Distributed Uplink Scheduling in CDMA Networks

Ashwin Sridharan[1], Ramesh Subbaraman[2] and Roch Guérin[2] ⋆

[1] Sprint Advanced Technology Labs, ashwin.sridharan@sprint.com
[2] University of Pennsylvania, {rameshrs, guerin}@seas.upenn.edu

Abstract. Ever more powerful mobile devices are handling a broader range of applications, so that giving them greater control in scheduling transmissions as a function of application needs is becoming increasingly desirable. Several standards have, therefore, proposed mechanisms aimed at giving devices more autonomy in making transmission decisions on the wireless *uplink*. This paper explores the impact this can have on total throughput in CDMA systems, where this control has traditionally been centralized. The investigation relies on a simple distributed policy that helps provide insight into the impact of distributed decisions on overall system efficiency, and identify guidelines on how to best mitigate it.

1 Introduction

With the power and versatility of mobiles[3] rivaling that of stationary platforms, the diversity and communication requirements of applications they run have also been expanding. This has resulted in a push to give mobiles more autonomy in making transmission decisions. This, however, often conflicts with the centralized operation of current wireless systems, e.g., the control exercized by base stations or the use of 802.11 RTS/CTS handshakes between devices and access points. In this paper, we explore the tension this creates in the uplink of CDMA systems.

Traditional CDMA base stations tightly control transmission schedules and power to maintain acceptable signal to interference levels. Several standards for modern 3G/4G cellular networks, e.g., 1xEV-DO Rev. A [1], HSUPA [2], have, however, introduced mechanisms that give devices significant autonomy in deciding when to transmit and at what rate. As stated in [3], a major driver was to define a *"wide-area-mobile wireless Ethernet,"* where devices had greater independence in making transmission decisions best matched to their applications. The price for this flexibility is potentially higher interferences, and a corresponding degradation in performance. Investigating this issue is what motivated this paper.

One proposed mechanism for allowing distributed transmission decisions while maintaining some control on resource sharing, is a *token bucket* [1] similar to that

⋆ [2]The work of these authors was supported in part by a Gift from the Sprint Corporation and through NSF Grant CNS-0627004.

[3] We use mobiles, devices and users interchangeably.

used in wired networks [4]. Each token grants access to a certain amount of "resources," with token generation *rate* and token bucket *depth* imposing limits on resource consumption. Mobile devices decide how to spend their tokens to achieve transmission rates (and latencies) best suited to their applications. Unlike wired networks where the token "currency" is in bytes, tokens are now in units of transmission power, the primary resource in a CDMA system.

This leads to resource sharing models fundamentally different from the "queueing systems" that capture buffer and bandwidth consumption in wired networks. Instead, as discussed in Section 4, the sharing of resources among users is measured through the resulting signal to interference and noise ratio (SINR). We develop models that reflect this sharing, with users making independent but constrained transmission decisions, where constraints arise from token bucket mechanisms. We first investigate a simple distributed policy, with users randomly and independently alternating between idle and active periods. Token bucket constraints are introduced next that limit the frequency of active periods. We derive expressions in both settings for the achieved user rates as functions of the frequency of active periods and rate selection. This enables us to explore the impact of distributed transmission decisions on performance and assess the efficacy and effect of token buckets. Our evaluation is carried out in the context of uplink transmission in a single cell. Extensions to the multi-cell scenario will be considered in future work.

The rest of the paper is structured as follows. Section 2 provides basic background on CDMA systems and reviews related work. Section 3 is a short tutorial on the operation of 3G/4G cellular uplinks. Our resource sharing model is covered in Section 4, while Section 5 describes our evaluation framework. Our analysis is covered in Section 6, with Section 7 comparing its results to simulations. Section 8 summarizes our findings and points to future work.

2 CDMA Reverse Link: Related Work

In CDMA systems [1, 2] the reverse link or uplink, i.e., from mobiles to the access network, is well-known to be interference limited. Because users share the same spectrum and their signals are not perfectly orthogonal, the throughput they see is a function of both their own transmission power and that of *other users* whose transmissions are perceived as interference. This introduces a trade-off as increasing either ones own signal strength or transmission frequency also increases interferences to others. A user must therefore first decide *what transmission power* to utilize and second, *when and how often* to transmit.

There is a large literature addressing optimum transmission power selection. See, for example [5–10] that address the problem of joint allocation of transmission power and associated QoS functions (e.g., rate). These studies however do not address the second issue of uplink scheduling. Off late, this topic has received some attention, notably [11–13] where authors proposed joint scheduling and power allocation algorithms that take multi-user interferences into account. These works, however, assume a centralized control of users transmissions, which places a heavy burden on the system in terms of signaling overhead and scalability.

3 The 3G/4G Cellular CDMA Reverse Link

We now describe two key features of the operation of 3G/4G CDMA uplinks, which play important roles in enabling and affecting distributed scheduling policies.

The first, *pilot assisted transmission*, governs the transmission power level of devices. Each device transmits a *pilot signal* on the uplink whose strength is set by the access network through a fast closed control loop to ensure that all pilot signals are *received* with equal power. When selecting an uplink transmission data rate, the device then sets its transmission power *relative* to the strength of the pilot signal. Specifically, if at time slot t the pilot strength of device i is $P_S^i(t)$, transmission at a data rate R requires a transmission power $P_D^i(R,t)$:

$$P_D^i(R,t) = TxT2P[R] \cdot P_S^i(t) \tag{1}$$

where $TxT2P[R]$ is an *a priori* specified proportionality factor function of the target rate R. This mechanism does not yield optimal transmission power selections, but ensures a level playing field to all devices by equalizing the strength of all pilot signals at the receiver. The fast power control loop allows the pilot signal to track variations of the wireless channel, and as a result *the transmission power for the data is de-coupled from the problem of coping with fading and attenuation on the wireless channel.* As we shall see later, this is a critical aspect of the system.

The second and more recent feature of CDMA systems is the use of a *token-bucket*[4] to control how devices access the uplink. Each MAC (layer 2) flow is assigned a token bucket which can hold σ "power-tokens" and is filled at a rate of ρ. To transmit at a data rate of R, a device must have $TxT2P[R]$ tokens that are then subtracted from its bucket[5]. The higher R, the higher $TxT2P[R]$, and hence the faster the token bucket drains. Through the token bucket parameters (σ_i, ρ_i), the network limits the maximum transmission power *and* frequency of devices, and therefore controls both the total uplink power and its allocation across devices. On the other hand, it relinquishes scheduling decisions to devices by letting them control the use of their power tokens.

4 System Model

The system consists of a single cell with $n+1$ homogeneous and continuously backlogged users sharing a time-slotted uplink. Denote the *data* transmission

[4] Also known as *Grants* in HSUPA [2].

[5] In practice there are only a finite set of rates a device can transmit at(e.g., six in [1]).

power of user i in slot t when transmitting at rate R_i as $P_D^i(R_i, t)$. Under the CDMA sharing model, the SINR of user i in slot t is then given by [14]:

$$S_i(R_i, t) = \frac{G(R_i) \cdot G_{loss}^i(t) \cdot P_D^i(R_i, t)}{\sigma^2 + \theta \sum_{j \neq i} G_{loss}^j(t) \cdot P_D^j(R_j, t)}, \tag{2}$$

where $\theta \in [0, 1]$ quantifies the orthogonality of the codes, σ^2 is the thermal noise, $G_{loss}^i(t)$ the path loss[6] of user i in time slot t, and $G(R_i) = W/R_i$ its processing gain, where W is the spread-spectrum bandwidth.

Recall from Section 3 that the pilot signal of each device is controlled by a fast control loop so that received pilot strengths are all identical at the base station. This can be modeled as each pilot signal seeking a *common* target SINR $1/\phi$. Let $P_S^i(t)$ be the pilot strength of user i in slot t. Assuming perfect power control and unconstrained transmission power, it is then easy to see that the pilot power control loop requires each device to set $P_S^i(t)$ such that,

$$G_{loss}^i(t) \cdot P_S^i(t) = \Delta = \frac{\sigma^2}{\phi - \theta_p n}, . \tag{3}$$

where θ_p is the orthogonality factor for the pilot signal. Since, the data transmission power $P_D^i(R, t)$ is relative to the pilot strength, using Eqn. (1) and Eqn. (3), Eqn. (2) can be written as

$$S_i(R_i, t) = \frac{G(R_i) \cdot TxT2P[R_i] \cdot \Delta}{\sigma^2 + \theta \sum_{j \neq i} TxT2P[R_j] \cdot \Delta}. \tag{4}$$

Eqn. (4) states that with perfect power control and no power constraint, the SINR of a user is influenced only by other users rate choices and not the channel.

5 Evaluation Framework

Given our assumption of continuously backlogged and homogeneous users, a metric of interest is long-term throughput. For simplicity, we first approximate the *effective* rate achieved by a user in a time slot as *linearly* proportional to its SINR. Specifically, the effective rate achieved by a user i in time slot t, *given* that it transmitted at rate R_i, is given by

$$C_i(t) = \frac{S_i(R_i, t)}{S_o} R_i \tag{5}$$

where $S_i(R_i, t)$ is given by Eqn. (4) and S_o is the target data SINR. The linearity assumption is typically valid at small SINR values [14] as long as the modulation scheme remains the same. A limitation of Eqn. (5) is, however, that whenever $S_i(R_i, t)$ is greater than S_o, it yields an effective rate *greater* than the

[6] Function of distance to base station and fast fading.

transmission rate R_i. This is clearly not possible, and hence, the above relation is modified to

$$C_i(t) = \min(R_i, \frac{S_i(R_i, t)}{S_o} R_i)\,. \tag{6}$$

We refer to Eqn. (5) as the *Linear Model* and Eqn. (6) as the *Bounded Model*. In both cases, the metric of interest is average user achieved rate, $\widehat{C} = E[C_i(t)]$.

Apart from $\widehat{C}$, for token bucket constrained systems, the *token efficiency,* i.e., the achieved effective rate per token expended, is another metric of interest. If T is the expected number of tokens expended per time slot, the token-bucket efficiency is defined as $\eta = \widehat{C}/T$.

5.1 The scheduling policy

We consider a transmission scheme where in each slot when a user has enough tokens it either transmits at rate R with probability p, or doesn't transmit at all with probability $1 - p$. All users are assumed independent and identical in their transmission behaviour, i.e., p and R are the same for all users.

Although simple, this policy is interesting for several reasons. First, it is inherently distributed, which when combined with its simplicity makes it eminently practical. Indeed, it has direct equivalents in wireline networks, e.g., Aloha, CSMA etc. Second, by virtue of easily controllable parameters, it lets us explore and understand key system properties, e.g., impact of cell load, transmission rate, etc., which we show can strongly influence performance. In addition, it captures a hybrid sharing model between pure CDMA (all users transmitting) and a slotted-system (one user transmitting at a time) that has the potential to enable distributed control while improving performance.

6 Analysis of Scheduling Behaviour

In this section, we analyze the performance of the on-off scheduler, and identify how the value of p that maximizes throughput depends on both the number of users in the cell $(n + 1)$ *and* their selected transmission rate R. The analysis is based on Eqn. (4) and assumes perfect power control. We also assume that S_o in Eqn. (5) is large enough that the rate obeys the *Linear Model*. The impact of these assumptions is explored in Section 7. Due to space limitations, we refer the reader to the technical report [15] for all proofs.

6.1 Scheduler Behaviour: No Token Bucket

From Section 3, the signal power to transmit at data rate R is $P_D^i(R, t) = TxT2P[R] \cdot P_S^i(t)$. For ease of exposition, assume that R is fixed and let $TxT2P[R] = K$.

Let K_i be the random variable denoting the transmission power used in a slot by user i. Under the on-off scheduler with no token constraints, K_i is a

Bernoulli random variable that takes values K and 0 with probability p and $1-p$, respectively. After minor algebraic manipulations of Eqn. (4) and using Eqn. (5) and $G(R) = W/R$, the expected achieved data rate $\widehat{C}(p)$ is found to be

$$\widehat{C}(p) = \frac{W}{\theta \cdot S_o} \cdot p \cdot \sum_{j=0}^{n} \frac{1}{j+\delta} \binom{n}{j} p^j (1-p)^{n-j}, \tag{7}$$

where $\delta = \dfrac{\gamma}{\theta K} = \dfrac{\phi - \theta_p n}{\theta K}$. We now state two propositions that capture and elicit the impact of δ on the scheduling parameter p and the achieved rate $\widehat{C}(p)$.

Proposition 1. *If $\delta \geq 1$, then the expected achieved rate $\widehat{C}(p)$ attains its maximum value at $p^* = 1$.*

Proposition 2. *If $\delta < 1$, then the expected achieved rate $\widehat{C}(p)$ has a* unique *maximum at $p = p^* < 1$.*

In either case, p^* satisfies the following equation:

$$\sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} \frac{1}{j+\delta} = \frac{1}{(n+1)p - 1 + \delta}\,. \tag{8}$$

Since $\delta = \dfrac{\phi - \theta_p n}{\theta K}$, assuming ϕ is fixed[7], Propositions 1 and 2 reflect the impact of the number of users *and* the selected rate on the optimal p. Specifically, with few users or low enough transmission rates so that $\delta \geq 1$, the optimal policy yields a *pure CDMA* system ($p^* = 1$) where everybody transmits, exploiting the orthogonality of the CDMA codes. As the load increases (n or $K \nearrow$) so that $\delta \leq 1$, the increased interference triggers a transition to a *hybrid* slot-division/CDMA allocation with only some users active in any given slot. This reflects the trade-off between reducing interference ($p \searrow$) and increasing transmission opportunities ($p \nearrow$).

Propositions 1 and 2 characterize the transition point precisely through δ. Similar results were obtained in [12], albeit in a centralized setting. Next, we study the impact of selecting different transmission rates in the *on* state on the optimal achieved rate $\widehat{C^*}$.

Proposition 3. *Let $\widehat{C_1^*}$ and $\widehat{C_2^*}$ be the optimal achieved rates when the transmission rates in the* on *state are R_1 and R_2, respectively. If $R_1 > R_2$, then $\widehat{C_1^*} > \widehat{C_2^*}$.*

The proposition states that under an on-off scheduler, increasing the transmission rate R always improves throughput. Hence, one should select the (R, p^*) combination with the highest R. However, this is true only for an *unconstrained system*, and need not hold when token constraints are present. In such a setting *token efficiency* matters, and lower rates may fare better than higher ones that consume more tokens.

[7] A typical target Pilot SINR $1/\phi$ is between -26 dB and -17 dB and the same for homogeneous users.

6.2 Incorporation of a Token Bucket

The previous section established that controlling transmission frequency and rate matters when devices make independent decisions. This can be realized by a token bucket with parameters (ρ, σ). The token rate ρ bounds transmission frequency and rate, while the bucket depth σ affords flexibility in scheduling decisions. Next, we use the results of Section 6.1 to explore how to spend tokens (transmissions at rate R cost K tokens) to maximize throughput.

Consider the on-off scheduler, but now operating under token bucket constraints. Specifically, if p is the *conditional* transmission probability *given* enough tokens ($\geq K$), the token bucket evolution can be modeled as a Markov chain to obtain the stationary distribution π_l of having l tokens in the bucket [15]. The *unconditional transmission probability* p_{tok} is then given by Eqn. (10) that together with Eqn. (7) can be used to approximate $\widehat{C}$, so that the optimum pair (K^*, p^*) is the solution of the non-linear program $\mathbf{N_1}$:

$$\mathbf{N_1} : \max_{p, K \in \mathcal{K}} \widehat{C}(p, K) \tag{9}$$

where

$$\begin{aligned} \widehat{C}(p, K) &= \frac{W}{\theta S_o} \cdot p_{tok} \cdot \sum_{j=0}^{n} \frac{1}{j + \delta} \binom{n}{j} p_{tok}^{j} (1 - p_{tok})^{n-j}, \\ p_{tok} &= p \cdot (1 - \sum_{k=1}^{K-1} \pi_k), \\ 0 &\leq p_{tok} \leq 1 \,. \end{aligned} \tag{10}$$

An algorithm that solves program $\mathbf{N_1}$ is described in [15], and evaluated in Section 7.2.

7 Simulation results

We explore the validity of the analysis of Section 6 and the roles of rate selection and transmission probability in both unconstrained (Section 7.1) and token bucket constrained systems (Section 7.2). Results are obtained using a detailed simulator of the uplink that incorporates key characteristics of the channel model and transmission system. The target pilot strength is set to -17dB, which allows up to 50 active users to share the uplink. Simulations for both perfect and imperfect power control yielded very similar results. Hence, only results for the former are presented. Results for the latter can be found in [15] together with additional details on the simulator itself. All results have a 90% confidence interval.

7.1 Unconstrained System Evaluation

Our first goal is the validation of $\delta = 1$ as a transition point for the optimal policy, i.e., from $p^* = 1$ to $p^* < 1$. We focus on the linear model on which the

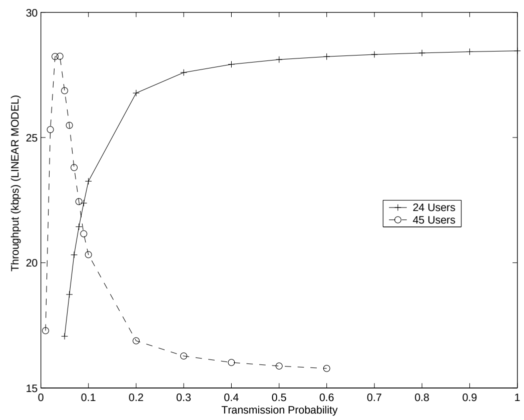

Fig. 1. Impact of δ

analysis of Section 6.1 is based, and plot in Fig. 1 the achieved throughput for two configurations: A lightly loaded system, $R = 76.8$ kbps and 24 users, for a load $\leq 50\%$ and $\delta > 1$; and a highly loaded system, $R = 76.8$ kbps and 45 users for a load $\approx 90\%$ and $\delta = 0.23 < 1$. The figure highlights the different optimal policy of each configuration ($p^* = 1$ for the former and $p* < 1$ for the latter) confirming predictions made in **Propositions 1** and **2**. Similar results were also obtained for the bounded rate model.

Next we explore the differences that exist between the linear and bounded rate models. We use a scenario with $\delta < 1$, and plot in Fig. 2 the achieved throughput as a function of p for the two rate models. For the linear model, the optimal $p^* \approx 0.03$ agrees with the solution of Eqn. (8). For the bounded rate model, the figure, however, highlights the impact of limiting the rate even as the SINR keeps increasing. The rate capping translates into a higher optimal $p^* \approx 0.2$, or in other words in allowing more simultaneously active users. The figure also illustrates for both rate models the benefits of the hybrid allocation of the on-off scheduler ($p < 1$) over both a pure CDMA system ($p = 1$) and a pure slot-based scheme (plotted on the right y-axis), where the latter was realized (for the bounded rate model) through a round robin scheduler that allowed only one user to be active in any time slot.

7.2 Token Bucket Constrained System Evaluation

Fig. 2 ($R = 153.6$ kbps, 45 users) and Fig. 1 ($R = 76.8$ kbps, 45 users) also validate **Proposition** 3, as they show that for the linear rate model, an (R, p^*) combination with a higher R is indeed better. As discussed earlier, this however ignores token efficiency. Indeed, Fig. 3 shows that $R = 76.8$ kbps has *higher* token efficiency for a 24 user system. We explore next how this affects throughput under token bucket constraints.

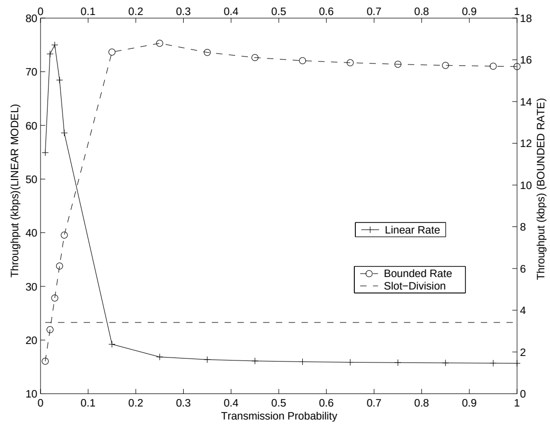

Fig. 2. Linear vs bounded rate – 45 users, $R = 153.6$kbps

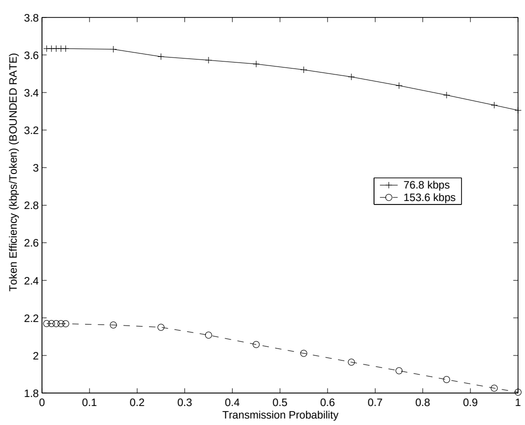

Fig. 3. Token Efficiency (24 users)

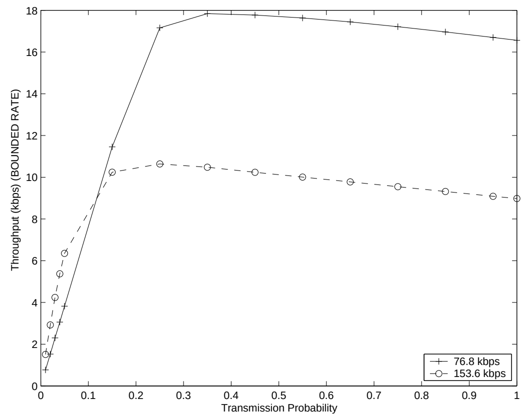

Fig. 4. Token bucket constrained system (24 users, bounded rate model)

Fig. 4 plots the achieved bounded rate throughput as a function of the *conditional* transmission probability p, i.e., probability of transmission *given* enough tokens in the bucket, for a 24 user system and a token bucket with $\rho = 7$ dB and $\sigma = 21.5$ dB. Based on the recommendations in [1], transmissions at $R = 153.6$ kbps require 18.5 dB worth of tokens and 13.5 dB at $R = 76.8$ kbps[8]. When p is low, $R = 153.6$ kbps yields better throughput than 76.8 kbps because the bucket is rarely exhausted and hence token efficiency is not critical. However, this is no longer true at higher values of p where the better token efficiency of 76.8 kbps yields a higher throughput. Overall, 76.8 kbps yields the highest achieved throughput because it provides a better compromise than $R = 153.6$ kbps between token efficiency and realized rate.

Last, we discuss the solution of program $\mathbf{N_1}$ of Section 6.2 that relies on the linear rate model and explore its differences with the bounded rate model (Fig. 4). Table 1 presents the optimal transmission probabilities p_A^* and achieved rates C_A^* obtained by solving $\mathbf{N_1}$ for both $R = 76.8$ kbps and $R = 153.6$ kbps in a 24 user system. It also gives the optimal values obtained by simulation for the bounded rate model. As expected, because the bounded rate model caps rates, its achieved rates are significantly lower. When it comes to optimal transmission probabilities however, the analytical results are in good agreement with simulations for $R = 153.6$ kbps. For $R = 76.8$ kbps the analytical $p^* = 1.0$ is higher than that predicted by simulations $p_{sim}^* = 0.35$. However, comparing the last column $C_{sim}(p_A^*)$ in Table 1, which shows the throughput achieved in simulations using the analytically computed p_A^*, with the optimal C_{sim}^*, we see that they are very close for both $R = 76.8, 153.6$ kbps. This indicates that the p_A^* obtained from solving $\mathbf{N}_1$ provide very reasonable estimates for setting the transmission probabilities in practice.

[8] See [15] for a full list of rate to token mappings.

Rate (kbps)	Token Bucket				
	Analysis :$\mathbf{N_1}$		Simulation		
	p_A^*	$\hat{C}_A^*$	p_{sim}^*	$\hat{C}_{sim}^*$	$C_{sim}(p_A^*)$
76.8	1.0	26.4	0.35	17.84	16.56
153.6	0.21	42.9	0.25	10.63	10.59

Table 1. Token Bucket - Bounded Rate Model

8 Conclusions and Future work

In this paper, we investigated the performance of a CDMA uplink, when transmissions decisions are distributed to mobiles. This was motivated by standard proposals [1, 2] that introduced support for such distribution. The investigation relied on a simple on-off scheduler to explore the impact of distributed transmission decisions, and identified both analytically as well as via simulations, key factors that affect system performance and how to account for them in designing a scheduling policy. The paper also investigated the realization of such distributed scheduling decisions through a token bucket, and how the token bucket operation affected the scheduler.

There are many possible extensions to this work, and we mention two we are currently exploring. The first is identification of the *optimal capacity* region for distributed decisions and how to achieve it. The second and possibly more important direction involves using the token bucket to provide *differentiated services* to users. Preliminary results can be found in [15].

References

1. QualComm: 1xEV: 1x EVolution, IS-856 TIA/EIA Standard (2004)
2. 3GPP TS 21.101 UTRAN-based 3GPP System, Rel. 6: High Speed Uplink Packet Access (2005)
3. Bhusha, N., Lott, C., R. Attar, P.B., Jou, Y.C., Fan, M., Ghosh, D., Au, J.: CDMA2000 1xEV-DO Revision A: A physical layer and MAC layer overview. IEEE Comm. Mag. **44**(2) (2006)
4. Turner, J.: New directions in communications (or which way to the information age?). IEEE Comm. Mag. **24**(10) (1986)
5. Foschini, G., Miljanic, Z.: A simple distributed autonomous power control algorithm and its convergence. IEEE Trans. Veh. Technol. **42**(4) (1993)
6. Holliday, T., Bambos, N., Goldsmith, A.J., Glynn, P.: Distributed power control for time varying wireless networks: Optimality and convergence. In: Proc. 41st Allerton Conf. (2003)
7. Hande, P., Rangan, S., Chiang, M.: Distributed Uplink Power Control for Optimal SIR Assignment in Cellular Data Networks. In: Proc. *INFOCOM.* (2006)
8. Price, J., Javidi, T.: Decentralized and Fair Rate Control in a Multi-Sector CDMA System. In: Proc. *WCNC.* (2004)
9. Yim, R., Shin, O.S., Tarokh, V.: A CDMA Reverse Link Rate Control algorithm with Fairness Gaurantee. In: Proc. *IWCT.* (2005)

10. Tinnakornsrisuphap, P., Lott, C.: On the fairness and stability of reverse-link MAC Layer in cdma2000 1xEV-DO. In: Proc. *ICC*. (2004)
11. Elbatt, T., Ephremides, A.: Joint Scheduling and Power Control for Wireless Ad-hoc Networks. In: Proc. *INFOCOM*. (2002)
12. Kumaran, K., Qian, L.: Uplink Scheduling in CDMA Packet-Data Systems. In: Proc.*INFOCOM*. (2003)
13. Cruz, R.L., Santhanam, A.V.: Optimal Routing, Link Scheduling and Power Control in Multi-hop Wireless Networks. In: Proc. *INFOCOM*. (2003)
14. Viterbi, A.J.: Principles of Spread Spectrum Communication. Addision-Wesley (1995)
15. Sridharan, A., Subbaraman, R., Guérin, R.: Distributed Uplink Scheduling in CDMA Systems. Research Report RR06-ATL12070139, Sprint ATL (2006)