ORCA-MRT: An Optimization-Based Approach for Fair Scheduling in Multirate TDMA Wireless Networks

Teerawat Issariyakul, *Student Member, IEEE*, and Ekram Hossain, *Member, IEEE*

***Abstract*—This paper presents an optimization-based approach to solve the wireless fair scheduling problem under a multirate time division multiple access (TDMA)-based medium access control (MAC) framework. By formulating the fair scheduling problem as an assignment problem, the authors propose the optimal radio channel allocation for multirate transmission (ORCA-MRT) algorithm for fair bandwidth allocation in wireless data networks that support MRT at the radio link level. The key feature of ORCA-MRT is that while allocating transmission rate to each flow fairly, it keeps the interaccess delay bounded under a certain limit. The authors investigate the performance of the proposed ORCA-MRT scheduler in comparison to another recently proposed multirate fair scheduling algorithm. They also propose two channel prediction models and perform extensive simulations to investigate the performance of ORCA-MRT for different system parameters such as channel state correlation, number of flows, etc.**

***Index Terms*—Adaptive transmission rate, finite state Markov channel, optimization, wireless fair scheduling.**

I. Introduction

THE PROBLEM of fair bandwidth allocation in a time division multiple access (TDMA)-based wireless environment has been studied quite extensively in recent literature. Nandagopal *et al.* [1] summarized 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 (WFS) algorithm]. An optimization-based approach, namely, optimal radio channel allocation (ORCA), was shown to provide improved performance over the above heuristic-based approaches [2]. 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 channel interference and fading conditions [3]. Dynamic transmission rate can be achieved, for example, through adaptive modulation [e.g., M-QAM (M-ary quadrature amplitude modulation)] and/or coding in TDMA-based systems and through variable spreading gain and/or multicode transmission in code division multiple access (CDMA)-based systems. Analysis of the fair-queuing problem under multirate transmission (MRT) would reveal the interesting interrelationships among the physical level transmission parameters and the radio link level performance measures.

Several works on fair scheduling in a multirate system have been reported in recent literature. The opportunistic auto rate (OAR) algorithm presented in [4] allows mobiles perceiving good channel condition to transmit several packets consecutively and ensures that all the mobiles acquire channel access to achieve the same long-term time shares (i.e., temporal fairness[1]). The algorithm presented in [5] stochastically fixes a fraction of time slot allocation to each mobile and maximizes the overall throughput in a multirate TDMA cellular system. Neither of the two above algorithms provides throughput fairness.[1] For a CDMA network, the multichannel fair scheduler (MFS) scheme, which maximizes system throughput while maintaining fairness among all the mobiles, was presented in [6]. In MFS, several mobiles are allowed to transmit at the same time as long as the total transmitted power does not exceed a certain threshold. However, MFS does not guarantee temporal fairness.

In this paper, we propose a framework, namely, ORCA-MRT, to solve the combined temporal-throughput fair scheduling problem in a multirate TDMA network. ORCA-MRT ensures fair time slot allocation in each frame and maximizes overall throughput without deteriorating throughput fairness. The problem is formulated as an assignment problem [7]. The properties of ORCA-MRT are analyzed by observing certain aspects of the modified assignment problem, and simulation results are presented in support of these mathematical observations. Also, two channel prediction methods are proposed to facilitate the optimal bandwidth allocation when channel state information cannot be perfectly known.

The remainder of the paper is organized as follows. Section II presents the background and motivation for this work. The architecture of the ORCA-MRT scheduling framework is presented in Section III. Section IV describes the simulation environment and the performance metrics. The simulation results are presented in Section V. Section VI discusses

Manuscript received September 10, 2003; revised April 17, 2004; accepted November 2, 2004. The editor coordinating the review of this paper and approving it for publication is Y. Fang. This work was supported in part by a scholarship from TR*Labs*, Winnipeg, MB, Canada, in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and in part by the University of Manitoba Graduate Fellowship (UMGF).

The authors are with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada (e-mail: teerawat@ee.umanitoba.ca; ekram@ee.umanitoba.ca).

Digital Object Identifier 10.1109/TWC.2005.858026

[1] See the definition in Section II-C.

1536-1276/$20.00 © 2005 IEEE

several issues for possible further studies. Conclusions are stated in Section VII. A list of the key mathematical notations used in this paper is given in Table I.

II. Background and Motivation of the Work

A. ORCA

In ORCA [2], the allocations for $T = \sum_{i=1}^{n} w_i$ time slots are calculated simultaneously, where w_i is the weight of flow i and n is the number of flows. The wireless channel is modeled by the two-state Gilbert–Elliott model in which the channel state can be either good or bad. Data transmission is assumed to be successful and unsuccessful when the channel is in good state and bad state, respectively. Assuming that the channel condition is known *a priori*, ORCA formulates a fair scheduling problem as an assignment problem. Mathematically

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

$$\text{subject to} \quad \sum_{i=1}^{k} x_{ij} = 1 \tag{2}$$

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

where the solution set of the assignment problem x_{ij} is interpreted as

$$x_{ij} = \begin{cases} 1, & \text{flow } i \text{ is assigned with time slot } j \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

In the above, Ω is the total transmission cost and k is the dimension of the cost matrix $(\mathbf{C})$ whose elements c_{ij} denote the transmission cost for flow i in time slot j calculated from

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

By setting the transmission cost of the bad state to be higher than that of the good state, the above optimization problem minimizes the number of unsuccessful transmissions (i.e., in bad time slots).

The solution set of the assignment problem can be obtained by using the Hungarian method [7] as follows.

Step 1) Subtract each element in the row by the minimum element in the row.

Step 2) Subtract each element in the column by the minimum element in the column.

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 element that are not covered by any line. Subtract this element from each element that is not covered by any line and add this element to each element that is covered by two lines. Go back to Step 3).

TABLE I
LIST OF KEY NOTATIONS

Notation	Meaning
$\mathbf{C}$=$[c_{ij}]$	cost matrix = [cost in row i and column j]
D_i	normalized inter-access delay for flow i
$\overline{D}$	average value of D_i
$\sigma(D)$	standard deviation of D_i
L_i	lag counter of flow i
M	number of channel states
n	number of flows
p_{ij}	transition probability from state i to state j
$P(s^{(m)})$	transmission state probability for channel state m
s_{ij}	actual channel state of flow i in time slot j
s'_{ij}	predicted channel state of flow i in time slot j
T	scheduling frame size
w_i	weight of flow i
ρ_{avg}	average channel state correlation
Ω	total cost
π_i	steady state probability for channel state i
γ_i	normalized throughput for flow i
$\overline{\gamma}$	average value of γ_i
$\sigma(\gamma)$	standard deviation of γ_i
$\Delta(t)$	average prediction error

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.

In order to allocate time slots among all the flows in proportion to their weights, ORCA replaces flow h, whose weight is w_h, with w_h identical flows, each with a weight of one. Flow h possesses w_h identical rows (the size of each row is T time slots) in the cost matrix, where the set of rows belonging to flow h is represented by $\mathbf{R}_h = (R_1 \; R_2 \; \cdots \; R_{w_h})$. 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.

The solution of the assignment problem might lead to transmission in bad time slots. Therefore, an explicit compensation mechanism (similar to that in WFS [8]) is utilized to defer the transmission of the flow experiencing bad channel condition and compensate for the deferred transmission later.

B. MFS

Two variants of MFS, namely, MFS-D and MFS-P, were proposed to solve two fairness problems [6]. MFS-D was designed to solve the deterministic fairness problem

$$\max_{X_i(k)} \quad \sum_{i=1}^{n} E\left[X_i(k)\right] \tag{6}$$

$$\text{subject to} \quad \frac{E\left[X_i(k)\right]}{w_i} = \frac{E\left[X_j(k)\right]}{w_j} \tag{7}$$

$$\sum_{i=1}^{n} s_i(k)X_i(k) \leqslant P \tag{8}$$

where $X_i(k) \in \{0, m_i^1, \ldots, m_i^{M_i}\}$ is the transmission rate of flow i in time slot k, $E[\cdot]$ is the expectation function, w_i is the weight of flow i, n is the total number of flows, and P is the maximum power limit in each time slot. The channel condition of flow i at time k, $s_i(k)$, is defined as

$$s_i(k) = 0.5 + d\cos(2\pi f_i k + \theta_i) + X_{\sigma_i}(k) \tag{9}$$

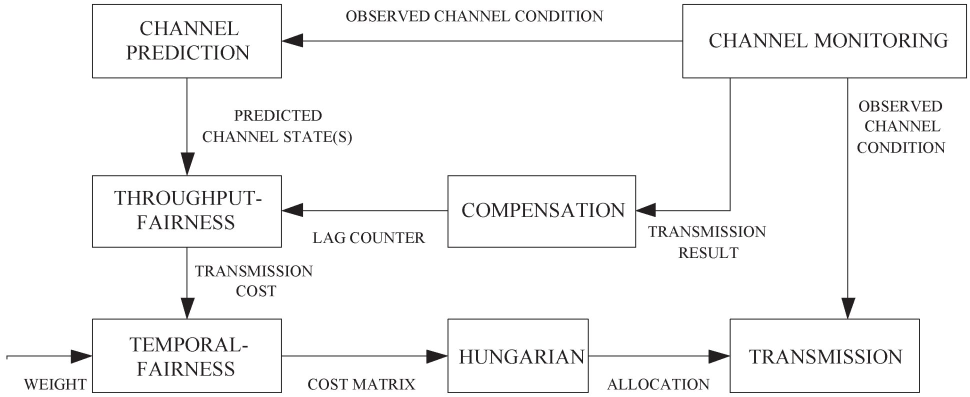

Fig. 1. Architecture of the ORCA-MRT scheduler.

where θ_i is a random variable uniformly distributed in $[0, 2\pi)$, $X_{\sigma_i}(t)$ models additive white Gaussian noise (AWGN) with variance σ_i^2, f_i expresses the channel correlation due to mobility over long time scale, and d is a scaling factor exhibiting the range of channel variation.

MFS-P was devised to solve the probabilistic fairness problem, and the constraint in (7) was replaced with the constraint

$$Pr\left(\left|\frac{E\left[X_i(k)\right]}{w_i} - \frac{E\left[X_j(k)\right]}{w_j}\right| > \delta\right) \leqslant \epsilon \qquad (10)$$

where δ is the service discrepancy defining the tolerable deviation from ideal fairness and ϵ is the corresponding probability.

Both MFS-D and MFS-P employ a stochastic approximation algorithm to establish a preference list. A greedy algorithm is then used to sequentially choose the next flows in the established list until the maximum power limit in (8) is reached. The details of these algorithms can be found in [6].

C. Motivation of the Work

ORCA is able to calculate the optimal channel allocation for fair scheduling in a single-rate TDMA network. For a multirate TDMA network, the opportunistic scheduling algorithm in [5] was designed to ensure temporal fairness (see the definition below) and to maximize overall throughput stochastically. However, neither ORCA nor the algorithm in [5] takes throughput fairness into account.

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-access delay.[2]

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.

MFS, on the other hand, was designed for throughput fair scheduling in a multirate CDMA network. MFS does not consider temporal fairness as one of its performance metrics. Also, the greedy algorithm and the stochastic approximation used in MFS may result in a suboptimal solution and slow convergence rate for the algorithm. Again, MFS assumes that the number of available codes is unlimited and the codes are perfectly orthogonal to each other. Therefore, the actual throughput is less likely to be as high as the throughput reported for MFS.

The primary objectives of the proposed ORCA-MRT framework are to ensure frame-based temporal fairness (i.e., all flows acquire temporal fair share in every frame) and bound the interaccess delay. Subject to these two constraints, our secondary objective is to maximize the overall throughput without deteriorating throughput fairness. We argue that such a combined temporal-throughput fair scheduling would be useful especially in situations where bounded delay is of utmost importance such as in the case of applications running over transmission control protocol (TCP) to avoid TCP timeouts.

The optimization problem is formulated by (1)–(4). The nature of a TDMA network and the primary objectives are realized by the hard constraints in (2) and (3) of the assignment problem. The secondary objective is achieved by designing appropriate cost function and minimizing the corresponding total cost (Ω) in (1).

III. System Model and Architecture of ORCA-MRT Scheduler

We consider a centralized multirate TDMA wireless network where only one transmitter is allowed to transmit in each time slot. It is assumed that there are M possible transmission rates, each of which corresponds to a certain level of channel condition. The scheduler (e.g., a base station in a cellular network) is responsible for performing optimization and assigning time slots among all the flows. The information about channel condition can be gathered either by the scheduler or by the mobiles. In the latter case, the channel information can be fed back to the scheduler via the reverse channel. Data transmission can be in both uplink and downlink. At the beginning of the scheduling frame, the allocation is calculated and broadcast to all the mobiles. Upon hearing the allocation, the mobiles know exactly when to transmit or receive data.

A general architecture of the ORCA-MRT framework is shown in Fig. 1. During each scheduling frame, the ORCA-MRT scheduler performs the following functions.

Step 1) The channel prediction block predicts the channel condition in the next scheduling frame.

[2]Inter-access delay for flow i is the interval between two consecutive opportunities at which flow i is allowed to transmit data packets.

Step 2) Based on the predicted channel state(s) and the lag counters,[3] the throughput-fairness block calculates transmission costs (c_{ij}).

Step 3) The temporal-fairness block receives the transmission cost from the throughput-fairness block and constructs an MRT cost matrix ($\mathbf{C}_{\text{MRT}}$).

Step 4) The Hungarian block solves an assignment problem with the above cost matrix $\mathbf{C}_{\text{MRT}}$.

Step 5) The transmission block calculates transmission rate for each time slot.

Step 6) Based on the transmission result, the compensation block updates the lag counters.

Step 7) The channel monitor block observes the transmission results and the channel conditions, and sends these to the compensation, channel prediction, and transmission blocks.

A. Channel Prediction Block

1) Channel Model and Parameter Estimation: A typical wireless channel can be modeled as a finite-state Markov channel (FSMC) [9]. An 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) in each state, there exists a maximum transmission rate that results in negligible probability of packet loss. Equivalently, for a fixed transmission rate, different states in an FSMC model represent different levels of received SNR [10].

An FSMC model is represented by a transition probability matrix $\mathbf{P}$ whose elements are p_{ij}. The estimation of FSMC parameters (i.e., p_{ij}) greatly depends on the underlying physical layer model. As an example, for M-QAM-based adaptive modulation with modulation index of {2, 4, 8, 16, 32, 64} over an AWGN channel, setting the SNR thresholds at {6, 10, 14, 18, 21, 24} can stabilize the average bit error rate (BER) of each state to 10^{-3} [3]. However, the estimation approach will be different in case of a Rayleigh fading channel [10]. To make our model applicable to any physical layer implementation, we do not make any assumption on the physical layer model and assume that the FSMC parameters are available to the channel prediction block.

In order to facilitate the understanding of the impact of channel parameters, we express the quality and the rate of change of the channel states in terms of the steady state probability (π_i) and the average channel state correlation (ρ_{avg}), which is defined as

$$\rho_{\text{avg}} = \sum_{i=1}^{M} \pi_i p_{ii} \tag{11}$$

where M is the number of FSMC states. Given $\pi_i(\forall i)$ and ρ_{avg}, all the transition probabilities p_{ij} can be calculated by solving the following optimization problem

$$\min_{p_{ij}} \quad \sum_{i=1}^{M} \pi_i \left(p_{ii} - \rho_{\text{avg}}\right)^2 \tag{12}$$

$$\text{subject to} \quad \sum_{j=1}^{M} p_{ij} = 1, \quad i = 1, \ldots, M \tag{13}$$

$$\sum_{i=1}^{M} \pi_i p_{ij} = \pi_j, \quad j = 2, \ldots, M \tag{14}$$

$$0 \leqslant p_{ij} \leqslant 1, \quad i,j = 1, 2, \ldots, M. \tag{15}$$

Along with the calculated p_{ij}, the channel state during the last time slot in the previous frame is utilized as the initial state in the M-state Markov chain to generate the channel state for flow h during time slot k, $s_{hk} \in \{1(\text{worst}), 2, \ldots, M(\text{best})\}$.

2) Channel Prediction: Assuming that the parameters of FSMC are known *a priori*, the channel prediction block calculates the predicted channel state (s'_{ij}) for flow i in slot j ($j = \{1, \cdots, T\}$) based on the following channel prediction models.

- Perfect channel prediction

$$s'_{ij} = s_{ij} \tag{16}$$

  where s_{ij} is the actual channel state.
- Simulation-based channel prediction: The channel state for flow i during the last time slot in the previous frame (s_{i0}) is utilized as the initial state in the FSMC to generate simulated channel state s'_{ij}.
- Expectation-based channel prediction: The expected channel state for flow i in time slot $k+t$ given that s_{ik} is known ($E[s_t|s_{ik}]$) can be calculated as

$$s'_{i,k+t} = E[s_t|s_{ik}] = \sum_{l=1}^{M} l p^{(t)}_{(s_{ik})l} \tag{17}$$

  where $p^{(t)}_{(s_{ik})l}$ (calculated from the Chapman–Kolmogorov equation [11]) is the probability that state s_{ik} will change to state l in next t time slots. For ORCA-MRT, we set $k = 0$ and $t = \{1, \ldots, T\}$.

B. Throughput-Fairness Block

By means of a cost function [which will be defined in (19)], the throughput-fairness block gives favor to flows that are lagging and/or perceiving good channel condition. The cost function is developed based on the following observations on the channel condition, the nature of the assignment problem, and the lag counter.

1) Channel Condition: A channel-aware cost function (c_{ij}^{CA}) could be obtained by generalizing the cost function of ORCA in (5) as

$$c_{ij}^{CA} = M - s'_{ij}. \tag{18}$$

[3]For a particular flow, the lag counter is the difference between the allocation for that flow and the maximum allocation for a flow over a time period. Formally, it is defined by (22).

Therefore, the cost function increases when the channel condition becomes worse.

2) Nature of the Assignment Problem: To minimize total cost (Ω) in (1), we set $x_{ij}=1$, where c_{ij} is the minimum. When each flow does not experience its minimum cost in the same slot, the problem is fairly simple in that the scheduler selects the flow with the minimum transmission cost in each slot. Hereinafter, we will consider only non-trivial cases where each flow experiences equal minimum cost in the same slot (which is referred to as the contention slot). In this case, all the flows must contend for the possession of that slot [because of (2)].

Consider a cost matrix $\mathbf{C}' = \begin{pmatrix} c & (c+H) \\ c & (c+I) \end{pmatrix}$, where column $j=1$ is the contention slot, and rows 1 and 2 represent transmission cost of flow h and i, respectively. If $I > H$, the contention slot $j=1$ will be given to flow i in row 2 because the solution $\{x_{12}=1, x_{21}=1\}$ leads to the minimum total transmission cost. In general, flow i will acquire the contention slot j if $c_{ik} > c_{hk}$ $(\forall k \neq j)$.

3) Lag Counter: A lagging flow is allowed to transmit in a slot with good channel condition even though other flows in the same slot perceive the same or better channel.

Observation 1 (Additive Cost Function With Respect to Lag Counter): The cost function cannot favor any lagging flow by just adding/subtracting a lag counter to all the elements in each row.

Proof: See Appendix I. ■

Therefore, a lagging flow cannot be favored only by decreasing its channel-aware costs in every time slot ($c_{ij} = c_{ij}^{CA} - L_i$, where L_i is the lag counter of flow i). From the nature of the assignment problem, the cost function will favor flow i in the contention slot if the transmission costs for flow i in all other slots are higher than those for all other flows. Therefore, the cost function is defined as

$$c_{ij} = c_{ij}^{CA}(L_i + 1). \tag{19}$$

Observation 2 ($\mathbf{C}_{\mathrm{MRT}}$): The cost function in (19) always favors lagging flows.

Proof: See Appendix II. ■

C. Temporal-Fairness Block

For flow i with weight w_i, the temporal-fairness block inserts into $\mathbf{C}$ w_i identical rows (each row has $T=\sum_{i=1}^{n} w_i$ columns) of transmission cost (c_{ij}) calculated by the throughput-fairness block. The set of rows belonging to flow i is denoted by $\mathbf{R}_i = (R_1, R_2, \ldots, R_{w_i})$.

Observation 3 (Properties of the Temporal-Fairness Block): Regardless of channel condition and/or lag counter, the temporal-fairness block ensures that

1) only one flow is allowed to transmit in each time slot;
2) flow i transmits in exactly w_i time slots in a scheduling frame;
3) the maximum interaccess delay for flow i is bounded by

$$\max d_i = 2\sum_{j\neq i} w_j + 1 \tag{20}$$

4) the maximum inter-access delay of each flow increases by $2w$ when a flow with the weight of w becomes active.

Proof: See Appendix III. ■

Note that the explicit compensation in ORCA gives the allocation of one flow to another. Therefore, some flows might not be allowed to transmit in a scheduling frame. To preserve the above properties of the temporal-fairness block, unlike ORCA, ORCA-MRT does not employ the explicit compensation mechanism.

D. Transmission Block

Like in WFS [8], the channel state is always known right before the transmission. Therefore, the transmission rate is calculated based on the actual channel state rather than the predicted channel state. In column (or time slot) j, flow i transmits at the rate of r_{ij}. In this paper, we assume that

$$r_{ij} = \begin{cases} r_{\min} + s_{ij} - 1, & x_{hj} = 1 \quad \forall h \in \mathbf{R}_i \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where $r_{\min}$ is the minimum number of packets that each flow can transmit in a particular time slot (when the channel condition is the worst). More realistic calculation of r_{ij} should take into account several factors such as target BER, modulation technique, and strength of error correcting code. However, this is out of the scope of this paper.

E. Compensation Block

The compensation block calculates the lag counter of flow i (L_i) by using

$$L_i = \max_{\forall k}\left\{\sum_{\forall j} r_{kj}\right\} - \sum_{\forall j} r_{ij} \tag{22}$$

where $\sum_{\forall j} r_{ij}$ is the number of packets successfully transmitted by flow i up to the current time slot.

IV. Simulation Environment

A. Performance Measures

We define the following weight-independent performance metrics which are similar to those in [6].

- Normalized throughput for flow i (γ_i)

$$\gamma_i = \frac{\sum_{j=1}^{T_{ob}} r_{ij}}{T_{ob} w_i}. \tag{23}$$

- Normalized interaccess delay for flow i (D_i)

$$D_i = \frac{w_i}{TK_i}\sum_{k=1}^{K_i} d_i(k) \tag{24}$$

where r_{ij} is the transmission rate of flow i in time slot j, w_i is the weight of flow i, K_i is the number of transmission opportunities of flow i over an observation period (T_{ob}), $d_i(k)$ is inter-access delay of flow i from $(k-1)$th to kth transmission opportunities, and T is the scheduling frame size. Note that γ_i is the average number of packets transmitted by flow i per time slot and D_i is independent of scheduling frame size.

We measure the system performances in terms of throughput and throughput fairness by calculating the average value of γ_i (denoted by $\overline{\gamma}$) and the standard deviation (SD) of γ_i [denoted by $\sigma(\gamma)$], respectively. Similarly, the performances in terms of inter-access delay and temporal fairness are measured by the average value of D_i (denoted by $\overline{D}$) and the standard deviation of D_i [denoted by $\sigma(D)$], respectively. Note that smaller values of $\sigma(\gamma)$ and $\sigma(D)$ imply better performance in terms of throughput fairness and temporal fairness, respectively. Under perfectly fair time slot allocation, $\overline{D} = 1$. For the sake of brevity, the terms throughput and delay will be used in place of average normalized throughput and average normalized inter-access delay hereafter.

B. Simulation Parameters and Methodology

To compare the performance of ORCA-MRT with that of MFS for different values of system parameters, the simulation environment is assumed to be similar to that in [6]. The number of active flows is 16 (12 flows each with weight of one and four flows each with weight two). Perfect channel prediction is assumed to eliminate the effect of prediction inaccuracy on performance evaluation. We assume a five-state FSMC (5-FSMC) with $r_{ij} = \{2, 3, 4, 5, 6\}$ (i.e., the average transmission rate $r_{\text{avg}} = 4$). Note that, ignoring the noise term and setting $d = 3$ in (9) (as in [6]), the channel condition varies approximately in the range [0.2, 0.8], and due to the power constraint in (8) with $P = 2$, the overall transmission rate is selected from the set $r_{ij} = \{2, 3, \ldots, 10\}$ for which $r_{\text{avg}} = 4$. We assume that each flow always has data to transmit. To show the effects of channel correlation, we set $\rho_{\text{avg}} = \{0.5, 0.6, 0.7, 0.8, 0.9\}$. The simulation runtime is 2000 time slots.

Next, we study the performance of the two proposed prediction models as well as the upper bound and lower bound of system throughput. The effect of channel condition is analyzed in two ways. One is by varying the number of states in an equally likely channel where 3-FSMC ($r_{ij} = \{2, 3, 4\}$) and 7-FSMC ($r_{ij} = \{2, 3, \ldots, 8\}$) are simulated in comparison with the above 5-FSMC. Another is by fixing the number of states to seven and changing the shape of the steady state probability distribution. Finally, we study the effects of the number of flows and the corresponding weights on system performance.

V. Simulation Results and Discussions

A. Performance of ORCA-MRT and MFS

For ORCA-MRT and MFS, the average value and the standard deviation of the normalized throughput (per time slot) among all the flows are shown in Fig. 2. We observe that the use of greedy algorithm and stochastic approximation in MFS results in suboptimal solutions, and therefore, poorer throughput fairness (compared to that for ORCA-MRT). However,

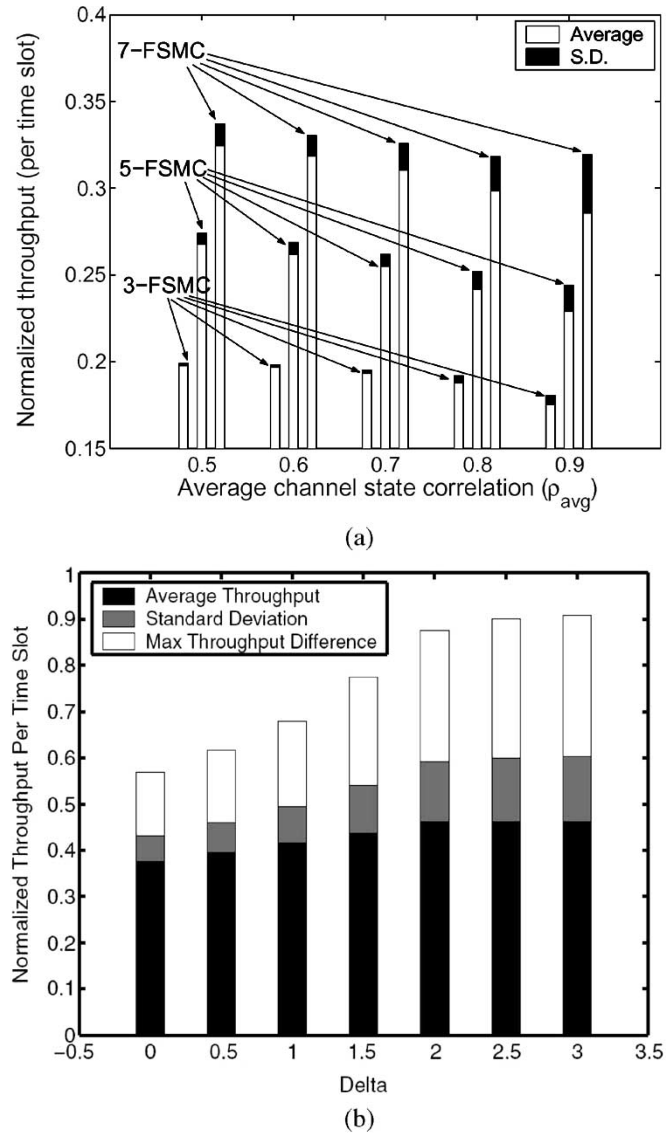

Fig. 2. Average value and standard deviation of normalized throughput for (a) ORCA-MRT and (b) MFS [6].

since MFS allows several flows to transmit simultaneously through CDMA, the resulting throughput for MFS is higher than that of ORCA-MRT. Note that, the actual throughput of MFS would not be as high as those reported in [6] when the non-orthogonality of the code channels is considered. Due to TDMA-based transmission, the inter-flow interference would be negligible for ORCA-MRT.

While MFS does not take into account the temporal-fairness constraint, ORCA-MRT shows robustness in delay and temporal fairness in that both $\overline{D}$ and $\sigma(D)$ are not affected by the channel condition. As can be observed from Fig. 3, the delay and the temporal-fairness performances of ORCA-MRT are fairly good ($\overline{D} \approx 1$ and $\sigma(D) < 0.005$) under all channel conditions.

B. Channel State Correlation

When the channel state correlation is small (small ρ_{avg}), a flow tends not to stay in the bad state for a long time. The

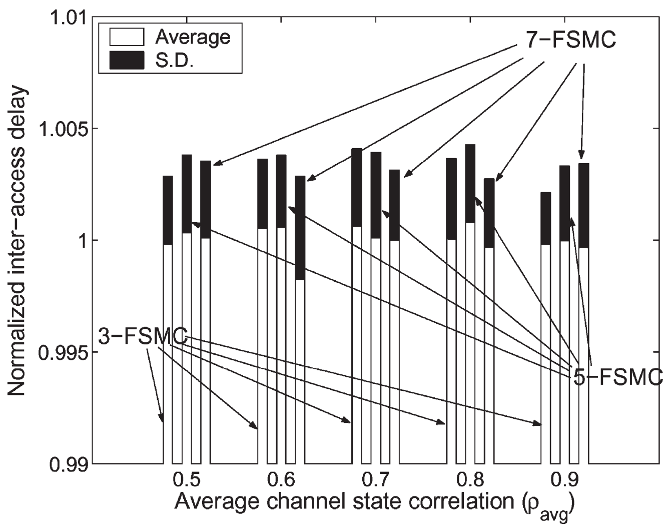

Fig. 3. Average value and standard deviation of normalized inter-access delay.

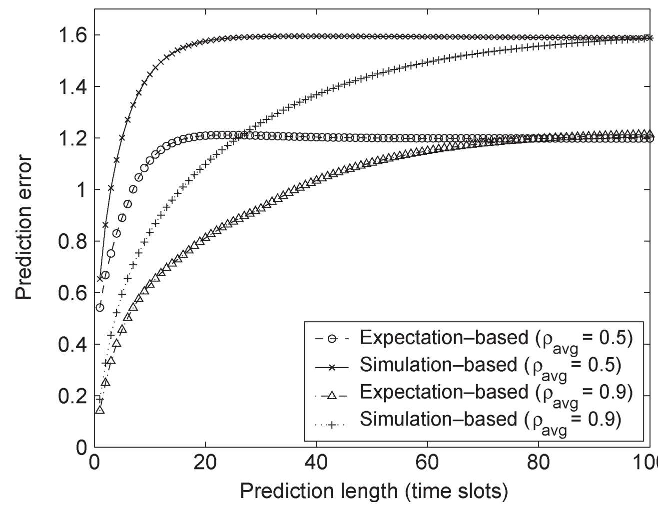

Fig. 4. Average prediction error versus prediction length.

scheduler is able to select the most suitable state, and therefore, is capable of improving performance in terms of both throughput and throughput fairness (Fig. 2).

C. Channel Prediction

Definition 3: The average prediction error $[\Delta(t)]$ is the mean absolute difference between predicted and actual channel states in any time slot t given that the last known state is in time slot $t = 0$. Mathematically

$$\Delta(t) = \sum_{i=1}^{M} \pi_i \sum_{s'=1}^{M} \sum_{s=1}^{M} |s - s'| P_{S,S'}^{(t)}(s, s') \tag{25}$$

where $P_{S,S'}^{(t)}(s, s')$ is the joint probability that in time slot t the actual state is s and the predicted state is s' given that the actual state in time slot $t = 0$ is known.

Theorem 1: For an FSMC model where the channel state at $t = 0$ is known, the average prediction error in time slot t for the expectation-based and the simulation-based prediction models can be calculated from (26) and (27), respectively

$$\Delta_E(t) = \sum_{i=1}^{M} \sum_{s=1}^{M} \pi_i \left| s - E[s_t|i] \right| p_{i,s}^{(t)} \tag{26}$$

$$\Delta_S(t) = 2 \sum_{i=1}^{M} \sum_{\tau=1}^{M-1} \sum_{s=1}^{M-\tau} \pi_i \tau p_{i,s}^{(t)} p_{i,s+\tau}^{(t)} \tag{27}$$

where the prediction length t is the interval (in terms of time slots) between the last known channel state and the predicted state, $E[s_t|i]$ is the predicted channel state obtained from the expectation-based prediction model [using (17)], and $p_{i,s}^{(t)}$ is the probability that state i will move to state s in t steps. In case of a random state channel model, where the channel state depends only on π_i $(\forall i)$, the average prediction error for expectation-based and simulation-based models can be calculated from

$$\Delta_E^{\text{rand}}(t) = \sum_{s=1}^{M} \pi_s \left| s - \sum_{i=1}^{M} i\pi_i \right| \tag{28}$$

$$\Delta_S^{\text{rand}}(t) = 2 \sum_{\tau=1}^{M-1} \sum_{s=1}^{M-\tau} \tau \pi_s \pi_{s+\tau}. \tag{29}$$

Proof: See Appendix IV. ■

Fig. 4 shows the effect of prediction length on the average prediction error when $M = 5$ and $\rho_{\text{avg}} \in \{0.5, 0.9\}$. We observe that in all the cases considered, the expectation-based channel prediction model always has a smaller prediction error than the simulation-based model. For small prediction length, the possible range of predicted states is limited (e.g., three possible states when $t = 1$) and the prediction is less likely to be erroneous. As the prediction length becomes larger, the possible values of the predicted states increase, the FSMC behaves more randomly, and the prediction becomes less accurate. When the prediction length is sufficiently long, the prediction of the FSMC converges to that of the random state channel that provides the upper bound for prediction error.

We observe from Fig. 4 that the average prediction error for both expectation-based and simulated-based models converges to $\Delta_E^{\text{rand}}(t) = 1.2$ and $\Delta_S^{\text{rand}}(t) = 1.6$ calculated from (28) and (29), respectively. We also observe that increased channel state correlation leads to better prediction accuracy because the channel tends to stay in the same state and the predicted channel states become more similar to the actual channel states. Increased channel state correlation also results in a smaller rate of increase in prediction error as prediction length increases.

With $M = 5$ and $\rho_{\text{avg}} \in \{0.5, 0.7, 0.9\}$, we observe that the simulation-based channel prediction model leads to inferior throughput and throughput-fairness performances due to prediction inaccuracy (Fig. 5). As channel states become more correlated, the prediction model performs better and the performance of ORCA-MRT with the simulation-based model becomes closer to that with the perfect channel prediction model.

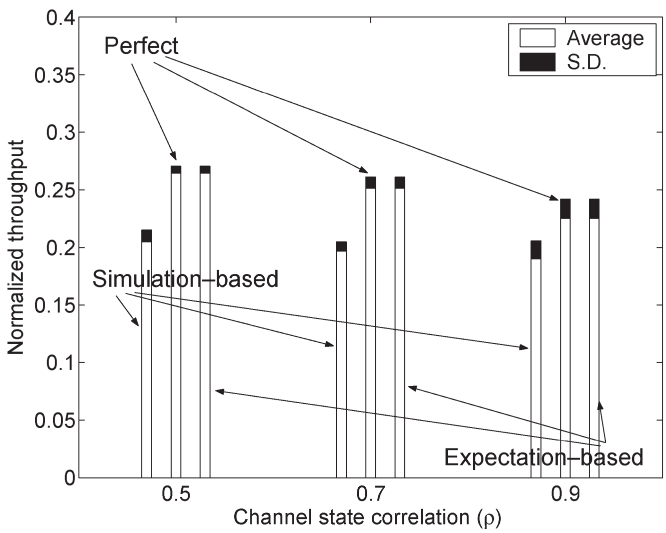

Fig. 5. Effect of imperfect channel prediction.

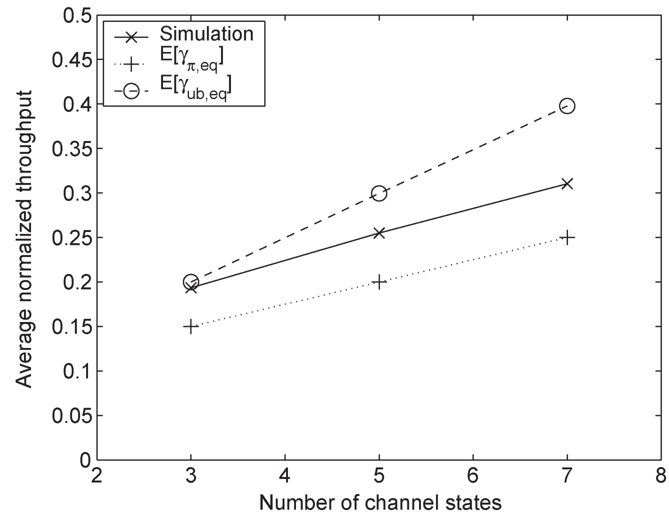

Fig. 6. Upper bound and lower bound of average normalized throughput.

The expectation-based model performs fairly good in that it almost results in the same throughput and throughput fairness as those obtained when the channel condition is known *a priori*. The performances in terms of delay and temporal fairness are very similar for all the prediction models (the results are omitted for brevity).

D. Upper Bound and Lower Bound of Average Normalized Throughput

Definition 4: Transmission state probability $(P(s^{(m)}))$ is the probability that a particular flow is allowed to transmit when the channel is in state m.

Let us consider a pure opportunistic scheduling algorithm where in each time slot the scheduler always allows the flow perceiving the best channel condition to transmit. It is well known that such an algorithm leads to maximum overall throughput at the expense of fairness.

Theorem 2: Consider a system at steady state (e.g., when T is sufficiently large) with n flows (each with weight one), where the scheduling frame size is $T = n$. For pure opportunistic scheduling, the upper bound of throughput $(E[\gamma_{ub}])$ and the corresponding $P(s^{(m)})$ can be obtained from (30) and (31), respectively. When steady-state probabilities are equally likely, both $(E[\gamma_{ub}])$ and $P(s^{(m)})$ reduce to $E[\gamma_{ub,eq}]$ and $P(s^{(m)})_{eq}$, as given by (32) and (33), respectively. That is

$$E[\gamma_{ub}] = \frac{1}{T}\left(r_{\min} - 1 + M - \sum_{m=1}^{M-1}(F_m)^n\right) \tag{30}$$

$$P\left(s^{(m)}\right) = (F_m)^n - (F_{m-1})^n \tag{31}$$

$$E[\gamma_{ub,eq}] = \frac{1}{T}\left(r_{\min} - 1 + M - \left(\frac{M-1}{2}\right)^n\right) \tag{32}$$

$$P\left(s^{(m)}\right)_{eq} = \frac{m^n - (m-1)^n}{M^n}. \tag{33}$$

In the above, M is the number of channel states and $F_m = \sum_{i=1}^{m} \pi_m$, where π_m is the steady-state probability that the channel is in state m.

Proof: See Appendix V. ■

In case of round-robin scheduling, where each flow is scheduled to transmit sequentially, $P(s^{(m)}) = \pi_m$ and the resulting throughput can be calculated from

$$E[\gamma_\pi] = \frac{1}{T}\left(r_{\min} - 1 + \sum_{m=1}^{M} m\pi_m\right). \tag{34}$$

Observation 4 (Upper Bound and Lower Bound of Throughput): The upper bound of throughput $(E[\gamma_{ub}])$ in (30) is always higher than $E[\gamma_\pi]$ obtained from (34).

Proof: See Appendix VI. ■

In Fig. 6, we plot both $E[\gamma_{ub,eq}]$ and $E[\gamma_{\pi,eq}]$ (i.e., the values of $E[\gamma_\pi]$ when the channel states are equally likely) obtained using (32) and (34), respectively. For comparison, we also plot the throughput of ORCA-MRT obtained via simulation with $\rho_{\text{avg}} = 0.7$, when $M = \{3, 5, 7\}$. We observe that the throughput for ORCA-MRT is lower bounded by $E[\gamma_{\pi,eq}]$ and upper bounded by $E[\gamma_{ub,eq}]$.

Fig. 7 plots the transmission state probabilities for the simulation scenario considered in Fig. 6. We observe that both $P(s^{(m)})_{eq}$: UB (upper bound) and $P(s^{(m)})_{eq}$: simulation (obtained via simulation) are shifted (from π or the lower bound case) toward the best channel state. Due to TDMA and fairness constraints in (2) and (3) as well as lag counter in (19), some flows perceiving good channel conditions might not be allowed to transmit. Therefore, the throughput of ORCA-MRT measured from the simulation is always lower than the upper bound in (30) when pure opportunistic scheduling is used, and higher than the lower bound in (34) when no optimization is performed.

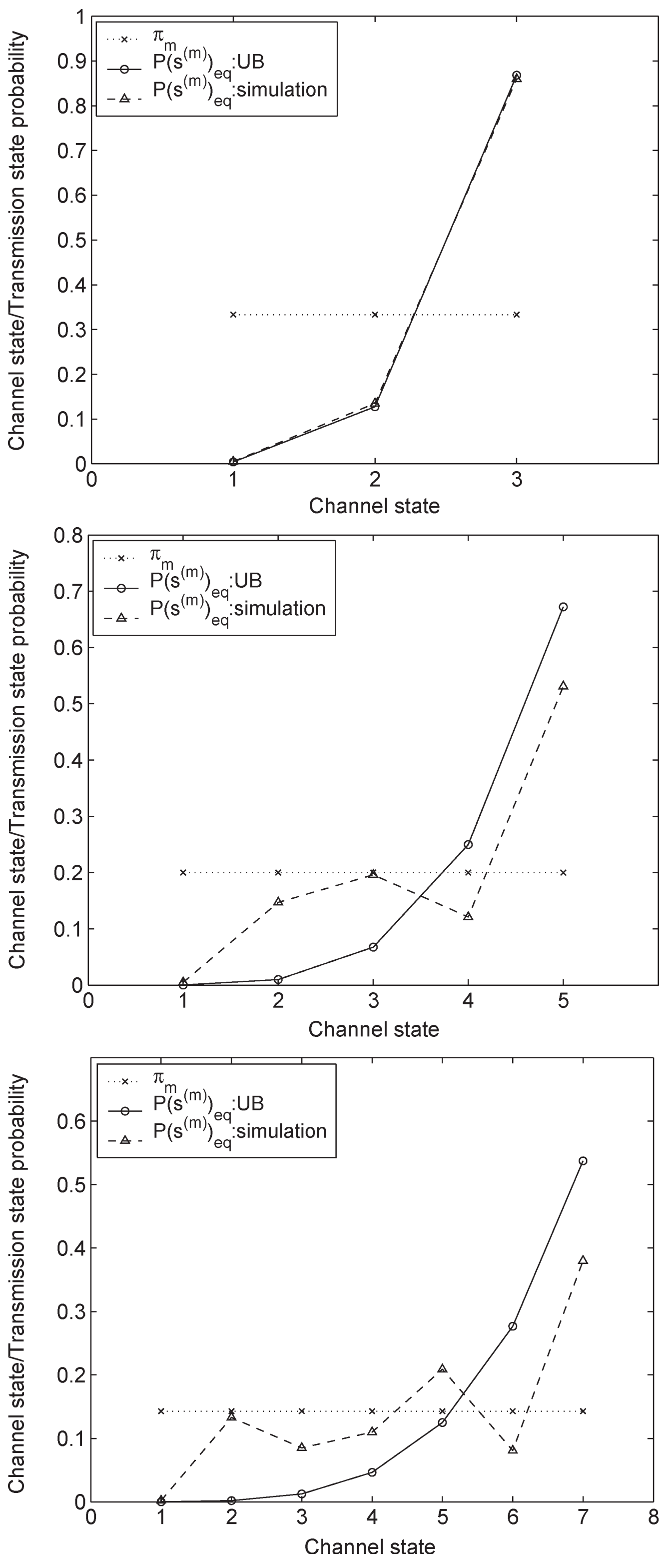

Fig. 7. Channel and transmission state probability versus channel state.

E. Number of FSMC States (M) in an Equally Likely Channel

In this section, we assume that the steady state probabilities are equally likely and study the effects of the number of FSMC states.

1) Throughput Performance: From Fig. 7, we observe that $P(s^{(m)})$ tends to cluster around higher states. Therefore, increasing the number of states in the FSMC model increases throughput (as in Fig. 6).

Due to (2), (3), and (19), the shape of $P(s^{(m)})$ obtained from the simulation of ORCA-MRT becomes less similar to the upper bound $P(s^{(m)})$ as the number of FSMC states increases (Fig. 7). As a result, the throughput performance diverges from the upper bound as the number of FSMC states increases (Fig. 6).

2) Fairness Performance: There are two key factors related to throughput fairness: equality in transmission rate and perfect compensation mechanism. When $P(s^{(i)}) = 1$ and $P(s^{(j)}) = 0$ $(\forall j \neq i)$, all flows transmit only when the channel is in state i and the allocation is perfectly fair. If there exists inequality in the transmission rate and the compensation mechanism is able to perfectly compensate for unequal allocation, there will be no unfairness.

In case of M-state equally likely channels, the conditional probability that all n flows (each with weight one) perceive the same $P(s^{(m)})$ during a scheduling frame, given that the allocation always satisfies (2) and (3), is

$$P_{\text{equal rate}}(M) = \sum_{m=1}^{M} n! \left(\frac{1}{M}\right)^n = n! \left(\frac{1}{M}\right)^{n-1}. \quad (35)$$

If the number of channel states is increased by a positive integer ξ, $P_{\text{equal rate}}(M+\xi) = n!(1/(M+\xi))^{n-1}$. Therefore, the relative incremental probability is

$$P_{\text{inc}} = \frac{P_{\text{equal rate}}(M+\xi)}{P_{\text{equal rate}}(M)} = \left(\frac{M}{M+\xi}\right)^{n-1}. \quad (36)$$

Since $P_{\text{inc}} \leqslant 1$, increasing the number of channel states $(\xi > 0)$ reduces the probability that all flows will be allocated the same transmission rate during a scheduling time frame.

In order to protect non-lagging flows from being disturbed, the compensation mechanism in ORCA-MRT is designed to be nonaggressive, where each flow is forced to transmit in exactly one time slot per scheduling frame. In each frame, the maximum number of compensations for a flow with the weight of one is limited to $M-1$ packets. Regardless of system parameters (e.g., M or n), this mechanism enables ORCA-MRT to preserve frame-based temporal fairness and bound inter-access delay (Fig. 3). Compared to the compensation mechanism, the inequality in transmission rate has stronger influence on fairness performance. Therefore, throughput fairness in ORCA-MRT degrades as M increases (Fig. 2).

F. Non-Equally Likely Channel With Fixed Number of States

In this section, we present simulation results for a 7-FSMC with $\rho_{\text{avg}} = 0.7$. We set the steady state probability for one of the states to 0.4 and that for all other states to 0.1. As the most likely state (with $\pi = 0.4$) moves toward better states, the throughput increases but the delay and temporal fairness remain unchanged. These results are fairly intuitive and are omitted for brevity.

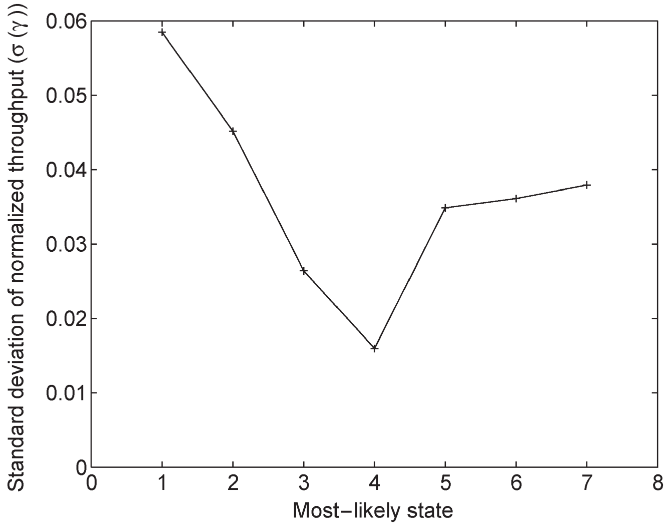

Fig. 8. Throughput fairness in a nonequally likely channel.

TABLE II
EFFECT OF WEIGHTS OF THE FLOWS

flow	1	2	3	4	5
w_i	1	2	3	4	5
$\max_{sim:\forall k}\{d_i(k)\}$	29	27	25	23	21
γ_i	0.228	0.263	0.224	0.276	0.246

Fig. 8 shows the effect of the location of the most likely state on throughput fairness. When data flows perceive good state for most of the time (e.g., $\pi_7 = 0.4$), lagging flows are not able to keep up with non-lagging flows due to the non-aggressive compensation mechanism of ORCA-MRT. On the other hand, when the channel is bad for most of the time (e.g., $\pi_1 = 0.4$), the lagging flows do not obtain proper compensation because they might perceive bad channel condition for an extended period of time. These two cases result in throughput unfairness (Fig. 8). When the channel is neither very good nor very bad (e.g., $\pi_4 = 0.4$), the non-lagging flows are more likely to transmit at a slower rate and the lagging flows are able to obtain proper compensation. From Fig. 8, we observe that the best throughput fairness is achieved when the most likely state is 4. When the most-likely state moves toward bad states, there will be more throughput unfairness compared to the case when it moves toward good states. The results for 3-FSMC and 5-FSMC have been observed to be similar to those for 7-FSMC and are omitted for brevity.

G. Number of Flows, Corresponding Weights, and Scheduling Frame Size

Table II shows the maximum inter-access delay $(\max_{sim:\forall k}\{d_i(k)\})$ for ORCA-MRT obtained from simulation when there are five flows with weights 1, 2, 3, 4, and 5, respectively, and the channel is defined by a 5-FSMC with $\rho_{\text{avg}} = 0.7$. We observe that the maximum inter-access delay is bounded by the values calculated from (20). Throughput fairness is observed to be fairly good in that the value of $\sigma(\gamma)$ calculated using γ_i is approximately 0.04 (Table II).

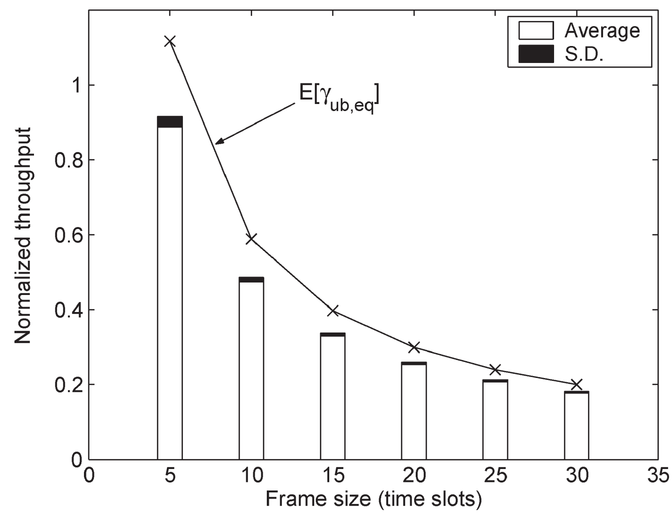

Fig. 9. Throughput versus scheduling frame size.

In ORCA-MRT, both the number of flows and the corresponding weights do not have direct impact on system performance as long as the frame size does not change. Let us assume that all flows have equal weight and the number of flows (n) is {5, 10, 15, 20, 25, 30}. The set of possible scheduling frame sizes (T) is {5, 10, 15, 20, 25, 30}. The values of $\overline{D}$ and $\sigma(D)$ are not affected by frame size since they are normalized by w_i and T. On the other hand, γ_i decreases as the frame size becomes larger because each flow has to wait longer for another transmission opportunity. From Fig. 9, we observe that the upper bound of throughput ($E[\gamma_{ub,eq}]$) calculated using (32) decreases as the frame size increases.

As the frame size becomes larger, the scheduler has more choices to allocate higher rate to each flow while maintaining fairness. The performance of ORCA-MRT with respect to the upper bound becomes better for a larger frame size.

VI. Directions for Further Studies

A. Decoupling of Temporal Fairness and Throughput Fairness

An ORCA-MRT-based scheduler treats temporal fairness and throughput fairness separately (Fig. 1). Temporal fairness is achieved by using the flows' weights while throughput fairness is implicitly achieved by the cost function. By introducing rate weight into the throughput-fair block [or (19)], we can explicitly change throughput fairness without affecting temporal fairness.

B. Designing Delay-Based Scheduler for Efficient End-to-End Performance

It is well known that, in case of TCP running over a semi-reliable radio link layer, timeouts at the senders (which trigger slow start) are largely responsible for end-to-end throughput degradation in wide-area wireless networks. The strength of ORCA-MRT is its ability to guarantee inter-access delay that can be utilized to assure that packets will be transmitted before TCP timeouts occur.

C. Complexity of the Scheduling Algorithm

The assignment problem can be considered as a special case of bipartite graph matching. The Hungarian method used in [2] has a time complexity of $O(k^3)$, where k is the number of rows/columns in the cost matrix. The readers are encouraged to refer to some other approaches such as successive shortest path algorithm, relaxation algorithm, cost scaling algorithm, or stable marriage problem [12], which can solve the assignment problem and might reduce the complexity of ORCA-MRT.

VII. Conclusion

In this paper, we have formulated the scheduling problem for fair bandwidth allocation as an assignment problem. Based on the Hungarian method to solve the assignment problem, the ORCA-MRT scheduler has been proposed for fair bandwidth allocation under MRT in a TDMA wireless network. ORCA-MRT utilizes optimization-based intraframe rate allocation along with fairness compensation using the lag counter. Also, to facilitate optimal bandwidth allocation, we have proposed two methods for channel prediction.

Simulation results have revealed that ORCA-MRT outperforms MFS in terms of throughput fairness and temporal fairness. The average throughput of ORCA-MRT is upper bounded and lower bounded by those obtained for the pure opportunistic and the round-robin scheduling algorithms, respectively. ORCA-MRT is able to ensure temporal fairness and bound the inter-access delay under a certain threshold. This feature can be utilized to prevent end-to-end throughput [e.g., TCP throughput] degradation in wide-area wireless networks. The proposed channel prediction methods perform almost as good as when the channel states are known *a priori*.

Appendix I
Proof of Observation 1

Let K be an integer. By adding a constant K to all elements, the minimum value in row i becomes $\min_i + K$. After step 1 in the Hungarian method, $c_{ij} = c_{ij}^{CA} + K - (\min_i + K) = c_{ij}^{CA} - \min_i$. Therefore, the transmission cost is not affected by adding a constant.

Appendix II
Proof of Observation 2

Consider a channel aware cost matrix $\mathbf{C}' = \begin{pmatrix} c & (c+H) \\ c & (c+I) \end{pmatrix}$. Assuming that $L_h = 0$ and applying (19) along with step 1 of the Hungarian method, $\mathbf{C}' = \begin{pmatrix} 0 & H \\ 0 & I(1+L_i) \end{pmatrix}$.

We observe that the cost of the lagging flow i in the second slot is increased by IL_i. Therefore, flow i is more likely to acquire the contention slot, or equivalently (19) can always favor the lagging flows. When the size of the cost matrix is more than 2, we can select the owner of the first contention slot, remove from the cost matrix the column and row corresponding to the owner of the first contention slot, and repeat the whole process with the reduced cost matrix.

Appendix III
Proof of Observation 3

From (2), there is exactly one flow transmitting in a time slot. There are w_i rows in the cost matrix corresponding to flow i. From (3), flow i transmits in exactly w_i time slots per frame. In two scheduling frames, flow i transmits in exactly $2w_i$ time slots. Therefore, the maximum interaccess delay is bounded to $(2\sum_{j\neq i} w_j + 1)$. When a flow with the weight of w becomes active, the scheduling frame size and, therefore, inter-access delay increase by $2w$.

Appendix IV
Proof of Theorem 1

Let $\Delta(t;i)$ be the average prediction error when the last known state is i and the probabilities that the predicted state (S') and the actual channel state (S) are s' and s in time slot t be represented by $P_{S'}^{(t)}(s')$ and $P_{S}^{(t)}(s)$, respectively. Since the predicted and the actual channel states are identical and independent to each other, $P_{S'}^{(t)}(s) = P_{S}^{(t)}(s) = p_{i,s}^{(t)}$ and

$$\Delta(t) = \sum_{i=1}^{M} \pi_i \Delta(t;i) \tag{37}$$

where

$$\Delta(t;i) = \sum_{s'=1}^{M}\sum_{s=1}^{M} |s-s'| p_{i,s}^{(t)} p_{i,s'}^{(t)}. \tag{38}$$

A. Proof of (26)

From (17)

$$\Delta_E(t) = \sum_{i=1}^{M} \pi_i \sum_{s=1}^{M} |s - E[s_t|i]|\, p_{i,s}^{(t)}. \tag{39}$$

B. Proof of (27)

Let $R_\tau(t;i)$ be the correlation function of $p_{i,s}^{(t)}$, which represents the probability that the states of two identical FSMC processes starting at state i will differ by τ after t step transitions. $R_\tau(t;i)$ can be calculated from

$$\begin{aligned} R_\tau(t;i) &= \sum_{0\leqslant s+\tau\leqslant M} p_{i,s}^{(t)} p_{i,s+\tau}^{(t)} + \sum_{0\leqslant s-\tau\leqslant M} p_{i,s}^{(t)} p_{i,s-\tau}^{(t)} \\ &= \sum_{s=1}^{M-\tau} 2 p_{i,s}^{(t)} p_{i,s+\tau}^{(t)}. \end{aligned} \tag{40}$$

The expectation of $R_\tau(t;i)$ over all values of τ $(E_\tau[R_\tau(t;i)])$ can be calculated from

$$\begin{aligned} E_\tau[R_\tau(t;i)] &= \sum_{\forall\tau} \tau R_\tau(t;i) \\ &= \sum_{\tau=1}^{M-1}\sum_{s=1}^{M-\tau} 2\tau p_{i,s}^{(t)} p_{i,s+\tau}^{(t)}. \end{aligned} \tag{41}$$

Equivalent to $\Delta(t;i)$, $E_\tau[R_\tau(t;i)]$ is an average prediction error in time slot t when the initial state is i. To verify this, we replace $s' = s+\tau$ in (38) and obtain $\Delta(t;i) = E_\tau[R_\tau(t;i)]$. Therefore

$$\Delta_S(t) = \sum_{i=1}^{M} \pi_i E_\tau\left[R_\tau(t;i)\right] = \sum_{i=1}^{M} \pi_i \sum_{\tau=1}^{M-1} \sum_{s=1}^{M-\tau} 2\tau p_{i,s}^{(t)} p_{i,s+\tau}^{(t)}. \tag{42}$$

C. Proof of (28) and (29)

In case of a random state channel model, $p_{i,s}^{(t)} = \pi_s (\forall i, t)$ and the average channel state is $\sum_{s=1}^{M} s\pi_s$. By replacing these $p_{i,s}^{(t)}$ and the average channel state into (37) and (38), (28) is proven. In common with the proof of (27), (29) can be proven by replacing $p_{i,s}^{(t)}$ and $p_{i,s+\tau}^{(t)}$ with π_s and $\pi_{s+\tau}$, respectively.

APPENDIX V
PROOF OF THEOREM 2

A. Proof of (31)

Let s_k be the channel state of flow k in a particular time slot. Then

$$P\left(s^{(m)}\right) = Pr\left\{(s_k = m \quad \exists k) \quad \text{and} \quad (s_k < m+1 \quad \forall k)\right\} = \left(1-(1-\pi_{m|m})^n\right)\left(\sum_{i=1}^{m}\pi_i\right)^n \tag{43}$$

where $\pi_{m|m}$ is the conditional steady state probability that the channel is in state m given that the best channel state experienced among all n flows is m. Therefore

$$\pi_{m|m} = \frac{\pi_m}{\sum_{i=1}^{m}\pi_i} \tag{44}$$

$$P\left(s^{(m)}\right) = \left(\sum_{i=1}^{m}\pi_i\right)^n - \left(\sum_{i=1}^{m-1}\pi_i\right)^n. \tag{45}$$

B. Proof of (30)

During T time slots, a flow with weight one transmits in exactly one time slot with the transmission rate defined in (21). Therefore

$$E[\gamma_{ub}] = \frac{1}{T}\left(r_{\min} - 1 + \sum_{m=1}^{M} mP\left(s^{(m)}\right)\right) = \frac{1}{T}\left(r_{\min} - 1 + M - \sum_{m=1}^{M-1}\left(\sum_{i=1}^{m}\pi_i\right)^n\right).$$

C. Proof of (32) and (33)

For an equally likely channel, $\pi_m = 1/M$ and $\sum_{i=1}^{m}\pi_i = m/M$. From (45) and (46)

$$P\left(s^{(m)}\right)_{eq} = \frac{m^n - (m-1)^n}{M^n}$$

$$E[\gamma_{ub,eq}] = \frac{1}{T}\left(r_{\min} - 1 + M - \frac{\left(\sum_{m=1}^{M-1} m\right)^n}{M^n}\right).$$

APPENDIX VI
PROOF OF OBSERVATION 4

Let $\Delta = E[\gamma_{ub}] - E[\gamma_\pi]$. Using (30) and (34)

$$\Delta = \frac{1}{T}\left(M - \left\{\sum_{m=1}^{M-1}\left(\left(\sum_{i=1}^{m}\pi_i\right)^n + m\pi_m\right) + M\pi_M\right\}\right) = \frac{1}{T}\left(M - X(n)\right) \tag{46}$$

where $X(n) = \{M\pi_M + \sum_{m=1}^{M-1}((\sum_{i=1}^{m}\pi_i)^n + m\pi_m)\}$ is a decreasing function in n

$$X(1) = M\pi_M + \sum_{m=1}^{M-1}(M-m)\pi_m + \sum_{m=1}^{M-1} m\pi_m = M. \tag{47}$$

Therefore, $\Delta \geqslant 0\ (\forall n)$ and $E[\gamma_{ub}] \geqslant E[\gamma_\pi]$.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Diamond for his valuable technical advice and the anonymous reviewers for their constructive comments.

REFERENCES

[1] T. Nandagopal, S. Lu, and V. Bhargharvan, "A unified architecture for the design and evaluation of wireless fair queuing algorithms," in *Proc. ACM Mobile Computing and Networks (MobiCom)*, Seattle, WA, Aug. 1999, pp. 132–142.

[2] T. Issariyakul and E. Hossain, "Optimal radio channel allocation for fair queuing in wireless data networks," in *Proc. IEEE Int. Conf. Communications (ICC)*, Anchorage, AK, May 2003, pp. 142–146.

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

[4] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proc. ACM Mobile Computing and Networks (MobiCom)*, Atlanta, GA, Sep. 2002, pp. 24–35.

[5] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

[6] Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channels," in *Proc. IEEE Annu. Conf. Computer Communications (INFOCOM)*, San Francisco, CA, Mar. 2003, pp. 1106–1115.

[7] L. Cooper and D. Steinberg, *Methods and Applications of Linear Programming*. Philadelphia, PA: Saunders, 1974.

[8] S. Lu, T. Nandagopal, and V. Bharghavan, "A wireless fair service algorithm for packet cellular networks," in *Proc. ACM Mobile Computing and Networks (MobiCom)*, Dallas, TX, Oct. 1998, pp. 10–20.

[9] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 43, no. 1, pp. 163–171, Feb. 1995.

[10] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[11] K. S. Trivedi, *Probability and Statistics With Reliability, Queuing and Computer Science Applications*. New York: Wiley, 2002.

[12] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

Teerawat Issariyakul (S'02) received the B.Eng. degree in electrical engineering from the Thammasat University, Bangkok, Thailand, in 1997, the M.Eng. degree in telecommunications from the Asian Institute of Technology (AIT), Pathumthani, Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Winnipeg, MB, Canada, in October 2005.

He was a recipient of the University of Manitoba Graduate Fellowship. Furthermore, he held a scholarship from TRLabs, Winnipeg, MB, Canada. Since November 2005, he has been with the University of British Columbia, Vancouver, Canada, as a Postdoctoral Fellow. His main research interests are in the area of modeling, analysis, and optimization of protocols for wireless networks.

Dr. Issariyakul is a Member of the IEEE Communications Society.

Ekram Hossain (S'98–M'00) received the B.Sc. and M.Sc. degrees in computer science and engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1995 and 1997, respectively, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, Canada, in 2001.

He is an Assistant Professor at the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada. He leads the Wireless Internet and Packet Radio Network Research Group at the Department of Electrical and Computer Engineering, University of Manitoba. His research interests include radio link control, transport layer protocol design, and cross-layer optimization issues for next-generation wireless networks.

Dr. Hossain served as one of the Guest Editors for the Special Issue of *Wireless Communications and Mobile Computing* (Wiley Interscience, 2002) on "Radio Link and Transport Protocol Engineering for Future-Generation Wireless Mobile Data Networks." He also served as one of the Guest Editors for the Special Issue of the *IEEE Canadian Journal of Electrical and Computer Engineering* on "Advances in Wireless Communications and Networking." He served as a Technical Program Committee Member for IEEE ICC'06, ICC'05, WCNC'05, WCNC'04, GLOBECOM'04, GLOBECOM'03, and IFIP Networking'05. He is serving as one of the Guest Editors for the Special Issue of *IEEE Communications Magazine* on "Cross-layer Protocol Engineering for Wireless Mobile Networks." He currently serves as the Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the *IEEE/KICS Journal of Communications and Networks*, *Wireless Communications and Mobile Computing* (Wiley Interscience), and *International Journal of Sensor Networks* (Inderscience Publishers). He was the recipient of the Lucent Technologies, Inc. research award for his contribution to the IEEE International Conference on Personal Wireless Communications (ICPWC), 1997. He was a University of Victoria Fellow.