Streaming Scalable Videos over Multi-Hop Cognitive Radio Networks

Donglin Hu and Shiwen Mao, *Senior Member, IEEE*

Abstract—We investigate the problem of streaming multiple videos over multi-hop cognitive radio (CR) networks. Fine-Granularity-Scalability (FGS) and Medium-Grain-Scalable (MGS) videos are adopted to accommodate the heterogeneity among channel availabilities and dynamic network conditions. We obtain a mixed integer nonlinear programming (MINLP) problem formulation, with objectives to maximize the overall received video quality and to achieve fairness among the video sessions, while bounding the collision rate with primary users under the presence of spectrum sensing errors. We first solve the MINLP problem using a centralized sequential fixing algorithm, and derive upper and lower bounds for the objective value. We then apply dual decomposition to develop a distributed algorithm and prove its optimality and convergence conditions. The proposed algorithms are evaluated with simulations and are shown to be effective in supporting concurrent scalable video sessions in multi-hop CR networks.

Index Terms—Cross-layer optimization, dynamic spectrum access, distributed algorithm, multi-hop cognitive radio networks, video streaming.

I. INTRODUCTION

A cognitive radio (CR) is an advanced radio device with interface(s) to sense the radio environment, an intelligent agent for decision-making based on radio environment and past experience, and a frequency-agile radio that can be tuned to a wide range of frequency bands and operate from there. CR represents a paradigm change in spectrum regulation and access, from exclusive use by licensed, or *primary*, users to sharing spectrum with, and dynamic spectrum access for unlicensed, or *secondary*, users. It has profound impact on how future wireless networks will be designed and operated.

The high potential of CRs has attracted substantial interest. The mainstream CR research has focused on developing effective spectrum sensing and access techniques (see [1] and [2] and reference therein). Although considerable understandings have been gained on various aspects of CR, the important problem of guaranteeing application performance has not been the focus of major CR research. To this end, we find spectrum-intensive and rate-adaptive multimedia, or video as
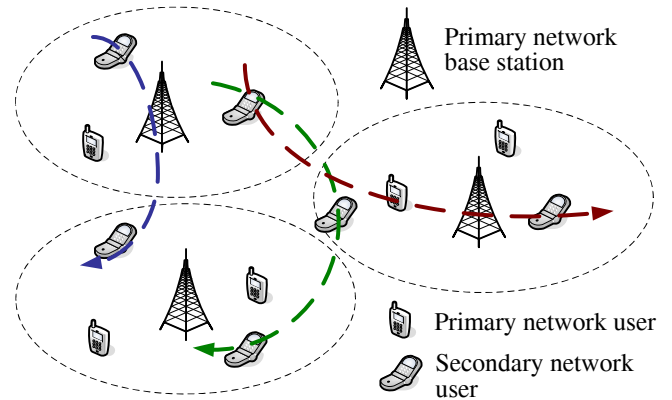
Fig. 1. Illustration of the multi-hop video CR network architecture.

a reference application, makes excellent use of the enhanced spectrum efficiency in CR networks. Unlike data, where each bit should be delivered, video is loss tolerant and rate adaptive. They are highly suited for CR networks, where the available bandwidth heavily depends on primary user behavior. Graceful degradation of video quality can be achieved as spectrum opportunities change over time.

In our prior work [3], we investigated the problem of multicasting Fine-Granularity-Scalablity (FGS) video in an infrastructure-based CR network and demonstrated the feasibility of video over CR networks. In this paper, we study the more challenging problem of video over multi-hop CR networks. As illustrated in Fig. 1, we consider an infrastructureless CR network co-located with one or more fixed primary networks. CR users non-intrusively exploit white spaces in the licensed bands for streaming multiple videos. The objective is two-fold: to maximize the overall video quality and to achieve fairness among the concurrent video sessions, subject to bounded interference to primary users.

We adopt FGS videos to accommodate heterogeneous channel availabilities and dynamic network conditions [4]. FGS video is coded into a *base layer* (BL) and an *enhancement layer* (EL). The EL can be truncated at any bit location, while all the remaining bits are still useful for decoding. This feature simplifies the design of video streaming systems. We also consider H.264/SVC Medium-Grain-Scalable (MGS) videos in this paper. MGS is shown to achieve better rate-distortion performance over MPEG-4 FGS, although MGS only has Network Abstraction Layer (NAL) unit-based granularity [5].

In order to model and guarantee end-to-end video performance, we adopt the *amplify-and-forward* relay approach for video data, which is well-studied in the context of cooperative communications [6]. Specifically, each CR node is equipped

with two transceivers operating on orthogonal channels. During data transmission, a relay CR node receives data from its upstream node using one transceiver on one channel, while simultaneously amplifies and forwards the received data to its downstream node using the other transceiver operating on a different channel. This is equivalent to establishing a virtual tunnel through a multi-hop multi-channel path, as illustrated in Fig. 2. It is also analogous to the "cut-through switching" approach for packet switching networks [7]. In addition to allowing a neat formulation of the challenging multi-hop video streaming problem, this approach also satisfies video's needs for low latency, low jitter, and high bandwidth. Its feasibility and practical considerations for multi-hop wireless networks have been addressed in [8].

The target problem is non-trivial due to the additional dimension of network dynamics (i.e., channel availability) and the additional uncertainty (i.e., spectrum sensing and sensing errors) found in CR networks. The stringent QoS requirements necessitate cross-layer optimization. The lack of centralized control also calls for distributed algorithms. We formulate streaming multiple videos over a multi-hop CR network as a *mixed integer nonlinear programming* (MINLP) problem, considering important design factors such as spectrum sensing and sensing errors, spectrum access and primary user protection, video quality and fairness, channel scheduling, and path selection. We first develop a centralized *sequential fixing* algorithm to derive upper and lower bounds for the achievable video quality. These bounds provide useful insights on performance limits of the CR video system. We then decompose the MINLP problem into a channel scheduling problem and a path selection problem. The channel scheduling problem is solved with a greedy algorithm. For path selection, we apply *dual decomposition* and develop a distributed algorithm. We prove the optimality of the proposed approach and derive the convergence condition for the distributed algorithm. The algorithms are evaluated with extensive simulations. The distributed algorithm is shown to be highly effective for supporting concurrent video sessions in multi-hop CR networks, as it can achieve a performance close to that of the centralized algorithm as well as the upper bound in the cases examined.

The remainder of this paper is organized as follows. The system model is described in Section II. We present the problem formulation and develop the centralized algorithm in Section III. We derive the distributed algorithm and analyze its optimality and convergence performance in Section IV. Simulation results are presented in Section V and related work is discussed in Section VI. Section VII concludes the paper.

II. SYSTEM MODEL

A. Network Model

We consider a spectrum band consisting of M orthogonal channels with identical bandwidth [9]. As shown in Fig. 1, the channels are shared by K primary networks and one multi-hop CR network. The primary network base stations provide data or multimedia service to primary users. There is no fixed infrastructure in the CR network; secondary users non-intrusively explore the spectrum opportunities for unicast video communications.

1) Primary Networks: We assume that the M channels are allocated to K primary networks, which cover different service areas. We further assume that the primary systems use a synchronous slot structure as in prior work [2], [10]. Due to primary user transmissions, the occupancy of each channel evolves following a discrete-time Markov process, as validated by recent measurement studies [2], [10], [11].

In primary network k, the status of channel m in time slot t is denoted by $S_m^k(t)$ with idle (i.e., $S_m^k(t) = 0$) and busy (i.e., $S_m^k(t) = 1$) states. Let λ_m^k and μ_m^k be the transition probability of remaining in state 0 and that from state 1 to 0, respectively, for channel m in primary network k. The utilization of channel m in primary network k, denoted by $\eta_m^k = \Pr(S_m^k = 1)$, is

$$\eta_m^k = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} S_m^k(t) = \frac{1 - \lambda_m^k}{1 - \lambda_m^k + \mu_m^k}. \tag{1}$$

2) The Multi-hop CR Network: Consider a multi-hop CR network co-located with the primary networks, within which N CR nodes are streaming S real-time videos. Let $\mathcal{U}^k$ denote the set of CR nodes that are located within the coverage of primary network k. A video session l may be relayed by multiple CR nodes if source z_l is not a one-hop neighbor of destination d_l. We assume a *common control channel* for the CR network [10]. We also assume the timescale of the primary channel process (or, the time slot durations) is larger than the broadcast delays on the control channel, such that feedbacks of channel information can be received at the source nodes in a timely manner.

For CR users, each time slot consists of three phases: the spectrum sensing phase, the data transmission phase, and the acknowledgment phase. Assume that each CR user has two transceivers. In the sensing phase, one transceiver is used to sense one of the M channels, while the other is tuned to the control channel to exchange channel information with other CR users. Each video source computes the optimal path selection and channel scheduling based on sensing results. In the transmission phase, the channels assigned to a video session l at each link along the path form a virtual "tunnel" connecting source z_l and destination d_l, as illustrated in Fig. 2. Each node can use one or more than one channels to communicate with other nodes using a channel aggregation technique such as Orthogonal Frequency Division Multiplexing (OFDM) [10], [12]. When multiple channels are available on all the links along a path, multiple tunnels can be established and used simultaneously for a video session. In the acknowledgment phase, the destination sends ACK to the source for successfully received video packets through the same tunnel.

We adopt amplify-and-forward for video transmission [6]. During the transmission phase, one transceiver of the relay node receives video data from the upstream node on one channel, while the other transceiver of the relay node amplifies and forwards the data to the downstream node on a different, orthogonal channel. There is no need to store video packets at the relay nodes. Error detection/correction will be performed at the destination node. As a result, we can transmit through the tunnel a block of video data with minimum delay and jitter in one time slot.

Compared to the traditional "hop-centric" approach, this scheme greatly reduces the collision, contention, processing,
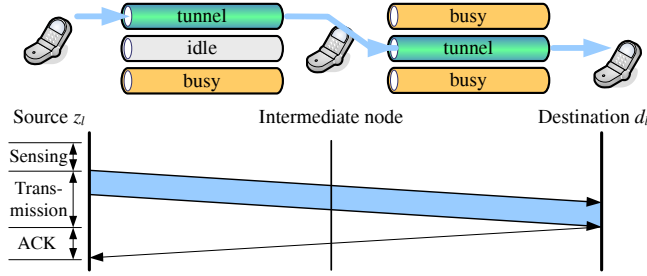
Fig. 2.   The cut-through switching model for video data.

and queueing delay induced at relay nodes [7], [8]. It is suitable for real-time data with tight delay and jitter requirements. It is especially amicable for FGS video, since a corrupted packet may still be useful for decoding. The viability, protocol-related issues, and practical considerations of this approach are addressed in [8]. The challenging issue, however, is how to set up the tunnels, while the available channels at the relays evolve over time due to primary user transmissions. We will address this issue in Section IV.

B. Spectrum Sensing

Although precise and timely channel state information is highly desirable, continuous full-spectrum sensing is hardware demanding. Without loss of generality, we assume each CR user periodically chooses one channel to sense in each time slot [13]. The index of the chosen channel by user i in time slot t is: $M_i(t) = [M_i(0) + t - 1] \mod M$, where $M_i(0)$ is the index of the channel sensed in time slot 0.

There are two types of spectrum sensing errors: with a *false alarm*, a spectrum opportunity will be wasted, while a *miss detection* may lead to collision with primary users. Without loss of generality, we assume that the same spectrum sensing mechanism is used with identical sensing error probabilities for all CR users. Let ϵ_m and δ_m denote the probabilities of false alarm and miss detection on channel m, respectively. For time slot t, we have:

$$\begin{cases} P(W_i^m(t) = 1 | S_m^k(t) = 0) = \epsilon_m, & m = 1, \cdots, M \\ P(W_i^m(t) = 0 | S_m^k(t) = 1) = \delta_m, & m = 1, \cdots, M, \end{cases} \quad (2)$$

where $W_i^m(t)$ is user i's sensing result for channel m.

In a multi-hop CR network, the sensing results from various users may be different. Denote H_0 as the hypothesis that channel m in primary network k is idle, and H_1 the hypothesis that channel m in primary network k is busy in time slot t. The conditional probability that channel m is available in primary network k, denoted by $a_m^k(t)$, can be derived as in (3), where θ_i^m represents a specific sensing result (0 or 1), $\mathcal{U}_m^k$ is the subset of users in $\mathcal{U}^k$ (i.e., the set of CR nodes that are located within the coverage of primary network k) that sense channel m, u_m^k is the number of users in $\mathcal{U}_m^k$ observing channel m is idle, π_m^k represents the history of channel m in primary

network k,[1] and φ_m^k and ϕ_m^k are defined as:

$$\begin{cases} \varphi_m^k = \frac{P(W_i^m = 0 | H_1)}{P(W_i^m = 0 | H_0)} = \frac{\delta_m}{1 - \epsilon_m}, & \text{when } \theta_i^m = 0 \\ \phi_m^k = \frac{P(W_i^m = 1 | H_1)}{P(W_i^m = 1 | H_0)} = \frac{1 - \delta_m}{\epsilon_m}, & \text{when } \theta_i^m = 1. \end{cases} \quad (4)$$

The third equality equality in (3) is due to independent sensing processes. The fourth equality is because sensing processes are independent of channel history. Based on the Markov chain channel model, we have (5), which can be recursively expanded:

$$\begin{cases} \Pr(H_0 | \pi_m^k) = \lambda_m^k a_m^k(t - 1) + \mu_m^k \left[1 - a_m^k(t - 1) \right] \\ \Pr(H_1 | \pi_m^k) = 1 - \Pr(H_0 | \pi_m^k). \end{cases} \quad (5)$$

C. Spectrum Access and Interference Modeling

Based on spectrum sensing results, a CR user determines which channel(s) to access for transmission of video data. Let κ_m^k be a threshold for spectrum access: channel m is considered idle if the estimate a_m^k is greater than the threshold, and is considered busy otherwise. The availability of channel m in primary network k, denoted as A_m^k, is

$$A_m^k = \begin{cases} 0, & a_m^k \geq \kappa_m^k \\ 1, & \text{otherwise.} \end{cases} \quad (6)$$

For each channel m, we can calculate the probability of collision with primary users as:

$$\Pr(A_m^k = 0 | H_1) = \sum_{i \in \psi_m^k} \binom{|\mathcal{U}_m^k|}{i} (1 - \delta_m)^{|\mathcal{U}_m^k| - i} (\delta_m)^i, \quad (7)$$

where set ψ_m^k is defined as:

$$\psi_m^k = \left\{ i \left| \left[1 + \varphi_m^i \phi_m^{|\mathcal{U}_m^k| - i} \frac{\Pr(H_1 | \pi_m^k)}{\Pr(H_0 | \pi_m^k)} \right]^{-1} \geq \kappa_m^k \right. \right\}. \quad (8)$$

For non-intrusive spectrum access, the collision probability should be bounded with a prescribed threshold γ_m^k. A higher spectrum access threshold κ_m^k will reduce the potential interference with primary users, but increase the chance of wasting transmission opportunities. For a given collision tolerance γ_m^k, we can solve $\Pr(A_m^k = 0 | H_1) = \gamma_m^k$ for κ_m^k. The objective is to maximize CR users' spectrum access without exceeding the maximum collision probability with primary users.

Let $\Omega_{i,j}$ be the set of available channels at link $\{i, j\}$. Assuming $i \in \mathcal{U}^k$ and $j \in \mathcal{U}^{k'}$, we have

$$\Omega_{i,j} = \left\{ m \left| A_m^k = 0 \text{ and } A_m^{k'} = 0 \right. \right\}. \quad (9)$$

D. Link and Path Statistics

Due to the amplify-and-forward approach for video data transmission, there is no queueing delay at intermediate nodes. Assume each link has a fixed delay $\omega_{i,j}$ (i.e., processing and propagation delays). Let $\mathcal{P}_l^A$ be the set of all possible paths from z_l to d_l. For a given delay requirement T_{th}, the set of feasible paths $\mathcal{P}_l$ for video session l can be determined as:

$$\mathcal{P}_l = \left\{ \mathcal{P} \left| \sum_{\{i,j\} \in \mathcal{P}} \omega_{i,j} \leq T_{th}, \mathcal{P} \in \mathcal{P}_l^A \right. \right\}. \quad (10)$$

---

[1]π_m^k represents the availability of channel m in primary network k in the previous time slot. If the channel was used in that time slot, π_m^k can be readily determined as 0 or 1, since the channel state was known (i.e., with or without ACKs). Otherwise, π_m^k can be estimated in the form of $a_m^k(t - 1)$ as in (3).

$$a_m^k(t) = \Pr(H_0 | W_i^m = \theta_i^m, \ i \in \mathcal{U}_m^k, \pi_m^k) = \frac{\Pr(W_i^m = \theta_i^m, i \in \mathcal{U}_m^k | H_0, \pi_m^k) \Pr(H_0 | \pi_m^k)}{\sum_{s \in \{0,1\}} \Pr(W_i^m = \theta_i^m, i \in \mathcal{U}_m^k | H_s, \pi_m^k) \Pr(H_s | \pi_m^k)}$$

$$= \frac{\Pr(H_0 | \pi_m^k) \prod_{i \in \mathcal{U}_m^k} \Pr(W_i^m = \theta_i^m | H_0, \pi_m^k)}{\sum_{s \in \{0,1\}} \Pr(H_s | \pi_m^k) \prod_{i \in \mathcal{U}_m^k} \Pr(W_i^m = \theta_i^m | H_s, \pi_m^k)} = \frac{\Pr(H_0 | \pi_m^k) \prod_{i \in \mathcal{U}_m^k} \Pr(W_i^m = \theta_i^m | H_0)}{\sum_{s \in \{0,1\}} \Pr(H_s | \pi_m^k) \prod_{i \in \mathcal{U}_m^k} \Pr(W_i^m = \theta_i^m | H_s)}$$

$$= \left[1 + \frac{\Pr(H_1 | \pi_m^k)}{\Pr(H_0 | \pi_m^k)} \prod_{i \in \mathcal{U}_m^k} \frac{\Pr(W_i^m = \theta_i^m | H_1)}{\Pr(W_i^m = \theta_i^m | H_0)} \right]^{-1} = \left[1 + (\varphi_m^k)^{u_m^k} (\phi_m^k)^{|\mathcal{U}_m^k| - u_m^k} \frac{\Pr(H_1 | \pi_m^k)}{\Pr(H_0 | \pi_m^k)} \right]^{-1}. \qquad (3)$$

Let $p_{i,j}^m$ be the packet loss rate on channel m at link $\{i,j\}$. A packet is successfully delivered over link $\{i,j\}$ if there is no loss on all the channels that were used for transmitting the packet. The link loss probability $p_{i,j}$ can be derived as:

$$p_{i,j} = 1 - \prod_{m \in \mathcal{M}} (1 - p_{i,j}^m)^{I_m}, \qquad (11)$$

where $\mathcal{M}$ is set of licensed channels and I_m is an indicator: $I_m = 1$ if channel m is used for the transmission, and $I_m = 0$ otherwise. Assuming independent link losses, the end-to-end loss probability for path $\mathcal{P}_l^h \in \mathcal{P}_l$ can be estimated as:

$$p_l^h = 1 - \prod_{\{i,j\} \in \mathcal{P}_l^h} (1 - p_{i,j}). \qquad (12)$$

E. Video Performance Measure

As discussed, both FGS and MGS videos are highly suitable for dynamic CR networks [3]. With FGS or MGS coding, each video l is encoded into one base layer with rate R_l^b and one enhancement layer with rate R_l^e. The total bit rate for video l is $R_l = R_l^b + R_l^e$.

We consider peak-signal-noise-ratio (PSNR) (in dB) of reconstructed videos. As in prior work [3], [14], the average PSNR of video l, denoted as Q_l, can be estimated as:

$$Q_l(R_l) = Q_l^b + \beta_l(R_l - R_l^b) = Q_l^0 + \beta_l R_l, \qquad (13)$$

where Q_l^b is the resulting PSNR when the base layer is decoded alone, β_l a constant depending on the video sequence and codec setting, and $Q_l^0 = Q_l^b - \beta_l R_l^b$. We verified the model (13) with several test video sequences using the MPEG-4 FGS codec and the H.264/SVC MGS codec and found it is highly accurate. The results are omitted for brevity.

Due to the real-time nature, we assume that each *group of pictures* (GOP) must be delivered during the next GOP window, which consists of N_G time slots. Beyond that, overdue data from the current GOP will be useless and will be discarded. We further assume fixed packet length L_p. Therefore the average rate in a GOP window is $R = (N_p L_p)/(N_G T_s)$, where T_s is the time slot duration and N_p is the number of successfully delivered packets.

We also aim to achieve fairness among the concurrent video sessions. It has been shown that *proportional fairness* can be achieved by maximizing the sum of logarithms of video PSNRs (i.e., utilities). Therefore, our objective is to maximize the overall system utility, i.e.,

$$\text{maximize:} \ \sum_l U_l(R_l) = \sum_l \log(Q_l(R_l)). \qquad (14)$$

III. PROBLEM STATEMENT

A. Multi-hop CR Network Video Streaming Problem

For the system described in Section II, the problem of video over multi-hop CR networks consists of path selection for each video session and channel scheduling for each CR node along the chosen paths. We define two sets of index variables. For channel scheduling, we have

$$x_{i,j,m}^{l,h,r} = \begin{cases} 1, & \text{at link } \{i,j\}, \text{ if channel } m \text{ is} \\ & \text{assigned to tunnel } r \text{ in path } \mathcal{P}_l^h \\ 0, & \text{otherwise.} \end{cases} \qquad (15)$$

For path selection, we have

$$y_l^h = \begin{cases} 1, & \text{if video session } l \text{ selects path } \mathcal{P}_l^h \in \mathcal{P}_l \\ 0, & \text{otherwise,} \end{cases} \qquad (16)$$

Note that the indicators, $x_{i,j,m}^{l,h,r}$ and y_l^h, are not independent. If $y_l^h = 0$ for path $\mathcal{P}_l^h$, all the $x_{i,j,m}^{l,h,r}$'s on that path are 0. If link $\{i,j\}$ is not on path $\mathcal{P}_l^h$, all its $x_{i,j,m}^{l,h,r}$'s are also 0. For link $\{i,j\}$ on path $\mathcal{P}_l^h$, we can only choose those available channels in set $\Omega_{i,j}$ to schedule video transmission. That is, we have $x_{i,j,m}^{l,h,r} \in \{0,1\}$ if $m \in \Omega_{i,j}$, and $x_{i,j,m}^{l,h,r} = 0$ otherwise. In the rest of the paper, we use $\mathbf{x}$ and $\mathbf{y}$ to represent the vector forms of $x_{i,j,m}^{l,h,r}$ and y_l^h, respectively.

As discussed, the objective is to maximize the expected utility sum at the end of N_G time slots, as given in (14). Since $\log(Q_l(\mathrm{E}[R_l(0)]))$ is a constant, (14) is equivalent to the sum of utility increments of all the time slots, as

$$\sum_l \log(Q_l(\mathrm{E}[R_l(N_G)])) - \log(Q_l(\mathrm{E}[R_l(0)]))$$
$$= \sum_t \sum_l \{\log(Q_l(\mathrm{E}[R_l(t)])) - \log(Q_l(\mathrm{E}[R_l(t-1)]))\} . \qquad (17)$$

Therefore, (14) will be maximized if we maximize the expected utility increment during each time slot, which can be written as:

$$\sum_l \log(Q_l(\mathrm{E}[R_l(t)])) - \log(Q_l(\mathrm{E}[R_l(t-1)]))$$
$$= \sum_l \log\left(1 + \beta_l \frac{\mathrm{E}[R_l(t)] - \mathrm{E}[R_l(t-1)]}{Q_l(\mathrm{E}[R_l(t-1)])}\right)$$
$$= \sum_l \sum_{h \in \mathcal{P}_l} y_l^h \log\left(1 + \sum_r \sum_m \frac{\beta_l L_p x_{z_l, z_l', m}^{l,h,r}}{N_G T_s Q_l^{t-1}} (1 - p_{l,h}^r)\right)$$
$$= \sum_l \sum_{h \in \mathcal{P}_l} y_l^h \log\left(1 + \rho_l^t \sum_r \sum_m x_{z_l, z_l', m}^{l,h,r} (1 - p_{l,h}^r)\right),$$

where z_l' is the next hop from z_l on path $\mathcal{P}_l^h$, $p_{l,h}^r$ is the packet loss rate on tunnel r of path $\mathcal{P}_l^h$, $Q_l^{t-1} = Q_l(\mathrm{E}[R_l(t-1)])$, and $\rho_l^t = \beta_l L_p / (N_G T_s Q_l^{t-1})$.

From (11) and (12), the end-to-end packet loss rate for tunnel r on path $\mathcal{P}_l^h$ is:

$$p_{l,h}^r = 1 - \prod_{\{i,j\} \in \mathcal{P}_l^h} \prod_{m \in \mathcal{M}} (1 - p_{i,j}^m)^{x_{i,j,m}^{l,h,r}}. \quad (18)$$

We assume that each tunnel can only include one channel on each link. When there are multiple channels available at each link along the path, a CR source node can set up multiple tunnels to exploit the additional bandwidth. We then have the following constraint:

$$\sum_m x_{i,j,m}^{l,h,r} \le 1, \quad \forall \{i,j\} \in \mathcal{P}_l^h. \quad (19)$$

Considering availability of the channels, we further have,

$$\sum_r \sum_m x_{i,j,m}^{l,h,r} \le |\Omega_{i,j}|, \quad \forall \{i,j\} \in \mathcal{P}_l^h, \quad (20)$$

where $|\Omega_{i,j}|$ is the number of available channels on link $\{i,j\}$ defined in (9).

As discussed, each node is equipped with two transceivers: one for receiving and the other for transmitting video data during the transmission phase. Hence a channel cannot be used to receive and transmit data simultaneously at a relay node. We have for each channel m:

$$\sum_r x_{i,j,m}^{l,h,r} + \sum_r x_{j,k,m}^{l,h,r} \le 1,$$
$$\forall\, m, l, \forall\, h \in \mathcal{P}_l, \forall\, \{i,j\}, \{j,k\} \in \mathcal{P}_l^h. \quad (21)$$

Let n_l^h be the number of tunnels on path $\mathcal{P}_l^h$. For each source z_l and each destination d_l, the number of scheduled channels is equal to n_l^h. We have for each source node

$$\sum_r \sum_m x_{z_l, z_l', m}^{l,h,r} = n_l^h y_l^h, \quad \forall\, h \in \mathcal{P}_l, \forall\, l. \quad (22)$$

Let d_l' be the last hop to destination d_l on path $\mathcal{P}_l^h$, we have for each destination node

$$\sum_r \sum_m x_{d_l', d_l, m}^{l,h,r} = n_l^h y_l^h, \quad \forall\, h \in \mathcal{P}_l, \forall\, l. \quad (23)$$

At a relay node, the number of channels used to receive data is equal to that of channels used to transmit data, due to flow conservation and amplify-and-forward. At relay node j for session l, assume $\{i,j\} \in \mathcal{P}_l^h$ and $\{j,k\} \in \mathcal{P}_l^h$. We have,

$$\sum_r \sum_m x_{i,j,m}^{l,h,r} = \sum_r \sum_m x_{j,k,m}^{l,h,r},$$
$$\forall\, h \in \mathcal{P}_l, \forall\, l, \forall\, \{i,j\}, \{j,k\} \in \mathcal{P}_l^h. \quad (24)$$

We also consider hardware-related constraints on path selection. We summarize such constraints in the following general form for ease of presentation:

$$\sum_l \sum_{h \in \mathcal{P}_l} w_{l,h}^g y_l^h \le 1, \forall\, g. \quad (25)$$

To simplify exposition, we choose at most one path in $\mathcal{P}_l$ for video session l. Such a single path routing constraint can be expressed as $\sum_h y_l^h \le 1$, which is a special case of (25) where $w_{l,h}^1 = 1$ for all h, and $w_{l',h}^g = 0$ for all $g \ne 1$, $l' \ne l$, and h. We can also have $\sum_h y_l^h \le \xi$ to allow up to ξ paths for each video session. In order to achieve optimality in the general case of multi-path routing, an optimal scheduling algorithm should be designed to dispatch packets to paths with different conditions (e.g., different number of tunnels and delays).

There are also disjointedness constraints for the chosen paths. This is because each CR node is equipped with two transceivers and both will be used for a video session if it is included in a chosen path. Such disjointedness constraint is also a special case of (25) with the following definition for $w_{l,h}^g$ for each CR node g:

$$w_{l,h}^g = \begin{cases} 1, & \text{if node } g \in \text{path } \mathcal{P}_l^h \\ 0, & \text{otherwise,} \end{cases} \quad (26)$$

Finally we formulate the problem of multi-hop CR network video streaming (OPT-CRV) as:

$$\max: \sum_l \sum_{h \in \mathcal{P}_l} y_l^h \log\left(1 + \rho_l^t \sum_r \sum_m x_{z_l, z_l', m}^{l,h,r}(1 - p_{l,h}^r)\right) \quad (27)$$
$$\text{subject to: } (15) \sim (25).$$

B. Centralized Algorithm and Upper/Lower Bounds

Problem OPT-CRV is in the form of MINLP (without continuous variables), which is NP-hard in general. We first describe a centralized algorithm to derive performance bounds in this section, and then present a distributed algorithm based on dual decomposition in the next section.

We first obtain a relaxed *non-linear programming* (NLP) version of OPT-CRV. The binary variables $x_{i,j,m}^{l,h,r}$ and y_l^h are relaxed to take values in [0,1]. The integer variables n_l^h are treated as nonnegative real numbers. It can be shown that the relaxed problem has a concave object function and the constraints are convex. This relaxed problem can be solved using a constrained nonlinear optimization problem solver. If all the variables are integer in the solution, then we have the exact optimal solution. Otherwise, we obtain an infeasible solution, which produces an upper bound for the problem. This is given in Lines 1~2 in Table I.

We also develop a *sequential fixing algorithm* (SF) for solving OPT-CRV. The pseudo-code is given in Table I. SF iteratively solves the relaxed problem, fixing one or more integer variables after each iteration [3], [15]. In Table I, Lines 3~7 fix the path selection variables y_l^h, and Lines 8~17 fix the channel scheduling variables $x_{i,j,m}^{l,h,r}$ and tunnel variables n_l^h. The tunnel variables n_l^h can be computed using (22) after $x_{i,j,m}^{l,h,r}$ and y_l^h are solved. When the algorithm terminates, it produces a feasible solution that yields a lower bound for the objective value.

IV. DUAL DECOMPOSITION

SF is a centralized algorithm requiring global information. It may not be suitable for multi-hop wireless networks, although the upper and lower bounds provide useful insights on the performance limits. In this section, we develop a distributed algorithm for Problem OPT-CRV and analyze its optimality and convergence performance.

A. Decompose Problem OPT-CRV

Since the domains of $x_{i,j,m}^{l,h,r}$ defined in (19)~(24) for different paths do not intersect with each other, we can decompose Problem OPT-CRV into two subproblems. The first subproblem deals with channel scheduling for maximizing the

TABLE I
SEQUENTIAL FIXING ALGORITHM (SF) FOR PROBLEM OPT-CRV

1 :	Relax integer variables $x_{i,j,m}^{l,h,r}$, y_l^h, and n_l^h;
2 :	Solve the relaxed problem using a constrained NLP solver;
3 :	**if** (there is y_l^h not fixed)
4 :	Find the largest $y_{l'}^{h'}$, where $[l', h'] = \arg\max\{y_l^h\}$, and fix it to 1;
5 :	Fix other y_l^h's according to constraint (25);
6 :	Go to Step 2;
7 :	**end if**
8 :	**if** (there is $x_{i,j,m}^{l,h,r}$ not fixed)
9 :	Find the largest $x_{i',j',m'}^{l',h',r'}$, where $[i', j', m', l', h', r'] =$ $\arg\max\{x_{i,j,m}^{l,h,r}\}$, and set it to 1;
10:	Fix other $x_{i,j,m}^{l,h,r}$'s according to the constraints;
11:	**if** (there is other variable that is not fixed)
12:	Go to Step 2;
13:	**else**
14:	Fix n_l^h's based on **x** and **y**;
15:	Exit with feasible solution $\{\mathbf{x}, \mathbf{y}, \mathbf{n}\}$;
16:	**end if**
17:	**end if**

TABLE II
GREEDY ALGORITHM FOR CHANNEL SCHEDULING

1 :	Initialization: tunnel $r = 1$, link $\{i, j\}$'s from z_l to d_l;
2 :	**if** ($|\Lambda_{i,j}| == 0$)
3 :	Exit;
4 :	**else if** ($|\Lambda_{i,j}| == 1$)
5 :	Assign the single channel in $\Lambda_{i,j}$, m', to tunnel r;
6 :	Check neighboring link $\{k, i\}$;
7 :	**if** ($p_{k,i}^{m'} \in \Lambda_{k,i}$)
8 :	Remove $p_{k,i}^{m'}$ from $\Lambda_{k,i}$, $i \leftarrow k$, $j \leftarrow i$ and go to Step 2;
9 :	**else**
10:	Go to Step 13;
11:	**end if**
12:	**else**
13:	Put $\Lambda_{i,j}$ in set Λ_l^h;
14:	**if** (node j is not destination d_l)
15:	$i \leftarrow j$, $j \leftarrow v$;
16:	Go to Step 2;
17:	**end if**
18:	**end if**
19:	**while** (Λ_l^h is not empty)
20:	Find the maximum value $p_{i',j'}^{m'}$ in set Λ_l^h $\{i', j', m'\} = \arg\min\{p_{i,j}^m\}$;
21:	Assign channel m' to tunnel r;
22:	Remove set $\Lambda_{i',j'}$ from set Λ_l^h;
23:	Check neighboring link $\{k, i\}$ and $\{j, v\}$;
24:	**if** ($p_{k,i}^{m'} \in \Lambda_{k,i}$ and $\Lambda_{k,i} \in \Lambda_l^h$)
25:	Remove $p_{k,i}^{m'}$ from $\Lambda_{k,i}$;
26:	**if** ($\Lambda_{k,i}$ is empty)
27:	Exit;
28:	**end if**
29:	**end if**
30:	**if** ($p_{j,v}^{m'} \in \Lambda_{j,v}$ and $\Lambda_{j,v} \in \Lambda_l^h$)
31:	Remove $p_{j,v}^{m'}$ from $\Lambda_{j,v}$;
32:	**if** ($\Lambda_{j,v}$ is empty)
33:	Exit;
34:	**end if**
35:	**end if**
36:	**end while**
37:	Compute the next tunnel: $r \leftarrow r + 1$ and go to Step 2;

expected utility on a chosen path $\mathcal{P}_l^h$. We have the *channel scheduling* problem (OPT-CS) as:

$$H_l^h = \max_{\mathbf{x}} \sum_r \sum_m x_{z_l, z_l', m}^{l,h,r}(1 - p_{l,h}^r) \qquad (28)$$

subject to: (19) $\sim$ (24), $x_{z_l, z_l', m}^{l,h,r} \in \{0, 1\}$, for all l, h, r, m.

In the second part, optimal paths are selected to maximize the overall objective function. Letting $F_l^h = \log\left(1 + \rho_l^T H_l^h\right)$, we have the following *path selection* problem (OPT-PS):

maximize: $f(\mathbf{y}) = \sum_l \sum_h F_l^h y_l^h$ (29)

subject to: $\sum_l \sum_{h \in \mathcal{P}_l} w_{l,h}^g y_l^h \leq 1$, for all g

                $y_l^h \in \{0, 1\}$, for all l, h.

B. Solve the Channel Scheduling Subproblem

We have the following result for assigning available channels at a relay node.

Theorem 1: Consider three consecutive nodes along a path, denoted as nodes i, j, and k. Idle channels 1 and 2 are available at link $\{i, j\}$ and idle channels 3 and 4 are available at link $\{j, k\}$. Assume the packet loss rates of the four channels satisfy $p_{i,j}^1 > p_{i,j}^2$ and $p_{j,k}^3 > p_{j,k}^4$. To set up two tunnels, assigning channels $\{1, 3\}$ to one tunnel and channels $\{2, 4\}$ to the other tunnel achieves the maximum expectation of successful transmission on path section $\{i, j, k\}$.

Proof: Let the success probabilities on the channels be $\tilde{p}_{i,j}^1 = 1 - p_{i,j}^1$, $\tilde{p}_{i,j}^2 = 1 - p_{i,j}^2$, $\tilde{p}_{j,k}^3 = 1 - p_{j,k}^3$, and $\tilde{p}_{j,k}^4 = 1 - p_{j,k}^4$. We have $\tilde{p}_{i,j}^1 < \tilde{p}_{i,j}^2$ and $\tilde{p}_{j,k}^3 < \tilde{p}_{j,k}^4$. Comparing the success probabilities of the channel assignment given in Theorem 1 and that of the alternative assignment, we have $\tilde{p}_{i,j}^1 \tilde{p}_{j,k}^3 + \tilde{p}_{i,j}^2 \tilde{p}_{j,k}^4 - \tilde{p}_{i,j}^1 \tilde{p}_{j,k}^4 - \tilde{p}_{i,j}^2 \tilde{p}_{j,k}^3 = (\tilde{p}_{i,j}^1 - \tilde{p}_{i,j}^2)(\tilde{p}_{j,k}^3 - \tilde{p}_{j,k}^4) > 0$. The result follows. ∎

According to Theorem 1, a greedy approach, which always chooses the channel with the lowest loss rate at each link when setting up tunnels along a path, produces the optimal overall success probability. More specifically, when there is only one tunnel to be set up along a path, the tunnel should consist of the most reliable channels available at each link along the path. When there are multiple tunnels to set up along a path, tunnel 1 should consist of the most reliable channels that are available at each link; tunnel 2 should consist of the second most reliable links available at each link; and so forth.

Define the set of loss rates of the available channels on link $\{i, j\}$ as $\Lambda_{i,j} = \{p_{i,j}^m | m \in \Omega_{i,j}\}$. The greedy algorithm is given in Table II, with which each video source node solves Problem OPT-CS for each feasible path. Lines 2$\sim$3 in Table II checks if there is more channels to assign and the algorithm terminates if no channel is left. In Lines 4$\sim$11, links with only one available channel are assigned to tunnel r and the neighboring links with the same available channels are removed due to constraint (21). In Lines 12$\sim$18, links with more than two channels are grouped to be assigned later. In Lines 19$\sim$21, the available channel with the lowest packet loss rate is assigned to tunnel r at each unallocated link, according to Theorem 1. To avoid co-channel interference, the same channel on neighboring links is removed as in Lines 22$\sim$35.

C. Solve the Path Selection Subproblem

To solve Problem OPT-PS, we first relax binary variables y_l^h to allow them take real values in [0,1] and obtain the following

TABLE III
DISTRIBUTION ALGORITHM FOR PATH SELECTION

1:	Initialization: set $\tau = 0$, $e_g(0) > 0$ and step size $s \in [0, 1]$;
2:	Each source locally solves the lower level problem in (32);
	if $(F_l^h - \sum_g d_{l,h}^g e_g(\tau)) > 0)$ $y_l^h = y_l^h + s$, $y_l^h = \min\{y_l^h, 1\}$;
	else $y_l^h = y_l^h - s$, $y_l^h = \max\{y_l^h, 0\}$;
3:	Broadcast solution $y_l^h(\mathbf{e}(\tau))$;
4:	Each source updates $\mathbf{e}$ according to (34) and broadcasts $\mathbf{e}(\tau + 1)$ through the common control channel;
5:	$\tau \leftarrow \tau + 1$ and go to Step 2 until termination criterion is satisfied;

relaxed path selection problem (OPT-rPS):

$$\text{maximize:} \quad f(\mathbf{y}) = \sum_l \sum_h F_l^h y_l^h \quad (30)$$
$$\text{subject to:} \quad \sum_l \sum_{h \in \mathcal{P}_l} w_{l,h}^g y_l^h \leq 1, \text{ for all } g$$
$$0 \leq y_l^h \leq 1, \text{ for all } h, l.$$

We then introduce positive Lagrange Multipliers e_g for the path selection constraints in Problem OPT-rPS and obtain the corresponding *Lagrangian function*:

$$\mathcal{L}(\mathbf{y}, \mathbf{e}) = \sum_l \sum_h F_l^h y_l^h + \sum_g e_g (1 - \sum_l \sum_h w_{l,h}^g y_l^h) \quad (31)$$
$$= \sum_l \sum_h (F_l^h y_l^h - \sum_g w_{l,h}^g y_l^h e_g) + \sum_g e_g$$
$$= \sum_l \sum_h \mathcal{L}_l^h(y_l^h, \mathbf{e}) + \sum_g e_g.$$

Problem (31) can be decoupled since the domains of y_l^h's do not overlap. Relaxing the coupling constraints, it can be decomposed into two levels. At the lower level, we have the following subproblems, one for each path $\mathcal{P}_l^h$,

$$\max_{0 \leq y_l^h \leq 1} \mathcal{L}_l^h(y_l^h, \mathbf{e}) = F_l^h y_l^h - \sum_g w_{l,h}^g y_l^h e_g. \quad (32)$$

At the higher level, by updating the dual variables e_g, we can solve the *relaxed dual problem*:

$$\min_{\mathbf{e} \geq 0} \quad q(\mathbf{e}) = \sum_l \sum_h \mathcal{L}_l^h \left((y_l^h)^*, \mathbf{e} \right) + \sum_g e_g, \quad (33)$$

where $(y_l^h)^*$ is the optimal solution to (32). Since the solution to (32) is unique, the relaxed dual problem (33) can be solved using the following *subgradient method* that iteratively updates the Lagrange Multipliers [16]:

$$e_g(\tau + 1) = \left[e_g(\tau) - \alpha(\tau)(1 - \sum_l \sum_h w_{l,h}^g y_l^h) \right]^+, \quad (34)$$

where τ is the iteration index, $\alpha(\tau)$ is a sufficiently small positive step size and $[x]^+$ denotes $\max\{x, 0\}$. The pseudo code for the distributed algorithm is given in Table III.

D. Optimality and Convergence Analysis

The distributed algorithm in Table III iteratively updates the dual variables until they converge to stable values. In this section, we first prove that the solution obtained by the distributed algorithm is also optimal for the original path selection problem OPT-PS. We then derive the convergence condition for the distributed algorithm.

Fact 1 ([16]): Consider a linear problem involving both equality and inequality constraints

$$\text{maximize:} \quad \mathbf{a}'\mathbf{x} \quad (35)$$
$$\text{subject to:} \quad \mathbf{h}_1'\mathbf{x} = b_1, \cdots, \mathbf{h}_m'\mathbf{x} = b_m$$
$$\mathbf{g}_1'\mathbf{x} \leq c_1, \cdots, \mathbf{g}_r'\mathbf{x} \leq c_r,$$

where $\mathbf{a}$, $\mathbf{h}_i$, and $\mathbf{g}_j$ are column vectors in $\mathcal{R}_n$, b_i's and c_j's are scalars, and $\mathbf{a}'$ is the *transpose* of $\mathbf{a}$. For any feasible point $\mathbf{x}$, the set of *active* inequality constraints is denoted by $\mathcal{A}(\mathbf{x}) = \{j | \mathbf{g}_j'\mathbf{x} = c_j\}$. If $\mathbf{x}^*$ is a maximizer of inequality constrained problem (35), $\mathbf{x}^*$ is also a maximizer of the following equality constrained problem:

$$\text{maximize:} \quad \mathbf{a}'\mathbf{x} \quad (36)$$
$$\text{subject to:} \quad \mathbf{h}_1'\mathbf{x} = b_1, \cdots, \mathbf{h}_m'\mathbf{x} = b_m$$
$$\mathbf{g}_j'\mathbf{x} = c_j, \forall j \in \mathcal{A}(\mathbf{x}).$$

Lemma 1: The optimal solution for the relaxed primal problem OPT-rPS in (30) is also feasible and optimal for the original Problem OPT-PS in (29).

Proof: According to Fact 1, the linearized problem of OPT-PS, i.e., OPT-rPS, can be rewritten as an equality constrained problem in the following form:

$$\text{maximize:} \quad \mathbf{F}'\mathbf{y} \quad (37)$$
$$\text{subject to:} \quad \mathbf{w}_j'\mathbf{y} = 1, \quad j \in \mathcal{A}(\mathbf{y}^*) \quad (38)$$
$$0 \leq y_l^h \leq 1, \text{ for all } h, l,$$

where $\mathbf{F}$, $\mathbf{w}_j$'s, and $\mathbf{y}$ are column vectors with elements F_l^h, $w_{l,h}^g$, and y_l^h, respectively. We apply *Gauss-Jordan elimination* to the constraints in (38) to solve for $\mathbf{y}$. Since there is not sufficient number of equations, some y_l^h's are free variables (denoted as y_i^f) and the rest are dependent variables (denoted as y_j^d). Assuming there are r free variables, the dependent variables can be written as linear combinations of the free variables after Gauss-Jordan elimination, as

$$y_j^d = \sum_{i=1}^r \bar{w}_j^i y_i^f + \bar{b}_j, \ j \in \mathcal{A}(y_i^*). \quad (39)$$

Due to Gauss-Jordan elimination and binary vectors $\mathbf{w}_j$'s, $\bar{w}_j^i$ and $\bar{b}_j$ in (39) are all integers. Therefore, if all the free variables y_i^f attain binary values, then all the dependent variables y_j^d computed using (39) will also be integers. Since $0 \leq y_j^d \leq 1$, being integers means that they are either 0 or 1, i.e., binaries. That is, such a solution will be feasible.

Next we substitute (39) into problem (37) to eliminate all the dependent variables. Then we obtain a unconstrained problem with only r free variables, as

$$\text{maximize:} \quad \sum_{i=1}^r \bar{F}_i y_i^f + \bar{b}_0 \quad (40)$$

Since the free variables y_i^f's take value in $\{0, 1\}$, this problem can be easily solved as follows. If the coefficient $\bar{F}_i > 0$, we set $y_i^f = 1$; otherwise, if $\bar{F}_i < 0$, we set $y_i^f = 0$. Thus (40) achieves its maximum objective value. Once all the free variables are determined with their optimal binary values, we computes the dependent variables using (39), which are also binary as discussed above. Thus we obtain a feasible solution, which is optimal. ∎

Lemma 2: If the relaxed primal Problem OPT-rPS in (30) has an optimal solution, then the relaxed dual problem (33) also has an optimal solution and the corresponding optimal values of the two problems are identical.

Proof: By definition, the problems in (31) and (33) are primal/dual problems. The primal problem always has an optimal solution because it is bounded. Since Problem OPT-rPS is an LP problem, the relaxed dual problem is also

bounded and feasible. Therefore the relaxed dual problem also has an optimal solution. We have the *strong duality* if the primal problem is convex, which is the case here since Problem OPT-rPS is an LP problem. ∎

We have Theorem 2 on the optimality of the path selection solution, which follows naturally from Lemmas 1 and 2.

Theorem 2: The optimal solution to the relaxed dual problem (32) and (33) is also feasible and optimal to the original path selection Problem OPT-PS given in (29).

As discussed, the relaxed dual problem (33) can be solved using the *subgradient method* that iteratively updates the Lagrange Multipliers. We have the following theorem on the convergence of the distributed algorithm given in Table III.

Theorem 3: Let $\mathbf{e}^*$ be the optimal solution. The distributed algorithm in Table III converges if the step sizes $\alpha(\tau)$ in (34) satisfy the following condition:

$$0 < \alpha(\tau) < \frac{2\left[q(\mathbf{e}(\tau)) - q(\mathbf{e}^*)\right]}{||G(\tau)||^2}, \quad \text{for all } \tau, \tag{41}$$

where $G(\tau)$ is the gradient of $q(\mathbf{e}(\tau))$.

Proof: Since $q(\mathbf{e}(\tau))$ is a linear function, we have subgradient equality, as

$$q(\mathbf{e}(\tau)) - q(\mathbf{e}^*) = [\mathbf{e}(\tau) - \mathbf{e}^*]' G(\tau).$$

It then follows that

$$||\mathbf{e}(\tau) - \alpha(\tau)G(\tau) - \mathbf{e}^*||^2$$
$$= ||\mathbf{e}(\tau) - \mathbf{e}^*||^2 - 2\alpha(\tau)[\mathbf{e}(\tau) - \mathbf{e}^*]'G(\tau) + (\alpha(\tau))^2||G(\tau)||^2$$
$$= ||\mathbf{e}(\tau) - \mathbf{e}^*||^2 - 2\alpha(\tau)[q(\mathbf{e}(\tau)) - q(\mathbf{e}^*)] + (\alpha(\tau))^2||G(\tau)||^2.$$
$$\tag{42}$$

If $\alpha(\tau)$ satisfy (41), the sum of the last two terms in (42) is negative. It follows that, $||\mathbf{e}(\tau) - \alpha(\tau)G(\tau) - \mathbf{e}^*|| < ||\mathbf{e}(\tau) - \mathbf{e}^*||$. Since the projection operation is *nonexpansive*, we have,

$$||\mathbf{e}(\tau + 1) - \mathbf{e}^*|| = ||[\mathbf{e}(\tau) - \alpha(\tau)G(\tau)]^+ - [\mathbf{e}^*]^+||$$
$$\leq ||\mathbf{e}(\tau) - \alpha(\tau)G(\tau) - \mathbf{e}^*|| < ||\mathbf{e}(\tau) - \mathbf{e}^*||,$$

which states the conditional convergence of the algorithm. ∎

Since the optimal solution $\mathbf{e}^*$ is not known a priori, we use the following approximation in the algorithm: $\alpha(\tau) = \frac{q(\mathbf{e}(\tau)) - \hat{q}(\tau)}{||G(\tau)||^2}$, where $\hat{q}(\tau)$ is the current estimate for $q(\mathbf{e}^*)$. We choose the mean of the objective values of the relaxed primal and dual problems for $\hat{q}(\tau)$.

E. Practical Considerations

The distributed algorithms are based on the fact that the computation is distributed on each feasible path. The OPT-CS algorithm requires information on channel availability and packet loss rates at the links of feasible paths. The OPT-PS algorithm computes the primal variable y_l^h for each path and broadcasts Lagrangian multipliers over the control channel to all the source nodes. We assume a perfect control channel such that channel information can be effectively distributed and shared, which is not confined by the time slot structure [10].

We assume relatively large timescales for the primary network time slots, and small to medium diameter for the CR network, such that there is sufficient time for timely feedback of channel information to the video source nodes and for
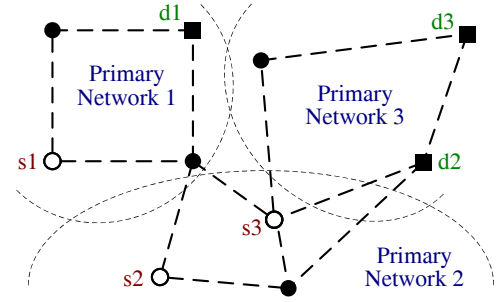

Fig. 3.   Topology of the multi-hop CR network. Note that only video source nodes, video destination nodes, and those nodes along the precomputed paths are shown in the topology.

the distributed algorithms to converge. Otherwise, channel information can be estimated using (5) based on delayed feedback, leading to suboptimal solutions. If the time slot is too short, the distributed algorithm may not converge to the optimal solution (see Fig. 7). We focus on developing the CR video framework in this paper, and will investigate these issues in our future work.

V. SIMULATION STUDIES

A. Methodology and Simulation Settings

We implement the proposed algorithms with a combination of C and MATLAB (i.e., for solving the relaxed NLP problems), and evaluate their performance with simulations. For the results reported in this section, we have $K = 3$ primary networks and $M = 10$ channels. There are 56, 55, and 62 CR users in the coverage areas of primary networks 1, 2, and 3, respectively. The $|\mathcal{U}_m^1|$'s are [5 4 6 4 8 7 5 6 7 4] (i.e., five users sense channel 1, four users sense channel 2, and so forth); the $|\mathcal{U}_m^2|$'s are [4 6 5 7 6 5 3 8 5 6], and the $|\mathcal{U}_m^3|$'s are [8 6 5 4 7 6 8 5 6 7]. The topology is shown in Fig. 3.

We choose $L_p = 100$, $T_s = 0.02$ and $N_G = 10$. The channel utilization is $\eta_m^k = 0.6$ for all the channels. The probability of false alarm is $\epsilon_m^k = 0.3$ and the probability of miss detection is $\delta_m^k = 0.2$ for all m and k, unless otherwise specified. Channel parameters λ_m^k and μ_m^k are set between $(0, 1)$. The maximum allowed collision probability γ_m^k is set to 0.2 for all the M channels in the three primary networks.

We consider three video sessions, each streaming a video in the Common Intermediate Format (CIF, 352×288), i.e., *Bus* to destination 1, *Foreman* to destination 2, and *Mother & Daughter* to destination 3. The frame rate is 30 fps, and a GOP consists of 10 frames. The duration of a time slot is 0.02 seconds and each GOP should be delivered in 0.2 seconds (i.e., 10 time slots).

We compare four schemes in the simulations: (i) the upper-bounding solution by solving the relaxed version of Problem OPT-CRV using an NLP solver, (ii) the proposed distributed algorithm in Tables II and III, (iii) the sequential fixing algorithm given in Table I, which computes a lower-bounding solution, and (iv) a greedy heuristic where at each hop, the link with the most available channels is used. Each point in the figures is the average of 10 simulation runs, with 95% confidence intervals plotted as error bars in the figures. The 95% confidence intervals are negligible in all the figures.
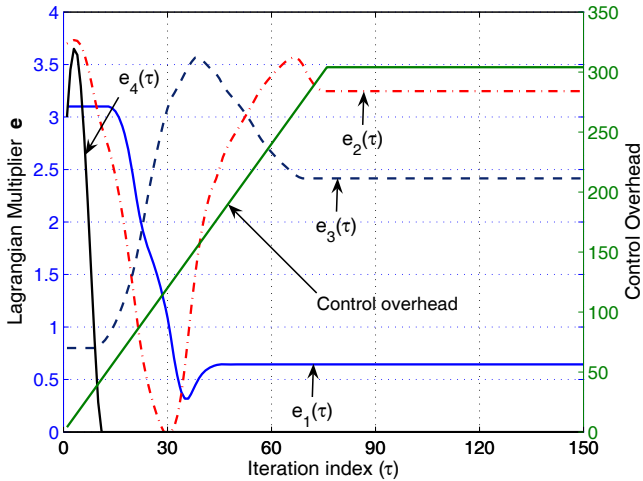
Fig. 4.   Illustrate the convergence of the distributed algorithm.

B. Simulation Results

1) Algorithm Performance: To demonstrate the convergence of the distributed algorithm, we plot the traces of the four Lagrangian multipliers in Fig. 4. We observe that all the Lagrangian multipliers converge to their optimal values after 76 iterations. We also plot the control overhead as measured by the number of distinct broadcast messages for $e_i(\tau)$ using the y-axis on the right-hand side. The overhead curve increases linearly with the number of iterations and gets flat (i.e., no more broadcast message) when all the Lagrangian multipliers converge to their optimal values.

We examine the impact of spectrum sensing errors in Fig. 5. We test six sensing error combinations $\{\epsilon_m, \delta_m\}$ as follows: $\{0.1, 0.5\}$, $\{0.2, 0.3\}$, $\{0.3, 0.2\}$, $\{0.5, 0.11\}$, $\{0.7, 0.06\}$, and $\{0.9, 0.02\}$, and plot the average PSNR values of the Foreman session. It is interesting to see that the best video quality is achieved when the false alarm probability ϵ_m is between 0.2 and 0.3. Since the two error probabilities are correlated, increasing one will generally decrease the other. With a larger ϵ_m, CR users are more likely to waste spectrum opportunities that are actually available, leading to lower bandwidth for videos and poorer video quality, as shown in Fig. 5. On the other hand, a larger δ_m implies more aggressive spectrum access and more severe interference to primary users. Therefore when ϵ_m is lower than 0.2 (and δ_m is higher than 0.3), the CR nodes themselves also suffer from the collisions and the video quality degrades.

2) Impact of Primary Network Parameters: In Fig. 6, we examine the impact of channel utilization η on received video quality. We focus on Session 2 with the Foreman sequence. The average PSNRs achieved by the four schemes are plotted when η is increased from 0.6 to 0.9 for all licensed channels. Intuitively, a smaller η allows more transmission opportunities for CR nodes, leading to improved video quality. This is illustrated in the figure where all the four curves decrease as η gets larger. The distributed scheme achieves PSNRs very close to that obtained by sequential fixing, and both of them are close to the upper bound. The heuristic scheme is inefficient in exploiting the available spectrum even when the channel utilization is low.
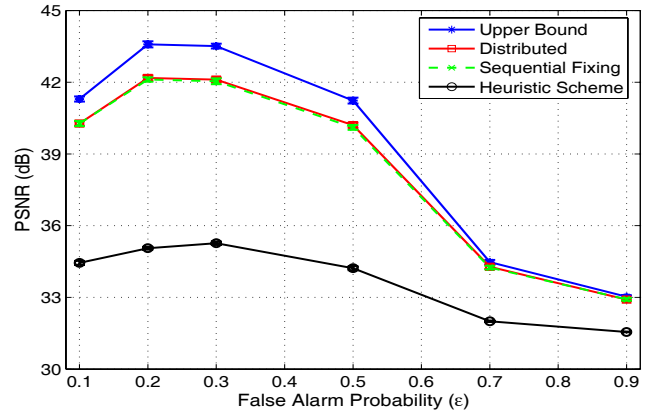

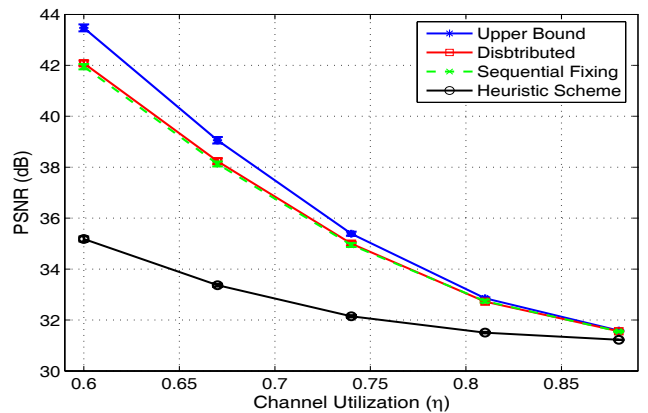Fig. 5.   Video PSNR versus spectrum sensing error.

Fig. 6.   Video PSNR versus primary user channel utilization η.

As discussed, the time slot duration is also an important parameter that may affect the convergence of the distributed algorithm. In Fig. 7, we keep the same network and video session settings, while increasing the time slot duration as 4 ms, 10 ms, 20 ms, 40ms and 100 ms. For a given time slot duration, we let the distributed algorithm run for 5% of the time slot duration, starting from the beginning of the time slot, and then stop. The solution that the algorithm produces when it is stopped will be used for video transmission in the remainder of this time slot. It can be seen that when the time slot is 4 ms, the algorithm does not converge after $5\% \times 4 = 0.2$ ms, and the PSNR produced by the distributed algorithm is low (but still higher than that of the heuristic algorithm). When the time slot duration is sufficiently large (e.g., over 10 ms), the algorithm can converge and the proposed algorithm produces very good video quality as compared to the upper bound and the lower bound given by the sequential fixing algorithm.

3) Comparison of MPEG-4 FGS and H.264/SVC MGS Videos: Finally, we compare MPEG-4 FGS and H.264/SVC MGS videos, while keeping the same settings. It has been shown that H.264/SVC has better rate-distortion performance than MPEG-4 FGS due to the use of efficient hierarchical prediction structures, the inter-layer prediction mechanisms, improved drift control mechanism, and the efficient coding scheme in H.264/AVC [5]. Although MGS has Network Abstraction Layer (NAL) unit-based granularity, it achieves similar rate-distortion performance as H.264/SVC FGS [5].
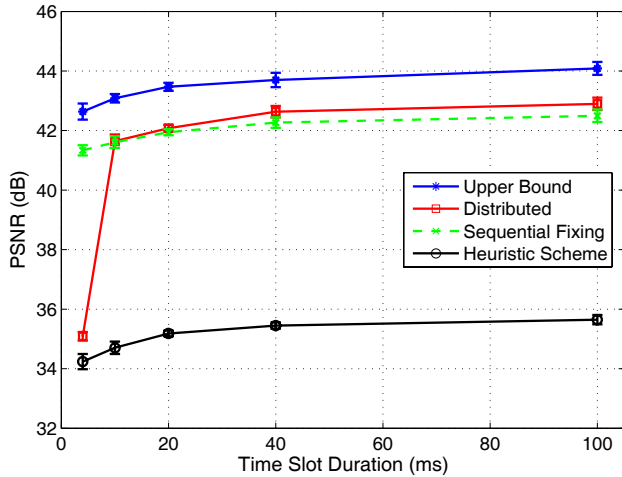
Fig. 7.   Impact of time slot duration on received video quality.

Fig. 9.   Comparison of MPEG-4 FGS video with H.264/SVC MGS video under various false alarm probabilities.
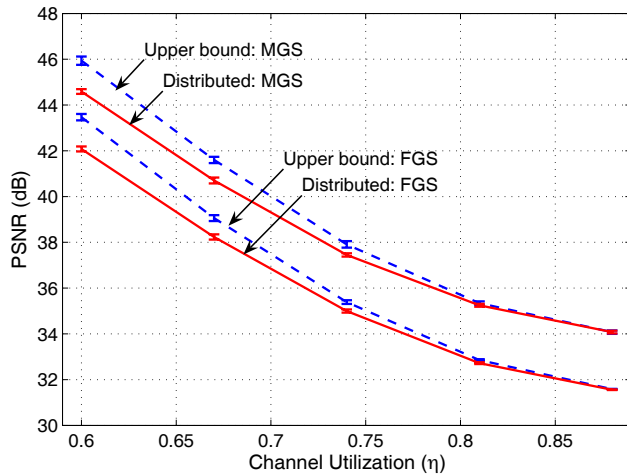

Fig. 8.   Comparison of MPEG-4 FGS video with H.264/SVC MGS video under various channel utilizations.

We plot the upper bounds and the distributed algorithm results in Figs. 8 and 9 for various channel utilizations and false alarm probabilities, respectively. From the figures, it can be observed that there is a gap about 2.5 dB between the H.264/SVC MGS and MPEG-4 FGS curves, which clearly demonstrates the rate-distortion efficiency of MGS over MPEG-4 FGS. The proposed algorithm can handle both MGS and FGS videos, and the same trend is observed in both cases.

VI. RELATED WORK

The high potential of CRs has attracted considerable interest from the wireless community [1], [2], [17]. The mainstream CR research has been focused on spectrum sensing and dynamic spectrum access issues [9]–[11], [13], [18]. Several papers have addressed the impact of spectrum sensing errors on the design of spectrum access schemes [18]–[21]. The approach of iteratively sensing a selected subset of available channels has been adopted in the design of CR MAC protocols (e.g., see [10], [13], [18]). Our work is complementary to this class of work by providing an important application for the enhanced spectrum efficiency achieved by spectrum sensing and access schemes reported in the literature.
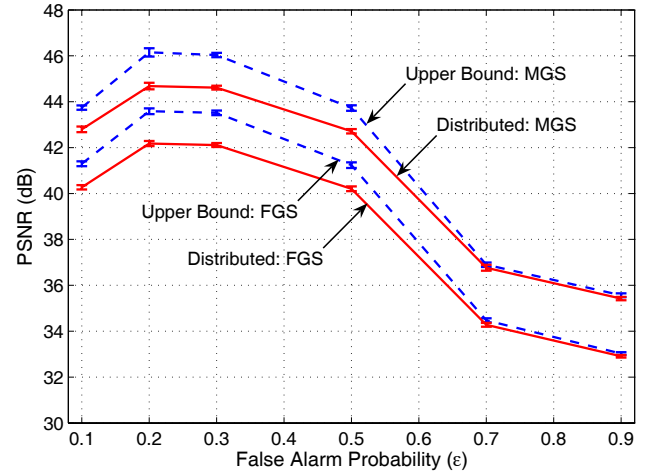
Multi-hop SDR or CR networks have been studied in a few recent works [15], [22], [23]. The authors formulate cross-layer optimization problems considering factors from the PHY up to the transport layer. Distributed algorithms are developed by applying the dual decomposition technique [16], [24]. We adopt similar methodology in this paper, but address the more challenging problem of streaming real-time videos.

The problem of QoS provisioning in CR networks has been considered in a few papers [10], [20], [25], where the focus is still on the so-called "network-centric" metrics such as maximum throughput and delay [10], [20]. In a recent work [20], Urgaonkar and Neely derive an interesting delay throughput trade-off for a multi-cell cognitive radio network, while primary user protection is achieved by stabilizing a virtual "collision queue."

In [25], a game-theoretic framework is described for resource allocation for multimedia transmissions in spectrum agile wireless networks. In this interesting work, each wireless station plays a resource management game, which is coordinated by a network moderator. A mechanism-based resource management scheme determines the amount of transmission time to be allocated to various users on different frequency bands such that certain global system metrics are optimized. In our prior work [3], we consider video multicast in an infrastructure-based CR network. We present effective greedy heuristic algorithms for scheduling video data, with proved optimality bound and low computational complexity. In this paper, we consider the challenging case of multi-hop CR networks, where distributed algorithms are highly appealing.

VII. CONCLUSION

We studied the problem of streaming multiple scalable videos in a multi-hop CR network. The problem formulation considered spectrum sensing and sensing errors, spectrum access and primary user protection, video quality and fairness, and channel/path selection for concurrent video sessions. We first solved the formulated MINLP problem using a sequential fixing scheme that produces lower and upper bounds on the achievable video quality. We then applied dual decomposition

to derive a distributed algorithm, and analyzed its optimality and convergence performance. Our simulations validated the efficacy of the proposed scheme.

REFERENCES

[1] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Netw. J.*, vol. 50, no. 9, pp. 2127-2159, Sep. 2006.

[2] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79-89, May 2007.

[3] D. Hu, S. Mao, and J. Reed, "On video multicast in cognitive radio networks," in *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2222-2230.

[4] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 53-68, Mar. 2001.

[5] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194-1203, Sep. 2007.

[6] N. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 3062-3080, Nov. 2004.

[7] R. Kesavan and D. K. Panda, "Efficient multicast on irregular switch-based cut-through networks with up-down routing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, no. 8, pp. 808-828, Aug. 2001.

[8] R. Ramanathan, "Challenges: a radically new architecture for next generation mobile ad-hoc networks," in *Proc. ACM MobiCom'05*, Cologne, Germany, Sep. 2005, pp. 132-139.

[9] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: a hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106-117, Jan. 2008.

[10] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118-129, Jan. 2008.

[11] A. Motamedi and A. Bahai, "MAC protocol design for spectrum-agile wireless networks: stochastic control approach," in *Proc. IEEE DySPAN'07*, Dublin, Ireland, Apr. 2007, pp. 448-451.

[12] H. Mahmoud, T. Yücek, and H. Arslan, "OFDM for cognitive radio: merits and challenges," *IEEE Wireless Commun.*, vol. 16, no. 2, pp. 6-14, Apr. 2009.

[13] Q. Zhao, S. Geirhofer, L. Tong, and B. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 36, no. 2, pp. 785-796, Feb. 2008.

[14] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1752-1763, Dec. 2003.

[15] Y. Hou, Y. Shi, and H. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 146-155, Jan. 2008.

[16] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.

[17] Y. Zhao, S. Mao, J. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: metrics, utility functions, and methodologies," *Proc. IEEE*, vol. 97, no. 4, pp. 642-659, Apr. 2009.

[18] D. Hu and S. Mao, "Design and analysis of a sensing error-aware MAC protocol for cognitive radio networks," in *Proc. IEEE GLOBECOM'09*, Honolulu, HI, Nov./Dec. 2009.

[19] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053-2071, May 2008.

[20] R. Urgaonkar and M. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 766-777, June 2009.

[21] T. Shu and M. Krunz, "Throughput-efficient sequential channel sensing and probing in cognitive radio networks under sensing errors," in *Proc. ACM MobiCom'09*, Beijing, China, Sep. 2009, pp. 37-48.

[22] Y. Hou, Y. Shi, and H. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," in *Proc. IEEE INFOCOM'07*, Anchorage, AK, Apr. 2007, pp. 1-9.

[23] Z. Feng and Y. Yang, "Joint transport, routing and spectrum sharing optimization for wireless networks with frequency-agile radios," in *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1665-1673.

[24] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439-1451, Aug. 2006.

[25] A. Fattahi, F. Fu, M. van der Schaar, and F. Paganni, "Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 601-612, Apr. 2007.

Donglin Hu received the M.S. degree from Tsinghua University, Beijing, China, in 2007 and the B.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China in 2004, respectively, all in electrical engineering. Since 2007, he has been pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL. His research interests include cognitive radio networks, cross-layer optimization, algorithm design for wireless network and multimedia communications.

Shiwen Mao (S'99-M'04-SM'09) received a Ph.D. in electrical and computer engineering from Polytechnic Institute of New York University, Brooklyn, NY (formerly known as Polytechnic University) in 2004. He was a research staff member with IBM China Research Lab from 1997 to 1998. He was a Research Scientist in the Bradley Department of Electrical and Computer Engineering at Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA from 2003 to 2006. Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL.

His research interests include cross-layer optimization of wireless networks and multimedia communications, with current focus on cognitive radio networks and free space optical networks. He is on the Editorial Board of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, *Elsevier Ad Hoc Networks Journal*, *Wiley International Journal of Communication Systems*, and *ICST Transactions on Mobile Communications and Applications*. He chairs the Interest Group on Cross-layer Design for Multimedia Communications of IEEE Communications Society's Multimedia Communications Technical Committee.

Dr. Mao is a coauthor of *TCP/IP Essentials: A Lab-Based Approach* (Cambridge, U.K.: Cambridge University Press, 2004). Dr. Mao received the US National Science Foundation (NSF) Faculty Early Career Development Award (CAREER) in 2010, the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems, and the Best Paper Runner-up Award from the Fifth International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine) 2008. He holds one US patent.