

Jointly Optimal Rate Control and Relay Selection for Cooperative Wireless Video Streaming

Zhangyu Guan, *Member, IEEE*, Tommaso Melodia, *Member, IEEE*, and Dongfeng Yuan, *Senior Member, IEEE*

Abstract—Physical-layer cooperation allows leveraging the spatial diversity of wireless channels without requiring multiple antennas on a single device. However, most research in this field focuses on optimizing physical-layer metrics, with little consideration for network-wide and application-specific performance measures. This paper studies cross-layer design techniques for video streaming over cooperative networks. The problem of joint rate control, relay selection, and power allocation is formulated as a mixed-integer nonlinear problem, with the objective of maximizing the sum peak signal-to-noise ratio (PSNR) of a set of concurrent video sessions. A global optimization algorithm based on the branch and bound framework and on convex relaxation of nonconvex constraints is then proposed to solve the problem. The proposed algorithm can provide a theoretical upper bound on the achievable video quality and is shown to provably converge to the optimal solution. In addition, it is shown that cooperative relaying allows nodes to save energy without leading to a perceivable decrease in video quality. Based on this observation, an uncoordinated, distributed, and localized low-complexity algorithm is designed, for which we derive conditions for convergence to a Nash equilibrium (NE) of relay selection. The distributed algorithm is also shown to achieve performance comparable in practice to the optimal solution.

Index Terms—Cooperative communication, rate control, relay selection, video streaming.

I. INTRODUCTION

THE NOTION of *spatial diversity* refers to the idea of using multiple transceiver antennas to effectively cope with fading in wireless channels. The underlying principle is that different propagation channels can be established with multiple transceiver pairs between a transmitting and a receiving

node. By sending signals that carry the same information through different channels, multiple faded copies of the same information can be obtained at the receiving node. The communication link reliability can then be considerably improved since, roughly speaking, the probability that all channels go down at the same time is low, resulting in higher data rate or lower power consumption.

Spatial diversity is traditionally exploited by using multiple transceiver antennas (i.e., *multiple-input-multiple-output* (MIMO) [2]). However, equipping a mobile device with multiple antennas may not be practical since the minimum required separation between the antennas is dictated by the operating radio wavelength. The concept of *cooperative communications* has been therefore proposed to achieve spatial diversity without requiring multiple transceiver antennas on the same node [3], [4]. In cooperative communications, in their *virtual multiple-input-single-output* (VMISO) variant, each node is equipped with a single antenna and relies on the antennas of neighboring devices to achieve spatial diversity. There is a vast and growing literature on information and communication theoretic results in cooperative communications. The reader is referred to [5], [6], and references therein for excellent surveys of the main results in this area. However, the common objective of most research in this field is to optimize physical-layer performance measures (i.e., bit error rate and link outage probability) from a broad system perspective, without considering in much detail how cooperation interacts with higher layers of the protocol stack to improve network performance measures. For example, [7] and [8] investigate the achievable rates and diversity gains of cooperative schemes focusing on a single source and destination pair. Some initial promising work on networking aspects of cooperative communications includes studies on medium access control protocols to leverage cooperation [4], [9]–[12], delay-centric cooperation [13], cooperative routing [14]–[16], optimal network-wide relay selection [17]–[19], and optimal stochastic control [20]. In these works, the main emphasis is not on the impact of cooperation on end-to-end video delivery. Moreover, different from [8], which derived the capacity for interference-aware cooperative relay networks, and different from most of the above works that consider interference-free network, i.e., orthogonal channels have been established *a priori* (e.g., [17] and [18]), we focus on interference-limited cooperative networks.

This paper studies cross-layer design techniques for video streaming over interference-limited cooperative wireless networks with distributed control. Specifically, we study strategies for joint control of the video encoding rate at the application layer, relay selection, and power control at the link and physical layers to maximize the sum peak signal-to-noise ratio (PSNR) of multiple concurrent video sessions. Cooperative video delivery

Manuscript received July 24, 2011; revised January 10, 2012; May 11, 2012; and August 18, 2012; accepted September 04, 2012; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor M. Reisslein. Date of publication March 21, 2013; date of current version August 14, 2013. This paper is based upon work supported in part by the National Science Foundation under Grant CNS1117121 and the Office of Naval Research under Grant N00014-11-1-0848. The work of Z. Guan was supported by the NSFC under Grant 61101120 and the Doctoral Fund of the Ministry of Education of China under Grant 20110131120028. A preliminary shorter version of this paper appeared in the Proceedings of the IEEE International Conference on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON), Salt Lake City, UT, USA, June 27–30, 2011.

Z. Guan is with the School of Information Science and Engineering, Shandong University, Jinan 250100, China, and also with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: zguan2@buffalo.edu).

T. Melodia is with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: tmelodia@buffalo.edu).

D. Yuan is with the School of Information Science and Engineering, Shandong University, Jinan 250100, China (e-mail: dfyuan@sdu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNET.2013.2248020

has attracted significant attention [21]–[24]. For video delivery, video quality is measured in terms of PSNR, which is a function of delay and throughput. Higher data rates result in higher video encoding rates and hence higher video quality, but at the same time cause higher delays. Higher delays result in higher packet losses because of violations of delay constraints, and consequently cause lower video quality. Therefore, delay and throughput need to be optimized jointly for high-quality video delivery.

The main contributions of this paper can be outlined as follows.

- 1) We first formulate the problem of *joint rate control, relay selection, and power control for video streaming in cooperative networks*. The problem turns out to be a nonlinear (nonconvex) and combinatorial optimization problem [i.e., a mixed-integer nonlinear problem (MINLP)].
- 2) Since MINLPs are in general NP-hard, we develop a solution algorithm with provable convergence based on the branch and bound (B&B) [25] framework and on relaxation of nonconvex problem constraints. The proposed algorithm searches for an ϵ -optimal solution iteratively. At each iteration, we relax the original nonconvex problem to a series of convex problems. We show that the proposed algorithm provides guaranteed convergence to the optimal solution.
- 3) In addition, through well-crafted numerical simulations, we show that cooperative relaying allows nodes to transmit at lower power without leading to a perceivable decrease in video quality. Intuitively, this happens because the effect of power control in interference-limited wireless networks is mitigated by cooperative relaying. Based on this observation, we design an uncoordinated, distributed, and localized low-complexity algorithm.
- 4) We study and demonstrate convergence of the distributed algorithm to a Nash equilibrium (NE) of relay selection and rate control. We compare the performance of the distributed algorithm to the optimal algorithm and show its excellent video quality performance in practice.

The rest of the paper is organized as follows. In Section II, we discuss related work in detail. In Section III, we briefly describe the communication system model and introduce the problem formulation. The centralized algorithm is described in detail in Section IV, while our distributed algorithm is then discussed in Section V. Examples and numerical results are discussed in Section VI. Finally, we draw the main conclusions in Section VII.

II. RELATED WORK

Resource allocation has long been an important research topic [26], [27], and there exists a solid body of works addressing resource allocation in cooperative networks at different layers of the protocol stack [4], [9]–[11], [13]–[20], [28], [29]. For example, Yang *et al.* [29] consider a system model where a relay node can be shared by multiple source–destination pairs. They propose an optimal algorithm that runs in polynomial time to solve the relay assignment problem to maximize the total capacity of all source–destination pairs. Sharma and Hou [14] study a joint problem of relay node assignment and multihop flow routing, with the objective to maximize the minimum rate among a set of concurrent sessions. Additionally, several

important papers have considered cross-layer design in cooperative networks (see [4], [16], [18], and [20]). For example, the authors of [4] proposed a cross-layer framework to exploit virtual MISO links in mobile ad hoc networks. Yeh *et al.* [20] formulate and solve an optimal stochastic control problem with cooperative relays. Cooperative communication has also been studied in the context of dynamic spectrum access or cognitive radio. For example, Zhang *et al.* [30] demonstrate that the network throughput in cooperative networks can be increased by jointly exploiting spatial and spectrum diversity. In this paper, we further consider the multimedia quality of information (QoI) at the application (APP) layer and jointly optimize the video rate control, relay node selection, and power control.

There are several excellent contributions in cooperative video delivery. For example, Mastrorarde *et al.* [21] proposed a solution based on cooperative coding, which warrants a uniformly better experience to the video users and needs relatively modest changes to the multiple access cross-layer optimization framework proposed by Fu and van der Schaar in [31]. Alay *et al.* [22] studied layered video multicast in a two-hop cooperative infrastructure-based networks, while Khalek and Dawy [23] derived the optimal resource allocation solution for energy efficient scalable video distribution over cooperative multihop networks. Xiao *et al.* [24] proposed a joint source-channel coding (JSCC) framework for video transmission. In this paper, we focus on optimizing the sum video quality of different video sessions by designing both centralized and distributed solution algorithms.

We rely on advanced optimization theoretic notions to design globally optimal and distributed algorithms, respectively. Applying optimization theory to solve complex resource allocation problems in cooperative networking has also received previous attention. Hou *et al.* [18] designed an optimal algorithm for joint flow routing and relay selection based on the branch-and-cut framework. Rossi *et al.* [19] proposed a focused real-time dynamic programming (FRTDP) approach to study the stochastic shortest-path problem in cooperative networks. Here, we design an optimal solution algorithm based on a combination of branch-and-bound framework and convex relaxation techniques. Different from [18], we study optimization of cooperative networks based on the signal-to-interference-plus-noise ratio (SINR) interference model, and different from [19], we focus on multiple concurrent flows.

III. PROBLEM FORMULATION

We consider a decentralized single-hop video streaming network, where each source node compresses a video sequence at a given rate. The scenario considered is, for example, representative of a multimedia sensor network [32]. The video content is enqueued at the source node buffer and then transmitted to the destination through a direct or cooperative link. If the video packet is not received before a predefined playout deadline, the packet is dropped. Video packets can also be dropped because of transmission errors caused by interference or channel fading. Cooperative communication techniques may be employed on each link to potentially increase the channel capacity, e.g., cooperative relaying, distributed space time coding, coded cooperation, and virtual spatial multiplexing. In this paper, we focus on decode-and-forward cooperative relaying.

There are multiple potential relay nodes, and each video session can select one of them (or none) as a relay node. The transmission time is divided into consecutive time slots, and if a session uses cooperative relay, a transmission is completed in two time slots. In the first time slot, the source node broadcasts its information to both destination and relay nodes, and in the second, the relay node forwards the received information to the destination. There are two common cooperation strategies: amplify-and-forward (AF) and decode-and-forward (DF) [3]. In AF, the relay simply amplifies the received signal and forwards it to the destination. With DF, the relay first decodes the received signal, then forwards it if it can be successfully decoded. In this paper, we concentrate on the DF strategy. However, the proposed ϵ -optimal algorithm can be extended to AF as discussed in Section IV. Distributed algorithms can be designed accordingly. The objective of the problem is to maximize the sum of the video qualities (expressed as the sum-PSNR) of multiple concurrent video sessions by jointly regulating the video encoding rate for each session, adjusting the transmission power for each source and relay node, and selecting the optimal relay node for each session. We start by introducing the link capacity model for direct transmission and for cooperative relaying in Section III-A, and the video distortion model in Section III-B. Then, in Section III-C, we formulate the MINLP problem.

A. Link Capacity

Denote \mathcal{S} as the set of video sessions and \mathcal{R} as the set of potential relay nodes. Define the vector of relay assignments as $\boldsymbol{\alpha} = \{\alpha_{s,r} \mid s \in \mathcal{S}, r \in \mathcal{R}\}$, where $\alpha_{s,r} = 1$ iff relay node r is selected as relay for video session s , and $\alpha_{s,r} = 0$ otherwise. Assume that each video session can select at most one relay in each cooperative transmission and each relay can at most be selected by one video session. We have

$$\alpha_{s,r} \in \{0, 1\} \quad \forall s \in \mathcal{S}, \quad \forall r \in \mathcal{R} \quad (1)$$

$$\sum_{r \in \mathcal{R}} \alpha_{s,r} \leq 1 \quad \forall s \in \mathcal{S} \quad (2)$$

$$\sum_{s \in \mathcal{S}} \alpha_{s,r} \leq 1 \quad \forall r \in \mathcal{R}. \quad (3)$$

Denote the maximum transmission power of each source and relay node as P_{\max}^s and P_{\max}^r , respectively. Then, it must hold

$$P_s \geq 0 \quad \forall s \in \mathcal{S} \quad (4)$$

$$P_r \geq 0 \quad \forall r \in \mathcal{R} \quad (5)$$

$$P_s \leq P_{\max}^s \quad \forall s \in \mathcal{S} \quad (6)$$

$$P_r \leq P_{\max}^r \quad \forall r \in \mathcal{R} \quad (7)$$

where P_s and P_r denote the transmission power of source s and relay r , respectively. Moreover, if a relay is not selected by any session, its transmission power should be zero so it does not cause interference to other nodes, otherwise it takes a value from $[0, P_{\max}^r]$. Therefore, we have

$$\left(\sum_{s \in \mathcal{S}} \alpha_{s,r} - 1 \right) P_r = 0 \quad \forall r \in \mathcal{R}. \quad (8)$$

In the proposed ϵ -optimal solution algorithm in Section IV, the constraint (8) will be used to search for a feasible solution, while

it can simply be omitted in the convex relaxation. Note that the omission does not affect optimality of the proposed algorithm since the constraint can be explicitly satisfied by setting the transmission power of a relay node not selected by any session to zero in the two-stage convex relaxation.

Now, assume that multiple concurrent transmissions are allowed on the same portion of the spectrum, e.g., through code division multiple access (CDMA) or time-hopping impulse-radio ultrawideband (TH-IR-UWB). Denote the total available bandwidth as B and the spreading gain as χ . We only consider long-term channel state information (CSI) and assume that the additive white Gaussian noise (AWGN) power perceived at each relay and destination node is equal to σ^2 .

The link capacity C_s for session s can be expressed as

$$C_s = \left(1 - \sum_{r \in \mathcal{R}} \alpha_{s,r} \right) C_{s2d}^s + \sum_{r \in \mathcal{R}} \alpha_{s,r} C_{\text{coop}}^s, \quad (9)$$

where C_{s2d}^s and C_{coop}^s represent the direct link capacity and cooperative link capacity for session s , respectively.

1) *Direct Link Capacity C_{s2d}^s* : In the case where session s uses only a direct link, i.e., $\sum_{r \in \mathcal{R}} \alpha_{s,r} = 0$, we have

$$C_{s2d}^s = B \log_2 \left(1 + \frac{\chi G_{s2d}^{r_s, s} P_s}{\sigma^2 + I_{\text{dst}}^s} \right) \quad (10)$$

where $G_{s2d}^{r_s, s}$ represents the average channel gain from the source node of session s to the corresponding destination node, and I_{dst}^s is the average interference perceived at the destination node, which can in turn be expressed as

$$I_{\text{dst}}^s = \sum_{w \in \mathcal{S}, w \neq s} I_{\text{dst}, s}^{\text{src}, w} + \sum_{r \in \mathcal{R}} I_{\text{dst}, s}^{\text{rly}, r}. \quad (11)$$

In (11), $I_{\text{dst}, s}^{\text{src}, w}$ denotes the interference at the destination of session s caused by the source of session w , and $I_{\text{dst}, s}^{\text{rly}, r}$ denotes the interference at the destination of session s caused by relay node r .

Interference on each link depends on power allocation and relay selection at each individual node, but also on the network scheduling strategy (i.e., the relative synchronization of transmission start times between different network communication links). To keep the model treatable, the interference at each receiver can be approximated in different ways, e.g., worst-case approximation assuming that all source and active relay nodes cause interference in both time slots, or average-based approximation, which considers the average effect of each interferer over the two time slots. Our investigation reveals that the average-based approximation approximates reality very well—*in-depth validation of the average-based interference model is discussed in detail in Appendix A*. The interference caused by the source of session w to the destination of session s can be expressed as

$$I_{\text{dst}, s}^{\text{src}, w} = \underbrace{\left(1 - \sum_{r \in \mathcal{R}} \alpha_{w,r} \right) G_{s2d}^{w, s} P_w}_{\text{direct link}} + \frac{1}{2} \underbrace{\left(\sum_{r \in \mathcal{R}} \alpha_{w,r} \right) G_{s2d}^{w, s} P_w}_{\text{cooperative link}} \quad (12)$$

where P_w represents transmission power of the source node of session w .

The other component in (11), $I_{\text{dst},s}^{\text{rly},r}$, represents the interference generated by relay node r at the destination of session s when it is used by another video session, i.e.,

$$I_{\text{dst},s}^{\text{rly},r} = \frac{1}{2} \sum_{w \in \mathcal{S}, w \neq s} \alpha_{w,r} G_{\text{r2d}}^{r,s} P_r \quad (13)$$

where $G_{\text{r2d}}^{r,s}$ is the average channel gain from relay r to the destination of session s .

2) *Cooperative Link Capacity C_{coop}^s* : Assume that the DF cooperative relaying mode is employed at each relay node. If session s uses the cooperative link, i.e., $\sum_{r \in \mathcal{R}} \alpha_{s,r} = 1$, we have

$$C_{\text{coop}}^s = \frac{1}{2} \min(C_{\text{s2r}}^s, C_{\text{sr2d}}^s) \quad (14)$$

where C_{s2r}^s is the capacity from source s to the selected relay node, and C_{sr2d}^s is the capacity achieved through maximal ratio combining of the received signals at the destination [3].

The link capacity from source node to relay node is modeled as

$$C_{\text{s2r}}^s = B \log_2 \left(1 + \frac{\chi \sum_{r \in \mathcal{R}} \alpha_{s,r} G_{\text{s2r}}^{s,r} P_s}{\sigma^2 + \sum_{r \in \mathcal{R}} \alpha_{s,r} I_{\text{rly}}^{r,s}} \right) \quad (15)$$

where $G_{\text{s2r}}^{s,r}$ is the average channel gain from the destination of session s to relay r , $I_{\text{rly}}^{r,s}$ is the interference perceived at relay r assuming that it is selected by session s , and consists of two components as

$$I_{\text{rly}}^{r,s} = I_{\text{rly}}^{r,s,\text{src}} + I_{\text{rly}}^{r,s,\text{rly}} \quad (16)$$

where $I_{\text{rly}}^{r,s,\text{src}}$ represents the component of $I_{\text{rly}}^{r,s}$ caused by all other source nodes, and $I_{\text{rly}}^{r,s,\text{rly}}$ represents the component of $I_{\text{rly}}^{r,s}$ caused by relay nodes. The two components can be expressed as

$$I_{\text{rly}}^{r,s,\text{src}} = \underbrace{\sum_{w \in \mathcal{S}, w \neq s} \left(1 - \sum_{r \in \mathcal{R}} \alpha_{wr} \right) G_{\text{s2r}}^{w,r} P_w}_{\text{direct link}} + \frac{1}{2} \underbrace{\sum_{w \in \mathcal{S}, w \neq s} \left(\sum_{r \in \mathcal{R}} \alpha_{wr} \right) G_{\text{s2r}}^{w,r} P_w}_{\text{cooperative link}} \quad (17)$$

$$I_{\text{rly}}^{r,s,\text{rly}} = \frac{1}{2} \sum_{u \in \mathcal{R}, u \neq r} \left(\sum_{w \in \mathcal{S}, w \neq s} \alpha_{w,u} \right) G_{\text{r2r}}^{u,r} P_u \quad (18)$$

where $G_{\text{r2r}}^{u,r}$ represents the average channel gain from relay u to relay r .

Finally, the link capacity achieved through maximal ratio combining of the received signals at the destination in (14) can be expressed as

$$C_{\text{sr2d}}^s = B \log_2 \left(1 + \frac{\chi \left(G_{\text{s2d}}^{s,s} P_s + \sum_{r \in \mathcal{R}} \alpha_{s,r} G_{\text{r2d}}^{r,s} P_r \right)}{(\sigma^2 + I_{\text{dst}}^s)} \right) \quad (19)$$

where I_{dst}^s is defined in (11).

B. Video Distortion Model

The video quality is measured in terms of PSNR, which is a monotonically decreasing function of the mean-square error (MSE) [33]

$$\text{PSNR}_s = 10 \log_{10}(D_{\text{max}}/D_s). \quad (20)$$

In (20), PSNR_s represents the PSNR of video session s , D_s is the corresponding video distortion, and D_{max} is a constant parameter representing the maximum of possible distortion.

Video distortion is caused by the interplay of lossy video compression, denoted as D_{enc}^s , and distortion caused by video packet loss, denoted as D_{los}^s ,

$$D_s = D_{\text{enc}}^s + D_{\text{los}}^s. \quad (21)$$

D_{enc}^s is a function of the video encoding rate R_s , modeled as [33]

$$D_{\text{enc}}^s(R_s) = D_0^s + \frac{\theta_s}{(R_s - R_0^s)} \quad D_0^s > 0, R_s > R_0^s \quad \forall s \in \mathcal{S} \quad (22)$$

where D_0^s , R_0^s , and θ_s are video-specific parameters.

Packet loss is caused by transmission errors and violations of the playout deadline caused by queuing delay. We denote the packet error rate for video session s as P_{err}^s . In this paper, we consider constant P_{err}^s for each session and assume that each session varies its physical-layer transmission scheme adaptively (e.g., modulation, coding) such that the transmission error probability is independent of the specific transmission strategy. For the sake of simplicity, we employ a simple $M/M/1$ model as in [34]—our model can be easily extended to account for more sophisticated delay models, e.g., a Chernoff-bound-based model [35]. The average queuing delay T_{dly}^s of video session s can then be expressed as [36]

$$T_{\text{dly}}^s = L_s / (C_s - R_s) \quad (23)$$

where L_s and R_s are the average packet length and video encoding rate, respectively, and C_s represents the link capacity available to session s as defined in (9). The probability that the queuing delay of a packet video session s exceeds the playout deadline T_0^s , measured in seconds, can be written as

$$P_{\text{dly}}^s = e^{-(C_s - R_s)T_0^s / L_s}. \quad (24)$$

The distortion caused by packet loss D_{los}^s in (21) can then be formulated as

$$D_{\text{los}}^s = k_s (P_{\text{err}}^s + (1 - P_{\text{err}}^s) P_{\text{dly}}^s) \quad (25)$$

where k_s is a parameter representing the sensitivity of a video sequence to packet loss, which can be measured offline or estimated in real time.

C. Sum-PSNR Maximization Problem

We can now formulate the problem of maximizing the sum-PSNR of multiple video sessions by jointly controlling the video encoding rate, power allocation, and relay selection, subject to a set of constraints as follows.

- *Minimum rate*: Each video session requires a minimum encoding rate R_0^s for video session s , i.e.,

$$R_s \geq R_0^s \quad \forall s \in \mathcal{S}. \quad (26)$$

- *Link capacity*: The video encoding rate for each video session cannot exceed the available link capacity

$$R_s \leq C_s \quad \forall s \in \mathcal{S}. \quad (27)$$

- *Queuing delay*: The average queuing delay for each video session cannot exceed the playout deadline,¹ i.e.,

$$T_{\text{dly}}^s \leq T_0^s \quad \forall s \in \mathcal{S}. \quad (28)$$

By substituting (23) into (28), we have

$$R_s \leq C_s - \frac{L_s}{T_0^s}. \quad (29)$$

Therefore, the link capacity constraint in (27) is implied by the queuing delay constraint.

- *Maximum transmission power*: The transmission power of each source and relay node is limited by the maximum transmission power as in (6) and (7).
- *Relay selection*: Relay assignment has to satisfy constraints expressed as in (1)–(3).

Define $\mathbf{P} = \{P_s, P_r \mid \forall s \in \mathcal{S}, r \in \mathcal{R}\}$, $\mathbf{R} = \{R_s \mid \forall s \in \mathcal{S}\}$, and $\boldsymbol{\alpha} = \{\alpha_{s,r} \mid s \in \mathcal{S}, r \in \mathcal{R}\}$ as the vectors of power allocation strategy, video encoding rate, and relay selection strategy, respectively. Then, the problem can be formulated as

$$\text{Given : } G_{s2d}^{w,s}, G_{r2d}^{r,s}, G_{r2r}^{u,r}, R_0^s, D_0^s, \theta_s, P_{\max}^s, P_{\max}^r, w, s \in \mathcal{S}, u, r \in \mathcal{R} \quad (30)$$

$$\text{Find : } \boldsymbol{\alpha}, \mathbf{P}, \mathbf{R} \quad (31)$$

$$\text{Maximize } f = \sum_{s \in \mathcal{S}} \text{PSNR}_s \quad (32)$$

$$\text{Subject to : } (1)–(22), (24)–(26), (29). \quad (33)$$

It is worth pointing out that, although the problem formulation in this section focuses on video streaming only, it can be extended to account for heterogeneous traffic sources. Additionally, the formulated problem in (30)–(33) can be easily extended to incorporate fairness among users, e.g., by maximizing the sum of weighted PSNRs or the sum of $\log(\text{PSNRs})$ without changing concavity of the objective function. Other forms of fairness e.g., max-min, will be considered in our future research.

IV. OPTIMAL SOLUTION ALGORITHM

The problem formulated in Section III-C is a nonlinear, non-convex combinatorial problem. In general, MINLP problems are NP-HARD, i.e., no existing algorithm can solve an arbitrary MINLP in polynomial time. We propose a solution algorithm based on the *branch-and-bound* framework and on convex relaxations. The algorithm is designed to solve the problem at hand with very low complexity in practice compared to an exhaustive search. This is the first algorithm that addresses optimal video rate control, relay selection, and power control in interference-limited wireless networks. In this section, we present details of the proposed algorithms, B&B framework, convex-relaxation, local search, and variable partition.

¹According to (23) and (28), the worst-case value of P_{dly}^s in (24) can be as large as e^{-1} , and the resulting individual video quality might become unacceptable. To enforce constraints on the quality of each individual user, one may add a constraint for each session imposing that the packet loss rate be lower than a given threshold. In this case, the resulting optimization problem might be infeasible, i.e., there may be instances in which the worst-case performance cannot be guaranteed for all users at the same time.

A. Overview of the Proposed Algorithm

We develop a nonheuristic method for global optimization of the problem introduced in Section III. The proposed algorithm searches for a globally optimal solution with predefined precision of optimality [25]. If we denote the globally optimal sum-PSNR objective function as f^* , $0 < \epsilon \leq 1$ as the optimality precision, then the algorithm searches for an ϵ -optimal solution f , which satisfies $f \geq \epsilon f^*$, with ϵ being arbitrarily close to 1.²

Denote $\mathcal{Q}_0 = \{\mathbf{P}, \mathbf{R}, \boldsymbol{\alpha}\}$ as the original search space, including all possible combinations of video rate control, power allocation, and relay selection. The proposed algorithm maintains a set of subdomains $\mathcal{Q} = \{\mathcal{Q}_i \subset \mathcal{Q}_0, i = 1, 2, \dots\}$, where i represents the iteration step of the algorithm. For any \mathcal{Q}_i , consider $\text{UP}(\cdot)$ and $\text{LR}(\cdot)$ as the upper and lower bounds on sum-PSNR over \mathcal{Q}_i . We refer to $\text{UP}(\mathcal{Q}_i)$ and $\text{LR}(\mathcal{Q}_i)$ as the local upper bound and local lower bound, respectively.

The B&B framework requires that, for given \mathcal{Q}_i , the $\text{UP}(\mathcal{Q}_i)$ and $\text{LR}(\mathcal{Q}_i)$ should be easy to calculate. To determine $\text{UP}(\cdot)$, we rely on relaxation, i.e., we relax the original nonconvex combinatorial problem into two convex problems assuming that: 1) the network is interference-free; and 2) relay assignment is fixed. Because of these two assumptions, the solution of the relaxed convex problem may not be feasible, e.g., the video encoding rate cannot be supported by the underlying network. For $\text{LR}(\cdot)$, we locally search for a *feasible solution* starting from the relaxed solution and set the corresponding sum-PSNR as the local lower bound. The convex-relaxation method and local search strategy will be described in detail in Sections IV-B and IV-C, respectively.

The proposed algorithm searches for the ϵ -optimal solution iteratively. At each iteration, the algorithm maintains a global upper bound UP_{glb} and a global lower bound LR_{glb} on the sum-PSNR such that

$$\text{LR}_{\text{glb}} \leq f^* \leq \text{UP}_{\text{glb}}. \quad (34)$$

At the beginning, i.e., $i = 0$, the set of subdomains \mathcal{Q} is initialized to $\{\mathcal{Q}_0\}$, i.e., $\mathcal{Q} = \{\mathcal{Q}_0\}$, and UP_{glb} and LR_{glb} are initialized to be $\text{UP}(\mathcal{Q}_0)$ and $\text{LR}(\mathcal{Q}_0)$, respectively. The algorithm partitions \mathcal{Q}_0 into two subdomains. For example, by assuming that relay 1 is assigned to session 2, \mathcal{Q}_0 can be divided into $\mathcal{Q}_1 = \{\boldsymbol{\alpha}, \mathbf{P}, \mathbf{R} \mid \alpha_{2,1} = 1\}$ and $\mathcal{Q}_2 = \{\boldsymbol{\alpha}, \mathbf{P}, \mathbf{R} \mid \alpha_{2,1} = 0\}$. Details of the partition strategy will be discussed in Section IV-D. For \mathcal{Q}_i , $i = 1, 2$, the algorithm calculates $\text{UP}(\mathcal{Q}_i)$ and $\text{LR}(\mathcal{Q}_i)$, respectively. If $\text{UP}(\mathcal{Q}_i) < \text{LR}_{\text{glb}}$, this indicates that the globally optimal solution f^* is not located in \mathcal{Q}_i . Hence, \mathcal{Q}_i is removed from \mathcal{Q} . Then, the algorithm updates the global upper and lower bounds as follows:

$$\text{UP}_{\text{glb}} = \max\{\text{UP}(\mathcal{Q}_i), i = 1, 2\} \quad (35)$$

$$\text{LR}_{\text{glb}} = \max\{\text{LR}(\mathcal{Q}_i), i = 1, 2\}. \quad (36)$$

We use UP_{glb} to drive the branch and bound technique and use LR_{glb} to check how close the obtained solution is to f^* and

²The proposed algorithm itself is optimal in the sense that the optimal algorithm can be obtained by setting ϵ arbitrarily close to 1. However, this may result in increased time and space complexity of the algorithm. In practice, ϵ is usually set to a value smaller than 1, and we call the resulting solution ϵ -optimal solution. With the above understanding, in this paper we use optimal for general description of the algorithm, while we use ϵ -optimal to refer to a specific solution.

decide when to terminate the algorithm. If $\text{LR}_{\text{glb}} \geq \epsilon \cdot \text{UP}_{\text{glb}}$, the algorithm terminates and sets the optimal sum-PSNR to $f = \text{LR}_{\text{glb}}$. Otherwise, the algorithm chooses one subdomain from \mathcal{Q} and further partitions it into two subdomains, calculates $\text{UP}(\cdot)$ and $\text{LR}(\cdot)$, and updates the UP_{glb} and LR_{glb} as in (35) and (36). In our algorithm, we select the \mathcal{Q}_i with the highest local upper bound from \mathcal{Q} , i.e.,

$$i = \arg \max_i \text{UP}(\mathcal{Q}_i). \quad (37)$$

As the domain-partition progresses, the algorithm converges to the optimal sum-PSNR f . This can be guaranteed by the following two properties of our B&B method.³

- 1) As $i \rightarrow \infty$, the measure of $\mathcal{Q}_i \in \mathcal{Q}$ goes to 0, and the transmission strategy in \mathcal{Q}_i becomes fixed. For example, as the partition progresses, more relay selection variables $\alpha_{s,r}$ become fixed to 0 or 1, and the allowed transmission power for P_s and P_r will be limited to a domain of smaller measure. As $i \rightarrow \infty$, each $\mathcal{Q}_i \in \mathcal{Q}$ contains only fixed relay selection, power allocation, and video encoding rate.
- 2) As the measure of $\mathcal{Q}_i \in \mathcal{Q}$ goes to 0, the gap between $\text{UP}(\mathcal{Q}_i)$ and $\text{LR}(\mathcal{Q}_i)$ also approaches 0. In Section IV-B, the original problem is relaxed to a standard convex problems, whose upper bound $\text{UP}(\mathcal{Q}_i)$ decreases monotonically with decreasing measure of \mathcal{Q}_i . As $i \rightarrow \infty$, the $\text{UP}(\mathcal{Q}_i)$ converges to $\text{LR}(\mathcal{Q}_i)$ with a fixed transmission strategy.

Based on the update criterion of UP_{glb} and LR_{glb} in (35) and (36), the gap between UP_{glb} and LR_{glb} converges to 0. Furthermore, from (34), UP_{glb} and LR_{glb} converge to the globally maximal sum-PSNR f^* .

B. Convex Relaxation

In this section, we derive a relaxation of the original problem through convex relaxations of nonconvex constraints. Through the proposed relaxation, at each iteration of the algorithm, the problem can be solved in polynomial time on a restricted domain using standard convex optimization techniques to provide an upper bound on the sum-PSNR.

First, we need to state the following proposition.

Proposition 1: The objective function in (32) is a concave function of video encoding rate R and link capacity C .

Proof: We need to show that the individual PSNR is a concave function of the video encoding rate R and link capacity C . Then, the sum of concave functions is a concave function. For the sake of conciseness, denote the total distortion as $f(R, C)$ and set $g(R, C) = \ln f(R, C)$. Since $\text{PSNR} = 10 \log_{10} D_{\text{max}}/f$, we have $\text{PSNR} = 10 \log_{10} D_{\text{max}} - \frac{10}{\ln 10} g(R, C)$. Hence, we only need to show that $g(R, C)$ is convex. This can be proven based on the property that a function is convex if and only if it is convex when restricted to any line in its domain [37]. ■

Based on: 1) Proposition 1; 2) the fact that PSNR is nondecreasing with respect to C_s ; and 3) the composition property that preserves convexity [37, Ch. 3.2.4, p. 83], to relax the original problem to a convex problem, we only need to relax C_s such that it is a concave function of P_s, P_r and $\alpha_{s,r}$, and hence the queuing delay constraint in (29) becomes convex.

From the link capacity model in Section III-A, we can see that the expression of the overall link capacity, defined in (9)–(19),

is rather convoluted. To simplify the convex relaxation, we designed a two-stage method that results in two mutually exclusive relaxations depending on the associated domain partition. In each stage, the original problem is relaxed to be a convex optimization problem, and a solution (which might be infeasible and is referred to as relaxed solution) can be obtained by solving the problem. This relaxed solution will then be used as a starting point to search for a feasible solution. The key idea of the relaxation method can be intuitively illustrated as follows.⁴

- 1) *Stage 1:* For a given subdomain \mathcal{Q}_i , if the relay selection strategy $\alpha_{s,r}$ in \mathcal{Q}_i is not fixed for all s and r , we only relax $\alpha_{s,r}$, but assume that the network is interference-free. Denote the relaxation in this stage as RLX_1 .
- 2) *Stage 2:* Given a subdomain \mathcal{Q}_i , if the relay selection strategy $\alpha_{s,r}$ in \mathcal{Q}_i is fixed for all s and r , we relax P_s and P_r . Denote the relaxation in this stage as RLX_2 .

Next, RLX_1 is taken as an example to show how to relax the original problem formulated in Section III to be convex, while relaxation for RLX_2 can be performed similarly.

RLX₁: Based on the assumption of interference-free network, each node can work at its maximum transmission power, and the link capacity is a function of $\alpha_{s,r}$ only. Then, the link capacity from source to destination C_{s2d}^s defined in (10), from source to relay C_{sr}^s defined in (15), and the combined capacity C_{sr2d}^s defined in (19), can be relaxed (by ignoring the effect of interference) as

$$C_{s2d}^s = B \log_2 (1 + \chi G_{s2d}^{s,s} P_{\text{max}}^s / \sigma^2) \quad (38)$$

$$C_{sr}^s = B \log_2 (1 + \chi P_{\text{tmp}}^s / \sigma^2) \quad (39)$$

$$C_{sr2d}^s = B \log_2 (1 + \chi (G_{s2d}^{s,s} P_{\text{max}}^s + P_{\text{tmp}}^r) / \sigma^2) \quad (40)$$

respectively, where $P_{\text{tmp}}^s = \sum_{r \in \mathcal{R}} G_{s2r}^{s,r} P_{\text{max}}^s \alpha_{s,r}$ and $P_{\text{tmp}}^r = \sum_{r \in \mathcal{R}} G_{r2d}^{r,s} P_{\text{max}}^r \alpha_{s,r}$.

The cooperative link capacity C_{coop}^s , defined in (14), can be expressed with two constraints as

$$C_{\text{coop}}^s \leq \frac{1}{2} C_{s2r}^s \quad (41)$$

and

$$C_{\text{coop}}^s \leq \frac{1}{2} C_{sr2d}^s. \quad (42)$$

Notice that the C_{coop}^s is nondecreasing with C_{s2r}^s and C_{sr2d}^s .

We also need to relax the overall link capacity C_s defined in (9). We first relax the relay selection strategy by allowing each relay to be assigned to multiple video sessions, and vice versa, allowing each video session to use multiple relays. Then, the constraint in (1) can be rewritten as

$$\alpha_{s,r}^{\min} \leq \alpha_{s,r} \leq \alpha_{s,r}^{\max} \quad \forall s \in \mathcal{S}, r \in \mathcal{R} \quad (43)$$

where $\alpha_{\min}^{s,r}$ and $\alpha_{\max}^{s,r}$ are the lower and upper bounds on $\alpha_{s,r}$, respectively. At the first iteration, they are set to

$$\alpha_{s,r}^{\min} = 0 \quad \text{and} \quad \alpha_{s,r}^{\max} = 1 \quad \forall s \in \mathcal{S}, r \in \mathcal{R}. \quad (44)$$

Denote $C_{\text{coop},s}^{\max}$, $C_{\text{coop},s}^{\min}$ as an upper and lower bound for C_{coop}^s , respectively, and further denote the nonlinear item in (9), $\alpha_{s,r} C_{\text{coop}}^s$ with $\beta_{\text{coop}}^{s,r}$. Then, according to the *reformulation*

⁴Tighter and uniform relaxation for both cases can be obtained (at the expense of higher complexity) to solve the resulting relaxed problem. The reader is referred to [38] for additional details of relaxation techniques.

³A formal proof of the convergence of branch and bound can be found in [25].

and linearization technique (RLT) [38], $\beta_{\text{coop}}^{s,r}$ can be relaxed using four linear constraints as

$$(\alpha_{s,r}^{\max} - \alpha_{s,r}) (C_{\text{coop},s}^{\max} - C_{\text{coop}}^s) \geq 0 \quad (45)$$

$$(\alpha_{s,r}^{\max} - \alpha_{s,r}) (C_{\text{coop}}^s - C_{\text{coop},s}^{\min}) \geq 0 \quad (46)$$

$$(\alpha_{s,r} - \alpha_{s,r}^{\min}) (C_{\text{coop},s}^{\max} - C_{\text{coop}}^s) \geq 0 \quad (47)$$

$$(\alpha_{s,r} - \alpha_{s,r}^{\min}) (C_{\text{coop}}^s - C_{\text{coop},s}^{\min}) \geq 0. \quad (48)$$

The lower bound for C_{coop}^s can be simply set to be $C_{\text{coop},s}^{\min} = 0$. An upper bound can be obtained as

$$C_{\text{coop},s}^{\max} = \frac{1}{2} \max (C_{s2r,s}^{\max}, C_{sr2d,s}^{\max}) \quad (49)$$

where $C_{s2r,s}^{\max} = B \log_2(1 + \chi P_{\text{tmp},s}^{\max}/\sigma^2)$ and $C_{sr2d,s}^{\max} = B \log_2(1 + \chi(G_{s2d}^{s,s} P_{\text{max}}^s + P_{\text{tmp},r}^{\max})/\sigma^2)$, with $P_{\text{tmp},s}^{\max} = \max_{r \in \mathcal{R}} (G_{s2r}^{s,r} P_{\text{max}}^s \alpha_{s,r}^{\max})$, $\forall s \in \mathcal{S}$ and $P_{\text{tmp},r}^{\max} = \max_{r \in \mathcal{R}} (G_{r2d}^{r,s} P_{\text{max}}^r \alpha_{s,r}^{\max})$, $\forall s \in \mathcal{S}$.

So far, we have completed the convex relaxation of stage RLX_1. Consequently, the overall link capacity C_s in (9) can be expressed as

$$C_s = \left(1 - \sum_{r \in \mathcal{R}} \alpha_{s,r}\right) C_{s2d}^s + \sum_{r \in \mathcal{R}} \beta_{\text{coop}}^{s,r} \quad (50)$$

and we have following proposition.

Proposition 2: The original problem formulated in (30)–(33) is relaxed to a convex optimization problem in standard form.

Proof: Through relaxation, we have the following.

- i) C_s in (50) is linear function of $\alpha_{s,r}$, C_{s2d}^s , and $\beta_{\text{coop}}^{s,r}$.
- ii) C_{s2d}^s in (38) is a constant.
- iii) $\beta_{\text{coop}}^{s,r}$ defined in (45)–(48) is constrained by $\alpha_{s,r}$ and C_{coop}^s through four linear functions.
- iv) C_{s2r}^s in (39) and C_{sr2d}^s in (40) are concave functions of $\alpha_{s,r}$.
- v) The two linear constraints, (41) and (42), result in a convex domain.

Together with Proposition 1, the relaxed problem is a convex optimization problem in standard form. ■

C. Local Search Method

At each iteration of our algorithm, we solve the RLX_1 or RLX_2 using interior-point algorithms [39]. However, the optimal relaxed solution may not be feasible. For example, in the case of RLX_1, the optimal $\alpha_{s,r}^*$ is likely to take an intermediate value between 0 and 1. As introduced in Section IV-B, we refer to the ϵ -optimal solution found in this way as “relaxed solution,” and we locally search for a feasible solution starting from the relaxed solution. The local search is based on the fact that the power allocation in the relaxed solution is always feasible. A feasible value of $\alpha_{s,r}$ can be set to its closest integer, i.e.,

$$\alpha_{s,r} = \text{round}(\alpha_{s,r}^*). \quad (51)$$

Note that because of the constraints in (2) and (3), there exists at most one r over all $s \in \mathcal{S}$, and one s over all $r \in \mathcal{R}$, such that $\text{round}(\alpha_{s,r}^*) = 1$. Then, a feasible power allocation can be obtained by setting P_r to zero for all relay nodes that are not selected by any sessions, i.e., $P_r = 0$ for each r with $\sum_{s \in \mathcal{S}} \alpha_{s,r} = 0$.

Given a feasible power allocation and relay selection, a feasible C_{s2d}^s , C_{coop}^s , and further C_s can be calculated using (10), (19), and (9), respectively.

Given C_s , the optimal video encoding rate R_s for session s can be calculated by solving the following convex problem:

$$\begin{aligned} & \text{Find : } R_s \quad \forall s \in \mathcal{S} \\ & \text{Maximize } \sum_{s \in \mathcal{S}} \text{PSNR}_s \\ & \text{Subject to : } (26), (29). \end{aligned} \quad (52)$$

If the above problem is feasible, the optimal sum-PSNR is set as the local lower bound $\text{LR}(\mathcal{Q}_i)$. Otherwise, it indicates that, for the present power allocation and relay strategy, there does not exist a video encoding rate $\{R_s, s \in \mathcal{S}\}$ satisfying the constraints of minimum video encoding rate and maximum queuing delay for all video sessions. In this case, the local lower bound is set to 0.

D. Domain Partition

Because of the relaxation in RLX_1 and RLX_2, in general there will be a performance gap between the local upper and lower bounds over each subdomain. In our algorithm, this gap is iteratively decreased by iteratively partitioning each subdomain into two smaller subdomains.

We first need to select a subdomain from \mathcal{Q} to be partitioned. Considering that the global upper bound is equal to the highest local lower bound, we select the \mathcal{Q}_i with the highest local upper bound to be partitioned, such that the global upper bound becomes smaller as partition progresses. To partition \mathcal{Q}_i into two subdomains, we need to select a variable among $\alpha_{s,r}$, P_s , and P_r , and then partition it from a middle value of its domain. Here, only variables that affect the sum-PSNR directly are partitioned, i.e., no intermediate variables are partitioned.

The variable partition is carried out corresponding to the two stages of relaxation. At the beginning of the algorithm, there is only one domain \mathcal{Q}_0 in \mathcal{Q} , and the relay selection is not fixed. In this case, the RLX_1 is employed to calculate the local upper bound. Denote $\alpha_{s,r}^*$ as the relaxed solution to $\alpha_{s,r}$. Then, α_{s^*,r^*} is selected for partition over all $s \in \mathcal{S}$ and $r \in \mathcal{R}$ such that

$$\alpha_{s^*,r^*} = \max_{s \in \mathcal{S}, r \in \mathcal{R}} \alpha_{s,r}^* \quad (53)$$

where $\alpha_{s,r}^*$ represent the original relaxed solution that have not been fixed through the rounding operation in (51). Since a higher value of $\alpha_{s,r}$ indicates that relay r is more likely to be assigned to video session s , the above criterion fixes the relay assignment decision with the highest value of $\alpha_{s,r}$. Recalling the relay selection constraints (1)–(3), i.e., each session can select at most one relay and each relay can be selected by one session, the selected α_{s^*,r^*} can be partitioned by setting

$$\alpha_{\text{max}}^{s^*,r^*} = 0 \quad \forall r \in \mathcal{R}, r \neq r^* \quad (54)$$

$$\alpha_{\text{max}}^{s,r^*} = 0 \quad \forall s \in \mathcal{S}, s \neq s^* \quad (55)$$

$$\alpha_{\text{min}}^{s^*,r^*} = 1 \quad (56)$$

for the first new subdomain and

$$\alpha_{\text{max}}^{s^*,r^*} = 0 \quad (57)$$

for the second new subdomain. The local upper and lower bound on sum-PSNR over two new subdomains are calculated by using RLX_1 and then local search, respectively. Finally, the two new

subdomains are added into \mathcal{Q} or pruned according to the prune strategy described in Section IV-A.

After a certain number of variable partitions, in the selected \mathcal{Q}_i the relay selection strategy could be fixed. In this case, \mathcal{Q}_i is partitioned by partitioning P_s or P_r , and RLX_2 is employed to calculate the local upper bound over the two new subdomains. The partition can be carried out as follows. Denote

$$\Delta P_{s^*} = \max_{s \in \mathcal{S}} \{P_{\max}^s - P_{\min}^s\} \quad (58)$$

$$\Delta P_{r^*} = \max_{r \in \mathcal{R}} \{P_{\max}^r - P_{\min}^r\}. \quad (59)$$

If $\Delta P_{s^*} \geq \Delta P_{r^*}$, P_{s^*} is partitioned by setting

$$P_{\max}^{s^*}|_{\text{new}} = P_{\min}^{s^*} + \Delta P_{s^*} / 2 \quad (60)$$

for the first new subdomain, and

$$P_{\min}^{s^*}|_{\text{new}} = P_{\min}^{s^*} + \Delta P_{s^*} / 2 \quad (61)$$

for the second new subdomain. Otherwise, the P_{r^*} is partitioned similarly. Here, $P_{\max}^{s^*}|_{\text{new}}$ and $P_{\min}^{s^*}|_{\text{new}}$ mean the upper and lower bounds for P_s in the new subdomains, respectively.

It is worth pointing out that variables can also be selected and partitioned in alternative ways. For example, one could select the $\alpha_{s,r}$ that is the least likely to be assigned, or partition P_s from its relaxed solution. Moreover, we can also first partition P_s or P_r and then partition $\alpha_{s,r}$. However, different variable partition methods do not affect the optimality of our algorithm, while they may affect the convergence speed. Alternative partition methods will be investigated in our future work to seek for faster convergence. We also notice that as the partition progresses, the relay selection strategy becomes fixed in each subdomain. Moreover, the gap between the upper and lower bounds for each P_s and P_r converges to 0. That is, the proposed algorithm has guaranteed convergence.

So far, we have designed an optimal algorithm for decode and forward cooperation based on a combination of the branch and bound framework and convex relaxation techniques. As stated in Section III, this can be (nontrivially) extended to the amplify and forward case, where the link capacity, denoted as C_{AF} , can be expressed as $C_{AF} = \frac{W}{2} \log_2(1 + \text{SINR}_{sd} + \frac{\text{SINR}_{sr} \cdot \text{SINR}_{rd}}{\text{SINR}_{sr} + \text{SINR}_{rd} + 1})$ with SINR_{sd} , SINR_{sr} , and SINR_{rd} representing the SINR corresponding to the links from source to destination, source to relay, and relay to destination, respectively. Since the expression of C_{AF} is rather complicated and nonconcave in general, it requires the design of an *ad hoc* convex relaxation method. This can be done by applying our proposed two-stage relaxation method again to relax the corresponding objective function and constraints so that they result in a convex optimization problem. Then, at each stage, the RLT can also be applied to relax those nonconvex items using a set of linear constraints.

V. DISTRIBUTED ALGORITHM

In this section, we propose a distributed solution algorithm for solving the problem formulated in Section III. Then, in Section VI, we evaluate the distributed algorithm by comparing it to the centralized algorithm proposed in Section IV.

The distributed algorithm is designed to achieve an NE point in the feasible domain [40]. The Nash equilibrium is a

well-known notion from noncooperative game theory often used as a tool for designing distributed algorithms in complex wireless communication systems [41]–[43]. There are two important characteristics of an NE solution: 1) At any NE solution point, no user has incentives to deviate from the current transmission strategy unilaterally; and 2) each user's utility is maximized, given the transmission strategies of all other users.

Considering that in cooperative wireless networks, the interference among video sessions depends on both power control and relay selection strategy, we design our distributed algorithm assuming that each source and relay node always transmits at maximum transmit power and relies on cooperative relaying for interference reduction. Then, we apply the distributed algorithm to different wireless networks and evaluate its performance by comparing it to the proposed centralized ϵ -optimal solution algorithm in Section IV.

With fixed maximum transmission power, the problem of video rate control and relay selection can be formulated as a game, in which each video session is a player. Each player optimizes its own video quality by selecting the best relay node and also deciding the optimal video encoding rate. Denote the vector of relay selection strategies for video session $s \in \mathcal{S}$ as $\alpha_s = (\alpha_{s,r})$, $r \in \mathcal{R}$, and the vector of relay selection strategy for all video sessions except s as $\alpha_{-s} = (\alpha_{w,r})$, $r \in \mathcal{R}$, $w \in \mathcal{S}$, $w \neq s$. Further define the vector of video encoding rates of all sessions except s as $\mathbf{R}_{-s} = (R_w)$, $w \in \mathcal{S}$, $w \neq s$. Then, the objective of the game is to find an NE solution for relay selection, denoted as $\alpha^* \triangleq (\alpha_s^*, \alpha_{-s}^*)$, and video encoding rate, denoted as \mathbf{R}^* , such that for each video session, say $s \in \mathcal{S}$

$$(\alpha_s^*, \mathbf{R}_s^*) = \arg \max_{\alpha_s, R_s} \text{PSNR}_s(\alpha_s, R_s, \alpha_{-s}^*, \mathbf{R}_{-s}^*) \quad (62)$$

Subject to : (1)–(3), (26), (29).

To this end, we propose an iterative best-response-based algorithm. In each iteration, each node locally optimizes its own video quality by solving the optimization problem in (62), with given fixed video encoding rate and relay selection strategy for all other video sessions. Each node continues to locally optimize its transmission strategy until any deviation from its last solution would imply a decrease in its performance.

Recall that the objective function of PSNR_s is a concave function of video encoding rate and underlying link capacity. With given fixed transmission power of all source and relay nodes, and also relay selection strategy for all other video sessions except s , the overall link capacity for video session s in (9) can be reformulated as a linear function of relay selection variable $\alpha_{s,r}$, $r \in \mathcal{R}$

$$C_s = \left(1 - \sum_{r \in \mathcal{R}} \alpha_{s,r}\right) C_{s2d}^s + \sum_{r \in \mathcal{R}} \alpha_{s,r} C_{\text{coop}}^{s,r} \quad (63)$$

where C_{s2d}^s and $C_{\text{coop}}^{s,r}$ are constant and represent the capacity of direct link and cooperative link if relay node r is used by video session s , respectively. Since composition preserves convexity [37], the objective function of PSNR_s is also a concave function of $\alpha_{s,r}$, $r \in \mathcal{R}$.

Considering that PSNR_s is also a monotonically increasing function of the underlying link capacity, maximizing PSNR_s can be decomposed into two subproblems: 1) maximizing link

capacity by solving a linear optimization problem with objective function defined in (63); and 2) maximizing PSNR_s by solving a convex optimization as follows:

$$\begin{aligned}
 & \text{Find : } R_s \\
 & \text{Maximize PSNR}_s \\
 & \text{Subject to : } (26), (29). \tag{64}
 \end{aligned}$$

Without loss of generality, we assume that in the first subproblem, $C_{s2d}^s \neq C_{\text{coop}}^{s,r}$, $r \in \mathcal{R}$ and $C_{\text{coop}}^{s,r} \neq C_{\text{coop}}^{s,q}$, $r, q \in \mathcal{R}$, $r \neq q$. Then, constraint in (1) is equivalent to $0 \leq \alpha_{s,r} \leq 1$, $s \in \mathcal{S}$, $r \in \mathcal{R}$. Hence, constraints in (1)–(3), (26), and (29) form a convex domain set, and the problem formulated in (62) is a convex optimization problem. This implies that in each iteration of the distributed algorithm, each node only needs to solve a convex optimization problem, which can be done in polynomial time through well-established techniques (i.e., interior-point algorithms [39]).

Theorem 1: The best-response-based distributed solution algorithm converges to an NE solution of relay selection game if the number of relay nodes is large compared to the number of video sessions and any two sessions in \mathcal{S} are located sufficiently far away from each other. In this case, the relay selection game, which is in general a quasi Nash equilibrium (QNE) problem, reduces to an NE problem.

Proof: Proof of the convergence to NE is shown in Appendix-B. ■

It is worth pointing out that the resulting NE is not necessarily an efficient working point. In our future research, we will seek to derive algorithms designed to provably select efficient Nash equilibria. In this paper, we verify the high efficiency of the solution obtained by our distributed algorithm via simulation. In Theorem 1, it is reasonable to assume large number of relay nodes compared to the number of video sessions. Since video transmission usually requires high-capacity links, within a given communication area and a portion of the spectrum, the number of concurrent video sessions must be limited. Additionally, the theorem provides a sufficient condition that guarantees convergence of the distributed relay selection algorithm to an NE with a limit on the effect of interference. Although for an arbitrarily given cooperative network it is usually not easy for the sufficient condition to be satisfied, the proposed distributed algorithm performs very well even when there is interference, which will be verified through numerical results in Section VI.

VI. SIMULATION RESULTS

A. System Setup

In this section, performance evaluation results are presented for the proposed centralized and distributed algorithms. We consider a single-hop wireless communication network with area of $1000 \times 600 \text{ m}^2$. Different network scales are considered including N nodes, with $N = 10, 20, 30, 50$, and 100 . Nodes are randomly distributed in the communication area. The path-loss coefficient between node s and node d is given by $G_{s,d}^2 = \|s - d\|^{-4}$, where $\|s - d\|$ represents the distance [m] between the two nodes, and 4 is the path-loss factor. AWGN noise power is set to 10^{-7} mW for each node. The bandwidth of the available spectrum is set to $B = 200$ kHz, and the spreading gain is set to $\chi = 10$. The maximum transmission power for each source

TABLE I
VIDEO PARAMETERS

| Video | D_0 | θ | R_0 | P_{err} | k | T_0 | L |
|-------|-------|----------|-------|-----------|-----|-------|------|
| FM | 0.38 | 2537 | 18.3 | 0.01 | 750 | 350 | 3040 |
| MD | 0 | 857 | 0.67 | 0 | 30 | 350 | 3040 |

and relay node is set to 1000 mW. Two video sequences are considered, *Foreman* (FM) and *Mother and Daughter* (MD), which are characterized by intense and moderate rate variability. We use the mathematical model described in Section III-B to calculate the PSNR with given video parameters and link capacity, which can be calculated with given SINR. The parameters of the two video sequences are reported in Table I, where the units of R_0 , T , and L are kb/s, ms, and bits, respectively. Moreover, we implement a slot-level simulator using MATLAB and perform the proposed algorithms for each given network topology.⁵ We first show convergence and complexity analysis for the proposed centralized and distributed algorithms. Then, we present example results of rate control, relay selection, and power control. Finally, we compare the performance of the two algorithms. All results presented in this section are obtained by averaging over 30 independent simulations.

B. Convergence and Complexity Analysis

Convergence performance of the proposed ϵ -optimal algorithm is illustrated in Fig. 1(a) for different video sessions, total number of nodes, and optimality precision. Network parameters are set to (FM, MD, 30, 98%)⁶ for the top figure, (FM, MD, FM, 30, 95%) for the middle figure, and (FM, MD, 100, 90%) for the bottom figure. The optimality precision is defined as the ratio of global lower bound on sum-PSNR to the global upper bound. In all three cases, the centralized algorithm converges to the predefined optimality precision.

Complexity of the terms of the ϵ -optimal algorithm is evaluated in average number of iterations, and accumulation probability of required number of iterations is shown in Fig. 1(b) in the case of (FM, MD, 30, 95%). We can see that the number of iteration varies from 7 to 3000 with an average of 752. We observe that around 80% of simulations can be finished in less 500 iterations. Recall that our algorithm searches for the ϵ -optimal solution by partitioning the original domain into a series of subdomains. Since there are two video sessions and 30 nodes, we have two source nodes and 26 potential relay nodes, and then the number of relay selection variables $\alpha_{s,r}$, $s \in \mathcal{S}$, $r \in \mathcal{R}$, is $2 \times 26 = 52$. Quantize the maximum transmission power of 1000 mW using a step of 100 mW, then each source or relay node has 10 possible choices of transmission power. Then, total number of possible transmission strategies can be calculated as $2^{52} \times (2 + 26)^{10} \approx 1.3 \times 10^{30}$, which is too large to search the original domain exhaustively. Therefore, compared to exhaustive search, the proposed centralized algorithm is quite efficient in computational complexity.

⁵In this paper, we focus on networks whose topologies vary slowly so that it suffices to perform the algorithms for a time duration much longer than each cooperative transmission.

⁶(FD, MD, 30, 98%) refers to two sessions transmitting FD and MD, respectively, $N = 30$ and $\epsilon = 98\%$. Similarly, (FM, MD, FM, 30, 95%) indicates three video sequences transmitting FM, MD, and FM, respectively, $N = 30$ and $\epsilon = 95\%$.

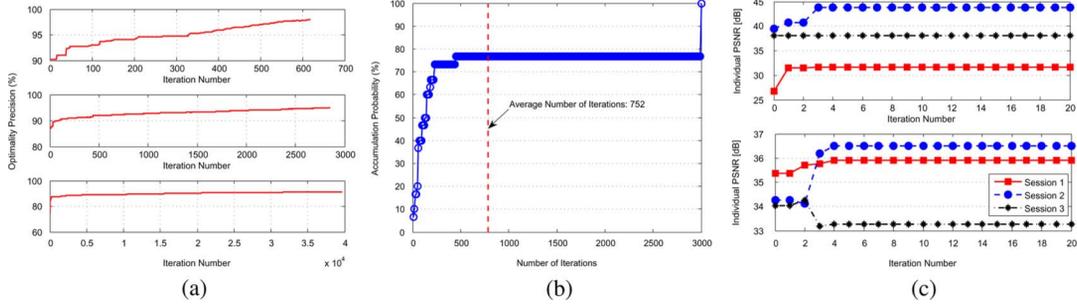


Fig. 1. (a) Convergence performance of the centralized algorithm. Parameters of video sequences, total number of nodes N , and maximum optimality precision ϵ are (top) (FM, MD, 30, 98%), (middle) (FM, MD, FM, 30, 95%), and (bottom) (FM, MD, 100, 90%). (b) Complexity performance of the centralized algorithm in the case of (FM, MD, 30, 95%), measured in accumulation probability of required number of iterations to converge. (c) Convergence performance of the distributed algorithm given in individual PSNR. Parameters are (FM, MD, FM, 20) for top figure and (FM, MD, FM, 50) for bottom figure.

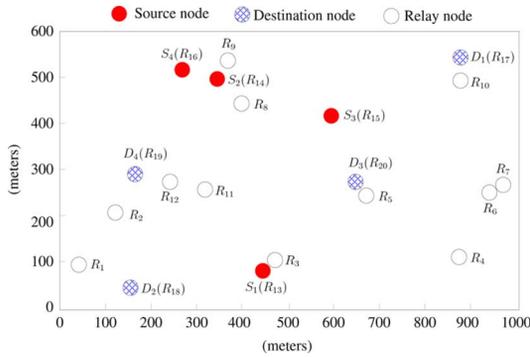


Fig. 2. Example of a 20-node network.

Convergence performance of the proposed distributed algorithm is evaluated with three video sessions, FM, MD, and FM, and in cases of total number of nodes of 20 and 50, respectively. Results indicate that the distributed algorithm converges very fast, i.e., within three iterations on average in both cases. Two examples of the convergence curves are shown in Fig. 1(c) for 20 (top) and 50 (bottom) nodes, respectively; we can see that the distributed algorithm converges in less than five iterations in the examples.

C. Example Results of Resource Allocation

We now discuss example results of the two algorithms with fixed network topology shown in Fig. 2.

1) *Example of Achieved Sum-PSNR*: First, we show two examples of the achieved sum-PSNR obtained using ϵ -optimal solution algorithm. In the first experiment, there is only one video session (S_1, D_1), which transmits video sequence FM. The result of the execution of the proposed algorithm is shown in Fig. 3(a). In this case, the ϵ -optimal PSNR is 35.2 dB. R_{20} is selected by S_1 as the relay node. The transmission powers of S_1 and R_{20} are 1000 and 1000 mW, respectively. The achieved link capacity and corresponding optimal video encoding rate are 306.4 and 241.9 kb/s, respectively. In the second experiment, there are two video sessions (S_i, D_i), $i = 1, 2$, which transmit the video sequences FM and MD, respectively. Precision of optimality is set to $\epsilon = 98\%$. In this case, the ϵ -optimal sum-PSNR is 79.38 dB. The individual PSNRs for the two video sessions are 34.9 and 44.4 dB, respectively. R_{20} is selected by S_1 as the relay node, while R_{12} is selected by S_2 as the relay node. The transmission powers of S_1, S_2, R_{20}, R_{12} are 561, 998, 980, and

TABLE II
EFFECTS OF RELAY SELECTION, POWER ALLOCATION, VIDEO RATE CONTROL
IN THE CASE OF TWO VIDEO SESSIONS

| | Source Power (mW) | | | Relay Node | Relay Power (mW) | | |
|---------|-------------------|------------|--|-----------------------------------|-----------------------------------|--|--|
| B&B | 1000 | 1000, 1000 | | R_4, R_{11}, \emptyset | 1000, 1000, \emptyset | | |
| woPC: 1 | 1000, 1000, 1000 | | | R_4, R_{11}, \emptyset | 1000, 1000, \emptyset | | |
| woPC: 2 | 500, 500, 500 | | | R_4, R_{12}, \emptyset | 1000, 1000, \emptyset | | |
| woPC: 3 | 1000, 1000, 1000 | | | R_4, R_{12}, \emptyset | 500, 500, \emptyset | | |
| woPC: 4 | 500, 500, 500 | | | R_4, R_2, \emptyset | 500, 500, \emptyset | | |
| woRS: 1 | 887, 889, 325 | | | $\emptyset, \emptyset, \emptyset$ | $\emptyset, \emptyset, \emptyset$ | | |
| woRS: 2 | 986, 967, 1000 | | | $R_5, \emptyset, \emptyset$ | 100, \emptyset, \emptyset | | |
| woRS: 3 | 833, 999, 325 | | | $\emptyset, R_{12}, \emptyset$ | $\emptyset, 986, \emptyset$ | | |
| woRS: 4 | 542, 734, 991 | | | R_5, R_{12}, \emptyset | 454, 101, \emptyset | | |
| woRC: 1 | 1000, 1000, 1000 | | | $\emptyset, \emptyset, \emptyset$ | $\emptyset, \emptyset, \emptyset$ | | |
| woRC: 2 | 945, 889, 216 | | | $R_{11}, R_{19}, \emptyset$ | 381, 606, \emptyset | | |
| woRC: 3 | 1000, 1000, 1000 | | | R_3, R_{19}, \emptyset | 1000, 1000, \emptyset | | |
| woRC: 4 | 606, 831, 493 | | | R_4, R_{11}, \emptyset | 606, 719, \emptyset | | |
| | Capacity (kb/s) | | | Video Rate (kb/s) | Sum PSNR (dB) | | |
| B&B | 140, 310, 1189 | | | 94, 261, 1098 | 112.4 | | |
| woPC: 1 | 140, 310, 1189 | | | 94, 261, 1098 | 112.4 | | |
| woPC: 2 | 112, 334, 1066 | | | 73, 284, 977 | 111.3 | | |
| woPC: 3 | 90, 333, 1272 | | | 56, 283, 1179 | 109.9 | | |
| woPC: 4 | 109, 175, 1139 | | | 70, 137, 1048 | 107.9 | | |
| woRS: 1 | 93, 159, 917 | | | 59, 124, 830 | 106.0 | | |
| woRS: 2 | 72, 211, 80 | | | 45, 170, 49 | 96.5 | | |
| woRS: 3 | 89, 320, 840 | | | 28, 44, 38 | 110.2 | | |
| woRS: 4 | 52, 344, 111 | | | 24, 293, 72 | 97.5 | | |
| woRC: 1 | 72, 162, 1210 | | | 30, 50, 100 | 92.5 | | |
| woRC: 2 | 72, 230, 878 | | | 50, 200, 300 | 104.4 | | |
| woRC: 3 | 69, 351, 1198 | | | 50, 300, 500 | 106.3 | | |
| woRC: 4 | 123, 337, 1069 | | | 100, 300, 1000 | 109.5 | | |

944 mW, respectively. The achieved link capacity and corresponding optimal video encoding rate are 285.9 and 222.9 kb/s for video session 1, and 430.4 and 370.3 kb/s for video session 2.

2) *Effects of Video Encoding Rate, Relay Selection, and Power Allocation*: Next, we examine the effects of relay selection, power allocation, video rate control, as well as the number of video sessions on the sum-PSNR. In this experiment, three video sessions are employed, FM, MD, and FM. Results of the proposed algorithm are compared to three cases: exhaustive search without relay selection (woRS), search without power control (woPC), search without rate control (woRC). Results are shown in Table II.

When the proposed algorithm is employed, R_4 and R_{11} are selected as relay for session 1 and 2, respectively. Session 3 uses direct transmission. All source and relay nodes transmit at the

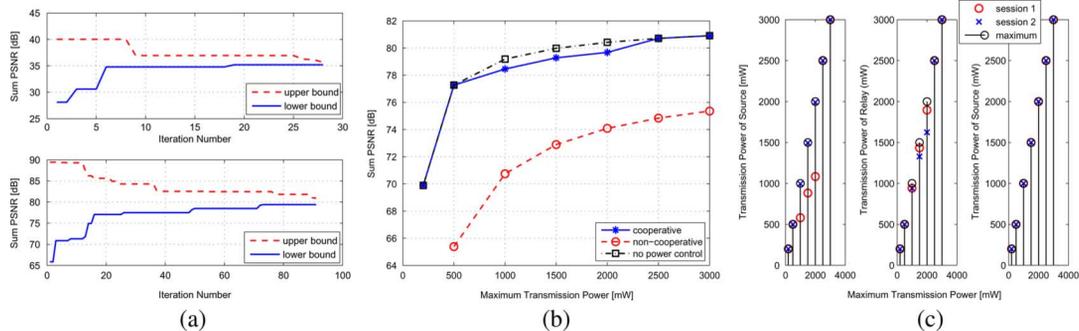


Fig. 3. (a) Upper and lower bounds of SUM PSNR in the case of (*top*) one and (*bottom*) two sessions. (b) Sum PSNR of two video sessions in the case of different maximum transmission power. (c) Optimal transmission power of (*left*) source and (*middle*) relay nodes, and (*right*) source power without power control, in the case of different maximum transmission power.

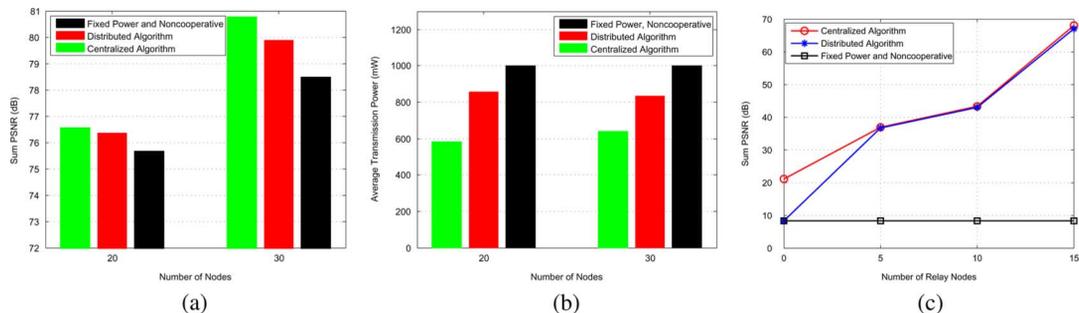


Fig. 4. (a) Performance in sum-PSNR for the proposed centralized and distributed algorithms. (b) Performance in average transmission power for the proposed centralized and distributed algorithms. (c) Sum-PSNR of six video sessions with different number of relay nodes.

maximum power. The achieved sum-PSNR is 112.4 dB. Due to the existence of session 3, R_{20} and R_5 are unsuitable to be relay for session 1. In the case of woPA, we fixed the transmission power of each node to the maximum power. We can see that the sum PSNR decreases as the maximum transmission power decreases. When all nodes transmit at 500 mW, the decrease in the sum-PSNR is about 4.5 dB compared to the proposed algorithm. In this case, it is also desirable for sessions 1 and 2 to use relays.

In the case of woRS, a session selects the node that is located between source and destination as relay. If R_5 is selected, the sum-PSNR decreases considerably. This is because R_5 causes interference to the destination of session 3. If no relay is used, the decrease in sum-PSNR is over 6 dB. Therefore, the relay should be selected considering its interference to other sessions. In the case of woRC, we can see that as the video encoding rates deviate considerably from the optimum, the sum-PSNR decreases significantly.

3) *Effects of Different Number of Video Sessions:* In Table III, we show the effects of the number of video sessions on relay selection, power allocation, and rate control. At most of four video sessions are employed, FM, MD, FM, and MD. We can see that relay selection is greatly influenced by the number of active sessions. For example, when $S = 2$, R_{20} and R_{12} are selected by S_1 and S_2 as relay, respectively. When $S = 3$, S_1 will not select R_{20} any longer because R_{20} will cause high interference to D_3 if selected. Instead, S_1 selects R_4 as the relay. A similar phenomenon happens when session 4 becomes active. From the table, we observe that the nodes usually transmit at the maximum power, except when $S = 2$. Actually, in the case of $S = 2$, if the maximum power is employed, the sum-PSNR increases slightly and also lies within the optimality precision.

TABLE III
COMPARISON OF PROPOSED ϵ -OPTIMAL ALGORITHM IN THE CASE OF DIFFERENT SESSION NUMBER

| S | Source Power (mW) | Relay Node |
|-----|---|---|
| 1 | 1000, -, -, - | R_{20} , -, -, - |
| 2 | 561, 998, -, - | R_{20} , R_{12} , -, - |
| 3 | 1000, 1000, 1000, - | R_4 , R_{11} , \emptyset , - |
| 4 | 1000, 1000, 1000, 1000 | R_4 , \emptyset , \emptyset , \emptyset |
| S | Relay Power (mW) | Capacity (kb/s) |
| 1 | 1000, -, -, - | 306, -, -, - |
| 2 | 980, 944, -, - | 285, 430, -, - |
| 3 | 1000, 1000, \emptyset , - | 140, 310, 1189, - |
| 4 | 1000, \emptyset , \emptyset , \emptyset | 136, 190, 1182, 635 |
| S | Video Rate (kb/s) | Sum (Individual) PSNR (dB) |
| 1 | 242, -, -, - | 35.2 (35.2, -, -, -) |
| 2 | 222, 375, -, - | 79.4 (34.9, 44.4, -, -) |
| 3 | 94, 261, 1098, - | 112.4 (31.6, 42.8, 38.0, -) |
| 4 | 91, 151, 1091, 573 | 156.1 (27.9, 37.1, 37.8, 45.2) |

4) *Effects of Transmission Power:* The effects of transmission power are further explored in Fig. 3(b) and (c). We see that enabling cooperation among nodes can result in a significant improvement in the sum-PSNR. Moreover, when cooperation is enabled, transmitting at the maximum power can lead to a slight increase in sum-PSNR. That is, cooperation can allow nodes to work at lower power without causing much decrease in the sum-PSNR. This is important in wireless multimedia ad hoc and sensor networks, where power efficiency is also of paramount importance.

D. Performance Evaluation for the Distributed Algorithm

Performances for the proposed centralized and also distributed algorithms are shown in Fig. 4(a) in terms of sum-PSNR. Results are obtained by averaging over 30 times

of independent simulations. In the figure, the algorithm with fixed maximum transmission power and without cooperation is employed as bottom-line performance. Two video sessions of FM and MD are employed, and total number of nodes is set to 20 and 30, respectively. We can see that the proposed centralized algorithm can achieve the highest sum-PSNR. In the case of 30 nodes, an improvement of greater than 2 dB can be achieved compared to the algorithm with fixed power and without cooperation. Sum PSNR achieved by the proposed distributed algorithm is very close to the centralized algorithm in the case of 20 nodes. Compared to the algorithm with fixed power and without cooperation, the distributed algorithm can achieve about 1.5 dB of improvement by using cooperative relaying in the case of 30 nodes.

The proposed two algorithms are also evaluated in terms of average transmission power, and corresponding results are shown in Fig. 4(b). From the figure, we can see that the proposed centralized algorithm results in the lowest average transmission power. This is achieved by using cooperative relaying since the transmission time for each source and relay is half compared to the case of direct transmission, and consequently, the total transmission power becomes more evenly distributed over the whole network. This is extremely important in wireless multimedia sensor networks to avoid depleting the battery of a single node and hence to avoid bottleneck links. Average transmission power is also given in Fig. 4(b). We can see that even if the maximum transmission power is used for each source and relay node, the distributed algorithm can achieve more even power consumption through cooperative relaying than in the noncooperative case.

E. Effects of Cooperative Relaying With Higher Interference

In this section, we study effects of cooperative relaying to the video quality in a larger network by setting the number of video sessions to six and 10, respectively. In the first experiment, there are six video sessions with three transmitting FM and the others transmitting MD. Sum PSNR of the six sessions is given in Fig. 4(c) by averaging over 30 times of simulation. Given an arbitrary network topology, it is possible that the minimum rates of all six sessions cannot be guaranteed simultaneously, and the sum PSNR is set to zero in this case. From Fig. 4(c), we can see that when there is no relay node, the distributed algorithm reduced to fixed power and noncooperative, and the centralized ϵ -optimal algorithm performs the best by using power control. As the relay number increases, the effects of power control are reduced, and performance of the distributed algorithm is very close to the ϵ -optimal algorithm.

In the second experiment, we use 10 video sessions, with six transmitting FM and the others transmitting MD. We found that it is very hard to generate a network topology that can satisfy the minimum rates of all sessions simultaneously. In this case, admission control is needed, or more spectrum should be used to support large number concurrent sessions.

In the third experiment, we consider a network with communication area of $200 \times 200 \text{ m}^2$ (which is much smaller than $1000 \times 600 \text{ m}^2$ used in above), 10 nodes in total, and two video sessions of MD and FM. As a result, the average interference between any two randomly located nodes in the network becomes much higher than that in previous experiments. Results show that the centralized ϵ -optimal solution algorithm could achieve

an average sum PSNR of 83.1 dB, while the distributed algorithm achieves 80 dB, which is more than 97% of the ϵ -optimum. That is to say, a good performance still can be achieved by applying the proposed distributed algorithm in interference-dominated cooperative networks.

VII. CONCLUSION

In this paper, we studied cross-layer design techniques for video streaming cooperative multimedia wireless networks. We formulated the joint control of the video encoding rate, relay selection, and power allocation as a nonconvex and combinatorial problem. Then, we proposed a global solution algorithm based on a combination of branch and bound and convex relaxation of the original problem, which is proven to converge to the optimal solution. Moreover, we have shown that through cooperative relaying, nodes are allowed to work at lower level of average transmission power without decrease in the sum video quality. We also proposed an iterative distributed algorithm in which each iteration can be executed in polynomial time and the overall algorithm converges very fast in practice. The proposed algorithm can be directly applied to a scenario with multiple co-existing preestablished source–destination pairs and can be also used to optimally control resource allocation for an independent set of transmissions with primary interference constraints (i.e., no transmitters and receivers in common) periodically scheduled by a separate scheduling algorithm, where idle nodes can be used as potential relays.

It is worth pointing out that in this work we focus on studying the problem of joint rate control and relay selection from an information-theoretic perspective, i.e., we do not make any assumptions on the actual modulation and coding scheme. In future work, we would like to consider real transmission techniques—e.g., specific modulations, SNR thresholds—and implement the proposed algorithm on a GNU Radio/USRP2 testbed available at SUNY Buffalo.

APPENDIX

A. Validation of the Average-Based Interference Model

We validate the average interference model by comparing it to the exact interference in practical cooperative wireless networks. Let us consider a cooperative network having multiple sessions that use cooperative relaying. Then, for a wireless link l that uses direct transmission only, the interference measured at its destination node comes from all source and relay nodes that transmit in the first time slot, while it comes from other source and relay nodes in the second. Then, the average capacity of the wireless link C_{rea} can be calculated as $C_{\text{rea}} = \frac{1}{2}(C_{\text{slot1}} + C_{\text{slot2}})$, where C_{slot1} and C_{slot2} represent the capacity in the first and second time slot, respectively.

We let C_{avg} represent capacity of the wireless link l calculated using the average-based interference model, and then we compare it to C_{rea} . A communication area of $500 \times 500 \text{ m}^2$ is considered, and the number of interfering cooperative sessions varies from 2 to 16 in increments of 2. The other simulation parameters are set as in Section VI. Results of the comparison in terms of $\frac{C_{\text{avg}}}{C_{\text{rea}}}$ are shown in Fig. 5(a) (Average/Exact). Every point was plotted by averaging over 10^3 simulations. The value of C_{avg} is slightly lower but very close to that of C_{rea} , e.g., around 98% and 97% of C_{rea} can be achieved when $N_{\text{itf}} = 2$

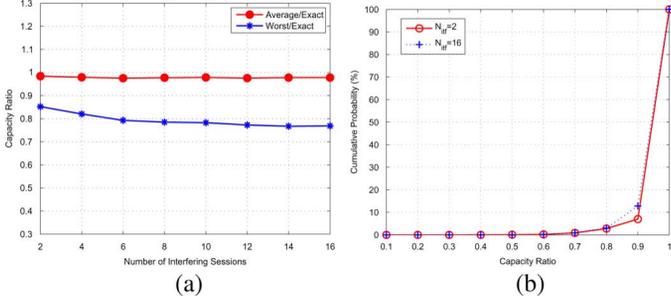


Fig. 5. (a) Comparison between the average-based/worst-case interference model and exact interference. (b) Cumulative probability of the capacity ratio corresponding to the average-based interference model.

and $N_{\text{itf}} = 16$, respectively. We observe that the value of $\frac{C_{\text{avg}}}{C_{\text{vea}}}$ decreases very slightly as the number of interfering sessions increases, implying that the accumulation of average performance degradation caused by the average-based interference model is negligible. Similar results can be also observed when wireless link l also uses cooperative relaying. For comparison, we also compare the worst-case approximation and the synchronization case. The performance of the worst-case approximation model deteriorates significantly as the number of interfering nodes increases, and less than 80% of the real capacity can be achieved when there are 16 interfering nodes. We also plot the cumulative probability of the capacity ratio corresponding to the average-based interference model in Fig. 5(b). A capacity ratio larger than 0.9 can be achieved with probability of 93%, 88% for $N_{\text{itf}} = 2$ and 16, respectively. Based on the above discussion, we can conclude that the average-based interference model provides a good approximation of the exact interference in practical cooperative wireless networks, especially when the number of interfering sessions is not large.

B. Proof for Convergence to NE for Distributed Algorithm

We only need to show that the relay selection problem formulated in Section V can be reformulated as a VI problem, and show that the proposed distributed algorithm converges to a VI solution that is also an NE solution.

VI Formulation: To this end, we assume that $\mathcal{R}'_s \cap \mathcal{R}'_w = \emptyset$, $\forall w, s \in \mathcal{S}$, $w \neq s$. The assumption is realistic since in a cooperative wireless network where the number of relay nodes is comparable or larger than the number of video sessions, relay selection for a video session only affects interference caused by this video session to the other sessions.⁷ Then, the overall link capacity C_s in (63) can be rewritten as $C_s = \alpha_{s,0}C_{s2d} + \sum_{r \in \mathcal{R}} \alpha_{s,r}C_{\text{coop}}^{s,r}$, and the constraint in (2) can be rewritten as $\sum_{r \in \mathcal{R}_s} \alpha_{s,r} \leq 1$, $\forall s \in \mathcal{S}$. It can be proven

that the domain set of the relay selection problem is compact and convex, and that the utility function of each session is continuously differentiable. Under these conditions, the problem of joint relay selection and video rate control in Section V can

⁷Multiple sessions may possibly want to select the same relay node as cooperative relay even in a network with only a limited number of sessions. The resulting NE problem is a QNE problem, where the domain sets of all sessions are coupled with each other and more advanced techniques, e.g., pricing-based algorithms, are needed to derive the equilibrium point. In this paper, we focus on standard NE analysis and leave the QNE for future work. Readers are referred to [44] for more details of QNE analysis.

be reformulated as a VI problem, and there exists at least one VI solution [45, p. 175, Theorem 2.4.4].

In the following, we present a sufficient condition for the distributed algorithm to converge, that can be derived based on the framework in [46]. We represent PSNR_s in (20) using U_s for simplicity. Then, the gradient vector of session s with respect to α_s can be written as $J_{\alpha_s}(U_s) = \left(\frac{\partial U_s}{\partial \alpha_{s,r}} \right)_{r=1}^{R_s}$, where R_s represents the number of potential relay nodes in set \mathcal{R}_s . Similarly, we represent the gradient vector of session g , with $g \in \mathcal{S}$, as $J_{\alpha_g}(U_g) = \left(\frac{\partial U_g}{\partial \alpha_{g,s}} \right)_{s=1}^{R_g}$. Furthermore, denote the Jacobi matrix of $J_{\alpha_s}(U_s)$ and $J_{\alpha_g}(U_g)$ with respect to α_s as $J_{\alpha_s \alpha_s}(U_s)$ and $J_{\alpha_g \alpha_s}(U_g)$, respectively. Here, $J_{\alpha_g \alpha_s}(U_g)$ is namely the Hessian matrix session s . Then, we can define a matrix $[\gamma]_{ij}$ as follows $[\gamma]_{sg} \triangleq \begin{cases} \eta_s^{\min}, & \text{if } s = g \\ -\beta_{sg}^{\max}, & \text{otherwise} \end{cases}$, where $\eta_s^{\min} \triangleq \inf_{\alpha \in \mathcal{X}} \lambda_{\text{least}}(J_{\alpha_s \alpha_s}(U_s))$ and $\beta_{sg}^{\max} \triangleq \sup_{\alpha \in \mathcal{X}} \|J_{\alpha_g \alpha_s}(U_g)\|$, with $\lambda_{\text{least}}(\mathbf{A})$ representing the eigenvalue of \mathbf{A} with the smallest absolute value. Then, to guarantee the convergence of the proposed distributed algorithm, we only need to show that the matrix $[\gamma]_{sg}$ is a P-matrix [46]. It can be shown that, for given transmission parameters (e.g., potential relay nodes of each session, maximum transmission power), there exists a threshold $L_{\text{th}} < \infty$ that can be calculated numerically, e.g., by using a bisection-based iterative method, to verify whether γ_{sg} is a P-matrix or not at each iteration.

REFERENCES

- [1] Z. Guan, T. Melodia, and D. Yuan, "Optimizing cooperative video streaming in wireless networks," in *Proc. IEEE SECON*, Salt Lake City, UT, USA, Jun. 2011, pp. 503–511.
- [2] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [4] G. Jakkari, S. V. Krishnamurthy, M. Faloutsos, P. V. Krishnamurthy, and O. Ercetin, "A cross-layer framework for exploiting virtual MISO links in mobile ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 579–594, Jun. 2007.
- [5] G. Kramer, I. Maric, and R. D. Yates, "Cooperative communications," *Found. Trends Netw.*, vol. 1, no. 3–4, pp. 271–425, Jun. 2007.
- [6] K. J. R. Liu, K. Ahmed, W. Su, and K. Andres, *Cooperative Communications and Netw.*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [7] A. H.-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.
- [8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [9] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. S. Panwar, "CoopMAC: A cooperative MAC for wireless LANs," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 340–354, Feb. 2007.
- [10] J. Zheng and M. Ma, "QoS-aware cooperative medium access control for MIMO ad-hoc networks," *IEEE Commun. Lett.*, vol. 14, no. 1, pp. 48–50, Jan. 2010.
- [11] S. Moh and C. Yu, "A cooperative diversity-based robust MAC protocol in wireless ad hoc networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 3, pp. 353–363, Mar. 2011.
- [12] Z. Guan, T. Melodia, D. Yuan, and D. A. Pados, "Distributed spectrum management and relay selection in interference-limited cooperative wireless networks," in *Proc. ACM MobiCom*, Las Vegas, NV, USA, Sep. 19–23, 2011, pp. 229–240.
- [13] C. Chen and S. Kishore, "Making cooperation valuable: A delay-centric, pricing-based user cooperation strategy," in *Proc. IEEE CISS*, Baltimore, MD, USA, Mar. 2011, pp. 1–6.

- [14] S. Lakshmanan and R. Sivakumar, "Diversity routing for multi-hop wireless networks with cooperative transmissions," in *Proc. IEEE SECON*, Rome, Italy, Jun. 2009, pp. 1–9.
- [15] J. Zhang and Q. Zhang, "Cooperative routing in multi-source multi-destination multi-hop wireless networks," in *Proc. IEEE INFOCOM*, Phoenix, AZ, USA, Apr. 2008, pp. 2369–2377.
- [16] L. Ding, T. Melodia, S. N. Batalama, and J. D. Matyas, "Distributed routing, relay selection, and spectrum allocation in cognitive and cooperative ad hoc networks," in *Proc. IEEE SECON*, Boston, MA, USA, Jun. 2010, pp. 1–9.
- [17] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM MobiHoc*, May 2008, pp. 3–12.
- [18] S. Sharma, Y. Shi, Y. T. Hou, H. D. Sherali, and S. Kompella, "Cooperative communications in multi-hop wireless networks: Joint flow routing and relay node assignment," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.
- [19] M. Rossi, C. Tapparello, and S. Tomasin, "On optimal cooperator selection policies for multi-hop ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 506–518, Feb. 2011.
- [20] E. M. Yeh and R. A. Berry, "Throughput optimal control of cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3827–3833, Oct. 2007.
- [21] N. Mastrorade, M. van der Schaar, A. Scaglione, F. Verde, and D. Darsena, "Sailing good radio waves and transmitting important bits: Relay cooperation in wireless video transmission," in *Proc. IEEE ICASSP*, Dallas, TX, USA, Mar. 2010, pp. 5566–5569.
- [22] O. Alay, P. Liu, Y. Wang, E. Erkip, and S. S. Panwar, "Cooperative layered video multicast using randomized distributed space time codes," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 1127–1140, Oct. 2011.
- [23] A. A. Khalek and Z. Dawy, "Energy-efficient cooperative video distribution with statistical QoS provisions over wireless networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 7, pp. 1223–1236, Jul. 2012.
- [24] H. Xiao, Q. Dai, and X. Ji, "Joint resources allocation for cooperative video transmission," in *Proc. IEEE ICIP*, Cairo, Egypt, Nov. 2009, pp. 921–924.
- [25] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, Jul. 1960.
- [26] T. J. Shepard, "A channel access scheme for large dense packet radio networks," in *Proc. ACM SIGCOMM*, 1996, pp. 219–230.
- [27] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 487–499, Aug. 2002.
- [28] A. Ulusoy, O. Gurbuz, and A. Onat, "Wireless model-based predictive networked control system over cooperative wireless network," *IEEE Trans. Ind. Inform.*, vol. 7, no. 1, pp. 41–51, Feb. 2011.
- [29] D. Yang, X. Fang, and G. Xue, "OPRA: Optimal relay assignment for capacity maximization in cooperative networks," in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011, pp. 1–6.
- [30] Q. Zhang, J. Jia, and J. Zhang, "Cooperative relay to improve diversity in cognitive radio networks," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 111–117, Feb. 2009.
- [31] F. Fu and M. van der Schaar, "A new systematic framework for autonomous cross-layer optimization," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1887–1903, May 2009.
- [32] S. Pudlewski, T. Melodia, and A. Prasanna, "Compressed-sensing-enabled video streaming for wireless multimedia sensor networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1060–1072, Jun. 2012.
- [33] K. Stuhlmüller, N. Fäber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.
- [34] X. Zhu, E. Setton, and B. Girod, "Congestion-distortion optimized video transmission over ad hoc networks," *EURASIP Signal Process., Image Commun.*, vol. 20, no. 8, pp. 773–783, Sep. 2005.
- [35] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 493–507, Dec. 1952.
- [36] D. Bertsekas and R. Gallager, *Data Networks*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.
- [37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [38] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*. Boston, MA, USA: Kluwer, 1999.
- [39] Y. E. Nesterov and A. S. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM, 2004.
- [40] J. Nash, "Equilibrium points in N-person games," *Proc. Nat. Acad. Sci.*, vol. 36, no. 1, pp. 48–49, Jan. 1950.
- [41] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1226–1237, Sep. 2008.
- [42] B. Wang, Z. Han, and K. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using Stackelberg game," *IEEE Trans. Mobile Comput.*, vol. 8, no. 7, pp. 975–990, Jul. 2009.
- [43] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative cognitive radio networks," in *Proc. ACM MobiHoc*, New Orleans, LA, USA, May 2009, pp. 23–32.
- [44] J.-S. Pang and M. Fukushima, "Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games," *Comput. Manage. Sci.*, vol. 2, no. 1, pp. 21–56, 2005.
- [45] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problem*. New York, NY, USA: Springer-Verlag, 2003.
- [46] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, , A. Rantzer and R. Johansson, Eds., "Monotone games for cognitive radio systems," in *Distributed Decision-Making and Control*, ser. Lecture Notes in Control and Information Sciences. New York, NY, USA: Springer-Verlag, 2011.



Zhangyu Guan (M'11) received the B.S. and Ph.D. degrees in communication engineering from Shandong University, Jinan, China, in 2003 and 2010, respectively.

He is a Lecturer with the School of Information Science and Engineering, Shandong University. From 2009 to 2010, he visited The State University of New York (SUNY) at Buffalo, Buffalo, NY, USA. His current research interests are in wireless network modeling and optimization, multimedia wireless networking, and mobile cloud computing.



Tommaso Melodia (M'07) received the "Laurea" (integrated B.S. and M.S.) and Doctorate degrees in telecommunications engineering from the University of Rome "La Sapienza," Rome, Italy, in 2001 and 2005, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2007.

He is an Associate Professor with the Department of Electrical Engineering, State University of New York (SUNY) at Buffalo, Buffalo, NY, where he directs the Wireless Networks and Embedded Systems

Laboratory. His current research interests are in modeling, optimization, and experimental evaluation of wireless networks, with applications to cognitive and cooperative networking, ultrasonic intrabody area networks, multimedia sensor networks, and underwater networks.

Dr. Melodia serves on the Editorial Boards of IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and *Computer Networks*, among others. He is the Technical Program Committee Vice Chair for IEEE GLOBECOM 2013 and the Technical Program Committee Vice Chair for Information Systems for IEEE INFOCOM 2013. He is a recipient of the National Science Foundation CAREER Award. He coauthored a paper that was recognized as the Fast Breaking Paper in the field of Computer Science for February 2009 by Thomson ISI Essential Science Indicators and a paper that received an Elsevier Top Cited Paper Award.



Dongfeng Yuan (SM'01) received the M.S. degree from Shandong University, Jinan, China, in 1988, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2000, both in electrical engineering.

Currently, he is a Full Professor and Dean with the School of Information Science and Engineering, Shandong University. His research interests are in the area of wireless communications, with a focus on cross-layer design and resource allocations.