DELAY CONTROL FOR CDF SCHEDULING WITH DEADLINES

PhuongBang C. Nguyen and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego

ABSTRACT

We propose two novel methods of service delay control for multiuser wireless systems that employ a scheduling scheme based on the Cumulative Density Function (CDF) of user channels in correlated fading environments. The first method allows for different hard delay deadlines for the users. The second method relaxes the deadline enforcement into "soft" deadlines in order to take better advantage of multiuser diversity. Both methods offer excellent performance while still preserving the temporal fair resource allocation property of the CDF scheduling policy.

***Index Terms*—** CDF, deadline, scheduling, delay

1. INTRODUCTION

Resource allocation has always played a critical role in any multiuser ecosystem. This task becomes even more challenging in the next generation heterogeneous networks (HetNets). The introductions of different base station tiers, different communication modes such as device-to-device or relay communications among others lead to large diversities between the channels experienced by the users, making it very difficult to satisfy fairness criteria while taking advantage of multiuser diversity gain. Many scheduling schemes have been developed to exploit multiuser diversity gain under some notion of fairness such as proportional fairness [1], [2], temporal fairness [3], [4], [5], and so on. All these schemes, however, are channel distribution-dependent and thus their behaviors become inconsistent across users in these new networks. In this paper, we focus on the CDF-based scheduling (CS) policy introduced in [6]. Beside the temporal fairness property, the CS scheme also features another very unique notion of fairness: the users are served when their channels are at their own best, regardless of whether they are strong or weak users. The user selections are independent of specific channel distributions making it particularly suited for HetNet environments where channel distributions vary widely among users [7]. In addition, this scheme lends itself to simple practical implementations [8] and much better analytical tractability for system performance analysis [9].

This research was supported in part by the UCSD Center for Wireless Communications and the NSF Grant CCF-1115645.

Despite all its salient features, the CS scheme is not designed to guarantee fairness on a short-term basis. Its fairness is only guaranteed in the long run. In a temporally correlated fading environment, such as a wireless system where users undergo correlated Rayleigh fading, the user selection decisions are correlated in time, leading to long delays between accesses (access starvation) for users who are in a weak fade of their channels. This issue is briefly addressed in the original work [6]. However, the authors examine only the case of a single fixed delay deadline where all users have the same access probability $w_k = 1/K$ and the user selections are independent. In our prior work [10], we address this delay issue with multiple deadlines using the Markov Decision Process (MDP) framework. This MDP approach, while effective in delay reduction, relaxes the temporal fairness guarantee. In this article, we introduce two delay control methods developed specifically for the CS scheme that not only provide a similar level of delay control effectiveness as the MDP approach but also guarantee the temporal fairness of user access.

2. CDF SCHEDULING BACKGROUND

In this section, we provide a brief summary of the CS policy and the original delay control approach to lay the groundwork for our proposed schemes. Consider a wireless system with K users sharing the same wireless channel. Time is divided into equal slots. Assume all the users experience independent fading. The scheduling decision is made every time slot. Let $X_k, k = 1 \dots K$, be the instantaneous SNR of user k. Let $U_k = F_{X_k}(X_k)$ be the CDF-transformed random variable for user k, where $F_{X_k}(x)$ is CDF of X_k. The CS policy in [6] grants the channel to the user according to $k^* = \underset{k}{\operatorname{argmax}}\, U_k^{(1/w_k)}$, where k^* is the index of the selected user and $\sum_{k=1}^{K} w_k = 1$. Each user k is guaranteed an access probability of w_k in the long run. In order to address the access starvation issue, the authors in [6] consider a single inter-service delay deadline t_{max} and identical user access time $w_k = 1/K$ and propose the following modifications to the original CS policy (a.k.a, starve-time-limited CS or CS-STL):

1. If there exists a user k such that its current delay $t_k = t_{max} - 1$, select user k.

2. Otherwise, follow the original CS policy.

This approach has two interesting features: it can enforce the inter-service delay to be no more than t_{max} and the long-termed user access probability remains $w_k = 1/K$. However, it does not address the case where users have different service delay requirements, which is necessary for practical systems. Nor does it have provisions for the case where the access probability w_k's are not identical. In addition, this approach does not consider "soft" deadlines where the users can tolerate certain fluctuation around the deadlines. In subsequent sections, we will introduce our CS-based policies that enable these new features.

3. HARD DEADLINE FORMULATION

In order to impose deadlines, we divide time into *Scheduling Frames* where frame length is F times of the time slot length. The scheduling frames must be long enough so that on average, each user is allocated at least one time slot per frame, i.e. $w_k F \geq 1$. Different deadlines $t_{k,max}$ can be assigned to different users. The scheme also supports users having no deadlines at all. The main idea is to enforce only expiring deadlines on each frame and select the remaining users opportunistically according to the CS policy. The proposed policy is described in table 1.

Table 1: Hard Deadline CS (CS-HD) Policy

1. At the beginning of each frame, get the set $\mathcal{E}$ of users whose deadlines will be expiring within the frame.
2. Mark off the slots where the deadlines will expire.
3. Let $E = |\mathcal{E}|$, the cardinality of set $\mathcal{E}$. Modify the scheduling weights according to the followings:
   (a) For expiring users: $\hat{w}_k = \frac{w_k F - 1}{F - E}$
   (b) For the rest of the users: $\hat{w}_k = \frac{w_k F}{F - E}$
4. Serve the expiring users at their corresponding marked slots.
5. For the remaining slots in the frame, select the user according the CS policy with the *modified weights* $\hat{w}_k$.

This delay control method addresses the two deficiencies of the CS-STL scheme: it allows both different deadlines and non-identical access probabilities to be used. Similar to the settings of CS-STL, we first consider the case where user selections are independent across time slots. In this case, this policy preserves the temporal fairness of the original CS policy as stated by proposition 1 below.

Proposition 1. *The hard-deadline CS (CS-HD) policy is temporally fair. That is, the average access probability for each user k is precisely w_k.*

Proof. For each scheduling frame, the average amount of resource allocated to a non-expiring user k is $(F-E) \times \frac{w_k F}{F-E} = w_k F$ time slots. For an expiring user, this number is $[1 + (F - E) \times \frac{w_k F - 1}{F - E}] = w_k F$ time slots as well. As a result, the average resource per time slot for all users is w_k. □

For correlated environments, user selections are not independent, which can cause some allocation errors due to the correlation between the user inter-service delays and the channel metrics in each scheduling frame. However, these allocation errors vanish as the number of users grows infinitely large as a direct result of theorem 1 below.

Theorem 1. *In a temporally correlated environment, when the number of users in the system grows infinitely large, the probability distribution of user metrics are independent of the user's past unselected events. That is,*

$$\lim_{K \to \infty} \Pr[X_{k,t} < x | k^*_{t-i} \neq k, i \geq 1] = \Pr[X_{k,t} < x]$$

Proof. This result is very intuitive: when the number of users is very large, the probability for user k not being selected in any time slot approaches unity, which is independent of the user metric. The proof is relegated to Appendix A. □

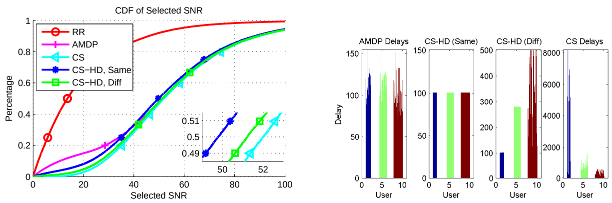
As a result of theorem 1, when the system grows larger, user distributions in each time slot become less dependent on the user inter-service delays. Thus, the allocation errors of the CS-HD policy become negligible for large systems.

Simulation 1. For performance evaluations, we simulated the CS-HD policy in a Rayleigh fading environment with the parameters listed in table 2. The Rayleigh channels follow the model described in [11]. The different amounts of channel temporal correlation are controlled via different user speeds.

Table 2: Simulation Parameters

Parameter	Value	
Number of users	10	K
Carrier frequency	1 GHz	f_c
Channel bandwidth	15 KHz	
Average channel SNR	30	ρ

In this simulation, the user speeds are selected to be in the linear set between 2 m/s and 30 m/s. The resource allocation weights are set to be in the linear set between 0.04 and 0.16. We simulate two different scenarios of delay deadlines. In the first scenario, all users having the same deadlines of $t_{k,max} = 100$ time slots (labeled "CS-HD, Same"). In the second scenario, the user delay deadlines vary linearly between 100 and 500 time slots (labeled "CS-HD, Diff"). Figures 1a and 1b illustrate the performance-delay tradeoff of the



a) Performance Tradeoff b) Delay Control

Fig. 1: Performance/Delay Tradeoff - CS-HD

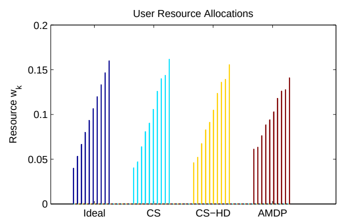


Fig. 2: Resource Allocation - CS-HD gives similar allocation as CS

CS-HD policy. The CS-HD policy can limit the inter-service delays for all users to be exactly less than or equal to maximum allowable deadlines with a small performance loss. As seen in figure 1a, the performance loss is smaller in the second scenario since the deadlines are less tight. Note that the CS-HD performance is comparable to that of the AMDP policy in [10] with much tighter delays. Figure 2 illustrates the ability of the CS-HD policy to provide accurate resource allocations. The CS-HD policy can achieve similar allocations as the original CS policy while the AMDP policy cannot as previously mentioned.

The hard deadlines lower multiuser diversity and thus reduce the system performance. When the user can tolerate some level of delay fluctuations (a.k.a soft deadlines), it is possible to gain back some multiuser diversity and improve the performance. This motivates the soft deadline scheme discussed in the next section.

4. SOFT DEADLINE FORMULATION

In order to motivate our formulation, we first consider a purely random selection scheme where the users are selected in a uniformly random manner. The selection is done via artificially generated i.i.d random metrics instead of user channel metrics. Let $\Pr[t_k > t_{k,max}]$ be the probability user k experiences a delay t_k that violates the deadline $t_{k,max}$ (a.k.a violation probability). Since the selection is based on i.i.d random data, no temporal correlation exists. Consequently, it can be seen intuitively that this scheme achieves the lowest delay deadline violation probabilities compared to any scheme that relies on temporal correlated metrics for user selection. The drawback of this random scheme is the lack of multiuser diversity exploitation, which significantly reduces the system performance. On the other hand, a channel-driven scheme such as the CS policy can experience much larger violation probabilities due to channel temporal correlations but achieve better system performance via multiuser diversity gain. This motivates us to consider a new policy as a combination of the purely random scheme and the CS policy. In this policy, we introduce a set of K virtual users whose metrics are generated as i.i.d random values. These users form a set of secondary participants in the selection competition. The proposed scheme is described in table 3.

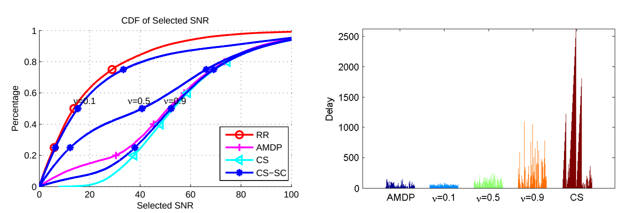
Table 3: Secondary Competition CDF Scheduling (CS-SC) Policy

1. For each user k, generate p_k from a uniform distribution on $[0,1]$. These represent virtual users.
2. Let $v_i = u_i^{1/(\nu_i w_i)}$ and $v_{K+i} = p_i^{1/[(1-\nu_i)w_i]}$ for $i = 1 \ldots K$, where $\nu_i \in [0,1]$.
3. Perform the selection $i^* = \underset{1 \le i \le 2K}{\operatorname{argmax}} \, v_i$
4. Grant the channel to user k if $i^* = k$ or $i^* = K + k$.

The mixing parameter ν_i controls tradeoff between the real and virtual users, which in turn controls the tradeoff between the user performance and deadline violation probability. Similar to the CS-HD policy, this policy also preserves the temporal fairness property as stated in proposition 2.

Proposition 2. *The Secondary Competition CS (CS-SC) policy is temporally fair. That is, the average access probability for each user k is precisely w_k.*

Proof. Since all users (real and virtual) have i.i.d metrics, according to the results in [6], the selection probability is $p_{k,r} = \nu_k w_k$ for the real user k and $p_{k,v} = (1-\nu_k)w_k$ for the corresponding virtual user $K+k$. The probability of a user k being selected at any time slot is the sum of these two probabilities: $p_k = p_{k,r} + p_{k,v} = w_k$. $\square$



a) Performance Tradeoff b) Delay Control

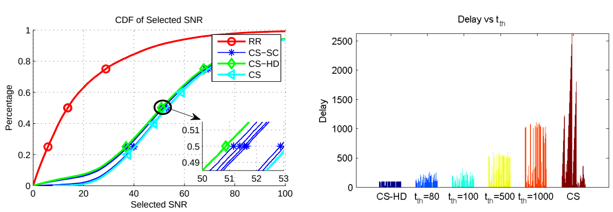
Fig. 3: Performance/Delay Tradeoff with a Fixed ν - CS-SC

Simulation 2. For performance evaluations, we simulate the CS-SC policy with the same parameters used in simulation 1 (table 2). For the ease of comparing between the proposed policies, the users all have the same speed $v_k = 2$ m/s, resource allocation weight $w_k = 0.1$ and mixing factor $\nu_k = \nu$. Figures 3a and 3b illustrate the performance-delay tradeoff for the CS-SC policy. As expected, lower values of ν_k yield better delay control but suffer higher performance loss. On the other hand, higher values of ν_k reduce the performance loss but the delay control is not as effective. Due to the use of fixed values for the mixing factors, the virtual users are present in all the resource competitions even when the user delays are low. This causes unnecessary performance loss, making delay reduction less effective. In order to improve the effectiveness of the delay control, we propose the following scheme (shown in table 4) for dynamically changing the values of ν_i in the CS-SC policy:

Table 4: Dynamic Parameters Adjustments for the CS-SC Policy

1. Set $\nu_i = 1$ when the user delay $t_i < t_{i,th}$.
2. Set $\nu_i = 0$ when the user delay $t_i \geq t_{i,th}$.

Here the thresholds $t_{i,th}$ are adjusted appropriately to achieve the desired violation probabilities or delay variances. With this parameter adjusting scheme, the delay control is much more effective. This improvement can be seen in figures 4a and 4b: the delays are at the same level as those in figure 3b (for $\nu = 0.5$ and $\nu = 0.9$) while the system performance is much better than those in figure 3a. Also the CS-SC policy can outperform the hard deadline CS-HD policy (figure 4a). Figure 5 shows the performance-delay variance tradeoff. The performance gets better as more delay variance is allowed. The CS policy has the highest variance.



a) Performance Tradeoff b) Delay Control

Fig. 4: Performance/Delay Tradeoff with Bimodal ν - CS-SC

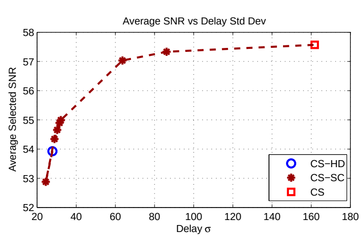


Fig. 5: Performance vs Delay Variance, CS-SC

It should be noted that the dynamic adjustments proposed in table 4 can cause some resource allocation errors in correlated fading environments. Similar to the case of the CS-HD policy, this effect is due to the correlation between the user inter-service delays and the channel metrics, which leads to the correlation between the adjustment of ν_i and the channel metrics. In the same fashion, this effect and thus the allocation errors vanish as the number of users in the system grows infinitely large due to the result of theorem 1.

5. CONCLUSIONS

In this article, we propose two high-performance delay control schemes that can effectively reduce the inter-service delays for the CDF-based scheduling policy. The first scheme, the CS-HD policy, enforces the hard deadlines for user delays, while the second scheme, the CS-SC policy, relaxes the deadline enforcement and achieves better performance. Mathematical performance analyses and parameter selections for these schemes will be considered in our future work.

A. PROOF OF THEOREM 1

Assume the user selection follows: $k^* = \underset{k}{\operatorname{argmax}}\, U_k$, where U_k is some utility function of X_k. For any $i \geq 1$, we have

$$\Pr[X_{k,t} < x | k^*_{t-i} \neq k] = \frac{\Pr[X_{k,t} < x; k^*_{t-i} \neq k]}{\Pr[k^*_{t-i} \neq k]},\ \text{where}$$

$$\begin{aligned}
&P_1 \triangleq \Pr[X_{k,t} < x; k^*_{t-i} \neq k] = \\
&= \Pr[X_{k,t} < x; U_{j,t-i} > U_{k,t-i}, \text{for some } j \neq k] \\
&\overset{(a)}{=} \int_0^\infty \Pr[X_{k,t} < x; U_{j,t-i} > y, \text{some } j \neq k | U_{k,t-i} = y] \\
&\qquad \times f_{U_{k,t-i}}(y) dy \\
&\overset{(b)}{=} \int_0^\infty \Pr[X_{k,t} < x | U_{k,t-i} = y] \\
&\qquad \Pr[X_{j,t-i} > y, \text{for some } j \neq k] f_{U_{k,t-i}}(y) dy \\
&= \int_0^\infty \Pr[X_{k,t} < x | U_{k,t-i} = y] \\
&\qquad \times (1 - \Pr[U_{j,t-i} < y, \forall j \neq k]) f_{U_{k,t-i}}(y) dy,
\end{aligned}$$

where (a) is from the rule of total probability, (b) from the product rule and the independence of user distributions. Also

$$\Pr[U_{j,t-i} < y, \forall j \neq k] = \prod_{j \neq k} F_{U_{j,t-i}}(y)$$

Since $F_{U_{j,t-i}}(y) \in [0, 1)$, $\forall y \in [0, \infty)$, we have

$$\begin{aligned}
&\lim_{K \to \infty} \Pr[U_{j,t-i} < y, \forall j \neq k] \\
&= \lim_{K \to \infty} \prod_{j=1, j \neq k}^{K} F_{U_{j,t-i}}(y) = 0. \qquad (1)
\end{aligned}$$

$$\begin{aligned}
&\therefore \lim_{K \to \infty} P_1 = \int_0^\infty \Pr[X_{k,t} < x | U_{k,t-i} = y] f_{U_{k,t-i}}(y) dy \\
&= \Pr[X_{k,t} < x]. \qquad (2)
\end{aligned}$$

Additionally, $P_2 \triangleq \Pr[k^*_{t-i} \neq k] = 1 - \Pr[k^*_{t-i} = k]$

$$= 1 - \int_0^\infty \Pr[U_{j,t-i} < y, \forall j \neq k] f_{U_{k,t-i}}(y) dy.$$

Thus, $\lim_{K \to \infty} P_2 = 1$ (from (1)). Finally, we get

$$\lim_{K \to \infty} \Pr[X_{k,t} < x | k^*_{t-i} \neq k] = \lim_{K \to \infty} \frac{P_1}{P_2} = \Pr[X_{k,t} < x].$$

B. REFERENCES

[1] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," *IEEE Vehicular Technology Conference Proceedings*, vol. 3, pp. 1854 – 1858, 2000.

[2] E. Chaponniere, P. Black, J. Holtzman, and D. Tse, "Transmitter directed code division multiple access system using path diversity to equitably maximize throughput," *US Patent 6,449,490*, 2002.

[3] Xin Liu, E. K P Chong, and N.B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 10, pp. 2053–2064, Oct 2001.

[4] S.S. Kulkarni and C. Rosenberg, "Opportunistic scheduling policies for wireless systems with short term fairness constraints," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, Dec 2003, vol. 1, pp. 533–537 Vol.1.

[5] S. Patil and G. de Veciana, "Managing resources and quality of service in heterogeneous wireless systems exploiting opportunism," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 5, pp. 1046–1058, Oct 2007.

[6] Daeyoung Park, Hanbyul Seo, Hojoong Kwon, and Byeong Gi Lee, "Wireless packet scheduling based on the cumulative distribution function of user transmission rates," *Communications, IEEE Transactions on*, vol. 53, no. 11, pp. 1919–1929, 2005.

[7] P. Nguyen and B. Rao, "Throughput improvements for cellular systems with device-to-device communications," in *Signals, Systems and Computers, 2013 Asilomar Conference on*, Nov 2013, pp. 1973–1977.

[8] AH. Nguyen, Yichao Huang, and B.D. Rao, "Learning methods for CDF scheduling in multiuser heterogeneous systems," *Signal Processing, IEEE Transactions on*, vol. 62, no. 15, pp. 3727–3740, Aug 2014.

[9] Yichao Huang and Bhaskar D. Rao, "Random beamforming with heterogeneous users and selective feedback: Individual sum rate and individual scaling laws," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 5, pp. 2080–2090, May 2013.

[10] P.C. Nguyen and B.D. Rao, "Delay control for CDF scheduling using Markov Decision Process," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 3469–3473.

[11] H. S. Wang and N. Moayeri, "Finite-state markov channel - a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 163 – 171, Feb 1995.