On Joint Routing and Server Selection for MD Video Streaming in Ad Hoc Networks

Shiwen Mao, *Member, IEEE,* Xiaolin Cheng, *Student Member, IEEE,*
Y. Thomas Hou, *Senior Member, IEEE,* Hanif D. Sherali, and Jeffrey H. Reed, *Fellow, IEEE*

***Abstract*— For media streaming in ad hoc networks, service replication has been demonstrated to be a quite effective countermeasure to streaming interruptions caused by fragile paths and dynamic topology. In this paper, we study the problem of joint routing and server selection for double description (DD) video streaming in ad hoc networks. We formulate the task as a combinatorial optimization problem and present tight lower and upper bounds for the achievable distortion. The upper bound provides a feasible solution to the formulated problem. Our extensive numerical results show that the bounds are very close to each other for all the cases studied, indicating the near-global optimality of the derived upper bounding solution. Moreover, we observe significant gains in video quality achieved by the proposed approach over existing server selection schemes. This justifies the importance of jointly considering routing and server selection for optimal MD video streaming.**

***Index Terms*— Ad hoc network, multiple description coding, routing, video streaming, server diversity.**

I. INTRODUCTION

AD hoc networks are infrastructureless wireless networks with mobile users. These two characteristics make ad hoc networks an excellent match for important military and civilian applications, all of which demand great simplicity and flexibility in deployment and operations. However, the other direct consequences of infrastructure-independence and user mobility are multi-hop, fragile wireless routes and dynamic network topology, posing great challenges for provisioning of content-rich multimedia streaming services in ad hoc networks.

With respect to such streaming services, a basic requirement is continuous delivery of media data, which translates to a continuous connectivity requirement between the media server and client. Furthermore, it would be highly desirable to have a graceful degradation of received media quality as network environment changes over time, i.e., a received quality commensurate with available network resources. Clearly, the traditional approach of accessing a single server through a single path could be hardly adequate, since the server could crash or be unavailable due to high workload or network partition, and the single path could be broken or congested. To address these issues, an effective solution is *service replication* [1], which has been widely used in the Internet to make service closer to clients and for balancing the workload among servers [2]–[4]. We believe an effective solution should jointly consider QoS provisioning mechanisms (e.g., server selection and QoS routing) as well as video coding, error control and concealment mechanisms for optimal streaming service in ad hoc networks.

Recent advances in Multiple Description (MD) coding has made it highly suitable for multimedia communications in wireless ad hoc networks [5]–[11], especially for distributed media deliveries. MD coding is a technique that generates multiple equally important descriptions, each giving a low, but acceptable video quality [5], [10]. The decoding independence among the descriptions permits a reconstruction of video from *any* subset of received descriptions, achieving a quality commensurate with the number of received descriptions. This feature makes MD video an excellent match for multimedia applications in ad hoc networks, where wireless links are unstable and reliable paths are hard to maintain.

MD for distributed storage has been suggested in [5], where a typical user would have fast access to the local video descriptions. For higher quality, one or more remote descriptions could be retrieved and combined with the local ones. An interesting and thorough study of MD streaming for content delivery networks (CDN) is presented in [6]. Particularly, three server selection algorithms, i.e., Shortest Path (SP), Heuristic, and Distortion are proposed for a client to select a pair of servers having complementary descriptions for improved video quality. Although these algorithms have been shown to be effective in content delivery networks, the first two simple algorithms only consider hop-counts of the paths when choosing servers. Such a network-centric approach may not necessarily guarantee good application layer performance, such as video quality [8], [9]. The third server selection algorithm, Distortion, selects servers based on the expected video distortion, but without considering the more difficult optimal routing problem. In wireless ad hoc networks, links are much more diverse in terms of quality (e.g., available

Manuscript received April 4, 2005; revised September 4, 2005; accepted April 15, 2006. The associate editor coordinating the review of this paper and approving it for publication was Dr. J. Hou. This research has been supported in part by the Office of Naval Research (ONR) under Grants N00014-03-1-0521 and N00014-05-1-0179, and by the National Science Foundation (NSF) under Grants CNS-0347390 and DMI-0552676. This work was presented in part at the IEEE International Conference on Communications, Seoul, Korea, May 2005.

S. Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 (email: smao@ieee.org).

Y. T. Hou and J. H. Reed are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 (email: {thou, reedjh}@vt.edu).

X. Cheng is with the Department of Computer Science, Univ. of California, Davis, CA 95616 (email: xicheng@vt.edu).

H. D. Sherali is with the Grado Department of Industrial and System Engineering, Virginia Tech, Blacksburg, VA 24061 (email: hanifs@vt.edu).

Digital Object Identifier 10.1109/TWC.2007.05236.

1536-1276/07$20.00 © 2007 IEEE

bandwidth and loss) than links in wireline networks. In such an environment, pure server selection-based algorithms, although effective in the Internet, may produce low video quality if the default routes happen to include links being congested or having high loss rates.

In this paper, we study the problem of joint routing and server selection for double description (DD) video streaming in wireless ad hoc networks. In addition to selecting a pair of servers, we also explore optimal routing strategies to find high quality paths to the servers. Such a joint routing and server selection scheme opens a new dimension of freedom for further improving the DD video quality, since it explores a much larger solution space than existing server selection schemes.

Specifically, we first formulate the joint routing and server selection task as a combinatorial optimization problem that minimizes the received video distortion. This approach is application-centric and cross-layer in nature since we optimize the application layer performance (video quality) via network layer operations (routing). Due to the highly complex nature of the formulated problem, exact solutions are hard to find. Rather, we present schemes to compute a lower bound and an upper bound on the best achievable video distortion based on the monotonicity properties of the objective function. The upper bound produces a near-optimal pair of servers and a pair of corresponding paths for the client. The proposed approach is computationally efficient and can be easily incorporated into existing routing protocols for ad hoc networks. Our extensive numerical results show that the upper and lower bounds are very close to each other for all the cases studied, indicating that they are very close to the global optimum. We also observe significant gains in video quality achieved by the proposed approach over existing server selection schemes, which justify the importance of jointly considering routing and server selection for MD video streaming in wireless ad hoc networks.

The remainder of this paper is organized as follows. In Section II, we present the problem formulation, and then present algorithms for computing a lower bound and an upper bound for the achievable optimal distortion in Section III. Our experimental studies are presented in Section IV. We discuss practical issues in Section V and related work in Section VI. Section VII concludes the paper.

TABLE I
NOTATION

Symbol	Definition
$\mathcal{G}\{V, E\}$:	graph representation of the network.
V :	set of vertices.
E :	set of edges.
$\mathcal{S}_h$:	server set that hosts Description h, $h = 1, 2$.
s_h :	a video server, $s_h \in \mathcal{S}_h$, $h = 1, 2$.
u :	a client node.
$\{i, j\}$:	a directed link from node i to node j.
b_{ij} :	available bandwidth of link $\{i, j\}$.
p_{ij} :	success probability of link $\{i, j\}$.
l_{ij} :	average loss burst length of link $\{i, j\}$.
α_{ij} :	"up" to "down" transition prob. of link $\{i, j\}$.
β_{ij} :	"down" to "up" transition prob. of link $\{i, j\}$.
x^h_{ij} :	routing index variables, defined in (3).
$\mathcal{P}_h$:	a path from s_h to u, $h = 1, 2$.
$\mathcal{J}(\mathcal{P}_1, \mathcal{P}_2)$:	joint portion of $\mathcal{P}_1$ and $\mathcal{P}_2$.
$\bar{\mathcal{J}}(\mathcal{P}_h)$:	the disjoint portion on $\mathcal{P}_h$, $h = 1, 2$.
p_{jnt} :	average success prob. of joint links.
p^h_{dj} :	average success prob. of disjoint links on $\mathcal{P}_h$.
T_{on} :	average "up" period of the joint links.
α :	transition probability from "up" to "down."
β :	transition probability from "down" to "up."
R_h :	rate of Description h in bits per pixel, $h = 1, 2$.
R :	rate for balanced descriptions, i.e., $R_1 = R_2 = R$.
a :	$a = 2^{-2R_1}$.
b :	$b = 2^{-2R_2}$.
ρ :	a constant for converting bandwidth to video rate.
γ :	a constant determined by chroma subsampling scheme.
W :	width of a video frame (in number of pixels).
H :	height of a video frame (in number of pixels).
f :	frame rate of a description.
d_0 :	distortion when both descriptions are received.
d_h :	distortion when only Description h is received, $h = 1, 2$.
σ^2 :	variance of the source.
P_{00} :	probability that both descriptions are received.
P_{01} :	probability of receiving Description 1 only.
P_{10} :	probability of receiving Description 2 only.
P_{11} :	probability that both descriptions are lost.
D :	average video distortion.
x^* :	the optimal solution.
x^*_u :	constructed upper bounding solution.
x^*_l :	constructed lower bounding solution.

II. Problem Description

In this section, we formulate the problem of joint routing and server selection for MD video streaming in wireless ad hoc networks. The notation is summarized in Table I.

A. Double Description Video Performance Measure

We focus on double-description (DD) video to illustrate the problem formulation, since it is most widely used in MD video streaming in practice [6]–[9], [12], [13]. For two descriptions, each generated for a sequence of video frames, let d_h denote the achieved distortion when only description h is received, $h = 1, 2$, and d_0 the distortion when both descriptions are received. Also, let P_{00} represent the probability of receiving both descriptions, P_{01} the probability of receiving Description 1 only, P_{10} the probability of receiving Description 2 only, and P_{11} the probability of losing both descriptions. Then, the average distortion of a received DD video could be expressed as:

$$D = P_{00} \cdot d_0 + P_{01} \cdot d_1 + P_{10} \cdot d_2 + P_{11} \cdot \sigma^2, \qquad (1)$$

where σ^2 is the variance of the source.

Let R_h be the rate in bits per sample of Description h, $h = 1, 2$. The rate-distortion region for an *i.i.d.* memoryless Gaussian source with the square error distortion measure was first introduced in [14]. For computational efficiency, in [12], Alasti *et al.* employed the following rate-distortion region, which is also used in the present paper.[1]

$$\begin{cases} d_0 = \frac{2^{-2(R_1+R_2)}}{2^{-2R_1}+2^{-2R_2}-2^{-2(R_1+R_2)}} \cdot \sigma^2 \\ d_1 = 2^{-2R_1} \cdot \sigma^2 \\ d_2 = 2^{-2R_2} \cdot \sigma^2. \end{cases} \qquad (2)$$

[1]Note that other empirical rate-distortion models, e.g., the model used in [6], can be incorporated into this framework as well.

B. Computing Distortion for Two Given Paths

A wireless mobile ad hoc network can be modeled as a probabilistic directed graph $\mathcal{G}\{V, E\}$, where V is the set of vertices and E the set of edges. We assume that nodes are reliable during the video session, but a link may fail with certain probabilities. Accurate and computationally efficient characterization of an end-to-end path in a wireless ad hoc network, with consideration of mobility, interference, and time-varying wireless channels, is extremely difficult and remains an open research problem. As an initial step, we focus on network layer characteristics in this paper, assuming that physical and MAC layer dynamics of wireless links can be translated into network layer parameters. For example, we could characterize a link $\{i, j\} \in E$ by: (i) b_{ij}, the available bandwidth of link $\{i, j\}$; (ii) p_{ij}, the "up" probability of link $\{i, j\}$; and (iii) l_{ij}, the average loss burst length on link $\{i, j\}$.

We assume (as in, e.g., [15]–[17]) that network dynamics occur at a relatively larger timescale such that topology and link metric changes could be propagated to the network in a timely fashion. Furthermore, traffic from other video sessions is regarded as background traffic, and is reflected in the available link bandwidth b_{ij}. Similarly, link layer retransmissions/buffering could actually be taken into account in the link parameters p_{ij} and l_{ij}. That is, p_{ij} and l_{ij} can be interpreted as network layer metrics, which are measured at the network layer and has already taken into consideration the lower layer dynamics. In practice, these parameters could be measured by nodes in the network. In proactive link state routing protocols, the measured parameters could be distributed throughout the network using Link State Advertisements (LSAs) [16]. In reactive routing protocols, such measured parameters could be piggybacked in route replies (RREP) in response to an on-demand route discovery [15].

Within the network, let there be two sets of streaming servers, denoted as $\mathcal{S}_h$, each hosting Description h of a video in their cache or public directory, $h = 1, 2$. Note that these two sets do not have to be disjoint. If $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \emptyset$, then nodes in $\mathcal{S}_1 \cap \mathcal{S}_2$ can offer both descriptions of the MD video. For video streaming applications, usually server nodes do not perform on-line coding. Therefore, we assume that the descriptions have fixed and balanced rates, i.e., both descriptions have the same rate R, $R_1 = R_2 = R$. Unbalanced descriptions, i.e., $R_1 \neq R_2$, can be easily handled in the proposed framework, which we have omitted for the sake of brevity.

Before we mathematically formulate the problem of joint routing and server selection, we need to compute the average distortion D as a function of link statistics for a *given* pair of servers and paths. We first define the indices for describing the choice of a pair of paths:

$$x_{ij}^h = \begin{cases} 1, & \text{if } \{i,j\} \in \mathcal{P}_h, \ \ \{i,j\} \in E, h = 1,2 \\ 0, & \text{otherwise}, \ \ \{i,j\} \in E, h = 1,2. \end{cases} \quad (3)$$

With these index variables, an arbitrary path $\mathcal{P}_h$ can be represented by a vector of $|E|$ elements, each of which corresponds to a link and has a binary value. The link bandwidth constraints can then be expressed as:

$$x_{ij}^1 \cdot R_1 + x_{ij}^2 \cdot R_2 \leq \rho \cdot b_{ij}, \ \ \{i,j\} \in E, \quad (4)$$

where ρ is a constant to convert bandwidth to video rate (bits/pixel), and $\rho = \gamma \cdot W \cdot H \cdot f$ for a video with frame rate f and frame size $W \times H$; γ is a constant determined by the chroma subsampling scheme (e.g., $\gamma = 1.5$ for the quarter common intermediate format (QCIF) videos).

We now consider how to compute the end-to-end success probabilities. Since we do not mandate "disjointedness" in routing, $\mathcal{P}_1$ and $\mathcal{P}_2$ may share nodes and links. We classify the links along the two paths into three sets: set one consisting of joint links shared by both paths, denoted as $\mathcal{J}(\mathcal{P}_1, \mathcal{P}_2)$, and the other two sets consisting of disjoint links on the two paths, denoted respectively as $\bar{\mathcal{J}}(\mathcal{P}_h)$, $h = 1, 2$. For disjoint portions of the paths, it suffices to model the packet losses as Bernoulli events, since the losses of the two descriptions are independent. Therefore, the success probabilities on the disjoint portions are:

$$p_{dj}^h = \begin{cases} \prod_{\{i,j\} \in \bar{\mathcal{J}}(\mathcal{P}_h)} p_{ij}, & \text{if } \bar{\mathcal{J}}(\mathcal{P}_h) \neq \emptyset, \ h = 1,2 \\ 1, & \text{otherwise}, \ h = 1, 2. \end{cases} \quad (5)$$

On the joint portion of the paths, the losses of the two streams are correlated. In order to capture such correlation, we model each link $\{i, j\}$ as an on-off process modulated by a discrete-time Markov chain. There is no packet loss when the link is in the "up" state, and the packet loss rate is one when the link is in the "down" state. Let α_{ij} and β_{ij} be the transition probability from "up" to "down" and from "down" to "up," respectively. The transition probabilities can be computed from the measured link statistics, as $\beta_{ij} = 1/l_{ij}$ and $\alpha_{ij} = (1 - p_{ij})/(p_{ij} l_{ij})$. If there are K shared links, the aggregate failure process of these links is a Markov process with 2^K states. In order to simplify the computation, we follow a similar approach in [6] and [13] to model the aggregate process as an on-off process. Observe that all the states that have at least one "down" link, are equivalent from the perspective of packet survivability. Therefore, we can lump such states into a single "down" state, and use the single remaining state where all the links are in the good condition as the "up" state.

Letting T_{on} be the average length of the "up' period, we have $T_{on} = 1/\left[1 - \prod_{\{i,j\} \in \mathcal{J}(\mathcal{P}_1, \mathcal{P}_2)} (1 - \alpha_{ij})\right]$. Furthermore, the average success probability of the joint portion is:

$$p_{jnt} = \begin{cases} \prod_{\{i,j\} \in \mathcal{J}(\mathcal{P}_1, \mathcal{P}_2)} p_{ij}, & \text{if } \mathcal{J}(\mathcal{P}_1, \mathcal{P}_2) \neq \emptyset \\ 1, & \text{otherwise}. \end{cases} \quad (6)$$

The transition probabilities of the aggregate on-off process are: $\alpha = 1/T_{on}$ and $\beta = p_{jnt}/\left[T_{on} \cdot (1 - p_{jnt})\right]$. Note that $\alpha = 0$ and $\beta = 0$ if $\mathcal{J}(\mathcal{P}_1, \mathcal{P}_2) = \emptyset$. The consolidated path model is illustrated in Figure 1, where $\mathcal{J}(\mathcal{P}_1, \mathcal{P}_2)$ is modeled as a two-state Markov process with parameters $\{\alpha, \beta\}$, and $\bar{\mathcal{J}}(\mathcal{P}_h)$ is modeled as a Bernoulli process with parameter $(1 - p_{dj}^h)$, $h = 1, 2$.

With the above path model, the joint probabilities of receiving the descriptions can be computed as:

$$\begin{cases} P_{00} = p_{jnt} \cdot (1 - \alpha) \cdot p_{dj}^1 \cdot p_{dj}^2 \\ P_{01} = p_{jnt} \cdot p_{dj}^1 \cdot \left[1 - (1 - \alpha) \cdot p_{dj}^2\right] \\ P_{10} = p_{jnt} \cdot p_{dj}^2 \cdot \left[1 - (1 - \alpha) \cdot p_{dj}^1\right] \\ P_{11} = 1 - p_{jnt} \cdot \left[p_{dj}^1 + p_{dj}^2 - (1 - \alpha) \cdot p_{dj}^1 \cdot p_{dj}^2\right]. \end{cases} \quad (7)$$

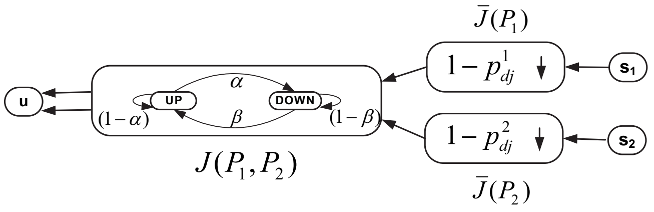

Fig. 1. A consolidated path model for double-description video.

Let $a = 2^{-2R_1}$ and $b = 2^{-2R_2}$. For balanced descriptions with rate R, we have that $a = b = 2^{-2R}$. The average video distortion can be derived by substituting (2) and (7) into (1):

$$\frac{D}{\sigma^2} = 1 + p_{jnt} \cdot \Big[(a-1) \cdot p_{dj}^1 + (b-1) \cdot p_{dj}^2 + (1-\alpha)\frac{(a+b)(a-1)(b-1)}{a+b(1-a)} \cdot p_{dj}^1 \cdot p_{dj}^2\Big]. \quad (8)$$

C. The Optimal Routing and Server Selection Problem

With the above preliminaries, we can mathematically formulate the joint routing and server selection problem for DD video (OPT-JRSS).

$$\textbf{Min}: \; D = P_{00} \cdot d_0 + P_{01} \cdot d_1 + P_{10} \cdot d_2 + P_{11} \cdot \sigma^2 \quad (9)$$

subject to:

$$\sum_{j:\{i,j\}\in E} x_{ij}^h = \begin{cases} \leq 1, & \text{if } i \neq u,\ i \in V, h = 1,2 \\ = 0, & \text{if } i = u,\ i \in V, h = 1,2 \end{cases} \quad (10)$$

$$\sum_{j:\{i,j\}\in E} x_{ij}^h - \sum_{j:\{j,i\}\in E} x_{ji}^h = \begin{cases} 1, & \text{if } i = s_h,\ i \in V, h = 1,2 \\ -1, & \text{if } i = u,\ i \in V, h = 1,2 \\ 0, & \text{otherwise},\ i \in V, h = 1,2 \end{cases} \quad (11)$$

$$x_{ij}^1 \cdot R_1 + x_{ij}^2 \cdot R_2 \leq \rho \cdot b_{ij}, \quad \{i,j\} \in E \quad (12)$$

$$x_{ij}^h \in \{0,1\}, \quad \{i,j\} \in E, h = 1,2 \quad (13)$$

$$s_h \in \mathcal{S}_h, \quad h = 1,2. \quad (14)$$

In Problem OPT-JRSS, $\{x_{ij}^h\}_{\{i,j\}\in E, h=1,2}$ and $\{s_h\}_{h=1,2}$ are optimization variables, representing the choice of a pair of servers and the links on a pair of paths from the chosen servers to the client. Constraints (10) and (11) guarantee that the paths are loop-free,[2] while constraint (12) guarantees that the links are stable. For a given pair of paths (and the pair of corresponding servers), the average video distortion D is determined by the end-to-end statistics and the correlation of the paths, as given in (1), (2), and (7). If multiple solutions are found having the minimum distortion, we can break ties by choosing the solution that has the largest bandwidth along the two selected paths [see (12)].

III. Lower and Upper Distortion Bounds

In the following, we first introduce several monotonicity properties of the objective function (9). We then construct a lower bound and an upper bound on the achievable video distortion.

[2]Note that although the feasible region permits disconnected subtours for any $h = 1,2$, the optimization problem model automatically precludes such a solution.

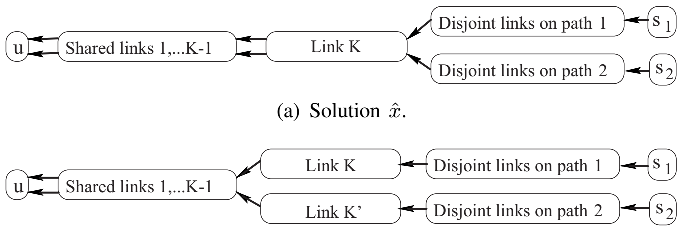

Fig. 2. The two solutions $\hat{x}$ and $\bar{x}$ satisfy the following conditions: (1) both solutions use the same pair of servers $\{s_1, s_2\}$; (2) each disjoint portion of $\hat{x}$ contains identical links as the corresponding disjoint portion in $\bar{x}$; (3) each shared link in $\bar{x}$ is identical to the corresponding shared link in $\hat{x}$; and (4) Link K in $\hat{x}$, Link K in $\bar{x}$, and Link K' in $\bar{x}$ all have the same set of parameters.

A. Properties of the Objective Function

The objective function of Problem OPT-JRSS, (9), has the following *monotonicity* properties.

Property 1: D is non-increasing with R_h, $h = 1,2$.

Proof: Recall that $a = 2^{-2R_1} \leq 1$ and $b = 2^{-2R_2} \leq 1$. From (1) and (2), we have

$$\frac{1}{\sigma^2}\frac{\partial D}{\partial R_1} = -P_{00}\frac{2\ln 2 \cdot ab^2}{(a+b-ab)^2} - 2\ln 2 \cdot P_{01} a \leq 0$$

Similarly, we have $\frac{\partial D}{\partial R_2} \leq 0$ due to the symmetry in (8). ■

Property 2: For two completely disjoint paths, D is non-increasing with p_{dj}^h, $h = 1,2$.

Proof: For a disjoint path set $\{\mathcal{P}_1, \mathcal{P}_2\}$, we have that $p_{jnt} = 1$ and $\alpha = 0$. Then, we have

$$\frac{1}{\sigma^2}\frac{\partial D}{\partial p_{dj}^1} = (a-1)\left[\frac{(1-p_{dj}^2)(a+b(1-a)) + b^2 p_{dj}^2}{a+b(1-a)}\right] \leq 0.$$

Similarly, we have $\frac{\partial D}{\partial p_{dj}^2} \leq 0$ due to the symmetry in (8). ■

Property 3: Consider the two solutions $\hat{x}$ and $\bar{x}$ shown in Figure 2. If the packet loss process of Link K is random or bursty, i.e., $\alpha_K + \beta_K \leq 1$, then $D(\hat{x}) \geq D(\bar{x})$.

Proof: Let there be K joint links with parameters $\{\alpha_k, \beta_k\}$, $k = 1, \cdots, K$ in solution $\hat{x}$, shown in Figure 2(a). Note that the two solutions $\hat{x}$ and $\bar{x}$ in Figure 2 are almost identical, except that Link K is shared in $\hat{x}$ but not shared in $\bar{x}$. Also note that Link K and Link K' have the same set of parameters. The difference between the two distortions is:

$$\frac{D(\hat{x}) - D(\bar{x})}{\sigma^2} = p_{jnt} \cdot p_{dj}^1 \cdot p_{dj}^2 \frac{(a+b)(1-a)(1-b)}{a+b(1-a)}\left[\prod_{k=1}^{K-1}(1-\alpha_k)\right] \cdot (1-\alpha_K-\beta_K)(1-p_K) \geq 0,$$

according to the "bursty" assumption, i.e., $\alpha_K + \beta_K \leq 1$. ■

The intuition behind Property 3 can be illustrated by examining the covariance of two consecutive failure events (e.g., X_m and X_{m+1}) on link K:

$$\text{Cov}\{X_m, X_{m+1}\} = \frac{\alpha_K \beta_K}{(\alpha_K+\beta_K)^2}(1-\alpha_K-\beta_K). \quad (15)$$

Algorithm ALG-LB

```
Remove link(s) {i,j} having ρ·b_ij < R, ∀{i,j} ∈ E to obtain a
  reduced graph G(V,E');
Set the cost of link {i,j} to log(1/p_ij), ∀{i,j} ∈ E';
Find the path P_h^l from a server s_h^l ∈ S_h to u in G(V,E') that has the
  minimum cost among all paths from any s_h ∈ S_h to u, h = 1,2;
Assuming P_1^l ∩ P_2^l = ∅, compute D(x_l^*), where x_l^* = {s_1^l, s_2^l, P_1^l, P_2^l}.
```

Fig. 3. Construct a lower bounding solution x_l^*.

If $\alpha_K + \beta_K < 1$, the two successive failures (or losing both descriptions sent back to back on this link) are *positively* correlated, i.e., the failure process is bursty, which, we argue, is typical in wireless ad hoc networks. When $\alpha_K + \beta_K = 1$, the two successive failures are uncorrelated, corresponding to *random* packet losses. When $\alpha_K + \beta_K > 1$, the successive failures are *negatively* correlated (called *sub-bursty*), which, we believe, is rare in wireless ad hoc networks. In Figure 2, if the Kth shared link has bursty losses, then $\bar{x}$ yields a lower distortion than $\hat{x}$; if the Kth shared link has random losses, then the two solutions yield the same distortion. This property is used in ALG-LB to construct a lower bounding solution for the achievable distortion, and is used in the proof of Proposition 1 (see Section III-B).

B. A Distortion Lower Bound

We are now ready to construct tight bounds on the average video distortion. From the monotonicity properties of D, we need to find a path pair (and the corresponding server pair) having the largest end-to-end bandwidths, the best loss characteristics, and are link-disjoint to each other. In this section, we first relax the bandwidth constraints [i.e., (12)] and path correlation to obtain a lower bound on D.

Algorithm ALG-LB in Figure 3 can be used to construct a solution that yields a lower bound for D. Let $x_l^* = \{s_1^l, s_2^l, \mathcal{P}_1^l, \mathcal{P}_2^l\}$ denote the lower bounding solution, i.e., a pair of servers $\{s_1^l, s_2^l\}$ and the corresponding pair of paths to them $\{\mathcal{P}_1^l, \mathcal{P}_2^l\}$. In ALG-LB, we first remove those links that do not have sufficient capacity to support a single description, since any solution containing such links would be infeasible [i.e., violating constraint (12)]. We then set the cost of a link $\{i,j\}$ to $\log(1/p_{ij})$, for all $\{i,j\} \in E'$. Then, the total cost of a path $\mathcal{P}$ is:

$$\sum_{\{i,j\}\in\mathcal{P}} \log\left(\frac{1}{p_{ij}}\right) = \log\left(\frac{1}{\prod_{\{i,j\}\in\mathcal{P}} p_{ij}}\right).$$

Applying a shortest path routing algorithm, we can find a min-cost path $\mathcal{P}_h^l$ from a server s_h^l for each of the server sets $\mathcal{S}_h$, $h = 1, 2$, to the user u such that the end-to-end success delivery ratio, $\prod_{\{i,j\}\in\mathcal{P}_h^l} p_{ij}$, is maximized. Finally, we compute the distortion achieved by the chosen solution $x_l^* = \{s_1^l, s_2^l, \mathcal{P}_1^l, \mathcal{P}_2^l\}$ while assuming $\mathcal{P}_1^l$ and $\mathcal{P}_2^l$ are disjoint.

According to (8) and Property 2, if the paths $\{\mathcal{P}_1^l, \mathcal{P}_2^l\}$ found in ALG-LB are really disjoint, then they are the optimal solution to Problem OPT-JRSS; otherwise, the computed distortion assuming that $\{\mathcal{P}_1^l, \mathcal{P}_2^l\}$ are disjoint will be a lower bound for the distortion achieved by the optimal solution (according to Property 3). Therefore, we have the following proposition holding true for the constructed solution x_l^*.

Proposition 1: The distortion, $D(x_l^*)$, of x_l^* constructed by ALG-LB is a lower bound for the average distortion D defined in (8).

Proof: Let the global optimal solution be $x^* = \{s_1^*, s_2^*, \mathcal{P}_1^*, \mathcal{P}_2^*\}$, i.e., a pair of optimal servers $\{s_1^*, s_2^*\}$ and the corresponding pair of optimal paths to the servers $\{\mathcal{P}_1^*, \mathcal{P}_2^*\}$. The formation of x^* could conform to one of the following two cases:

Case I: If x^* is comprised of a pair of disjoint paths, then from the construction procedure, we have that $p_{dj}^1(\mathcal{P}_1^l, \mathcal{P}_2^l) \geq p_{dj}^1(\mathcal{P}_1^*, \mathcal{P}_2^*)$ and $p_{dj}^2(\mathcal{P}_1^l, \mathcal{P}_2^l) \geq p_{dj}^2(\mathcal{P}_1^*, \mathcal{P}_2^*)$, where $p_{dj}^h(\mathcal{P}_1, \mathcal{P}_2) = \prod_{\{i,j\}\in\bar{\mathcal{J}}(\mathcal{P}_h)} p_{ij} = \prod_{\{i,j\}\in\mathcal{P}_h} p_{ij}$ for disjoint paths $\{\mathcal{P}_1, \mathcal{P}_2\}$, $h = 1, 2$ [see (5)]. From Property 2, we have that $D(x_l^*) \leq D(x^*)$.

Case II: If $\mathcal{P}_1^*$ and $\mathcal{P}_2^*$ share K links, we can construct a virtual solution $\bar{x}^* = [\bar{\mathcal{P}}_1^*, \bar{\mathcal{P}}_2^*]$, by (i) appending a copy of the shared link k to each of the two disjoint portions; (ii) removing the shared link k from the shared portion, $k = 1, \cdots, K$ (see Figure 2). That is, we construct a solution $\bar{x}^*$ with disjoint paths and identical links to x^* by duplicating each shared link in x^*. Note that as a result, $\bar{x}^*$ may not be realizable. By applying Property 3 repeatedly for K times, we have that $D(\bar{x}^*) \leq D(x^*)$. Finally, from Case I, we have that $D(x_l^*) \leq D(\bar{x}^*) \leq D(x^*)$. ■

In ALG-LB, $\mathcal{P}_h^l$ can be found by applying Dijkstra's algorithm to first find the lowest cost paths to each server in $\mathcal{S}_h$, with a time complexity of $O(|\mathcal{S}_h| \cdot (|E| + |V| \cdot \log|V|))$, and then choosing the server (and the corresponding path) having the minimum cost path among all servers in $\mathcal{S}_h$, with a time complexity of $O(|\mathcal{S}_h|)$, $h = 1, 2$. Note that although the computed $\mathcal{P}_1^l$ and $\mathcal{P}_2^l$ may share links, we assume that they are completely disjoint in order to obtain a distortion lower bound. As a result, the solution x_l^* that achieves the lower bound may not be realizable.

C. A Distortion Upper Bound

Although the above lower bound is very useful in providing a close approximation for the minimum achievable distortion by jointly selecting the optimal servers and the corresponding optimal paths, Algorithm ALG-LB may not provide a usable set of servers and paths for client u. In this section, we present an algorithm to construct a feasible solution that yields an upper bound on D.

Algorithm ALG-UB in Figure 4 can be used to construct an upper bounding solution. Let the upper bounding solution be $x_u^* = \{s_1^u, s_2^u, \mathcal{P}_1^u, \mathcal{P}_2^u\}$. The first three steps in ALG-UB are exactly the same as those in ALG-LB. However, we take into consideration link bandwidth constraints when choosing the second server/path, in order to make a feasible solution, and consider the path correlation when computing distortion. Specifically, we remove those links on $\mathcal{P}_1^u$ that do not have sufficient bandwidth to support both descriptions in Step 6, and then compute the optimal path to the second server set. Finally, when the upper bounding solution $x_u^* = \{s_1^u, s_2^u, \mathcal{P}_1^u, \mathcal{P}_2^u\}$ is available, we compute distortion by taking into consideration of the correlation of the two paths $\{\mathcal{P}_1^u, \mathcal{P}_2^u\}$. In contrast to ALG-LB, both link bandwidth constraint and path correlation are accounted for in ALG-UB.

Algorithm ALG-UB

Remove link(s) $\{i,j\}$ having $\rho \cdot b_{ij} < R, \forall \{i,j\} \in E$ to obtain a
  reduced graph $G(V, E')$;
Set the cost of link $\{i,j\}$ to $\log(1/p_{ij}), \forall \{i,j\} \in E'$;
Find the path $\mathcal{P}_1^u$ from a server $s_1^u \in \mathcal{S}_1$ to u in $G(V, E')$ that has
  the minimum cost among all the paths from any server in $\mathcal{S}_1$;
From $G(V, E')$, remove link(s) $\{i,j\}$ having $\rho \cdot b_{ij} < 2R, \forall \{i,j\}$
  $\in \mathcal{P}_1^u$ to obtain a further reduced graph $G(V, E'')$;
Find the path $\mathcal{P}_2^u$ from a server $s_2^u \in \mathcal{S}_2$ to u in $G(V, E'')$ that has the
  minimum cost among all the paths from any server in $\mathcal{S}_2$;
Compute $D(x_u^*)$, where $x_u^* = \{s_1^u, s_2^u, \mathcal{P}_1^u, \mathcal{P}_2^u\}$.

Fig. 4. Construct an upper bounding solution x_u^*.

As in ALG-LB, $\mathcal{P}_1^u$ can be found by applying Dijkstra's algorithm with a time complexity of $O(|\mathcal{S}_1| \cdot (|E| + |V| \cdot \log|V|))$ and $\mathcal{P}_2^u$ can be found with a time complexity of $O(|\mathcal{S}_2| \cdot (|E| + |V| \cdot \log|V|))$. For the constructed solution x_u^*, we have the following proposition holding true.

Proposition 2: The distortion of x_u^* constructed in ALG-UB, $D(x_u^*)$, is an upper bound for the average distortion D defined in (8).

Proof: Clearly, $x_u^* = \{s_1^u, s_2^u, \mathcal{P}_1^u, \mathcal{P}_2^u\}$ is a feasible solution to Problem OPT-JRSS, since it satisfies all the constraints (10)–(14). Therefore, $D(x_u^*)$ must be an upper bound for D, which is the distortion of the global optimal solution x^*. ■

The four-tuple $\{s_1^u, s_2^u, \mathcal{P}_1^u, \mathcal{P}_2^u\}$ provides a usable solution to Problem OPT-JRSS. We will show that the lower and upper bounds derived in this section are very close to each other. In other words, the upper bound is near-optimal in all the cases that we examined.

IV. Simulation Results

In this section, we present simulation studies on the performance of the proposed algorithms using a customized simulator developed by the authors. We first examine the quality of the lower and upper bounds, and then compared the proposed algorithms to three existing server selection schemes.

A. Simulation Settings

In each experiment, we generate an ad hoc network topology by uniformly placing a number of nodes at random locations in a square region. Connectivity is determined by the distance coverage of each nodes transmitter (set to 250 m in all the following experiments). In order to obtain connected networks, the area of the network is adjusted for different numbers of nodes to achieve an appropriate node density (e.g., the "magic number" in [18]). For the experiments reported in this section, we choose the area such that there are eight nodes in a node's transmission coverage.

The client node and server nodes are randomly chosen. For large-sized networks, we avoid the trivial cases where the servers are within two hops from the client node. For all the experiments reported, the success probability p_{ij} is uniformly chosen from $[0.9, 0.995]$, for all $\{i,j\} \in E$. The proposed algorithms, and the algorithms used for comparison, are executed on such networks to compute a pair of servers and a pair of paths to the servers for a DD video session. We set the variance σ^2 to 1, since it does not influence routing and server selection decisions. Other parameter settings will be introduced in the following when the results are discussed. The computation time is in tens of milliseconds for all the experiments.

B. Optimality of the Distortion Bounds

One important performance concern is the optimality of the proposed lower and upper distortion bounds. Table II present the distortion bounds found by ALG-UB and ALG-LB for three 10-node and three 15-node networks, as well as the global optimal distortion values found by exhaustive search (ES). For the 10-node networks, we have $|\mathcal{S}_1| = 2$ and $|\mathcal{S}_2| = 3$; for the 15-node networks, we have $|\mathcal{S}_1| = 3$ and $|\mathcal{S}_2| = 4$. The available bandwidth of each wireless link b_{ij} is uniformly chosen from $[128Kb/s, 448Kb/s]$, in steps of $64Kb/s$, for all $\{i,j\} \in E$. The video description rate and mean burst length of the links are varied to demonstrate their impact on the distortion bounds.

We observe that for all the cases, the global optimal distortion found by ES always lies between the corresponding lower and upper bounds. In addition, the difference between the lower and upper bounds is negligible. When l_{ij} is in the range of two and six, the largest difference between the lower and upper bounds is 0.012, giving a relative difference of 4.7% [computed as $(UB - LB)/LB$]. When l_{ij} is in the range of 10 and 25, the largest difference between the two bounds is 0.011, yielding a relative difference of 4.2%. Such small gap between the lower and upper bounds indicates that they are both close to the global optimum.

We also perform extensive simulations for larger sized networks where exhaustive search is impractical. The distortion values presented in Table III are obtained for a 50-node network, an 80-node network, and a 100-node network. There are 10, 13, and 15 servers in each server set for the 50-, 80-, and 100-node networks, respectively. The description rate R is increased from $64Kb/s$ to $384Kb/s$, in steps of $64Kb/s$, while l_{ij} is uniformly chosen between two and six for each link. Again, the proposed bounds were very close to each other in all of the cases examined. In several cases, the lower and upper bounds yield the same distortion value, implying that they are actually the global optimal. For example, the maximum relative difference between the lower and upper bounds in Table III is about 10% (the 100-node network, $R = 384Kb/s$ case), while in several cases (e.g., the 80-node network, $R = 64Kb/s$ case) ALG-UB does find the global optimal solution.

We find that the proposed bounds can provide an excellent estimate for the global optimal solution. The servers and the corresponding paths found by ALG-UB yield a highly competitive solution to Problem OPT-JRSS. In addition, since ALG-UB is based on Dijkstra's algorithm, the computation time for each run is in tens of milliseconds using a Pentium-4 2.4 GHz computer (with 512 MB memory), even for a 100-node network (70-90 ms). Indeed, the proposed algorithms are computationally efficient and are suitable for joint routing and server selection for large-sized ad hoc networks.

C. Comparison with Existing Algorithms

In order to compare with the existing server selection schemes, we implement the following three server selection algorithms proposed in [6] for MD video streaming in CDN.

TABLE II
COMPARISON OF AVERAGE DISTORTIONS FOR SMALL-SIZED NETWORKS

Range of l_{ij}	$2 \leq l_{ij} \leq 6$						$10 \leq l_{ij} \leq 25$					
RKb/s	64		192		384		64		192		384	
Network Size	10	15	10	15	10	15	10	15	10	15	10	15
UB	0.756	0.757	0.477	0.475	0.266	0.264	0.756	0.757	0.534	0.476	0.266	0.265
ES	0.756	0.757	0.475	0.472	0.254	0.264	0.756	0.757	0.533	0.472	0.255	0.265
LB	0.755	0.756	0.471	0.470	0.254	0.256	0.755	0.756	0.514	0.470	0.255	0.256
SP	0.806	0.757	0.685	0.614	0.429	0.414	0.806	0.757	0.701	0.614	0.429	0.414
Heuristic	0.806	0.760	0.685	0.614	0.429	0.414	0.806	0.760	0.677	0.614	0.429	0.414
Distortion	0.775	0.757	0.674	0.582	0.429	0.414	0.785	0.757	0.605	0.582	0.429	0.414

1) Shortest Path (SP): pick the closest server (in terms of hop count) from each server set.
2) Heuristic: compute a score, $r_{mn} = (L_m + L_n)/2 + L^J_{mn}$, for each pair of servers $\{s_m, s_n\}$ that host complementary descriptions, where L_m (L_n) is the path length in hop-count from server s_m (s_n) to u (for a given path), and L^J_{mn} is the number of shared links between these two paths. The server pair with the lowest score is selected.
3) Distortion: calculate the expected distortion for each server pair having complementary descriptions. The server pair with the lowest distortion is selected.

For these three schemes, SP and Heuristic does not explicitly require measure link parameters. However, since both schemes need to identify links that does not have sufficient available bandwidth, and exclude such links from being used, certain kind of measurement should still be adopted. The Third scheme, Distortion, requires link parameters for computing video distortion. It has the similar requirements in link parameter measurement and distribution as the proposed algorithm.

1) Varying Loss Burst Lengths: The distortion value obtained by the three algorithms are presented in Table II for six small-sized networks. We find that the Distortion algorithm has the best performance among the three, since it explicitly optimizes the received video distortion. SP and Heuristic are simple heuristic algorithms and there is no general rule on which one is better than the other. Since they only consider path length or path correlation, they do not achieve good performance as compared to Distortion.

Another interesting observation is that sometimes a distortion value found by an algorithm remains the same for different mean burst lengths (i.e., l_{ij}). This is because the paths to the chosen servers in these cases are completely disjoint, where the average distortion D does not depend on mean burst lengths [see (5), (6), and (7)]. Finally, in several cases SP and Heuristic yield the same distortion value. This is because when the shortest paths to the two servers, each belonging to a server set, are disjoint, these servers and the corresponding paths will be chosen by both algorithms.

In Tables II, ALG-UB outperforms all the three existing algorithms with a significant margin. We observe the similar trend in results for large-sized networks (50-, 80-, and 100-node networks). The corresponding results are presented in Table III. In wireless ad hoc networks, links have highly diverse qualities. Therefore, only considering hop-count in server selection would not produce good received MD video quality. For the Distortion algorithm, although it selects servers based on the computed distortion values, it does not necessarily provide good results since it only considers the given routes from the servers to the client. It may not be efficient in handling the cases when there are low quality links (e.g., low available bandwidth or high loss rates) in the given routes.

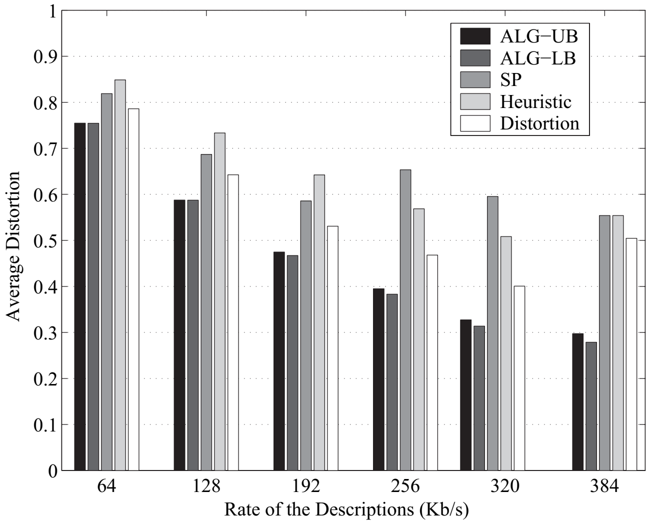

Fig. 5. Average distortions for increasing description rate R. From left to right for each value of R: ALG-UB, ALG-LB, SP, Heuristic, Distortion.

2) Increasing Video Rates: We examine the impact of the description rate R in this section. For a 100-node network having $|\mathcal{S}_1| = |\mathcal{S}_2| = 15$, we compute the average video distortion values using the algorithms, while increasing R from $64Kb/s$ to $384Kb/s$ in steps of $64Kb/s$. The link bandwidth b_{ij} is uniformly chosen from $[64Kb/s, 576Kb/s]$, also in steps of $64Kb/s$, and l_{ij} is in the range of two and six.

The distortion values obtained by the algorithms are plotted in Figure 5. Again, we find that the upper and lower bounds are very close to each other for all the rates examined, although the gap increases slightly when R gets large. In the worst case (when $R = 384Kb/s$), the difference between the bounds is 0.0188, giving a 6.3% relative difference; while in the best case (when $R = 64Kb/s$), the relative difference is only 0.01%. The proposed algorithms also outperform the existing three schemes by a large margin for all the description rates

TABLE III
COMPARISON OF AVERAGE DISTORTIONS FOR LARGE-SIZED NETWORKS: $l_{ij} \in [2, 6], \forall \{i, j\} \in E$

RKb/s	64	128	192	256	320	384
UB (50-n)	0.770	0.619	0.546	0.463	0.437	0.407
LB (50-n)	0.768	0.607	0.538	0.452	0.415	0.382
Dist. (50-n)	0.832	0.751	0.741	0.726	0.682	0.643
UB (80-n)	0.782	0.628	0.513	0.480	0.453	0.449
LB (80-n)	0.782	0.628	0.513	0.462	0.438	0.424
Dist. (80-n)	0.816	0.694	0.607	0.589	0.576	0.555
UB (100-n)	0.790	0.647	0.548	0.496	0.430	0.376
LB (100-n)	0.786	0.635	0.521	0.464	0.396	0.341
Dist. (100-n)	0.813	0.697	0.654	0.617	0.561	0.514

examined.

When R is increased, there is lower distortion due to the lossy encoder and the average distortion will decrease. On the other hand, a larger R makes more links having $b_{ij} < R$ unusable in routing and may cause path changes (since a previously chosen path may become infeasible for an increased rate). For example, for the SP scheme, the achieved distortion first decreases with the increased rate. When the rate further increases (e.g., when $R = 256Kb/s$), one of the shortest paths becomes infeasible. In this case, a new shortest path in the reduced graph (by removing links having $b_{ij} < R$) is computed and used. Such path change will cause a sudden increase in the distortion, since the new path is longer than the original path. Similar trend can be observed for the Heuristic and Distortion schemes, since they all operate on a set of given paths. For ALG-UB, since server selection and routing are jointly performed, there is much more freedom in selecting which links to use and no such sudden increase in distortion is observed.

3) Quality of Individual Frames: Since the Distortion algorithm has the best performance among the three existing algorithms, we further compare its performance with ALG-UB by transmitting double description (DD) video in a 50-node ad hoc network. There are 10 servers in each server set. We choose a time-domain partitioning coding scheme [6]–[9], [13], where two descriptions are generated by separating the even and odd-numbered frames and coding them separately (with a 10% macroblock level intra-refreshment). An H.263+ codec is used to generate the descriptions. The 400-frame QCIF [176×144 Y pixels/frame, 88×72 Cb/Cr pixels/frame] sequence "Foreman" is encoded at 15 fps and $192Kb/s$ for each description. The descriptions are then packetized (one GOB per packet) and transmitted over the paths found by the algorithms.

The average peak signal to noise ratios (PSNR) of the reconstructed video frames are plotted in Figure 6. During the period of Frame 65 to 92 and the period of Frame 270 to 290, the ALG-UB curve suffers big drops. By examining the packet loss trace, we find that there were bursty, concurrent loss of packets from both descriptions during these intervals. Furthermore, the ALG-UB curve is well above the Distortion curve for most of the frames. The average PSNRs obtained by

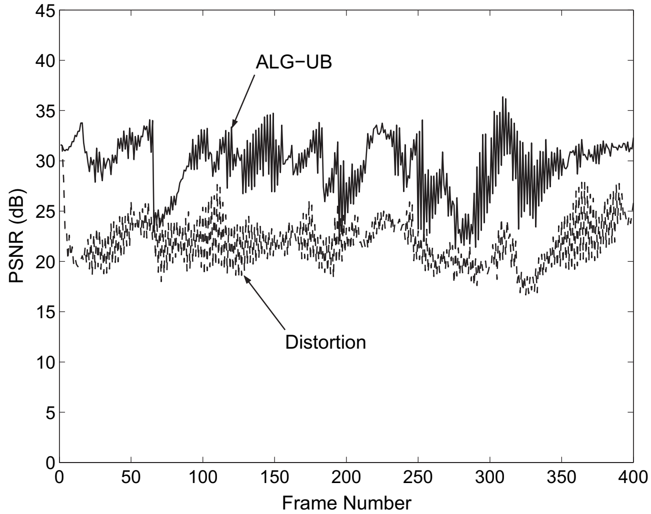

Fig. 6. PSNRs of reconstructed frames obtained by Algorithm ALG-UB and the Distortion scheme.

ALG-UB and Distortion are $29dB$ and $21.9dB$, respectively. By jointly optimizing the routing and server selection decisions, an $8.1dB$ gain in average PSNR has been achieved. Such significant gains demonstrate the efficacy of the joint routing and server selection approach for MD video in wireless ad hoc networks.

V. PRACTICAL IMPLICATIONS

In practice, the joint routing and server selection scheme can be incorporated into existing distributed routing protocols for wireless ad hoc networks. Existing routing protocols can be roughly categorized as *proactive*, where a consistent and up-to-date view of the network is always maintained, and *reactive*, where route discovery is performed on-demand. For proactive routing protocols (e.g., OLSR [16]), we can define a new type of Link State Advertisement (LSA), in addition to the original types that report link states and statistics, to disseminate the availability of video descriptions at each node. Link parameters can be piggybacked in the LSAs. Then, a client node can determine the two server sets from the received LSAs and use Algorithm ALG-UB to quickly find near-optimal servers and paths to them.

Under reactive routing protocols (e.g., DSR [15]), we can let the client node broadcast Video Request (VREQ) messages [rather than Route Request (RREQ) messages in the original DSR] to the network in order to discover nodes that host one or both of the video descriptions. Such a node, after receiving the VREQ message, will return a Video Reply (VREP) message to the client, carrying information on which description(s) it has, link statistics, and path information. After receiving a number of such VREPs, the client can construct a partial view of the network and the server sets, and then run Algorithm ALG-UB to select the best servers along with associated routes to them.

Note that the link parameters are piggybacked in routing control messages, and only those link parameters that have changed need to be disseminated. The additional overhead of distributing the network layer parameters should be moderate. In addition, since video applications usually have traffic changing over longer timescales, such overhead will be amortized.

Accurate measurement/estimate of available link capacity is a difficult problem. There are several possible approaches to address it. For instance, many wireless network interface cards work with several discrete, fixed rates. They adjust the current data rates according to the signal strength or the SNR (e.g., in WLAN cards). A rough estimate of the available link capacity could be made by measuring the SNR values and the traffic rates to the corresponding neighbor averaged over a time window. Second, estimating path parameters based on end-to-end measurements has been an active research area for years. There exist many effective techniques in the literature (e.g., see [19], [20]). We conjecture that these techniques could be adapted to provide accurate estimate of available link capacity as well.

So far, we have only considered joint routing and server selection from a client's perspective. For the traditional server selection problem, load balancing for servers is a major design objective in order to reduce congestion and server response time [2]–[4]. Such a load balancing issue, when a large number of clients exist, can be easily addressed in our proposed framework by defining a virtual link for each server as the first link for any path starting from the server. For a server s_i, the virtual link has an available bandwidth in proportion to the server's available processing capability. Running Algorithm ALG-UB on this augmented network will ensure load balancing among the servers.

VI. Related Work

One of the related papers, [6], has been discussed in Sections I and IV. In the following, we briefly discuss other related work in the literature.

Caching and service replication are effective techniques for providing scalable distributed service over the Internet. The single server selection problem, i.e., how to select a server from a set of mirror sites for a client request so as to provide the "best" service for the client, has been studied over the years (e.g., see [2]–[4] and references therein). In existing server selection schemes, either the client or server monitor server load and/or network performance (e.g., round trip times from the servers to the client) and then select a "best" server based on these measurements. These schemes are mainly designed for data applications (e.g., web service) and do not explicitly attempt to optimize video quality. Moreover, the optimal routing problem has not been addressed.

End system based mechanisms have been explored for media streaming or data download from multiple mirror sites in parallel [21]–[23]. In [21], Tornado Codes are used to enable a client to access a file from multiple mirror sites in parallel to speed up the download process. In [22], the authors propose a packet scheduling algorithm for DD video streaming from a pair of servers. The objective is to minimize video distortion under the given rate constraints. In [23], Nguyen and Zakhor propose a receiver-driven protocol for simultaneous video streaming from multiple senders to a single receiver. The objectives of this work are to achieve higher throughput, and to increase tolerance to packet loss and delay caused by congestion. These works demonstrate the benefits of concurrently accessing multiple servers, such as improved error resilience and reduced downloading time. A fixed pair of servers and given pair of paths are used in all of these studies.

In our previous work, we studied the problem of optimal routing for MD video for point-to-point communications [8] and multicast applications [9] in ad hoc networks. We introduced a Genetic Algorithm (GA)-based approach for the optimal routing problems. The GA-based approach in [8] and [9] could be adapted for Problem OPT-JRSS addressed in this paper. Since the constructed upper bound and lower bound are very close to each other for all the cases studied, the upper bound suffices to produce a highly competitive solution to Problem OPT-JRSS.

VII. Concluding Remarks

In this paper, we studied the problem of joint routing and server selection for DD video streaming in wireless ad hoc networks. Based on the monotone properties of the average video distortion, we derived a lower bound and an upper bound for the best achievable video distortion. The upper bound was demonstrated to produce a near-optimal pair of servers along with a pair of corresponding paths. Numerical results show that the bounds are very close to each other for all the cases studied, indicating the near-global optimality of the derived upper bounding solution. We also observed significant gains in video quality achieved by the proposed approach over existing server selection schemes. This justifies the importance of jointly considering routing and server selection. The proposed approach are computationally efficient and can be incorporated into existing routing protocols for optimal DD video streaming in wireless ad hoc networks.

Acknowledgments

The authors are grateful to editor Prof. Jennifer Hou and the anonymous reviewers for their constructive comments that improved the quality of this paper.

References

[1] B. Li and K. H. Wang, "Nonstop: continuous multimedia streaming in wireless ad hoc networks with node mobility," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1627–1641, Sept. 2003.

[2] R. L. Carter and M. E. Crovella, "Server selection using dynamic path characterization in wide-area networks," in *Proc. IEEE INFOCOM 1997*, pp. 1014–1021.

[3] Z.-M. Fei, S. Bhattacharjee, E. W. Zegura, and M. H. Ammar, "A novel server selection technique for improving the response time of a replicated service," in *Proc. IEEE INFOCOM 1998*, pp. 783–791.

[4] Y. Dong, Z.-L. Zhang, and Y. T. Hou, "Server-based dynamic server selection algorithms," *Springer-Verlag Lecture Notes in Computer Science (LNCS)*, vol. 2344, pp. 575–584, 2002.

[5] V. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, pp. 74–93, Sept. 2001.

[6] J. Apostolopoulos, T. Wong, W. Tan, and S. Wee, "On multiple description streaming with content delivery networks," in *Proc. IEEE INFOCOM 2002*, pp. 1736–1745.

[7] S. Mao, S. Lin, S. S. Panwar, Y. Wang, and E. Celebi, "Video transport over ad hoc networks: Multistream coding with multipath transport," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1721–1737, Dec. 2003.

[8] S. Mao, Y. T. Hou, X. Cheng, H. D. Sherali, S. F. Midkiff, and Y.-Q. Zhang, "On routing for multiple description video over wireless ad hoc networks," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 1063–1074, Oct. 2006.

[9] S. Mao, X. Cheng, Y. T. Hou, and H. D. Sherali, "Multiple description video multicast in wireless ad hoc networks," *ACM/Kluwer Mobile Networks and Applications J. (MONET)*, vol. 11, no. 1, pp. 63–73, Jan. 2006.

[10] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proc. IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[11] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Commun. Mag.*, vol. 12, no. 3, pp. 59–65, June 2005.

[12] M. Alasti, K. Sayrafian-Pour, A. Ephremides, and N. Farvardin, "Multiple description coding in networks with congestion problem," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 891–902, Mar. 2001.

[13] A. C. Begen, Y. Altunbasak, O. Ergun, and M. H. Ammar, "Multi-path selection for multiple description encoded video streaming," *EURASIP Sig. Proc.: Image Commun.*, vol. 20, no. 1, pp. 39–60, Jan. 2005.

[14] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 84–91, Dec. 1980.

[15] D. B. Johnson, D. A. Maltz, and Y.-C. Hu, "The dynamic source routing protocol for mobile ad hoc networks (DSR)," Apr. 2003, IETF Internet Draft, draft-ietf-manet-dsr-09.txt.

[16] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol," Oct. 2003, IETF RFC 3626.

[17] H. Badis and K. Al Agha, "QOLSR: QoS routing for ad hoc wireless networks using OLSR," *European Trans. Telecommun.*, vol. 15, no. 4, 2005.

[18] T. K. Philips, S. S. Panwar, and A. N. Tantawi, "Connectivity properties of a packet radio network model," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1044–1047, Sept. 1989.

[19] R. L. Carter and M. E. Crovella, "Measuring bottleneck link speed in packet-switched networks," *Perform. Eval.*, vol. 27/28, pp. 297–318, Oct. 1996.

[20] R. Kapoor, L.-J. Chen, L. Lao, M. Gerla, and M. Y. Sanadidi, "Capprobe: a simple and accurate capacity estimation technique," in *Proc. ACM SIGCOMM 2004*, pp. 67–78.

[21] J. W. Byers, M. Luby, and M. Mitzenmacher, "Accessing multiple mirror sites in parallel: using Tornado codes to speed up downloads," in *Proc. IEEE INFOCOM 1999*, pp. 275–283.

[22] A. C. Begen, Y. Altunbasak, and M. Begen, "Rate-distortion optimized on-demand media streaming with server diversity," in *Proc. IEEE ICIP 2003*, pp. III.657–III.660.

[23] T. Nguyen and A. Zakhor, "Multiple sender distributed video streaming," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 315–326, Apr. 2004.

Shiwen Mao (S'99–M'04) received the B.S. and the M.S. degree from Tsinghua University, Beijing, P.R. China in 1994 and 1997, respectively, both in Electrical Engineering. He received the M.S. degree in System Engineering and the Ph.D. degree in Electrical and Computer Engineering from Polytechnic University, Brooklyn, NY, in 2000 and 2004, respectively.

He was a Research Member at IBM China Research Lab, Beijing from 1997 to 1998. In the summer of 2001, he was a research intern at Avaya Labs-Research, Holmdel, NJ. He has been a Research Scientist in the Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA since December 2003. Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering at Auburn University, Auburn, AL.

Dr. Mao's research interests include cross-layer design and optimization in multi-hop wireless networks, as well as multimedia communications. He is a co-recipient of the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is the co-author of a textbook, *TCP/IP Essentials: A Lab-Based Approach* (Cambridge University Press, 2004).

Xiaolin Cheng (S'04) received his B.E. and M.E. in Automation from Tsinghua University, China in 1997 and 2000 respectively, and his M.S. in Computer Engineering from Virginia Tech in 2005.

He worked as a Senior Engineer in Panasonic Beijing Labs for two and half years. Currently, he is working toward his Ph.D. degree in the Computer Science Department at UC Davis. His research focuses on multimedia delivery over wireless networks and routing algorithms for wireless ad hoc networks.

Y. Thomas Hou (S'91–M'98–SM'04) obtained his B.E. degree from the City College of New York in 1991, the M.S. degree from Columbia University in 1993, and the Ph.D. degree from Polytechnic University, Brooklyn, New York, in 1998, all in Electrical Engineering. From 1997 to 2002, He was a researcher at Fujitsu Laboratories of America, Sunnyvale, California. Since Fall 2002, he has been an Assistant Professor at Virginia Tech, the Bradley Department of Electrical and Computer Engineering, Blacksburg, Virginia.

His current research interests are radio resource (spectrum) management and networking for software-defined radio wireless networks, optimization and algorithm design for wireless ad hoc and sensor networks, and video communications over dynamic ad hoc networks. In recent past, he also had work on scalable architectures, protocols, and implementations for differentiated services Internet, service overlay networking, video streaming, and network bandwidth allocation policies and distributed flow control algorithms. He has published over 100 journal and conference papers in the above areas. He is a member of ACM.

Hanif D. Sherali is W. Thomas Rice Endowed Chaired Professor of Engineering in the Industrial and Systems Engineering Department at Virginia Polytechnic Institute and State University. His area of research interest is in discrete and continuous optimization, with applications to location, transportation, and engineering design problems. He has published about 200 papers in Operations Research journals, has co-authored four books in this area, and serves on the editorial board of eight journals. He is a Fellow of the IIE, a Fellow of INFORMS, and a member of the National Academy of Engineering.

Jeffrey H. Reed (S'78–M'80–SM'98–F'05) is the Willis G. Worcester professor in the Bradley Department of Electrical and Computer Engineering. Dr. Reed's areas of expertise are in software radios, smart antennas, and ultra wideband. His book *Software Radio: A Modern Approach to Radio Engineering* was published by Prentice Hall in 2002 and his book *An Introduction to Ultra Wideband Communication Systems* was published by Prentice Hall in May 2005. Dr. Reed is a Fellow of the IEEE.