.11 DFC provides long-term temporal fairness. *Opportunistic Auto Rate* maintains the temporal fairness property of IEEE 802.11 and maximizes the overall system throughput [6]. The basic idea behind this protocol is that the coherence time in a particular environment is much longer than a packet size. As a result a mobile perceive good channel condition is allowed to transmit several packets consecutively.

From (4), we observe that flow i with weight w_i is always allocated exactly w_i time slots. The assignment problem has a unique property to guarantee delay-fairness when the channel condition is known ahead of time. However, the explicit compensation utilized in ORCA nullifies this property. In addition, the assumption of perfect channel knowledge in ORCA is fairly unrealistic. In this paper, we proposed two channel prediction models and develop *Optimal Radio Channel Allocation for Multi-Rate Transmission* (ORCA-MRT) to support throughput-fairness in a multi-rate environment.

[1]Inter-packet transmission delay for flow i (d_i) is the interval between two consecutive packets transmitted by flow i

III. Architecture of ORCA-MRT Scheduler

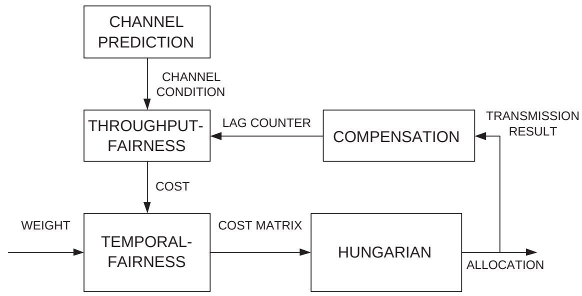

Fig. 1. Architecture of the ORCA-MRT scheduler.

Fig. 1 is the block diagram illustrating how ORCA works. Each component performs as follows:

A. Channel Prediction Block

A.1 Channel Model

The channel model is modelled as a *Finite-State Markov Channel* (FSMC) [7]-[8], which is the generalized model of the Gilbert-Elliot model used in ORCA. A FSMC is represented by a discrete-time Markov chain in which transitions only between adjacent states are allowed. Each state in the Markov chain corresponds to a channel state in a particular time slot with a specific level of signal-to-noise ratio (SNR) at which a certain transmission rate results in negligible probability of packet loss. For the M-state FSMC, we assume M possible different transmission rates within a time slot.

The most important parameter in FSMC model is the transition probability matrix $\mathbf{P}$ whose element is $p_{ij} \in \mathbb{R}^{[0,1]}$. Given steady state probabilities ($\pi^{(i)} \in \mathbb{R}^{[0,1]}$ $(i = 1, 2, \cdots, M)$) and average channel state correlation ($\rho_{avg} \in \mathbb{R}^{[0,1]}$) which is defined as follows:

$$\rho_{avg} = \frac{1}{M} \sum_{i=1}^{M} p_{ii} \tag{6}$$

, the transition probabilities p_{ij} are calculated by solving the following optimization problem:

$$\min_{p_{ij}} \sum_{i=1}^{M} \left(p_{ii} - \rho_{(avg)}\right)^2 \tag{7}$$

$$s.t. \quad \sum_{j=1}^{M} p_{ij} = 1; \quad i = 1, \cdots, M. \tag{8}$$

$$\sum_{i=1}^{M} \pi_i \cdot p_{ij} = \pi_i; \quad j = 2, \cdots, M, \tag{9}$$

$$0 \leqslant p_{ij} \leqslant 1; \quad i, j = 1, 2, \cdots, M. \tag{10}$$

A.2 Channel Prediction

The throughput-fairness block takes predicted channel states (s'_{ij}) to calculated transmission cost. Other than *perfect channel prediction* where s'_{ij} is equal to actual channel state (s_{ij}), we propose two following realistic channel prediction models:

- *Simulated-based channel prediction* utilizes last known state (s_{i0}) as an initial state to simulate $s'_{ij}; j \in \{1, \cdots, T\}$,where T is scheduling frame size.
- *Expectation-based channel prediction* calculates the predicted channel state of flow i in time slot $k + t$ given that s_{ik} is known ($s'_{ij} = E[s_{i,k+t}|s_{ik}]$) as follows:

$$s'_{ij} = E[s_{i,k+t}|s_{ik}] = \sum_{l=1}^{M} l \cdot p_{il}^{(t)} \tag{11}$$

  where $p_{il}^{(t)}$ is the probability that state i will be in state l in next t time slots. The expected state in each slot in a scheduling frame can be calculated by setting $k = 0$ and varying t to the values in the set of $\{1, \cdots, T\}$.

B. Throughput-Fairness Block

The responsibility of the throughput-fairness block is to give favor to a flow which is lagging and/or perceives good channel condition. The cost function is formulated as follows:

$$c_{ij} = (M - s'_{ij}) \cdot (L_i + 1). \tag{12}$$

where c_{ij} and s'_{ij} is the transmission cost and predicted channel state of flow i in time slot j, M is the number of FSMC states, and L_i is the lag counter of flow i. The weights of flows are ignored in this block and are taken into consideration in the temporal-fairness block.

C. Temporal-Fairness Block

ORCA-MRT is a frame-based scheduling where the scheduling frame size is $T = \sum_i w_i$ time slots. For each flow i with the weight of w_i, the temporal-fairness block calculates a row vector of c_{ij} (derived from the throughput-fairness block) with the length of T, and inserts w_i identical rows ($\in \mathbf{R_i}$) into the cost matrix.

Observation: (PROPERTIES OF TEMPORAL-FAIRNESS BLOCK) *Regardless of channel condition and/or lag counter, the temporal-fairness block ensures that*

1. *Flow i transmits in exactly w_i time slots in a scheduling frame.*
2. *The maximum inter-packet transmission delay for flow i is bounded by*

$$d_{i,max} = 2\sum_{\forall_{j \neq i}} w_j + 1. \tag{13}$$

3. *The maximum inter-packet transmission delay of each flow increases by $2w$ if a flow with the weight of w becomes active.*

Proof: Due to constraints (3) and (4), each row in the cost matrix is allocated with exactly 1 time slot. By inserting w_i rows into the cost matrix, flow i is allocated with exactly w_i time slots. Therefore, temporal-fairness is not affected by the channel condition.

Consider two consecutive scheduling frames. The maximum inter-packet transmission delay occurs when an allocation for a flow is located on the beginning of the first frame and the end of the second frame. In this case, for flow i, $max\{d_i\} = 2(T - w_i) + 1 = 2\sum_{\forall_{j \neq i}} w_j + 1$. It can easily be observed that $d_{i,max}$ increases by $2w$, if summation of all the weights increases by w.

■

D. Hungarian Block

The Hungarian block receives the cost matrix from the temporal-fairness block and uses the algorithm presented in section II-A to solve the assignment problem. In column (or time slot) j, flow i will transmit with the rate of

$$r_{ij} = \begin{cases} r_{min} + s_{ij} - 1, & x_{hj} = 1 \ \forall_{h \in \mathbf{R_i}} \\ 0 & , \ otherwise \end{cases} \tag{14}$$

where r_{min} is the minimum rate that can be transmit (when the channel state is worst). We assume that channel state is always known right before the transmission. The transmission rate is therefore calculated based on the actual channel, not the predicted channel.

E. Compensation Block

Due to the channel variation, each flow might be allocated at different transmission rate. The resulting unfairness is monitored by the compensation block via the lag counter calculated as the difference between total transmitted packets of flow i and the maximum (among all the flows) total transmitted packets.

F. Summary

During each scheduling frame, the ORCA-MRT scheduler performs the following procedure:

- **Step 1:** Channel prediction block predicts the channel condition in the next scheduling frame.
- **Step 2:** Based on the predicted channel condition and the lag counter, the throughput-fairness block calculates for each flow transmission cost in each time slot, by using (12).
- **Step 3:** Temporal-fairness block receives transmission cost from the throughput-fairness block. For flow i, w_i identical rows of transmission costs are inserted into the cost matrix.
- **Step 4:** Hungarian block solves the assignment problem based on the input cost matrix and gives the allocation to each flow.
- **Step 5:** Compensation block observes the transmission result and calculates lag counters as a function of the observed transmission result.

IV. Simulation Environment

A. Performance Measures

In this paper, we define the following performance matrics based on [5]:

- **Normalized throughput for flow** i (γ_i):

$$\gamma_i = \frac{\sum_{k=1}^{F} r_i(k)}{FTw_i}. \tag{15}$$

- **Normalized inter-packet transmission delay for flow** i (D_i):

$$D_i = \frac{w_i}{T\dot{K}_i} \cdot \sum_{k=1}^{K_i} d_i(k). \tag{16}$$

where F is total number of scheduling frames, each with the size of T time slots, $r_i(k)$ is the number of packets successfully transmitted by flow i (with the weight of w_i) in frame k, and $d_i(k)$ is inter-packet transmission delay of flow i from packet k^{th} to $k+1^{th}$, and K_i is the number of packets transmitted by flow i. Note that the above values are normalized so that they are independent of both w_i and T.

We measure throughput and throughput-fairness by means of the average value ($\overline{\gamma}$) and standard deviation ($\sigma(\gamma)$) of γ_i. Similarly, delay and temporal-fairness are measured by the average value ($\overline{D}$) and standard deviation ($\sigma(D)$) of D_i. Before proceeding to the simulation, the following observations are noteworthy:

1. Higher values of $\overline{\gamma} \leqslant M/T$ imply better throughput.
2. Smaller values of $\sigma(\gamma)$ imply a better throughput-fairness performance.
3. Under perfectly fair time slot allocation, $\overline{D} = 1$.
4. Small $\sigma(D)$ expresses better temporal-fairness performance.

B. Simulation Parameters and Methodology

In this paper, MFS is used as a reference for the performance comparison. The simulation environment is set up to be similar to that in [5]. More specifically, we run ORCA-MRT with maximum lag counter and maximum compensation frames of ∞. We set r_{min} to 2 packets and run the simulation over 3-FSMC ($r = \{2,3,4\}$), 5-FSMC ($r = \{2,3,\cdots,6\}$), and 7-FSMC ($r = \{2,3,\cdots,8\}$) with equally-likely steady state probability for 2000 time slots as suggested in [5]. We vary ρ_{avg} to a range of $\{0.5, 0.6, 0.7, 0.8, 0.9\}$. Unless otherwise specified, perfect channel prediction is assumed to eliminate the effect of prediction inaccuracy on the performance.

V. Simulation Results and Discussions

A. Performance of ORCA-MRT and MFS

As in MFS, the number of active flows is set to 12 flows with the weight of 1 and 4 flows with the weight of 2. The average value and standard deviation of the normalized throughput among all the flows are shown in Fig. 2.

Throughput of MFS ranges from 0.35 to 0.45 depending on the service discrepancy (δ),

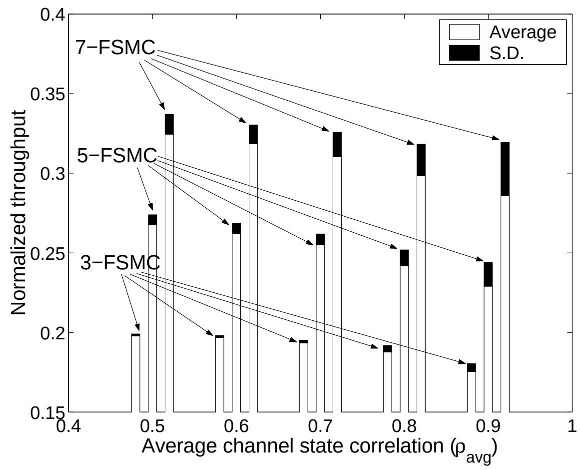

Fig. 2. Average and standard deviation of normalized throughput.

while ORCA-MRT has maximum throughput of 0.33. However, the MFS incurs standard deviation of 0.05-0.11, while ORCA-MRT achieves better throughput-fairness of about 0.006-0.03.

B. Channel State Correlation

As can be seen from Fig. 2, throughput and throughput-fairness improve as channel states become less correlated (small ρ_{avg}). When the channel is less correlated, a flow tends not to stay in the bad state for a long time. The scheduler is able to select the most suitable state in terms of both throughput and throughput-fairness for all the flows.

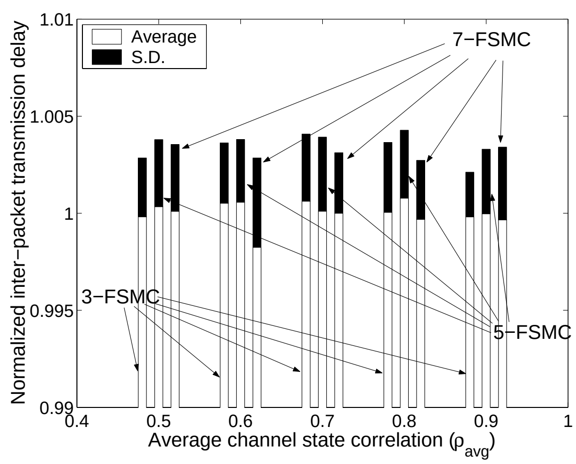

Fig. 3. Average and standard deviation of normalized inter-packet transmission delay.

Fig. 3 reveals that delay and temporal-fairness performance is fairly good: $\overline{D} \approx 1$ and $\sigma(D) < 0.005$ for all the channel condition. ORCA-MRT expresses delay robustness in that both $\overline{D}$ and $\sigma(D)$ are not affected by the channel condition.

C. Channel Prediction

Fig. 4 shows the effect of different channel prediction when $M = 5$ and $\rho \in \{0.5, 0.7, 0.9\}$. It can be observed that the simulation-based channel prediction give worse throughput and throughput-fairness due to the prediction inaccuracy. As channel becomes more correlated, the channel tends to stay in the same state. The simulated channel becomes fairly similar to the actual channel. As a result, the difference of throughput and throughput-fairness of the simulated-based prediction and that of perfect channel prediction reduces.

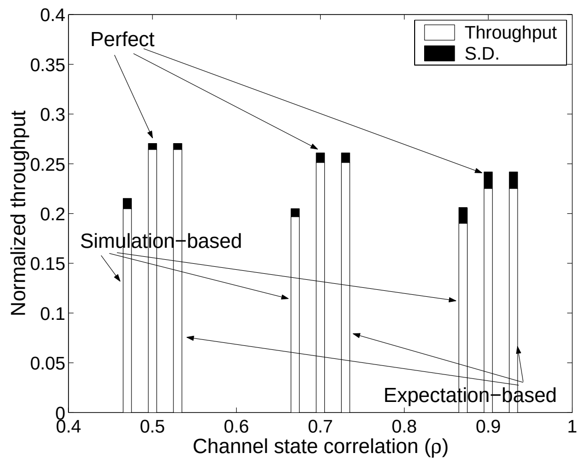

Fig. 4. Effect of imperfect channel prediction.

The expectation-based channel prediction performs fairly good in that it almost results in the same throughput and throughput-fairness as that obtained when the channel is known ahead of time. The performance in terms of delay and temporal-fairness is very similar for all the prediction models (the results are omitted for the brevity of the paper).

D. Number of FSMC States (M)

As the number of possible transmission rates increases, throughput increases as shown in Fig. 2. Throughput-fairness, on the other hand, degrades because there is less possibility for all the flows to perceive the same channel state in the same frame. From Fig. 3, we can see that delay and temporal-fairness are not affected by number of FSMC states.

E. Weights of the Flows

To investigate the effect of the weights of flows on system performance, we assume 5 flows with the weights of 1, 2, 3, 4, and 5, and run the simulation on a channel with 5-FSMC and $\rho_{avg} = 0.7$. In this case, the scheduling frame size (T) is 15 time slots.

TABLE I

EFFECT OF WEIGHT ON INTER-PACKET TRANSMISSION DELAY

flow	1	2	3	4	5
w_i	1	2	3	4	5
$max_{sim}\{d_i\}$	29	27	25	23	21
$d_{i,max}(13)$	29	27	25	23	21
$\overline{d_i}$	15.01	7.50	5.00	3.74	3.00
D_i	1.0	1.0	1.0	1.0	1.0
$\gamma_i \cdot w_i$	0.228	0.527	0.671	1.104	1.23

Table I shows that the maximum inter-packet transmission delay ($max_{sim}\{d_i\}$) is bounded by the value calculated from (13) ($d_{i,max}(13)$). We can observe that the average inter-packet transmission delay of flow i ($\overline{d_i}$) is proportional to w_i. Equivalently, the normalized inter-packet transmission delay approaches its value in ideal fairness condition ($D_i = 1$). Also, we can observe that the actual throughput ($\gamma_i \cdot w_i$) is proportional to the weight of each flow.

F. Number of Flows

We assume that all the flows have equal weight and vary the number of flows as in the set $\{5, 10, 15, 20, 25, 30\}$. Scheduling frame size is calculated by $T = \sum_i w_i = \{5, 10, 15, 20, 25, 30\}$.

As the number of flows increases, the frame size increases and there are more transmission opportunities in a scheduling frame. A scheduler has more choices to allocate higher rate to each flow while maintaining fairness. Throughput-fairness therefore improves as number of flows increases. However, the expanding frame size forces a flow to wait longer for the next transmission opportunities. As a result, the throughput decreases as the number of flows increases.

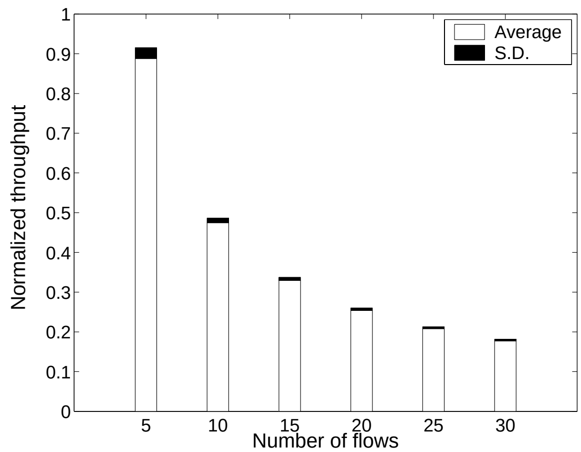

Fig. 5. Throughput and throughput-fairness vs. number of flows.

VI. Conclusions

By formulating the scheduling problem for fair bandwidth allocation as an assignment problem and based on the Hungarian method to solve the assignment problem, the ORCA-MRT methods has been proposed for fair bandwidth allocation under multi-rate transmission scenario. ORCA-MRT protocols utilize optimization-based intra-frame rate allocation along with the fairness compensation using lag counter. We also propose two channel prediction models, which perform as good as when channel states are known ahead of time. Simulation results have revealed that, due to such optimization the ORCA-MRT outperforms the opportunistic fair scheduling, MFS, in terms of throughput-fairness. More preferable feature of ORCA-MRT is its ability to ensure temporal-fairness and bound inter-packet transmission delay under a certain threshold, which can be utilized to prevent end-to-end throughput degradation in wide-area wireless networks.

References

[1] T. Nandagopal, S. Lu, and V. Bhargharvan, "A unified architecture for the design and evaluation of wireless fair queuing algorithms," *Proc. of ACM MobiCom'99*, pp. 132-142, Aug. 1999.

[2] T. Issariyakul and E. Hossain, "Optimal radio channel allocation for fair queuing in wireless data networks," *Proc. of IEEE ICC'03*, May 2003.

[3] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems, " *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 884-895, Jun. 1999.

[4] A. J. Goldsmith and S.-G. Chua, "Variable rate variable power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218-1230, Oct. 1997.

[5] Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channel," *Proc. of IEEE INFOCOM'03*, Mar. 2003.

[6] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," *Proc. of ACM MobiCom'02*, Sept. 2002.

[7] H. S. Wang and N. Moayeri, "Finite-state markov channel - a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163-171, Feb. 1995.

[8] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1688-1692, Nov. 1999.