

Opportunistic Scheduling for QoS Guarantees in 3G Wireless Links

Jeong Geun Kim and Een-Kee Hong

Department of Electronics and Information,
Kyung Hee University, Suwon, Republic of Korea
{jg_kim,ekhong}@khu.ac.kr

Abstract. Proportional Fair (PF) share policy has been adopted as a downlink scheduling scheme in CDMA2000 1xEV-DO standard. Although it offers optimal performance in aggregate throughput conditioned on equal time share among users, it cannot provide a bandwidth guarantee and a strict delay bound, which is essential requirements of real-time (RT) applications. In this study, we propose a new scheduling policy that provides quality-of-service (QoS) guarantees to a variety of traffic types demanding diverse service requirements. In our policy data traffic is categorized into three classes, depending on sensitivity of its performance to delay or throughput. And the primary components of our policy, namely, Proportional Fair (PF), Weighted Fair Queuing (WFQ), and delay-based prioritized scheme are intelligently combined to satisfy QoS requirements of each traffic type. In our policy all the traffic categories run on the PF policy as a basis. However the level of emphasis on each of those ingredient policies is changed in an adaptive manner by taking into account the channel conditions and QoS requirements. Such flexibility of our proposed policy leads to offering QoS guarantees effectively and, at the same time, maximizing the throughput. Simulations are used to verify the performance of the proposed scheduling policy. Experimental results show that our proposal can provide guaranteed throughput and maximum delay bound more efficiently compared to other policies.

1 Introduction

With the explosive growth of the Internet and rapid proliferation of personal communication services, the demand for data services in wireless networks has been ever-increasing. Responding to this demand, wireless networks have been evolving toward packet-switched architectures that are more flexible and efficient in providing packet data services. The CDMA2000 1xEV-DO standard (abbreviated hereafter as 1xEV-DO) is one of such architectures, that is designed to support data services in the third generation wireless network [1]. Compared to previous wireless networks often characterized by circuit-switched and voice-oriented architecture, the 1xEV-DO system has several unique features. Notably among them is the “opportunistic” scheduling used to schedule downlink (or equivalently forward link) transmission.

The basic idea behind the opportunistic scheduling is that temporal channel variation of multiple users is taken into account in scheduling [2]. Thus at a given time the scheduling decision favors a mobile user currently seeing a better channel. If all the traffic is “elastic” (i.e., having flexible service requirements), such opportunistic scheduling mechanisms then greatly improve throughput performance and yield higher bandwidth utilization.

The 1xEV-DO standard adopted an opportunistic scheduling called Proportional Fair (PF) share policy as a downlink scheduler. Although the PF policy offers optimal performance in aggregate throughput conditioned on equal time share among users [3], it cannot provide bandwidth guarantee and strict delay bound, which is essential requirements of real-time applications. As the 1xEV-DO system expands its service realm from non-real-time (NRT) to real-time (RT) applications like video streaming, such limitation triggers a lot of research efforts toward expanding its capability for supporting various quality-of-service (QoS) requirements of RT traffic.

The authors in [4] proposed a scheduling discipline called the Exponential Rule in which the queue with larger weighted delay of head-of-line (HOL) packet gets higher priority in transmission. This rule behaves like the PF policy when the weighted delay difference is not relatively large. As the difference becomes significant, this policy gracefully adapts from the PF policy to the Exponential rule. However, this policy suppresses the chance the HOL packet with relative lower weighted delay gets selected for transmission even when it currently sees a better channel. This may lead to lower throughput performance. In [5], Liu et al. proposed an utility-based scheduling algorithm in which utility is a decreasing function of packet delay. By scheduling the transmission at each time in a way the total utility rate is maximized, this scheme can provide a strict delay bound. However, it is questionable to how the utility function can be devised for throughput-sensitive applications. Furthermore existence of the utility functions and their derivatives is not obvious for applications with diverse requirements in reality. To address the issues of fairness and throughput guarantee in the PF policy, a dynamic rate control algorithm was proposed in [6]. In this scheme, target throughput is set along with the associated minimum and maximum bounds, and the scheduler attempts to maintain user’s perceived throughput within those bounds. However this scheme can only work with throughput-sensitive applications, but may not suit other traffic categories.

The main contribution of this paper is that we propose a novel “opportunistic” service rule that can flexibly support a wide range of traffic categories. Most of previous works in [5]-[6] focuses on applications sensitive to either delay or throughput. In contrast, our scheme can deal with QoS needs from both traffic categories using a simple metric. In addition, throughput guarantee can be supported, depending on urgency, in either a strict or relaxed manner via configurable parameters. Such capability is useful when throughput guarantee is not necessarily required over microscopic time-scale, but rather needed on long-term basis. Moreover in order to maximize the link capacity, the RT traffic runs on the PF policy as long as delay or throughput performance does not deviated much

from its prescribed target value. However, as its QoS target is more likely to be missed, the corresponding traffic gains more weight in scheduling decision and increases the chance for transmission. This feature of graceful adaptation from the PF policy to the prioritized mode can provide maximized throughput and, at the same time, offers an accurate tool for exercising QoS. Our numerical results indicate that the proposed rule can deliver QoS guarantees in multiservice wireless environments that are often characterized by heterogeneous channel conditions and mixed QoS requirements. Moreover, throughput gains are observed against other schemes and the flexible throughput guarantees are proven to work effectively as designed and give a capacity improvement.

The rest of the paper is organized as follows. In Section 2 we give overview of CDMA2000 1xEV-DO standard and describe the proposed scheduling algorithm. Numerical results and simulations are reported in Section 3, followed by concluding remarks in Section 4.

2 Scheduling Algorithm

2.1 Background

To utilize temporal channel variation, the PF share policy in the 1xEV-DO requires the mobile terminals (MTs) to report continuously their channel state information to the scheduler in the access point (AP). Accordingly, the MTs report back the channel condition to the network every 1.667 msec. The pilot bursts from the AP enable the MT to accurately estimate the channel conditions. The channel state information is sent back to the AP in the form of data rate request (see Table 1) through data rate request channel (DRC) in the reverse link [1]. Once data on DRC from each MT is gathered, the PF share scheduler selects the MT i^* which satisfies

$$i^* = \arg \max_i \frac{\text{DRC}_i(t)}{\hat{R}_i(t)} \quad (1)$$

where $\text{DRC}_i(t)$ is the data rate request of MT i at time t , $\hat{R}_i(t)$ is average transmission rate of MT i by time t .

In 1xEV-DO, downlink bandwidth is shared among multiple MTs in time-division multiplexing (TDM) basis, where fixed-size time-slots are dedicated to each MT based on scheduling policy. In particular, different modulation schemes including QPSK, 8PSK, and 16QAM are employed in an adaptive manner, depending on the channel condition of the target MT. The set of available data rates for the corresponding channel conditions is listed in Table 1. To provide fairness in channel usage among the MTs, the average rate $\hat{R}_i(t)$ of the i th MT is updated every time-slot as follows:

$$\hat{R}_i(t+1) = \left(1 - \frac{1}{t_c}\right) \hat{R}_i(t) + \frac{1}{t_c} R_i(t)$$

where t_c is the time constant set to 1000 slots [2], and $R_i(t)$ is the current rate of the i th MT. At the end of current transmission the scheduler determines the next MT to transmit based on the criterion in (1) and this procedure repeats.

Table 1. Data rate options in CDMA2000 1xEV-DO.

Nominal Data Rate (Kbps)	Nominal Slots per PHY Packet	Total Bits per PHY Packet	E_c/N_t (dB) thresholds for DRC selection
38.4	16	1024	-13.5
76.8	8	1024	-10.5
153.6	4	1024	-7.4
307.2	2	1024	-4.3
614.4	1	1024	-1.0
921.6	2	3072	1.5
1228.8	1	2048	3.7
1843.2	1	3072	7.1
2457.6	1	4096	9.1

2.2 Algorithm

In a majority of previous works, prioritized services and associated QoS capabilities have been implemented by incorporating a weighting function $\mathbf{F}_w(\cdot)$ into the PF metric in (1). The modified metrics in general have the following form:

$$i^* = \arg \max_i \frac{\text{DRC}_i(t)}{\hat{R}_i(t)} \mathbf{F}_w(\cdot).$$

Different weight functions $\mathbf{F}_w(\cdot)$'s have been chosen depending on which performance factor (e.g., bandwidth, delay) is prioritized. The Exponential rule in [4] has a weight function given by

$$\mathbf{F}_w(\cdot) = \exp\left(\frac{a_i W_i(t) - \overline{aW}}{1 + \sqrt{aW}}\right) \quad (2)$$

where a_i is the weight of the i th flow, $W_i(t)$ is the waiting time of the packet, and \overline{aW} is mean weighted delay. In [6], $\mathbf{F}_w(\cdot)$ is defined as a group of functions whose values are proportional to deviation from the target rate. See [6] for details.

As pointed earlier, these rules lack the flexibility of handling diverse QoS requirements. To overcome this limitation, our rule is designed to support three primary types of traffic into which most network applications can be categorized:

- Class I: Delay-sensitive
- Class II: Throughput-sensitive
- Class III: Best-effort.

Such capability of providing differentiation in service is essential in operating an anticipating multi-service wireless networks where a wide array of traffic types coexist demanding diverse requirements. In our rule, the metric is simple to calculate and the control parameters are flexibly configured depending on the traffic category.

We propose the rule named the ‘‘Adaptive policy’’ whose metric is given by:

$$i^* = \arg \max_i \frac{\text{DRC}_i(t)}{\hat{R}_i(t)} \left(C_{wd}(t)(\overline{D}_i(t) + 1)^\alpha + 1 \right)^{w_i(t)} \quad (3)$$

In the above equation $C_{wd}(t)$ is a coefficient that determines the impact of the weight function and is set to the maximum ratio of DRC over the average rate at each scheduling epoch, that is,

$$C_{wd}(t) = \max_i \frac{\text{DRC}_i(t)}{\hat{R}_i(t)}.$$

α is a configurable parameter for prioritizing traffic sensitive to either delay or throughput, and $w_i(t)$ is the weighting factor controlling the level of emphasis on the weighting function in (3). Configuration of these parameters will be discussed later in the next section.

$\overline{D}_i(t)$ is the normalized waiting time given by

$$\overline{D}_i(t) = \frac{D_i(t) - D_{i,max}}{D_{i,max}} \quad (4)$$

where $D_i(t)$ is the waiting time of the HOL packet in the i th flow (destined to MT i) and $D_{i,max}$ is the maximum tolerable delay of the HOL packet. $D_{i,max}$ is specified as a QoS parameter for each class and is given by:

$$D_{i,max} = \begin{cases} D_{i,max}, & \text{for Class I} \\ \max(t_{a,i}, F_{i,-1}) + \frac{L_i}{R_i}, & \text{for Class II} \\ \infty, & \text{for Class III.} \end{cases} \quad (5)$$

For delay-sensitive traffic (Class I), the maximum delay is specified by $D_{i,max}$ as input QoS parameter. For best-effort traffic (Class III), the maximum delay is set to infinity. Thus the weighting function in (3) becomes the unity and the bandwidth is shared by the PF policy among the MTs. For Class II traffic which is sensitive to throughput, $D_{i,max}$ is set to the finish time of the HOL packet. The notion of finish time is central to service discipline called weighted fair queueing (WFQ) and finish time for the i th flow is given by [7]:

$$F_i(t) = \max(t_{a,i}, F_{i,-1}) + \frac{L_i}{R_i} \quad (6)$$

where $t_{a,i}$ is the arrival time of the HOL packet, $F_{i,-1}$ is the finish time of the previous packet, L_i is the packet length, and R_i is the promised rate (or bandwidth). WFQ scheme attempts to emulate packet flow in ideal fluid model by

calculating the departure time of a packet (i.e., finish time) in a corresponding fluid model and using this virtual time stamp to schedule packets. The advantage in offering throughput guarantees via WFQ-like policy against the periodic counter in [4] is more accurate and fair in distribution of bandwidth among the competing flows.

Since there often exists mismatch between physical packet size and data size from the higher layer, the expression of finish time in (6) needs to be modified to take into account the case in which multiple data packets are transmitted over a single packet. In this case, the representative arrival time $t_{a,i}$ for a group of packets including the HOL and its subsequent ones is set to that of the HOL packet. Once such a flow is selected for transmission, the packets behind the HOL enjoy a free ride and the flow gets more bandwidth than necessary. To correct this, the actual finish time is calculated after transmission. The actual finish time $F_{i,k+n-1}$ for n packets conditioned that the k th packet of the i th flow is at HOL, is

$$F_{i,k+n-1}(t) = \max(t_{a,(i,k)}, F_{i,k-1}(t)) + \sum_{j=k}^{k+n-1} \frac{L_{i,j}}{R_i} + \sum_{j=k}^{k+n-2} (t_{a,(i,j+1)} - F_{i,j}(t)) \cdot \mathbf{1}_{\{t_{a,(i,j+1)} > F_{i,j}(t)\}} \quad (7)$$

where $\mathbf{1}_x$ is the indicator function whose value is 1 if the condition x is satisfied and 0 otherwise. Previous finish time $F_{i,-1}$ in (6) becomes equivalent to (8).

2.3 Configuration of Parameters

As delay of the HOL packet approaches the prescribed target, how the weighting function $(C_{wd}(t)(\overline{D_i(t)} + 1)^\alpha + 1)$ varies is depicted in Fig. 1. Here, $C_{wd}(t)$ is set to 63 for illustration. As the figure indicates, the rule is designed so that the flows are scheduled on the PF policy as long as delay of the HOL packet has a sufficient margin from the target, but the delay-based priority part in the metric gradually overrides the PF policy as the delay approaches to the target. By scheduling this way our proposed rule can utilize temporal channel variation and consequently maximize the channel capacity. In contrast, the Exponential rule (2) suppresses selection of the HOL packets whose weighted delays are less than mean value, regardless of the channel conditions.

The weighting factor $w_i(t)$ is another parameter configured at the connection setup phase, according to the service requirements of traffic. The values of $w_i(t)$ for each traffic class is set as follows:

$$w_i(t) = \begin{cases} 1, & \text{for Class I} \\ \text{variable (0-1)}, & \text{for Class II} \\ 0, & \text{for Class III.} \end{cases} \quad (8)$$

The $w_i(t)$ is fixed at 1 for Class I. In this case the resulting rule behaves like a delay-based priority scheme for lagging flows ($\overline{D_i(t)} > 0$), whereas leading or in-sync flows ($\overline{D_i(t)} \leq 0$) still run on the PF policy. For Class III, $w_i(t)$ is fixed

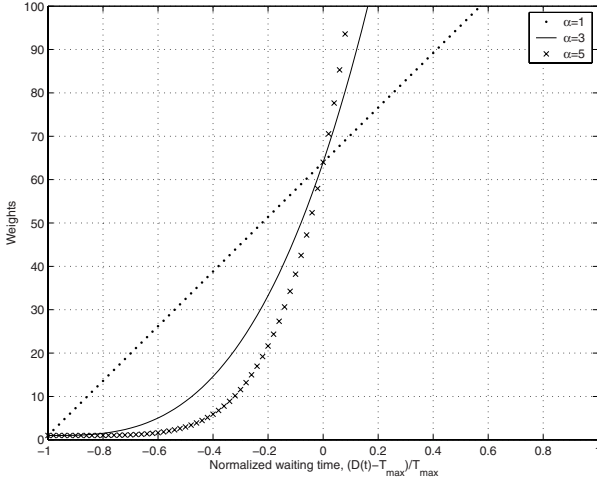


Fig. 1. Weight function $(C_{wd}(t)(\overline{D_i(t)} + 1)^\alpha + 1)$ versus $\overline{D_i(t)}$.

at 0, suppressing the weighting function and equalizing the rule to the PF policy. For Class II, $w_i(t)$ varies depending on how strictly bandwidth guarantee is required. Together with $w_i(t)$ an associated parameter denoted by Δ is used to indicate tolerable deviation from the target throughput. The rationale behind introducing this parameter is to control the way of allocating bandwidth over time. Depending on the type of applications, MTs may have different level of expectation on bandwidth guarantees. For applications that require a constant bandwidth over time, the rule needs to behave like WFQ rule. For the opposite case, some applications, e.g., Web browsing, may care about just the total amount of bandwidth allocated over the entire session, but not about strict guarantees over a microscopic time scale. For Class II, $w_i(t)$ is initially set to 0, and is reset to w_0 as the average rate $R_i(t)$ deviates from the promised rate \widetilde{R}_i by Δ amount. From then on $w_i(t)$ is multiplied with the constant β until it reaches 1 or the average rate exceeds the promised rate. The following summarizes the algorithm:

1. Initialize $w_i(t)$,

$$w_i(t) = 0$$

2. If the average rate deviates from the promised rate by larger than the tolerance Δ , reset $w_i(t)$ to a non-zero value w_0 .

$$w_i(t) = w_0, \quad \text{if } \frac{R_i(t) - \widetilde{R}_i}{\widetilde{R}_i} < -\Delta$$

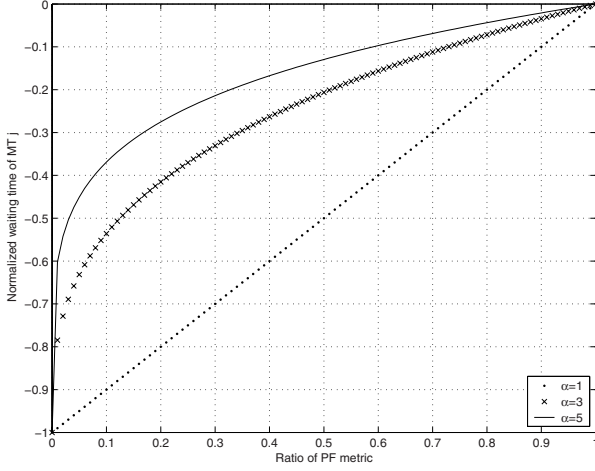


Fig. 2. Prioritization of leading flow over in-sync flow.

3. Update $w_i(t)$ at each scheduling epoch, i.e.,

$$w_i(t) = \begin{cases} \beta \times w_i(t), & w_i(t) < 1, \\ 1, & w_i(t) \geq 1 \end{cases}$$

4. Repeat Step 3 until $R_i(t) > \widetilde{R}_i$
5. If $R_i(t) > \widetilde{R}_i$, then $w_i(t) = 0$

Whereas the parameter $C_{wd}(t)$ is designed to prioritize Class I and II traffic over Class III, the parameter α is introduced to prioritize among the flows belonging to Class I or II. To analyze the impact of this parameter, we derive the condition in which a leading flow j gets scheduled before an in-sync flow i , i.e., $\overline{D}_i(t) = 0$. Assuming both the flow i and j belong to Class I for brevity, the metric of the flow j must satisfy:

$$\frac{\text{DRC}_i(t)}{\hat{R}_i(t)} (C_{wd}(t) + 1) < \frac{\text{DRC}_j(t)}{\hat{R}_j(t)} \left(C_{wd}(t) (\overline{D}_j(t) + 1)^\alpha + 1 \right).$$

Since $C_{wd}(t) \gg 1$, $1/C_{wd}(t) \approx 0$ and $\overline{D}_j(t)$ must satisfy

$$\overline{D}_j(t) > \left(\frac{\text{DRC}_i(t)}{\hat{R}_i(t)} / \frac{\text{DRC}_j(t)}{\hat{R}_j(t)} \right)^{1/\alpha} - 1.$$

Figure 2 shows the conditions in which a leading flow j gets prioritized over an in-sync flow i for $\alpha = 1, 3, 5$. As α becomes larger, the leading flow must see much better channel and the waiting time of the HOL packet must further approach the target delay bound. Larger α leads to much stricter delay-based

scheduling among Class I and II flows. However there is a tradeoff in choosing α between exploiting temporal channel variation and delivering more accurate QoS requirements as Fig. 1 indicates.

3 Numerical Results and Simulations

In this section, we present simulation results for the scheduling policy described in the previous section. We consider a CDMA cell in which several MTs run different applications under time-varying channel conditions. The channel models are Rayleigh fading, which is further approximated into finite-state Markov channel model following the approach in [8]. The Markov channel model has 10 channel states that correspond to the DRC rates listed in Table 1. The E_c/N_t thresholds for DRC selection in Table 1 are used to determine the ranges of E_c/N_t mapping onto the states of the discrete Markov channel model [9]. The speed of a MT determines how fast the channel varies, which will be characterized by the transition probability of the Markov channel model.

We simulate a scenario of mixed traffic in which MTs with RT traffic or with NRT traffic are uniformly positioned in a cell, competing for downlink bandwidth. We assume that NRT traffic generates packets following a Poisson process, whereas RT source traffic is modelled by the real-time video streaming model in [10]. In this model, a video streaming session consists of a sequence of frames with the period of 0.1 seconds. The number of packets generated in each frame is fixed at 8, and the packet inter-arrival time and packet size is distributed by truncated Pareto distribution:

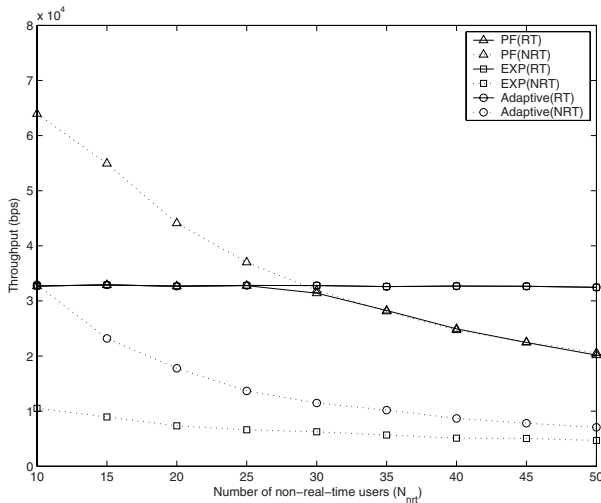
$$F(x) = \begin{cases} 1 - \frac{K^a}{x^a}, & x < m, \\ 1, & x \geq m, \end{cases}$$

with $a = 1.2$, $K = 0.0025$ seconds, $m = 0.00125$ seconds for the inter-arrival time, and $a = 1.2$, $K = 20$ bytes, $m = 125$ bytes for the packet size, respectively. With these parameters the mean rate of a single video traffic is 32 Kbps. For the sake of simplicity, no packet error is assumed. Since DRC rate is selected to target at 1 % packet error rate [1], we believe the impacts of packet error on system performance can be safely ignored. Unless noted otherwise, the results are obtained with $\alpha = 3$. All the simulation results are reported based on 1000 seconds run, equivalent to 1,670,000 slots. Table 2 summarizes the values of the various parameters in simulation.

Figure 3 depicts throughput for three different scheduling policies: PF, EXP (Exponential rule in [4]), and Adaptive rule proposed in this paper. All the MTs are under homogeneous channel conditions, i.e., mean $E_c/N_t = 0$ dB and $v = 3$ Km/h. The number of MTs with RT application is fixed at 5. In terms of throughput, the PF policy shows the best performance due to its property of placing absolute priority on the MT seeing the best channel. However, throughput of RT traffic decreases proportionally as the shared link becomes loaded with more traffic. Obviously it is because RT traffic in the PF policy is equally

Table 2. Parameter values used in the simulations.

Parameters	Symbol	Value
Number of MTs with RT application	N_{rt}	3 - 5
Number of MTs with NRT application	N_{nrt}	10 - 50
Maximum inter-arrival time of NRT traffic	-	0.01 seconds
Mean packet size of NRT traffic	-	640 bits
Video frame period	-	0.1 seconds
Number of packets in a video frame	-	8
Mean rate of video traffic	-	32 Kbps
Mean inter-arrival time of video packets	-	0.006 seconds
Maximum inter-arrival time of video packets	-	0.0125 seconds
Mean packet size of video traffic	-	50 bytes
Maximum packet size of video traffic	-	125 bytes
Speed of MT (Km per hour)	v	3 - 120 Km/h

**Fig. 3.** Throughput of MTs with RT and NRT traffic ($v = 3$ Km/h).

treated with other NRT traffic. In contrast, the EXP and Adaptive policies guarantee a constant throughput (i.e., 32 Kbps) irrespective of traffic conditions. In particular, the Adaptive policy yields higher throughput in NRT traffic than the EXP policy. For $N_{nrt} = 10$, the Adaptive policy achieves twice the throughput of NRT traffic that the EXP policy does.

Figure 4 and 5 shows throughput and delay CDF of RT traffic, respectively, under heterogeneous channel conditions in which mean E_c/N_t of 0 dB, -3 dB, and 3 dB is given to three MTs. For the case of $v = 3$ Km/h in Fig 4, throughput decreases proportionally with worse channel conditions under the PF policy, whereas the Adaptive policy offers a steady throughput irrespective of channel conditions. In delay CDF, both the EXP and the Adaptive policies miss 0.2

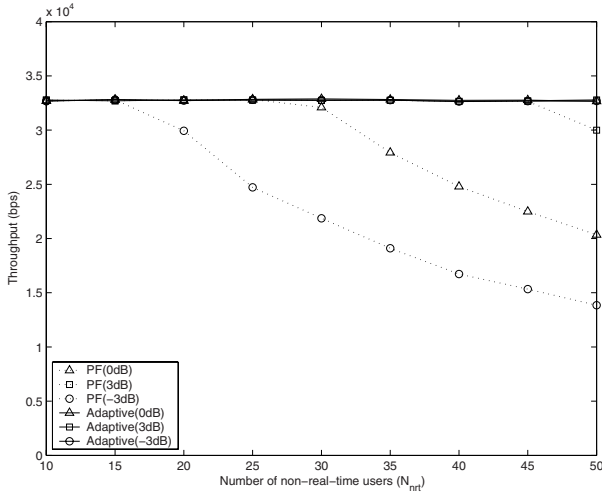


Fig. 4. Throughput under heterogeneous channel environments ($v = 3$ Km/h, $N_{nrt} = 30$).

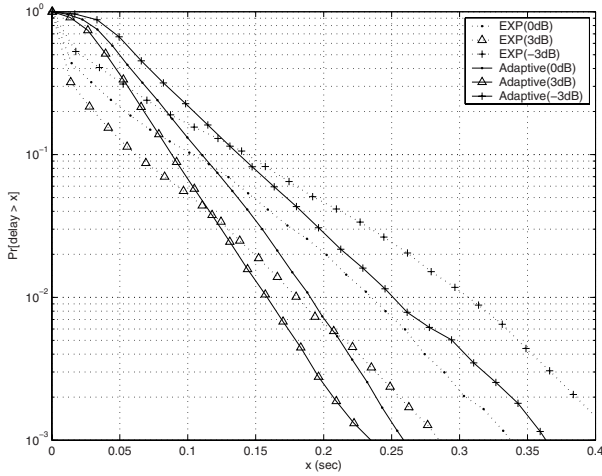


Fig. 5. Delay CDF under heterogeneous channel environments ($v = 3$ Km/h, $N_{nrt} = 30$).

seconds delay bound. However, the EXP policy fails at 0 and -3 dB, but the Adaptive policy does only at -3 dB.

Delay performance under heterogeneous QoS requirements is shown in Fig. 6 and 7. Five MTs having RT traffic set maximum delay bound at 0.1, 0.2, 0.3, 0.4, and 0.5 seconds, respectively. In the Adaptive policy, those bound are set to the parameter of $D_{i,max}$. In the EXP policy, the weight factor is set following

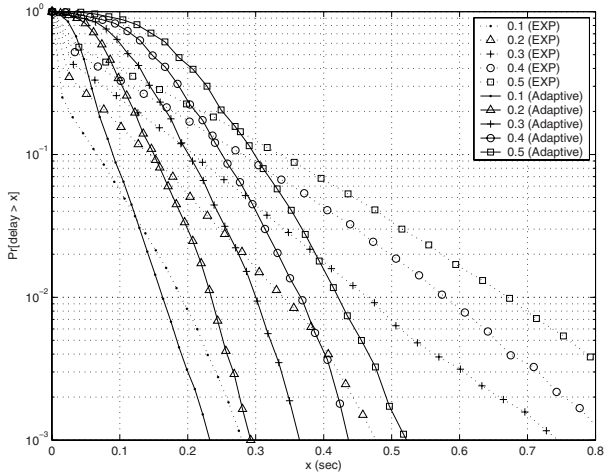


Fig. 6. Delay CDF with heterogeneous QoS requirements ($v = 3$ Km/h, $N_{nrt} = 50$).

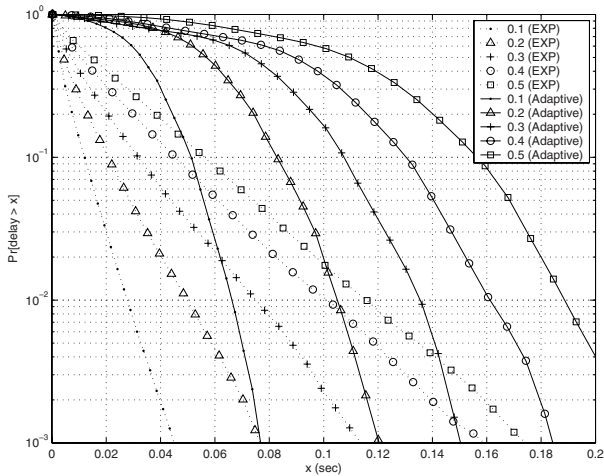


Fig. 7. Delay CDF with heterogeneous QoS requirements ($v = 50$ Km/h, $N_{nrt} = 50$).

the formula $a_i = -\log(\delta_i)/T_i$ suggested in [4]. Here δ_i and T_i is derived from the following delay requirement:

$$P[\text{Delay} > T_i] \leq \delta_i.$$

Thus, δ_i is set to 0.01 and T_i is set to from 0.1 to 0.5 as above. For the case of $v = 3$ Km/h, both policies did not deliver successfully the specified QoS requirements. However when it comes to the level of deviation, the EXP policy far exceeds the desired delay bound, whereas the Adaptive policy offers a delay bound relatively close to the specified values. For $v = 50$ Km/h, both policies

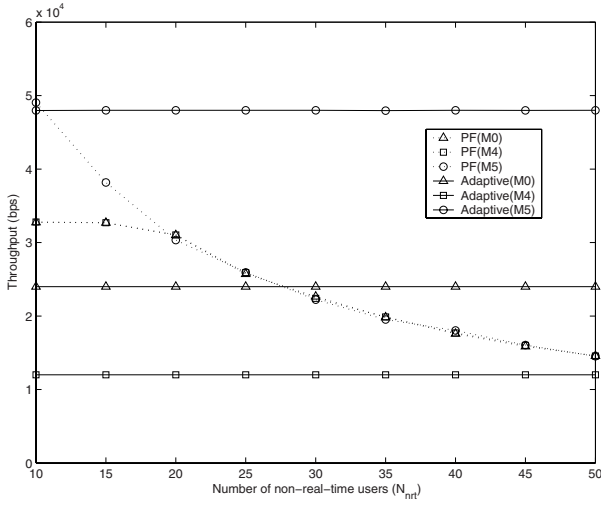


Fig. 8. Throughput performance for MTs with heterogeneous requirements ($v = 3$ Km/h, Mean $E_c/N_t = -3$ dB).

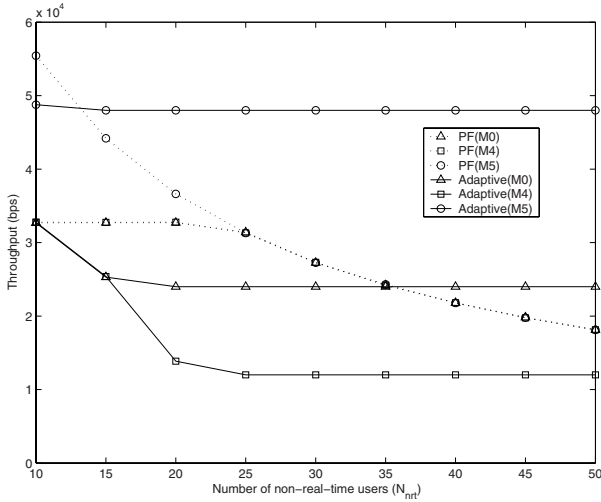


Fig. 9. Throughput performance for MTs with heterogeneous requirements ($v = 50$ Km/h, Mean $E_c/N_t = -3$ dB).

meet the maximum delay bound as specified, but in a somehow conservative manner. In particular, the EXP policy exercises too much bandwidth as shown in Fig. 7. From those results, the Adaptive policy can offers a maximum delay bound more accurately relative to the EXP policy.

Figure 8 and 9 shows throughput performance of the Adaptive policy in a scenario where MTs have traffic requiring a strict throughput guarantee. Two MTs, M0 and M4, have mean data rate of 32 Kbps for both and target throughput of 12 Kbps and 24 Kbps, respectively. And the MT M5 has the mean data of 64 Kbps and target throughput of 48 Kbps. Under the PF policy all MTs yield basically equivalent throughput given the same channel conditions. Due to this property of fair sharing, throughput per a MT decreases as the number of MTs increases for $v = 3, 50$ Km/h. Thus, it is not possible to guarantee the requested bandwidth with the PF policy. In contrast, the Adaptive policy, as shown in the figures, provides a steady bandwidth without being affected by traffic conditions. Notably two MTs, M0 and M4, requiring the target of 12 and 24 Kbps, respectively, get some extra bandwidth at $N_{nrt} = 10, 20$ (See Fig. 8 and 9) since overall link capacity increases with faster speed and the system is not highly loaded. This indicates the Adaptive policy basically runs on the PF policy, and switches to WFQ-like mode when QoS cannot be satisfied with the PF policy. Such capability of adaptation is significant in that it can exploit better throughput performance of the PF policy while offering QoS guarantees when needed.

4 Conclusions

In this paper, we proposed a new scheduling policy that provides QoS guarantees to a variety of traffic types demanding diverse service requirements. In our proposal, data traffic is categorized into three classes, depending on sensitivity of its performance to delay or throughput. And the primary components of our policy, namely, the PF, WFQ, and delay-based prioritized scheme are intelligently combined to satisfy QoS requirements of each traffic type. Our policy changes the level of emphasis on each of those ingredient policies in an adaptive manner, taking into account the channel conditions and QoS requirements. Such flexibility, as shown in numerical results, leads to offering QoS guarantees effectively and, at the same time, maximizing the throughput.

References

1. Bender, P., Black, P., Grob, M., Padovani, R., Sindhusayana, N., Viterbi, A.: CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Commun. Mag.* **38** (2000) 70–77
2. Jalali, A., Padovani, R., Pankaj, R.: Data throughput of CDMA-HDR: a high efficiency, high data rate personal communication wireless system. *Proc. IEEE VTC '00*, **3** (2000) 1854–1858
3. Kelly, F.: Charging and rate control for elastic traffic. *European Transactions on Telecommunications* **8** (1997) 33–37
4. Shakkottai, S., Stolyar, A. Scheduling algorithms for a mixture of real-time and non-real-time data in HDR. *Proc. International Teletraffic Congress '2001* (2001)
5. Liu, P., Berry, R., Honig, M.: Delay-sensitive packet scheduling in wireless networks. *Proc. IEEE WCNC '03*, **3** (2003) 1627–1632

6. Huang, C., Su, H., Vitebsky, S., Chen, P.: Schedulers for 1xEV-DO: third generation wireless high-speed data systems. Proc. IEEE VTC '03 **3** (2003) 1710–1714
7. Zhang, H., Service discipline for guaranteed performance service in packet-switching networks. Proc. IEEE **83** (1995) 1374–1396
8. Wang, H., Moayeri, N.: Finite-state Markov channel—a useful model for radio communication channels. IEEE Trans. Veh. Technol. **44** (1995) 163–171
9. Paranchych, D., Yavuz, M.: Analytical expression for 1xEV-DO forward link throughput. Proc. IEEE WCNC '03 **1** (2003) 508–513
10. 3GPP2: 3GPP2 1xEV-DV Evaluation Methodology (2001)