Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2006, Article ID 43759, Pages 1–15
DOI 10.1155/WCN/2006/43759

Adaptive Rate-Scheduling with Reactive Delay Control for Next Generation CDMA Wireless Mobile Systems

Oliver Yu, Emir Saric, and Anfei Li

Department of ECE, University of Illinois at Chicago, 851 S. Morgan Street, 1020 SEO, Chicago, IL 60607, USA

Received 1 October 2005; Revised 11 March 2006; Accepted 26 May 2006

To minimize QoS degradations during nonstationary packet loadings, predictive rate schedulers adapt the operation according to anticipated packet arrival rates deduced via specified estimation algorithm. Existing predictive rate schedulers are developed under the assumption of perfect estimation, which may not be possible in future CDMA-based cellular networks characterized with highly nonstationary and bursty traffic. Additional shortcoming of existing rate schedulers is the coupling of delay and bandwidth, that is, close interdependence of delay and bandwidth (rate), whereby controlling one is accomplished solely by changing the other. In order to mitigate for the arrival rate estimation errors and delay-bandwidth coupling, this paper presents the feedback-enhanced target-tracking weighted fair queuing (FT-WFQ) rate scheduler. It is an adaptive rate scheduler over multiclass CDMA systems with predictive adaptation control to adapt to nonstationary loadings; and feedback-enhanced reactive adaptation control to counteract arrival rate estimation errors. When the predictive adaptation control is not able to maintain long-term delay targets, feedback information will trigger reactive adaptation control. The objective of FT-WFQ scheduler is to minimize deviations from delay targets subject to maximum throughput utilization. Analytical and simulation results indicate that FT-WFQ is able to substantially reduce degradations caused by arrival rate estimation errors and to minimize delay degradations during nonstationary loading conditions.

1. INTRODUCTION

Next generation CDMA-based cellular wireless networks are slated to provide wide range of integrated multimedia services with a guaranteed quality of service (QoS) (e.g., voice, video, high-speed data). This, in turn, will create heterogeneous traffic environment characterized with highly nonstationary and bursty transmissions. The universal mobile telecommunication system (UMTS) is a 3rd generation (3G) mobile communication system developed by 3rd generation partnership project (3GPP). It defines "per-class" QoS provisioning, and classifies all traffic into four QoS classes, namely conversational, streaming, interactive, and background [1]. Each class has its own connection-level (or call-level) QoS requirements in terms of connection blocking/dropping probabilities, as well as application-level QoS requirements in terms of delay, jitter, throughput, BER, and burstiness. QoS provisioning with performance differentiation in a heterogeneous nonstationary environment requires efficient call admission control (CAC) and medium access control (MAC) protocols.

Given limited wireless resources, CAC enables connection-level QoS guarantees by implementing class-prioritized admission control. It also enables minimum application-level performance guarantees by limiting the total number of admitted connections. However, due to the bursty nature of packet traffic (especially from the connections of nonreal-time classes) CAC alone is not adequate to provide optimal resource utilizations and application-level performance. MAC algorithm that includes efficient *packet scheduler* needs to accompany an admission controller. It is responsible for provisioning differentiated application-level QoS requirements to admitted connections by providing optimal resource allocations. This paper focuses on packet scheduler part of MAC algorithm that accompanies admission controller proposed in [2].

Efficient packet scheduler is crucial for QoS provisioning in an integrated multiclass packetized network. Some of the desirable properties of a packet scheduler providing "per-class" QoS support in a wireless network include efficient link utilization with optimal resource distributions, delay bound guarantees for each class, bit-error-rate (BER) guarantees, throughput guarantees, delay-bandwidth decoupling, and low complexity.

Many packet scheduling algorithms have been proposed for CDMA-based wireless networks. The capacity of CDMA systems (especially in uplink) is interference-limited, subject to the variation of signal-to-interference ratios (SIRs), and bandwidth demands of users with limited power constraints. CDMA system loading factor can be derived to denote interference-based CDMA resources occupied by transmitting users. The schemes in [3–6] utilize interference-based loading and the variants of generalized processor sharing (GPS) fair scheduling discipline to dynamically allocate transmission rates and schedule packets in a CDMA-based system. Specifically, the authors in [3, 4] propose code-division GPS (CDGPS) scheduling scheme that maximizes throughput by providing "weighted fairness" (i.e., relative provisioning) in terms of the rate and signal-to-interference ratio (SIR) guarantees. Similarly the scheme in [5] proposes a rate scheduler with explicit BER guarantees in a wideband CDMA system. The scheme in [6] is a rate scheduler based on the adjusted GPS concept that explicitly takes into account current channel conditions. It maximizes total throughput by providing "weighted-fair" rate allocations with BER guarantees. The scheduler in [7] controls transmission power and dynamically allocates transmission rates so as to maximize the number of users whose BER is satisfied. To solve such an optimization problem, the authors suggest search procedure based on the genetic algorithm. One of the drawbacks of the aforementioned schemes [3–7] is the *delay-bandwidth coupling* whereby interdependence of delay and bandwidth (e.g., reducing delay implies a larger bandwidth allocation) could lead to resource underutilizations. The importance of delay-bandwidth decoupling is even more signified in the future multimedia wireless networks supporting traffic with similar delay but considerably different bandwidth requirements or vice versa.

Schedulers in [8, 9] dispense with delay-bandwidth coupling in a time-division-duplex (TDD) CDMA system by utilizing packet-prioritization that arrange transmissions in order to explicitly reduce packet delays. Similarly, authors in [10, 11] propose token bank fair queuing (TBFQ) scheduling algorithm that provides soft QoS guarantees. TBFQ keeps track of previous transmissions and introduces a priority index that determines which connections can utilize excess resources.

Time varying fair queuing (TVFQ) scheme in [12] is motivated by the delay-bandwidth decoupling problem. It extends dynamic (weighted) fair queuing concept into multicode (MC) CDMA systems. TVFQ decouples delay and bandwidth by solving a nonlinear integer programming problem that explicitly minimizes queuing delays and produces optimal weight (rate) assignments on a time-varying basis. The authors present computationally efficient solution method based on dynamic programming. However, the problem with TVFQ algorithm as well as the adaptive rate schedulers in [3, 4] is that they rely upon the perfect estimations of the future traffic arrival rates (or queue size); estimation errors would degrade their performance. Due to nonstationary traffic expected in the future wireless networks, arrival rate estimation errors are imminent. Consequently, estimation errors could lead to inefficient and erroneous resource distributions (i.e., rate assignments) whereby over-provisioning of some traffic classes might occur even when other classes are not meeting QoS targets. Moreover, TVFQ adapts weights (or rates) based on the future queue size (and predefined priority indices) without any regard to absolute delay targets. In a highly nonstationary environment characterized with frequent packet bursts, however, it is possible to have a connection with large instantaneous queue size (due to sudden arrival burst) but whose *mean* delay is significantly below its delay target. Hence, to utilize resources efficiently in a nonstationary traffic environment, adaptive rate scheduler needs additional delay *target-tracking* constraints so as to minimize delay deviations from absolute targets.

In order to dispense with delay-bandwidth coupling as well as to counteract arrival rate estimation errors and to achieve efficient resource distributions with absolute delay target-tracking, this paper proposes feedback-enhanced target-tracking weighted fair queuing (FT-WFQ) scheduler. It dynamically adapts transmission rates on a "per-class" basis such as to minimize overall delay deviations from absolute delay targets subject to maximum throughput utilization. FT-WFQ utilizes predictive adaptation control based on estimated arrival rates, but it also implements concurrent feedback-enhanced reactive control that detects imperfections, such as estimation errors, and counteracts them. Feedback control unit monitors average delays of each class and if it detects that a class is degraded (possibly because of estimation errors) it corrects the problem in order to achieve efficient resource distributions and minimize overall delay deviations from corresponding delay targets.

This paper is organized as follows. In Section 2, system model as well as problem statements are described. Then, in Section 3, CDMA "bandwidth" in terms of the interference-based loading is derived. Also, maximum loading-capacity is computed. The proposed FT-WFQ scheduler is thoroughly presented in Section 4. In Section 5 analysis and simulation models for performance evaluation are presented. Section 6 displays numerical results and comparison. Finally, Section 7 concludes the paper.

2. SYSTEM MODEL AND PROBLEM STATEMENTS

2.1. System model

Uplink scheduling in a single cell of a wireless cellular system that uses CDMA is considered. The cell contains mobile users requesting packet transmission (i.e., seeking access to CDMA resources) and the base station (BS) which centrally implements scheduling algorithm and optimally allocates resources on a dynamic basis (i.e., every scheduling time interval). Transmission is packetized with fixed-length packets and time is divided into frames of equal length T_f (e.g., $T_f = 10$ ms in UMTS). A class-based system is assumed where packets of each user belong to one of N traffic-classes (e.g., $N = 4$ in UMTS). Packets of each class have a distinct time-out value (i.e., delay requirement denoted by a delay target) measured in frames and if not transmitted by this time, they are useless. It is assumed that each mobile user has a large enough buffer, so that packets are lost only if not scheduled and transmitted on time (time-out expiration), and not due to buffer overflow.

Let the maximum uplink capacity of CDMA system (i.e., resource capacity measured in terms of interference-based loading) in the nth time interval be denoted by $\eta_T[n]$. The maximum capacity in terms of CDMA loading, subject to BER constraints, is analytically derived in the next section (Section 3). In each scheduling time interval n, the job of a packet scheduler is to optimally allocate the available capacity among active (i.e., transmitting) mobile users.

It is implied that the admission control has been conducted previously, and only users that are admitted into the system can send packet transmission requests. Admission control is such that each traffic class i ($i = 1, 2, \ldots, N$) is guaranteed minimum allocation rate $R_{i,\min}$ (packets/frame). Packet scheduling is performed to further exploit bursty nature of user traffic.

Consequently, whenever active admitted users have packets ready for transmission in the *next* time interval, they send packet transmission requests (i.e., small signaling packets) to the base station (BS) in the *current* time interval on special uplink random access request channels. It is assumed in this paper that some efficient random access technique is employed and that random access delay is negligible. Users seeking medium access indicate the number of packets ready for transmission in the next time interval, as well as traffic class of each packet.

BS collects all transmission requests (i.e., small signaling packets) for the following time interval. It first *classifies* all requests according to their traffic class and then places classified packet requests (one for each packet requested) into N traffic queues on a first-come-first-served basis (see Figure 1). Note that besides new packet requests each queue may also contain unexpired backlogged packet requests that were invoked in previous intervals but were not accommodated for transmission yet.

At the end of the scheduling time interval BS performs the proposed adaptive scheduling algorithm as explained in next sections. The algorithm returns the optimal rate allocations (in packets/frame) for each traffic-queue that would minimize delay cost function, namely R_i^*, $i = 1, 2, \ldots, N$. Based on these, BS notifies the owners (i.e., users) of the R_i^* head-of-line packet requests in the traffic-queue i ($i = 1, 2, \ldots, N$) that they are granted permission to transmit in the next interval. The notification is through a downlink broadcast control channel. After listening to broadcast control channel, mobile users, which are granted permission to transmit, forward their packets to BS on uplink dedicated channels in the corresponding frame of time interval $n + 1$. The whole process is repeated every scheduling time interval.

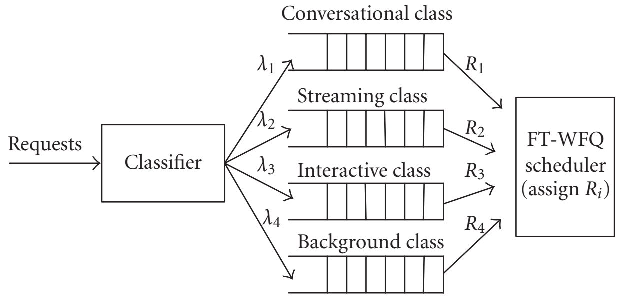

Figure 1: Packet scheduler at base station (with $N = 4$ traffic classes).

2.2. Problem statements

2.2.1. Packet arrival rate estimation errors and efficient resource distributions

Adaptation of the existing dynamic scheduling schemes such as time varying fair queuing (TVFQ) [12] is highly sensitive on the real-time estimation of future packet arrival rates (or some other measure of future traffic). Traffic in the future wireless networks is, however, expected to be highly nonstationary. Due to small cell size and increased handoff rates even traffic of real-time classes (conversational and streaming) observed at BS is expected to fluctuate and be nonstationary. The performance of the adaptive scheduler degrades in the presence of arrival rate estimation errors inherent in nonstationary environment.

Estimation errors could lead to inefficient resource (i.e., rate) distributions and unequal delay deviations from the targets. For instance, classes whose arrival rate is overestimated will (erroneously) allocate more resources than needed to keep their delays at the corresponding targets. This, in turn, will capture resources from other classes whose delay as a result might rise above targets. Consequently, this could lead to a situation where for some classes, large negative deviations from delay targets could be present even when positive delay deviations are observed for other classes. Ideally, however, there should not be any negative deviations when positive ones are observed.

Authors in [13] suggest three prediction techniques for estimating packet arrival rate $\widehat{\lambda}_i[n]$ of class i at the current time interval n. First one is to use the arrival rate observed at the previous time interval. The second one is to use average arrival rate based on observed history (i.e., $\widehat{\lambda}_i[n] = \sum_{j=1}^{n-1} \lambda_i[j]/(n-1)$). The third method they suggest is based on moving average of the first two methods. From these estimation techniques it is evident that they are prone to errors in a highly nonstationary environment subject to sudden bursts of packet arrivals.

Furthermore, TVFQ scheduler does not consider absolute delay targets when dynamically adapting weights (or rates). In a highly nonstationary environment even under perfect traffic estimations it is possible to have connections whose traffic queue size is large due to sudden traffic bursts but whose mean delay is significantly below corresponding delay target. Thus, for efficient rate adaptations, target tracking constraints that minimize delay deviations need to be incorporated.

2.2.2. Delay-bandwidth coupling

One of the major shortcomings of dynamic rate scheduler, such as the ones based on GPS, is the coupling of delay and bandwidth. It refers to close interdependence of delay and rate (i.e., bandwidth) parameters, whereby provisioning one parameter (e.g., delay) can only be accomplished by changing the other (e.g., rate). For instance, in GPS, the delay of a class-queue is controlled by changing its allocated rate (i.e., bandwidth). Since delay and bandwidth cannot be modified independently, the BS scheduler would allocate high rate to a class-queue with low delay requirement even if this class has low bandwidth requirement. This would lead to high bandwidth underutilizations. Delay-bandwidth coupling problem is even more signified in a future multiclass environment where classes with similar delay requirements might have significantly different bandwidth requirements (e.g., voice and video). In order to utilize resources efficiently, a dynamic scheduler needs to decouple delay and bandwidth such that both parameters can be guaranteed independently.

3. LOADING AND MAXIMUM LOADING CAPACITY IN CDMA SYSTEM

This section presents the concept of loading as an integrated measure of resource-usage in a multiclass CDMA system. The maximum possible loading capacity subject to BER constraints is also derived. These results are used by the dynamic resource monitor of the proposed scheduler as explained in detail in Section 4.

3.1. CDMA interference-based loading

Let $G_{p,i}$ be the processing gain (or the spreading factor) of a user that belongs to traffic-class i ($i = 1, \dots, N$), defined as $G_{p,i} = W/r_i$, where W is the system bandwidth in Hz (or chip rate), and r_i is the bit rate of a *user* of traffic class i. The signal energy per bit to noise-plus-interference ratio $(E_b/I_0)_i$ of a user of class i, $i = 1, 2, \dots, N$ (observed at BS) is given as

$$\left(\frac{E_b}{I_0}\right)_i = G_{p,i} \cdot \frac{S_i}{I_{\text{total}} - S_i}, \tag{1}$$

where S_i is the received signal power of a user of class i, and I_{total} is the total received wideband power including thermal noise power P_N in the BS. Assume a perfect power control such that the received power levels S_i of all users belonging to the same class i are equal. Let γ_i be the minimum value of $(E_b/I_0)_i$ required for acceptable BER (for a user of class i). Therefore, for satisfactory BER, the following constraints need to be satisfied ($\forall i$):

$$\left(\frac{E_b}{I_0}\right)_i = G_{p,i} \cdot \frac{S_i}{I_{\text{total}} - S_i} \geq \gamma_i. \tag{2}$$

It can be shown that the received power levels are minimized when the above equation is satisfied with equality. Let S_i^* be the received power level of a user of class i such that the above equation is satisfied with equality. Thus,

$$S_i^* = \frac{1}{1 + G_{p,i}/\gamma_i} \cdot I_{\text{total}}. \tag{3}$$

Note, however, that the received power level S_i is bounded by the maximum value $S_{i,\max}$ which is dependent on (mobile) transmit power, and achieving feasible $S_i^* \leq S_{i,\max}$ is a requirement that limits maximum interference I_{total} that a system is able to tolerate, as elaborated in the next subsection. Let the load factor increment $\Delta\eta_i$ of a user of class i be defined as $\Delta\eta_i \equiv S_i^*/I_{\text{total}}$.Therefore,

$$\Delta\eta_i = \frac{1}{1 + G_{p,i}/\gamma_i}. \tag{4}$$

Assuming N_i users of class i are in the system, I_{total} is given as

$$I_{\text{total}} = \sum_{i=1}^{N} N_i \cdot S_i^* + P_N. \tag{5}$$

Using terminology of the last section, note that the bit rate of the "class" i is given as $R_i = N_i \cdot r_i$. Let noise rise NR be defined as the ratio of total received wideband noise power in BS to the thermal noise power ($NR = I_{\text{total}}/P_N$). Substituting into the above formulas,

$$NR = \frac{I_{\text{total}}}{P_N} = \frac{1}{1 - \sum_{i=1}^{N} N_i \cdot \Delta\eta_i} = \frac{1}{1 - \eta}, \tag{6}$$

where η ($\eta \geq 0$) is defined as *loading*:

$$\eta = \sum_{i=1}^{N} N_i \cdot \Delta\eta_i = 1 - \frac{P_N}{I_{\text{total}}} \leq \eta_T. \tag{7}$$

The loading represents the amount of resources used in a CDMA system (when corresponding bit rates are allocated), and it defines the so-called "CDMA bandwidth."

3.2. Maximum loading capacity

Theoretically, the maximum loading, denoted as η_T, is 1. In reality, however, η_T is limited by utmost interference (or loading) a system is able to tolerate (given BER and limited power $S_{i,\max}$ constraints). From the above (5) and (7), the total interference I_{total} can be expressed in terms of loading η as $I_{\text{total}} = P_N/(1-\eta)$. Then, the BER constraints of (2) become

$$G_{p,i} \cdot \frac{S_{i,\max}}{[P_N/(1-\eta)] - S_{i,\max}} \geq \gamma_i, \quad \forall i. \tag{8}$$

Equivalently,

$$\frac{P_N}{1-\eta} \leq \frac{G_{p,i} \cdot S_{i,\max}}{\gamma_i} + S_{i,\max}, \quad \forall i \tag{9}$$

or, in terms of loading η,

$$\eta \leq \left(1 - \frac{P_N}{(G_{p,i} \cdot S_{i,\max})/\gamma_i + S_{i,\max}}\right) < 1, \quad \forall i. \tag{10}$$

Therefore, loading bound, or the maximum loading η_T tolerated by a system is given as

$$\eta_T = \min_{\forall i}\left[1 - \frac{P_N}{(G_{p,i} \cdot S_{i,\max})/\gamma_i + S_{i,\max}}\right]. \tag{11}$$

4. FEEDBACK-ENHANCED TARGET-TRACKING WEIGHTED FAIR QUEUING (FT-WFQ)

Two versions of FT-WFQ rate scheduling scheme are proposed, namely, heuristic and optimal. The proposed scheme is characterized with the following features.

(i) It supports a multiclass prioritized adaptive rate scheduling with "per-class" QoS support including guaranteed rate, delay, and BER. To maintain QoS guarantees the proposed scheme adapts to changing traffic conditions by employing predictive adaptation based on estimation of future packet arrival rates as well as feedback-enhanced reactive adaptation control.

(ii) It exploits feedback-enhanced reactive control in order to maintain delay targets (target tracking) and to counteract arrival rate estimation errors. When predictive adaptation fails to maintain delay targets (due to arrival rate estimation errors or high congestions) feedback information is utilized to correct rate allocations. Feedback control ensures that deviations from delay targets are minimized by efficient allocation of resources during failure condition.

(iii) It decouples delay and bandwidth (i.e., rate) parameters. Maintaining delay targets and rate allocation are accomplished through a separate control. Total scheduling delay is explicitly minimized while rate guarantees are still met.

(iv) It utilizes cross-layered design, whereby dynamic resource monitor ensures that allocated rates are feasible in the sense that BER is satisfied for all transmitting users. Interference-based loading is used to denote resource usage in a CDMA system.

4.1. FT-WFQ architecture

The unifying architecture that applies to both versions (heuristic and optimal) of feedback-enhanced target-tracking weighted fair queuing (FT-WFQ) scheduler is shown in Figure 2. The scheduler consists of *feedback-enhanced scheduling unit* (F-SU) fed and controlled by *arrival rate estimator block* (AE), *feedback control unit* (FCU) and *dynamic resource monitor* (DRM). F-SU defines an optimization problem that optimally allocates transmission rates every scheduling time interval. The optimization problem within F-SU is shaped by the information provided by AE, FCU, and DRM, and its objective is to minimize delay cost function as defined in the next subsections. AE block provides estimated arrival rates for the following time interval, while FCU monitors average delay incurred by each class, and adjusts optimization problem within F-SU if delays exceed pre-defined targets (i.e., it provides a corrective feedback). The feedback adjustment (as well as optimization problem within F-SU) is *heuristic* or *optimal* depending on the version of scheduler and as elaborated in the following subsections. DRM on the other hand dynamically recalculates total resources (i.e., CDMA capacity) available and checks if scheduling assignment is feasible by adding (cross-layer) resource constraint in the optimization problem.

4.2. Heuristic-based scheme

Let $\hat{\lambda}_i[n]$ be the estimated arrival rate of class i ($i = 1, 2, \ldots, N$) for the nth scheduling time interval measured in packets per frame (note that the actual estimation method is not considered in this paper). It is provided by the arrival rate estimator block (AE) (Figure 2). Also, let $Q_i[n]$ be the queue size (in packets) of class i at the *beginning* of the nth (scheduling) time interval. Note that $Q_i[n]$ is known to the BS scheduler as it represents the current packet backlog. Considering the nth time interval in isolation, the scheduling delay (in frames) of class i packet-queue is given by

$$D_i[n] = \frac{Q_i[n] + \hat{\lambda}_i[n] \cdot T}{R_i[n]}, \tag{12}$$

where $R_i[n]$ is the allocated rate (in packets/frame) to class i packet-queue in the nth time interval and T is the scheduling time interval duration measured in frames ($T = 1$ if scheduling is done on a frame-by-frame basis). The objective of the (heuristic) F-SU in the nth scheduling time interval is to allocate rates $R_i[n]$ ($i = 1, 2, \ldots, N$) such as to minimize overall delay cost function $\sum_{i=1}^{N} D_i[n]$, while keeping mean delay of all classes as close as possible to their respective delay targets. Note, however, that the delay cost function defined above is highly dependent on the estimated arrival rates $\hat{\lambda}_i[n]$. Even slight estimation errors by AE block could degrade performance, and lead to erroneous rate assignments with inefficient resource distributions.

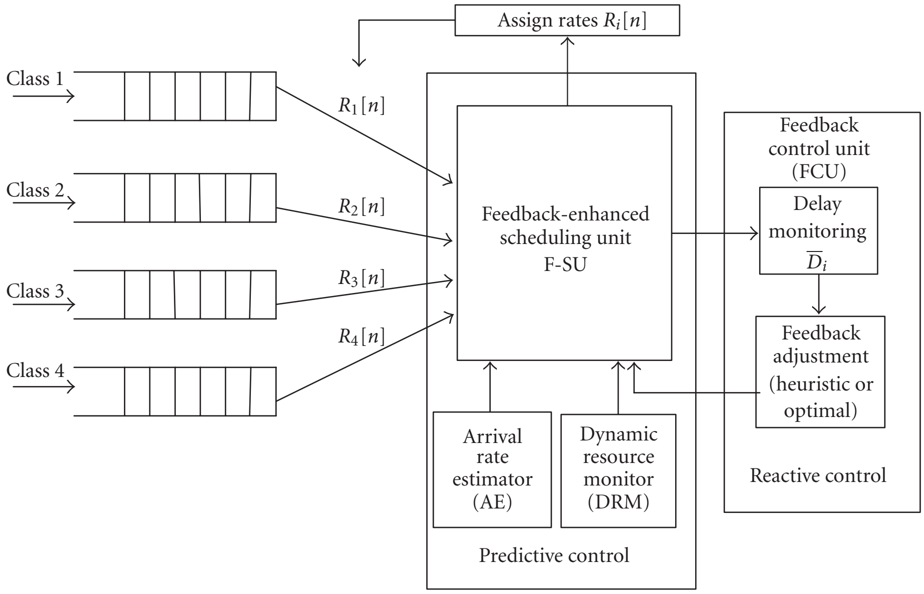

Figure 2: Architecture of FT-WFQ scheduler (with four traffic classes).

In order to mitigate for the estimation error, as well as to meet mean delay objectives as efficiently as possible, the following *heuristic-based* feedback control unit (FCU) that initiated adjustment of the optimization problem in F-SU is proposed. Let $T_{d,i}$ denote mean delay target for packets of class i (measured in frames). It is an operator specific value based on the level of QoS guarantee provided. The FCU monitors *mean* packet scheduling delays of each class. Let the running average of monitored packet delay of class i at the time interval n be denoted as $\overline{D}_i[n]$. Starting from the highest priority class (class 1) with descending priority, FCU finds class i (if any) whose delay $\overline{D}_i[n]$ is above targeted threshold $T_{d,i}$ (i.e., $\overline{D}_i[n] > T_{d,i}$). This signals that the estimation error occurred (with high probability) and that class i was degraded due to wrong assignments. FCU then "preempts" all classes $j \neq i$ whose mean delay $\overline{D}_j[n]$ is below corresponding targeted threshold (i.e., all classes j for which $\overline{D}_j[n] < T_{d,j}$). A "preempted" class is constrained to minimum guaranteed rate and it is prevented from sharing excess resources (in that time interval). "Preemption" is conducted by sending feedback information that changes corresponding constraints in optimization problem within (heuristic) F-SU in the nth interval. This ensures that class j receives only minimum guaranteed service rate until delay of class i has stabilized. The pseudocode of FCU-initiated heuristic adjustment is shown in Figure 3.

Let the set of preempted classes (in the nth time interval) be denoted by $\mathcal{P}$. Let $\eta_T[n]$ denote the total capacity available as evaluated by dynamic resource monitor (DRM), and let constant p_i indicate different priorities in the system, such that if class i has higher priority than class j, then $p_i > p_j$. Then, the optimization problem of (heuristic) F-SU in the interval n is formulated (for clarity of presentation index n is dropped) as follows.

Find the optimal rate allocations R_i^*, $i = 1, 2, \ldots, N$, so as to

$$\text{minimize} \sum_{i=1}^{N} \frac{p_i \cdot (Q_i + \hat{\lambda}_i \cdot T)}{R_i} \tag{13}$$

subject to

$$R_i \geq \min\left(R_{i,\min}, (Q_i + \hat{\lambda}_i \cdot T)/T_{d,i}\right) \quad \forall i \notin \mathcal{P}, \tag{14a}$$

$$R_i = \min\left(R_{i,\min}, (Q_i + \hat{\lambda}_i \cdot T)/T_{d,i}\right) \quad \forall i \in \mathcal{P}, \tag{14b}$$

$$\sum_{i=1}^{N} \frac{1}{1 + (W/R_i)/\gamma_i} \leq \eta_T, \tag{14c}$$

$$R_i \geq 0 \quad i = 1, 2, \ldots, N. \tag{14d}$$

The term $(Q_i + \hat{\lambda}_i) \cdot T/T_{d,i}$, appearing in constraints of (14a), (14b), represents the rate needed to keep class i delay below its delay target $T_{d,i}$. However, in order to make the solution feasible in the case of unpredicted bursts, each class is only guaranteed service rate $R_{i,\min}$, which is the minimum rate for class i guaranteed by admission control (see $\min(\cdot)$ term in (14a) and (14b)). Note that if class i is preempted by heuristic FCU the inequality constraint in (14a) is changed to the corresponding *equality* constraint in (14b). The constraint in (14c) is due to DRM. It ensures that the rate allocation is feasible in the sense that BER is satisfied for all transmitting users. DRM constraint in (14c) follows from Section 3 with class i rate given as $R_i = N_i \cdot r_i$ and with the maximum loading η_T given by (11).

```
Heuristic FCU in interval n:
(1) i = 1
(2) if  (D_i[n] > T_di) {
(3)     Preempt Classes j for which D_j[n] < T_dj
(4)     → Set R_j[n] = R_j,min
(5) DONE
(6)   }
(7) else {
(8)       i = i + 1
(9)       GO TO 2
(10)  }
```

FIGURE 3: Pseudocode of FCU-initiated heuristic.

4.3. Optimal scheme

The objective of the optimal scheduling scheme (i.e., F-SU) is to minimize the overall delay and in the case of arrival rate estimation errors or high loading congestions to minimize mean delay deviations from the corresponding targeted objectives. It is "optimal" in the sense that it explicitly minimizes delay deviations from targeted objectives and as such allocates resources as efficiently as possible. It is, however, not overall optimal as it only considers single time interval in isolation, whereas the overall optimal scheme would consider a larger time horizon.

The optimization problem is defined as follows. Let the *indicator function* $I_D[n]$ in the nth time interval be defined as

$$I_D[n] = \begin{cases} 0, & \text{if } \overline{D}_i[n] \leq T_{d,i}\ \forall i = 1, 2, \ldots, N, \\ 1, & \text{otherwise}, \end{cases} \tag{15}$$

where as in the last subsection $\overline{D}_i[n]$ denotes the mean (FCU) monitored scheduling delay of class i at the time interval n, and $T_{d,i}$ is the mean delay target for packets of class i as measured in frames. Therefore, the binary indicator function $I_D[n]$ is set to 1 if mean delay of *any* class exceeds its delay target $T_{d,i}$. This signals that resources were assigned erroneously either due to arrival rate estimation errors or due to very high congestion. The indicator function is set by the (optimal) feedback control unit (FCU) (recall that FCU explicitly monitors *mean* packet scheduling delays $\overline{D}_i[n]$ of each class i). Using the same terminology as in the last subsection, the optimization problem of the optimal F-SU in the interval n is formulated (for clarity of presentation index n is dropped) as follows.

Find the optimal rate allocations R_i^*, $i = 1, 2, \ldots, N$, so as to

$$\text{minimize } (1 - I_D) \cdot \sum_{i=1}^{N} \frac{p_i \cdot (Q_i + \widehat{\lambda}_i \cdot T)}{R_i} + I_D \cdot \sum_{i=1}^{N} p_i \cdot \left(\frac{(Q_i + \widehat{\lambda}_i \cdot T)}{R_i} - T_{d,i} \right)^2 \tag{16}$$

subject to

$$R_i \geq \min\left(R_{i,\min}, (Q_i + \lambda_i) \cdot T/T_{d,i}\right) \quad \forall i, \tag{17a}$$

$$\sum_{i=1}^{N} \frac{1}{1 + (W/R_i)/\gamma_i} \leq \eta_T, \tag{17b}$$

$$R_i \geq 0 \quad i = 1, 2, \ldots, N. \tag{17c}$$

Note that the proposed optimization problem will minimize total deviations from delay targets if FCU detects that mean delay of any class exceeds corresponding delay target (i.e., if the indicator function $I_D[n]$ is set to 1), otherwise it will minimize the total delay (i.e., if the indicator function $I_D[n]$ is set to 0). The reasoning behind this is that if the mean delay of all classes is below their respective delay targets, then the objective is to minimize the overall delay, whereas if delay of any class is above its corresponding delay target, the resources should be redistributed so as to keep delay of all classes as close to their delay targets as possible. In other words if there is any class whose mean delay is above its corresponding delay target, there should be no classes whose mean delay is below theirs.

The constraints in (17a) and (17b) are analogous to the corresponding constraints in a heuristic-based problem of the last subsection with constraint in (17b) due to DRM.

5. PERFORMANCE ANALYSIS AND SIMULATION MODELS

In this section, the analysis and simulation models are developed for the proposed FT-WFQ scheduler in nonstationary traffic environment. Four traffic classes (i.e., $N = 4$) defined in UMTS network are considered (see Table 1). Performance measures are mean delay and service rate assigned to each class.

5.1. Delay analysis with nonstationary packet arrival rate and estimation error

Assume that packets of class i ($i = 1, 2, 3, 4$) arrive according to a *nonstationary* Poisson arrival process with mean arrival rate of $\lambda_i(t)$ (packets/frame). Nonstationary Poisson arrival process is characterized by time-varying mean arrival rate $\lambda_i(t)$ modeled as follows. Time is divided into equal length (scheduling) time intervals of duration T frames. In the nth interval ($n = 1, 2, 3, \ldots$) mean arrival rate $\lambda_i(n \cdot t)$ (denoted as $\lambda_i[n]$) takes a random value according to a uniform distribution. It retains this value for the duration of interval n. Without any loss of generality, the four aforementioned nonstationary Poisson arrival processes are assumed to be independent.

Let $\lambda_i[n]$ as defined above be the *actual* mean arrival rate of class i arrival process for the nth time interval. Let $\widehat{\lambda}_i[n]$ be the *estimated* arrival rate that is observed at BS and used by

Table 1: Numerical values of QoS parameters for each class.

Traffic class i	Traffic type (UMTS QoS class)	Delay tolerance (frames) or packet timeout value T_{di}	QoS requirements	
			Minimum rate $R_{i,\min}$	BER requirement
1	Conversational	< 1	12 kbps	10^{-3}
2	Streaming	1–2	128 kps	10^{-4}
3	Interactive	2–4	32 kbps	10^{-5}
4	Background	> 8	0	10^{-7}

the scheduling algorithm (in the nth interval). As discussed previously the estimator is not perfect, and consequently it is assumed that an additive white Gaussian error ε_n is introduced in each interval n, that is,

$$\hat{\lambda}_i[n] = \lambda_i[n] + \varepsilon_n. \tag{18}$$

As noted above, ε_n is a white Gaussian random process with mean ε, and variance $0.1 \cdot \varepsilon$, for all n. Also, $E[\varepsilon_n \cdot \varepsilon_k] = 0$ for all $n \neq k$ ($E[\cdot]$ is the expectation operator).

In the nth time interval, class i ($i = 1, 2, 3, 4$) packet-queue receives service rate $R_i^*[n]$ (packets/frame) in accordance with the solution of the optimization problem defined in (13) for the heuristic-based scheduler or (16) for the optimal scheduler. Hence, each class i queue can be considered in isolation with time-varying arrival rate $\hat{\lambda}_i[n]$ and time-varying service rate $R_i^*[n]$. Such a queue can be represented by an $M[n]/D[n]/1$ system, where $M[n]$ represents the nonstationary packet arrival process as defined above and $D[n]$ stands for *deterministic* server operating at optimal rates of $R_i^*[n]$. Because of its time-varying nature, it is very difficult to analyze $M[n]/D[n]/1$ system directly (i.e., to solve Kolmogorov forward equations). However, various approximations have been proposed in the literature. One very simple approximation is called *point-wise stationary approximation* (PSA) also known as quasistationary approximation [14, 15]. According to PSA, in each time interval n, $M[n]/D[n]/1$ system can be approximated by a stationary $M/D/1$ model where the current value of $\hat{\lambda}_i[n]$ is used as "stationary" arrival rate and the current value of $R_i^*[n]$ is used as deterministic service rate in that particular interval. For PSA approximation to be valid, duration of the time interval (T) should be 4–5 times greater than the packet service time, so that the system can asymptotically reach a steady-state. Consequently, in analytical approximation T frames (for some large enough T) constitute one time interval n.

Assuming $M/D/1$ model in each time-interval n ($n = 1, 2, \ldots$), *instantaneous* PSA delay for class i (denoted as $D_i[n]$) is given by Pollaczek-Khinchin delay formula [16]:

$$D_i[n] = \frac{\hat{\lambda}_i[n]/\left(R_i^*[n]\right)^2}{2\left(1 - \hat{\lambda}_i[n]/R_i^*[n]\right)} + \frac{1}{R_i^*[n]}. \tag{19}$$

Then, PSA *running average* delay of class i used by feedback control unit (FCU) is defined as

$$\overline{D}_i[n] = \frac{\overline{D}_i[n-1] \cdot (n-1) + D_i[n]}{n}. \tag{20}$$

In the accordance with the proposed scheduler, FCU monitors PSA running average delay of each class i (20) and adjusts optimization problem in the nth time interval accordingly, as explained in Section 4. Hence, the optimization problem is solved in each time interval n as given in (13) for the heuristic-based or (16) for the optimal scheduler. MATLAB (optimization toolbox) was utilized to solve the actual optimization problem in the nth time interval.

5.2. Simulation model with nonstationary packet arrival rate and estimation error

Proposed scheduling scheme was simulated in a nonstationary environment using an event-driven simulation tool OPNET [17]. The model consists of four traffic generators (one for each class), and the base station (BS) where the scheduling algorithm is implemented. As in the analysis section, time is divided into equal-duration intervals n of length T frames. Traffic generators generate traffic according to four independent nonstationary Poisson processes as in the last subsection. As in the analysis, it is assumed that the additive white Gaussian estimation error ε_n is present when estimating the actual arrival rate. The mean ε of the estimation error was used as a simulation parameter.

In order to solve the optimization problem in (13) or (16) using optimization toolbox provided by MATLAB, a co-simulation interface model of OPNET and MATLAB was developed. The "mx" interface provided by MATLAB was used, as explained in detail in [18]. (This is very useful if one needs to use MATLAB algorithms when simulating complex communications systems with discrete event simulator.) The running average delay statistic was collected for each class during simulation run-time. In accordance with the proposed scheduling scheme, this information was used by feedback control unit (FCU) to adjust optimization problem in each time interval.

6. ANALYSIS AND SIMULATION NUMERICAL RESULTS

The numerical parameters used in the analysis as well as in simulations are summarized in Table 2. The proposed heuristic-based and optimal FT-WFQ scheduling schemes are evaluated in nonstationary packet arrival environment with and without the presence of arrival rate estimation error. The proposed scheduling schemes are compared to the TVFQ scheme without reactive control as originally

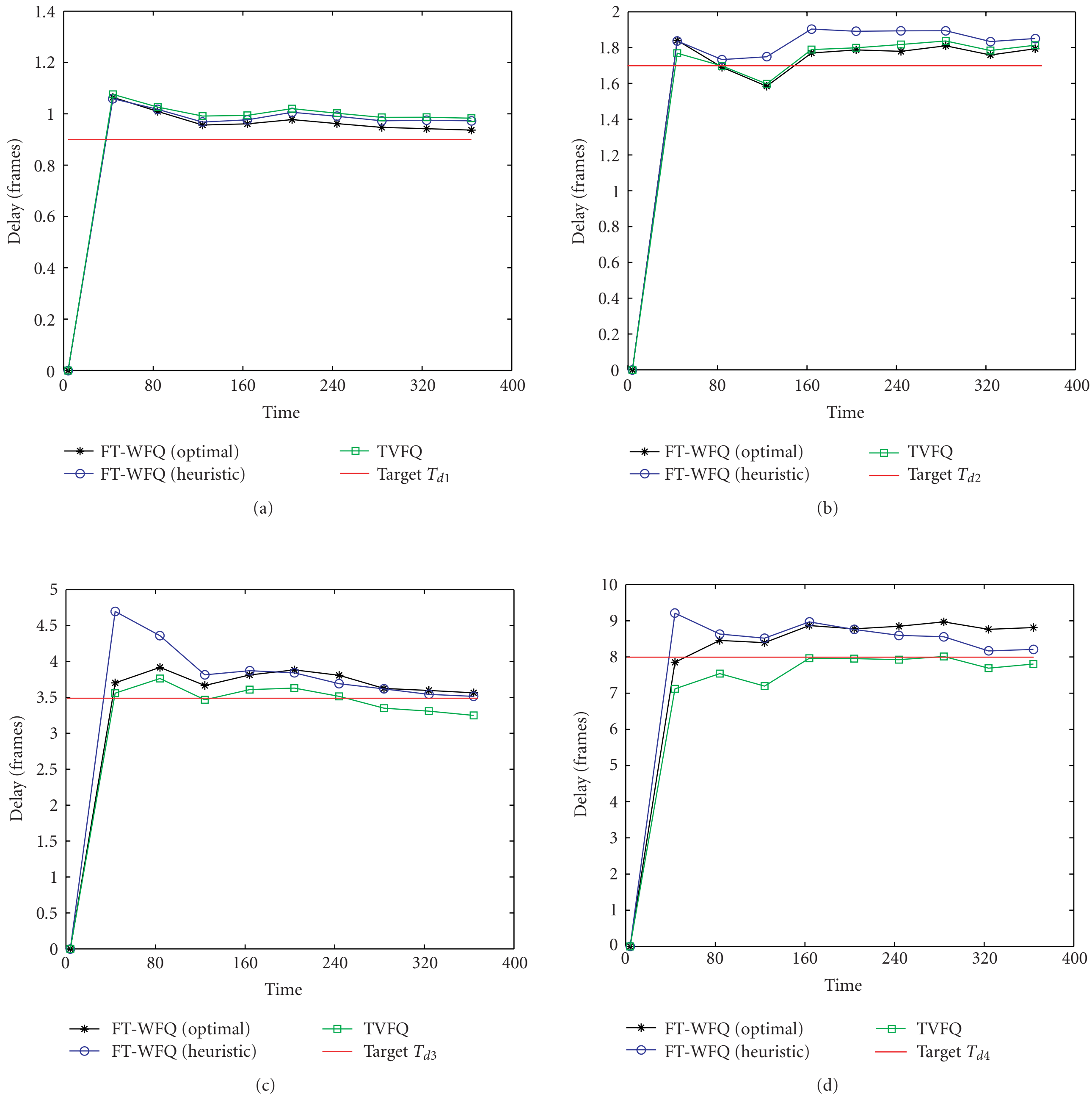

Figure 4: Analysis: delay for classes 1–4 (no estimation error).

proposed in [12]. Consistent with the last section, four traffic classes are considered.

6.1. No arrival rate estimation error

Due to nonstationary traffic conditions, even under perfect traffic estimations, the selection of priority weights p_i needed to maintain delay targets becomes a difficult task for the TVFQ scheduler. In this subsection, the performance of TVFQ scheme [12] is compared to the proposed scheduling schemes under such conditions (i.e., nonstationary arrivals with estimation error $\varepsilon_n = 0$). Priority weights p_i are selected such that under average stationary (arrival) conditions delay targets are met; the numerical values are listed in Table 2. The running average of delay (measured in frames) versus simulation time (i.e., time instant) for each class are obtained for the compared schemes following analysis and simulations models presented in the last section. Note that, as mentioned before, TVFQ scheme is the one without reactive adaptation control. Delay results for the compared schemes, obtained from an analytical model, are shown in Figure 4 for classes 1–4, respectively. The results are further compared by the

Table 2: Summary of analysis and simulation parameters.

Parameter	Value
Class 1 arrival process	Poisson, uniform range: $0 \le \lambda_1 \le 0.4 \cdot \eta_T$
Class 2 arrival process	Poisson, uniform range: $0 \le \lambda_2 \le 0.3 \cdot \eta_T$
Class 3 arrival process	Poisson, uniform range: $0 \le \lambda_3 \le 0.38 \cdot \eta_T$
Class 4 arrival process	Poisson, uniform range: $0 \le \lambda_4 \le 0.25 \cdot \eta_T$
Delay targets T_{di} (in frames)	$T_{d1} = 0.9$ $T_{d2} = 1.7$ $T_{d3} = 3.5$ $T_{d4} = 8$
Interval size T (in frames)	20
Maximum power $S_{i,\max}$	100 (mW)
Mean estimation error ε	-0.5 (also varied)
Priority weight p_i	$p_1 = 6$ $p_2 = 4$ $p_3 = 2$ $p_4 = 1$

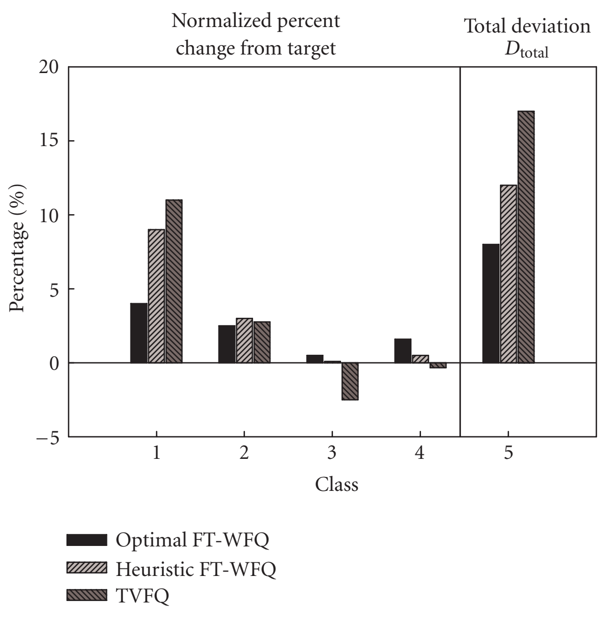

Figure 5: Analysis: delay percent change from target (no estimation error).

bar charts shown on the left part of Figure 5 that show normalized mean delay deviation of each class from the respective delay target (normalization is with respect to priority weights, that is, for class i shown is $p_i/p_1 \cdot$ *actual deviation*). As evident from Figure 4 and the bar chart on the left part of Figure 5, TVFQ scheme performs the worst resource allocations among the compared schemes as it does not implement any target-tracking constraints. It forces high positive delay deviations from targets for classes 1 and 2, respectively, even when large negative delay deviations for classes 3 and 4 are present (see left part of Figure 5). Thus it wastes resources by over-feeding classes 3 and 4 during their light packet arrivals, when these excess resources could have been allocated to classes 1 and 2, respectively. From Figure 4 and bar chart in Figure 5, it can be seen that the proposed FT-WFQ schemes (heuristic and optimal) achieve far better resource distributions and that the total delay deviations from the targets are minimized. By utilizing feedback-enhanced reactive control designed to explicitly minimize delay deviations from the corresponding delay targets, the heuristic-based and optimal FT-WFQ schemes slightly increase the mean delay of classes 3 and 4, respectively by reducing resources (i.e., rates) allocated to them, but nevertheless keeps them close to their respective targets. As evident from Figure 4 and the bar chart on the left of Figure 5, this in turn provides more resources to accommodate heavy traffic arrival from classes 1 and 2, respectively, thereby reducing their mean delay deviations from the targets. It can also be seen that the optimal FT-WFQ scheme achieves better resource allocations than heuristic-based scheme as its objective is to explicitly minimize delay deviations.

Total performance gain/loss is quantified as follows. From the bar charts on the left in Figure 5, the *total deviation* from targets D_{total} is defined and calculated as $D_{\text{total}} = \sum |DV_i|$ where DV_i is the normalized deviation of class i ($i = 1, 2, 3$, and 4). Hence, evaluating from the left part of Figure 5 for TVFQ: $D_{\text{total}} = 0.11(11\%) + 0.0275(2.75\%) + 0.025(2.5\%) + 0.0025(0.25\%) = 0.165(16.5\%)$. Similarly, the total deviations of heuristic and optimal FT-WFQ schemes can be obtained as 12.6% and 8.5%, respectively. The total deviation D_{total} bar chart is shown on the right part of Figure 5.

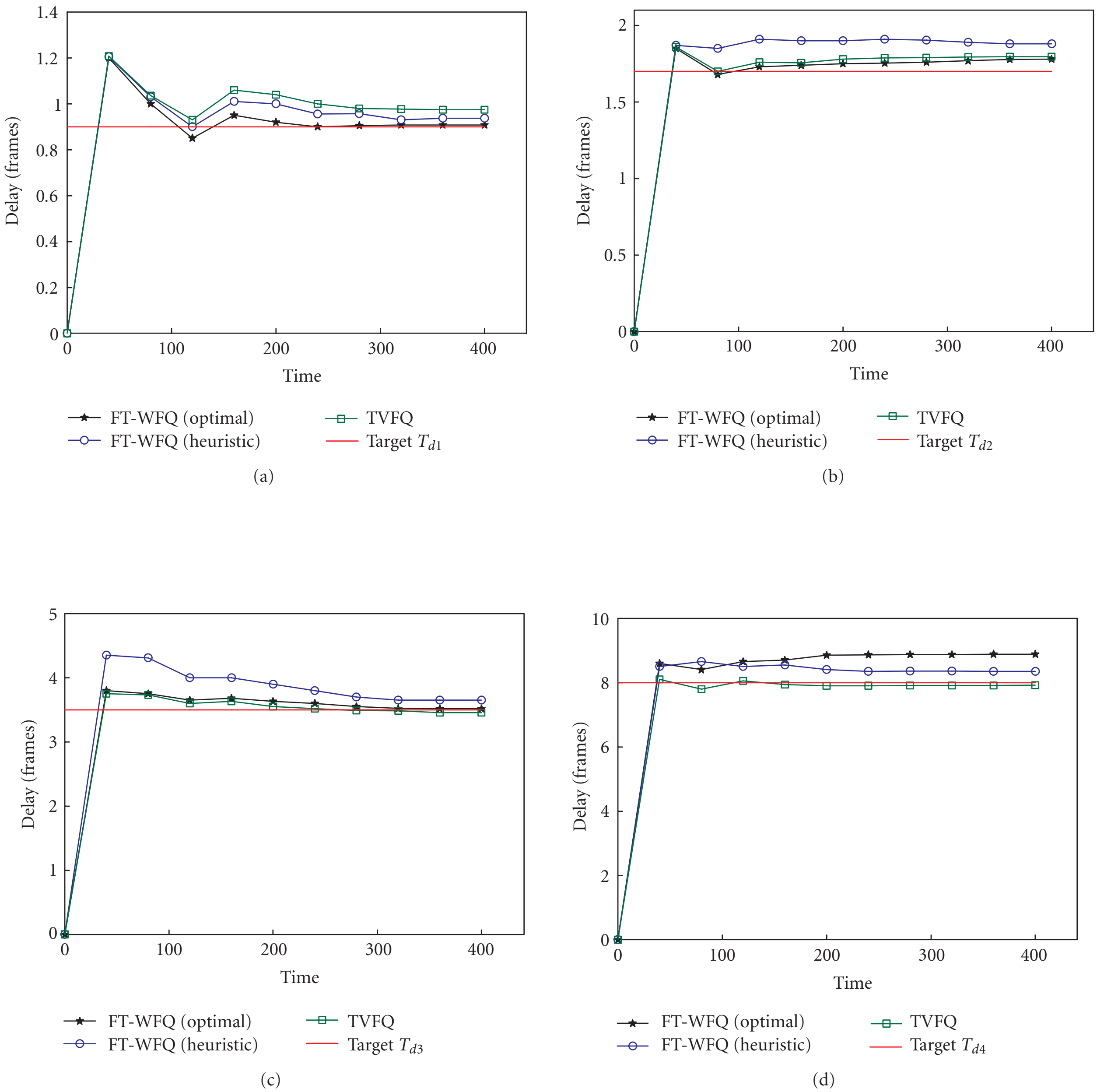

FIGURE 6: Delay: for classes 1–4 (no estimation error).

The corresponding OPNET simulations results, namely, running averages of delays and delay deviations are shown in Figures 6 and 7, respectively. The results are in close agreement with the corresponding analytical results (from Figures 4 and 5) with differences in instantaneous values mainly due to highly random nature of nonstationary arrival process as defined in the previous section. As the analytical results, simulations from Figures 6 and 7 demonstrate that the proposed schemes are able to minimize deviations from corresponding delay targets and allocate resources efficiently. The total deviations D_{total} observed form Figure 7 for TVFQ, heuristic, and optimal FT-WFQ are 16.5%, 11.7%, and 8.9%, respectively. It is also worth mentioning that analysis and simulation results in Figures 4 and 6, respectively, show that the system stabilizes relatively quickly.

6.2. *Results under arrival rate estimation error*

PSA analysis and OPNET simulations are employed to compare scheduling schemes (TVFQ, FT-WFQ optimal and heuristic) in nonstationary environment with imperfect arrival rate estimation. Consequently, class 1 arrival rate estimation

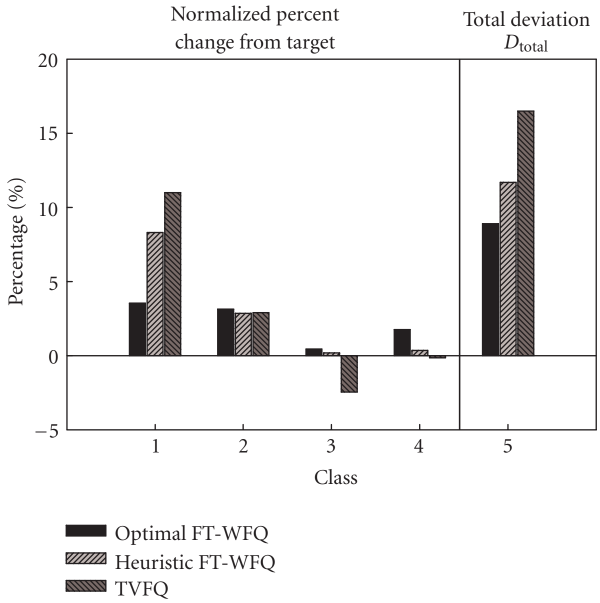

FIGURE 7: Simulation: delay percent change from target (no estimation error).

error is assumed in this subsection with mean error ε of -0.5. Subsequently, the study was repeated using different error-mean values.

The analytical delay running averages (measured in frames) versus time for classes 1–4, respectively, are shown in Figure 8 (for $\varepsilon = -0.5$). The corresponding bar charts showing (normalized) mean delay deviations from delay targets for each class as well as the total deviations D_{total} are shown on the left and right parts, respectively, of Figure 9. As expected and evident from Figures 8 and 9, the regular TVFQ is the most affected by an estimation error as it is not able to correctly adapt resource (i.e., rate) allocations. From Figure 9 it can be seen that class 1 is affected the most as TVFQ degrades its mean delay to 15% above corresponding delay target. Note that resources not allocated to class 1 are utilized by lower priority classes as evident by unintended decrease of their mean delay (see Figure 9). Thus, TVFQ wastes resources by allowing positive (normalized) deviations from the targets (classes 1 and 2) even when negative (normalized) deviations are present (classes 3 and 4). By employing feedback-enhanced reactive adaptation control, proposed FT-WFQ schemes achieve much better resource distributions thereby keeping mean (normalized) delays closer to corresponding targets (see Figures 8 and 9, respectively) and are thus able to substantially mitigate for the arrival rate error effect. From Figure 9, mean (normalized) delay deviations of class 1 for the heuristic-based and optimal FT-WFQs are 14% and 7.5%, respectively. Calculating similarly as in the last subsection from Figure 9, the total deviation from targets D_{total} for TVFQ, heuristic, and optimal FT-WFQs are 29.5%, 24.5%, and 19.5%, respectively (shown on the right part of Figure 9). Thus, the heuristic FT-WFQ achieves a reduction in mean deviation as compared to regular TVFQ, while the optimal FT-WFQ as expected distributes resources the most efficiently and achieves the largest reduction in the total deviation. It is also worth mentioning that when other mean-error values ε are used instead, similar results are observed. These results are verified by OPNET simulation model, but are not shown due to space limitations. It is worth mentioning that simulation and analytical results are in strong agreement.

To compare results under different error conditions, the above analysis (and simulation) is repeated for other mean values of error ε. Figure 10 shows the total deviation (D_{total}) versus mean-estimation error (ε) observed for the compared schemes using PSA analysis. Shown are results for both underestimation ($\varepsilon < 0$) and overestimation ($\varepsilon > 0$) cases. Observe that the optimal FT-WFQ achieves the lowest total deviations for all mean-estimation errors (ε), and as the magnitude of error increases so does the gain of the optimal FT-WFQ compared to the other schemes. Almost identical results are obtained by OPNET simulation (not shown).

7. CONCLUSIONS

In this work, two versions of feedback-enhanced target-tracking weighted fair queuing (FT-WFQ) schedulers, namely, heuristic and optimal, were presented. FT-WFQ is adaptive rate scheduler able to provide QoS guarantees on a "per-class" basis. It dynamically adapts allocation rates based on the predictive and feedback-enhanced reactive adaptation

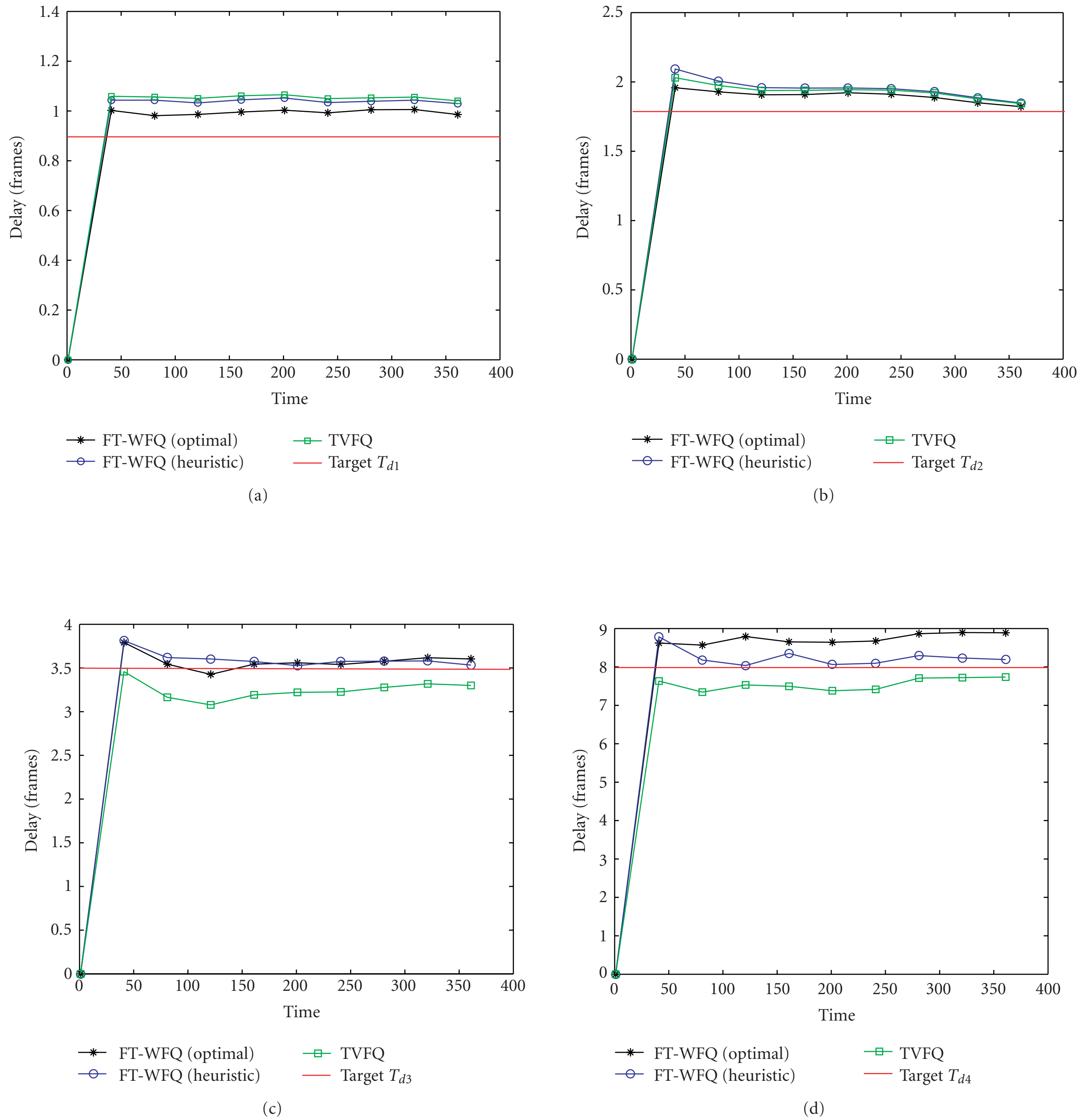

Figure 8: Analysis: delay for classes 1–4 (estimation error).

control. Predictive control adjusts rate allocations based on the anticipated network conditions reflected by estimated arrival rates. When the predictive adaptation control fails to maintain delay targets (due to estimation errors or extremely high loading), the corrective feedback-enhanced reactive control ensures that deviations from delay targets are minimized.

Performance of the proposed scheduler was evaluated analytically and by simulation in a nonstationary environment with and without arrival rate estimation error. PSA approximation was used for analytical delay modeling. It was shown that the proposed schemes are able to substantially reduce the error effect and minimize deviations from delay targets.

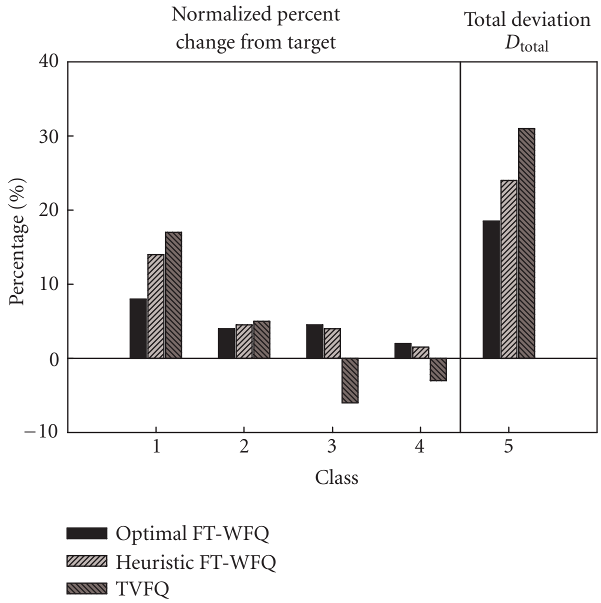

FIGURE 9: Analysis: delay percent change from target (estimation error).

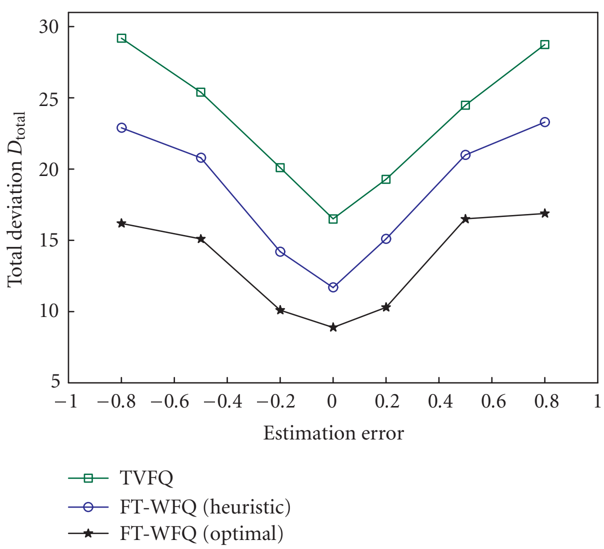

FIGURE 10: Analysis: total deviation versus estimation error.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (NSF) under Contract SCI-0129527. The authors are thankful to the editors of this special issue and to the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] H. Holma and A. Toskala, *WCDMA for UMTS*, John Wiley & Sons, West Sussex, UK, 2000.

[2] O. Yu, E. Saric, and A. Li, "Fairly adjusted multimode dynamic guard bandwidth admission control over CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 579–592, 2006.

[3] L. Xu, X. Shen, and J. W. Mark, "Dynamic bandwidth allocation with fair scheduling for WCDMA systems," *IEEE Wireless Communications*, vol. 9, no. 2, pp. 26–32, 2002.

[4] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 60–73, 2004.

[5] X. Wang, "An FDD wideband CDMA MAC protocol with minimum-power allocation and GPS-scheduling for wireless wide area multimedia networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 1, pp. 16–28, 2005.

[6] C. Li and S. Papavassiliou, "Fair channel-adaptive rate scheduling in wireless networks with multirate multimedia services," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1604–1614, 2003.

[7] M. Moustafa, I. Habib, and M. N. Naghshineh, "Efficient radio resource control in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 2385–2395, 2004.

[8] I. F. Akyildiz, D. A. Levine, and I. Joe, "A slotted CDMA protocol with BER scheduling for wireless multimedia networks," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 146–158, 1999.

[9] X. Wang, "A new scheduling scheme for the wideband TD-CDMA MAC protocol," in *Proceedings of IEEE International Conference on Communications (ICC '01)*, vol. 3, pp. 974–978, Helsinki, Finland, June 2001.

[10] W. K. Wong, H. Zhu, and V. C.M. Leung, "Soft QoS provisioning using the token bank fair queuing scheduling algorithm," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 8–16, 2003.

[11] W. K. Wong, H. Y. Tang, and V. C. M. Leung, "Token bank fair queuing: a new scheduling algorithm for wireless multimedia services," *International Journal of Communication Systems*, vol. 17, no. 6, pp. 591–614, 2004.

[12] A. Stamoulis, N. D. Sidiropoulos, and G. B. Giannakis, "Time-varying fair queueing scheduling for multicode CDMA based on dynamic programming," *IEEE Transactions on Wireless Communications*, vol. 3, no. 2, pp. 512–523, 2004.

[13] C.-C. Li, S.-L. Tsao, M. C. Chen, Y. Sun, and Y.-M. Huang, "Proportional delay differentiation service based on weighted fair queuing," in *Proceedings of the 9th International Conference on Computer Communications and Networks (ICCCN '00)*, pp. 418–423, Las Vegas, Nev, USA, October 2000.

[14] O. B. Jennings and W. A. Massey, "A modified offered load approximation for nonstationary circuit switched networks," *Telecommunication Systems*, vol. 7, no. 1–3, pp. 229–251, 1997.

[15] L. Green and P. Kolesar, "The pointwise stationary approximation for queues with nonstationary arrivals," *Management Science*, vol. 37, no. 1, pp. 84–97, 1991.

[16] T. Robertazzi, *Computer Networks and Systems: Queueing Theory and Performance Evaluation*, Springer, New York, NY, USA, 2000.

[17] I. Katzela, *Modeling and Simulating Communication Networks: A Hands-on Approach Using OPNET*, Prentice Hall, Upper Saddle River, NJ, USA, 1998.

[18] V. Dham, L. A. DaSilva, A. Annamalai, and V. Srivastava, "A phased array antenna model for space division multiple access in wireless networks," in *Proceedings of OPNETWORK Conference*, Washington, DC, USA, August 2002.

Oliver Yu received the B.A.S. (1981), M.A.S. (1991), and Ph.D. (1997) degrees in electrical and computer engineering from the University of British Columbia in Canada. He had extensive industry research and development experiences. He had held technical leadership and engineering management positions in Microtel, Nortel, Hughes, and Motorola. He is currently a Research Associate Professor at the Department of Electrical and Computer Engineering in the University of Illinois at Chicago. His research interests are in the areas of fourth-generation wireless systems, all-optical photonic networks, next generation Internet, and computing grid systems. He has been working on research projects that are funded by the National Science Foundation (NSF) and the Department of Energy (DOE). He received the Career Award from DOE in 2003.

Emir Saric received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Chicago (UIC) in 2001, and 2004, respectively. He is currently working towards the Ph.D. degree in electrical and computer engineering at UIC. Since 2004, he has been a Research Assistant for Professor O. Yu at the Networking and Wireless Communications Laboratory, UIC. His research interests include QoS provisioning (MAC and admission control) and performance evaluation of future cellular wireless networks and cognitive wireless networks.

Anfei Li received the B.S. degree from University of Science and Technology of China, Hefei, 2002. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, the University of Illinois at Chicago. His research interests include medium access control (MAC) in 3G and 4G wireless communication systems, wireless ad hoc and sensor networks, all-optical networks, and computing grids.