ANALYSIS OF MULTI-USER CONGESTION CONTROL FOR VIDEO STREAMING OVER WIRELESS NETWORKS

Xiaoqing Zhu and Bernd Girod

Information Systems Laboratory,Stanford University, CA, U.S.A.
{zhuxq,bgirod}@stanford.edu

ABSTRACT

When multiple video sources are live-encoded and transmitted over a common wireless network, each stream needs to adapt its encoding parameters to wireless channel fluctuations, so as to avoid congesting the network. We present a stochastic system model for analyzing multi-user congestion control for live video coding and streaming over a wireless network. Variations in video content complexities and wireless channel conditions are modeled as independent Markov processes, which jointly determine the bottleneck queue size of each stream. Interaction among multiple users are captured by a simple model of random traffic contention. Using the model, we investigate two distributed congestion control policies: the approach based on *stochastic dynamic programming* (SDP) and a greedy heuristic. Compared to fixed-quality coding with no congestion control, performance gains in the range of 0.5-1.3 dB in average video quality are reported for both schemes from simulation results under various video and channel characteristics.

1. INTRODUCTION

The advance of wireless networking technologies and the availability of low-cost video cameras are compelling for many applications, ranging from surveillance camera networks to visual communication in search-and-rescue operations. In such scenarios, multiple video streams are captured at different locations, and need to be live-encoded and transmitted over the network with low latency.

The wireless channel fluctuation and temporal variation of video contents, however, typically introduce queuing delays. Moreover, multiple simultaneous video streams can easily congest a shared wireless network. It is therefore important for each user to perform *congestion control*, by adapting its encoding parameters to video and channel variations and striving for the best tradeoff between network congestion and video distortion.

We present a stochastic model for such systems, and formulate the congestion control problem in the context of controlled Markov decision chain [1]. The system evolution is expressed as the transitional probability of system *states*, including video complexity, channel condition and bottleneck queue size of each stream, given a permissable *action*, namely the choice of encoder quantization step size. Interaction among multiple users are captured by a probabilistic

model for traffic contention among nodes with non-empty queues. The expected performance of a congestion control policy can then be evaluated based on the equilibrium state distribution of the system. To minimize the expected network congestion and video distortion of all streams, a distributed congestion control policy is derived by performing *stochastic dynamic programming* (SDP) at each user. A greedy heuristic is also proposed to achieve most of the performance gain with reduced computational complexity.

After a brief survey of related work in Section 2, we explain the system model in Section 3 and discuss both the distributed SDP and the heuristic schemes for congestion control in Section 4. In Section 5, the performance of both schemes are compared against fixed-quality coding under simulated scenarios with various video and channel conditions.

2. RELATED WORK

Congestion control for live video coding over wireless networks can be traced back to rate control, where the encoder adjusts the quantization step size on a per-frame or per-macroblock basis to meet a target average rate. Many rate control algorithms have been proposed for different video coding standards, such as TM5 for MPEG-2 [2] and [3] for H.264/AVC. We, instead, seek to optimize the tradeoff between network congestion and video distortion as in [4], for multi-user video streaming over wireless networks.

Formulation of the rate control problem over a time-varying wireless channel has been studied in [5], where the authors also propose a optimal solution based on stochastic dynamic programming. In [6], the SDP approach is applied to live video streaming, where the sender prunes the pre-encoded video packets and the receiver adjusts the playout deadline in a jointly optimized fashion. The problem formulation in both works for the single-stream case has inspired our investigation into the multi-stream scenario, for which we further introduce a probabilistic model to capture the interaction among multiple streams due to traffic contention.

3. SYSTEM MODEL

Consider S simultaneous video streams over a wireless network. Transmission opportunities are slotted into durations of ΔT, and the frame arrival interval is $T_f = M\Delta T$. Each user sends a sequences of frames with complexity $X_{s,n}$ over
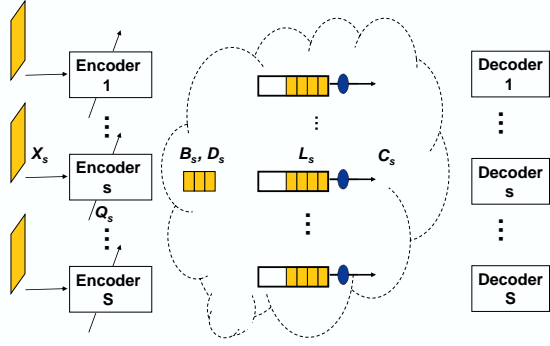
Fig. 1. System diagram for multi-user video streaming over wireless network. Each encoder performs congestion control by adjusting the quantization step size Q_s. End-to-end delay is dominated by the bottleneck backlog L_s.

a given route. The encoded frame size $B_{s,n}$ and distortion $D_{s,n}$ are controlled by the encoder quantization step size $Q_{s,n}$. The end-to-end delay $T_{s,n}$ for delivering each frame is assumed to be dominated by the bottleneck node on the path, with transmission rate $C_{s,t}$ and backlog size $L_{s,t}$ at each time slot t. The various system states are modeled as stochastic processes in the following subsections, and are illustrated in Fig. 1.

3.1. Video content variation

We characterize video content complexity $X_{s,n}$ of each frame by its mean absolute difference (MAD) per pixel with respect to the previous frame. Its evolution is modeled as a stationary first-order Gauss-Markov process with mean μ_x and variance $\sigma_x{}^2$:

$$X_{s,n+1} - \mu_x = a(X_{s,n} - \mu_x) + N_{s,n}, \tag{1}$$

where $N_{s,n} \sim \mathcal{N}(0, (1-a^2)\sigma_x{}^2)$.

Following observations in [7], the encoded frame size and distortion given a quantization step size Q can be approximated as:

$$B_{s,n} = f_B(X_{s,n}, Q) = \alpha_s \frac{X_{s,n}}{Q} \tag{2}$$

$$D_{s,n} = f_D(X_{s,n}, Q) = \beta_s Q. \tag{3}$$

The parameters α_s and β_s need to be fitted with empirical data from actual encodings for each video stream.

3.2. Wireless channel fluctuation

The wireless channel condition at each bottleneck node is modeled as a Markov process, specified by state transitional probabilities $P_{t+1|t}(\mathbf{C}|\mathbf{c})$ [1] and average channel rates in each state. Note that wireless nodes close to each other typically

---

[1] We denote random variables in capital letters and their previous-time-slot realizations in the conditional probabilities in lowercases. Random vectors in bold faces refer to the collection over all S streams.

experience similar channel conditions. The correlation coefficients depend on their physical locations and the processes need to be jointly modeled. For simplicity, we only consider the special case where all bottleneck nodes observe the *same* channel condition, and henceforth drop the s indices in the notation. The analysis and optimization, on the other hand, can extend to the more general case.

3.3. Bottleneck backlog

Given current queue size $L_{s,t}$ and observed channel rate C_t at bottleneck node, backlog in the next time slot $L_{s,t+1}$ can be predicted as follows:

$$L_{s,t+1} = \begin{cases} \max\{L_{s,t} + B'_{s,t} - C_t \Delta T, \ 0\} & w.p. \ \frac{1}{S'} \\ L_{s,t} + B'_{s,t} & w.p. \ 1 - \frac{1}{S'}. \end{cases} \tag{4}$$

In (4), $B'_{s,t}$ corresponds to the encoded frame size $B_{s,n}$ at times when a new frame arrives:

$$B'_{s,t} = \begin{cases} B_{s,n} = f_B(X_{s,n}, Q_{s,n}) & t = nM \\ 0 & mod(t, M) \neq 0, \end{cases} \tag{5}$$

and S' denotes the number of bottleneck nodes with non-empty backlogs competing with equal opportunities for wireless media access. This simple probabilistic model is motivated by the CSMA/CA media access control protocols used by wireless LAN cards following the IEEE 802.11 standard, where nodes with data for transmission compete for channel access following the same random backoff procedures.

3.4. System evolution

Combining the model components together, evolution of the entire system state space can be expressed as:

$$P_{t+1|t}(\mathbf{X}, C, \mathbf{L}|\mathbf{x}, c, \mathbf{l}, \mathbf{q}) = \prod_{s=1}^{S} (P_{t+1|t}(X_s|x_s) P_{t+1|t}(C|c) \\ P_{t+1|t}(\mathbf{L}|\mathbf{l}, c, \mathbf{x}, \mathbf{q}), \tag{6}$$

given current system states $\mathbf{x}$, c, $\mathbf{l}$ and action $\mathbf{q}$.

In (6), the decomposition of the joint state transitional probability assumes of statistical independence between video content of different streams and the wireless channel condition. The conditional probability for X_s can be calculated from (1) and the evolution of C is specified in Subsection 3.2. The last term $P_{t+1|t}(\mathbf{L})$ is computed according to (2), (4) and (5).

4. CONGESTION CONTROL POLICIES

Given current system state, the encoder quantization step size is chosen according to the control policy $\mathbf{Q} = f_Q(\mathbf{X}, \mathbf{C}, \mathbf{L})$, which can either be a deterministic mapping or a randomized decision. This in turn determines the state transitional probability by substituting $f_Q(.)$ for $\mathbf{Q}$ into (6). The equilibrium state distribution $\mathcal{P}$ can then be calculated as the Perron-Frobenius eigenvector of the state transition matrix [1].

We propose to optimize both metrics of distortion and network congestion in a lagrangian formulation:

$$\min_{f_Q(.)} \sum_s \mathbf{E}D_s + \lambda \sum_s \mathbf{E}B_s \mathbf{E}T_s, \qquad (7)$$

where network congestion in the second term is the sum of expected delay of each stream weighted by its own traffic rate. The expected video distortion $\mathbf{E}D_s$, rate $\mathbf{E}B_s$ and delay $\mathbf{E}T_s$ are calculated according to the equilibrium state distribution $\mathcal{P}$.

4.1. Distributed SDP

There remains the task of finding the optimal control policy for (7). Full search over the product space of all possible system states and candidate policies would be computationally prohibitive. Furthermore, knowledge of the system states such as video complexity and bottleneck queue size of other users may simply not be available in practice. Alternatively, one can approximate (7) by minimizing a finite-horizon cost function at each user:

$$\min_{q_{s,n}, n=1,\ldots N'} J = \frac{1}{N'} \sum_{n=1}^{N'} (d_{s,n} + \lambda b_{s,n} t_{s,n}), \qquad (8)$$

This can be solved by recursively minimizing the cost given local system state at each frame interval:

$$
\begin{aligned}
V_n(x_s, c_s, l_s) &= \min_{q'_n:n'=n\ldots N'} \sum_{n'=n}^{N'} d_{s,n'} + \lambda b_{s,n'} t_{s,n'} \\
&= \min_{q_n} (d_{s,n} + \lambda b_{s,n} t_{s,n} + \mathbf{E}V_{n+1}). \quad (9)
\end{aligned}
$$

Here, the estimated delay for frame n of Stream s is $t_{s,n} = (b_{s,n} + l_{s,n})S/c_n$, assuming equal channel access among all S streams on average. Note that in (9), the choice of the quantization step size is determined not only by the tradeoff between distortion and expected congestion for the current frame, but also influenced by the impact on the expected cost in the future.

4.2. Heuristic scheme

To further reduce computational complexity, one can use a greedy heuristic that ignores the impact on the future:

$$q_{s,n}^* = \arg\min_q (d_{s,n} + \lambda b_{s,n} t_{s,n}). \qquad (10)$$

In this case, the optimal $q_{s,n}^*$ only depends on observations of current frame complexity $x_{s,n}$, encoder RD performance α_s and β_s, current channel state c_n and bottleneck queue size $l_{s,n}$. It should be noted that the dependency of $q_{s,n}^*$ on $l_{s,n}$ is a merit of the chosen optimization objective, since it incorporates the impact of *past* choices into the *current* decision. An alternative lagrangian formulation involving only expected delay and distortion as in $D + \lambda T$, for instance, would have inappropriately ignored the influence of bottleneck queue size $l_{s,n}$ on the current choice of $q_{s,n}$.

5. SIMULATION RESULTS

The proposed system is simulated under various scenarios combining different video and channel characteristics. The wireless channel fluctuates between a good and a fading state, each with a random throughput following normal distribution $\mathcal{N}(C_G, \sigma_G^2)$ or $\mathcal{N}(C_F, \sigma_F^2)$. Average state duration is 1.3 seconds for both. Parameters for the frame complexity process and for the encoder RD performances are fitted with actual data from encoding the *Foreman* and *Mother and Daughter* CIF sequences with the H.264/AVC codec [8]. We evaluate the congestion-distortion tradeoff of both the distributed SDP scheme and the greedy heuristic by varying the value of λ in the optimization objectives. As a basis for comparison, results from a fix-quality scheme are also included, where the videos are encoded with fixed quantization step sizes without congestion control.

Figure 2 compares the delay-quality tradeoff of the three schemes for streaming only *Foreman* or *Mother and Daughter* over the wireless channel. The proposed heuristic and SDP schemes outperform the fix-quality scheme by 0.5-1.0 dB in PSNR in terms of average video quality given same average delay. Additional gain up to 1.3 dB is achieved by the SDP scheme over the heuristic for *Foreman*, and up to 0.8 dB for *Mother and Daughter*. The improvement using SDP is more pronounced in the former case, where both video and channel variations are higher.
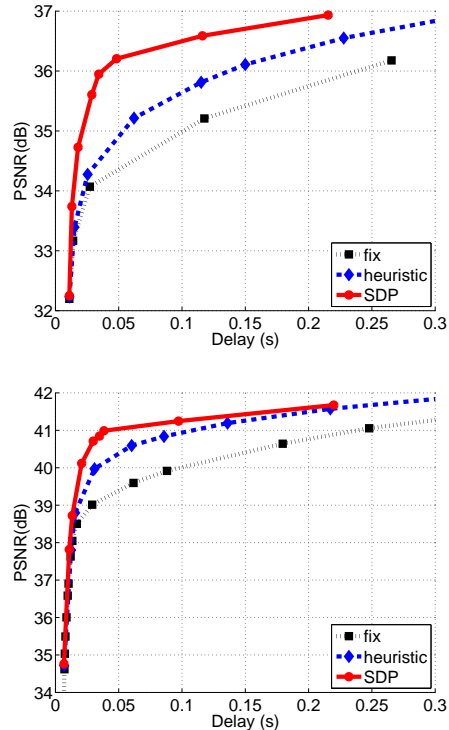

Fig. 2. Delay-PSNR tradeoff when transmitting only *Foreman* (top) or *Mother and Daughter* (bottom), encoded at 30 fps, over a wireless channel. For *Foreman*, $C_G = 900$ kbps, $C_F = 300$ kbps, $\sigma_G = \sigma_F = 50$ kbps, whereas for the later, $C_G = 600$ kbps, $C_F = 200$ kbps, $\sigma_G = \sigma_F = 20$Kbps.
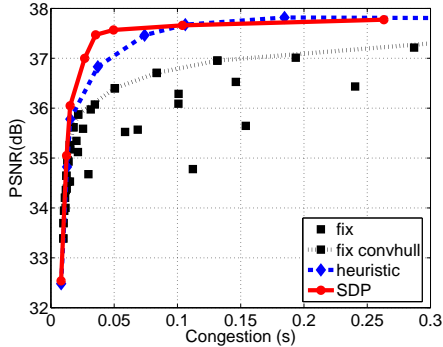
Fig. 3. Tradeoff between overall video quality and network congestion with two streams over the network. $C_G = 900$ kbps, $C_F = 500$ kbps, $\sigma_G = \sigma_F = 20$ kbps.

Fig. 4. Delay-PSNR tradeoff experienced by each of the users.The upper left set of the curves correspond to *Mother and Daughter*, and the lower ones are for *Foreman*.

When both video streams are transmitted simultaneously over a common network, the relative performance of the three schemes are shown in Fig. 3. The PSNR values correspond to the average video distortion of both streams, and the overall network congestion is measured as average delay weighted by traffic rate of each stream. For the fix-quality scheme, the different data points correspond to different pairs of quantization step sizes used for the two streams. The best possible performance is represented by the convex hull plotted in dotted line. By adjusting quantization step sizes according to current observations of video complexity, channel rate and backlog size, the heuristic scheme achieves superior video quality by up to 0.8 dB over fix-quality coding. Additional improvement up to 0.5 dB is achieved by the distributed SDP scheme, within a very limited range of network congestion. It is also observed in the experiments that the SDP scheme tends to be more conservative. Given the same λ, SDP yields a lower cost of (7) than the heuristic, with slightly inferior video quality rewarded by lower average delay.

Figure 4 shows the tradeoff between video quality and delay experienced by each stream, while both try to access the common wireless media. For the fix-quality scheme, only results using equal quantization step sizes in both streams are displayed. Due to the higher video complexity, the *Foreman* sequence experiences higher average delay and lower video quality, when competing over the same wireless media with the *Mother and Daughter* sequence with slower motion. By congestion control, the proposed SDP and heuristic schemes can improve the delay-quality tradeoff at *both* streams simultaneously.

6. CONCLUSIONS

We present a stochastic system model for analyzing multi-user congestion control schemes for live video coding and streaming over wireless networks. The model incorporates the effects of video and channel variation, as well as multi-stream traffic contention, into the evaluation of different congestion control policies. A distributed scheme based on stochastic dynamic programming at each user and a greedy heuristic are proposed to minimize the tradeoff between
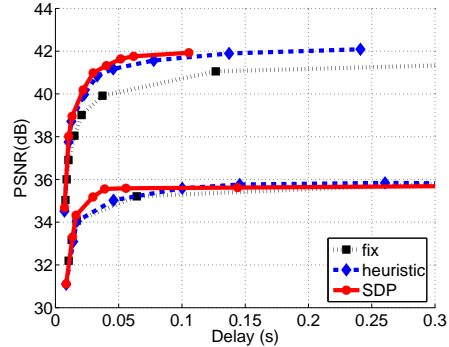
average video distortion and network congestion. Simulation results under various video and channel conditions show gains ranging 0.5-1.3 dB in average video quality achieved by the distributed SDP and heuristic schemes over fix-quality coding.

7. REFERENCES

[1] D. Bertsekas, *Dynamic Programming and Optimal Control, Vol I.*, Athena Scientific, Belmont, Massachusetts, USA, 2000.

[2] ISO/IEC/JTC1SC29WG11, *MPEG 2 Test Model 5, Rev. 2, Section 10: Rate Control and Quantization Optimization*, Apr. 2003.

[3] S. Ma, W. Gao, P. Gao, and Y. Lu, "Rate control for advance video coding (AVC) standard," *Proc. International Symposium on Circuits and Systems, Bankok, Thailand*, vol. 2, pp. 892–895, May 2003.

[4] X. Zhu, E. Setton, and B. Girod, "Content-adaptive coding and delay-aware rate control for a multi-camera wireless surveillance network," *Proc. IEEE International Workshop on Multimedia Signal Processing, Shanghai, China*, Oct. 2005.

[5] J. Cabrera and A. Ortega, "Stochastic rate control of video coders for wireless channels," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 496–510, June 2002.

[6] Y. Li, A. Markopoulou, J. Apostolopoulos, and N. Bambos, "Packet transmission and content-dependent playout for video streaming over wireless networks," *Proc. IEEE International Workshop on Multimedia Signal Processing, Shanghai, China*, Oct. 2005.

[7] S. Takamura and N. Kobayashi, "MPEG-2 one-pass variable bit rate control algorithm and its LSI implementation," *Proc. IEEE International Conference on Image Processing, Thessaloniki, Greece*, vol. 1, pp. 942–945, Oct. 2001.

[8] ITU-T and ISO/IEC JTC 1, *Advanced Video Coding for Generic Audiovisual services, ITU-T Recommendation H.264 - ISO/IEC 14496-10(AVC)*, 2003.