LIMITED RETRANSMISSION OF REAL-TIME LAYERED MULTIMEDIA

Matthew Podolsky[1], Martin Vetterli[1,2], and Steven McCanne[1]

[1]Dept. of Elect. Engin. & Comp. Sci.
Univ. of California, Berkeley
Berkeley, CA, USA

[2]Laboratoire de Comm. Audiovisuelles
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

Abstract -   In contrast to multimedia applications that involve human-to-human communication, streaming media over the Internet enjoys relaxed delay constraints. Thus, streaming media servers are at liberty to retransmit missing packets to avoid unnecessary signal corruption. While state-of-the-art media servers employ such strategies, no work to date has proposed an optimal strategy for delay-constrained retransmissions of streaming media. In this paper, we propose a framework for streaming media retransmission based on layered media representations and explore the performance advantage of integrating layered signal structure into the retransmission strategy. In our approach, the source must choose between transmitting an older layer that expires sooner and a newer layer that expires later but is more important. To arrive at the proper mix of these two extreme strategies, we derive an optimal strategy for transmitting layered data over a binary erasure channel with instantaneous feedback. To provide a quantitative performance comparison of different transmission policies, we conduct a Markov-chain analysis, which shows that the best transmission policy is time-invariant and thus does not change as the layers approach their expiration times.

INTRODUCTION

A common class of traffic on the Internet is so-called "streaming media," where real-time signals like audio and video are delivered from a server somewhere in the network to a human user that interactively views the material. Unlike human-to-human communication, which requires relatively tight and consistent end-to-end delays for good interactive performance, server-to-human communication can afford a certain level of artificial delay. As a result, streaming media applications often have sufficient time to recover from lost packets through retransmission and thereby avoid unnecessary degradation in reconstructed signal quality. While this form of delay-constrained automatic repeat/request (ARQ) has been exploited in research protocols like STORM [5] and MESH [3] and in commercial products like RealNetworks clients and servers, there is no existing solution to the problem of how a sender optimally chooses what available data to retransmit when the receiver indicates loss.

In this paper, we propose a framework to solve this optimization problem. In our scheme, the sender represents its signal in a layered format and at any given time transmits the "most important" information conditioned on receiver feedback and constrained by the available bit rate (which is either pre-configured or inferred from a companion congestion control algorithm [4]). If the packet loss rate is sufficiently

low that the effective channel capacity is larger than the source rate, then simple ARQ would more or less suffice and our problem would be solved. Hence, we instead assume and study the case that packet loss is high and the source rate of the highest-quality version of the signal exceeds the available capacity. We also assume that the layered stream is easily segmented into equal "time slices" or frames (*e.g.*, 20 ms of an audio signal), each of which is carried in a packet. Finally, we assume that hierarchy is imposed on layers and that distortion is reduced as more layers are used to reconstruct the signal.

To optimize the delivery of the signal, the source finds the best transmission policy for the layered data. This policy controls what information the source transmits (or retransmits) at each opportunity as dictated by the capacity constraint. We define the optimal policy π^* as the transmission policy that minimizes the distortion (*i.e.*, maximizes signal quality) of the signal reconstructed from the successfully received layers. To solve this problem, the source must carefully decide between two seemingly conflicting goals: either to (re)transmit older, less important layers that expire soon and/or to send newer and more important layers that expire later.

To illustrate this tradeoff, suppose we have a multimedia signal that is encoded into N layers. The signal is further segmented in time into frames, and each layer of a frame is sent in a separate packet. We define one time unit to be the length of time it takes to transmit one layer of one frame. For convenience we will use seconds to denote unit time. Let one frame length be T seconds, so that every Ts the N layers of a frame "arrive" at the source for transmission. We require $T \geq N$ so that we have at least one chance to transmit each layer. Let the lifetime of each frame be L, so that each layer of a frame must successfully arrive at the receiver within Ls after it was produced. Note that all layers of the same frame have the same expiration time. To capture the effect of losses of layers by the network (*e.g.*, packet drops), we use a binary erasure channel (BEC) with feedback as our transmission channel. For this initial study we ignore delay in both directions. In order to have a chance to transmit all layers and permit some number of transmissions, we require $L > N$. Figure 1 illustrates this transmission system.

As a specific example let $N = 2$, so that there is only a high priority layer (HPL) and a low priority layer (LPL). The HPL has a greater decoding benefit (*e.g.*, a greater distortion reduction) than the LPL. Let $T = 3$ and $L = 5$. Every 3s a frame and its two associated layers are produced, and these live for 5s. Since $L > T$
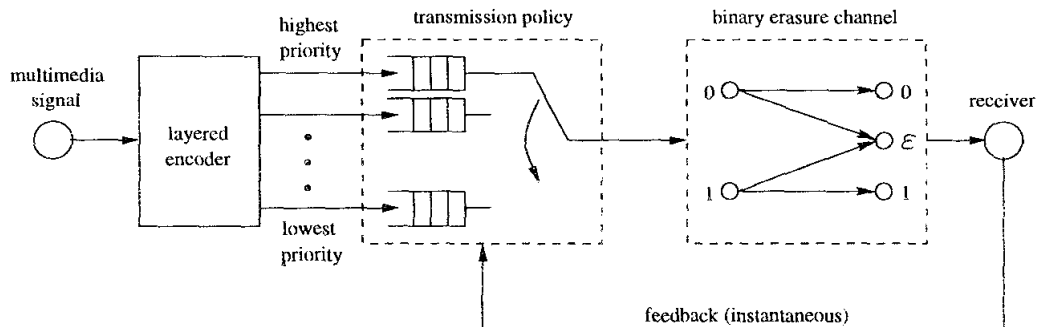

Figure 1: System diagram of layered transmission over a BEC with feedback.

there is overlap between the lifetimes of consecutive frames. Let A be the first frame of the signal, and A_H and A_L be A's HPL and LPL, respectively. When A is produced at time 0, the transmission policy is obvious: A_H should be transmitted before A_L, because A_H has more decoding benefit and both layers expire at the same time. Likewise, if the first transmission attempt of A_H at time 0 is erased, then at time 1 A_H should be retransmitted, in lieu of transmitting A_L. Assume that retransmissions of A_H at time 1 and 2 are also erased. At time 3 frame B is also produced, so now there are 4 layers awaiting transmission: A_H and A_L (expiring at time 5), and B_H and B_L (expiring at time 8). Assuming the decoding benefit of HPL's or LPL's is constant across frames, it is easily shown that A_H is the optimal choice for transmission at time 3. This is because no other frame has a shorter time to extinction (STE), and no other frame has a greater decoding benefit. If the retransmission of A_H succeeds at time 3, then at time 4 the transmission choice is no longer clear. A choice must be made between transmitting A_L and B_H (B_L is not a candidate for transmission, since A_L provides the same benefit but has a shorter STE). A_L expires sooner than B_H, so it might be optimal to transmit it first. But A_L also provides less benefit than B_H, and transmitting A_L reduces the number of opportunities to get B_H across. The best transmission policy is not obvious, and we consider ways of finding it in the next section.

Before continuing, we note that Bhattacharya and Ephremides examined similar queuing problems in [1] and [2]. However, their work differs in that the arrival times of packets (*i.e.*, production times of layers) are random and geometrically distributed; in our case we have known deterministic and periodic arrival times.

ANALYSIS

In order to determine the average distortion obtained with a particular transmission policy π, we must determine the probabilities of successfully transmitting the i highest priority layers of a frame by the time the frame expires, for all $i \in \{0, 1, ..., N\}$. For a BEC with *i.i.d.* erasures, the layered transmission discussed in the previous section can be described by a cyclostationary process with period T. Since every Ts one frame expires and one frame is produced, the problem can be analyzed by keeping track of frames by their relative arrival times during a cycle of length T, rather than by their absolute arrival times (*e.g.*, a frame would be denoted as the oldest one still "alive" during a cycle, rather than the frame that was produced at some specific time). Let ϕ denote the phase within a cycle of T; $0 \leq \phi < T$. We choose $\phi = 0$ to be the arrival time of the new frame during the cycle. The maximum number of frames K alive at any time is given by $K = \lceil L/T \rceil$.

If the lifetime L is not an exact multiple of T, there will be only $K-1$ frames alive during the last $KT - L$ seconds of a cycle; the oldest frame expires at time $\phi = L - (K-1)T$. Let frame 1 be the oldest frame alive within a cycle, frame 2 the second oldest, *etc.*, so frame K is the newest one, produced at $\phi = 0$. Let the K-tuple $\mathbf{n} = (n^{(1)}, n^{(2)}, ..., n^{(K)})$ be the number of layers that have been successfully transmitted for each frame so far; $0 \leq n^{(i)} \leq N$. The current state of the transmission process

at any time t can then be described by the pair $s = (\phi, \mathbf{n})$. Because $\mathbf{n}$ describes what layers have been successfully transmitted to date and ϕ describes how close to their expiration they are, the optimal transmission choice only depends on s. With the above definition of s the process forms a finite-state Markov chain.

Transitions from a state $s_\phi = (\phi, \mathbf{n}_\phi)$ to $s_{\phi+1} = (\phi + 1, \mathbf{n}_{\phi+1})$ can be described by a sparse $M \times M$ state transition matrix P_ϕ, where $M = (N+1)^K$ is the maximum number of possible values of $\mathbf{n}$.[1] The only transitions with non-zero probabilities are those in which $\mathbf{n}_{\phi+1}$ differs from $\mathbf{n}_\phi$ in at most one element i. In other words, during one time unit either the transmitted layer is erased ($\mathbf{n}_{\phi+1} = \mathbf{n}_\phi$) or the highest priority layer remaining of frame i is successfully transmitted ($n^{(i)}_{\phi+1} = n^{(i)}_\phi + 1$). The state transition matrix will depend on the channel erasure probability ε and the decision policy π. A state transition matrix can be defined for each of the T possible values of ϕ. Additionally, at time $\phi = L - (K-1)T$ the oldest frame expires, so matrices $P_{L-(K-1)}$ through P_{T-1} have no transitions in which $n^{(1)}$ changes. Finally, a simple matrix P_a is needed to to reflect the relative "aging" of the frames at the end of the cycle; after time $\phi = T - 1$ the state changes from $(n^{(1)}, n^{(2)}, \ldots, n^{(K)})$ to $(n^{(2)}, n^{(3)}, \ldots, n^{(K)}, 0)$. With these definitions, the transition probabilities of moving from a state $\mathbf{n}_t$ at time t to $\mathbf{n}_{t+T}$ at time $t + T$ given by the $M \times M$ matrix

$$P_\phi P_{\phi+1} \cdots P_{T-1} P_a P_0 P_1 \cdots P_{\phi-1}, \tag{1}$$

where $\phi = (t \bmod T)$ and the matrices are defined for row vector states. The process $\{\mathbf{n}_\phi, \mathbf{n}_{\phi+T}, \mathbf{n}_{\phi+2T}, \ldots\}$ is stationary and hence a stationary distribution for $\mathbf{n}$ can be found analyzing the matrix of Equation 1. By solving for the stationary distribution η when $\phi = L - (K - 1)T$, the probability $\nu^{(i)}$ of transmitting the i most important layers of a frame is given by

$$\nu^{(i)} = \sum_{n_2=0}^{N} \sum_{n_3=0}^{N} \cdots \sum_{n_K=0}^{N} \eta^{(i,n_2,n_3,\ldots,n_K)} \tag{2}$$

Finally, given a rate-distortion function $D(R)$ such that $D(i)$ is the distortion incurred in reconstructing a frame from its i highest priority layers, $0 \le i \le N$, we can compute the average distortion per frame for a transmission policy π by

$$D_\pi = \sum_{i=0}^{N} \nu^{(i)} D(i). \tag{3}$$

We initially analyzed the problem for the case of two layers ($N = 2$) and a maximum overlap of two frames ($K = 2$). For this case there are 9 possible values of $\mathbf{n} = (n^{(1)}, n^{(2)})$. Recall from our earlier example that the only state requiring a transmission choice is $\mathbf{n} = (1, 0)$; the choice is between the older frame's LPL and the newer frame's HPL. Since this choice can vary as a function of the phase within a cycle, there are 2^{L-T} possible different transmission policies.

---

[1] The number of *achievable* states may actually be more limited than this.

For fixed values of L and T, the optimal policy will depend not only on the erasure rate ε, but also on the rate-distortion function $D(R)$. Since D_π (the distortion associated with a policy given by Equation 3) is a linear function of $D(R)$, we can arbitrarily translate or positively scale $D(R)$ without changing which policy results in minimum distortion. We thus arbitrarily choose $D(0) = 1$ and $D(2) = 0$, and let $D(1) = d$, where $0 < d \leq 0.5$. If d is close to 0.5 the HPL and LPL are of near equal importance; if d is near 0 the HPL has much more associated benefit than the LPL.
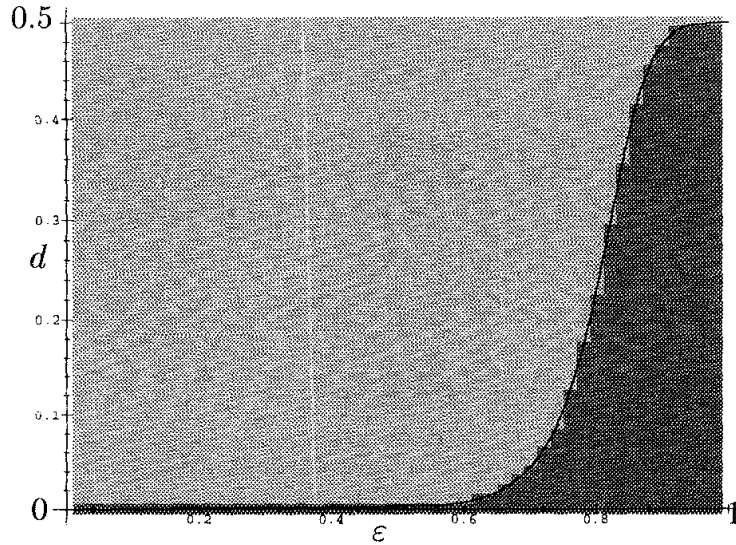

Figure 2: Optimal decision policy as a function of ε and d.

Figure 2 shows the optimal transmission policy for the case of $L = 18$ and $T = 10$ as a function of ε and d. Our results indicate that of the 256 possible transmission policies, only the two extremes are ever optimal: either always choosing to transmit the older LPL over the newer HPL, or always choosing to transmit the newer HPL instead. In other words the choice for this state does not vary as the frames get closer to expiring—the best policy is time (phase) invariant. The graph of Figure 2 was produced by sampling the ε-d plane, evaluating the distortion of all policies at the sampled points, and shading a rectangle about each point according to the optimal choice. For the lighter area on the left, the optimal policy is transmitting the older LPL; for the darker area to the right, it is the newer HPL. The smooth curve illustrating the border between these two optimal policies was obtained by setting these policies' distortion functions equal and solving for ε as d was varied.

The crossover in Figure 2 between the two policies makes intuitive sense. If ε is small then it is best to transmit the old LPL, since it is unlikely that it will take many attempts to successfully transmit the new HPL. Conversely, if ε is high it is best to choose the newer HPL, since it may take a large number of tries. Regarding d, the smaller it is the less the LPL matters, so the new HPL should be favored. Conversely, if d is high then the old LPL is of near equal importance to the new HPL, and so the old LPL should be favored since it expires sooner.

APPLICATIONS AND FUTURE WORK

We have observed results similar to Figure 2 in the two layer case ($N = 2$) with other values of L and T for which $K \geq 2$; the best policy is always one of two time-invariant ones. These results can be applied to the development of a protocol which adapts the transmission policy as network conditions vary. Since d and T are fixed for a given multimedia encoding, the value of ε at which the optimal policy changes can be pre-computed for various values of L. A protocol could adaptively estimate the data lifetime L (which can be defined to account for variations in the network delay) to determine the crossover value of ε, and compare this to the current estimate of ε to choose the best policy.

Another area for future work is to study the transmission problem under the more general cases of multiple layers ($N > 2$) and delay in the feedback. The "curse of dimensionality" is inherent in both of these, since increasing these factors causes an exponential increase in the size of the state space. Hence, another avenue of future investigation is examining other ways of analyzing the problem, such as using approximations or iterative algorithms for finding the optimal transmission policy.

CONCLUSION

We have examined optimal transmission policies for real-time layered multimedia data, and we looked at the specific case of using limited retransmission over a BEC with instantaneous feedback. We illustrated that there are non-trivial transmission choices to be made between older, lower priority layers and newer, higher priority layers. We presented a Markov chain analysis for evaluating the efficacy of a specific transmission policy. The results of our analysis indicate that, for fixed network conditions, the best transmission policy is time-invariant and does not change as layers approach their expiration. We then discussed how these results could be used to develop a transmission protocol which can adapt to changing network conditions.

References

[1] P. Bhattacharya and A. Ephremides. Optimal scheduling with strict deadlines. *IEEE Trans. Automat. Control*, 34(7):721–728, July 1989.

[2] P. Bhattacharya and A. Ephremides. Optimal allocation of a server between two queues with due times. *IEEE Trans. Automat. Control*, 36(12):1417–1423, December 1991.

[3] M. Lucas, B. Dempsey, and A. Weaver. MESH: distributed error recovery for multimedia streams in wide-area multicast networks. In *Proc. IEEE Int. Conf. on Commun.*, volume 2, pages 1127–32, Montreal, Que., June 1997.

[4] R. Rejaie, M. Handley, and D. Estrin. RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the internet. *Submitted to ICNP 98*.

[5] X. Xu, A. Myers, H. Zhang, and R. Yavatkar. Resilient multicast support for continuous-media applications. In *Proc. NOSSDAV*, pages 183–194, St. Louis, MO, May 1997.