Massive Live Video Distribution using Hybrid Cellular and Ad Hoc Networks

Ngoc Minh Do*†, Cheng-Hsin Hsu*, Jatinder Pal Singh*, and Nalini Venkatasubramanian†
*Deutsche Telekom Inc. R&D Laboratories, 5050 El Camino Real 221, Los Altos, CA 94022
†Department of Computer Science, University of California, Irvine, CA 92697

***Abstract*—This paper addresses the problem of disseminating multiple live videos to mobile users by using a hybrid cellular and ad hoc network. Specifically, we develop techniques to optimize the overall quality of video delivery by: (a) exploiting the flexibility of layered videos for in-network adaptation to reduce the gap between video coding rate and network capacity, and (b) alleviating the load of individually handling a large number of flows at the cell tower by using device-to-device ad hoc connectivity to deliver videos. We study the problem of optimally choosing the mobile devices that will serve as gateways from the cellular to the ad hoc network, the ad hoc routes from the gateway to individual devices, and the layers to deliver on these ad hoc routes. We develop a Mixed Integer Linear Program (MILP) based solution to the considered problem. We also develop a heuristic algorithm to select the devices, routes, and layers more efficiently than the ideal, but potentially time-consuming MILP-based algorithm. We evaluate the proposed techniques via through simulations. The simulation results show that the proposed algorithms significantly outperform the current solution in terms of overall video quality, transmission latency, delivery ratio, and missed frame ratio. For example, compared to the current cellular network, the MILP-based and the heuristic algorithms result in up to 20 dB higher video quality. Furthermore, the heuristic algorithm runs efficiently yet achieves near-optimal quality: at most 2.3 dB gap across all experiments.**

***Keywords*-wireless networks, video streaming, quality optimization, resource allocation**

I. Introduction

Recent market forecasts predict that mobile data traffic will increase 39 times over a span of five years, and 66% of the increase will be attributed to mobile video traffic [1]. Cellular service providers have already had a hard time to keep up with the staggering increase in data traffic [2], [3], and will have to carefully engineer their networks of supporting the tremendous amount of mobile video traffic in the future. Today, cellular networks are unable to handle large scale live video distributions since existing cellular deployments do not natively support multicast and broadcast. In fact, a measurement study reports that each UMTS HSDPA cell can only support 4 to 6 mobile video users at 256 kbps [4], which renders massive live video distribution, e.g., for soccer and other sports games, less commercially-viable.

Cellular service providers may address the capacity issue by: (i) deploying more base stations, (ii) upgrading their base stations, e.g., to support Multimedia Broadcast Multicast Services (MBMS) [5], or (iii) building dedicated broadcast networks, such as Digital Video Broadcast–Handheld (DVB-H) [6]. However, these solutions incur high infrastructure costs and may not be compatible with current mobile devices. Hence, a better solution is desirable. Since modern mobile devices are equipped with multiple network interfaces, cellular service providers may *offload* mobile video traffic to an *auxiliary* network. As illustrated in Fig. 1, we consider a *hybrid* cellular and ad hoc network consisting of a base station and multiple mobile devices. Mobile devices relay video data among each other using ad hoc links, exploiting such a *free* mechanism of distribution alleviates bottlenecks and reduces cost for cellular service providers. Note that distributing videos over an ad hoc network is more complex than distribution over a centralized network, such as a WiMAX network or a WiFi hotspot; we expect to be able to easily generalize our analysis to centralized auxiliary networks, which is one of our future focus areas.

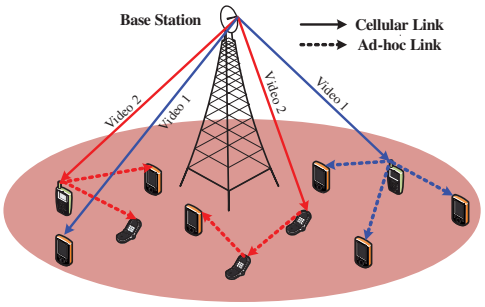


Figure 1: A hybrid cellular and ad hoc network.

While a hybrid cellular and ad hoc network has potential to capitalize on the complementary features of both networks for low-cost yet reliable massive live video distribution, transmission of video data must adhere to the timing needs inherent in the delivery and playback of video content. Traditionally, computationally complex *transcoders* [7] are used by video servers to reduce the video coding rates in order to guarantee ontime delivery of video data. However, in a hybrid network, real-time transcoding is not feasible on resource-constrained mobile devices, and thus we employ scalable videos [8] for in-network video adaptation [9]. More precisely, at the base station, scalable coders encode each video into a scalable stream consisting of multiple layers, and each mobile device can selectively forward some layers

978-1-4577-0351-5/11/$26.00

to its neighbors in a timely fashion.

In this paper, we study the problem to optimally determine: (i) the mobile devices that will inject video data from the cellular network into the ad hoc network, (ii) the multi-hop ad hoc routes for disseminating video data, and (iii) the subsets of video data each mobile device relays to the next hops under capacity constraints. We formulate the optimization problem into an Mixed Integer Linear Program (MILP), and propose an MILP-based algorithm to optimally solve it. Solving the MILP problem for hybrid networks with large numbers of mobile users is complex and time-consuming. To address the dual challenges of scalability and performance, we also propose an efficient heuristic algorithm. We conduct extensive simulations to evaluate the MILP-based and heuristic algorithms. The simulation results show that: (i) the hybrid network indeed improves the overall video quality, by up to 20 dB improvement, (ii) the proposed heuristic algorithm achieves near-optimal performance with at most 2.3 dB gap across all simulations, and (iii) the proposed heuristic is fast, and is more practical for hybrid networks with a large number of mobile devices.

The rest of this paper is organized as follows. We present related work in Sec. II and present the overall problem in Sec. III. We mathematically formulate the optimization problem in Sec. IV and propose solutions in Sec. V. The evaluation of the proposed approaches using trace-driven simulation results is presented in Sec. VI. We conclude in Sec. VII with future research directions.

II. Related Work

Using an auxiliary ad hoc network to increase the capacity of a cellular network has been considered in the literature.

- **Unicast Data Transfer.** Luo et al. [10] design a hybrid network that can route cellular data via other mobile devices with higher cellular data rates using a WiFi ad hoc network. Through simulations, Hsieh and Sivakumar [11] show that generic ad hoc protocols do not work well in hybrid cellular and WiFi ad hoc networks, and propose two approaches to improve the efficiency of ad hoc protocols. First, the base station can run optimization algorithms for the WiFi ad hoc network. Second, mobile devices connected to other access networks can offload traffic from the cellular network to those access networks.
- **Multicast Data Transfer.** Law et al. [12] evaluate a hybrid network in which some mobile devices act as gateways and relay data to mobile devices outside the range via a multi-hop ad hoc network. Lao and Cui [13] propose a hybrid network, in which each multicast group is either in the cellular mode or in the ad hoc mode. Park and Kasera [14] consider the gateway node discovery problem, and model ad hoc interference as a graph coloring problem. Bhatia et al. [15] formulate a problem of finding the relay forest to maximize the overall data rate, and they propose an approximation algorithm to solve it.

Unlike above works [10]–[15], we focus on delay sensitive live video distribution over a hybrid network–a problem that has not been throughly addressed. Qin and Zimmermann [16] present an adaptive strategy for live video distribution to determine the number of quality layers to be transmitted between two mobile devices. Hua et al. [17] formulate an optimization problem in a hybrid network to determine the cellular broadcast rate of each quality layer. In the ad hoc network, a flooding routing protocol is used to discover neighbors and a heuristic is employed to forward video data. Our work differs from Hua et al. [17] in several aspects: (i) we propose a unified optimization problem that jointly finds the optimal gateway mobile devices, ad hoc routes, and video adaptation, (ii) we consider existing cellular base stations without *MBMS* support, and (iii) we employ Variable-Bit-Rate (VBR) streams.

III. Live Video Distribution in Hybrid Networks

A. System Overview and Notations

We consider a hybrid network (see Fig. 1), which consists of a cellular base station and several mobile devices. The base station concurrently transmits K videos to U mobile devices, where each mobile device receives and renders a video chosen by its user. Throughout this paper, we use *node* to refer to both the base station and mobile devices. All mobile devices are equipped with two network interfaces for cellular and ad hoc networks, respectively. Mobile devices can always receive video data from the base station via cellular links. Unlike cellular networks, ad hoc connectivity is not guaranteed because a typical ad hoc network has a shorter range than cellular networks.

Distributing live videos in a hybrid network is challenging because: (i) wireless networks are dynamic in terms of connectivity, latency, and capacity, and (ii) video data requires high throughput and low latency. To cope with these challenges, we employ layered video coding [8], such as H.264/SVC [18], to encode each video into L layers. Layer 1 is referred to as the base layer, which provides a basic video quality. Layers $2,3,\ldots,L$ are enhancement layers, which provide incremental quality improvements. An enhancement layer is decodable if all layers below it are received. With layered videos, we can dynamically adjust the number of layers sent to each mobile device. While the adjustments may be done very frequently, a subject user study [19] reveals that frequent quality changes lead to degraded viewer experience. Therefore, we divide each video into multiple D-sec video segments, where D is a small integer. Quality changes are only allowed at boundaries of segments. We let S be the total number of segments of every video, and we let $t_{k,s,l}$ ($1 \le k \le K, 1 \le s \le S, 1 \le l \le L$) be the *transmission unit* of video k, segment s, and layer l.

We study an optimization problem in a recurring scheduling window of W segments. We refer to a solution as a schedule and we call an algorithm that runs at the base station to compute schedules as a scheduler. The scheduler on the base station takes feedback from networks, and computes a new schedule every DW' secs ($1 \leq W' \leq W$). The feedback includes transmission unit availability $y^u_{k,s,l}$ and mobile device location $\omega_u = (\omega_{u,x}, \omega_{u,y})$. We let $y^u_{k,s,l} = 1$ if mobile device u holds unit $t_{k,s,l}$, and $y^u_{k,s,l} = 0$ otherwise. We use $\omega_{u,x}$ and $\omega_{u,y}$ to denote the longitude and latitude of u, which can be derived from Global-Positioning-System (GPS) functionality, cellular network triangulations, and WiFi fingerprints.

Each mobile device u reports its $y^u_{k,s,l}$ and ω_u to the base station, and the base station maintains the state of availability and device location for all mobile devices $1 \leq u \leq U$. Given that the base station maintains a global view of the hybrid cellular and ad hoc network, the scheduler on the base station has a potential to find global optimum solutions.

The base station sends a new schedule to all mobile devices every DW' secs. The mobile devices then distribute transmission units following the schedule. To maintain the tractability, our schedule does not explicitly specify the transmission time of each transmission unit. Rather, we take a *precedence list* $\mathbf{P} = \{(p_{s,i}, p_{l,i}) \mid 1 \leq i \leq WL\}$ as an input, where $p_{s,i}$ and $p_{l,i}$ represent the relative segment number and layer number of precedence i transmission unit in each scheduling window. More specifically, let s_c be the first segment of the current scheduling window, mobile devices transmit scheduled transmission units in the following order: $t_{k,s_c+p_{s,1},p_{l,1}}, t_{k,s_c+p_{s,2},p_{l,2}}, \ldots, t_{k,s_c+p_{s,WL},p_{l,WL}}$, for any $1 \leq k \leq K$. Mobile devices skip transmission units that haven't been received, and check their availability again whenever a transmission unit is completely sent. For concrete discussion, we employs the following precedence list: $\{(0,1),(1,1),\ldots,(W-1,1),(0,2),(1,2),\ldots,(W-1,2),\ldots,(0,L),(1,L),\ldots,(W-1,L)\}$ if not otherwise specified.

B. Problem Statement and Hardness

Problem 1 (Scheduling in a Hybrid Network): Given K videos concurrently distributed from a cellular base station to a large number of mobile devices over a hybrid cellular and ad hoc network. Each video k is coded into multiple transmission units, while each unit $t_{k,s,l}$ represents layer l of segment s. Every DW' secs, we compute the schedule for a recurring window of W segments and for every network link, in order to maximize the overall video quality across all mobile devices. The resulting schedule should be feasible in the sense that the scheduled units can be delivered in DW' secs.

This scheduling problem is fairly general because: (i) any mobile device can relay any transmission unit to other mobile devices and (ii) each transmission unit can be disseminated over different multicast trees.

Lemma 1 (Hardness): The scheduling problem in a hybrid cellular and ad hoc network is NP-Hard.

The proof is omitted due to the space limitation. Since the scheduling problem in a hybrid network is NP-Hard, we formulate it as an MILP in the next section.

IV. System Models and Problem Formulation

A. Rate-Distortion Model

Our objective is to maximize the perceived video quality under network bandwidth constraints. A popular method to achieve such quality-optimized system is to use a rate-distortion (R-D) model, which describes the mapping between video rates and degrees of quality degradation in reconstructed videos. R-D models capture the diverse video characteristics and enable media-aware resource allocation. The distortion caused by not sending a transmission unit $t_{k,s,l}$ to a mobile device can be divided into two parts [20]: (i) truncation distortion and (ii) drifting distortion. Truncation distortion refers to the quality degradation of pictures in segment s itself, and drifting distortion refers to the quality degradation of pictures in other segments due to imperfect reconstruction of reference pictures. We assume each segment s contains a multiple of groups-of-picture (GoPs) and thus can be independently decoded. This practical assumption eliminates the needs to model drifting distortion.

We let $q_{k,s,l}$ to be the quality improvement when receiving $t_{k,s,l}$ in addition to the previously received $t_{k,s,l'}$, where $l' = 1,2,\ldots,l-1$. While quality improvement can be in any video quality metric, we use peak signal-to-noise ratio (PSNR) for concrete discussion. We let $z_{k,s,l}$ be the size of $t_{k,s,l}$. The sets $\mathbf{Q}_k = \{q_{k,s',l'} \mid 1 \leq s' \leq S, 1 \leq l' \leq L\}$ and $\mathbf{Z}_k = \{z_{k,s',l'} \mid 1 \leq s' \leq S, 1 \leq l' \leq L\}$ model the R-D characteristics of video stream k. $\mathbf{Q}_k$ and $\mathbf{Z}_k$ are computed during the encoding time, and sent to the base station as meta-data along with the video stream k itself.

B. Network Capacity

In wireless networks, interference can be described by a widely adopted model [21], where a transmission is successful if and only if the signal-to-interference-and-noise ratio (SINR) at the receiver is higher than a threshold. The *SINR* between nodes i and j is modeled by $SINR_{i,j} = \frac{e_i/(d_{i,j})^\alpha}{N+\sum_{k\in\mathbf{K}} e_k/(d_{k,j})^\alpha}$, where N is the noise, e_i denotes the transmission power of node i, $d_{i,j} = \sqrt{(\omega_{i,x}-\omega_{j,x})^2+(\omega_{i,y}-\omega_{j,y})^2}$ is the euclidean distance between i and j, $\mathbf{K}$ is the subset of nodes simultaneously transmitting at some time instants over the same channel, and α is the path loss exponent. The value of α depends on wireless environments, and is typically in the range of $(2,6]$. We consider an interference dominated environment [12] by letting $N = 0$ if not otherwise specified. We adopt Gaussian co-channel interference model as in [12], [21] with

the link throughput bounded by Shannon capacity. For the purpose of this work and for tractability of analysis we do not model wireless protocol specific implementations, adaptive techniques that might be deployed in the practical wireless network settings, and wireless channel fading and shadowing dynamics. Hence, we write the link capacity $c_{i,j}$ as $c_{i,j} = W \log_2(1 + SINR_{i,j})$, where W is the bandwidth in hertz.

In cellular networks, the base station runs a centralized algorithm to allocate δ_u air-time to mobile device u, where $1 \leq u \leq U$, and $\delta = \sum_{u=1}^{U} \delta_u$ is the total air-time reserved for mobile data, which is a system parameter. Let node 0 be the base station, the effective cellular capacity between it and node u is $\delta_u c_{0,u}$, where δ_u $(1 \leq u \leq U)$ are variables of our optimization problem.

In ad hoc networks, the air-time allocation is done by distributed media access control (MAC) protocols, which can be modeled by conflict graphs [15], [21], [22]. A conflict graph shows the sets of links that cannot be simultaneously activated and can be derived from the network graph. Let $G(\mathbf{V}, \mathbf{E})$ be the network graph, where $\mathbf{V}$ and $\mathbf{E}$ are nodes and edges. Its corresponding conflict graph $G(\bar{\mathbf{V}}, \bar{\mathbf{E}})$ is constructed as follows. We first create a vertex $\bar{v}_{i,j}$ in $\bar{\mathbf{V}}$ for each edge $\varepsilon_{i,j} \in \mathbf{E}$, and we add an edge connecting $\bar{v}_{i,j}$ and $\bar{v}_{k,l}$ to $\bar{\mathbf{E}}$ if nodes i or j are in nodes k or j's transmission range.

Each independent set selected from a conflict graph $G(\bar{\mathbf{V}}, \bar{\mathbf{E}})$ corresponds to a set of edges in the communication graph $G(\mathbf{V}, \mathbf{E})$ that can be simultaneously activated without interfering with each other. An independent set is called a maximal independent set if adding any vertex to it leads to a non-independent set. We let $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_Q$ be all maximal independent sets. Given Q maximal independent sets, distributed MAC protocols allocate λ_q air-time to maximal independent set $\mathbf{I}_q$, where $\sum_{q=1}^{Q} \lambda_q \leq 1$ [22], [23]. λ_q $(1 \leq q \leq Q)$ are variables of our optimization problem. For a link $\varepsilon_{i,j} \in \mathbf{E}$ between two mobile devices i and j, the effective ad hoc capacity is therefore $\sum_{1 \leq q \leq Q,\, \bar{v}_{i,j} \in \mathbf{I}_q} \lambda_q c_{i,j}$.

C. Controlling Distribution Latency

Our optimization problem only determines which transmission units to send in the current scheduling window, but does not model the fine-grained delivery time of each transmission unit. We should mention that the unit delivery time could be modeled using time-indexed Integer Linear Program (ILP) [24]. In time-indexed ILP formulations, all time intervals are expressed as (rounded to) multiples of a sufficiently small time slot. In these formulations, short time slots are essential for good performance, but short time slots also lead to a large number of decision variables and render the formulation computationally intractable.

We do not employ time-indexed ILP in our formulation, but use two other approaches to control latency. First, we limit each unit to be sent over at most H hops in each scheduling window, where H is a small integer and a system parameter. Second, we trim unnecessary indirect paths as follows. Let $\mathbf{A}_{k,s,l}$ be the set of nodes that already have unit $t_{k,s,l}$. Nodes in $\mathbf{A}_{k,s,l}$ are potential sources for distributing $t_{k,s,l}$ and all other mobile devices are receivers of that transmission unit. For a source $a \in \mathbf{A}_{k,s,l}$ and an arbitrary receiver u, there are many paths between them for distributing $t_{k,s,l}$. To avoid inefficient paths, we only consider the paths that follow the breadth-first tree from the source a to all mobile devices not in $\mathbf{A}_{k,s,l}$. We let $\mathbf{N}_{k,s,l}^{a,h}$ be the receiving mobile devices that are h hops away from $a \in \mathbf{A}_{k,s,l}$ in the breadth-first tree, where $1 \leq h \leq H$. Mathematically, we write:

$$\mathbf{N}_{k,s,l}^{a,h} = \bigcup_{u \in \mathbf{N}_{k,s,l}^{a,h-1}} \mathbf{N}_{k,s,l}^{u,1} \setminus \mathbf{N}_{k,s,l}^{a,h-1} \setminus \mathbf{N}_{k,s,l}^{a,h-2}, \tag{1}$$

where $\mathbf{N}_{k,s,l}^{a,1}$ are a's neighbors and $\mathbf{N}_{k,s,l}^{a,0} = \{a\}$.

Distributing transmission units over breadth-first trees not only limits the distribution latency and avoid loops, but also reduces the complexity of the considered problem without sacrificing its optimality. This is because paths that do not follow breadth-first trees are inefficient and should be avoided.

D. Formulation

We define $x_{k,s,l}^{a,v,u} \in \{0,1\}$ to be a decision variable: $x_{k,s,l}^{a,v,u} = 1$ if transmission unit $t_{k,s,l}$ is scheduled to be sent from node v to node u over the breadth-first tree rooted at node a; $x_{k,s,l}^{a,v,u} = 0$ otherwise. The scheduling problem in a hybrid cellular and ad hoc network is formulated in Eq. (2). In this formulation, we refer to the base station as node 0. The objective function in Eq. (2a) is the average video quality achieved by all U mobile devices. The objective function contains two terms (within the square brackets): the first term considers the breadth-first tree rooted at the base station, and the second term considers the breadth-first trees rooted at mobile devices that directly receive the transmission unit from the base station. We note that, if the objective function only has the first term, a new transmission unit would only have a breadth-first tree rooted at the base station with height one because all mobile devices are one hop away from the base station. Consequently, the transmission units would never be exchanged over ad hoc networks. Hence, the second term is critical to the utilization of the ad hoc network, and overall video quality.

The constraints in Eqs. (2b) and (2c) model the air-time allocation in the cellular network, and the constraints in Eqs. (2d) and (2e) model the air-time allocation in the ad hoc network. The constraints in Eq. (2f) guarantee the dependency among layers. The constraints in Eq. (2g) ensures a mobile device receives each transmission unit from a single sender over a single breadth-first tree. The constraints in Eq. (2h) makes sure that a mobile device sends

$$\max \quad \frac{1}{U}\sum_{k=1}^{K}\sum_{s=s_c}^{s_c+W}\sum_{l=1}^{L} q_{s,k,l}\Big[\sum_{a\in\mathbf{A}_{k,s,l}\setminus\{0\}}\sum_{h=1}^{H}\sum_{v\in\mathbf{N}_{k,s,l}^{a,h-1}}\sum_{u\in\mathbf{N}_{k,s,l}^{a,h}} x_{k,s,l}^{a,v,u} + \sum_{a'\in\mathbf{N}_{k,s,l}^{0,1}}\sum_{h=1}^{H-1}\sum_{v\in\mathbf{N}_{k,s,l}^{a',h-1}}\sum_{u\in\mathbf{N}_{k,s,l}^{a',h}} x_{k,s,l}^{a',v,u}\Big] \tag{2a}$$

$$\text{s.t.} \qquad \sum_{k=1}^{K}\sum_{s=s_c}^{s_c+W}\sum_{l=1}^{L}\frac{z_{k,s,l}x_{k,s,l}^{0,0,\hat{u}}}{c_{0,\hat{u}}DW'} - \delta_{\hat{u}} = 0; \tag{2b}$$

$$\sum_{u=1}^{U}\delta_u - \delta \le 0; \tag{2c}$$

$$\sum_{k=1}^{K}\sum_{s=s_c}^{s_c+W}\sum_{l=1}^{L}\sum_{a\in\mathbf{A}_{k,s,l}}\frac{z_{k,s,l}x_{k,s,l}^{a,\hat{u},\hat{v}}}{c_{\hat{u},\hat{v}}DW'} - \sum_{1\le q\le Q,\,\bar{v}_{\hat{u},\hat{v}}\in\mathbf{I}_q}\lambda_q = 0; \tag{2d}$$

$$\sum_{q=1}^{Q}\lambda_q - 1 \le 0; \tag{2e}$$

$$\sum_{a'\in\mathbf{A}_{\hat{k},\hat{s},\hat{l}+1}}\sum_{h=1}^{H}\sum_{t\in\mathbf{N}_{\hat{k},\hat{s},\hat{l}+1}^{a',h-1}} x_{\hat{k},\hat{s},\hat{l}+1}^{a',t,\hat{u}} - y_{\hat{k},\hat{s},\hat{l}}^{\hat{u}} - \sum_{a\in\mathbf{A}_{\hat{k},\hat{s},\hat{l}}}\sum_{h=1}^{H}\sum_{t\in\mathbf{N}_{\hat{k},\hat{s},\hat{l}}^{a,h-1}} x_{\hat{k},\hat{s},\hat{l}}^{a,t,\hat{u}} \le 0; \tag{2f}$$

$$\sum_{a\in\mathbf{A}_{\hat{k},\hat{s},\hat{l}}}\sum_{h=1}^{H}\sum_{t\in\mathbf{N}_{\hat{k},\hat{s},\hat{l}}^{a,h-1}} x_{\hat{k},\hat{s},\hat{l}}^{a,t,\hat{u}} - 1 \le 0; \tag{2g}$$

$$\sum_{a'\in\mathbf{A}_{\hat{k},\hat{s},\hat{l}}}\sum_{t'\in\mathbf{N}_{\hat{k},\hat{s},\hat{l}}^{a',\hat{h}}} x_{\hat{k},\hat{s},\hat{l}}^{a',t',\hat{u}} - \sum_{a\in\mathbf{A}_{\hat{k},\hat{s},\hat{l}}}\sum_{t\in\mathbf{N}_{\hat{k},\hat{s},\hat{l}}^{a,\hat{h}-1}} x_{\hat{k},\hat{s},\hat{l}}^{a,t,\hat{u}} \le 0; \tag{2h}$$

$$\forall\ 1\le\hat{u},\hat{v}\le U,\ 1\le\hat{k}\le K,\ 1\le\hat{s}\le S,\ 1\le\hat{l}\le L,\ 1\le\hat{h}\le H.$$

a transmission unit only if it receives that unit in current or earlier scheduling windows.

V. Solutions

A. An MILP-based Solution: POPT

The formulation in Eq. (2) is an MILP problem and may be solved by MILP solvers. However, Eqs. (2d) and (2e) include all the maximal independent sets $\mathbf{I}_q$ ($1 \le q \le Q$) in the conflict graph, and finding all $\mathbf{I}_q$ itself is an NP-Complete problem [25]. Therefore, it is computationally impractical to consider all Q maximal independent sets. Jain et al. [26] propose a random search algorithm for deriving a subset of maximal independent sets that is sufficient for optimal schedulers. Li et al. [23] show that this random search algorithm is inefficient, and propose a priority-based algorithm to find the maximal independent sets that will be used in the optimal schedule with high probability. While the priority-based algorithm is defined for the throughput optimization problem in a multi-radio, multi-channel wireless network, it can be extended to other conflict graph based optimization problems by revising the definition of the scheduling priority. Readers are referred to Li et al. [23] for more details on this algorithm.

We define a new priority function for each ad hoc link to achieve four design heuristics:

1) The links into mobile devices with more descendants in breadth-first trees are given higher priorities.
2) The links into mobile devices on breadth-first trees of transmission units with higher quality improvement values are given higher priorities.
3) The links with higher ad hoc link capacities are given higher priorities.
4) The links from mobile devices with higher cellular link capacities are given higher priorities.

Specifically, we define the priority function $f(v,u)$ of an edge from v to u as:

$$f(v,u) = f_a(u) + f_c(v), \tag{3}$$

where $f_a(u)$ and $f_v(v)$ are the "importance" factors due to the ad hoc network and the cellular network, respectively. They are computed as:

$$f_a(u) = c_{v,u}\sum_{k=1}^{K}\sum_{s=s_c}^{s_c+W}\sum_{l=1}^{L}\sum_{a\in\mathbf{A}_{k,s,l}\setminus\{0\}} m_{k,s,l}^{a,u}q_{k,s,l}, \tag{4}$$

$$f_c(v) = \delta c_{0,v}\sum_{k=1}^{K}\sum_{s=s_c}^{s_c+W}\sum_{l=1}^{L} m_{k,s,l}^{v,v}q_{k,s,l}, \tag{5}$$

where $m_{k,s,l}^{a,u}$ is the number of descendants of mobile device u on the breadth-first tree rooted at node a for video k, segment s, and layer l.

With the priority function $f(v,u)$, we leverage the priority-based algorithm [23] to generate a small set of $\hat{Q}$ ($1 \le \hat{Q} \le Q$) maximal independent sets that will be employed by the optimal schedules with high probability. We then apply a practical simplification on the formulation in Eq. (2) by only considering the $\hat{Q}$ maximal independent sets in the constraints in Eqs. (2d) and (2e). Unlike the original formulation that may consist of exponentially many maximal independent sets, the simplified formulation has a reasonable number of maximal independent sets, and can be solved by MILP solvers. We use an MILP solver to solve the simplified formulation, and we refer to it as Prioritized Optimization (POPT) algorithm.

1. **for** i = 1 to WL // step 1
2. **let** $s = s_c + p_{s,i}$; $l = p_{l,i}$
3. **for** k = 1 to K
4. **let** $\mathbf{R} = \{u \mid 1 \le u \le U\}$ // all mobile devices
5. more1:
6. **foreach** root a in $\mathbf{A}_{k,s,l}$
7. **foreach** edge between u and v following the
7. breadth-first tree rooted at a, where $v \in \mathbf{R}$
8. **let** $\tilde{c}_{u,v} = \tilde{c}_{u,v} + z_{k,s,l}/(DW')$; $\mathbf{R} = \mathbf{R} \setminus \{v\}$
9. **if** $\mathbf{R} \neq \emptyset$
10. **let** $\hat{u}$ be the device in $\mathbf{R}$ with the highest $c_{0,\hat{u}}$
11. **let** $\mathbf{R} = \mathbf{R} \setminus \{\hat{u}\}$
12. **goto** more1 with $\hat{u}$ as the next root
13. **if** $\tilde{c}_{u,v}$ for all $1 \le u,v \le U$ is infeasible **break**
14. **let** $\hat{c}_{u,v} = \tilde{c}_{u,v}$ for all $1 \le u,v, \le U$ // best demands
15. **solve** LP in Eq. (6) with $\hat{c}_{u,v}$ for $c^*_{u,v}$ // opt. b/w
16. **let** $\bar{c}_{u,v} = c^*_{u,v}$ for $1 \le u,v \le U$ // step 2, remaining b/w
17. **for** i = 1 to WL // precedence list
18. **let** $s = s_c + p_{s,i}$; $l = p_{l,i}$
19. **sort** K videos on $q_{k,s,l}/z_{k,s,l}$
20. **foreach** video k in the descending order
21. **let** $\mathbf{R} = \{u \mid 1 \le u \le U\}$ // all mobile devices
22. **sort** $\mathbf{A}_{k,s,l} \setminus \{0\}$ on $m^{a,a}_{k,s,l}$ // on tree size
23. more2:
24. **foreach** root a in the descending order
25. **foreach** edge between u and v following the
25. breadth-first tree rooted at a, where $v \in \mathbf{R}$
26. **if** $\bar{c}_{u,v} \ge z_{k,s,l}/(DW')$
27. **let** $x^{a,u,v}_{k,s,l} = 1$; $\bar{c}_{u,v} = \bar{c}_{u,v} - z_{k,s,l}/(DW')$;
27. $\mathbf{R} = \mathbf{R} \setminus \{v\}$
28. **else**
29. **let** $x^{a,u,v}_{k,s,l} = 0$; **truncate** the tree at v
30. **if** $\mathbf{R} \neq \emptyset$
31. **sort** $u \in \mathbf{R}$ on $c_{0,u}$ // cellular capacity
32. **foreach** $\hat{u}$ in the descending order
33. **if** $z_{k,s,l}/c_{0,\hat{u}} \le \delta - \sum_{u'=1}^{U} \delta_{u'}$
34. **let** $x^{a,0,\hat{u}}_{k,s,l} = 1$; $\delta_{\hat{u}} = \delta_{\hat{u}} + \frac{z_{k,s,l}}{c_{0,\hat{u}}}$; $\mathbf{R} = \mathbf{R} \setminus \{\hat{u}\}$
35. **goto** more2

Figure 2: MTS: the proposed efficient scheduling algorithm.

B. An Efficient Algorithm: MTS

Since MLIP problems are NP-Complete, the POPT algorithm may not scale well with the number of mobile devices. Through extensive simulations (see Sec. VI), we find that the POPT algorithm runs efficiently for hybrid networks with up to 20 mobile devices. For hybrid networks with more mobile devices, we present an efficient heuristic algorithm in the following. The algorithm first probes the maximum feasible ad hoc network capacity based on transmission unit availability. It then greedily schedules transmission units until the ad hoc and cellular network capacities are both saturated. We refer to this algorithm as Maximum Throughput Scheduling (MTS) algorithm.

Fig. 2 gives the pseudo-code of the proposed MTS algorithm. This algorithm consists of two steps. In step 1, we derive the *demand capacity* $\hat{c}_{u,v}$ for each link from mobile device u to v. We iterate through the transmission units following the precedence list, which generally starts from lower to higher layers and from earlier to later segments. For each transmission unit, we schedule it to be delivered to all mobile devices that do not have that unit yet, and we employ the cellular network to help the ad hoc network if needed. We accumulate the required ad hoc capacity $\tilde{c}_{u,v}$ for each ad hoc link by calculating the ratio of total unit size sent over it and the rescheduling frequency DW'. After considering each transmission unit, we check whether the current ad hoc capacity $\tilde{c}_{u,v}$ is feasible using the conflict graph $G(\bar{\mathbf{V}}, \bar{\mathbf{E}})$, and we let demand capacity $\hat{c}_{u,v} = \tilde{c}_{u,v}$ be the last feasible ad hoc link capacity.

Upon getting demand capacity $\hat{c}_{u,v}$, we compute the maximum ad hoc network capacity by solving a Linear Program (LP):

$$\max \quad \sum_{1 \le q \le \hat{Q}} \sum_{\bar{v}_{u,v} \in \mathbf{I}_q} c_{u,v} \lambda q \tag{6a}$$

$$\text{s.t.} \quad \sum_{q=1}^{\hat{Q}} \lambda_q \le 1; \tag{6b}$$

$$c_{u,v} \sum_{1 \le q \le \hat{Q}, \bar{v}_{u,v} \in \mathbf{I}_q} \lambda_q \ge \hat{c}_{u,v}. \tag{6c}$$

This formulation maximizes the total ad hoc capacity in Eq. (6a), while guaranteeing the demand capacity is met in Eq. (6c). Eq. (6) can be efficiently solved by LP solvers. Let λ^*_q ($1 \le q \le \hat{Q}$) be the optimum air-time allocation, we compute the optimum effective ad hoc link capacity between mobile device u and v as:

$$c^*_{u,v} = c_{u,v} \sum_{1 \le q \le \hat{Q}, \bar{v}_{u,v} \in \mathbf{I}_q} \lambda^*_q. \tag{7}$$

In step 2, we traverse through the precedence list, and we go through the transmission units of different videos on the descending order of the ratio of quality improvement and unit size. We consider the transmission units with higher ratios earlier, because sending them leads to more efficient schedule in the R-D fashion. Next, for each transmission unit, we sort the mobile devices that already hold that transmission unit on the numbers of descendants on their breadth-first trees. We iterate through these mobile devices, and schedule the transmission unit as long as the remaining link capacity permits. We stop once the current transmission unit is distributed to all mobile devices. If the transmission unit can not be received by some mobile devices, we use the cellular links to help. The algorithm stops upon both maximum ad hoc link capacity and cellular data air-time allocation δ are saturated.

The next lemma (proof is omitted due to space limitations) shows that the MTS algorithm runs in polynomial time.

Lemma 2 (Complexity): The MTS algorithm given in Fig. 2 runs in polynomial time in the worst-case, if the LP formulation in Eq. (6) is solved by a polynomial time LP solver. For example, with Karmarkar's interior point method [27], the MTS algorithm has a time complexity of $O\left[\hat{Q}^{5.5}E^2 + WLK(\log K+1)U(\log U+1)\right]$, where $E = |\mathbf{E}|$ is the number of edges in the network graph. Hence, the MIT algorithm scales to hybrid networks with a large number of mobile devices.

VI. Evaluation

A. Setup

We have implemented a trace-driven simulator using a combination of C, MATLAB, and CPLEX [28]. Specifically, we use: (i) C to implement the simulator framework, (ii) MATLAB to prepare the objective functions and constraints of the MILP and LP formulations, and (iii) CPLEX to solve the MILP and LP problems. We have implemented the proposed POPT and MTS algorithms in the simulator. We have also implemented a cellular only optimal scheduler for comparison, which is referred to as Current in the figures. We emphasize that Current is not a naive algorithm; rather, it achieves optimal cellular air-time allocation, but it does not leverage on the auxiliary ad hoc network. We cannot compare our work against other data transfer solutions [10]–[15], which are not media aware, nor [17], which assumes base stations support multicast/broadcast.

The simulator takes two type of traces as inputs: (i) mobility traces and (ii) video traces. The mobility traces are synthetically generated using the random waypoint model with a maximum speed of 10 m/sec, a minimum speed of 1 m/sec, and a pause time of 30 sec. We adopt the video traces of H.264/SVC layered videos from an online video library [29]. In this paper, we report sample simulation results of distributing *Crew* video. However, the proposed formulation and solutions are general and also work for the scenarios where mobile devices watch different videos.

We consider $D = 2$ sec segments, and we vary the scheduling window $W = 2,4,6$. We let $W' = W/2$ if not otherwise specified. We consider the number of mobile devices $U = 5,10,15,20,30,40,60,80$ watching live videos. The nodes are randomly distributed over a 600×600 m^2 cell. The cellular network data rates are $R_c = 384,600,1024,2048,4096$ kbps, and the ad hoc data rates are $R_a = 2,11$ Mbps. Mobile data air-time δ in the cellular network is 0.75, and ad hoc transmission range is 100 m. We assume an initial buffering time $T_0 = 3$ secs when determining whether a transmission unit misses its playout deadline. For the POPT and MTS algorithms, we let maximum hop count $H = 1,2,3,4,5$. Furthermore, we use the default settings of CPLEX [28] when solving MILP problems. All the experiments are run on a Linux workstation with an Intel 3.2 GHz CPU.

We conduct experiments with different parameters, including: (i) number of mobile devices U, (ii) maximum hop count H, and (iii) scheduling window size W. In each experiment, we vary one parameter, and fix all other parameters. Each simulation lasts for $10DW'$ secs.

Metrics investigated in our experiments include: (i) perceived video quality in PSNR (dB), (ii) delay in sec, (iii) delivery ratio, which is the ratio of the ontime units over all scheduled units, and (iv) missed segment ratio, which is the fraction of undecodable segments due to misses of base layers.

B. Simulation Results

Performance Improvement. We investigate the performance improvement achieved by the hybrid network with different number of mobile devices U. For the POPT algorithm, we stop at $U = 30$ because it takes prohibitively long time when solving a problem with more mobile devices. We set $R_c = 600$ kbps, $R_a = 11$ Mbps, and $H = 3$. Fig. 3(a) plots the video quality. This figure shows that, for a PSNR requirement of 20 dB, the Current scheduler can only support 10 mobile devices, while the POPT and MTS algorithms can support 80+ mobile devices. Fig. 3(b) plots the same quality results with 95% confidence intervals. This figure reveals that the POPT and MTS algorithms achieve better fairness among mobile devices, while the Current scheduler leads to up to 10 dB quality difference. We plot the missed segment ratio in Fig. 3(c), which illustrates that, for the Current scheduler, the cellular network capacity is quickly saturated with increasingly more mobile devices. For example, with only 10 mobile devices, the Current scheduler fails to deliver about 17% base layer units, which will lead to unacceptable user experience.

Next, we plot the delivery ratio of different schedulers in Fig. 4(a), and mean delivery delay in Fig. 4(b). We observe that the delivery ratio is always higher than 96% in Fig. 4(a), and the delivery delay is fairly stable with increasingly more mobile devices in Fig. 4(b). These two figures show that although our formulation doesn't explicitly model delivery delay, nearly all scheduled units are delivered in time.

Time Complexity. We plot the running time of the POPT and MTS algorithms in Fig. 4(c). This figure shows that the POPT algorithm runs fast with 20 or fewer mobile devices. However, its running time dramatically increases with more than 20 mobile devices. While the MTS algorithm is more efficient, it achieves near-optimal video quality: at most 2.3 dB gap is reported in Fig. 3(a). We only consider the MTS algorithm in the rest of this section.

Maximum Hop Count. We study the implication of maximum hop count H on the MTS algorithm. We let $R_c =$ 2048 kbps and $R_a = 2$ Mbps, $U = 30,40$, and vary H. We plot the video quality in Fig. 5(a), which shows that the best H value is 4 for $U = 30$, and is 3 for $U = 40$. Fig. 5(b) gives the running time of different H values, which illustrates that the MTS algorithm scale well with H in terms of running time.

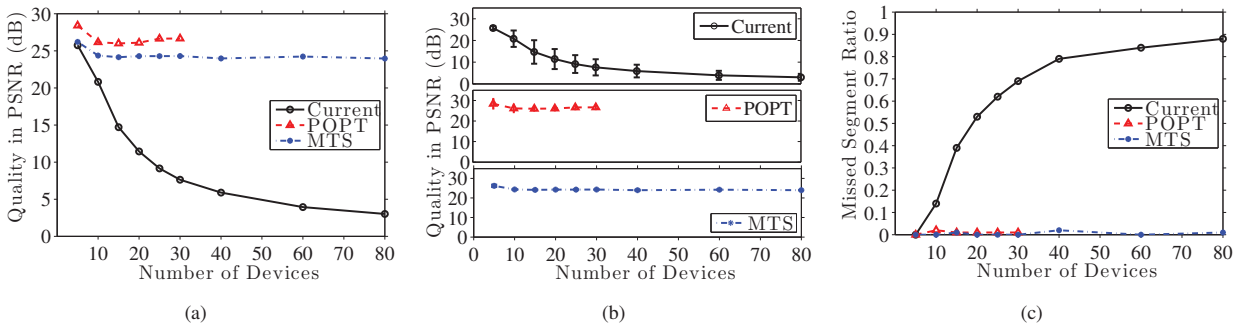


Figure 3: Performance improvement: (a) video quality, (b) video quality with 95% confidence interval, and (c) missed segment ratio.

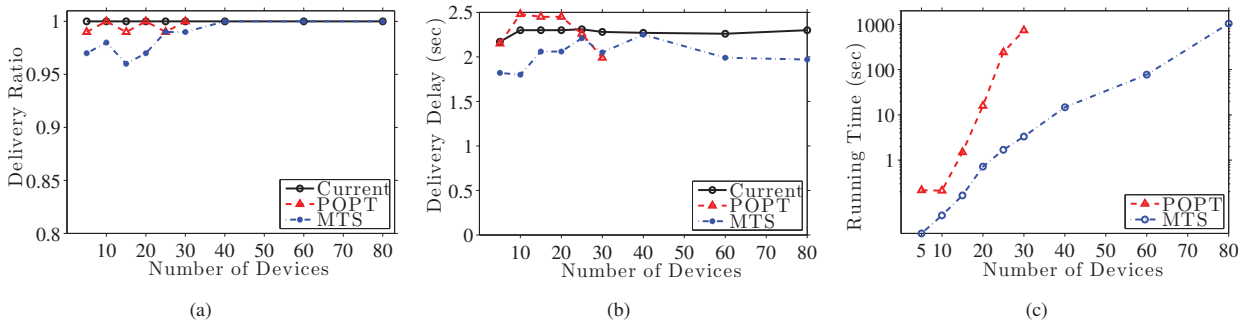


Figure 4: Performance comparison: (a) delivery ratio, (b) delivery delay, and (c) running time.

Scheduling Window Size. Selecting a good scheduling windows size W is challenging: a too small W may constrain the optimization algorithm to redistribute network resources among segments, while a too large W may lead to low delivery ratio because of device mobility. We let $U = 30$, $R_c = 600$ kbps, $R_a = 11$ Mbps, and $W = 2, 4, 6$. Table I summarizes the results. It shows that $W = 2$ is too small and results in low video quality, while larger W leads to better video quality and slightly longer running time.

VII. Conclusions and Future Work

We studied the problem of optimally leveraging an auxiliary ad hoc network to boost the overall video quality of mobile users in a cellular network. We formulated this problem as an MILP problem to jointly solve the gateway selection, ad hoc routing, and video adaptation problems for a global optimum schedule. We proposed two algorithms: (i) an MILP-based algorithm, POPT and (ii) a heuristic algorithm, MTS. Extensive simulation results indicate that the POPT and MTS algorithms significantly outperform the current mechanisms that do not leverage an auxiliary ad hoc network. For example, the POPT and MTS algorithms achieve up to 20 dB higher video quality than the Current scheduler.

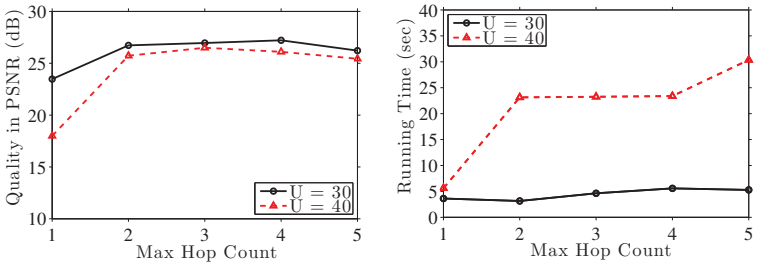
Our work can be extended to apply to multiple domains. From a broader viewpoint, massive delivery of rich information is useful in a range of mission critical scenarios such as military command-and-control and emergency response applications where existing infrastructure may be damaged, inaccessible, or overloaded. For example, customized notifications in emergency alerting situations will make it possible for users to receive rich alerts (such as evacuation maps and traffic routes) based on their current context for a more effective response. The ability to combine multiple infrastructure and ad hoc connectivities to achieve faster and improved information exchange on a societal scale, as demonstrated in this paper, is key to enabling richer mobile applications for the future.

References

[1] "Cisco visual networking index: Forecast and methodology, 20092014," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, 2010.

[2] "AT&T faces 5,000 percent surge in traffic," http://www.internetnews.com/mobility/article.php/3843001, 2009.

[3] "T-Mobile's growth focusing on 3G," http://connectedplanetonline.com/wireless/news/t-mobile-3g-growth-0130, 2009.



(a) (b)

Figure 5: Implication of maximum hop count: (a) quality, (b) running time.

Table I: Implication of scheduling window size W.

W	Quality (dB)	Delay (sec)	Run Time (sec)
2	15.23	0.68	3.19
4	24.17	1.31	3.28
6	26.66	1.99	3.32

[4] F. Hartung, U. Horn, J. Huschke, M. Kampmann, T. Lohmar, and M. Lundevall, "Delivery of broadcast services in 3G networks," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 188–199, March 2007.

[5] S. Parkvall, E. Englund, M. Lundevall, and J. Torsner, "Evolving 3G mobile systems: Broadband and broadcast services in WCDMA," *IEEE Communications Magazine*, vol. 44, no. 2, pp. 30–36, February 2006.

[6] M. Kornfeld and G. May, "DVB-H and IP Datacast – broadcast to handheld devices," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 161–170, March 2007.

[7] J. Xin, C. Lin, and M. Sun, "Digital video transcoding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 84–97, January 2005.

[8] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communications*, 1st ed. Prentice Hall, 2001.

[9] I. Kofler, M. Prangl, R. Kuschnig, and H. Hellwagner, "An H.264/SVC-based adaptation proxy on a WiFi router," in *Proc. of ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'08)*, Braunschweig, Germany, May 2008, pp. 63–68.

[10] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "UCAN: a unified cellular and ad-hoc network architecture," in *Proc. of ACM International Conference on Mobile Computing and Networking (MobiCom'03)*, San Diego, CA, September 2003, pp. 353–367.

[11] H. Hsieh and R. Sivalumar, "On using peer-to-peer communication in cellular wireless data networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 57–72, March 2004.

[12] L. Law, K. Pelechrinis, S. Krishnamurthy, and M. Faloutsos, "Downlink capacity of hybrid cellular ad hoc networks," *IEEE Transactions on Networking*, vol. 18, no. 1, pp. 243–256, February 2010.

[13] L. Lao and J. Cui, "Reducing multicast traffic load for cellular networks using ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 317–329, May 2006.

[14] J. Park and S. Kasera, "Enhancing cellular multicast performance using ad hoc networks," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'05)*, New Orleans, LA, March 2005, pp. 2175–2181.

[15] R. Bhatia, L. Li, H. Luo, and R. Ramjee, "ICAM: Integrated cellular and ad hoc multicast," *IEEE Transactions on Mobile Computing*, vol. 5, no. 8, pp. 1004–1015, August 2006.

[16] M. Qin and R. Zimmermann, "An adaptive strategy for mobile ad hoc media streaming," *IEEE Transactions on Multimedia*, vol. 12, no. 4, pp. 317–329, June 2010.

[17] S. Hua, Y. Guo, Y. Liu, H. Liu, and S. Panwar, "SV-BCMCS: Scalable video multicast in hybrid 3G/ad-hoc networks," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM'09)*, Honolulu, HI, November 2009, pp. 4662–4667.

[18] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, September 2007.

[19] P. Ni, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Fine-grained scalable streaming from coarse-grained videos," in *Proc. of International workshop on Network and Operating Systems support for Digital Audio and Video*, Williamsburg, VA, September 2009, pp. 103–108.

[20] C. Hsu, N. Freris, J. Singh, and X. Zhu, "Rate control and stream adaptation for scalable video streaming over multiple access networks," in *Proc. of International Packet Video Workshop (PV'10)*, Hong Kong, China, December 2010.

[21] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[22] Y. Wang, W. Wang, X. Li, and W. Song, "Interference-aware joint routing and TDMA link scheduling for static wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 12, pp. 1709–1726, 2008.

[23] H. Li, Y. Cheng, C. Zhou, and P. Wan, "Multi-dimensional conflict graph based computing for optimal capacity in MR-MC wireless networks," in *Proc. of IEEE International Conference on Distributed Computing Systems (ICDCS'10)*, Boston, MA, August 2010, pp. 774–783.

[24] M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 3rd ed. Springer, 2008.

[25] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, 1st ed. W. H. Freeman and Company, 1979.

[26] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. of ACM International Conference on Mobile Computing and Networking (MobiCom'03)*, San Diego, CA, September 2003, pp. 66–80.

[27] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, vol. 4, no. 4, pp. 373–395, December 1984.

[28] "IBM ILOG CPLEX," http://www.ibm.com/software/integration/optimization/cplex-optimizer, 2010.

[29] "Video trace files and statistics: Video traces 2," http://trace.eas.asu.edu/videotraces2/cgs, 2009.