Queue Length Analysis for Multicast: Limits of Performance and Achievable Queue Length with Random Linear Coding

Randy Cogill
Dept. of Systems and Information Engineering
University of Virginia
Charlottesville, Virginia 22904
Email: rcogill@virginia.edu

Brooke Shrader
Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, Massachusetts 02420
Email: brooke.shrader@ll.mit.edu

Abstract

In this work we analyze the average queue backlog at a source node serving a single multicast flow consisting of M destination nodes. In the model we consider, the channel between the source node and each receiver is an independent identically distributed packet erasure channel. We first develop a lower bound on the average queue backlog achievable by any transmission strategy; our bound indicates that the queue size must scale as at least $\Omega(\ln(M))$. We then analyze the queue backlog for a strategy in which random linear coding is performed over groups of packets in the queue; this strategy is an instance of the random linear network coding strategy introduced in [9]. We develop an upper bound on the average queue backlog for the packet-coding strategy to show that the queue size for this strategy scales as $O(\ln(M))$. Our results demonstrate that in terms of the queue backlog, the packet coding strategy is order-optimal with respect to the number of receivers.

I. Introduction

In this paper we analyze the size of the queue backlog for a scheme in which packets are communicated from source to destination nodes by sending algebraic combinations of a batch of packets. This random linear coding scheme was introduced in [9] and shown to provide improved throughput over uncoded transmission for multicast traffic. While random linear coding has been shown to increase multicast throughput in many works, the performance in terms of queue backlog is not fully understood. In this work we provide an upper bound on the average queue backlog for random linear coding and show that as the number of receivers scales up, random linear coding performs as well as any other multicast transmission strategy.

A number of recent works deal with queueing analysis for random linear coding schemes. The work in [10] considers finite-capacity buffers and presents numerical results on the queue blocking probability and delay. A bulk-service queueing model for random linear coding is developed in [13] and numerical results on queueing delay are also presented. The work in [16] proposes a new packet acknowledgment strategy for random linear coding based on acknowledging degrees of freedom; a queue backlog analysis for the policy is provided and the queue size is shown to grow more slowly with load factor than a baseline acknowledgment strategy. The work in [6] provides analytical bounds on the completion time and stable throughput for random linear coding across multiple multicast sessions; the results indicate that although coding across sessions requires unintended recipients to decode packets, the coding strategy can provide larger throughput than uncoded transmission. Recent work in [1] shows that while significant delay penalties are incurred for synchronized coding across flows, asynchronous coding of packets across flows can reduce the queueing delay. Our work differs from previous work in that we present analytical bounds on the average queue backlog and show that random linear coding is order-optimal with respect to the number of receivers.

II. The Problem

In this paper we consider a problem involving multicast transmission of data packets from a single transmitter. We model the transmitter as a single queue, where packets arrive in this queue according to a Bernoulli process with

This work was sponsored by the Department of the Air Force under Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

rate λ. In each time slot, a packet is transmitted to all of the M receivers. Each receiver receives the transmitted packet independently with probability q, independent of past receptions. This system is depicted in Figure 1. Our goal is to devise a transmission scheme that minimizes the expected number of packets in the queue in the steady state.

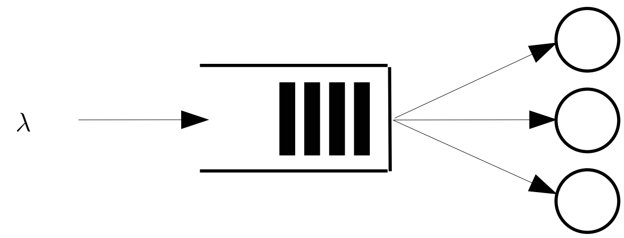

Fig. 1. Multicast queueing system.

The problem of minimizing expected queue length is deceptively difficult. To see why this is the case, first consider the simple scheme that retransmits the head of line packet until it has been received by all M receivers. As shown in [6], the expected number of time slots that a packet stays at the head of the queue is greater than or equal to

$$\frac{\ln(M+1)+0.3}{-\ln(1-q)}. \tag{1}$$

Therefore, this scheme cannot stabilize arrival rates satisfying

$$\lambda > \frac{-\ln(1-q)}{\ln(M+1)+0.3}. \tag{2}$$

This means that the expected queue length becomes arbitrarily large for arrival rates approaching some value less than or equal to the right-hand side of (2). In other words, for any λ, one can choose M sufficiently large so that the expected queue length under the retransmission strategy is arbitrarily far from optimal.

Another approach is to view the problem of minimizing expected queue length as a purely control theoretic problem. Rather than simply transmitting the head of line packet, in each time slot any packet in the queue can be transmitted. When the state of each channel connecting the transmitter to each receiver is known before transmitting a packet, a controller would use this information, together with the reception history of each packet in the queue to decide which packet to send next. However the complexity and information requirements of such a scheme are clearly very high. On the other hand, without channel state information, throughput of such a scheme is no better than the throughput of the simple retransmission scheme. To see why this is true, if the system is ignorant of the channel states before transmitting a packet, the expected number of times a packet must be transmitted before successful reception still must be greater than or equal to (1), even if transmissions are attempted out of order.

It is well known that the queue can be stabilized for all arrival rates satisfying $\lambda < q$ using simple random linear coding schemes (see [7], for example). These schemes generally operate by collecting large blocks of packets in the queue, then transmitting encoded packets formed from this block until all receivers can decode all packets in the block. To stabilize rates λ approaching q, arbitrarily large blocks of packets must be formed. The block coding operation must create a backlog that grows with the size of the block. So, it seems that schemes that code over fixed-length blocks of packets are not well suited for the problem of minimizing expected backlog.

III. Main Results

In the previous section, we argued that the problem of minimizing expected steady state queue length for multicast is surprisingly complex. In this paper we establish two results for this problem:

- We show that the expected steady-state queue length of any strategy must satisfy

$$\liminf_{t\to\infty} \mathbf{E}[Q(t)] \geq \frac{3}{4}\left(\frac{\lambda\ln(M)-1}{-\ln(1-q)}\right),$$

where $Q(t)$ is the number of packets in the queue at time t. This is true even for strategies that exploit channel state information.

- We show the queue length process of a simple random linear coding scheme (at packet departure times) satisfies
$$\limsup_{n\to\infty} \frac{1}{n+1}\sum_{k=0}^{n} \mathbf{E}[Q(t_k)] \;\le\; 3.12\ln(M)\left(\frac{\lambda}{q-\lambda}\right)^2 + \left(7.35\sqrt{\ln(M)}+5.22\right)\frac{\lambda}{q-\lambda}.$$
Here, t_n is the time at which the n-th block of packets departs the system.

So, if the queue length process at departure times corresponds to the true queue length process, the expected queue length of the random linear coding strategy is order optimal with respect to the number of receivers.

A LOWER BOUND ON ACHIEVABLE BACKLOG

Here we will present a lower bound on the minimum achievable steady state expected queue length for multicast. Specifically, we show that backlog must scale at least logarithmically with M, the number of receivers. In the next section we show that the backlog of the code over queue contents strategy scales logarithmically with M. So, in this section we show that coding over the queue contents is order-optimal with respect to the number of receivers. The theorem that will be proved in this section is the following:

Theorem 1: Under any strategy
$$\liminf_{t\to\infty} \mathbf{E}[Q(t)] \ge \frac{3}{4}\left(\frac{\lambda\ln(M)-1}{-\ln(1-q)}\right).$$

The proof of this theorem is at the end of this section, after several supporting lemmas are established.

The lower bound presented in this section makes very few assumptions on the strategy used. We will start by stating these assumptions in words, then give a more precise condition that must be satisfied by all policies.

To make the motivation for our assumptions clear, first consider a system composed of a single transmitter and a single receiver. Packets arrive at the transmitter according to a Bernoulli process with rate λ. In each time slot the transmitter is connected to the receiver with probability q, independent of past connections.

Various strategies could be used to transmit packets to the receiver. Regardless of the strategy used, we always assume that the system possesses several properties:

(1) At any time, the total number of packets that have been removed from the queue does not exceed the total number of packets that have entered the queue.

(2) At any time, the total number of packets that have been removed from the queue does not exceed the total number of time slots where the transmitter has been connected to the receiver. So, even if coding is applied, we do not consider schemes that compress packets. We can only transmit m packets to the receiver if the transmitter and receiver have been connected in at least m time slots.

(3) Connections cannot be used to transmit future arrivals. In other words, the strategy must be *causal*. We cannot send information about a packet that has not yet arrived in the queue.

(4) The queue starts out empty. Since we are concerned with steady-state queue length, this is without loss of generality.

To make these conditions precise, we'll introduce some notation. Let $Q(t)$ be the length of the queue at time t. Let $A(t)$ be the random variable with $A(t) = 1$ if there is an arrival at time slot t and $A(t) = 0$ otherwise. Let $S(t)$ be the random variable with $S(t) = 1$ if the receiver is connected at time slot t and $S(t) = 0$ otherwise. Finally, let $D(t)$ be the number of departures from the queue at time t.

In terms of A, D, and S, properties (1), (2), and (3) are captured by the following condition. For all $\tau \le t$, we require that D satisfies
$$\sum_{k=0}^{t-1} D(k) \le \sum_{k=0}^{\tau-1} A(k) + \sum_{k=\tau}^{t-1} S(k). \tag{3}$$

Properties (1) and (2) imply this condition for $\tau = t$ and $\tau = 0$, respectively. Showing that property (3) implies this condition for all other τ is a little less obvious. To show this, consider the set of packets that arrived in the system during the interval $[0, t-1]$. We can partition this set into two sets $\mathcal{A}_1$ and $\mathcal{A}_2$, where $\mathcal{A}_1$ is the set of packets arriving in the interval $[0, \tau - 1]$ and $\mathcal{A}_2$ is the set of packets arriving in the interval $[\tau, t-1]$. The total number of departures in the interval $[0, t-1]$ equals the total number of packets from $\mathcal{A}_1$ that depart the system in the interval $[0, t-1]$ plus the total number of packets from $\mathcal{A}_2$ that depart the system in the interval $[0, t-1]$. The total number of departures from $\mathcal{A}_1$ is, by property (1), less than or equal to the total number of packets in $\mathcal{A}_1$, which is

$$\sum_{k=0}^{\tau-1} A(k).$$

Since all packets in $\mathcal{A}_2$ arrived on or after τ, property (3) implies that no connection prior to τ can be used to serve a packet in $\mathcal{A}_2$. This, together with property (2), implies that the total number of departures from $\mathcal{A}_2$ is less than or equal to the total number of connections in the interval $[\tau, t-1]$, which is

$$\sum_{k=\tau}^{t-1} D(k).$$

Finally, summing these upper bounds on the number of departures from $\mathcal{A}_1$ and $\mathcal{A}_2$ gives (3).

Using property (3), we have the following lemma. This Lemma is very similar to a well known result that obtains a queue length process by applying a reflection mapping to a netput process (see Proposition 6.3 in [2], for example).

Lemma 1: If $Q(0) = 0$, the queue length at time t satisfies

$$Q(t) \geq \max_{t \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t-1} \big(A(k) - S(k)\big) \right\}$$

for any service policy with departures satisfying (3). Moreover, this inequality is tight when packets depart the queue as services occur.

Proof: At any time t,

$$Q(t) = \sum_{k=0}^{t-1} \big(A(k) - D(k)\big)$$

Since

$$\sum_{k=0}^{t-1} D(k) \leq \sum_{k=0}^{\tau-1} A(k) + \sum_{k=\tau}^{t-1} S(k)$$

for all $\tau \in [0, t]$, the queue length satisfies

$$\begin{aligned} Q(t) &\geq \sum_{k=0}^{t-1} A(k) - \sum_{k=0}^{\tau-1} A(k) - \sum_{k=\tau}^{t-1} S(k) \\ &= \sum_{k=\tau}^{t-1} \big(A(k) - S(k)\big) \end{aligned}$$

for all $\tau \in [0, t]$. Since this holds for all τ, clearly

$$Q(t) \geq \max_{t \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t-1} \big(A(k) - S_j(k)\big) \right\}$$

The fact that the inequality is tight if packets depart the queue as services occur can be shown by induction. Since we start with $Q(0) = 0$,

$$\begin{aligned} Q(1) &= \max\{0, A(0) - S(0)\} \\ &= \max_{1 \geq \tau} \left\{ \sum_{k=\tau}^{0} \big(A(k) - S(k)\big) \right\}. \end{aligned}$$

Now suppose $Q(t)$ satisfies

$$Q(t) = \max_{t \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t-1} \big(A(k) - S(k)\big) \right\}.$$

Then the queue length at time $t+1$ is

$$\begin{aligned} Q(t+1) &= \max\{0, Q(t) + A(t) - S(t)\} \\ &= \max \left\{ 0,\ \max_{t \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t-1} \big(A(k) - S(k)\big) \right\} + A(t) - S(t) \right\} \\ &= \max_{t+1 \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t} \big(A(k) - S(k)\big) \right\}. \end{aligned}$$

■

We will now extend these results to the multicast case. Let $S_j(t)$ be the random variable with $S_j(t) = 1$ if the transmitter is connected to the j-th receiver at time slot t, and $S_j(t) = 0$ otherwise. Let $R_j(t)$ be the number of packets received by receiver j at time t. Each $R_j(t)$ must satisfy the property ((3)). That is, for all $\tau \leq t$, the total number of packets received by receiver j up to time $t-1$ satisfies

$$\sum_{k=0}^{t-1} R_j(k) \leq \sum_{k=0}^{\tau-1} A(k) + \sum_{k=\tau}^{t-1} S_j(k).$$

Also, since packets do not leave the queue until they have been received by all receivers, the total number of departures in the interval $[0, t-1]$ satisfies

$$\sum_{k=0}^{t-1} D(k) = \min_j \left\{ \sum_{k=0}^{t-1} R_j(k) \right\}$$

Under this condition alone, we can establish a lower bound on the achievable expected queue backlog. The next lemma gives a lower bound in terms of the expected value of the minimum of binomial random variables.

Lemma 2: Let $Q(t)$ be the queue length of the multicast system at time t. Under any policy,

$$\liminf_{t \to \infty} \mathbf{E}[Q(t)] \geq \sup_{n \geq 0} \{\lambda n - f_M(n)\},$$

where $f_M(n)$ is the expected minimum of M independent binomial random variables, each with parameters (q, n).

Proof: The total number of packets in the queue at time t is

$$Q(t) = \sum_{k=0}^{t-1} A(k) - \sum_{k=0}^{t-1} D(k).$$

Note that for all $t \geq \tau \geq 0$,

$$\begin{aligned}\sum_{k=0}^{t-1} D(k) &= \min_j \left\{ \sum_{k=0}^{t-1} R_j(k) \right\} \\ &\leq \sum_{k=0}^{\tau-1} A(k) + \min_j \left\{ \sum_{k=\tau}^{t-1} S_j(k) \right\}.\end{aligned}$$

By Lemma 1, $Q(t)$ satisfies

$$Q(t) \geq \max_{t \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t-1} A(k) - \min_j \left\{ \sum_{k=\tau}^{t-1} S_j(k) \right\} \right\}.$$

By Jensen's inequality,

$$\begin{aligned}\mathbf{E}[Q(t)] &\geq \mathbf{E}\left[\max_{t \geq \tau \geq 0} \left\{ \sum_{k=\tau}^{t-1} A(k) - \min_j \left\{ \sum_{k=\tau}^{t-1} S_j(k) \right\} \right\} \right] \\ &\geq \max_{t \geq \tau \geq 0} \left\{ \mathbf{E}\left[\sum_{k=\tau}^{t-1} A(k) - \min_j \left\{ \sum_{k=\tau}^{t-1} S_j(k) \right\} \right] \right\} \\ &= \max_{t \geq \tau \geq 0} \{ \lambda(t-\tau) - f_M(t-\tau) \} \\ &= \max_{t \geq n \geq 0} \{ \lambda n - f_M(n) \}\end{aligned}$$

where $f_M(n)$ is the expected minimum of M independent binomial random variables, each with parameters (q, n). Finally,

$$\begin{aligned}\liminf_{t \to \infty} \mathbf{E}[Q(t)] &\geq \lim_{t \to \infty} \left(\max_{t \geq n \geq 0} \{ \lambda n - f_M(n) \} \right) \\ &= \sup_{n \geq 0} \{ \lambda n - f_M(n) \}.\end{aligned}$$

■

Now we are ready to prove Theorem 1.

Proof of Theorem 1: Recall that for any $n \geq 0$,

$$\liminf_{t \to \infty} \mathbf{E}[Q(t)] \geq \lambda n - f_M(n),$$

where

$$f_M(n) = \mathbf{E}[\min\{X_1, \ldots, X_M\}],$$

the expected minimum of M independent binomial random variables $X_1, \ldots, X_M$ each with parameters (q, n).

Let

$$\widehat{n}(M) = \frac{3}{4} \left(\frac{\ln(M)}{-\ln(1-q)} \right).$$

The proof will proceed by showing that

$$f_M(\widehat{n}(M)) \leq \frac{3}{4} \left(\frac{1}{-\ln(1-q)} \right) \tag{4}$$

for all $M \geq 1$. Therefore, for all $M \geq 1$ we will have the lower bound

$$\lambda\widehat{n}(M) - f_M(\widehat{n}(M)) \geq \frac{3}{4}\left(\frac{\lambda\ln(M) - 1}{-\ln(1-q)}\right).$$

Note that the largest value taken by any (q, n) binomial random variable is n. To show (4), we will use the upper bound

$$\mathbf{E}[\min\{X_1, \ldots, X_M\}] \leq n\mathbf{P}(\min\{X_1, \ldots, X_M\} > 0).$$

It is the case that $\min\{X_1, \ldots, X_M\} > 0$ if and only if $X_i > 0$ for all i. The event $X_i > 0$ occurs with probability $1 - (1-q)^n$. Since $X_1, \ldots, X_M$ are independent,

$$\begin{aligned}
\mathbf{P}(\min\{X_1, \ldots, X_M\} > 0) &= \mathbf{P}(X_1 > 0) \cdots \mathbf{P}(X_M > 0) \\
&= \left(1 - (1-q)^n\right)^M \\
&= \left(1 - e^{n\ln(1-q)}\right)^M.
\end{aligned}$$

Using $n = \widehat{n}(M)$ gives

$$\begin{aligned}
1 - e^{\widehat{n}(M)\ln(1-q)} &= 1 - e^{-(3/4)\ln(M)} \\
&= 1 - M^{-3/4} \\
&= 1 - \frac{M^{1/4}}{M}.
\end{aligned}$$

Since $M^{1/4} \leq M$ for $M \geq 1$, we can use Lemma 3 in the Appendix to obtain

$$\begin{aligned}
\mathbf{P}(\min\{X_1, \ldots, X_M\} > 0) &= \left(1 - \frac{M^{1/4}}{M}\right)^M \\
&\leq e^{-M^{1/4}}.
\end{aligned}$$

To bound $f_M(\widehat{n}(M))$,

$$\begin{aligned}
f_M(\widehat{n}(M)) &\leq \widehat{n}(M)\mathbf{P}(\min\{X_1, \ldots, X_M\} > 0) \\
&\leq \frac{3}{4}\left(\frac{\ln(M)}{-\ln(1-q)}\right)e^{-M^{1/4}} \\
&\leq \frac{3}{4}\left(\frac{1}{-\ln(1-q)}\right)\ln(M)e^{-M^{1/4}}.
\end{aligned}$$

Note that

$$\begin{aligned}
\ln(M)e^{-M^{1/4}} &= e^{\ln(\ln(M))}e^{-M^{1/4}} \\
&= e^{\left(\ln(\ln(M)) - M^{1/4}\right)}
\end{aligned}$$

Since $x^{1/2} \geq \ln(x)$ for $x \geq 0$ and both $x^{1/2}$ and $\ln(x)$ are monotonically increasing for $x \geq 0$, $M^{1/4} \geq \ln(\ln(M))$ for all $M \geq 1$. Therefore, $\ln(\ln(M)) - M^{1/4} \leq 0$ for all $M \geq 1$, or equivalently $\ln(M)e^{-M^{1/4}} \leq 1$ for all $M \geq 1$. So,

$$f_M(\widehat{n}(M)) \leq \frac{3}{4}\left(\frac{1}{-\ln(1-q)}\right)$$

for all $M \geq 1$, giving the bound

$$\begin{aligned}
\liminf_{t\to\infty} \mathbf{E}[Q(t)] &\geq \lambda\widehat{n}(M) - f_M(\widehat{n}(M)) \\
&\geq \frac{3}{4}\left(\frac{\lambda\ln(M) - 1}{-\ln(1-q)}\right).
\end{aligned}$$

■

Another interpretation of the bound presented here relates to the response time of a parallel fork-join queue as described in [3]. The fork-join queue is a model employed for parallel processing systems in which an arriving job or customer is directed to M parallel, independent servers and service completion can only take place when all M servers have completed the task. In [3] results for a continuous-time fork-join queue are presented and it is shown that the average response time (or waiting time in the queue) is $O(\ln(M))$. The multicast problem we consider in this work might also be modeled as a fork-join queue in which each parallel server represents the process of transmission to one of the M destination nodes. By Little's Law, the average queue length is equal to λ times the average response time, so our result that the average queue length scales logarithmically with M is supported by the previous results on fork-join queues.

A SIMPLE STRATEGY AND AN UPPER BOUND

Now we will analyze the behavior of the queue under a simple random linear coding strategy that we call 'code over queue contents'. This strategy sequentially performs rounds of encoding, each of which lasts several time slots. When a round of encoding begins, all packets in the queue are selected. Let C be the total number of packets in the queue at the start of a round of encoding. Encoded packets formed from random linear combinations of these packets C are then sent to the receivers. Any arrivals to the queues during a round of encoding will not be considered until the next round of encoding. The round of encoding ends when all receivers can decode all C packets. These C packets are removed from their queue, then the next round of encoding begins.

Each encoded packet is formed by randomly and uniformly selecting coefficients $a_i \in \{0, 1\}$ for $i = 1, \dots, C$ and taking a linear combination of the C head-of-line packets, where the ith packet in the linear combination is multiplied by a_i. Therefore, each encoded packet is a random linear combination of the C packets in the current coding block. A receiver can recover the original C packets once it has received C linearly independent combinations of encoded packets.

Recall that $Q(t)$ denotes the number of packets in the queue at time t. The queue length process $Q(t)$ generally does not evolve as a Markov chain. This is because each round of encoding lasts several time slots, and the distribution of the length of a round of encoding is not memoryless. Although the process $Q(t)$ is not a Markov chain, the process $Q(t_0), Q(t_1), Q(t_2), \dots$ is a Markov chain, where $t_0, t_1, t_2, \dots$ are the starting times of successive rounds of encoding. Thus, we can apply tools for Markov chains to analyze the average value of the *embedded Markov chain* $Q(t_n)$. This will provide the steady-state average value of the queue backlog at the start of each round of encoding.

To analyze the average backlog of the 'code over queue contents' strategy, we present the following theorem.

Theorem 2: The steady-state average of the embedded Markov chain $Q(t_k)$ satisfies

$$\limsup_{n\to\infty} \frac{1}{n+1}\sum_{k=0}^{n} \mathbf{E}[Q(t_k)] \;\le\; 3.12\ln(M)\left(\frac{\lambda}{q-\lambda}\right)^2 + \left(7.35\sqrt{\ln(M)} + 5.22\right)\frac{\lambda}{q-\lambda}.$$

Proof: Throughout this proof, we will let $\rho = \lambda/q$ denote *load factor* associated with the queue. To prove the upper bound, we will use the Lyapunov technique described in Lemma 4 in the Appendix. Specifically, we will use the Lyapunov function

$$h(x) = \frac{2x}{1-\rho}.$$

By applying Lemma 4 we get

$$x + \mathbf{E}[h(Q(t_{i+1}) \,|\, Q(t_i) = x] - h(x) = x + \frac{2(\lambda\mathbf{E}[T(x)] - x)}{1-\rho}$$

where $\mathbf{E}[T(x)]$ denotes the expected time to transmit a coding block containing x packets. By Theorem 3 in the Appendix,

$$\mathbf{E}[T(x)] \le \frac{1}{q}\big(x + 2\sqrt{(0.78x + 3.37)\ln(M)} + 2.61\big).$$

Using this bound on $\mathbf{E}[T(x)]$, we get the upper bound

$$x + \mathbf{E}[h(Q(t_{i+1}) \,|\, Q(t_i) = x] - h(x) \leq \frac{\rho(2\alpha(x) - 1) - 1}{1 - \rho} x$$

where

$$\alpha(x) = \frac{x + 2\sqrt{(0.78x + 3.37)\ln(M)} + 2.61}{x}.$$

So,

$$\begin{aligned} \frac{\rho(2\alpha(x) - 1) - 1}{1 - \rho} x &= \frac{\rho}{1 - \rho}\left(4\sqrt{(0.78x + 3.37)\ln(M)} + 5.22\right) - x \\ &\leq \frac{\rho}{1 - \rho}\left(4\sqrt{0.78\ln(M)x} + 4\sqrt{3.37\ln(M)} + 5.22\right) - x. \end{aligned}$$

The value of $x \geq 0$ that maximizes this expression is

$$x = 3.12\ln(M)\left(\frac{\rho}{1 - \rho}\right)^2,$$

and the associated maximum is

$$3.12\ln(M)\left(\frac{\rho}{1 - \rho}\right)^2 + \left(7.35\sqrt{\ln(M)} + 5.22\right)\frac{\rho}{1 - \rho}.$$

Therefore,

$$\limsup_{n \to \infty} \frac{1}{n + 1}\sum_{k=0}^{n} \mathbf{E}[Q(t_k)] \;\leq\; 3.12\ln(M)\left(\frac{\rho}{1 - \rho}\right)^2 + \left(7.35\sqrt{\ln(M)} + 5.22\right)\frac{\rho}{1 - \rho}.$$

■

This theorem shows that the average queue length at departure times for the 'code over queue contents' strategy scales as $O(\ln(M))$.

IV. Conclusions

In this paper we considered a simple multicast model, where a single transmitter sends packets to M receivers over lossy links. The transmitter is equipped with a queue, and our goal was to find a transmission strategy that minimizes the expected number of packets in the queue. While finding the queue length minimizing strategy is still an open problem, here we found a lower bound on achievable performance and an upper bound on the performance for a random linear coding strategy. Specifically, we have shown that queue length must scale as $\Omega(\ln(M))$, and that queue length under the random linear coding strategy scales as $O(\ln(M))$. Hence, the random linear coding strategy is order-optimal with respect to the number of receivers.

References

[1] O. Abdelrahman and E. Gelenbe. Queueing models for network coding. *IEEE Information Theory Workshop*, 2009.

[2] S. Asmussen. *Applied Probability and Queues*. Springer, 2003.

[3] F. Baccelli and A. M. Makowski. Queueing models for systems with synchronization constraints. *Proceedings of the IEEE*, 77(1):138–161, 1989.

[4] J. Byers, M. Luby, M. Mitzenmacher, and A. Rege. A digital fountain approach to reliable distribution of bulk data. *Proceedings of ACM SIGCOMM 98*, 1998.

[5] R. Cogill, B. Shrader, and A. Ephremides. Stability analysis of random linear coding across multicast sessions. *Proceedings of the 2008 IEEE International Symposium on Information Theory*, 2008.

[6] R. Cogill, B. Shrader, and A. Ephremides. Stable throughput for multicast with random linear coding. *Submitted*, 2008.

[7] C. Fragouli, D. Lun, M. Medard, and P. Pakzad. On feedback for network coding. *Proceedings of the 2007 Conference on Information Sciences and Systems*, 2007.

[8] T. Ho, R. Koetter, M. Medard, D.R. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. *Proceedings of the IEEE International Symposium on Information Theory*, 2003.

[9] D.S. Lun, P. Pakzad, C. Fragouli, M. Medard, and R. Koetter. An analysis of finite-memory random linear coding on packet streams. *Proceedings of the Fourth International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt06)*, 2006.

[10] D.J.C. MacKay. Fountain codes. *Proceedings of the Fourth IEE Workshop on Discrete Event Systems (WODES98)*, 1998.

[11] B. Shrader and A. Ephremides. On the queueing delay of a multicast erasure channel. *Proceedings of the IEEE Information Theory Workshop*, 2006.

[12] B. Shrader and A. Ephremides. A queueing model for random linear coding. *IEEE Military Communications Conference (MILCOM)*, 2007.

[13] J. Sundararajan, D. Shah, and M. Medard. Queueing advantages of network coding. *MIT LIDS Technical Report*, 2007.

[14] J.K. Sundararajan, D. Shah, and M. Medard. On queueing in coded networks - queue size follows degrees of freedom. *Proceedings of the 2007 IEEE Workshop on Information Theory for Wireless Networks*, 2007.

[15] J.K. Sundararajan, D. Shah, and M. Medard. ARQ for network coding. *Proceedings of the 2008 IEEE International Symposium on Information Theory*, 2008.

V. APPENDIX

Here we present a number of supporting results that are used to prove our main results. Most of these results appear elsewhere [7], but they are presented here to keep our treatment self-contained.

The first Lemma is used in the proof of Theorem 1.

Lemma 3: For all $x \geq a > 0$,

$$\left(1 - \frac{a}{x}\right)^x \leq e^{-a}$$

Proof: It is well known that

$$\lim_{x \to \infty} \left(1 - \frac{a}{x}\right)^x = e^{-a}.$$

In the remainder of the proof, we will show that $\left(1 - \frac{a}{x}\right)^x$ is monotonically increasing for $x \geq a$.

Taking the derivative yields

$$\frac{d}{dx}\left(1 - \frac{a}{x}\right)^x = \left(1 - \frac{a}{x}\right)^x \left(\ln\left(1 - \frac{a}{x}\right) + \frac{a}{x-a}\right).$$

The logarithmic part can be rewritten as

$$\begin{aligned} \ln\left(1 - \frac{a}{x}\right) &= -\ln\left(\frac{x}{x-a}\right) \\ &= -\ln\left(\frac{x-a+a}{x-a}\right) \\ &= -\ln\left(1 + \frac{a}{x-a}\right). \end{aligned}$$

For all $z \geq 0$, clearly $z \geq \ln(1+z)$. So,

$$\frac{a}{x-a} - \ln\left(1 + \frac{a}{x-a}\right) \geq 0$$

for all $x \geq a$. Also,

$$\left(1 - \frac{a}{x}\right)^x \geq 0$$

for all $x \geq a$. So,

$$\frac{d}{dx}\left(1 - \frac{a}{x}\right)^x \geq 0$$

for all $x \geq a$. ■

The next lemma is used in the proof of Theorem 2.

Lemma 4: Let $X(t)$ be a Markov chain with countable state space $\mathcal{X}$. Let $r : \mathcal{X} \to \mathbb{R}$ be a cost associated with being in each state in $\mathcal{X}$, and let $h : \mathcal{X} \to \mathbb{R}_+$ be a nonnegative function on $\mathcal{X}$. Then

$$\limsup_{n \to \infty} \frac{1}{n+1} \sum_{k=0}^{n} \mathbf{E}[r(X(k))] \leq \sup_{x \in \mathcal{X}} \{r(x) + \mathbf{E}[h(X(t+1)) \mid X(t) = x] - h(x)\}.$$

Proof: Let

$$\beta = \sup_{x\in\mathcal{X}}\{r(x) + \mathbf{E}[h(X(t+1)) \mid X(t)=x] - h(x)\}.$$

Then for any n

$$\begin{aligned}
\beta &\geq \frac{1}{n+1}\sum_{k=0}^{n}\mathbf{E}\big[r(X(k)) + \mathbf{E}[h(X(k+1)) \mid X(k)] - h(X(k))\big] \\
&= \frac{1}{n+1}\sum_{k=0}^{n}\Big(\mathbf{E}[r(X(k))] + \mathbf{E}\big[\mathbf{E}[h(X(k+1)) \mid X(k)]\big] - \mathbf{E}[h(X(k))]\Big) \\
&= \frac{1}{n+1}\sum_{k=0}^{n}\mathbf{E}[r(X(k))] + \frac{1}{n+1}\big(\mathbf{E}[h(X(n+1))] - \mathbf{E}[h(X(0))]\big) \\
&\geq \frac{1}{n+1}\sum_{k=0}^{n}\mathbf{E}[r(X(k))] - \frac{1}{n+1}\mathbf{E}[h(X(0))].
\end{aligned}$$

The second equality above follows from the telescoping sum of h terms. The last inequality follows from the fact that $\mathbf{E}[h(X(n+1))] \geq 0$ for all n. Also,

$$\lim_{n\to\infty}\frac{1}{n+1}\mathbf{E}[h(X(0))] = 0.$$

So,

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}\mathbf{E}[r(X(k))] &= \limsup_{n\to\infty}\left(\frac{1}{n+1}\sum_{k=0}^{n}\mathbf{E}[r(X(k))] - \frac{1}{n+1}\mathbf{E}[h(X(0))]\right) \\
&\leq \beta.
\end{aligned}$$

■

The next theorem is one of the main results of [7]. This theorem provides an upper bound on the expected time to transmit a block of packets to M receivers using random linear coding. We use this result in the proof of Theorem 2.

Theorem 3: Consider the random linear coding strategy. For any $C \geq 21\ln(M_i) - 4$, the expected time to transmit C packets from queue i satisfies

$$\mathbf{E}[T_i(C)] \leq \frac{C + 2\sqrt{\big(0.78C + 3.37\big)\ln(M_i)} + 2.61}{\min_j\{q_{ij}\}}.$$

Proof: Let X_{ij} be the time required for queue i to transmit C linearly independent combinations of C packets to receiver j. By Lemma 5 in this Appendix, an upper bound on the expected time for all M_i receivers to receive C linearly independent packets is

$$\begin{aligned}
\mathbf{E}[T_i(C)] &= \mathbf{E}[\max\{X_{i1},\ldots,X_{iM_i}\}] \\
&\leq \frac{1}{t}\big(\ln(\mathbf{E}[e^{tX_{i1}}] + \cdots + \mathbf{E}[e^{tX_{iM_i}}])\big). \qquad (5)
\end{aligned}$$

Each X_{ij} is a sum of C independent geometric random variables $Y_{i,j,0},\ldots,Y_{i,j,C-1}$, where $\mathbf{E}[Y_{i,j,k}] = \frac{1}{(1-d^{k-C})q_{ij}}$. So, the moment generating function associated with X_{ij} is

$$\mathbf{E}[e^{tX_{ij}}] = \prod_{k=0}^{C-1}\left(\frac{(1-d^{k-C})q_{ij}e^t}{(1-d^{k-C})q_{ij}e^t - (e^t-1)}\right).$$

For $t > 0$, this quantity is increases as q_{ij} decreases. So, denoting $q = \min_j\{q_{ij}\}$, we have the upper bound

$$\mathbf{E}[e^{tX_{ij}}] \leq \prod_{k=0}^{C-1} \left(\frac{(1-d^{k-C})qe^t}{(1-d^{k-C})qe^t - (e^t - 1)} \right).$$

for all j. By upper bounding the right-hand side of (5) we obtain

$$\begin{aligned} \mathbf{E}[T_i(C)] &\leq \frac{1}{t}\big(\ln(\mathbf{E}[e^{tX_{i1}}] + \cdots + \mathbf{E}[e^{tX_{iM_i}}]\big) \\ &\leq \frac{1}{t}\left(\ln\big(M_i \max_j \left\{\mathbf{E}[e^{tX_{ij}}]\right\}\big)\right) \\ &\leq \frac{1}{t}\left(\ln(M_i) - \sum_{k=0}^{C-1} \ln\left(1 - \frac{1-e^{-t}}{(1-d^{k-C})q}\right)\right) \end{aligned}$$

In the upper bound, we will use $t = -\ln(1-qz)$ for some parameter z. By Lemma 6 in this Appendix, $-\ln(1-qz) \geq qz$ for all $0 \leq z < 1$. So, the right hand side of (5) is less than or equal to

$$\frac{1}{qz}\left(\ln(M_i) - \sum_{k=0}^{C-1} \ln\left(1 - \frac{z}{1-d^{k-C}}\right)\right).$$

Let $\nu_k = z(1-d^{k-C})^{-1}$. Then

$$-\frac{1}{qz}\sum_{k=0}^{C-1} \ln\left(1 - \frac{z}{1-d^{k-C}}\right) = -\frac{1}{q}\sum_{k=0}^{C-1}\frac{\ln(1-\nu_k)}{(1-d^{k-C})\nu_k}$$

By Lemma 6 in this Appendix, $-\frac{1}{\nu}\ln(1-\nu)$ is convex and $\lim_{\nu\to 0} -\frac{1}{\nu}\ln(1-\nu) = 1$. So, for $0 < \nu \leq 1/2$,

$$-\frac{1}{\nu}\ln(1-\nu) \leq 1 + \big(4\ln(2) - 2\big)\nu.$$

This gives

$$\begin{aligned} &-\frac{1}{q}\sum_{k=0}^{C-1} \frac{\ln(1-\nu_k)}{(1-d^{k-C})\nu_k} \\ &\quad\leq \frac{1}{q}\sum_{k=0}^{C-1} \frac{1 + \big(4\ln(2) - 2\big)\nu_k}{1-d^{k-C}} \\ &\quad= \frac{1}{q}\sum_{k=0}^{C-1}\left(\frac{1}{1-d^{k-C}} + z\frac{4\ln(2)-2}{(1-d^{k-C})^2}\right) \\ &\quad\leq \frac{1}{q}\big(C + 1.61 + (0.78C + 3.37)z\big) \end{aligned}$$

if all ν_k satisfy $0 < \nu_k \leq 1/2$, which holds if $0 < z \leq 1/4$. So, this gives the bound

$$\mathbf{E}[T_i(C)] \leq \frac{1}{q}\left(\frac{\ln(M_i)}{z} + (0.78C + 3.37)z + C + 1.61\right)$$

for all $0 < z \leq 1/4$. Ignoring this constraint on z for the moment, this upper bound is minimized by using

$$z = \sqrt{\frac{\ln(M_i)}{0.78C + 3.37}}.$$

So, if $C \geq 21\ln(M_i) - 4$ then $z \leq 1/4$ and

$$\mathbf{E}[T_i(C)] \leq \frac{1}{q}\left(C + 2\sqrt{\big(0.78C + 3.37\big)\ln(M_i)} + 2.61\right).$$

■

The last two lemmas appear in [6], but are repeated here to keep our treatment self-contained.

Lemma 5: Suppose $X_1, \ldots, X_N$ are arbitrary random variables. Provided the moment generating function of each X_i exists,

$$\mathbf{E}[\max\{X_1, \ldots, X_N\}] \leq \frac{1}{t} \ln\left(\mathbf{E}[e^{tX_1}] + \cdots + \mathbf{E}[e^{tX_N}]\right).$$

for all $t > 0$.

Proof: For any $x_1, \ldots, x_N$ and $t > 0$,

$$\begin{aligned} \max\{x_1, \ldots, x_N\} &= \frac{1}{t} \ln\left(\max\{e^{tx_1}, \ldots, e^{tx_N}\}\right) \\ &\leq \frac{1}{t} \ln\left(e^{tx_1} + \cdots + e^{tx_N}\right). \end{aligned}$$

Therefore, for the random variables $X_1, \ldots, X_N$ we have

$$\begin{aligned} &\mathbf{E}[\max\{X_1, \ldots, X_N\}] \\ &\quad\leq \frac{1}{t}\mathbf{E}\left[\ln\left(e^{tX_1} + \cdots + e^{tX_N}\right)\right] \\ &\quad\leq \frac{1}{t} \ln\left(\mathbf{E}[e^{tX_1}] + \cdots + \mathbf{E}[e^{tX_N}]\right), \end{aligned}$$

where the second inequality follows from Jensen's inequality. ■

The following lemma is used in the proof of Theorem 3 in this Appendix.

Lemma 6: The function $-\frac{1}{\nu}\ln(1-\nu)$ satisfies the following:

(i) $\lim_{\nu \to 0} -\frac{1}{\nu}\ln(1-\nu) = 1$

(ii) $-\frac{1}{\nu}\ln(1-\nu) \geq 1$ for all $\nu \in (0,1)$

(iii) $-\frac{1}{\nu}\ln(1-\nu)$ is convex on $(0,1)$

Proof: We will show all parts by using the Taylor series expansion for $-\ln(1-\nu)$ on $[0,1)$:

$$-\ln(1-\nu) = \sum_{i=1}^{\infty} \frac{\nu^i}{i}.$$

To show part (i), note that

$$-\frac{1}{\nu}\ln(1-\nu) = \sum_{i=1}^{\infty} \frac{\nu^{i-1}}{i}.$$

So, it is clear that

$$\lim_{\nu \to 0} \left(1 + \sum_{i=2}^{\infty} \frac{\nu^{i-1}}{i}\right) = 1.$$

To show part (ii), note that all terms of the series for $-\frac{1}{\nu}\ln(1-\nu)$ are non-negative for $\nu \in (0,1)$. So, for all $\nu \in (0,1)$ we have

$$\begin{aligned} -\frac{1}{\nu}\ln(1-\nu) &= 1 + \sum_{i=2}^{\infty} \frac{\nu^{i-1}}{i} \\ &\geq 1. \end{aligned}$$

Finally, to show part (iii), note that

$$\begin{aligned} -\frac{d}{d\nu}\frac{1}{\nu}\ln(1-\nu) &= \frac{d}{d\nu}\sum_{i=1}^{\infty}\frac{\nu^{i-1}}{i} \\ &= \sum_{i=2}^{\infty}\frac{i-1}{i}\nu^{i-2}. \end{aligned}$$

Since all terms of this sum have positive coefficients, it is increasing with increasing ν. Hence, by the first-order conditions for convexity, $\frac{1}{\nu}\ln(1-\nu)$ is convex. ■