Network coding broadcast delay on erasure channels

Nan Xie and Steven Weber
Drexel University, Dept. of ECE, Philadelphia, PA 19104, USA
nx23@drexel.edu, sweber@coe.drexel.edu

***Abstract*— We consider a broadcast erasure channel where a single transmitter broadcasts to M receivers, where the channel between the transmitter and each receiver is an independent erasure channel with success probabilities $\mathbf{q} = (q_1, \ldots, q_M)$. Prior work on this model by Cogill, Shrader and Ephremides (2011) established bounds on the expected delays incurred in transmitting a collection of C packets to each of the M receivers using both scheduling and random linear network coding (RLNC). In this work we extend these results in several ways. First, we provide exact expressions and tight upper and lower bounds for the expected delay under uncoded transmissions (UT) for arbitrary q. Second, we provide a recurrence relationship for the expected delay under RLNC that allows one to compute the exact delay for any finite C. Third, we prove convergence in probability and in rth mean of the normalized expected delay under RLNC as $C \to \infty$ to $1/\min_j q_j$, with the implication that the normalized delay is solely a function of the quality of the worst erasure channel. Finally, we note our results are of general interest in probability theory as they address the moments of the maximum of (sums of) independent geometric random variables (UT, RLNC) and the convergence of moments of functions of sums of independent geometric random variables (RLNC).**

***Index Terms*— erasure channel; broadcast channel; network coding; scheduling; delay.**

I. Introduction

In this paper we will focus on the expected number of time slots required to broadcast a collection of C packets to M receivers, where each transmitter to receiver channel is an independent erasure channel with success probabilities $\mathbf{q} = (q_1, \ldots, q_M)$ (see Fig. 1). In particular, the delay associated with the transmission is the number of time slots until each receiver has all C packets. We consider two approaches to achieving this goal, namely uncoded transmissions (UT) and random linear network coding (RLNC).

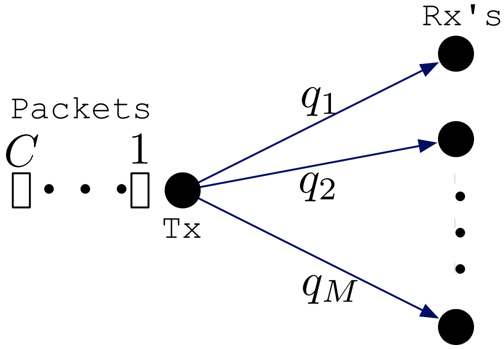

Fig. 1. The heterogeneous broadcast erasure channel consists of a transmitter and M receivers connected over independent erasure channels with success probabilities $\mathbf{q} = (q_1, \ldots, q_M)$. Of interest is the time required for all M receivers to receive all C packets. The homogeneous channel has $q_j = q$.

Our original motivations come from the recent papers by Cogill, Shrader and Ephremides [1], [2], [3] ([1] in particular; also see references therein for related work) where the authors consider a base station serving a collection of receivers. The base station must select one of several flows for transmission at each time slot, where each flow is separately queued at the base station and each flow has a possibly distinct set of receivers, and each receiver is connected to the base station over an independent erasure channel. Their focus is on contrasting the delay associated with broadcasting a collection of C packets under $i)$ scheduling and $ii)$ RLNC. Scheduling means selecting a flow and transmitting the head of line packet from the flow's queue, and RLNC means forming random linear combinations of the first C packets within a flow's queue. They do a comprehensive investigation under static and time-varying erasure channel models, including coding overhead considerations, and also consider whether or not it is beneficial to code across queues. Their main contribution is to establish configurations where RLNC outperforms *any* scheduling strategy, meaning there are instances where an inner bound of the stable throughput region under RLNC is strictly larger than an outer bound of that of scheduling.

Our focus in this paper is on deriving exact expressions, tight bounds, asymptotic results and recurrences for the expected delay per packet under both UT and RLNC for a simplified version of this problem. Namely, we restrict our attention to the case of a single flow, so that "scheduling" is simply retransmitting the same packet over the channel until all M receivers have successfully received the packet, which is what we mean by uncoded transmissions. In particular, packet c is retransmitted a random $Z = \max(Z_1, \ldots, Z_M)$ times where $Z_j \sim \mathrm{Geo}(q_j)$ is the geometric random variable (RV) giving the number of transmissions required until receiver j has the packet. It follows that the expected delay per packet under UT is $\mathbb{E}[Z]$. In contrast, the expected delay per packet under RLNC for a block of C packets is $\mathbb{E}[Y]/C$ where $Y = \max(Y_1, \ldots, Y_M)$ and $Y_j = \sum_{k=1}^{C} X_{jk}$ is the sum of C independent geometric RVs. That is, Y_j is the number of independent trials required to obtain C successes where success k has probability q_{jk} for $k \in [C]$. The intuition for why RLNC is superior to UT in terms of expected delay per packet is that it removes some of the inefficiency associated with UT. Namely, under UT receivers that are done with a packet can only "wait" for all the other receivers to receive the packet before moving on to the next packet. In contrast, under RLNC, any receiver that has not yet received C coded combinations can still benefit from the transmission by possibly receiving an additional innovative combination, thus partially reducing the inefficiency of UT. This also explains why the delay per packet decreases in C.

Our contributions in this paper are:

1) Exact expressions and tight upper and lower bounds on the expected delay under UT for the heterogenous broadcast erasure channel; more generally these expressions and bounds hold for the moments of maximum of independent heterogenous geometric RVs.
2) Recurrence expression for the expected delay under RLNC for the heterogeneous broadcast erasure channel; more generally this recurrence holds for the moments of the maximum of independent heterogenous negative binomial RVs.
3) Furthermore, under certain assumptions (partially required to apply the weak law of large numbers), we establish convergence in probability and in rth mean of the random delay per packet under RLNC as the number of packets C grows large to $1/\min_j q_j$. This implies that the asymptotic (in C) expected delay per packet under RLNC for broadcasting to M receivers is the same as transmitting to a single receiver over an erasure channel with success probability $\min_j q_j$. Note convergence in mean was established in [1] using a different proof technique.

The rest of this paper is structured as follows. In §II we derive bounds and an expression for the exact delay of UT. In §III we establish a recurrence and convergence results in C for RLNC. In §IV we compare UT and RLNC and exhibit the delay advantage of the latter. §V concludes the paper.

II. Delay under uncoded transmission (maximum of independent geometric RVs)

Cogill, Shrader and Ephremides [1] provide a lower bound (Lemma 4 in [1]) on the expected delay for scheduling for the special case of a homogeneous broadcast channel, i.e., $q_j = q$ for all $j \in [M]$. Recall that scheduling across flows in [1] reduces to serial uncoded transmission of packets in our context of a single flow. Under this restriction, we are able to extend their result by establishing both upper and lower bounds (the bounds have a difference of one) for the heterogeneous channel with arbitrary $\mathbf{q}$ (Prop. 1). We also provide expressions for the exact expected delay (Prop. 2) which is then used to establish that RLNC always outperforms UT in §IV (Prop. 7).

A. Upper and lower bounds on expected delay $\mathbb{E}[Z]$ under UT

Recall $Z \equiv \max(Z_1, \ldots, Z_M)$ represents the delay per packet under UT, where $Z_1, \ldots, Z_M$ are independent and each $Z_j \sim \mathrm{Geo}(q_j)$ represents the delay per packet under UT for receiver j. Here and throughout the paper we employ the notation $[M]_k$ for the set of all subsets of $[M] \equiv \{1, \ldots, M\}$ of size k, and define the function $\phi(q) \equiv -\ln(1-q)$.

Proposition 1: The expectation of the maximum of M independent geometric RVs $\mathbb{E}[Z]$ has lower and upper bounds

$$\psi(\mathbf{q}) \leq \mathbb{E}[Z] \leq \psi(\mathbf{q}) + 1, \tag{1}$$

where

$$\psi(\mathbf{q}) \equiv \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \left(\sum_{j \in A} \phi(q_j) \right)^{-1}. \tag{2}$$

Proof: See Appendix. ∎

Remark 1: Recall the classic (sequential) coupon-collector problem ([4] §3.6) asks for the expected time to receive all of M coupons, when in each time slot a new coupon is selected uniformly at random from $[M]$. The broadcast delay problem can be considered as a "parallel" coupon-collector problem in that at each time slot each of the M coupons is received independently with probabilities $\mathbf{q}$, i.e., a successful reception by receiver j corresponds to collecting a coupon of type j.

Remark 2: The above remark interpreted $\mathbb{E}[Z]$ as the delay in a parallel coupon collector problem with probabilities $\mathbf{q}$. In addition, the lower bound $\psi(\mathbf{q})$ may also be interpreted as the "normalized" delay in a parallel coupon collector problem with probabilities $\tilde{\boldsymbol{\phi}} = (\tilde{\phi}_1, \ldots, \tilde{\phi}_M)$ where each $\tilde{\phi}_j = \phi_j / \sum_{j=1}^{M} \phi_j$. This interpretation is discussed in [5]; for completeness we briefly summarize it here. Observe

$$\psi(\mathbf{q}) \cdot \left(\sum_{j=1}^{M} \phi_j \right) = \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \frac{1}{\sum_{j \in A} \tilde{\phi}_j}. \tag{3}$$

At each time slot, each type of coupon $j \in [M]$ is collected independently with probability $\tilde{\phi}_j$. Define $(\tilde{Z}_1, \ldots, \tilde{Z}_M)$ with $\tilde{Z}_j \sim \mathrm{Geo}(\tilde{\phi}_j)$ as the independent times to collect each coupon j. Then $\tilde{Z} \equiv \max(\tilde{Z}_1, \ldots, \tilde{Z}_M)$ is the time to collect all the coupons. Define for each size k subset $A \in [M]_k$, the RV $\tilde{Z}^A \equiv \min_{j \in A} \tilde{Z}_j$ as the time required to collect at least one coupon in $A \in [M]_k$. Due to the independence of $(\tilde{Z}_1, \ldots, \tilde{Z}_M)$, $\tilde{Z}^A \sim \mathrm{Geo}\left(\sum_{j \in A} \tilde{\phi}_j \right)$. According to the min-max identity ([5] §4.6),

$$\mathbb{E}[\tilde{Z}] = \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \mathbb{E}[\tilde{Z}^A], \tag{4}$$

where the RHS of (4) equals the RHS of (3). Thus we see the expected delay $\mathbb{E}[\tilde{Z}]$ is given by (3), i.e., a scaling of the lower bound $\psi(\mathbf{q})$.

Specializing Prop. 1 to the homogeneous channel case with $q_j = q$ for all $j \in [M]$ yields simpler expressions for the lower and upper bounds after using the combinatorial identity $\sum_{k=1}^{M} (-1)^k \binom{M}{k} (-\frac{1}{k}) = \sum_{j=1}^{M} \frac{1}{j}$. We observe the lower bound in Cor. 1 is the same as Lemma 4 in [1].

Corollary 1: The expectation of the maximum of iid geometric RVs with parameter q is bounded as:

$$\frac{H_M}{\phi(q)} \leq \mathbb{E}[Z] \leq \frac{H_M}{\phi(q)} + 1.$$

for $H_M \equiv 1 + 1/2 + \cdots + 1/M$ the Mth harmonic number.

B. Exact expressions for moments of the delay Z under UT

In this subsection we obtain an exact delay expression via min-max identity [5] (see Remark 2 and (4)) and the property that the minimum of independent geometric RVs is also a geometric RV.

Proposition 2: The expectation of the maximum of M independent geometric RVs with parameters $\mathbf{q}$ is

$$\mathbb{E}[Z] = \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \left(1 - \prod_{j \in A} (1 - q_j) \right)^{-1}. \tag{5}$$

In the homogeneous case ($q_j = q$), this expression simplifies:

$$\mathbb{E}[Z] = \sum_{k=1}^{M} (-1)^{k+1} \binom{M}{k} \frac{1}{1-(1-q)^k}. \quad (6)$$

Proof: By independence, for any size k subset $A \in [M]_k$, the RV $Z^A \equiv \min_{j \in A} Z_j$ is geometric with success probability $q^A \equiv 1 - \prod_{j \in A}(1-q_j)$ and thus $\mathbb{E}[Z^A] = 1/q^A$. Applying the min-max identity (4) we have:

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \mathbb{E}[Z^A] \\ &= \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \left(1 - \prod_{j \in A}(1-q_j)\right)^{-1} . (7) \end{aligned}$$

■

It is clear that the proof technique above extends to higher moments. The next proposition gives the second moment for Z. Recall if $Z_j \sim \mathrm{Geo}(q_j)$ that $\mathbb{E}[Z_j^2] = (2-q_j)/q_j^2$.

Proposition 3: The second moment of Z is

$$\mathbb{E}[Z^2] = \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \frac{1 + \prod_{j \in A}(1-q_j)}{\left(1 - \prod_{j \in A}(1-q_j)\right)^2}. \quad (8)$$

Proof:

$$\begin{aligned} \mathbb{E}[Z^2] &= \mathbb{E}[(\max(Z_1, \ldots, Z_M))^2] \\ &= \mathbb{E}[\max(Z_1^2, \ldots, Z_M^2)] \\ &= \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \mathbb{E}\left[\min_{j \in A} Z_j^2\right] \\ &= \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \mathbb{E}\left[(Z^A)^2\right] \quad (9) \end{aligned}$$

Observe $Z^A \sim \mathrm{Geo}(q^A)$ with $\mathbb{E}[(Z^A)^2] = (2-q^A)/(q^A)^2$ and $q^A = 1 - \prod_{j \in A}(1-q_j)$. Substitution yields (8). ■

Before closing this section we comment briefly on related work. First, recent work [6] has extended the min-max identity ([5] §4.6) to a more general "sorting" identity in a non-probabilistic setting. Second, we point out there is a large literature bounding the expectation of the maximum of random variables. A representative result along these lines is found in [5] (§4.6) which states that for general (continuous or discrete, dependent, heterogeneously distributed) non-negative random variables $(X_1, \ldots, X_M)$:

$$\mathbb{E}[\max(X_1, \ldots, X_M)] \leq c + \sum_{i=1}^{M} \int_c^{\infty} \mathbb{P}(X_i > y) \mathrm{d}y, \quad (10)$$

where the bound is tightest when

$$\sum_{i=1}^{M} \mathbb{P}(X_i > c^*) = 1. \quad (11)$$

Finally, we observe there are also recent results on the expectation of the maximum of M iid geometric RVs by Eisenberg [7] (2008) (Corollary 2). An example result is for $(Z_1, \ldots, Z_M)$ iid geometric RVs with parameter q and $Z = \max(Z_1, \ldots, Z_M)$,

$$\mathbb{E}[Z] = \frac{1}{2} + \frac{H_M}{q} - \sum_{m \neq 0} \frac{1}{2\pi m \mathrm{i}} \prod_{k=1}^{M} \left(1 + \frac{2\pi m \mathrm{i}}{kq}\right)^{-1}, \quad (12)$$

where $\mathrm{i} = \sqrt{-1}$. This result improves an earlier result by Szpanknowski and Rego [8] (1990) (Eq. 2.8). Observe the second sum involves an infinite number of terms.

III. Delay under RLNC (Maximum of Sums of Geometric RVs)

Under RLNC at each time slot the transmitter forms a new random linear combination of the C packets which are then broadcast to the receivers. A receiver is able to decode all C packets once it has received C linearly independent combinations. As with UT, the delay is the number of time slots until all receivers have all C packets, i.e., all M receivers have received C linearly independent combinations.

Remark 3: The random delay for a block of C packets under RLNC is denoted $Y(C)$ (or Y when C is understood) where $Y = \max(Y_1, \ldots, Y_M)$. The expected delay per packet is then $\mathbb{E}[Y(C)]/C$. Each $Y_j = \sum_{k=1}^{C} X_{jk}$ is the sum of independent geometric RVs $(X_{j1}, \ldots, X_{jC})$ where each $X_{jk} \sim \mathrm{Geo}(q_{jk})$. The $M \times C$ matrix $\mathbf{Q}$ with entries q_{jk} for $j \in [M]$ and $k \in [C]$ holds the success probability indexed by receiver (j) at "state" (k). We will use the phrase *state-independent* RLNC when $q_{jk} = q_j$ and *state-dependent* RLNC for the more general case. Motivation for consideration of time-varying receiver state is given in the following remark.

Remark 4: Our motivation for considering $\{X_{jk}\}$ that are heterogeneous in k as well as j is the fact that the probability of the random linear combination sent by the transmitter being linearly independent of the combinations received thus far by receiver j, say, is decreasing in the number of linearly independent combinations already received by j. In particular, a common model for this phenomenon is

$$q_{jk} = (1 - d^{k-1-C}) q_j, \quad (13)$$

where q_j is the time-invariant channel erasure probability for channel j, d is the field size, and $1 - d^{k-1-C}$ is the probability that the linear combination sent by the transmitter is in fact independent of the $k-1$ linear combinations already received by receiver j. We prefer to use the more general setting of an arbitrary matrix $\mathbf{Q}$, noting that (13) is a particular instantiation of $\mathbf{Q}$ of interest.

Remark 5: Observe that RLNC reduces to UT for the case $C = 1$.

In this section we provide i) exact expressions for $\mathbb{E}[Y(C)]$ in §III-A (Prop. 4) which involves a sum over an infinite number of terms, ii) a recurrence for $\mathbb{E}[Y(C)]$ in §III-B (Prop. 5) which allows for easy computation of $\mathbb{E}[Y(C)]$ using a finite number of steps, and iii) a proof that the RV $Y(C)/C \to (\min_j q_j)^{-1}$ as $C \to \infty$ both in probability and in rth mean in §III-C (Prop. 6).

A. Exact expressions of the expected delay $\mathbb{E}[Y]$ under RLNC

Prop. 4 gives an expression for the exact delay $\mathbb{E}[Y(C)]$ under RLNC for the state-dependent case as well as two equivalent expressions for $\mathbb{E}[Y(C)]$ under RLNC for the state-independent case (recall Remark 3).

Proposition 4: For the state-dependent case:

$$\mathbb{E}[Y] = C + \sum_{n=C}^{\infty} \left(1 - \prod_{j=1}^{M} \sum_{t=C}^{n} \sum_{\boldsymbol{\alpha} \in \mathcal{A}_t} \prod_{k=1}^{C} (1 - q_{jk})^{\alpha_k - 1} q_{jk} \right), \tag{14}$$

where $\mathcal{A}_t$ is the finite set of all C-vectors of positive integers that sum to t, i.e.,

$$\mathcal{A}_t = \left\{ \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_C) \ : \ \alpha_k \in \mathbb{N}, \ \sum_{k=1}^{C} \alpha_k = t \right\}. \tag{15}$$

For the state-independent case, (14) simplifies to

$$\mathbb{E}[Y] = C + \sum_{n=C}^{\infty} \left(1 - \prod_{j=1}^{M} \sum_{t=C}^{n} \binom{t-1}{C-1} (1 - q_j)^{t-C} q_j^C \right). \tag{16}$$

An alternative expression for the state-independent case is

$$\mathbb{E}[Y] = C + \sum_{n=C}^{\infty} \left(1 - \prod_{j=1}^{M} I_{q_j}(C, n - C + 1) \right), \tag{17}$$

where in the last equation, $I_x(\alpha, \beta)$ is the *regularized incomplete Beta function* which is both the CDF of the beta distribution as well as the tail probability of the binomial RV $W \sim \mathrm{Bin}(n, p)$:

$$\mathbb{P}(W > k) = \sum_{r=k+1}^{n} \binom{n}{r} p^r (1-p)^{n-r} = I_p(k+1, n-k). \tag{18}$$

Proof: We first derive (14). Start by using the well-known expression for expectation of a RV with support $\{C, C+1, \ldots\}$ in terms of its complementary CDF (CCDF).

$$\begin{aligned} \mathbb{E}[Y] &= C + \sum_{n=C}^{\infty} \mathbb{P}(Y > n) \\ &= C + \sum_{n=C}^{\infty} (1 - \mathbb{P}(Y \le n)) \\ &= C + \sum_{n=C}^{\infty} \left(1 - \prod_{j=1}^{M} \mathbb{P}(Y_j \le n) \right) \end{aligned} \tag{19}$$

Next, fix $j \in [M]$ and observe $\{Y_j \le n\}$ means that there are C successes in the first t trials, for some $t \in \{C, \ldots, n\}$. Conditioning on t, the C successes occurred over trials $\{1, \ldots, t\}$ with the number of trials between successive successes given by a C-vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_C)$ such that $\alpha_k \in \mathbb{N}$ and $\alpha_1 + \cdots + \alpha_C = t$. Fixing the particular sequence of successes and failures over trials $\{1, \ldots, t\}$, that sequence of outcomes has probability $\prod_{k \in [C]} (1 - q_{jk})^{\alpha_k - 1} q_{jk}$. This yields (14).

To obtain (16) from (14) set $q_{jk} = q_j$ and observe

$$\begin{aligned} \sum_{\boldsymbol{\alpha} \in \mathcal{A}_t} \prod_{k=1}^{C} (1 - q_j)^{\alpha_k - 1} q_j &= \sum_{\boldsymbol{\alpha} \in \mathcal{A}_t} q_j^C (1 - q_j)^{t-C} \\ &= q_j^C (1 - q_j)^{t-C} |\mathcal{A}_t|, \end{aligned} \tag{20}$$

where $|\mathcal{A}_t|$ is the cardinality of $\mathcal{A}_t$, i.e., the number of C-vectors of positive integers that sum to t. It can be shown that $|\mathcal{A}_t| = \binom{t-1}{C-1}$; this yields (16).

To derive (17) which holds for the state-independent case, observe Y_j is a negative binomial RV, i.e., the number of trials required to obtain C successes where each trial is successful with probability q_j. Let $W_j \sim \mathrm{Bin}(n, q_j)$ be a binomial RV counting the number of successes in n trials, each trial having success probability q_j. Observe the equivalence of the events $\{Y_j > n\} = \{W_j < C\}$. Taking probabilities of both sides, and applying (18) yields:

$$\begin{aligned} \mathbb{P}(Y_j > n) &= \mathbb{P}(W_j < C) = 1 - \mathbb{P}(W_j \ge C) \\ &= 1 - \sum_{t=C}^{n} \mathbb{P}(W_j = t) \\ &= 1 - \sum_{t=C}^{n} \binom{n}{t} q_j^t (1 - q_j)^{n-t} \\ &= 1 - I_{q_j}(C, n - C + 1) \end{aligned} \tag{21}$$

Starting from (19) and substituting (21) yields (17):

$$\begin{aligned} \mathbb{E}[Y] &= C + \sum_{n=C}^{\infty} \left(1 - \prod_{j=1}^{M} (1 - \mathbb{P}(Y_j > n)) \right) \qquad (22) \\ &= C + \sum_{n=C}^{\infty} \left(1 - \prod_{j=1}^{M} I_{q_j}(C, n - C + 1) \right). \end{aligned}$$

■

B. Recurrence for the moments of delay Y under RLNC

In this subsection, we offer a recurrence equation for the RV Y that permits calculation (by computer or a patient human) of the exact value for, say $\mathbb{E}[Y(C)]$, using only a finite number of terms. In fact it is convenient to generalize our setting slightly, in that we now suppose that receiver j requires reception of C_j packets; previous to this we have assumed $C_j = C$ for all $j \in [M]$. Let $\mathbf{C}^0 = (C_1^0, \ldots, C_M^0)$ be the M-vector of required number of successes for each receiver. The recurrence will be in terms of the generic vector $\mathbf{C} \le \mathbf{C}^0$ (component-wise), interpreted as the number of successes left to go for each receiver, as explained below.

We introduce some shorthand notation. First, define $[M(\mathbf{C})] \equiv \{j \in [M] : C_j > 0\}$ as the set of active receivers, i.e., those still requiring an additional reception to complete. Second, define the probabilities of success and failure by the active receiver subsets S and $[M(\mathbf{C})] \setminus S$ respectively, to be:

$$\begin{aligned} q(S, \mathbf{C}) &\equiv \prod_{j \in S} q_{j, C_j^0 - C_j + 1} \\ \bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C}) &\equiv \prod_{j \in [M(\mathbf{C})] \setminus S} (1 - q_{j, C_j^0 - C_j + 1}). \end{aligned} \tag{23}$$

In words, $q(S, \mathbf{C})$ is the probability of success for active receivers with indices in S where the "successes to go" C_j and initial number of successes required C_j^0 determine the state index $k = C_j^0 - C_j + 1$ for receiver j. Further, $\bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C})$ is the probability of failure for the active receivers not indexed by S, where again the state index for each such receiver j is $k = C_j^0 - C_j + 1$.

Third, define $\mathbf{1}_S$ to be the M-vector with ones in the positions indexed by S and zero elsewhere. We reiterate that in the state-dependent case with success probability matrix $\mathbf{Q}$, we have $Y = \max(Y_1, \ldots, Y_M)$ where $(Y_1, \ldots, Y_M)$ are independent with $Y_j = \sum_{k=1}^{C_j} X_{jk}$ and the $X_{jk} \sim \mathrm{Geo}(q_{jk})$ are themselves independent in j and k. In the state-independent case we have $X_{jk} \sim \mathrm{Geo}(q_j)$ for $k \in [C_j]$.

Proposition 5: For the state-dependent case the RV $Y = Y(\mathbf{C}^0)$ defined above admits the recurrence

$$Y(\mathbf{C}) = 1 + \sum_{S \subseteq [M(\mathbf{C})]} q(S, \mathbf{C}) \bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C}) Y(\mathbf{C} - \mathbf{1}_S), \tag{24}$$

with boundary condition $Y(\mathbf{0}) = 0$. For the state-independent case the RV $Y = Y(\mathbf{C}^0)$ defined above admits the recurrence

$$Y(\mathbf{C}) = 1 + \sum_{S \subseteq [M(\mathbf{C})]} q(S, \mathbf{C}^0) \bar{q}([M] \setminus S, \mathbf{C}^0) Y(\mathbf{C} - \mathbf{1}_S), \tag{25}$$

again with boundary condition $Y(\mathbf{0}) = 0$.

Proof: The recurrence is obtained by conditioning on the outcomes of the next trial starting from the given "state" $\mathbf{C}^0$ indicating the number of successes required by each receiver. The set of outcomes for the first trial is all subsets S of the active receivers $[M(\mathbf{C})]$, i.e., all possible subsets of potential successes. The probability of success by active receivers in S and failure by active receivers not in S is $q(S, \mathbf{C}) \bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C})$. The effect of successes by receivers in S is to reduce the number of required succeses by those receivers by one, i.e., $\mathbf{C} \to \mathbf{C} - \mathbf{1}_S$, thus advancing the active receivers in S to the next "state". That is, their success probability advances from q_{j,k_j} where $k_j = C_j^0 - C_j + 1$ to q_{j,k_j+1}. This gives (24). To obtain (25), observe that in the state-independent case, the probability of success by S and failure by $[M(\mathbf{C})] \setminus S$ is $\prod_{j \in S} q_j$ times $\prod_{j \in [M(\mathbf{C})] \setminus S} (1 - q_j)$, which is the same as $q(S, \mathbf{C}^0) \bar{q}([M] \setminus S, \mathbf{C}^0)$. ∎

Remark 6: The recurrence hinges critically on the fact that the X_{jk} are geometric RVs, i.e., on the memoryless property. Conditioning on the number of successes affects the state $\mathbf{C}$ but aside from that the system probabilistically restarts itself.

Remark 7: The empty set is included in the set of subsets of $[M(\mathbf{C})]$ in (24) and (25), which must be "subtracted out" to express the recurrence for $Y(\mathbf{C})$ in terms of strictly smaller vectors $\mathbf{C}' < \mathbf{C}$ (in at least one component). Doing this gives, for (24), $Y(\mathbf{C}) =$

$$\frac{1 + \sum_{S \subseteq [M](\mathbf{C}) \setminus \emptyset} q(S, \mathbf{C}) \bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C}) Y(\mathbf{C} - \mathbf{1}_S)}{1 - q(\emptyset, \mathbf{C}) \bar{q}([M(\mathbf{C})], \mathbf{C})}, \tag{26}$$

where $[M(\mathbf{C})] \setminus \emptyset$ denotes all non-empty subsets of $[M(\mathbf{C})]$.

Taking the expectations of (24) and (25) yields recurrences on $\mathbb{E}[Y(C)]$ for the state dependent and independent cases.

Corollary 2: The expectation $\mathbb{E}[Y(C)]$ for the state-dependent case admits the recurrence

$$\mathbb{E}[Y(\mathbf{C})] = 1 + \sum_{S \subseteq [M(\mathbf{C})]} q(S, \mathbf{C}) \bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C}) \mathbb{E}[Y(\mathbf{C} - \mathbf{1}_S)], \tag{27}$$

with boundary condition $\mathbb{E}[Y(\mathbf{0})] = 0$. The expectation $\mathbb{E}[Y(C)]$ for the state-independent case admits the recurrence

$$\mathbb{E}[Y(\mathbf{C})] = 1 + \sum_{S \subseteq [M(\mathbf{C})]} q(S, \mathbf{C}^0) \bar{q}([M] \setminus S, \mathbf{C}^0) \mathbb{E}[Y(\mathbf{C} - \mathbf{1}_S)], \tag{28}$$

again with boundary condition $\mathbb{E}[Y(\mathbf{0})] = 0$.

In fact it is straightforward to see that for $\mathbf{C} = C_j \mathbf{e}_j$ (the M-vector of all zeros with value C_j in position j), in the state-independent case we have the boundary condition $\mathbb{E}[Y(C_j \mathbf{e}_j)] = C_j / q_j$. That is, when there is only one receiver left to receive, the expected duration is the expectation of a negative binomial RV, which of course is C_j succcesses required times $1/q_j$ times slots on average per success.

Remark 8: A parallel recurrence holds for the RV $T = T(\mathbf{C}^0)$ where $T = \min(Y_1, \ldots, Y_M)$ is the random time until the *first* receiver, say j, receives its target number of successes C_j^0 (as opposed to counting until all receivers reach their targets). For the state-dependent case the RV $T = T(\mathbf{C}^0)$ defined above admits the recurrence

$$T(\mathbf{C}) = 1 + \sum_{S \subseteq [M]} q(S, \mathbf{C}) \bar{q}([M] \setminus S, \mathbf{C}) T(\mathbf{C} - \mathbf{1}_S), \tag{29}$$

with boundary condition $T(\mathbf{C}) = 0$ for any $\mathbf{C}$ containing one or more zeros. For the state-independent case the RV $T = T(\mathbf{C}^0)$ defined above admits the recurrence

$$T(\mathbf{C}) = 1 + \sum_{S \subseteq [M]} q(S, \mathbf{C}^0) \bar{q}([M] \setminus S, \mathbf{C}^0) T(\mathbf{C} - \mathbf{1}_S), \tag{30}$$

again with boundary condition $T(\mathbf{C}) = 0$ for any $\mathbf{C}$ containing one or more zeros.

Remark 9: In fact we can use the memoryless property to establish recurrences to compute arbitrary moments of Y, i.e., $\mathbb{E}[Y(\mathbf{C}^0)^l]$. To see this, by the memoryless property:

$$\begin{aligned} \mathbb{P}(Y(\mathbf{C}) = y \mid S \in [M(\mathbf{C})] \text{ succeed in c.t.s.}) &= \\ \mathbb{P}(1 + Y(\mathbf{C} - \mathbf{1}_S) = y). \end{aligned} \tag{31}$$

where "c.t.s." stands for *current time slot*. That is, the event that the RV $\{Y(\mathbf{C}) = y\}$ conditioned on the event of successes for active receivers S in the current time slot, is the same as the event $\{1 + Y(\mathbf{C} - \mathbf{1}_S) = y\}$, which is just an example of the common first-step analysis technique in probability. But this allows us to equate arbitrary powers l of both sides, i.e.,

$$\begin{aligned} \mathbb{P}(Y(\mathbf{C})^l = y^l \mid S \in [M(\mathbf{C})] \text{ succeed in c.t.s}) &= \\ \mathbb{P}((1 + Y(\mathbf{C} - \mathbf{1}_S))^l = y^l). \end{aligned} \tag{32}$$

On this basis, our recurrence can be generalized to one for higher moments of Y, e.g., for the state-dependent case:

$$\mathbb{E}[Y(\mathbf{C})^l] = \sum_{S \subseteq [M](\mathbf{C})} q(S, \mathbf{C}) \bar{q}([M(\mathbf{C})] \setminus S, \mathbf{C}) \mathbb{E}[(1 + Y(\mathbf{C} - \mathbf{1}_S))^l], \tag{33}$$

with boundary condition $\mathbb{E}[Y(\mathbf{0})^l] = 0$.

Remark 10: Recurrences for maximum of geometric RVs and in fact maximum of sums of geometric RVs are found in the literature. We mention in particular [8] (1990 (2.5)) which provided the inspiration for our recurrence, but in the simpler context of the expected maximum of iid geometric RVs. Further, recent work [9] (2010) on the expected delay of network coded packets routed simultaneously over two paths employed a similar recurrence to ours, but with significant differences. In particular, since in their context there is a single receiver looking for C innovative packets over $M = 2$ disjoint routes, their recurrence is univariate in the scalar C. The general problem of expressing the maximum of negative binomial RVs is addressed in [10].

C. Convergence of RLNC delay as $C \to \infty$

In this section we revert to our prior assumption that all receivers j require the same number of successes, i.e., $C_j = C$. We retain the general framework of independent RVs $(Y_1, \ldots, Y_M)$, where each $Y_j = \sum_{k=1}^{C} X_{jk}$ is state-dependent in that the $X_{jk} \sim \mathrm{Geo}(q_{jk})$ are independent but heterogeneously distributed in both j, k. The main result is Prop. 6 which asserts that $Y(C)/C \to (\min_j q_j)^{-1}$ as $C \to \infty$ in both probability and rth mean assuming the rows of the matrix $\mathbf{Q}$ holding the parameters q_{jk} are well-behaved. Normalize $(Y_1(C), \ldots, Y_M(C))$ as $(\tilde{Y}_1^C, \ldots, \tilde{Y}_M^C)$ with

$$\tilde{Y}_j^C \equiv Y_j(C)/C = \frac{1}{C} \sum_{k=1}^{C} X_{jk}, \ j \in [M]. \tag{34}$$

Similarly, normalize $Y(C)$ as $\tilde{Y}(C) \equiv Y(C)/C$.

Proposition 6: Given the $M \times C$ matrix $\mathbf{Q}$ with success probabilities q_{jk} (with C growing to infinity), suppose there exist M-vectors $\mathbf{q} = (q_1, \ldots, q_M)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_M^2)$ such that the following hold for all $j \in [M]$:

$$\begin{aligned} \lim_{C \to \infty} \frac{1}{C} \sum_{k=1}^{C} \mathbb{E}[X_{jk}] &= \frac{1}{q_j} < \infty \\ \lim_{C \to \infty} \frac{1}{C} \sum_{k=1}^{C} \mathrm{Var}(X_{jk}) &= \sigma_j^2 < \infty \end{aligned} \tag{35}$$

Then, as $C \to \infty$, $\tilde{Y}^C$ converges in probability :

$$\tilde{Y}^C \xrightarrow{\mathbb{P}} (\min_j q_j)^{-1}. \tag{36}$$

Further, fix $r \in [1, \infty)$ and suppose that each X_{jk} has uniformly bounded moments up to order $2r$. Then (35) along with this assumption ensure that, as $C \to \infty$, $\tilde{Y}^C$ converges in rth mean:

$$\tilde{Y}^C \xrightarrow{L^r} (\min_j q_j)^{-1}. \tag{37}$$

Proof: We first provide the outline of the proof. To establish convergence in probability of $\tilde{Y}^C$ we first establish convergence in probability of each component of the vector $(\tilde{Y}_1^C, \ldots, \tilde{Y}_M^C)$ via the weak law of large numbers (WLLN) for independent but not necessarily identically distributed RVs. Convergence in probability of each component of a sequence of vector-valued random variables to a limit ensures convergence in probability of the vector as a whole. Next, use the fact that convergence in probability is preserved under continuous functions, and in particular the function $\max(y_1, \ldots, y_M)$. This establishes convergence in probability of $\tilde{Y}^C$.

Next, to establish convergence in rth mean of $\tilde{Y}^C$, we first use the fact that if there exists a finite constant, say R, such that the 2nd moment of each element of the sequence is uniformly bounded by R, then the sequence itself is *uniformly integrable* (UI). Then, we employ the fact that a sequence of RVs that converges in probability and is UI with parameter r must converge in rth mean.

We now establish convergence in probability. The WLLN for independent but not necessarily identically distributed RVs (e.g., Theorem 1.1 of Chapter 7 in [11]), states (35) guarantees

$$\tilde{Y}_j^C \xrightarrow{\mathbb{P}} \mathbb{E}\left[\lim_{C \to \infty} \frac{1}{C} \sum_{k=1}^{C} X_{jk}\right] = \frac{1}{q_j}, \ j \in [M]. \tag{38}$$

Convergence in probability of each component $j \in [M]$ ensures convergence in probability of the vector as a whole:

$$(\tilde{Y}_1^C, \ldots, \tilde{Y}_M^C) \xrightarrow{\mathbb{P}} \left(\frac{1}{q_1}, \ldots, \frac{1}{q_M}\right). \tag{39}$$

Since convergence in probability is preserved under continuous functions, and $\max(y_1, \ldots, y_M)$ is a continuous function, it follows that

$$\tilde{Y}^C = \max(\tilde{Y}_1^C, \ldots, \tilde{Y}_M^C) \xrightarrow{\mathbb{P}} \max\left(\frac{1}{q_1}, \ldots, \frac{1}{q_M}\right). \tag{40}$$

This establishes convergence in probability of $\tilde{Y}^C$ in C to $1/\min_j q_j$.

We next establish convergence in rth mean. We first establish the sequence $(\tilde{Y}^C)$ in C is uniformly integrable for all $r \in [1, \infty)$. A sufficient condition for this is to show $\mathbb{E}[\left((\tilde{Y}^C)^r\right)^2] < R$ for all C. We establish this as follows:

$$\begin{aligned} (\tilde{Y}^C)^r &= \left(\max\left(\tilde{Y}_1^C, \ldots, \tilde{Y}_M^C\right)\right)^r \\ &= \max\left((\tilde{Y}_1^C)^r, \ldots, (\tilde{Y}_M^C)^r\right) \\ &\le \sum_{j=1}^{M} (\tilde{Y}_j^C)^r = \sum_{j=1}^{M} \left(\frac{1}{C} \sum_{k=1}^{C} X_{jk}\right)^r. \end{aligned} \tag{41}$$

We now establish the sequence of RVs $(\tilde{Y}^C)^r$ is UI. It suffices ([11] Thm. (4.2) in §7) to show $\mathbb{E}[(\tilde{Y}^C)^{2r}] < R$ for all C.

$$\mathbb{E}[(\tilde{Y}^C)^{2r}] \le \mathbb{E}\left[\sum_{j=1}^{M} \left(\frac{1}{C} \sum_{k=1}^{C} X_{jk}\right)^{2r}\right]. \tag{42}$$

It follows that there exists a constant $R < \infty$ such that the RHS of (42) is bounded above by R provided each X_{jk} has uniformly bounded moments up to order $2r$. Next, it is a theorem ([11] Thm. (4.1) in §7) that convergence in probability and uniform integrability with parameter r ensures convergence in L^r. ■

Remark 11: In particular for $r = 1$ it easily follows Prop. 6 that $\mathbb{E}[\tilde{Y}^C] \to (\min_j q_j)^{-1}$ as $C \to \infty$. Note [1] establishes this fact as well, but by quite different means. Namely, they derive upper and lower bounds on $\mathbb{E}[\tilde{Y}^C]$ for each C and show that these bounds converge in C to $(\min_j q_j)^{-1}$. Although our approach is in some sense cleaner, it does require certain conditions hold, as stipulated in the proposition statement. Note, however, that the convergence conditions in our proposition hold automatically in the state-independent case.

Remark 12: Indeed, for $r = 1$, L^1 convergence implies "interchanging limit and expectation" is justified, hence:

$$\lim_{C \to \infty} \mathbb{E}\left[\tilde{Y}^C\right] = \mathbb{E}\left[\lim_{C \to \infty} \tilde{Y}^C\right] = \left(\min_j q_j\right)^{-1}. \tag{43}$$

Remark 13: The convergence in probability in Prop. 6 holds for any continuous function, not just the $\max(y_1, \ldots, y_M)$ function.

Remark 14: Prop. 6 reveals something important about RLNC over the erasure broadcast channel. Namely, since $\mathbb{E}[\tilde{Y}^C] \to (\min_j q_j)^{-1}$ depends upon $\mathbf{q}$ (and thus M) only through the statistic $\min_j q_j$. Thus, in particular, the average delay per packet for M channels with success probabilities $\mathbf{Q}$ is asymptotically in C the same as the average delay per packet of a single erasure channel (i.e., $M' = 1$) with $q_1 = \min_{j \in M} q_j$. In short, the performance of the broadcast erasure channel is limited by the bottleneck channel.

IV. Comparing UT with RLNC

Recall Remark 5 observed that UT and RLNC are equivalent under our definitions in the special case of $C = 1$ packet. That is, $\mathbb{E}[Z] = \mathbb{E}[Y(1)]/1$. Prop. 7 below establishes that asymptotically in C, RLNC outperforms UT in terms of expected delay per packet.

Proposition 7: The asymptotic expected delay per packet under RLNC is superior to the expected delay per packet under UT:

$$\lim_{C\to\infty} \frac{\mathbb{E}[Y(C)]}{C} = \left(\min_{j\in M} q_j\right)^{-1} \leq \mathbb{E}[Z]. \qquad (44)$$

Proof: Fix the number of receivers M and the success probabilities $\mathbf{q} = (q_1, \ldots, q_M)$. Suppose $i \in \arg\min_j q_j$ and fix an arbitrary other index $k \in [M] \setminus i$. Consider two copies of this channel, one running RLNC with $C \to \infty$ for all M receivers with success probabilities $\mathbf{q}$, and the other running UT on with only the pair of 2 receivers, i, k, selected from $[M]$, with success probabilities $0 < q_i \leq q_k$. The expected delay per packet under RLNC is $1/q_i$, and the expected delay per packet under UT on the pruned systems with 2 receivers is, using Prop. 2,

$$\begin{aligned}\mathbb{E}[Z] &= (-1)^{1+1}\left(\frac{1}{1-(1-q_i)} + \frac{1}{1-(1-q_k)}\right) + \\ &\quad (-1)^{2+1}\frac{1}{1-(1-q_i)(1-q_k)}. \end{aligned} \qquad (45)$$

Observe that if we can show $(\min_j q_j)^{-1} \leq \mathbb{E}[Z]$ for the pruned system, then clearly it also holds for the original system with all M receivers. Simple algebra establishes that $1/q_i \leq \mathbb{E}[Z]$, where the inequality is tight only in the degenerate cases when either $q_i = 0$ or $q_k = 1$. ■

Remark 15: The above proof attempted using the lower bound $\psi(\mathbf{q})$ from Prop. 1, instead of the exact expression for $\mathbb{E}[Z]$ from Prop. 2 would fail as there are non-degenerate choices for $\mathbf{q}$ for which the lower bound on expected delay per packet under UT could be superior to that of RLNC. This is in fact a key motivation for developing exact expressions for $\mathbb{E}[Z]$. The desired inequality does go through when using the upper bound on $\mathbb{E}[Z]$, namely $\psi(\mathbf{q}) + 1$, from Prop. 1, but this inequality is inconclusive as it only relates an upper bound on UT to the asymptotic performance of RLNC.

V. Conclusion

In this paper we have addressed the expected delay per packet when transmitting C packets over a broadcast erasure channel to M receivers with independent success probabilities $\mathbf{q}$ under both uncoded transmission (UT) and random linear network coding (RLNC). Our contributions lie in $i)$ tight bounds and exact expressions for expected delay under UT, and $ii)$ exact expressions, recurrences, and asymptotic properties of the delay per packet under RLNC. These investigations naturally lead to more general questions in probability theory about the expected value of the maximum of geometric RVs (in UT) and the maximum of sums of geometric RVs (in RLNC). Such results are not just pertinent to our problem; see for instance [7] for the motivation and importance of these questions in bioinformatics. More generally this belongs to the study of *order statistics* of geometric RVs and negative binomial RVs. Although it is possible to write their distributions and express the moments, it is not always obvious how to derive an identical more explicit expression that does not involve infinite sums, say, and/or is easily computable. See [10] to appreciate some difficulty of these problems (even for the homogeneous cases). In the past, many efforts from statisticians and other mathematicians have been devoted to approximating using tools from complex analysis typically [8], [10], or the idea that exponential RV can approximate geometric RV, etc.

We close with a conjecture which is the focus of our ongoing work in this area.

Conjecture 1: The sequence $\mathbb{E}[Y(C)]/C$ in the state-independent case is monotone decreasing in C, with $\mathbb{E}[Y(1)] = \mathbb{E}[Z]$, converging in C to $(\min_j q_j)^{-1}$.
This paper establishes partial results along these lines, namely $\mathbb{E}[Y(1)] = \mathbb{E}[Z]$, and $\mathbb{E}[Z] > \lim_{C\to\infty} \mathbb{E}[Y(C)]/C$. This conjecture is important in that it establishes the average delay per packet is decreasing in the block length C, and in particular that any finite block length C yields a performance improvement over uncoded transmission.

References

[1] R. Cogill, B. Shrader, and A. Ephremides, "Stable throughput for multicast with random linear coding," *IEEE Trans. Info. Theory*, vol. 57, no. 1, pp. 267–281, January 2011.

[2] R. Cogill and B. Shrader, "Multicast queueing delay: Performance limits and order-optimality of random linear coding," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 5, pp. 1075–1083, May 2011.

[3] ——, "Delay bounds for random linear coding in multihop relay networks," in *Proceedings of the 2011 Conference on Information Science and Systems (CISS)*, 2011.

[4] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995.

[5] S. M. Ross and E. A. Pekoz, *A Second Course in Probability*. Probability Bookstore, 2007.

[6] V. M. Zaskulnikov, "Statistical mechanics of fluids in a step potential," *arXiv: 1205.6546v1*, 2012.

[7] B. Eisenberg, "On the expectation of the maximum of iid geometric random variables," *Statistics & Probability Letters*, no. 78, pp. 135–143, 2008.

[8] W. Szpankowski and V. Rego, "Yet another application of binomial recurrence: order statistics," *Computing*, no. 43, pp. 401–410, 1990.

[9] T. K. Dikaliotis, A. G. Dimakis, T. Ho, and M. Effros, "On the delay advantage of coding in packet erasure networks," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, 2010.

[10] P. J. Grabner and H. Prodinger, "Maximum statistics of n random variables distributed by the negative binomial distribution," *Combinatorics, Probability and Computing*, vol. 6, no. 2, pp. 179–183, 1997.

[11] A. Gut, *An Intermediate Course in Probability*, 2nd ed. Springer, 2009.

[12] J. Jacod and P. Protter, *Probability Essentials*, 2nd ed. Springer, 2003.

VI. Appendix

A. *Proof of Prop. 1*

The proof is a generalization of the technique employed in the proof of Lemma 4 in [1]. Each RV Z_j has PMF $\mathbb{P}(Z_j = n) = (1-q_j)^{n-1}q_j$ and complementary CDF (CCDF) $\mathbb{P}(Z_j > n) = (1-q_j)^n$ for $n \in \mathbb{N}$ and $j \in [M]$. Recall $\phi(q) \equiv -\ln(1-q)$ and define $\boldsymbol{\phi} \equiv (\phi_1, \ldots, \phi_M)$ with $\phi_j \equiv \phi(q_j)$ for $j \in [M]$; observe $1 - q_j = \mathrm{e}^{-\phi_j}$. Recall $[M]_k$ denotes all subsets of $[M]$ with cardinality k. The outline of the proof is as follows. First, we establish an expression for the PMF of Z, i.e., $\mathbf{p}_Z = (p_Z(n), n \in \mathbb{N})$ where $p_Z(n) = \mathbb{P}(Z = n)$, in terms of $\boldsymbol{\phi}$ in (47). Next, we apply the definition of expectation $\mathbb{E}[Z]$ in terms of $\mathbf{p}_Z$. The key inequality to obtain both the lower and upper bound is

$$\int_{n-1}^{n} x f(x)\mathrm{d}x \leq n \int_{n-1}^{n} f(x)\mathrm{d}x \leq \int_{n-1}^{n} (x+1) f(x)\mathrm{d}x, \tag{46}$$

when $f(x) \geq 0$.

We first obtain $\mathbf{p}_Z$ in terms of $\boldsymbol{\phi}$ by expressing the PMF in terms of the difference in the CDF, and recognizing the event $\{Z \leq n\}$ equals the intersection of independent events $\bigcap_{j\in[M]}\{Z_j \leq n\}$, which may be expressed in terms of the CCDF of each Z_j:

$$\begin{aligned} p_Z(n) &= \mathbb{P}(Z \leq n) - \mathbb{P}(Z \leq n-1) \\ &= \mathbb{P}(Z_1 \leq n, \ldots, Z_M \leq n) - \\ &\quad \mathbb{P}(Z_1 \leq n-1, \ldots, Z_M \leq n-1) \\ &= \prod_{j=1}^{M} \mathbb{P}(Z_j \leq n) - \prod_{j=1}^{M} \mathbb{P}(Z_j \leq n-1) \\ &= \prod_{j=1}^{M} (1 - \mathbb{P}(Z_j > n)) - \prod_{j=1}^{M} (1 - \mathbb{P}(Z_j > n-1)) \\ &= \prod_{j=1}^{M} (1 - (1-q_j)^n) - \prod_{j=1}^{M} \left(1 - (1-q_j)^{n-1}\right) \\ &= \prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j n}\right) - \prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j (n-1)}\right) \end{aligned} \tag{47}$$

We apply the PMF $\mathbf{p}_Z$ in the definition of $\mathbb{E}[Z]$ and use the fundamental theorem of calculus in the form

$$\int_{n-1}^{n} \frac{\mathrm{d}}{\mathrm{d}x} g(x)\mathrm{d}x = g(n) - g(n-1). \tag{48}$$

Namely:

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{n=1}^{\infty} n \mathbb{P}(Z = n) \\ &= \sum_{n=1}^{\infty} n \left(\prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j n}\right) - \prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j (n-1)}\right) \right) \\ &= \sum_{n=1}^{\infty} n \int_{n-1}^{n} \frac{\mathrm{d}}{\mathrm{d}x} \left(\prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j x}\right) \right) \mathrm{d}x \end{aligned} \tag{49}$$

Next, use the lower bound in (46) and recognize that the sum of integrals over disjoint intervals is the integral over the union of the intervals:

$$\begin{aligned} \psi(\mathbf{q}) &= \sum_{n=1}^{\infty} \int_{n-1}^{n} x \cdot \frac{\mathrm{d}}{\mathrm{d}x} \left(\prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j x}\right) \right) \mathrm{d}x \\ &= \int_{0}^{\infty} x \cdot \frac{\mathrm{d}}{\mathrm{d}x} \left(\prod_{j=1}^{M} \left(1 - \mathrm{e}^{-\phi_j x}\right) \right) \mathrm{d}x \end{aligned} \tag{50}$$

Next we apply the multi-binomial theorem in the form

$$\prod_{j=1}^{M} (1 - a_j) = \sum_{k=0}^{M} (-1)^k \sum_{A \in [M]_k} \prod_{j \in A} a_j. \tag{51}$$

Namely:

$$\begin{aligned} \psi(\mathbf{q}) &= \int_{0}^{\infty} x \cdot \frac{\mathrm{d}}{\mathrm{d}x} \left(\sum_{k=0}^{M} (-1)^k \sum_{A \in [M]_k} \mathrm{e}^{-\left(\sum_{j\in A} \phi_j\right)x} \right) \mathrm{d}x \\ &= \sum_{k=1}^{M} (-1)^k \sum_{A \in [M]_k} - \left(\sum_{j \in A} \phi_j \right) \int_{0}^{\infty} x \mathrm{e}^{-\left(\sum_{j\in A} \phi_j\right)x} \mathrm{d}x \\ &= \sum_{k=1}^{M} (-1)^k \sum_{A \in [M]_k} - \left(\sum_{j \in A} \phi_j \right) \frac{1}{\left(\sum_{j\in A} \phi_j\right)^2} \\ &= \sum_{k=1}^{M} (-1)^{k+1} \sum_{A \in [M]_k} \left(\sum_{j \in A} \phi_j \right)^{-1} \end{aligned} \tag{52}$$

This is the lower bound $\psi(\mathbf{q})$ in (2). To prove the upper bound, i.e., $\mathbb{E}[Z] \leq \psi(\mathbf{q}) + 1$, apply the upper bound in (46) and proceed as above. ■