On ARQ for Packet Erasure Channels with Bernoulli Arrivals

Dinkar Vasudevan, Vijay G. Subramanian and Douglas J. Leith
Hamilton Institute, National University of Ireland, Maynooth

Abstract—We study packet streaming over an erasure channel with delayed feedback. We consider the lag in playback between the sender and the receiver as the performance criterion and propose and analyze schemes to minimize the lag. We show that at lower delays in feedback, purely retransmission based schemes are better than random linear coding schemes and also analyze the tradeoff of the lag with the delay in feedback.
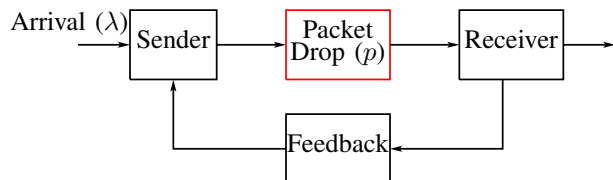
I. INTRODUCTION

Real-time data streaming applications are characterized by low delay and in-order delivery requirements. Feedback can be crucially used in such applications to encode the data packets in a manner so as to prioritize the decoding of the earlier data packets without losing efficiency (capacity) of communication. While feedback will not increase the capacity, it can dramatically increase the error exponent yielding the same performance as in the non-feedback case with much smaller block-lengths, i.e., effectively lower delays. Furthermore, in an online setting, the coding must also be adaptable to variability in the arrival of packets. Most of the literature [1],[2],[3],[4] has focused on packet communication over channels in single as well as multiuser settings, assuming that feedback on transmission is available at the sender *without any delay*. Moreover, such studies have emphasized "queue management" at the sender with the objective of reducing the average delay suffered by a packet in reaching the receiver.

In this paper, we consider delayed feedback. Moreover, in contrast to average packet delay, we focus on the average lag, which captures the number of packets by which playback at the receiver lags the playback at the sender. Thus, the emphasis is not on the number of packets delivered, but on the number of packets delivered *in-order*. This metric better reflects the performance of coding schemes for streaming applications. We study a natural purely retransmission based scheme as well as a random linear coding based scheme for point to point communication over a packet erasure channel with Bernoulli packet arrivals and study the performance of these schemes as a function of the delay in feedback. Our main aim is to demonstrate that in terms of the average packet lag, packet retransmission schemes can be bettered by coding schemes when the feedback delay is not very small. The next section describes the problem in greater detail.

II. PROBLEM SETUP

We consider communication over a slotted packet erasure channel (refer Figure 1), where in each slot a packet sent from the transmitter has an independent probability p of erasure. If

Fig. 1. Problem Setup

the packet is not erased, it is received perfectly by the receiver after a fixed delay of D_f slots. We assume that the receiver regularly feedbacks information on the packets received, back to the transmitter. We assume this feedback is received, again after a fixed delay D_b, by the sender. Thus, the feedback at time i indicates whether packet transmitted at time $i - D$ was successfully received, where D is the sum of the packet transmission time D_f and the feedback time D_b. We assume that the feedback is erasure free and available at every time slot. Figure 2 illustrates what is known at the sender at time slot i.

Information packets arrive at the sender independently in every time slot according to a Bernoulli(λ) process. The sender wishes to stream incoming information packets in such a way that the *lag* at the receiver is minimized. The lag refers to the difference in the total number of packets that have arrived at the transmitter and the number of *contiguous* packets starting from the first packet that have been received at the decoder. As defined, the lag is composed of two terms, packets queued at the transmitter that are yet to be transmitted and packets within the re-ordering buffer (also called play-out buffer) at the receiver that await older packets for playback. The information that is available at the sender is the feedback from the receiver. Based on the feedback, the sender decides what packet to transmit. Depending on whether or not the sender encodes the information, there are two classes of schemes:

1) Retransmission based schemes: The sender only retransmits original packets when the feedback says that a packet was erased; and
2) Coding based schemes: Information about a packet is encoded in a stream of packets. The coding scheme we study is one where the packet sent is a random linear combination of information packets arrived so far at the sender which have not been decoded by the reciever. Feedback is used to determine the latter information.

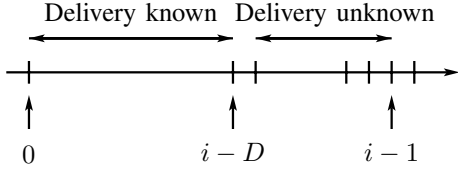Thinking of each packet as a symbol, the former scheme is

Fig. 2. At time i, the sender knows delivery status of packets till time $i-D$.

repetition coding while the latter is parity-check-based coding which, in general, encompasses the former.

The advantage of one scheme over the other becomes evident when one consider extremes of delay in the feedback. If the feedback is instantaneous ($D = 1$), then the retransmission based scheme (in this case, a "retransmit-packet-until-successful" scheme) has a lesser lag as compared to the coding scheme. If the feedback is non-existent (infinite delay), then the coding scheme has lesser lag as compared to the retransmission based scheme.

The organization of the rest of the paper is as follows: We consider a particular retransmission based scheme called "Max-strat" and study the tradeoff between the lag and the delay in feedback D for this strategy (Section III). In particular, we show that average lag is bounded as $O(D^2)$ whenever $\lambda < 1 - p$, where $1 - p$ is the capacity of the erasure channel. We also study a causal random linear coding scheme and analytically express the average lag (Section IV). In Section V, we compare the performance of the schemes from simulations and conclude.

III. MAX-STRAT

Let $A(i)$ and $Z(i)$ denote, respectively, the indicator of an arrival of a new packet at the transmitter at time i and the indicator of the channel being OFF (or the erasure of a packet, if sent) at time i. $A(i)$ and $Z(i)$ are distributed, respectively, as i.i.d Bernoulli(λ) and Bernoulli(p) random variables. Let $C(i) = \sum_{j=0}^{i} A(j)$ denote the cumulative number of arrivals until i. If a packet arrives at time i, it is assigned the number $C(i)$. Let β_i denote the number of the packet transmitted in slot i. We assume that a packet arrival at time i as well as the feedback on the transmission of packet β_{i-D} (sent at time $i - D$) arrives at the beginning of slot i. Time starts from 0 onwards and we let $A(i) = Z(i) = \beta_i = 0$ for $i < 0$. We now define the scheme Max-strat. The sender decides the packet β_i to be sent in slot i as

$$\beta_i = \begin{cases} \beta_{i-D} & Z(i-D) = 1 \\ \max\{\beta_{i-1}, \ldots, \beta_{i-D}\} + 1 & \text{if } Z(i-D) = 0, C(i) > \\ & \max\{\beta_{i-1}, \ldots, \beta_{i-D}\} \\ \max\{\beta_{i-1}, \ldots, \beta_{i-D+1}\} & \text{else} \end{cases} \quad (1)$$

The scheme is explained as follows: If packet β_{i-D} is erased, it is resent. Otherwise if there is a fresh packet (indicated by $C(i) > \max\{\beta_{i-1}, \ldots, \beta_{i-D}\}$), it is sent, and if not, then the sender sends the most recent packet from among

the past $D - 1$ packets. The latter case is the reason why we name the scheme "Max-strat". Note that packets in transit, $\{\beta_{i-1}, \ldots, \beta_{i-D}\}$, can be received correctly, so an intelligent choice needs to be made so that capacity is not sacrificed. One could alternatively propose a different scheme such as Min-strat where the sender sends the oldest packet in the transit window, or Div-strat where the sender selects a packet that has been retransmitted the least number times. We discuss this in Section V.

Let $d(i)$ denote the maximum number of *contiguous* packets that are decoded by the receiver at time i starting from the first packet. In other words, $d(i)$ is such that the receiver can decode packets $1, \ldots, d(i)$, but cannot decode packet $d(i)+1$. The lag at time i is denoted by $L(i)$ and is given by $L(i) = C(i) - d(i)$. The time-averaged lag (alt. expected lag) is defined as

$$\bar{L} = \lim_{i \to \infty} \frac{1}{i} \sum_{j=0}^{i} L(j) = \lim_{i \to \infty} \frac{1}{i} \sum_{j=0}^{i} \left(C(j) - d(j)\right). \quad (2)$$

Our objective is to express $\bar{L}$ as a function of D.

Case $D = 1$: Feedback is received at the beginning of the next slot. All strategies are the same here, namely to retransmit the packet until successful. The expected lag is just the expected queue length of a FIFO $Geom/Geom/1$ queue with a rate λ of arrivals and service rate of $\mu \triangleq 1 - p$ and is given by [1], $\bar{L} = \frac{\lambda(1-\mu)}{\mu-\lambda}$, when $\lambda < 1 - p$. If $\lambda > 1 - p$, the average lag is unbounded.

Case $D > 1$: We prove the following theorem.

Theorem 1. *The time-averaged lag of Max-strat for a given feedback delay of D is bounded as*

$$\bar{L} \leq \frac{\lambda(1-\mu)}{\mu-\lambda} + D\left[\sum_{k=1}^{D}\binom{D}{k}(-1)^{k+1}\frac{p^k}{1-p^k}\right] + 1 \quad (3)$$

Proof: The lag is bound as

$$L(i) = C(i) - d(i) \overset{(a)}{\leq} C(i) - (\min\{\beta_i, \ldots, \beta_{i-D+1}\} - 1)$$
$$= C(i) - \max\{\beta_i, \ldots, \beta_{i-D+1}\} + \max\{\beta_i, \ldots, \beta_{i-D+1}\}$$
$$- (\min\{\beta_i, \ldots, \beta_{i-D+1}\} - 1)$$
$$\overset{(b)}{=} Q(i) + E(i)$$

The inequality (a) follows since the oldest backlogged packet $d(i) + 1 \in \{\beta_i, \ldots, \beta_{i-D+1}\}$. The last equality follows from defining the queue length $Q(i)$ at time i to be

$$Q(i) = C(i) - \max\{\beta_i, \ldots, \beta_{i-D+1}\} \quad (4)$$

and the quantity $E(i)$ to be

$$E(i) = \max\{\beta_i, \ldots, \beta_{i-D+1}\} - \min\{\beta_i, \ldots, \beta_{i-D+1}\} + 1. \quad (5)$$

The quantity $E(i)$ is an excess over and above the queue length $Q(i)$, which contributes to the lag. It includes all the packets which lie in the interval $[\min\{\beta_i, \ldots, \beta_{i-D+1}\}, \max\{\beta_i, \ldots, \beta_{i-D+1}\}]$. For ease

of notation, we define $\text{Max}(i) \triangleq \max\{\beta_i, \ldots, \beta_{i-D+1}\}$ and $\text{Min}(i) \triangleq \min\{\beta_i, \ldots, \beta_{i-D+1}\}$.

We will now show the following: The first term in the R.H.S of the bound (3) is a bound on the time averaged $Q(i)$'s while the remaining terms are a bound on the time averaged $E(i)$'s. We first focus on $Q(i)$, which evolves as

$$Q(i+1) = Q(i) + C(i+1) - C(i) - \big(\text{Max}(i+1) - \text{Max}(i)\big)$$
$$\overset{(a)}{=} Q(i) + A(i+1) - 1_{C(i+1) > \text{Max}(i)}\big(1 - Z(i-D+1)\big)$$
$$\overset{(b)}{=} Q(i) + A(i+1) - 1_{C(i+1) > C(i) - Q(i)}\big(1 - Z(i-D+1)\big)$$
$$= Q(i) + A(i+1) - 1_{Q(i)+A(i+1) > 0}\big(1 - Z(i-D+1)\big)$$
$$= [Q(i) + A(i+1) + Z(i-D+1) - 1]^+$$

The equality (a) follows from the choice of β_{i+1} (Refer (1)). The equality (b) follows from the definition of $Q(i)$ (see (4)). Thus, the change in the queue length is due to two factors, increase due to the fresh arrival, $A(i+1)$, and decrease if there was no erasure of the packet sent at time $(i+1-D)$. Observe that $Q(i)$ is stochastically equivalent to $\tilde{Q}(i)$ satisfying the recursion

$$\tilde{Q}(i+1) = [\tilde{Q}(i) + A(i+1) + Z(i+1) - 1]^+ \qquad (6)$$

The above follows since the sequence $\{Z(i)\}$ is i.i.d. Note this is just the recursion for the standard FIFO $Geom/Geom/1$ queue for the case $D = 1$. It therefore holds that

$$\lim_{i \to \infty} \frac{1}{i} \sum_{j=0}^{i} Q(j) = \frac{\lambda(1-\mu)}{\mu - \lambda}. \qquad (7)$$

We now focus on the sequence $\{E(i)\}$:

$$E(i) = \text{Max}(i) - \text{Min}(i) + 1 \triangleq V(i) + 1 \qquad (8)$$

where we define $V(i) \triangleq \text{Max}(i) - \text{Min}(i)$. In the following, we focus on the evolution of $V(i)$. Consider the window of packets $(\beta_i, \ldots, \beta_{i-D+1})$. We define a corresponding window of lags $(x_i, \ldots, x_{i-D+1})$ where each $x_j \triangleq \text{Max}(i) - \beta_j$. Note, $V(i)$ is the maximum element in the window of lags $(x_i, \ldots, x_{i-D+1})$. From (1), this window evolves as $(x_i, \ldots, x_{i-D+1}) =$

$$\begin{cases} (x_{i-D}, x_{i-1}, \ldots, x_{i-D+1}) & Z(i-D) = 1 \\ (0, x_{i-1}+1, \ldots, x_{i-D+1}+1) & Z(i-D) = 0 \text{ and} \\ & C(i) > \text{Max}(i-1) \\ (0, x_{i-1}, \ldots, x_{i-D+1}) & \text{else} \end{cases}$$
$$(9)$$

We will now show that the window $(x_i, \ldots, x_{i-D+1})$ has a stationary distribution, and bound the expectation of $V(i)$ with respect to this distribution. This will give us a bound on the time averaged $V(j)$'s, i.e., $\lim_{i\to\infty} \frac{1}{i}\sum_{j=0}^{i} V(j)$. We introduce some notation connected to Markov chains from [5]. For two states x, y, let $\rho_{xy} = \text{Pr}_x(T_y < \infty)$, where the R.H.S denotes the probability that the first time to visit the state y starting from x is finite. Let S be the vectors $y \in \mathbb{Z}^d$ which can be reached from the state $x_0 = (0, 1, \ldots, D-1)$. We say that the state y is reachable from x if $\rho_{xy} > 0$. The state x is said

to be recurrent if $\rho_{xx} = 1$ and positive recurrent if $\mathbb{E}_x T_x < \infty$, where $\mathbb{E}_x T_x$ denotes the expected time of first return to state x starting from x. We prove the following statements:

1) The state x_0 is positive recurrent, i.e., $\mathbb{E}_{x_0} T_{x_0} < \infty$; and
2) The Markov chain is irreducible on the set S.

Lemma 2. $\mathbb{E}_{x_0} T_{x_0} < \infty$.

Proof: We have

$$\mathbb{E}_{x_0} T_{x_0} = \sum_{j=1}^{\infty} \text{Pr}_{x_0}(T_{x_0} \geq j) \overset{(a)}{\leq} \sum_{j=1}^{\infty} \big(1 - (\lambda\bar{p})^D\big)^{\lfloor \frac{j}{D} \rfloor} < \infty$$

The reason for (a) is the following: Divide the j timeslots into j/D chunks of length D. Suppose there is an arrival in each slot of a chunk and successful transmission in each slot of the previous chunk, this implies a fresh packet is sent in each slot and that we reach state x_0 at the end of the chunk. The probability of this event for every chunk is independently $(\lambda\bar{p})^D$. Thus, the event $\{T_{x_0} \geq j\}$ implies that for each chunk the aforementioned event does not occur. ∎

Lemma 3. *The Markov chain is irreducible on the set S.*

Proof: We have to show that $\rho_{xy} > 0$ for any $x, y \in S$. This is true, by definition, if $x = x_0$. Furthermore, it holds that if x_0 is recurrent and $\rho_{x_0 y} > 0$, then $\rho_{y x_0} = 1$ [5]. For $x \neq x_0, y \neq x_0$, it holds that $\rho_{xy} \geq \rho_{x x_0} \rho_{x_0 y} > 0$. ∎

From the above two lemmas [5], the following holds.

Lemma 4. *The Markov chain defined by (9) has a unique stationary distribution.*

Let π denote the stationary distribution. Let $S_l \subseteq S$ denote the subset of states where $V(s) \geq l$. We rewrite $\sum_{s \in S_l} \pi(s)$ as follows:

$$\sum_{s \in S_l} \pi(s) \overset{(a)}{=} \sum_{\tilde{s} \in S} \pi(\tilde{s}) \text{Pr}_{\tilde{s}}^l(S_l).$$

In (a), the quantity $\text{Pr}_{\tilde{s}}^l(S_l)$ denotes the probability of hitting a state in S_l in l steps starting from the state $\tilde{s}$. We now derive a bound on $\text{Pr}_{\tilde{s}}^l(S_l)$ that is independent of $\tilde{s}$. Hitting a state $s \in S_l$ implies that $s_j \geq l$ for some $j \in 0, \ldots, D-1$. Without loss of generality let $Z(0), \ldots Z(l-1)$ denote the indicator sequence of channel erasures over the l steps. It follows that for some k, $Z(k+iD) = 1$ for all i s.t. $0 \leq k + iD \leq l-1$. Indeed, from (9), the lag at any time can exceed l only if the any packet is re-sent over the past l time slots, i.e., the packet suffers erasure at least $\lfloor \frac{l}{D} \rfloor$ times. It therefore follows that

$$\text{Pr}_{\tilde{s}}^l(S_l) \leq \text{Pr}(\cup_k \{\forall i \text{ s.t. } 0 \leq k + iD \leq l-1 : Z(k+iD) = 1\})$$
$$= 1 - \prod_{k=0}^{D-1} (1 - \text{Pr}\{\forall i \text{ s.t. } 0 \leq k + iD \leq l-1 : Z(k+iD) = 1\})$$
$$= 1 - (1 - p^{\lfloor \frac{l}{D} \rfloor})^D, \text{ and}$$
$$\pi(S_l) = \sum_{\tilde{s} \in S} \pi(\tilde{s}) \text{Pr}_{\tilde{s}}^l(S_l) \leq 1 - (1 - p^{\lfloor \frac{l}{D} \rfloor})^D.$$

From the bound on $\pi(S_l)$ and $\mathbb{E}V(i) = \sum_{l=1}^{\infty} \pi(S_l)$, we

have

$$\mathbb{E}V(i) \leq \sum_{l=1}^{\infty} \left(1 - (1 - p^{\lfloor \frac{l}{D} \rfloor})^D\right) = D \sum_{k=1}^{+\infty} \left(1 - (1 - p^k)^D\right)$$

$$= D \left[\sum_{k=1}^{D} \binom{D}{k} (-1)^{k+1} \frac{p^k}{1 - p^k} \right]$$

From the above relation and from (8), it follows that

$$\lim_{i \to \infty} \frac{1}{i} \sum_{j=0}^{i} E(j) \leq D \left[\sum_{k=1}^{D} \binom{D}{k} (-1)^{k+1} \frac{p^k}{1 - p^k} \right] + 1 \quad (10)$$

Theorem 1 now follows from (7) and (10). ∎

Remark 5. *The average lag is bounded if $\lambda < 1 - p$. The dependence of the bound (3) is $O(D \log(D))$ as D gets large.*

IV. ONLINE CODING

The performance of retransmission based schemes deteriorates with increasing delay in feedback (see Fig. 5). This motivates examining coding-based schemes. The particular scheme we consider is the "Drop-when-decoded" scheme proposed in [1]. The packet sent at any time is a linear combination of all packets that are currently in the queue, i.e., packets which have not been decoded by the receiver. Feedback is used to obtain this latter information. The coded packet is formed by computing a random linear combination of all packets currently in the queue. If the field size is large, the sent packet is (with high probability) linearly independent to the previously sent packets. We therefore ignore the probability that the coded packet is not innovative. At the receiver, the packets in queue at the sender are unknowns, and each received linear combination is an equation in these unknowns. Decoding is possible whenever the number of linearly independent equations is equal to the number of unknowns.

We reuse some notation from the previous section, namely, $A(i)$ is the indicator of packet arrival at the transmitter at time i and $Z(i)$ is the indicator of the event that the channel is OFF in time-slot i. We introduce variable $S(i) = 1 - Z(i)$ which is an indicator that the channel is ON in time i. We define the following Markov Chain on $\mathbb{Z}^2$ with the state space described by variables (L, R). $(L(0), R(0)) = (0, 0)$ and successive transitions are governed by: $(L(i+1), R(i+1)) =$

$$\begin{cases} (L(i), R(i)) & A(i+1) = 0, S(i+1) = 0 \\ (L(i)+1, R(i)) & A(i+1) = 1, S(i+1) = 0 \\ (L(i), R(i)+1) & A(i+1) = 0, S(i+1) = 1, \\ & L(i) - R(i) > 1 \\ (0,0) & A(i+1) = 0, S(i+1) = 1, \\ & L(i) - R(i) = 0, 1 \\ (L(i)+1, R(i)+1) & A(i+1) = 1, S(i+1) = 1, \\ & (L(i), R(i) \neq (0,0)) \\ (0,0) & A(i+1) = 1, S(i+1) = 1, \\ & (L(i), R(i) = (0,0)) \end{cases}$$
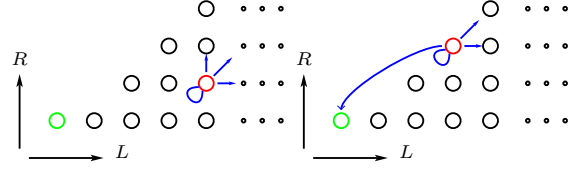
(11)

Fig. 3. Transitions for $L - R > 1$ (left) and $L - R = 1$ (right)

The chain is illustrated in Figure 3. The quantity $L(i)$ denotes the lag (the number of packets not decoded by the receiver) at time i, while the quantity $R(i)$ denotes the number of received packets *since* the last time the receiver was able to decode packets. The lag increases with every arrival until $R(i) = L(i)$, at which point (the receiver is able to decode) it resets to 0 along with $R(i)$. We also define the (more familiar) Markov chain corresponding to the quantity $L(i) - R(i)$. If $L(i) - R(i) = k$, then $L(i+1) - R(i+1) =$

$$\begin{cases} k & A(i) = 0, S(i) = 0 \text{ or } A(i) = 1, S(i) = 1 \\ k+1 & A(i) = 1, S(i) = 0 \\ k-1 & A(i) = 0, S(i) = 1, k > 1 \\ 0 & A(i) = 0, S(i) = 1, k = 0 \end{cases}$$

(12)

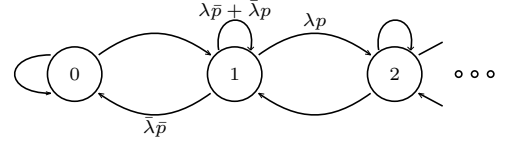The chain is illustrated in Figure 4. Let N denote the first time

Fig. 4. The $L - R$ Markov Chain

to reach the state $(0, 0)$ from the state $(1, 0)$ for the Markov chain (L, R) described by (11) (alternately first time to reach state 0 from 1 for the Markov Chain $L - R$ described by (12)) and let $C(N)$ denote the cumulative number of arrivals in this interval. We have the following theorem.

Theorem 6.

$$\bar{L} = \frac{\mathbb{E}[NC(N)] - \frac{\lambda}{2}(\mathbb{E}[N^2] + \mathbb{E}[N])}{\mathbb{E}[N] + \frac{1}{\lambda p}} \quad (13)$$

where

$$\mathbb{E}[N] = \frac{1}{1 - p - \lambda}, \quad \mathbb{E}[N^2] = \frac{1 - \lambda^2 - p^2}{(1 - p - \lambda)^3}, \text{ and}$$

$$\mathbb{E}[NC(N)] = \frac{(1-p)(1-\lambda)(1-p+\lambda)}{(1-p-\lambda)^3} \quad (14)$$

Proof: We compute the time averaged lag as

$$\bar{L}(i) = \frac{1}{i} \sum_{j=0}^{i} L(j) \overset{(a)}{=} \frac{1}{i} \sum_{j=0}^{i} (\min(i, N(j)) - j) A(j)$$

$$\overset{(b)}{\leq} \frac{1}{i} \sum_{j=0}^{i} (N(j) - j) A(j) \quad (15)$$

In the above, $N(j) = \min\{l \geq j : (L(l), R(l)) = (0,0)\}$, i.e., $N(j)$ is the time slot after time j when the state $(0, 0)$ is

first visited. The reason for the equality (a) is the following: A packet that arrives at time j contributes to the lag over all the future times that the packet is not decoded. This time is the interval $[j, N(j) - 1]$ (At time $N(j)$, the state $(0,0)$ is reached and therefore the packet does not contribute to the lag at time $N(j)$). It might be that $N(j)$ exceeds the length of time, i, under consideration. In this case the contribution to the lag of the packet arriving at time j is $i - j$. The inequality (b) follows by upperbounding $\min(i, N(j))$ with $N(j)$. Let $j_1, j_2 \ldots$ represent the time slots corresponding to the transition from $(0,0) \to (1,0)$, i.e., $(L(j_l-1), R(j_l-1)) = (0,0)$ and $(L(j_l), R(j_l)) = (1,0)$. We refer to the interval $[j_l, j_{l+1})$ as an epoch. Let $e(i)$ denote the number of epochs in the interval $[0, i]$. In particular, the interval from $[0, i]$ splits as $[0, j_1) \cup [j_1, j_2) \cdots \cup [j_{e(i)}, i]$. Thus it is assumed that $j_{e(i)} \leq i < j_{e(i)+1}$. It holds that

$$\frac{1}{i} \sum_{j=0}^{i} (N(j) - j)A(j) \overset{(c)}{\leq} \frac{1}{i} \sum_{l=1}^{e(i)} \sum_{j=j_l}^{N(j_l)} (N(j_l) - j)A(j)$$

$$\overset{(d)}{\leq} \frac{\sum_{l=1}^{e(i)} \sum_{j=j_l}^{N(j_l)} (N(j_l) - j)A(j)}{\sum_{l=1}^{e(i)-1} (j_{l+1} - j_l)} \quad (16)$$

The first inequality (c) holds since, in the numerator, we are adding the lags induced by the packets potentially arriving in the interval $[i + 1, N(j_{e(i)})]$. The inequality (d) holds since we are undercounting i as the sum of the first $e(i) - 1$ epoch times. We note that for each epoch $[j_l, j_{l+1})$, the sum lag in the epoch given by $\sum_{j=j_l}^{N(j_l)} (N(j_l) - j)A(j)$ as well as the duration of the epoch $(j_{l+1} - j_l)$ is i.i.d.

Consider the epoch duration term $j_{l+1} - j_l$. We compute its expectation as

$$\mathbb{E}[j_{l+1} - j_l] = \mathbb{E}[N(j_l) - j_l] + \mathbb{E}[j_{l+1} - N(j_l)]$$
$$= \mathbb{E}[N] + \mathbb{E}[j_{l+1} - N(j_l)]$$

We now compute the expected accumulated lag in the interval $[j_l, j_{l+1})$, $\mathbb{E}[\sum_{j=j_l}^{N(j_l)} (N(j_l) - j)A(j)]$. For convenience, we drop the dependence on l and assume $j_l = 0$ in the following computation.

$$\mathbb{E}[\sum_{j=0}^{N} (N - j)A(j)] = \mathbb{E}[N \sum_{j=0}^{N} A(j) - \sum_{j=0}^{\infty} j 1_{N \geq j} A(j)]$$

$$\overset{(a)}{=} \mathbb{E}[NC(N)] - \sum_{j=0}^{\infty} j \mathbb{E}[A(j)] \mathbb{E}[1_{N \geq j}]$$

$$= \mathbb{E}[NC(N)] - \frac{\lambda}{2} (\mathbb{E}[N^2] + \mathbb{E}[N]).$$

The equality (a) follows from the independence of $A(j)$ and $1_{N \geq j}$ and by defining $C(N)$ to be the cumulative number of arrivals in the epoch. Combining (15) and (16) and taking the limit of $\bar{L}(i)$ as $i \to \infty$, we have

$$\lim_{i \to \infty} \bar{L}(i) \leq \lim_{i \to \infty} \left(\frac{\sum_{l=1}^{e(i)} \sum_{j=j_l}^{N(j_l)} (N(j_l) - j)A(j)}{e(i)} \frac{e(i)}{e(i) - 1} \right.$$

$$\left. \times \frac{e(i) - 1}{\sum_{l=1}^{e(i)-1} (j_{l+1} - j_l)} \right)$$

$$= \frac{\mathbb{E}[NC(N)] - \frac{\lambda}{2} (\mathbb{E}[N^2] + \mathbb{E}[N])}{\mathbb{E}[N] + \frac{1}{\lambda p}} \quad (17)$$

where the last equality follows since $e(i) \to \infty$ as $i \to \infty$ due to the positive recurrence of the Markov Chain and from $\lim_{i \to \infty} x_i/y_i = \lim_{i \to \infty} x_i / \lim_{i \to \infty} y_i$ if both y_i and $\lim_{i \to \infty} y_i$ is non-zero. Similarly one can prove the reverse inequality (proof omitted for lack of space)

$$\lim_{i \to \infty} \bar{L}(i) \geq \frac{\mathbb{E}[NC(N)] - \frac{\lambda}{2} (\mathbb{E}[N^2] + \mathbb{E}[N])}{\mathbb{E}[N] + \frac{1}{\lambda p}} \quad (18)$$

The statement of the theorem now follows from (17) and (18). The proof of the analytical expressions (14) (omitted for lack of space) follows from skip-free property of chain and noting that the transition probabilities don't depend on the state. ∎

V. DISCUSSION

We had briefly mentioned alternatives to Max-strat, namely Min-strat and Div-strat. The Figure (5) is a simulated plot of the average lag as a function of delay for four strategies: Div-strat, Max-strat and Min-strat and the online coding scheme. Observe that when the delay in feedback is low,
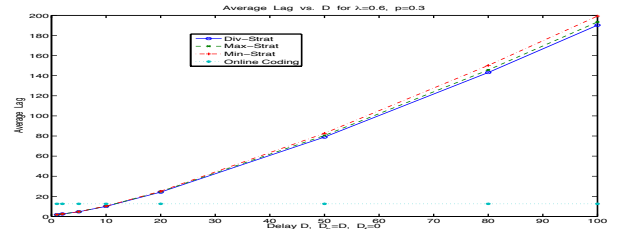
Fig. 5. Average lag versus D for $\lambda = 0.6, p = 0.3$

retransmission schemes have lesser average lag than the coding scheme. On the other hand, the performance of retransmission schemes degrades with delay, and beyond a certain delay it is more advantageous to use the coding scheme. The implicit assumption here is that $D_f = 0, D_b = D$, i.e., all delay is attributed to the reverse link. In Theorem 1, we have shown that the dependence of the average lag on D for the Max-strat strategy is no more than $O(D^2)$. We do not plot this bound (3) in Fig. 5 as it is very loose for this case.

There are a number of open issues to be resolved. The strategy Max-strat was analyzed for its relative amenability to analysis. It would be of interest to analyze Div-strat and indeed, to determine what is the optimal retransmission strategy for this problem. Even for Max-strat, we believe that the $O(D^2)$ upper bound is loose and could be sharpened, so it is important to determine lower bounds to the average lag.

REFERENCES

[1] J. Sundararajan, D. Shah, M. Medard, "Feedback-based online network coding", submitted to *IEEE Trans. Info. Theory*, Apr 2009.

[2] A. Sahai, "Why delay and block length are not the same thing for channel coding with feedback", *ITA, Invited paper*, Feb 2006.

[3] R. N. Swamy and T. Javidi, " Delay Analysis of Block Coding over a Noisy Channel with Limited Feedback," in *Asilomar Conference on Signals, Systems and Computers*, 2008.

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

[5] R. Durrett, Probability: Theory and Examples, 3rd Edition, Duxbury press.