COLLABORATIVE RESOURCE MANAGEMENT FOR VIDEO OVER WIRELESS MULTI-HOP MESH NETWORKS

Nicholas Mastronarde[1], Deepak S. Turaga[2], and Mihaela van der Schaar[3]

[1]Electrical and Computer Engineering
University of California, Davis
nhmastro@ucdavis.edu

[2]IBM Research
Hawthorne, NY
turaga@us.ibm.com

[3]Electrical Engineering
University of California, Los Angeles
mihaela@ee.ucla.edu

ABSTRACT

In this paper, we consider the problem of real-time multimedia transmission among several peers (users). The peers use a heterogeneous wireless multi-hop mesh network for the delivery of these high-bandwidth streams. One of the main challenges of the considered problem is the division of the scarce wireless resources among the various peers. To address this problem, we propose an efficient, distributed and collaborative framework for wireless resource exchanges that enables peers to divide available wireless resources among themselves based on their quality of service (QoS) requirements, the underlying channel conditions and network topology. The scalable coding of the video content and decomposition of video flows into various sub-flows (priorities) allow peers to transfer the video at different quality levels, depending on the network load. Users collaboratively decide which of their sub-flows to admit, and which paths these sub-flows should be transmitted on in order to maximize a system defined utility. Our results show that with user collaboration, these distributed algorithms provide system and user performance comparable to a centralized exhaustive implementation.

Index Terms— *Multimedia communication*

1. INTRODUCTION

The resource management problem for wireless networks involves distributing available wireless resources (air-time at the wireless stations) among users given their QoS requirements, underlying channel conditions, and the network topology. Different centralized and distributed approaches have been proposed to solve this resource management problem. Centralized approaches solve the end-to-end routing and path selection problem as a combined optimization using algorithms designed for Multi-Commodity Flow [1] problems. Such an optimization ensures that the end-to-end throughput is maximized while constraints on individual link capacities are satisfied. In contrast, distributed approaches use, for instance, game theoretic algorithms to resolve resource allocation issues for wireless networks [2]. However, previous research has not considered the benefits of collaborative and distributed resource and information exchanges among wireless peers.

We are concerned with multi-hop wireless mesh topologies. We assume that the mesh network topology is fixed over the duration of the video sessions and that the various nodes of the mesh network employ polling-based (reservation-based) admission control similar to that employed in 802.11e WLAN networks [3]. Furthermore, we assume an overlay network topology [4] that can convey information about the expected SNR, as well as the guaranteed bandwidth under dynamically-changing physical layer modulation to each wireless node. To improve the system utilization (number of admitted users) as well as the QoS for admitted users, we use scalable video coding schemes that enable each video flow (bit-stream) to be divided into several sub-flows (layers) depending on their relative priority in terms of the overall impact on the decoded video distortion. Under this framework we design distributed algorithms for *collaborative resource exchanges* where given the average (or worst case) underlying channel conditions, source peers collaboratively solve a resource management problem by determining how many sub-flows to admit, and which paths these sub-flows should be transmitted on. Subsequently, given the resource allocation decisions, each peer optimizes its own video quality using cross-layer adaptation: adaptive modulation strategy at the physical layer (PHY), adaptive MAC retransmission limit, and distortion-optimized scheduling at the application-layer (APP) layer [5][6].

This paper is organized as follows. We introduce the notion of sub-flow, then discuss the partitioning of a scalable video flow into multiple sub-flows and introduce a quality-rate model to guide the various resource allocation tradeoffs in Section 2. In Section 3 we present distributed algorithms for collaborative path partitioning and air-time reservation. We present simulation results in Section 4 and conclude in Section 5.

2. SUB-FLOWS AND QUALITY-RATE MODEL

2.1. Partitioning scalable bit-stream into sub-flows

In [5], it has been shown that partitioning a scalable video flow into several prioritized sub-flows can improve the number of admitted flows in an 802.11e enabled wireless network, as well as the overall quality received by the peers. Each sub-flow may be viewed as a separate *quality layer* of the scalable bit-stream. We use a 3D wavelet codec that uses a spatio-temporal wavelet transform followed by embedded coding [7]. In this paper, we label data belonging to each temporal decomposition level as a separate quality layer [6]. Each sub-flow has an associated priority based on its distortion impact and delay-deadline constraint. Furthermore, each sub-flow is admitted independently by the wireless network despite dependencies among sub-flows that may limit the gains from decoding some if others are not received. The collection of sub-flows belonging to one source-destination pair is referred to as an *Aggregate Flow*.

Let there be N_p aggregate flows in the network. We label aggregate flow y, as set Ψ_y (with $1 \leq y \leq N_p$). We partition these N_p aggregate flows into N total sub-flows and label sub-flow *x* as f_x $1 \leq x \leq N$. Sub-flow f_x is a part of the aggregate flow Ψ_y, if $f_x \in \Psi_y$.

2.2. Quality-rate model: utility for collaborative resource exchange

1-4244-0481-9/06/$20.00 ©2006 IEEE

Different rate-distortion models may be used to derive a distortion metric for scalable and fair resource allocation. We define a generic quality-rate model $q_x(b_x)$: where the bit-rate b_x corresponds to the rate requirement of the sub-flow, and the function q_x depends on the sub-flow content characteristics, encoding parameters etc. The quality received by aggregate flow Ψ_y may be computed as $Q_y = \sum_{f_x \in \Psi_y} q_x(b_x)$.

Collaborative resource exchange algorithms are used to admit sub-flows and determine a path for each admitted sub-flow. In particular, for each sub-flow f_x we define an indicator variable $w_x \in \{0,1\}$ that takes values 1 and 0 based on whether the sub-flow is successfully admitted into the network, or not. Hence, the decoded video quality for the aggregate flow Ψ_y based on all its sub-flows that are either admitted or denied admission is:

$$\hat{Q}_y(P_y) = \sum_{f_x \in \Psi_y} w_x q_x(b_x), \tag{1}$$

where P_y is the set of paths selected with one path $p_x \in P_y$ per sub-flow f_x. We combine the individual quality experienced by each aggregate flow into one end-to-end utility function:

$$U_{MTQ} = \sum_y \hat{Q}_y(P_y) = \sum_y \sum_{f_x \in \Psi_y} w_x q_x(b_x), \tag{2}$$

which maximizes the total quality (MTQ) of all N_p aggregate flows. When maximizing U_{MTQ} we may admit more sub-flows from one aggregate flow than another (because sub-flows are admitted strictly in order of their contribution to the overall utility), potentially leading to a large variation in quality across the different aggregate flows. To avoid being unfair to one (or more) aggregate flows we can also maximize the minimum quality (MMQ) experienced by any aggregate flow in the system, thereby minimizing the quality variation across aggregate flows. The MMQ utility may be written as:

$$U_{MMQ} = \min_y\left(\hat{Q}_y(P_y)\right) = \min_y\left(\sum_{f_x \in \Psi_y} w_x q_x(b_x)\right), \tag{3}$$

In this paper, we use the following quality-rate model [8]:

$$q_x(b_x) = q_x^0 + \lambda_x \log(b_x), \tag{4}$$

where q_x^0 and λ_x (typically increasing with the sub-flow's distortion impact) are flow-specific parameters that depend on the underlying video characteristics, encoding parameters etc. Note, however, that other models could be used. In section 3 we describe our designed resource management algorithms that maximize these utilities.

3. COLLABORATIVE RESOURCE EXCHANGES

We allow peers to jointly determine which sub-flows they can admit into the network, and what paths they should select for each admitted sub-flow, based on the average underlying channel conditions, the bit-rate requirements of each sub-flow and their contribution to the different utilities as defined in equations (1) and (2). Peers collaborate in terms of exchanging information about selected paths, bandwidth requirements, and time reservation [3] at intermediate nodes. In this paper, all collaborative resource exchange decisions are made only when either a new source-destination pair is established, or an existing connection is terminated, hence, negligible computational cost is incurred. Let us consider an intermediate peer v_a that receives a set of sub-flows from its neighbors, where each sub-flow f_x has an associated rate b_x, and experiences different channel conditions based on which neighboring node it is transmitted from. Consider that the neighboring peer selects PHY mode θ_x^a while transmitting f_x to v_a. θ_x^a determines the bit error rate $e(\theta_x^a)$, the equivalent packet error rate for a packet size L_x, $\varepsilon_x^a(L_x,\theta_x^a)$, and the maximum achievable PHY layer transmission rate T_x^a as in [9]. Hence, the expected goodput $\overline{g}_x^a$ experienced over this link is

$$\overline{g}_x^a = (1 - \varepsilon_x^a(L_x,\theta_x^a))T_x^a(\theta_x^a). \tag{5}$$

We assume that the service interval (SI) t_{SI} for each peer is partitioned into a listening service interval $t_{SI}^{(RX)}$ and a transmission service interval $t_{SI}^{(TX)}$. We assume that t_{SI}, $t_{SI}^{(RX)}$ and $t_{SI}^{(TX)}$ are determined a priori by the deployed protocol and are the same at each node. Furthermore, we assume that $t_{SI}^{(RX)} << t_{SI}^{(TX)}$ such that congestion exists mainly because of the limited listening time.

Each receiving peer polls its neighbors for a different amount of time (within its fixed $t_{SI}^{(RX)}$) based on the number of sub-flows competing for this air-time, their priorities, and rate requirements. In particular, at v_a, a fraction r_x^a $(0 \le r_x^a \le 1)$ of the peer's listening time is allocated to f_x. This corresponds to providing sub-flow f_x a transmission opportunity $r_x^a t_{SI}^{(RX)}$ within every service interval t_{SI}. Hence, the expected rate for f_x arriving at v_a is $\overline{g}_x^a r_x^a \frac{t_{SI}^{(RX)}}{t_{SI}}$. If $\overline{g}_x^a r_x^a \frac{t_{SI}^{(RX)}}{t_{SI}} < b_x$ then f_x does not receive its rate requirement. We decide to drop sub-flows that do not receive their rate requirement as the incremental benefit of decoding partially received sub-flows outweighs the cost in terms of resources assigned to them. Consequently, we also drop all the sub-flows that depend on this sub-flow.

Given a selection of paths (with one path per sub-flow) we can determine which sub-flows may be admitted into the network, based on the available air-time at each peer, such that the desired utility is maximized. In order to determine which sub-flows should be admitted, we need to sort sub-flows in order of their impact on the end-to-end utility function.

In order to maximize U_{MTQ}, we admit sub-flows in decreasing order of the fraction $\lambda_x/b_x \left(= \partial q_x/\partial b_x\right)$. This fraction represents the tradeoff between quality (proportional to $\log b_x$) and resources allocated (proportional to b_x).

In order to maximize U_{MMQ}, we admit sub-flows for all aggregate flows quality layer by quality layer (i.e. in order of the temporal level to which the sub-flow corresponds). Within sub-flows corresponding to the same quality layer we admit them in decreasing order of the fraction λ_x/b_x.

For each sub-flow in order of this sorted list we determine the time reservation fraction at all intermediate peers. At each intermediate peer v_a along the path of this sub-flow, we determine the desired time fraction $r_x^a = \frac{b_x t_{SI}}{\overline{g}_x^a t_{SI}^{(RX)}}$. If that is less than the remaining available listening time fraction ρ^a ($0 \leq \rho^a \leq 1$), at the peer then the flow is admitted at v_a. If the flow can be admitted at all peers along its path then the flow is admitted into the network, and the time reservation fractions at each intermediate peer are modified appropriately (i.e. this fraction r_x^a is reserved at node v_a, and the listening available time fraction ρ^a is decremented appropriately). If the bandwidth requirement for the sub-flow is violated at any intermediate node, we discard it from the list, and also all other sub-flows that depend on this sub-flow.

3.1. Centralized algorithm for optimal sub-flow admission control and path provisioning

We may use the aforementioned admission control system in conjunction with a centralized algorithm in order to determine the optimal set of paths. An exhaustive approach to determining the optimal set of paths is to consider all sets of combinations of paths (with one path per sub-flow) and pick the set that maximizes the appropriate end-to-end utility. If there are a total of K_x available paths per sub-flow, then we must compare the utilities of all $\prod_{x=1}^{N} K_x$ possible path-sub-flow assignments before determining the optimal set of paths. Since the total number of sub-flows, and correspondingly paths per sub-flow can be large, an exhaustive search can be prohibitively expensive.

3.2. Distributed algorithms for optimal sub-flow admission control and path provisioning

In order to avoid having a centralized and computationally expensive algorithm, we design distributed solutions to the wireless multi-user admission control and path provisioning. In these approaches, each source peer solves the path provisioning for each of its sub-flows individually, in a greedy manner. We thus decompose the joint optimization into a set of successive path selection problems, where in each step the source peer determines one path for each sub-flow given the paths selected for the other sub-flows. The complexity of this distributed approach is $\sum_{x=1}^{N} K_x$, which is significantly lower than that of the exhaustive approach.

Under our proposed *collaborative distributed path provisioning algorithm,* source peers collaboratively determine the sorted list of sub-flows in decreasing order of their contribution to the utility function being considered (U_{MTQ} or U_{MMQ}). For each sub-flow in this sorted order, the corresponding source determines (based on the air-time available at intermediate peers) if there exists a path that provides it with its desired bandwidth end-to-end. If no such path exists, the source peer discards this sub-flow and all other sub-flows that depend on it. Alternatively, if multiple such paths exist, the source peer selects the path that leads to the smallest amount of congestion (the congestion at a node v_a may be written as $1 - \rho^a$, with ρ^a being the fraction of available listening time). In this paper, the source peer selects among multiple available paths based on two different end-to-end congestion metrics. The first choice is *bottleneck air-time congestion*. Bottleneck congestion ξ_x^i for path p_x^i (the *i*-th path of sub-flow *x*) is defined as the maximum congestion experienced at any node along p_x^i. The second choice is *mean end-to-end air-time congestion.* Mean congestion ϕ_x^i for path p_x^i is the mean congestion experienced across all nodes along the path. Bottleneck path congestion is preferable when the network topology contains bottleneck nodes (i.e. nodes through which several sub-flows must necessarily pass, as in a star topology). Otherwise, it is preferable to use mean end-to-end congestion metrics.

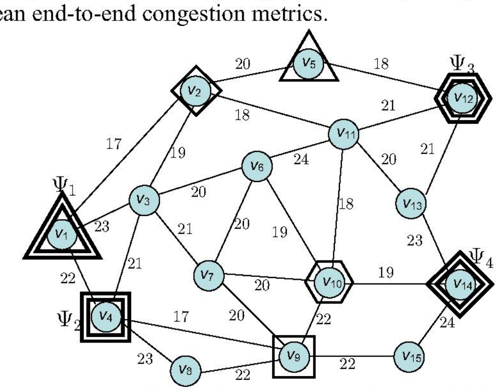

Figure 1. Network topology, average channel SNR (dB)

4. SIMULATION RESULTS

The network topology we used for our experiments, as proposed in [10], is shown in Figure 1 with average channel SNRs (in dB) and source-destination pairs marked by like shapes (with sources marked by a double border). We consider the transmission of four different aggregate flows, with different sequence characteristics and bit-rates, over this network infrastructure. The sequences selected are CIF (352×288) with 300 frames at 30 frames per second. We use a Group of Pictures (GOP) structure with 16 frames in each GOP, and a temporal decomposition with four temporal levels. Aggregate flows Ψ_1 (Coastguard) and Ψ_3 (Foreman) were encoded at 1.5 Mb/s, Ψ_2 (Mobile) at 2.0 Mb/s, and Ψ_4 (Hall) at 500 Kb/s. The quality parameter λ_x is determined for each sub-flow as the average of the priorities of the packets within the sub-flow. Each packet's priority is determined in terms of distortion impact upon loss or gain of the packet. We normalize the individual packet priorities to a maximum of 1000. More details on this may be obtained from [5]. In Table 1 we show numerical results for the optimization of U_{MTQ} using both the centralized optimal and collaborative algorithms with both end-to-end congestion metrics and under medium and high network load. In our experiments, the centralized optimal algorithm admits ~69% (11) of 16 sub-flows in the "medium" network load case, and ~44% (7) of 16 sub-flows in the "high" network load cases.

From Table 1, when optimizing U_{MTQ} for the medium load case, the use of the bottleneck congestion metric is slightly more beneficial (<0.1%) than the use of mean congestion because one more sub-flow of aggregate flow Ψ_4 is admitted. Instead, in the high load case, the mean congestion metric leads to an improved performance (>1.0%) over bottleneck congestion.

Table 1. Admitted sub-flows for optimization of U_{MTQ} under medium and high network load

Load	Algorithm	Congestion Metric	Optimized U_{MTQ}	Number of admitted sub-flows				Resulting U_{MMQ}
				Ψ_1	Ψ_2	Ψ_3	Ψ_4	
Medium	*Centralized Optimal*	-	1157.86	3	3	2	3	217.33
	Collaborative Distributed	Bottle-neck	1152.92	2	3	2	4	218.49
		Mean Air-time	1151.75	2	3	2	3	217.33
High	*Centralized Optimal*	-	1134.43	2	2	1	2	215.38
	Collaborative Distributed	Bottle-neck	1119.72	1	2	1	3	217.33
		Mean Air-time	1134.43	2	2	1	2	215.38

Table 2. Comparison of decoded PSNR, and utility contribution

Aggregate Flow	Load	Decoded PSNR	Contribution to U_{MTQ}
Ψ_3	Medium	26.08	243.88
	High	22.42	233.05
Ψ_4	Medium	34.03	218.49
	High	33.65	217.33

Increased load levels require each sub-flow to incur minimal impact across *all* nodes traversed, thus, the mean air-time metric more closely approaches the optimal solution, and in this case, actually obtains the optimal solution. Under both medium and high load, the distributed algorithm performance lies within 2.0% of the optimal solution in terms of the utility.

When optimizing U_{MMQ} we found that the distributed algorithm with bottle-neck metric has the same performance as the optimal one, and both achieved a U_{MMQ} of 218.49 under medium load, and 217.33 under high load. These numbers are similar to resulting U_{MMQ} results in Table 1. This suggests that there are several path-sub-flow combinations that optimize U_{MMQ}. In general, the almost monotonic relationship between the quality layer of the sub-flow and λ_x / b_x for real video streams ensures that optimization of U_{MTQ} often results in optimizing U_{MMQ}.

Table 2 shows the decoded PSNR and the contribution to U_{MTQ} of aggregate flows Ψ_3 and Ψ_4 (the flows with the smallest and largest number of admitted sub-flows, respectively). The shown decoded PSNR is the average PSNR of the Y channel over 3 simulations. Each simulation concurrently transmits all four aggregate flows over the network using the time reservations determined using the collaborative distributed algorithm and bottle-neck congestion metric in order to optimize U_{MTQ}. Independent error patterns and PHY layer rates corresponding to the underlying channel SNR (as in Figure 1) are generated for each simulation. Table 2 shows that there is a >3 dB decrease in average decoded PSNR of Ψ_3 under the high load case, for which only the base quality-layer is admitted, compared to the medium load case where the base layer and one enhancement layer are admitted. Ψ_4 loses <1 dB in average decoded PSNR between medium and high load. Observing the contributions of Ψ_3 and Ψ_4 to U_{MTQ} in table 2, it is clear that lower utility contributions directly lead to lower decoded quality. Note, however, that because the sub-flows are characterized by unique parameters q_x^0 and λ_x from the quality-rate model, a particular sequence may have a lower impact on the overall system utility than another, while having a higher decoded PSNR (e.g. Ψ_4).

5. CONCLUSION

We consider a scenario with multiple pairs of peers transmitting scalably encoded video, partitioned into multiple sub-flows, to each other over a shared infrastructure. We propose distributed algorithms for resource management and exchange that allow peers to collaboratively perform admission control, path provisioning and air-time reservation at intermediate nodes in order to optimize particular network utilities pertaining to the experienced decoded video quality over all users in the network. We compare these algorithms under networks experiencing both high and medium load and provide some insights into determining the best choice of parameters for our collaborative distributed algorithms under these conditions. We show that the distributed resource management algorithms perform within 2% of the centralized optimal solution for these utilities.

7. REFERENCES

[1] B. Awerbuch and T. Leighton, "Improved approximation algorithms for the multi-commodity flow problem and local competitive routing in dynamic networks," Proc. 26th ACM Symposium on Theory of Computing, May 1994.

[2] R. La and V. Anantharam, "Optimal Routing Control: Repeated Game Approach," *IEEE Trans. Automatic Control*, Vol. 47, No. 3, pp. 437 -450, March 2002.

[3] IEEE 802.11e/D5.0, Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), June 2003.

[4] D. Krishnaswamy, and J. Vicente, "Scalable adaptive wireless networks for multimedia in the proactive enterprise." Intel Technology Journal:http://developer.intel.com/technology/itj/2004/volume08issue04/art04scalingwireless/p01_abstract.htm

[5] M. van der Schaar, S. Shankar, "Cross-Layer Wireless Multimedia Transmission: Challenges, principles, and new paradigms", IEEE Wireless Communications Magazine, vol. 12, no. 4 pp. 50-58, August 2005.

[6] D. S. Turaga, M. van der Schaar, "Cross-layer Aware Packetization Strategies for Optimized Wireless Multimedia Transmission," IEEE ICIP 2005, Genova, September 2005.

[7] Y. Andreopoulos, A. Munteanu, *et al*, "In-band motion compensated temporal filtering," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 653-673, Aug. 2004.

[8] M. Dai, D. Loguinov, and H. Radha, "Rate-Distortion Modeling of Scalable Video Coders," IEEE International Conference on Image Processing (ICIP), October 2004.

[9] D. Krishnaswamy, "Network-assisted link adaptation with power control and channel reassignment in wireless networks," 3G Wireless Conference, pp. 165-170, 2002.

[10] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop mesh networks with orthogonal channels," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 868-80, August 2005.