Stability Analysis of Random Linear Coding Across Multicast Sessions

Randy Cogill
Dept. of Systems and Information Engineering
University of Virginia
rcogill@virginia.edu

Brooke Shrader and Anthony Ephremides
Dept. of Electrical and Computer Engineering
University of Maryland
{bshrader,etony}@umd.edu

***Abstract*— We consider a problem of managing separate multicast sessions from a single transmitter. Each of K sessions has an associated packet stream, and a single transmitter must transmit these packet streams to a group of receivers. The multicast sessions are separate in the sense that each receiver only wants packets from one of the K streams. We will compare the maximum stable arrival rates that can be supported with and without using random linear coding across the K sessions.**

Intuitively, it seems that coding across sessions is not beneficial. Coding across sessions appears to introduce unnecessary additional delay since each receiver does not receive its next packet until it can decode the head-of-line packets from *all* K streams. However, we show that in many cases the maximum stable arrival rate that can be supported when coding across sessions is significantly greater than maximum stable arrival rate that can be supported when not coding across sessions. We provide a sufficient condition that indicates when coding across sessions is preferable. This condition is expressed in terms of the number of sessions, the number of receivers per session, and the reliability of the channels connecting the transmitter to the receivers.

I. INTRODUCTION

In this paper we consider a problem of managing separate multicast sessions from a single transmitter. The system we consider is modeled by a single transmitter with a set of packet streams to send to a set of receivers. Each stream is to be multicast to a separate set of receivers. The transmitter is connected to each receiver by an unreliable channel, so transmissions are not always successfully received. This is similar to the systems considered in [12] and [4].

We consider the stability of transmission schemes that encode packets across sessions. In particular, we will analyze a transmission scheme based on the use of random linear Fountain codes [10]. Rather than transmitting packets individually, when using random linear coding we transmit random linear combinations of the packets awaiting transmission. After a receiver obtains sufficiently many linearly independent combinations of the packets, it can decode the received packets to obtain the original packets.

We will explore the benefits of performing random linear coding across separate multicast sessions. Several recent papers have considered the performance of coding across sessions over lossless networks [3], [5]. Here we will examine not only how the benefits of coding are affected by the number of sessions and receivers, but also how coding improves performance over unreliable channels. Specifically, we will compare two transmission schemes. The first is a standard retransmission scheme, where unencoded packets are transmitted until they have been successfully received by their assigned receivers. The second scheme transmits random linear combinations of the head-of-line packets in all streams until each receiver can decode *all* of the head-of-line packets. It will be shown that, surprisingly, the second scheme can often support higher packet arrival rates than the first scheme.

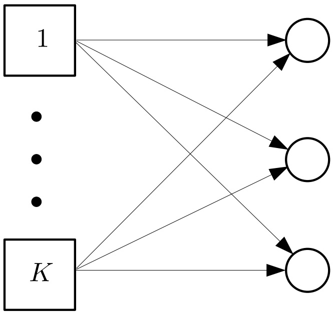

Fig. 1. Multiple overlapping multicast sessions. K packets are to be transmitted to a set of M receivers. There is an edge connecting packet i to receiver j if packet i must be received by receiver j. In this case, each packet must be received by all M receivers.

To show why the results of this paper are surprising, we will compare them to known results for some special cases. The setup considered in this paper is similar to those in [12] and [4]. In [12], a single queue of packets is considered, and all of the packets are to be multicast to a common set of M receivers. We can consider a similar setup in the framework of this paper, where instead we have K queues. Packets from each session must be received by all M receivers, as depicted in Figure 1. When coding over the K head-of-line packets, the maximum stable arrival rate is the same as the maximum stable arrival rate when coding over the first K packets of a single queue, as in [12].

As another case, we could consider a setup where we have K separate unicast sessions. Here we have K queues and K receivers. For all i, receiver i only wants packets in queue i as depicted in Figure 2. In this case, the maximum stable arrival rate is always higher if we do not perform coding across sessions, as noted in [4].

In this paper we analyze the case depicted in Figure 3, where we have K separate multicast sessions. Here each packet is destined for M receivers. However, unlike the setup shown in Figure 1, each packet is multicast to a *distinct* set of receivers. That is, each receiver wants packets from only

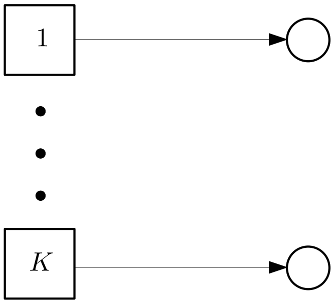

Fig. 2. Separate unicast sessions. K packets are to be transmitted to a set of K receivers. Packet i must be received by receiver i. In this case, coding across sessions is never preferable.

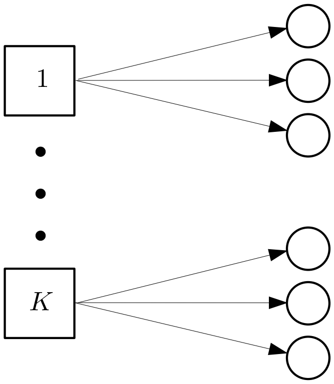

Fig. 3. Separate multicast sessions. K packets are to be transmitted to a set of MK receivers. Packet i must be received by receivers $M(i-1)+1,\ldots,Mi$. For $M>1$, coding across sessions is often preferable.

one of the sessions. Intuition suggests that this is closer to the unicast setup shown in Figure 2 than the multicast setup in Figure 1. By coding across all K sessions, it seems that we introduce additional delay since packets do not leave the queues until each receiver can decode *all* K packets. While each receiver wants all K packets in the case shown in Figure 1, this is clearly an unnecessary requirement for the cases shown in Figures 2 and 3. However, for the case shown in Figure 3, we will show that coding across all sessions can still often support significantly higher stable arrival rates than when not coding. We provide a sufficient condition that indicates when coding across sessions is preferable. This condition is expressed in terms of the number of sessions, the number of receivers per session, and the reliability of the channels connecting the transmitter to the receivers.

II. SYSTEM MODEL

The system we consider is modeled by a single transmitter with K packet streams to send to a set of receivers. Each packet stream has an associated queue. Time is slotted, and in each time slot at most one packet may arrive in each queue. Arrivals are independent and identically distributed over time. The expected number of packets arriving in a queue in each time slot is the *arrival rate* for that queue. In our model, each of the K queues have the same arrival rate. The *sum* arrival rate is the expected total number of packets arriving in all K queues in each time slot.

The set of receivers is denoted $\mathcal{R}$, and each stream is to be multicast to some subset of the receivers in $\mathcal{R}$. We let $\mathcal{R}_i \subseteq \mathcal{R}$ denote the set of receivers that the packets in queue i must be sent to. In this paper, we will focus on the case where $|\mathcal{R}_i| = M$ for all i and $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ for all $i \neq j$.

The transmitter is connected to each receiver by an erasure channel. In each time slot, a single packet may be transmitted. A transmitted packet is successfully received by each receiver with probability q, and successes are independent across the receivers. We will consider the set of sum arrival rates for which this system is stable for two transmission schemes:

(a) **Retransmission:** The head-of-line packet in a selected session is transmitted repeatedly until it is successfully received by the M receivers assigned to this packet. This packet is then removed from the queue, and the next session in round-robin order is selected.

(b) **Random linear coding:** In each time slot, we create a random linear combination of the K head-of-line packets by selecting a random subset of these packets and adding them bitwise. This encoded packet is then transmitted. Random linear combinations of the K head-of-line packets are transmitted repeatedly until all MK receivers can decode all K head-of-line packets. If some queues are unoccupied when a round of encoding begins, we only code over the current head-of-line packets. Any arrivals to those queues during transmission will not be moved to the head of the queue until the current head-of-line packets are decoded.

One of the key differences between these two schemes is that retransmission only requires that each packet is successfully received by the receivers assigned to it, but random linear coding implicitly requires that each packet is successfully received by every receiver. However, we will see that random linear coding still often supports higher arrival rates than retransmission.

III. MAIN RESULTS

The goal of this paper is to determine when random linear coding across sessions can support higher arrival rates than retransmission. When coding across sessions, let $\mathbf{E}[T_1]$ be the expected number of time slots required to transmit K head-of-line packets. For a given arrival rate λ, we can define the *traffic intensity* as

$$\rho_1 = \frac{\lambda \mathbf{E}[T_1]}{K}.$$

It is well known [7] that the queues are stable if and only if $\rho_1 < 0$. When coding across sessions, the maximum stable arrival rate is defined as

$$\begin{aligned} \lambda_1 &= \sup\{\lambda \,|\, \rho_1 < 1\} \\ &= \frac{K}{\mathbf{E}[T_1]}. \end{aligned}$$

Similarly, when not coding across sessions, let $\mathbf{E}[T_2]$ be the expected number of time slots required to transmit one packet. For a given arrival rate λ, the traffic intensity is

$$\rho_2 = \lambda \mathbf{E}[T_2].$$

The maximum stable arrival rate when not coding across sessions is

$$\begin{aligned} \lambda_2 &= \sup\{\lambda \mid \rho_2 < 1\} \\ &= \frac{1}{\mathbf{E}[T_2]}. \end{aligned}$$

By the end of this section we will provide a condition that indicates when $\lambda_1 \geq \lambda_2$. Theorem 1, which provides a lower bound on λ_1, is based on the general inequality in Lemma 1. The proof of this lemma also appears in [1], but is provided here to keep our treatment self-contained.

Lemma 1: Suppose $X_1, \ldots, X_N$ are identically distributed, but not necessarily independent random variables. For any $t > 0$,

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

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

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

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

Now we provide a lower bound on λ_1, the maximum stable arrival when coding across all sessions.

Theorem 1: The maximum stable arrival rate using random linear coding across sessions satisfies

$$\lambda_1 \geq \frac{-K\ln\left(1 - qe^{-1}\right)}{0.46K + \ln(M) + \ln(K) + 1.26}$$

Proof: We will start by computing an upper bound on $\mathbf{E}[T_1]$, the expected number of time slots required to transmit K head-of-line packets. Let X_i be the number of transmissions until receiver i successfully receives K linearly independent combinations of the K head-of-line packets. $X_1, \ldots, X_{MK}$ are identically distributed, and for each i we can write X_i as

$$X_i = Y_{0,i} + \cdots + Y_{K-1,i}.$$

Here, $Y_{j,i}$ is the time for receiver i to receive the next linearly independent combination after successfully receiving j linearly independent combinations. Since the channel connecting the transmitter to receiver i is memoryless, $Y_{0,i}, \ldots, Y_{K-1,i}$ are independent. Each transmitted packet is successfully received by receiver i with probability q. The combination associated with this packet is linearly independent of the combinations received so far if it is not the sum of a subset of the j packets received so far. After receiving j linearly independent combinations, there are $2^K - 2^j$ linearly independent combinations remaining. Since each transmitted packet is sampled uniformly from among the 2^K possible combinations of the head-of-line packets,

$$Y_{j,i} \sim \text{Geom}\big((1 - 2^{j-K})q\big).$$

Here we will let $N = MK$. Using the inequality in Lemma 1, we have

$$\mathbf{E}[\max\{X_1, \ldots, X_N\}] \leq \frac{1}{t}\big(\ln(N) + \ln(\mathbf{E}[e^{tX_1}])\big) \tag{1}$$

for all $t > 0$. The moment generating function of X_1 is

$$\begin{aligned} \mathbf{E}[e^{tX_1}] &= \mathbf{E}\big[e^{t(Y_{0,1} + \cdots + Y_{K-1,1})}\big] \\ &= \mathbf{E}[e^{tY_{0,1}}] \cdots \mathbf{E}[e^{tY_{K-1,1}}] \\ &= \prod_{j=0}^{K-1}\left(\frac{(1 - 2^{j-K})qe^t}{1 - (1 - (1 - 2^{j-K})q)e^t}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \ln\big(\mathbf{E}[e^{tX_1}]\big) &= \sum_{j=0}^{K-1}\ln\left(\frac{(1 - 2^{j-K})qe^t}{1 - (1 - (1 - 2^{j-K})q)e^t}\right) \\ &= -\sum_{j=0}^{K-1}\ln\left(1 - \frac{1 - e^{-t}}{(1 - 2^{j-K})q}\right). \end{aligned}$$

Applying this to (1) gives

$$\mathbf{E}[T_1] \leq \frac{1}{t}\left(\ln(N) - \sum_{j=0}^{K-1}\ln\left(1 - \frac{1 - e^{-t}}{(1 - 2^{j-K})q}\right)\right).$$

Using $t = -\ln(1 - qe^{-1})$ gives

$$\mathbf{E}[T_1] \leq \frac{-1}{\ln\left(1 - qe^{-1}\right)}\left(\ln(N) - \sum_{j=0}^{K-1}\ln\left(1 - \frac{e^{-1}}{1 - 2^{j-K}}\right)\right).$$

Since

$$\begin{aligned} &-\sum_{j=0}^{K-1}\ln\left(1 - \frac{e^{-1}}{1 - 2^{j-K}}\right) \\ &\quad\leq -K\ln(1 - e^{-1}) - \sum_{j=1}^{\infty}\ln\left(\frac{1 - \frac{e^{-1}}{1 - 2^{-j}}}{1 - e^{-1}}\right) \\ &\quad\leq 0.46K + 1.26, \end{aligned}$$

this finally gives

$$
\begin{aligned}
\mathbf{E}[T_1] &\leq \frac{\ln(N)+0.46K+1.26}{-\ln\left(1-qe^{-1}\right)} \\
&= \frac{\ln(M)+\ln(K)+0.46K+1.26}{-\ln\left(1-qe^{-1}\right)}
\end{aligned}
$$

The maximum stable arrival rate then satisfies

$$
\begin{aligned}
\lambda_1 &= \frac{K}{\mathbf{E}[T_1]} \\
&\geq \frac{-K\ln\left(1-qe^{-1}\right)}{\ln(M)+\ln(K)+0.46K+1.26}
\end{aligned}
$$

■

Now that we have a bound on λ_1, we would like to compare this to λ_2. The following theorem gives an upper bound on λ_2, the maximum stable arrival rate when not coding across sessions.

Theorem 2: The maximum stable arrival rate using unencoded retransmission satisfies

$$
\lambda_2 \leq \frac{-\ln\left(1-q\right)}{\ln(M+1)+0.3}
$$

Proof: We will start by computing a lower bound on $\mathbf{E}[T_2]$, the expected number of time slots required to transmit a packet from queue 1. Without loss of generality, we will focus our attention on queue 1. Let Z_i be the number of attempted transmissions to receiver i until it successfully receives its packet. Since the number of attempted transmissions until success for a given receiver is geometrically distributed,

$$
\mathbf{P}(Z_i \leq n) = 1-(1-q)^n.
$$

The time to transmit a packet from queue 1 is $\max\{Z_1,\ldots,Z_M\}$. Since the channels connecting the transmitter to receivers $1,\ldots,M$ are independent,

$$
\begin{aligned}
\mathbf{P}(\max\{Z_1,\ldots,Z_M\}\leq n) &= \mathbf{P}(Z_1\leq n,\ldots,Z_M\leq n) \\
&= \mathbf{P}(Z_1\leq n)\cdots\mathbf{P}(Z_M\leq n) \\
&= \left(1-(1-q)^n\right)^M.
\end{aligned}
$$

So,

$$
\begin{aligned}
&\mathbf{E}[\max\{Z_1,\ldots,Z_M\}] \\
&= \sum_{n=1}^{\infty} n\mathbf{P}(\max_i\{Z_i\}=n) \\
&= \sum_{n=1}^{\infty} n\big(\mathbf{P}(\max_i\{Z_i\}\leq n)-\mathbf{P}(\max_i\{Z_i\}\leq n-1)\big) \\
&= \sum_{n=1}^{\infty} n\Big(\big(1-(1-q)^n\big)^M-\big(1-(1-q)^{n-1}\big)^M\Big).
\end{aligned}
$$

Letting $\lambda = -\ln(1-q)$,

$$
\begin{aligned}
&\mathbf{E}[\max\{Z_1,\ldots,Z_M\}] \\
&= \sum_{n=1}^{\infty} n\Big(\big(1-e^{-\lambda n}\big)^M-\big(1-e^{-\lambda(n-1)}\big)^M\Big) \\
&= \sum_{n=1}^{\infty} n\int_{n-1}^{n}\frac{d}{dx}\big(1-e^{-\lambda x}\big)^M dx \\
&= \lambda M\sum_{n=1}^{\infty} n\left(\int_{n-1}^{n}\big(1-e^{-\lambda x}\big)^{M-1}e^{-\lambda x}dx\right) \\
&\geq \lambda M\int_{0}^{\infty} x\big(1-e^{-\lambda x}\big)^{M-1}e^{-\lambda x}dx
\end{aligned}
$$

The inequality above follows from Hölder's inequality. As shown in [11], the last integral above can be evaluated to give

$$
\begin{aligned}
&\lambda M\int_{0}^{\infty} x\big(1-e^{-\lambda x}\big)^{M-1}e^{-\lambda x}dx \\
&\qquad = \frac{1}{\lambda}\sum_{i=1}^{M}\frac{1}{i} \\
&\qquad \geq \frac{1}{\lambda}+\frac{1}{\lambda}\sum_{i=2}^{M}\int_{i}^{i+1}\frac{1}{x}dx \\
&\qquad = \frac{1}{\lambda}+\frac{1}{\lambda}\int_{2}^{M+1}\frac{1}{x}dx \\
&\qquad = \frac{\ln(M+1)+1-\ln(2)}{-\ln\left(1-q\right)} \\
&\qquad \geq \frac{\ln(M+1)+0.3}{-\ln\left(1-q\right)}.
\end{aligned}
$$

The first inequality above follows from the fact that $\frac{1}{x}$ is decreasing. So,

$$
\mathbf{E}[T_2] \geq \frac{\ln(M+1)+0.3}{-\ln\left(1-q\right)}
$$

The maximum stable arrival rate then satisfies

$$
\begin{aligned}
\lambda_2 &= \frac{1}{\mathbf{E}[T_2]} \\
&\leq \frac{-\ln\left(1-q\right)}{\ln(M+1)+0.3}
\end{aligned}
$$

■

Our ultimate goal is to determine when it is preferable to code across sessions. That is, we want to know when $\lambda_1 \geq \lambda_2$. The following corollary gives a simple sufficient condition indicating when coding is preferable. After establishing this corollary we are able to make several general statements relating the benefits of coding to the parameters q, M, and K.

Corollary 1: $\lambda_1 \geq \lambda_2$ if

$$
\frac{0.46K+\ln(M)+\ln(K)+1.26}{K\big(\ln(M+1)+0.3\big)} \leq \frac{\ln(1-qe^{-1})}{\ln(1-q)}. \qquad (2)
$$

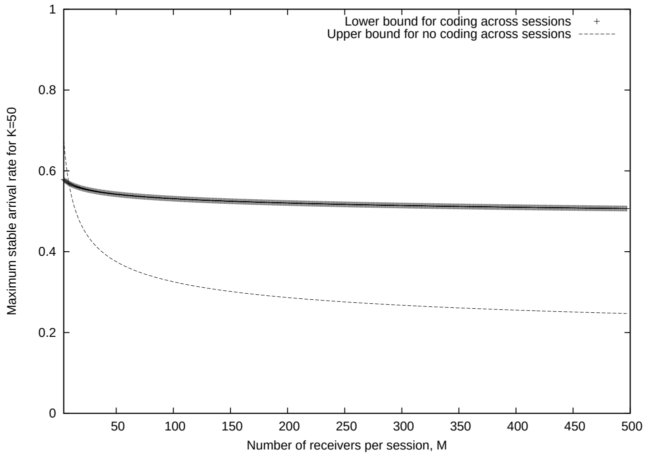

Fig. 4. Comparison of the maximum stable arrival rates for 50 sessions with and without coding. In this case, $q = 0.8$ and we compare stable rate as a function of M, the number of receivers per session. Transmitting without coding is only preferable for very small M. For large M, the maximum stable arrival rate using coding is more than twice the maximum stable rate without coding.

This corollary is proven by simply comparing the lower and upper bounds for λ_1 and λ_2 that were obtained in Theorems 1 and 2. Several observations regarding the condition (2):

(i) **Coding is preferable when M is large.** The left-hand side of (2) is decreasing with increasing M. So, if coding is preferable for some q, K, and M' then coding is also preferable for q, K, and all $M > M'$.

(ii) **For all $M > 1$, there are q and K such that coding is preferable.** As noted in the introduction, coding across sessions is never preferable when $M = 1$. However, for $M = 2$, there are q and K such that coding is preferable. For example, coding is preferable for $M = 2$, $q = 0.2$, and $K = 500$. From item (i), we also see that coding is preferable for $q = 0.2$, $K = 500$, and all $M > 1$.

(iii) **Coding is preferable when K is large.** The left-hand side of (2) is decreasing with increasing K. So, if coding is preferable for some q, M, and K', then coding is also preferable for q, M, and all $K > K'$.

(iv) **Coding is preferable when q is small.** The right hand side of (2) is decreasing with increasing q, and approaches zero as $q \to 1$. So, if coding is preferable for some m, K, and q', then coding is also preferable for M, K, and all $q < q'$. On the other hand, for any M and K, there is q sufficiently close to 1 such that the condition (2) fails to hold. This is to be expected, since as $q \to 1$, coding is not preferable. To see why this is true, without coding the expected transmission time is K when $q = 1$. With coding, the expected transmission time is strictly greater than K for all $q \leq 1$ since each receiver must receive K linearly independent columns.

IV. CONCLUSIONS

In this paper we considered a problem of managing several separate multicast sessions from a single transmitter. We compared the maximum stable arrival rates when using a standard retransmission scheme and when using a scheme that performs random linear coding across the head-of-line packets of the queues associated with the K sessions. Since the multicast sessions are separate, it seems that coding across sessions is not beneficial. However, we have shown that in many cases the maximum stable arrival rate that can be supported when coding across sessions is significantly greater than maximum stable arrival rate that can be supported when not coding across sessions. We provided a sufficient condition in terms of M, K, and q that indicates when coding across sessions supports higher arrival rates.

The results of this paper immediately raise several questions. In particular:

- The random linear coding approach requires the transmission of additional overhead information to identify the random subset of packets used in forming a random linear combination. What is the effect of this coding overhead on stable arrival rates?
- Are similar results obtained in a model where there are an equal number of receivers in each multicast group?
- How does the maximum stable rate we can achieve through coding compare with the maximum stable rate that can be achieved using scheduling with channel state information?

These issues are addressed in the upcoming paper [2].

REFERENCES

[1] R. Cogill. Randomized load balancing with non-uniform task lengths. In *Fourty-Fifth Allerton Conference on Communication, Contol, and Computing*, 2007.
[2] R. Cogill, B. Shrader, and A. Ephremides. Stable throughput for multicast with inter-session network coding. In preparation.
[3] A. Eryilmaz and D.S. Lun. Control for inter-session network coding. *Submitted to IEEE Transactions on Information Theory*, 2007.
[4] A. Eryilmaz, A. Ozdaglar, M. Medard, and E. Ahmed. On the delay and throughput gains of coding in unreliable networks. *Submitted to IEEE Transactions on Information Theory*, 2007.
[5] T. Ho, Y.-H. Chang, and K. Han. On constructive network coding for multiple unicasts. In *Forty-fourth Allerton Conference on Communication, Control, and Computing*, 2006.
[6] T. Ho, R. Koetter, M. Medard, D.R. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. In *IEEE International Symposium on Information Theory*, 2003.
[7] L. Kleinrock. *Queueing Systems, Volume I: Theory*. Wiley, New York, 1975.
[8] S.-Y.R. Li, R.W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49, 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. In *Fourth International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt06)*, 2006.
[10] D.J.C. MacKay. Fountain codes. In *Fourth IEE Workshop on Discrete Event Systems (WODES98)*, 1998.
[11] S. Ross. *Introduction to Probability Models*. Academic Press, 2007.
[12] B. Shrader and A. Ephremides. On the queueing delay of a multicast erasure channel. In *IEEE Information Theory Workshop*, 2006.