Relay Scheduling and Cooperative Diversity for Delay-Sensitive and Bursty Traffic

Somsak Kittipiyakul and Tara Javidi

***Abstract*— This paper considers a problem of scheduling and cluster size with N i.i.d. source nodes trying to transmit data to a common destination node (e.g. a gateway or a processing center), with the help of some subset of nodes in the form of relays. The source nodes share the network in a TDM manner, using a round-robin scheduling scheme. There are two causes of bit errors in the system: channel decoding and delay violation. We are interested in the optimal negative SNR exponent, i.e., the asymptotic decay rate, of the total probability of bit error. In finding the SNR exponent, we optimize the channel transmission rate as well as the number of cooperative relay nodes.**

I. Introduction

We consider a cooperative wireless network consisting of multiple nodes, each with an independent information source, and a common destination-node. We are interested in a cross-layer queue-channel optimization problem for bursty and delay-sensitive information sources. Each node is a bursty source of information bits concatenated with an infinite buffer and a constant-rate quasi-static relay channel to the common destination. More specifically, we consider a network where, at any particular time, one node's bits are transmitted to the destination, with the help of other nodes in the form of relays. The cross-layer performance metric of interest is the total bit loss probability where loss can be due to decoding errors as well as delay violation.

It is well-known that cooperation among nodes in a slow-fading wireless network can substantially improve the reliability of communication [2], [3]. Although there are many ways to cooperate in the network with multiple sources (see e.g. [4], [5]), we consider a simple time-sharing cooperation scheme among the information-source nodes. At each time, only one node is an information source and some of the other (pre-assigned) nodes act as relays to help transmission of the source node. We call this set of an information source and its relay nodes as a *cooperative cluster*. The choice of which information source (and its cooperative cluster) is active at any particular time is determined by a scheduler. In this paper, we consider a simple round-robin scheduler where each node periodically becomes the information source.

In each cooperative cluster, the improvement due to cooperation among the source and relays, known as *cooperative diversity* gain [2], is a result of encoding across independent spatial channels. It is important to note that the cooperative diversity gain is fully achieved only with coding the information over some minimum number of channel uses. This required coding time is a monotone increasing function of the number of cooperative relay nodes. In cases where delay requirements (in units of channel uses) are much larger than the total number of relays, the cooperative diversity gain improves the system performance by decreasing the probability of channel outage. On the other hand, when the bit-arrival processes are stochastic and bursty and the bits have a strict delay requirement which is of the same order as the size of the network, the required coding time results in an increase in the end-end delay any bit faces. This potentially can increase the probability of delay violation. In such a setting, increasing cooperative diversity might or might not be desirable. In this paper we are interested in finding the optimal values of the cooperative cluster size, as well as the transmission rate of the relay channels, so that the total bit loss probability is minimized.

Since it is difficult to derive the exact relationship between the interested parameters and the probabilities of channel outage and the delay violation, we choose to study an asymptotic approximation when the signal-to-noise ratio (SNR) is asymptotically high. The first advantage of this choice is that there exists an asymptotic high-SNR analysis for the channel outage probability, known as the *diversity-multiplexing-tradeoff* (DMT) analysis [1]. Another advantage of a high SNR analysis is that, with a proper scaling of the source statistics with SNR, we can derive an asymptotic approximation of the delay violation probability that is valid even when the delay requirement is finite.[1] Having the asymptotic expressions for the probabilities of channel outage and delay violation enables the main contribution of this paper: the formulation of a cross-layer optimal operating point for cooperative wireless network with multiple bursty sources and delay constraints under the static round-robin scheduler.

This paper is a part of an on-going research in jointly considering channels and queues for delay-sensitive data over wireless channels (see [13] for references). This paper is an extension of our previous work in [12], where we studied a cooperative network with a single source at high-SNR. The high-SNR approximation in this work was motivated by a cross-layer study in [9], where the DMT result [1] was first used for a study of joint source-channel optimization.

The remainder of the paper is organized as follows. In Section II, we provide mathematical models for the bit-arrival

The authors are with the ECE Department, University of California - San Diego, La Jolla, CA, 92093. This research is supported in part by the Center for Wireless Communications, UCSD and UC Discovery Grant No. Com04-10176, ARO-MURI Grant No. W911NF-04-1-0224, NSF CAREER Award No. CNS-0533035, and AFOSR Grant No. FA9550-05-01-0430.

[1] This derivation is based on the *many-flow large-deviations* results (see e.g. [6]–[8]).

process, the cooperative communication with an amplify-and-forward protocol, and the round-robin scheduler. Section III describes the asymptotic performance measures of the system and provides the existing high-SNR asymptotic result, also known as DMT, for the channel outage in cooperative relay channel. Section IV gives the asymptotic probability of delay violation under the round-robin scheduler. Section V presents the main result of the paper where we derive the optimal transmission rate and the cooperative cluster size that minimize the asymptotic total bit loss probability. A discussion of the results is also provided. Section VI concludes the paper. The proofs are given in the appendices.

For the rest of the paper, we use the notation $\doteq$ corresponding to the exponential equality, i.e. $y \doteq \rho^x$ is equivalent to $\lim_{\rho\to\infty} \frac{\log(y)}{\log(\rho)} = x$. We refer to the set $\{1, 2, \ldots, N\}$ as $\mathcal{S}$ and use $[q]_a^b := \min\{\max\{a, q\}, b\}$ when $a < b$.

II. System Model and Problem Formulation

We consider a cooperative (uplink) network consists of N nodes, denoted by 1 to N, and a common destination node E, shown in Figure 1. The system is time-slotted into discrete *timeslots*, for which all nodes' transmissions are assumed to be synchronized. At each timeslot, each source-node receives information according to a stochastic process and stores the bits that cannot be sent immediately in a buffer which is assumed to be infinite.

The delay requirement asks that each bit of information be decoded at the destination-node E within a *maximum allowable delay* of D time-slots from the time it arrives at its source-node. Otherwise, the bit will be obsolete, discarded, and counted as erroneous. We assume no retransmission of unsuccessful transmissions.

Next we provide more details and notations for a particular cooperative communication scheme of interest, as well as the arrival processes. In addition, we define the performance measure.

A. Cooperative Communication with OAF Protocol and Relay Scheduling

Communication takes place in the presence of additive receiver noise, and in the presence of spatially independent and identically distributed quasi-static fading. We assume complete knowledge of the fading channels at the receiver of the final destination, and no knowledge of the fading at the receivers of the assisting relays. Each node has a single receive and transmit antenna, operating in half-duplex. We assume, without loss of generality, that a timeslot contains one channel use.

The nodes cooperate within units of T consecutive timeslots, called *cooperation frame* [5]. Without loss of generality, we assume cooperation frames start at time mT, $m \in \mathbb{Z}$. Within any cooperation frame, we have a relay channel, where only one node is an information source and the other $n-1$ nodes (out of $N-1$ nodes) are relays. The assignment of relays to source is fixed. We denote the set of an information source and its relays as a cooperative cluster. Every cooperative cluster has a fixed size of n.

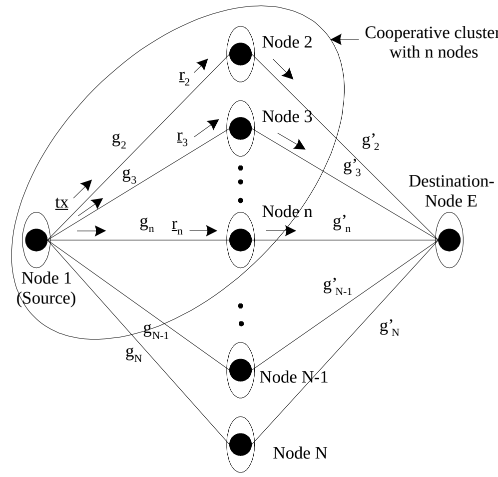


Fig. 1. Cooperative network with a common destination-node E and N source nodes. At this snapshot, the n-node cooperative cluster consists of node 1 as the information source and nodes 2 to n as the relays.

Figure 1 illustrates the network at a particular cooperation frame where the active cooperative cluster size consists of node 1 as the information source and nodes 2 to n as the relay nodes. At the beginning of each cooperation frame, a node is selected as an information source, according to a scheduling scheme. In other words, a cooperative cluster is selected to be active during each cooperation frame.

Within a cooperation frame, the source node and the relay nodes in the selected cooperative cluster cooperate using the symmetric and minimum-delay, orthogonal amplify-and-forward (OAF) protocol [4]. In the symmetric OAF protocol, an information-source node transmits for a duration of $T/2$ timeslot (broadcast phase of the round, duration $T/2$), and, after the broadcast phase, all $n-1$ relay nodes in the cooperative cluster start forwarding the vectors they have received (cooperative-relay phase, duration $T/2$). The source node remains silent during the second phase. We focus on the minimum-delay version of this protocol [11], where the cooperation frame duration is $T = 2n$.

When a node is selected as the information source, exactly RT oldest bits at the head of its buffer are removed.[2] Thus, within each T-timeslot cooperation frame, an amount of RT bits are transmitted to the destination. The average information transmission rate R bits per channel use or timeslot (bpcu) and the cluster size n are fixed at all time.

In particular, the statistics of the sources and channels, as well as the value of the delay requirement D, are assumed to be known and fixed for all time. The statistical descriptions of the channel and the source are used to arrive at the appropriate network parameters. The choice of R as well as n (or T which is equal to $2n$) will be the outcome of an optimization of the overall system performance.

[2] If the buffer contains less than RT bits, null bits are used. It is easy to show that the use of null-bits does not impact the asymptotic performance measure of interest.

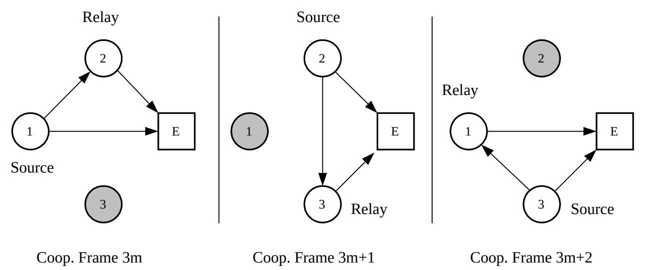


Fig. 2. An example illustrating cooperative cluster of size $n = 2$ for the network with $N = 3$ source nodes, under the round-robin scheduler. Different cluster is active at different time, e.g. during cooperation frame $3m$, where m is an integer, the active cooperative cluster is composed of node 1 as the information source and node 2 as the only relay.

B. Bit-Arrival Processes

For each node $k \in \mathcal{S}$, we consider a family of bit arrival processes $(A^{k,\rho} = (A_t^{k,\rho}, t \in \mathbb{Z}), \rho \in \mathbb{N})$, indexed by SNR ρ, where $A_t^{k,\rho}$ is the number of bits arriving in time-slot $t \in \mathbb{Z}$. Notice that indexing our arrival process by SNR ρ is in anticipation of our high-SNR asymptotic analysis. We assume $A_t^{k,\rho}$ to be i.i.d. over time and nodes. In this paper, we consider $A_t^{k,\rho}$ to be a compound Poisson source with exponential packet size, denoted by CPE(λ, μ, ρ). Specifically, we consider

$$A_t^{k,\rho} = \sum_{i=1}^{N_t^{k,\rho}} Y_{i,t},$$

where $N_t^{k,\rho}$ is the number of packets that have arrived over the t^{th} time-slot and $Y_{i,t}$ is the size of packet i. $N_t^{k,\rho}$ is drawn from an independent Poisson distribution with mean of $\mu\lambda \log \rho$, and $Y_{i,t}$ is drawn independently from an exponential distribution with mean of $1/\mu$ (independent of ρ) bits per packet. This compound Poisson random variable $A_t^{k,\rho}$ has a log moment generation function of

$$\Lambda_A(\theta) := \log E[e^{\theta A_1^{k,\rho}}] = \frac{\theta\mu\lambda \log \rho}{\mu - \theta}, \text{ for } \theta < \mu. \quad (1)$$

In particular, the average bit arrival is

$$\mathbb{E}\{A_t^{k,\rho}\} = \lambda \log \rho \qquad \text{(bits per time-slot).}$$

C. Source and Relay Scheduling

We consider a simple round-robin (RR) scheduler, also known as static time-division-multiplexing (TDM), according to an arbitrary ordering of nodes. Without loss of generality, we assume an increasing ordering of nodes 1 to N. In other words, the cooperation frame when node $k \in \mathcal{S}$ is the information source starts at time $(mN+k-1)T$, $m \in \mathbb{Z}$. Figure 2 illustrates an example of cooperative clusters, which are active at different cooperation frames, according to the RR scheduler.

D. Performance Measures and System Objective

The overall performance measure is the total probability of bit loss, P_{tot}, where loss can occur due to channel decoding error or the end-to-end delay violation, i.e.,

$$P_{\text{tot}} := P_{\text{ch}} + (1 - P_{\text{ch}})P_{\text{dv}}. \quad (2)$$

where P_{ch} denotes the probability of channel decoding error and P_{dv} denotes the probability of delay violation.

III. High-SNR Asymptotic Analysis

Since it is difficult to derive the exact relation between the probabilities of channel error and delay violation and the interested parameters, we choose to study an asymptotic approximation when SNR is asymptotically large. The benefit of this choice is that we can use the known results in information theory, known as DMT, to describe the performance of a communication scheme. In particular, we have

a) Asymptotic Performance Measure for Decoding Errors: Given a choice of transmission rate R and cluster size n, the performance of the decoder at the receiver of the destination node is given by the probability P_e of codeword decoding error and is described in terms of the DMT [1]. Given that the information-source node operates at rate R bpcu and at the average received SNR ρ, the performance is described in terms of the *diversity gain*

$$d_{\text{ch}}(r, n) := -\lim_{\rho\to\infty} \frac{\log P_e(\rho, r, n)}{\log \rho} \quad (3)$$

as a function of the channel *multiplexing gain*

$$r := R/\log(\rho).$$

The DMT analysis of the 3-node OAF protocol with large and equal durations of the first and second phases, was first provided in [4]. For the $(n + 1)$-node $(n \geq 2)$ symmetric minimum-delay $(T = 2n)$ OAF protocol, the negative SNR exponent of the probability of codeword error, which coincides with the negative SNR exponent of the *probability of bit error*, $P_{\text{ch}}(r, n)$, is given in [11] as

$$d_{\text{ch}}(r, n) = \begin{cases} n(1 - 2r), & 0 \leq r \leq \frac{1}{2}, \\ 0, & r > \frac{1}{2}. \end{cases} \quad (4)$$

This together with (3) gives

$$P_{\text{ch}}(r, n) \doteq \rho^{-n(1-2r)}, \text{ for } 0 \leq r \leq \frac{1}{2}. \quad (5)$$

Having identified P_{ch}, we next discuss P_{dv}.

b) Asymptotic Performance Measure for Delay Violations: The delay violation probability P_{dv} for the channel multiplexing rate r and the cluster size n is defined as

$$P_{\text{dv}}(r, n) := P[\text{steady-state end-to-end delay} > D], \quad (6)$$

where the end-to-end delay is the time interval between a bit's arrival to a node and its departure from the decoder at the destination node.

Note that given a fixed multiplexing rate r, the total traffic load, i.e.,

$$\frac{NE[A_1^{1,\rho}]}{R} = \frac{N\lambda \log \rho}{r \log \rho} = \frac{N\lambda}{r},$$

is kept independent of ρ. To ensure stability and the existence of the steady-state probabilities, we assume that

$$N\lambda < r_{\max} := 1/2.$$

As we will see, we use large-deviations techniques to arrive at the asymptotic formula for P_{dv}. We will show in Theorem 1 that

$$P_{\text{dv}} \doteq \rho^{-I^{\text{RR}}(r,n,\lambda,\mu,N,D)},$$

where I^{RR} is a good rate function describing the negative SNR exponent of P_{dv} and is closely related to the log moment generating function Λ_A.

c) Asymptotic Overall Performance Measure: Having identified the two elements affecting the overall performance, we proceed to describe our objective precisely: By symmetry, eqn. (2) in the asymptotic high-SNR regime becomes

$$P_{\text{tot}}(r,n) \doteq \Pr[\text{bit error for user 1}] \doteq P_{\text{ch}}(r,n) + P_{\text{dv}}(r,n), \tag{7}$$

where

$$P_{\text{ch}}(r,n) \quad := \quad \Pr[\text{channel decoding error of user 1}] \tag{8}$$
$$P_{\text{dv}}(r,n) \quad := \quad \Pr[\text{delay violation of user 1}] \tag{9}$$

We are interested in finding the optimal values of channel multiplexing rate r and the cooperative cluster size n, as a function of the parameters λ, μ, N, and D, which maximize the negative SNR exponent of the overall probability of bit loss. As a result, we have the optimal negative SNR exponent of $P_{\text{tot}}(r,n)$ as

$$d^* = \max_{r,n} \lim_{\rho\to\infty} \frac{-\log P_{\text{tot}}(r,n)}{\log \rho}, \tag{10}$$

and the optimizing r^* and n^*. In the next section, we provide an asymptotic analysis for the probability of delay violation under the RR scheduler.

IV. Asymptotic Analysis of Probability of Delay Violation

Here we find the asymptotic probability of delay violation $P_{\text{dv}}(r,n)$, in term of the multiplexing gain r and the cooperative cluster size n. Since it is more intuitive to describe P_{dv} in term of r and T, we use the relation $n = T/2$ and focus on $P_{\text{dv}}(r, T/2)$. Also note that the dependencies on ρ, λ, μ, N, and D are implied in the probability.

Theorem 1: Given the individual bit-arrival process as a compound Poisson process CPE(λ, μ, ρ) and the RR scheduler, the asymptotic delay violation probability P_{dv} is given as

$$\lim_{\rho\to\infty} \frac{\log P_{\text{dv}}(r,T/2)}{\log \rho} = -I^{\text{RR}}(r,T,D), \tag{11}$$

where

$$I^{\text{RR}}(r,T,D) = \min_{\substack{t\in\mathbb{N}:\\ NTt+V>0}} (NTt+V)\Lambda^*\left(\frac{r}{N} + \frac{r(D+1-(N+1)T)}{N(NTt+V)}\right), \tag{12}$$

where $V = T-1-D (\bmod NT)$ and

$$\Lambda^*(x) = \mu(\sqrt{x} - \sqrt{\lambda})^2. \tag{13}$$

Proof: See Appendix I. ■

Note that, under the RR scheduler, the queuing system of user 1 is equivalent to a batch-service queue where services happen every NT timeslots. Each service serves RT oldest bits and takes T timeslots to complete (i.e. the bits leave the decoder).

Approximation 1: By relaxing the integer optimization in (12), we have

$$I^{\text{RR}}(r,T,D) \geq \mu(\frac{r}{N} - \lambda)(D+1-(N+1)T)$$

and hence we use the following approximation:

$$I^{\text{RR}}(r,T,D) \approx \mu(\frac{r}{N} - \lambda)(D+1-(N+1)T). \tag{14}$$

Proof: See Appendix II. ■

We will see via numerical examples in Section V-C that the approximation in (14) is surprisingly precise for our purpose. Hence, we use the following approximation for the rest of the paper:

$$P_{\text{dv}}(r,n) \dot{\approx} \rho^{-\mu(\frac{r}{N}-\lambda)(D+1-(N+1)2n)}, \tag{15}$$

where the asymptotic approximation $\dot{\approx}$ is due to the integer-relaxation approximation of the exponent of P_{dv}.

V. Main Result: Asymptotically Optimal Cluster Size

In this section, we find the asymptotic total probability of bit error, in term of the channel multiplexing gain r and the cluster size n. Given such relation, we find the optimal choice of the parameters r and n and the negative SNR exponent of the optimal total probability of bit error. Note that the following results on r^*, n^*, and d^* are approximations and based on relaxations of integer-constraint optimization problems.

A. Asymptotic Total Probability of Bit Error

Plugging in P_{ch} from (5) and the approximation of P_{dv} from (15) into (7), we get

$$\begin{aligned} P_{\text{tot}}(r,n) &\doteq P_{\text{ch}}(r,n) + P_{\text{dv}}(r,n) \\ &\dot{\approx} \rho^{-n(1-2r)} + \rho^{-\mu(\frac{r}{N}-\lambda)(D+1-(N+1)2n)} \\ &\doteq \rho^{-\min\left\{n(1-2r), \mu(\frac{r}{N}-\lambda)(D+1-(N+1)2n)\right\}}. \end{aligned}$$

B. Asymptotically Optimal Cluster Size and Optimal Negative SNR Exponent

The optimal choice of r and n and the optimal negative SNR exponent under the RR discipline, d^*, is posed as a mixed integer optimization problem:

$$d^* = \max_{r,n} \min\left\{n(1-2r), \mu(\frac{r}{N}-\lambda)(D+1-(N+1)2n)\right\}, \tag{16}$$

subject to[3]

$$\lambda N < r < 1/2;\ n \in \left\{2, 3, \ldots, \left\lfloor \frac{D}{2(N+1)} \right\rfloor \wedge N\right\}.$$

[3]The limited range of n is a result of the conditions that the expression for P_{ch} in (5) is valid when $n \in \{2, \ldots, N\}$ and P_{dv} in (15) is reasonable when $n \in \{1, \ldots, \left\lfloor \frac{D}{2(N+1)} \right\rfloor\}$. To have a well-defined problem, we require $\left\lfloor \frac{D}{2(N+1)} \right\rfloor \geq 2$.

Let r^* and n^* be the optimizing parameters. We notice that r has opposite effects on the two terms inside the minimum. Hence, we solve the above optimization problem by first fixing n and finding

$$\begin{aligned} & r^*(n) \\ & := \arg_r \left(n(1-2r) - \mu\left(\frac{r}{N}-\lambda\right)(D+1-(N+1)2n) = 0 \right) \\ & = \frac{1}{2} - \frac{\frac{1}{2}-\lambda N}{1+\frac{2nN}{\mu(D+1-(N+1)2n)}}, \end{aligned}$$

where $r^*(n)$ is such that the two elements inside the $\min$ function are equal. Note that $r^*(n)$ is an increasing function of n. Next, we maximize

$$n(1-2r^*(n)) = \frac{\frac{1}{2}-\lambda N}{\frac{1}{2n}+\frac{N}{\mu(D+1-(N+1)2n)}}, \quad (17)$$

with respect to n. Letting n^*_{ir} be the solution to the differential equation:

$$\frac{d}{dn}\left(\frac{1}{2n}+\frac{N}{\mu(D+1-(N+1)2n)}\right) = 0,$$

we see that n^* is equal to either $\lfloor n^*_{\text{ir}} \rfloor$ or $\lceil n^*_{\text{ir}} \rceil$, depending on which value results in a higher value of (17). In other words, n^* can be approximated by its integer-relaxed version (n^*_{ir}) as:

$$n^* \approx \left[\frac{D+1}{2(N+1)\left(1+\sqrt{\frac{N}{(N+1)\mu}}\right)} \right]_2^{\left(\left\lfloor \frac{D}{2(N+1)} \right\rfloor \wedge N\right)}. \quad (18)$$

Plugging in this approximation of n^* in $r^*(n)$ approximately gives the optimal channel multiplexing gain

$$r^* := r^*(n^*) \approx \left[\lambda N + \frac{\frac{1}{2}-\lambda N}{1+\sqrt{\frac{(N+1)\mu}{N}}} \right]_{r^*(n=2)}^{r^*\left(n=\left(\left\lfloor \frac{D}{2(N+1)} \right\rfloor \wedge N\right)\right)} \quad (19)$$

since $r^*(n)$ is an increasing function of n. Finally, we have the optimal negative SNR exponent of P_{tot} as[4]

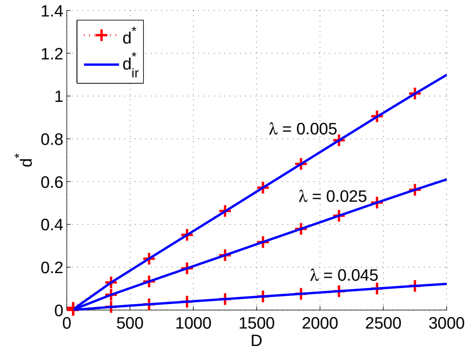
$$d^* = n^*(1-2r^*) \approx \frac{(D+1)(\frac{1}{2}-\lambda N)}{(1+\sqrt{\frac{N}{(N+1)\mu}})^2(N+1)}. \quad (20)$$

Next, we confirm that Approximation 1 performs well. Then, we discuss some interesting observations of the above results.

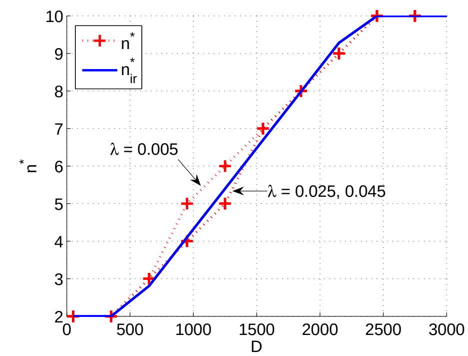
C. Numerical Comparison of the Approximations

Here we want to see if the above optimal cluster size n^*, the multiplexing rate r^*, and the negative SNR exponent d^* are well represented by their approximations, which are the results of relaxing the integer constraints in calculating I^{RR} and n^*. In Figure 3, we compare the exact values of d^*, n^*, and r^* with their approximations (denoted as d^*_{ir}, n^*_{ir}, and r^*_{ir}) given in (18)-(20), at various values of D and λ. We

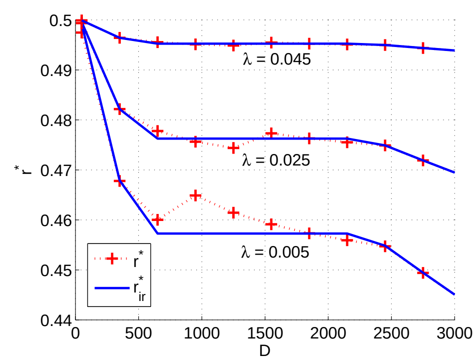
[4]To be more precise, we need to limit the range of d^* to the range of $n^*(1-2r^*(n^*))$ when $n^* \in \left[2, \left\lfloor \frac{D}{2(N+1)} \right\rfloor \wedge N\right]$.



(a) d^* and d^*_{ir} vs D



(b) n^* and n^*_{ir} vs D



(c) r^* and r^*_{ir} vs D

Fig. 3. Comparison of the exact values of d^*, n^*, r^* and their approximations ($d^*_{\text{ir}}, n^*_{\text{ir}}, r^*_{\text{ir}}$), at various delay bounds D and arrival rates ($\lambda = 0.005, 0.025, 0.045$). There are $N = 10$ source nodes and the average packet size $1/\mu$ is 100. Dotted lines with markers correspond to the exact solution and solid lines represent the approximated solution.

consider $N = 10$ and $1/\mu = 100$. It is observed that the approximated values match well with the exact values as D is sufficiently large. The matching is very good for d^* and d^*_{ir}.

D. Discussion of the Main Result

We first make some observations regarding the optimal cluster size n^* and the optimal negative SNR exponent d^* given in (18) and (20), respectively.

- The dependency of n^* on D confirms our intuition that, as the delay requirement becomes stricter, the

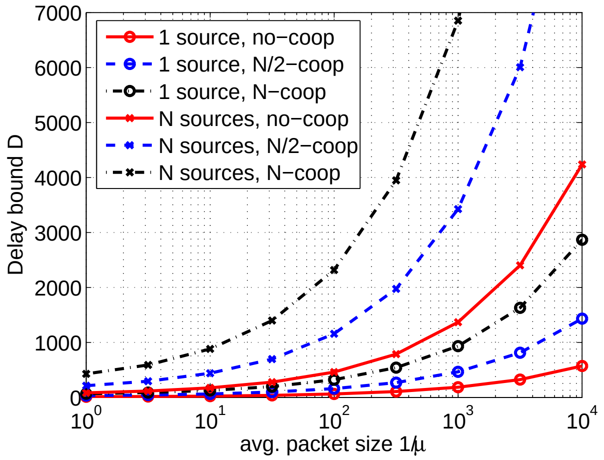


Fig. 4. D vs average packet size $1/\mu$, for N sources and for single source, where $N = 10$.

optimal cluster size should get smaller (given fixed source burstiness, i.e. average packet size $1/\mu$, and the number of sources). This is because cooperation in a larger cluster size, although reduces the probability of channel outage, requires longer coding time and hence increases the delay violation probability.

- Furthermore, for fixed number of sources and source burstiness, the optimal cluster size n^* and the optimal negative SNR exponent d^* increase linearly with the delay bound D. In addition, d^* decreases with the average packet size.

Our results, in addition, provide insight on the advantage of cooperation among many users. As discussed in Section V-B, for any value of λ, μ, D, N, one can compute the optimal cluster size of cooperative cluster, n^*. Here, we contrast this result with those presented in [12], where relays are never sources of information.

Figure 4 shows this comparison when $N = 10$. In this figure, the lines marked by "*" identify those parameters of $(1/\mu, D)$ such that $n^* = 1, n^* = N/2$, and $n^* = N$, when all N nodes in the network are sources of information and they transmit their information in a round-robin fashion. In contrast, the lines marked by "o" represent parameters $(1/\mu, D)$ for which the optimal cluster sizes are $n^* = 1, n^* = N/2$, and $n^* = N$, when only one node is an information source relying on potentially $N-1$ nodes to relay its information. Note that the parameters $(1/\mu, D)$ below the line corresponding to $n^* = 1$ represent those parameters $(1/\mu, D)$ such that it is optimal for the information-source node to directly transmit to the destination node without using any relays, i.e., cooperation is not beneficial in this range of parameters $(1/\mu, D)$. It is interesting that in the case of N information sources, the delay requirement D has to be significantly relaxed before any cooperation will become beneficial.

To be more precise, let us consider the case where the average packet size is 10^3. In a network with a single information source and 10 nodes, when $D = 466$, it is optimal to form cooperation cluster of size 5 (or form cluster size of 10 when $D = 933$.) In contrast, when all nodes are sources of information who are scheduled to transmit according to an RR schedule, no cooperation is optimal for all values of $D \leq 1370$. This is because when all nodes are information sources, the increase in the coding time associated with cooperation degrades P_{dv} in two ways: 1) the larger cluster size increases the duration of the cooperative frame (the batch service time), and also 2) it increases the interval between two cooperative frames in which a given node is transmitting.

Note that the above observations are independent of the arrival rate λ since n^* is always independent of the loading of the system. However, the overall performance, d^*, is a function of λ in both systems. For instance, if the loading is kept constant across the two systems (i.e., $\lambda_{\text{1-source}} = N\lambda_{\text{N-sources}}$), then

$$\frac{d^*_{\text{1-source}}}{d^*_{\text{N-sources}}} = \frac{(1+\sqrt{\frac{N}{(N+1)\mu}})^2(N+1)}{2(1+\frac{1}{\sqrt{2\mu}})^2}.$$

The ratio is approximately N when N and $1/\mu$ are sufficiently large.

VI. Conclusion

In this paper, we considered a cross-layer queue-channel optimization problem for multiple bursty and delay-sensitive information sources. The cross-layer performance metric of interest is the asymptotic total bit loss probability where loss can be due to either delay violation or decoding errors. Under the static round-robin scheduling discipline, we found the optimal values of the transmission rate of the relay channel and the cooperative cluster size such that the negative SNR exponent of the total bit loss probability is maximized.

The work in this paper is only a first step in studying queue-channel performance, in high-SNR asymptotic approximation, for multiple sources and finite delay requirement. Future extensions might include an extension to a dynamic scheduler such as the largest-delay-first scheduler [10]. This will improve the probability of delay violation significantly due to statistical multiplexing gain.

Another interesting topic of research is the cooperative MAC networks in which data from multiple sources is relayed simultaneously. For example, the cooperative MAC protocol proposed in [5] will certainly provide a better performance.

Appendix I
Proof of Theorem 1

Proof: As noted earlier, the RR scheduler provides a constant batch-service to user 1. The server serves user 1 only for T timeslots out of every NT timeslots. That is, each service takes T timeslots to complete and the server takes a vacation for $(N-1)T$ timeslots between consecutive services, no matter if there are bits waiting in the buffer of not. Since the batch service under the RR scheduler is a generalization of the batch service considered in Lemma 2 in [13] (where there is only one information source), we can follow its proof closely.[5]

We start the proof by letting $Q_t^{1,\rho}$ denote the queue length (in bits) of user 1 at timeslot t, when the SNR is ρ. The queue dynamics is

$$Q_t^{1,\rho} = \begin{cases} \left[Q_{t-1}^{1,\rho} + A_t^{1,\rho} - RT\right]^+ & \text{if } t = mNT,\ m \in \mathbb{Z}, \\ Q_{t-1}^{1,\rho} + A_t^{1,\rho} & \text{otherwise,} \end{cases} \tag{21}$$

where $Q_{-\infty}^{1,\rho} \equiv 0$. Since the arrival process is stationary, the system is stable and reaches a steady state, and the system started empty at time $-\infty$, then $Q_i^{1,\rho}$ has the same steady-state distribution as that of $Q_{mNT+i}^{1,\rho}$, $m \in \mathbb{Z}$, for each $i = 0, \ldots, NT-1$. The delay at time i also has the same steady-state distribution as the delay at time $mNT+i$. Since $P_{\text{dv}}(r,T)$, as a function of r, T, is defined as the probability of the steady-state delay greater than D, we have

$$\begin{aligned} &P_{\text{dv}}(r,T) \\ &:= P[\text{(s-s delay of user 1)} > D] \\ &= \frac{1}{NT}\sum_{i=0}^{NT-1} P[\text{delay of any bit of user 1 arrive at time } i > D], \end{aligned}$$

where the equality holds since the arrivals are independent across time. Now, using the result of Lemma 4 in [13] that the delay violation probability of *any* bit arriving at time i is asymptotically equal to the delay violation probability of the *last* bit arriving at time i, we have

$$P_{\text{dv}}(r,T) \doteq \frac{1}{NT}\sum_{i=0}^{NT-1} P(\mathcal{Q}_i^{\text{RR},\rho}) \doteq \sum_{i=0}^{NT-1} P(\mathcal{Q}_i^{\text{RR},\rho}), \tag{22}$$

where we define the event $\mathcal{Q}_i^{\text{RR},\rho}$ to be the event that the last bit of user 1 arriving at timeslot i violates the delay bound D, under the RR discipline and use the fact that NT is a constant, independent of ρ. Hence, we see that P_{dv} is asymptotically equal to the sum of $P(\mathcal{Q}_i^{\text{RR},\rho})$.

Next we try to relate the event $\mathcal{Q}_i^{\text{RR},\rho}$ to a condition on the queue length $Q_i^{1,\rho}$, for $i = 0, \ldots, NT-1$. To do this, we need to look at how the delay of the last bit arriving at timeslot i violates the delay bound D. Upon arrival, this last bit sees $Q_i^{1,\rho}$ bits (including itself) waiting in the queue, Since the batch service happens exactly at time in multiple of NT, the bit must wait $NT-i$ timeslots for the next service to start and another $\left\lceil \frac{Q_i^{1,\rho}}{RT} \right\rceil T + (\left\lceil \frac{Q_i^{1,\rho}}{RT} \right\rceil - 1)(NT-T)$ timeslots for the bit to get served and be decoded. This is because it takes $\left\lceil \frac{Q_i^{1,\rho}}{RT} \right\rceil$ services to serve all $Q_i^{1,\rho}$ bits. Each service takes T timeslots to complete and the next service comes $NT-T$ timeslots after the completion of the current one. Hence, the total waiting time between services is $(\left\lceil \frac{Q_i^{1,\rho}}{RT} \right\rceil - 1)(NT-T)$. As a result, the last bit violates the delay bound D if and only if

$$NT - i + \left\lceil \frac{Q_i^{1,\rho}}{RT} \right\rceil T + \left(\left\lceil \frac{Q_i^{1,\rho}}{RT} \right\rceil - 1\right)(NT-T) > D.$$

Let $\Omega^{1,\rho}$ contain all measurable sample paths of the arrival process $A^{1,\rho}$. The condition above implies that the delay violation event for the last bit is given as

$$\mathcal{Q}_i^{\text{RR},\rho} = \left\{\omega \in \Omega^{1,\rho} : T - i + \left\lceil \frac{Q_i^{1,\rho}(\omega)}{RT} \right\rceil NT > D \right\}.$$

Now using an important observation that, for any $x, y \in \mathbb{R}$,

$$\lceil x \rceil > y \Leftrightarrow \lceil x \rceil > \lfloor y \rfloor \Leftrightarrow x > \lfloor y \rfloor,$$

we have

$$\mathcal{Q}_i^{\text{RR},\rho} = \left\{\omega \in \Omega^{1,\rho} : Q_i^{1,\rho}(\omega) > RT \left\lfloor \frac{D+i-T}{NT} \right\rfloor \right\}. \tag{23}$$

By defining $k := D(\bmod NT)$ (note that $\frac{D-k}{NT}$ is an integer), the term $\left\lfloor \frac{D+i-T}{NT} \right\rfloor$ can be expressed as

$$\left\lfloor \frac{D+i-T}{NT} \right\rfloor = \begin{cases} \frac{D-k-NT}{NT}, & \text{if } k \in [0, T-1], i \in [0, T-1-k] \\ \frac{D-k}{NT}, & \text{if } \begin{cases} k \in [0, T-1], i \in [T-k, NT-1] \text{ or} \\ k \in [T, NT-1], i \in [0, NT+T-1-k] \end{cases} \\ \frac{D-k+NT}{NT}, & \text{if } \begin{cases} k \in [T, NT-1], \\ i \in [NT+T-k, NT-1] \end{cases} \end{cases} \tag{24}$$

Next, we use the results in (21)-(24) and follow the techniques used in the proof of Lemma 2 in [13] to get $\lim_{\rho\to\infty} \frac{\log P_{\text{dv}}(r,T)}{\log \rho}$.[6] We consider two cases, depending on the value of k, as follows:

Case 1: $k \in [0, T-1]$. Using (21)-(24), it can be shown that the delay violation probability of the last bit of user 1 arriving at timeslot $T-1-k$ asymptotically dominates P_{dv}, i.e.,

$$I^{\text{RR}}(r,T) := -\lim_{\rho\to\infty} \frac{\log P_{\text{dv}}(r,T)}{\log \rho} = -\lim_{\rho\to\infty} \frac{\log P(\mathcal{Q}_{T-1-k}^{\text{RR},\rho})}{\log \rho}.$$

Now using (23) and (24) with $k \in [0, T-1]$ and $i = T-1-k$, we have

$$I^{\text{RR}}(r,T) = -\lim_{\rho\to\infty} \frac{\log P(Q_{T-1-k}^{1,\rho} > (D-k-NT)R/N)}{\log \rho}.$$

By using the queue dynamics (21) recursively, the queue length $Q_i^{1,\rho}$ is related to the arrivals $A_j^{1,\rho}$, $j \le i$, in the following manner

$$Q_i^{1,\rho} = \sup_{t \in \mathbb{N}} \left\{ \sum_{j=-tNT+1}^{i} A_j^{1,\rho} - tRT \right\}. \tag{25}$$

[5]In fact, the result of Lemma 2 in [13] can be thought of a special case of Theorem 1 when there is only one source.

[6]For brevity we will show only the main steps and omit the detailed reasonings.

By using this expression and the large-deviation principle of the arrival process $(A^{1,\rho})$, it can be shown that

$$\begin{aligned}
&I^{\text{RR}}(r,T)\\
&= -\lim_{\rho\to\infty}\frac{1}{\log\rho}\cdot\\
&\log P\left(\sup_{t\in\mathbb{N}}\left\{\sum_{j=-tNT+1}^{T-1-k}A_j^{1,\rho}-tRT\right\}>(D-k-NT)\frac{R}{N}\right)\\
&= \min_{t\in\mathbb{N}}(tNT+V)\Lambda^*\left(\frac{r(tNT+D-k-NT)}{N(tNT+V)}\right)\\
&= \min_{t\in\mathbb{N}}(tNT+V)\Lambda^*\left(\frac{r}{N}+\frac{(D+1-(N+1)T)\frac{r}{N}}{tNT+V}\right),
\end{aligned}$$

where we use the short-hand notation $V=T-1-k$.

Case 2: $k\in[T,NT-1]$. Similarly, it can be shown that

$$\begin{aligned}
&I^{\text{RR}}(r,T)\\
&= -\lim_{\rho\to\infty}\frac{\log P(\mathcal{Q}^{\text{RR},\rho}_{NT+T-1-k})}{\log\rho}\\
&= -\lim_{\rho\to\infty}\frac{\log P(Q^{1,\rho}_{NT+T-1-k}>(D-k)R/N)}{\log\rho}\\
&= \min_{t\in\mathbb{N}}(NTt+NT+V)\Lambda^*\left(\frac{r}{N}+\frac{(D+1-(N+1)T)\frac{r}{N}}{NTt+NT+V}\right).
\end{aligned}$$

We can combine the results of the above two cases into a general one for any k to write

$$\begin{aligned}
&I^{\text{RR}}(r,T)\\
&= \min_{\substack{t\in\mathbb{N}:\\ NTt+V>0}}(NTt+V)\Lambda^*\left(\frac{r}{N}+\frac{(D+1-(N+1)T)\frac{r}{N}}{NTt+V}\right),
\end{aligned}$$

where, from (1), the scaled log moment generating function Λ of CPE(λ,μ,ρ) bit-arrival process is given as

$$\Lambda(\theta):=\frac{\log E[e^{\theta A_1^{1,\rho}}]}{\log\rho}=\frac{\mu\lambda\theta}{\mu-\theta},\ \theta<\mu,\forall\rho, \qquad (26)$$

and hence Λ^* is given as

$$\Lambda^*(x):=\sup_{\theta\in\mathbb{R}}\theta x-\Lambda(\theta)=\mu(\sqrt{x}-\sqrt{\lambda})^2.$$

■

APPENDIX II
PROOF OF APPROXIMATION 1

Proof: We relax the integer optimization in I^{RR} to get a lower bound on I^{RR}:

$$\begin{aligned}
I^{\text{RR}}(r,T) &\geq \min_{\tau\in\mathbb{R}^+}\tau\Lambda^*\left(\frac{r}{N}+\frac{(D+1-(N+1)T)\frac{r}{N}}{\tau}\right)\\
&= \delta_{\frac{r}{N}}\frac{r}{N}(D+1-(N+1)T),
\end{aligned}$$

where we use the result of Lemma 1.7 in [7] with $\delta_{\frac{r}{N}}$ defined as the solution to $\Lambda(\delta_{\frac{r}{N}})=\frac{r}{N}\delta_{\frac{r}{N}}$. Using Λ in (26), we get

$$\delta_{\frac{r}{N}}=\mu(1-\frac{\lambda N}{r})$$

and hence

$$I^{\text{RR}}(r,T)\geq\mu\left(\frac{r}{N}-\lambda\right)(D+1-(N+1)T).$$

■

REFERENCES

[1] L. Zheng and D. Tse, "Diversity-multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Info. Theory*, v. 49, no. 5, May 2003.

[2] A. Sendonaris, E. Erkip and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," *In Proceedings of 1998 IEEE ISIT*, pp. 156, August 1998.

[3] A. Sendonaris, E. Erkip and B. Aazhang, "User cooperation diversity-Part I: System description," *IEEE Trans. Communications*, vol. 51, no. 11, pp. 1927-1938, November 2003.

[4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Info. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[5] K. Azarian, H. El Gamal, and P. Schniter, "On the Achievable Diversity-Multiplexing Tradeoff in Half-Duplex Cooperative Channels," *IEEE Trans. Info. Theory*, vol. 51, no. 12, pp. 4152-4172, Dec. 2005.

[6] D. D. Botvich and N. G. Duffield, "Large deviations, the shape of the loss curve, and economies of scale in large multiplexers," *Queuing Systems*, vol. 20, pp. 293–320, 1995.

[7] A. Ganesh, N. O'Connel, and D. Wischik, *Big Queues*, Springer-Verlag, Berlin, 2004.

[8] A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd edition, Springer-Verlag, New York, 1998.

[9] T. Holliday and A. Goldsmith, "Joint source and channel coding for MIMO systems," in *Proc. Allerton Conf. Communications, Control and Computing*, Oct. 2004.

[10] A. L. Stolyar and K. Ramanan, "Largest weighted delay first scheduling: Large deviations and optimality." *Ann. Applied Probability*, v. 11, pp. 1-48, 2001.

[11] P. Elia, K. Vinodh, M. Anand and P. V. Kumar, "D-MG Tradeoff and Optimal Codes for a Class of AF and DF Cooperative Communication Protocols," submitted to *IEEE Trans. Info. Theory*, January 2007, available at: http://arxiv.org/pdf/cs.IT/0611156

[12] P. Elia, S. Kittipiyakul, and T. Javidi, "Cooperative diversity in wireless networks with stochastic and bursty traffic," Proceeding of ISIT 2007, June 2007.

[13] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with bursty and delay-limited information," in preparation.