Performance of greedy policies for downlink scheduling in networks with relay stations

Seung Jun Baek\textsuperscript{1a)} and JaeSheung Shin\textsuperscript{2}

\textsuperscript{1} College of Information and Communications, Korea University, 5–1, Seongbuk-gu, Anam-dong, Seoul, 131–701, Korea  
\textsuperscript{2} Electronics and Telecommunications Research Institute,  
138 Gajeongno, Yuseong-gu, Daejeon, 305–700, Korea  
a) sjbaek@korea.ac.kr

Abstract: We investigate a greedy policy for delay-optimal scheduling in broadband wireless networks using relay stations (RS). The key idea behind the policy is that users should greedily exploit opportunism from time varying channels, while the basestation should prioritize balancing queues at RS. A fluid model is presented to show that such policy is optimal in the fluid regime. By simulation we show that the proposed scheme reduces mean delay by up to 73\% compared to backpressure algorithm.

Keywords: heterogeneous networks, relay station, scheduling, optimization

Classification: Science and engineering for electronics

References


1 Introduction

In LTE Advanced, heterogeneous network architecture such as PicoCell, FemtoCells and relay stations (RS) have been proposed to overcome the limitation of cellular networks having single basestation (BS) [1]. In this letter we study downlink scheduling problems of minimizing mean delay in networks using RS. In networks with RS there exists an Over-The-Air (OTA) backhaul (‘backhaul’ in short) between RS and BS for relaying packets destined to users. We consider a case such that users are associated only with RS but not with BS. Backhauls typically have higher capacity than wireless links from RS to users (‘RS links’ in short) due to planned deployment of RS, e.g., RS may be in line-of-sight to BS. We assume that RS operates at half-duplex, i.e., only one link among backhaul and RS links may be activated at any instant. The channels are time-varying, thus there exists a set of jointly achievable long-term rates among the backhaul and RS links. BS and RS have a separate set of queues, one for each user both at BS and RS, which store packets to be transmitted in the downlink: see Fig. 1.

Fig. 1. Downlink packet scheduling with a relay station.

How to design a delay-optimal scheduler for such multi-hop queuing networks with time-varying service rates is an open question. Even for single-hop multi-user case, e.g., systems with single BS, only a limited number of delay-optimal policies are known, e.g., Longest Connected Queue (LCQ) [2] and Longest Queue Highest Possible Rate (LQHPR) [3]. These schemes are greedy in that they try to maximize the number of nonempty queues in order to maintain gains from multi-user diversity (MUD). However in our case if MUD has to be compromised by backhaul, it is no longer clear whether such greedy policy is optimal. In this paper we investigate a greedy scheduling policy for symmetric systems to partly answer the question. We propose a scheme called Longest Sum-queue Highest Possible Rate – Balance Queues with Lowest Possible Rate (LSHPR–BQLPR). The key idea is that, since the channel quality of backhaul is more likely to be in favorable condition than that of RS links due to planned deployment of RS networks, one should prioritize opportunism at RS links. We investigate the policy using a recently developed technique called fluid model [4]. We show that an analogous policy to the proposed one is optimal for the fluid model. Even for packetized networks we demonstrate that the policy reduces mean delay by up to 73% relative to the well-known backpressure (BP) policy [5, 6] by simulation. The significance of this work is showing that, by exploiting architectural proper-
ties of multi-hop networks, one can reap further performance gains. Thus the proposed scheme may aid future design of scheduling algorithms for multi-hop networks with similar topology.

2 Fluid model

Suppose there are \( n \) users in the system, thus there exist \( n \) queues at BS and RS respectively. Let \( I = \{1, 2, \ldots, n\} \) be the set of user indices. The arrival of data bits at BS is stationary and the arrival rates among users are assumed to be identical. By \( \mathcal{C} \in \mathbb{R}^{n+1} \) we denote the set of long-term rates jointly achievable by backhaul and RS links where \( \mathcal{C} \) is assumed to be a convex polyhedron. For some \( v \in \mathcal{C} \), \( v_i \) (the \( i \)-th entry of \( v \)) represents the rate achievable by \( i \)-th user (RS link), and \( v_{n+1} \) for that at backhaul. We assume \( \mathcal{C} \) is symmetric among RS links however not necessarily is so between backhaul and an RS link. A rate vector \( v \in \mathcal{C} \) is called maximal if for any \( a \in \mathcal{C} \), \( a \succeq v \) implies \( a = v \) where \( \succeq \) denotes the entrywise inequality between two vectors. Let \( \mathcal{M} \) denote the set of maximal vectors in \( \mathcal{C} \): see Fig. 2 (a).

Consider a time-slotted system and at time \( t \geq 0 \), let us denote the queue lengths, or the number of untransmitted data bits, at BS and RS for User \( i \) by \( L_i(t) \) and \( M_i(t) \) respectively. The following is the proposed policy for rate allocation to backhaul and RS links.

LSHPR–BQLPR Policy:

1. Sort the users in the order of \( L_j(t) + M_j(t) \). Let \( \mathcal{P} \) be a permutation of \( I \) such that user \( \mathcal{P}(i) \) is currently \( i \)-th largest among \( L_j(t) + M_j(t) \), \( j = 1, \ldots, n \).

2. Define sets \( V^{(i)} \subset \mathbb{R}^{n+1} \) recursively by \( V^{(i)} = \{ \hat{v} | \hat{v} = \arg\max_{w \in V^{(i-1)}} [a \cdot \mathbf{e}_{\mathcal{P}(i)}] \} \) for \( i = 1, \ldots, n \), and \( V^{(0)} = \mathcal{M} \) where \( \mathbf{e}_k \in \mathbb{R}^{n+1} \) denotes a unit vector whose \( k \)-th entry is 1. Let \( v \) be the element of \( V^{(n)} \). Note that such choice of \( v \) allocates the lowest possible rate to \( v_{n+1} \) from \( \mathcal{M} \).
3. Allocate $v_i$ to $i$-th RS link (User $i$) for $i \in I$ and allocate $v_{n+1}$ to backhaul.

4. BS transmits data to the shortest queue at RS. Specifically let $J = \{j|L_j(t) > 0\}$, then at rate $v_{n+1}$ BS transmits data for User $j^* = \arg\min_{j \in J} [M_j(t)]$.

The rationale behind the policy is as follows. Firstly, since the arrival and service statistics are symmetric, it is important to balance the combined queues of BS and RS. Thus we allocate higher rates to users with longer sum of RS and BS queue lengths. Secondly, BS should transmit data to the shortest queue in RS in order to keep as many queues busy at RS so as not to lose MUD gain at RS links. The third point is subtle: the policy assigns highest possible rate to user with the longest sum-queue however allocates lowest possible rate to the backhaul – is this desirable? If so, under what condition? We use a fluid model to partly answer that question.

Fluid model is a relaxation of optimal control of stochastic queuing networks as deterministic optimization problems [4]. The model treats data bits as a continuum of fluid and aims at finding an optimal policy to ‘drain’ fluid from the system. Note stochastic control problems are hard to solve, even numerically. However by using fluid model we often obtain a good approximation to the true optimal solution [4]. We show that an analogous policy to the proposed one is optimal in the fluid model. A formulation of fluid model is given as follows. From now on we assume $n = 2$ for simplicity, while the arguments hereafter are readily extended to the case for larger $n$. Let us denote the amount of fluid at User $i$’s BS queue (resp. RS queue) at time $t \geq 0$ by $l_i(t)$ (resp. $m_i(t)$) which are called trajectories. Let $x(t) \triangleq (l_1(t), l_2(t), m_1(t), m_2(t))$ which is called the state of the system. We reuse the notation $C$ as set of joint service rates for fluid queues, i.e., a vector $v \in C$ is a set of rates at which fluid is drained from the queues. A control is defined to be allocating service rates from $C$ depending on the system state.

Fluid arrives at BS queues at a constant rate of $\lambda$: see Fig. 2 (b). The problem of minimizing the sum queue length, which is equivalent to minimizing mean delay in stationary systems by Little’s law, can be cast into an optimal draining problem given an initial state $x(0)$ as follows.

**Fluid problem (FP):**

\[
\text{Minimize:} \quad \int_0^\infty \sum_{i=1}^2 \{l_i(t) + m_i(t)\} \, dt \tag{1}
\]

Subject to:

\[
\frac{dl_i(t)}{dt} = \lambda - u_i(x(t)), \quad \text{for } i = 1, 2, \tag{2}
\]

\[
\frac{dm_i(t)}{dt} = u_i(x(t)) - w_i(x(t)), \quad \text{for } i = 1, 2, \tag{3}
\]

\[
x(0) = (l_0^1, l_0^2, m_0^1, m_0^2), \quad x(t) \in \mathbb{R}_+^4 \tag{4}
\]

where the control $u_i(\cdot): \mathbb{R}_+^4 \to \mathbb{R}_+$ (resp. $w_i(\cdot)$) maps a state to a service rate at User $i$’s BS queue (resp. RS queue). We define a policy as $\pi(\cdot): \mathbb{R}_+^4 \to C$
such that $\pi(\cdot) = (w_1(\cdot), w_2(\cdot), u_1(\cdot) + u_2(\cdot))$. The condition (2) (resp. (3)) represents the net rate of change of User $i$’s BS queue (resp. RS queue) length. (4) represents the amounts of fluid initially present in the system. In general the optimal policy for FP depends on the initial condition (4). We show that, however, the proposed policy is always optimal for FP under the following condition: for any $a, v \in \mathcal{M}$, we have that $a_{n+1} > \max_{i \in [1]} [v_i]$. We denote this condition by (C1). It states that the minimum rate achievable at backhaul in $\mathcal{M}$ is greater than the maximum rate achievable for any individual RS link.

The underlying assumption is that the rate associated with backhaul tends to be substantially higher than the rates offered to users.

**Theorem 1** Under (C1), LSHPR–BQLPR policy is optimal for FP.

*Proof:* Let us denote $a \land b = \min(a, b)$. Under our policy define the time $T$ such that

$$T = \inf \{ t | m_1(t) \land m_2(t) = 0 \}. \quad (5)$$

We show that $l_1(T) \land l_2(T) = 0$ by contradiction. Without loss of generality (WLOG) assume $m_1(T) < m_2(T)$. Suppose $m_1(T) = 0$ but $l_1(T) > 0$. Since fluid trajectories are continuous [4], there exists small $\epsilon > 0$ such that, at $\tau = T - \epsilon$, $m_1(t) < m_2(t)$ and $l_1(t) > 0$ for all $t \in [\tau, T]$. This means that BS transmits fluid to queue 1 in RS during $[\tau, T]$ at the lowest possible rate from $\mathcal{M}$ (BQLPR). Let us denote this rate by $r$, or $u_1(x(t)) = r$, $\forall t \in [\tau, T]$. However for any rate that is assigned to drain User 1’s queue at RS, say $\rho(t)$, we have that, from (3) and due to (C1),

$$\frac{dm_1(t)}{dt} = u_1(x(t)) - w_1(x(t)) = r - \rho(t) > 0, \quad \forall t \in [\tau, T]. \quad (6)$$

From (6) we have that

$$m_1(T) = \int_\tau^T \frac{dm_1(t)}{dt} dt + m_1(\tau) > m_1(\tau), \quad (7)$$

however this contradicts the assumption $m_1(T) = 0$. This implies that at any time $m_1(t) = 0$ implies $l_1(t) = 0$ or equivalently $l_1(t) + m_1(t) = 0$. Thus $T$ is also given by

$$T = \inf \{ t | [l_1(t) + m_1(t)] \land [l_2(t) + m_2(t)] = 0 \}. \quad (8)$$

Namely is $T$ also the minimum time such that one of the sum-queue becomes empty. Let us define $s_i(t) = l_i(t) + m_i(t)$ for $i = 1, 2$, and WLOG assume $s_1(0) \geq s_2(0)$. We show that $\forall t \in [0, T]$, $s_1(t) \geq s_2(t)$ as follows. Suppose there exists $\hat{t} = \inf \{ t \in [0, T] | s_1(t) = s_2(t) \}$. Our policy assigns highest possible rate to User 1, $\forall t \in [0, \hat{t})$, e.g., $\pi(x(t)) = \theta$ in Fig. 2 (a). At $t = \hat{t}$, $s_1(t) = s_2(t)$, and from this time on, the policy will alternate between two RS queues in time-sharing manner to allocate highest possible rates, e.g., $\pi(x(t)) = (\theta + v)/2$ in Fig. 2 (a), $\forall t \in [\hat{t}, T]$. Thus $s_1(t) = s_2(t)$, $\forall t \in [\hat{t}, T]$. If such $\hat{t}$ does not exist, we have that $s_1(t) > s_2(t)$, $\forall t \in [0, T]$. In summary
User 1’s queue at RS is continuously served at the highest possible rate either until \( t = T \), i.e., the entire fluid for User 2 is drained from the system, or until sum-queue of User 1 and 2 become equal. Now let us take a different view of the system as if BS and RS queues for each user are ‘merged’ into a single queue which has the trajectory \( s_i(t), i = 1, 2 \). As summarized above the trajectory \( s_i(t) \) evolves as if such merged queues are jointly served under LQHPR policy [3]. Since LQHPR is proven to minimize \( s_1(t) + s_2(t), \forall t \geq 0, \) and \( s_1(t) + s_2(t) = \sum_{i=1}^{2} [l_i(t) + m_i(t)] \), our policy is also optimal for FP. ■

In the next section we apply the proposed policy to packetized networks.

3 Simulation

In our simulations we consider a time-slotted discrete time system. At each time slot a packet of unit size arrives for each user with probability \( p_a \). The number of packets that can be served for each time slot on the backhaul is according to a random variable \( R_r \) where \( R_r = 6 \) with probability (w.p.) \( p_r \) and \( R_r = 4 \) w.p. \( 1 - p_r \). The number of packets that can be served by RS links is given by a random variable \( R_u \) where \( R_u = 2 \) w.p. \( p_u \) and \( R_u = 1 \) w.p. \( 1 - p_u \). Note these parameters account for the superior channel quality of backhaul relative to that of RS links. All arrival and channel statistics are mutually independent. At any time only one of the links can be activated. Clearly the long-term rates achievable by users from this channel form a convex polyhedron. Note that the channels are in either ‘good’ (w.p. \( p_r \)) or ‘bad’ (w.p. \( 1 - p_r \)) state. The following describes the proposed policy applied to this system. If any of RS links are in good state, we serve the user with the longest sum-queue among such users with good channels. This is LSHPR part of the proposed policy. If all of RS links are in bad state and the backhaul link is in good state, we relay packets via backhaul to the shortest queue at RS, which corresponds to BQLPR. If all the links are in bad state, we use a heuristic: we try to improve the overall balance between BS and RS queues. Specifically if the sum of the queue lengths over all users at RS is greater than that at BS, the queue at RS of the user with longest sum-queue is served. Otherwise the backhaul link is activated and BQLPR is performed.

Fig. 3 shows a plot of the mean number of packets in the system which is proportional to mean delay by Little’s law, versus normalized arrival rates for \( n = 4 \) and 8 respectively. The normalized arrival rate \( \alpha \) is defined such that, the probability of a packet arrival for each user is given by \( \alpha/n \). There are two sets of curves for each plot: those grouped by ‘Ch1’ (resp. ‘Ch2’) correspond to the case where \( p_u = 0.3 \) (resp. \( p_u = 0.7 \)). Namely the RS links for ‘Ch2’ is more favorable to users than ‘Ch1’. We set \( p_r = 0.9 \) for the backhaul. In all cases the proposed scheme outperforms BP algorithm. For \( n = 4 \) and 8, the reduction in mean delay relative to BP algorithm are 50–66% and 65–73% respectively. Note that the overall performance gain relative to BP algorithm increases with \( n \). This is due to the fact that the average sum-rate of RS links increases with \( n \), i.e., we obtain higher MUD.
gain. Since the proposed scheme is more aggressive in exploiting MUD gain, we observe the greater reduction in delay with higher MUD gain, compared to BP algorithm. We also see that the proposed scheme performs increasingly better than BP policy with increasing arrival rates for fixed $n$. This can be explained as follows. As the arrival rate increases, more RS links become busy on average, which also causes an increase in MUD gain. Thus we again observe that such queuing dynamics has a favorable effect on the proposed policy.

4 Conclusion

We studied the delay-optimal joint scheduling of BS and RS by establishing a queuing model in networks with relay stations. We observed that it is crucial to emphasize opportunism among users at RS while the backhaul should focus on balancing the queues at RS so that multi-user diversity is preserved. The argument has been supported by proving the optimality in the analogous fluid model. The simulation results show that the proposed policy leads to significant reduction in mean delay.

Fig. 3. Mean delay performance of the proposed policy.