On the Throughput of Multicasting with Incremental Forward Error Correction

I-Chung Lee   Cheng-Shang Chang   Ching-Ming Lien
Institute of Communications Engineering
National Tsing Hua University
Hsinchu 300, Taiwan, R.O.C.
Email: iclee@gibbs.ee.nthu.edu.tw/cschang@ee.nthu.edu.tw
keiichi@gibbs.ee.nthu.edu.tw

Abstract

In this paper, we consider a multicasting model that uses incremental FEC (Forward Error Correction). In this model, there is one sender and r^n receivers. The sender uses an ideal $(n, n(1-p), np)$ FEC code to code a group of $n(1-p)$ data packets with additional np redundant packets so that any set of $n(1-p)$ packets received by a receiver can be used to recover the original $n(1-p)$ data packets. Packets to the receivers are lost independently with probability q. For this model, we prove several strong laws of large numbers for the asymptotic throughput as $n \to \infty$. The asymptotic throughput is characterized by the unique solution of an equation in terms of p, q and r. These strong laws not only provide theoretical justification for several important observations made in the literature, but also provide insights that might have impact on future design of multicasting protocols.

Index Terms

Multicast, throughput, forward error correction, large deviations, law of large numbers

This research is supported in part by the National Science Council, Taiwan, R.O.C., under Contract NSC-89-2213-E007-127 and the program for promoting academic excellence of universities 89-E-FA04-1-4.

I. INTRODUCTION

Using FEC (Forward Error Correction) for multicasting in the Internet has received a lot of attention recently (see e.g., [18], [5], [19], [25], [22], [15]). Several important observations were made in the literature based on numerical comparisons of various protocol schemes that use FEC. For instance, it is noted in [18] that multicasting protocols with FEC scale well for a broad range of receivers and that the amount of redundancy should be "matched" to the size of data in order to obtain high throughput. Also, it is observed in [22] that the throughput for a multicasting problem with heterogeneous groups of receivers is dominated by the group with high loss. Motivated by providing rigorous justification and insightful explanations for these important observations, we take a theoretical approach to study a basic multicasting problem with incremental forward error correction. In this paper, we make the following assumptions for our model (see Figure 1).

(A1) There is an ideal incremental forward error correction code $(n, n(1-p), np)$, where n, $n(1-p)$, and np are the block length, the amount of data symbols, and the amount of redundant symbols, respectively. Any set of $n(1-p)$ symbols among the n symbols can be used to recover the whole n symbols and thus the original $n(1-p)$ data symbols.

(A2) There are one sender and r^n receivers for some $r > 1$. The sender sends fixed size packets to the receivers. Packets to the receivers are lost *independently* with probability $q < 1$. To ensure all the receivers receive the correct data from the sender, np redundant packets are added to every group of $n(1-p)$ data packets using the forward error correction code in Assumption (A1). By so doing, any set of $n(1-p)$ packets received by a receiver can be used to recover the original $n(1-p)$ data packets. The sender then sends the n packets one by one until **all** the receivers receive at least $n(1-p)$ distinct packets. If all the n packets are sent and there are still receivers that have not received enough distinct packets, the sender then resends the n packets one by one from the first packet. This procedure is carried out repeatedly until all the receivers receive enough distinct packets.

Here we implicitly assume that there is a feedback channel so that the sender always knows whether there is a receiver that has not received at least $n(1-p)$ distinct packets. However, the sender does not know how many of them. The reason for the limited feedback assumption is

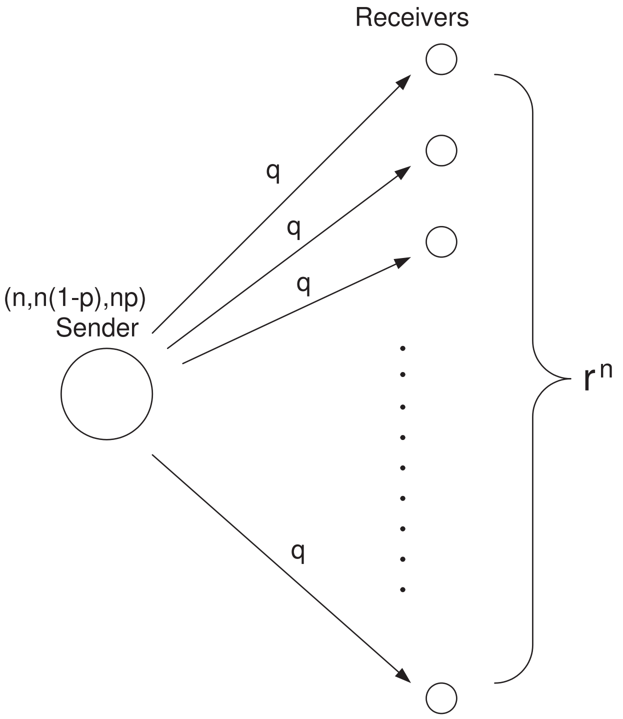

Fig. 1. The model

the well-known feedback implosion problem (see e.g., [24]). If all receivers send requests for retransmission as soon as they lose data, these requests would overwhelm the sender. In practice, feedbacks need to be aggregated or suppressed to solve the feedback implosion problem. As such, the sender only knows that some receivers do not receive the data, but the sender does not know who do not receive the data.

When all the n packets are sent and there are still receivers that have not received enough distinct packets, it is possible that the sender could create more redundant symbols and transmit these symbols, instead of resending the n symbols repeatedly. The analysis for such a scheme is a special case of our analysis, and we believe the readers should not have problems adapting our analysis for such a scheme.

Our model is almost identical to the idealized integrated FEC in [18]. The difference is that the integrated FEC in [18] stops the multicasting process after sending n packets. Here we allow the sender to resend the packets until all the receivers receive enough distinct packets. In our model, there are three key parameters: the fraction of redundancy for the forward error correction code p, the packet loss probability q, and the exponent coefficient of the number of receivers $\log r$. We will show in this paper how these three parameters affect the throughput of the multicasting problem when n is large. To be precise, let $T(n)$ be the number of packets sent by the sender until all the receivers receive at least $n(1-p)$ distinct packets. Then we can define the throughput of the sender $S(n)$ as the ratio of the number of data packets to the number of transmitted packets, i.e.,

$$S(n) = \frac{n(1-p)}{T(n)}. \tag{1}$$

One of our main results in this paper is a strong law of large numbers that shows $S(n)$ converging to an asymptotic throughput S when $n \to \infty$. The asymptotic throughput S is characterized by the unique solution of an equation in terms of the three key parameters p, q and $\log r$. The strong law not only provides theoretical justification for some observations in [18], [22], but also provides insights that might have impact on future design of multicasting protocols.

We summarize our contributions of this paper:

(i) In Section II, we prove a strong law of large numbers for the asymptotic throughput S in terms of the three key parameters p, q and $\log r$. The strong law shows how the length of the FEC code n scales with the number of receivers r^n while keeping a nonzero asymptotic throughput.

(ii) In Section III, we prove a strong law of large numbers for all the receivers to receive enough distinct packets. The strong law proves that all the receivers will receive enough distinct packets (with probability 1) if a sufficient number of packets are sent by the sender. One important implication of the strong law is that *feedback* may not be needed for a highly reliable multicasting system if the sender knows the three key parameters p, q and $\log r$. In the case that the packet loss probability q is unknown, we suggest a *partial feedback* protocol by selecting a subset of receivers with feedback capability and use the throughput formula in Section II to estimate q. The key idea behind the partial feedback protocol is that after the subset of receivers with feedback capability has received enough distinct packets, sending a few more packets will ensure that the remaining receivers (without feedback capability) will receive enough distinct packets.

(iii) When both q and $\log r$ are known to the sender, the sender can maximize the throughput over the fraction of redundancy p. In Section IV, we show that p should be chosen so that every packet is sent the same number of times. Based on this, we derive an efficient algorithm for computing the optimal fraction of redundancy p. Our numerical results provide an answer to the question that redundancy should be "matched" to the size of data in [18].

(iv) We prove in Section V that the asymptotic throughput for a multicasting problem with heterogeneous groups of receivers is determined by the asymptotic throughput of the group with the worst throughput. This provides theoretical justification for the observation made in [22].

(v) In Section VI, we extend the strong law of large numbers in Section II to multicasting with a common shared link (the satellite model). For such a model, the number of received packets of each receiver now become spatially correlated. From various numerical examples, we show that spatial loss correlation can increase the asymptotic throughput of a multicast system. This is consistent with the observation made in [18].

II. A STRONG LAW OF LARGE NUMBERS FOR THE ASYMPTOTIC THROUGHPUT

In Theorem 1, we prove a strong law of large numbers for the asymptotic throughput. The strong law shows how the length of the FEC code n scales with the number of receivers r^n while keeping a nonzero asymptotic throughput.

Theorem 1 *Let*

$$\Lambda(\theta,\alpha) = \delta \log(q^{k+1}e^{\theta} + (1-q^{k+1})) + (1-\delta)\log(q^k e^{\theta} + (1-q^k)), \quad \alpha > 0, \tag{2}$$

where $k = \lfloor \alpha \rfloor$ is the integral part of α and $\delta = \alpha - k$ is the fractional part of α. Also, let

$$\Lambda^*(\gamma,\alpha) = \sup_{\theta \ge 0}[\theta\gamma - \Lambda(\theta,\alpha)] \tag{3}$$

be the one-sided Legendre transform of $\Lambda(\theta,\alpha)$. Then

$$\lim_{n\to\infty} S(n) = \lim_{n\to\infty} \frac{n(1-p)}{\mathsf{E}T(n)} = S, \quad a.s., \tag{4}$$

where $0 < S < 1 - q$ *is the unique solution of the following equation*

$$\Lambda^*(p, \frac{1-p}{S}) = \log r. \tag{5}$$

Our approach for Theorem 1 is based on the large deviation principle (see e.g., [4], [10], [23]) and its proof is given in Appendix B. Such a technique was previously developed by Kingman [17] and Biggins [3] for the last birth problem and it was also used in [7] for bounding the speedup of parallel processing systems.

The formula in Theorem 1 is not as complicated as it looks and in fact the one-sided Legendre transform in (3) has a closed form expression as shown in Lemma 2 below. This is due to many well-known properties of the Legendre transform (also known as the convex transform, see e.g., Rockafellar [21], Avriel [2], and Hiriart-Urruty and Lemaaréchal [13]). For instance, as $\Lambda(0,\alpha) = 0$, $\Lambda^*(\gamma,\alpha) \geq 0\gamma - \Lambda(0,\alpha) = 0$. Moreover, as $\Lambda(\theta,\alpha)$ is convex and differentiable for all θ and $\Lambda(0,\alpha) = 0$, it is known that $\Lambda^*(\gamma,\alpha)$ is strictly increasing, convex, and continuous in γ when $0 < \Lambda^*(\gamma,\alpha) < \infty$. In Lemma 2, we show some additional properties of the one-sided Legendre transform that can be used for computing the asymptotic throughput S in Theorem 1 (and for the proof of Theorem 1). The proof of Lemma 2 is given in Appendix A. Using the closed form formulas in Lemma 2(iii) and (iv), in Figure 2 we plot the asymptotic throughput S in Theorem 1 as a function of $\log r$ for fixed $p = 0.1, 0.3, 0.5, 0.7, 0.9$ and $q = 0.1$. As shown in Figure 2, both the curves from $p = 0.1$ and $p = 0.9$ yield lower throughput than others. In the case $p = 0.1$, the redundancy of the FEC code is too low to recover from packet losses as the packet loss probability q is also 0.1. We may view this as an *underprotected* case. On the other hand, the redundancy for the FEC code with $p = 0.9$ is so high that most of the time the sender is sending redundant information. This is an *overprotected* case. Also, there does not exist a curve that is always better than others for the whole range of $\log r$. A natural question is what fraction of redundancy p should be used to maximize the asymptotic throughput. We will discuss this in Section IV.

Lemma 2 (i) $\Lambda(\theta,\alpha)$ *is continuous in* $\alpha > 0$.

(ii) *For* $\theta > 0$, $\Lambda(\theta,\alpha)$ *is strictly decreasing in* α *to 0.*

(iii) *For* $\gamma \leq \Lambda'(0,\alpha)$, $\Lambda^*(\gamma,\alpha) = 0$.

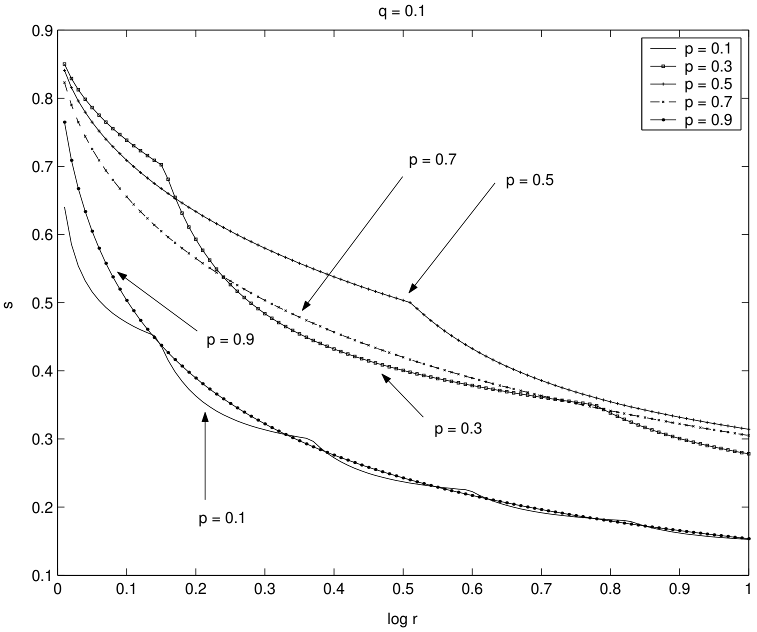

Fig. 2. The asymptotic throughput S as a function of $\log r$

(iv) *For* $\Lambda'(0,\alpha) < \gamma < 1$,

$$\Lambda^*(\gamma,\alpha) = \theta(\gamma,\alpha)\gamma - \Lambda(\theta(\gamma,\alpha),\alpha) > 0, \tag{6}$$

where

$$\theta(\gamma,\alpha) = \log(\frac{b+\sqrt{b^2+4ac}}{2a}) > 0, \tag{7}$$

$$a = (1-\gamma)q^{k+1}q^k, \tag{8}$$

$$b = (\gamma-\delta)q^{k+1}(1-q^k) + (\gamma-1+\delta)q^k(1-q^{k+1}), \tag{9}$$

$$c = \gamma(1-q^{k+1})(1-q^k). \tag{10}$$

(v) *For a fixed* $0 < \gamma < 1$,

$$\Lambda^*(\gamma,\alpha) = \theta(\gamma,\alpha)^+\gamma - \Lambda(\theta(\gamma,\alpha)^+,\alpha), \tag{11}$$

where $x^+ = \max(0,x)$ *and* $\theta(\gamma,\alpha)$ *in (7).*

(vi) *For a fixed* $0 < \gamma < 1$, $\Lambda^*(\gamma,\alpha)$ *is continuous in* $\alpha > 0$.

(vii) *For a fixed* $0 < \gamma < 1$, *if* $\Lambda^*(\gamma,\alpha) > 0$, *then* $\Lambda^*(\gamma,\alpha)$ *is strictly increasing in* $\alpha > 0$.

(viii) *For a fixed* $0 < \gamma < 1$,

$$\lim_{\alpha\to\infty} \Lambda^*(\gamma,\alpha) = \infty.$$

We note that when $\alpha = k$ is an integer ($\delta = 0$), the one-sided Legendre transform in (3) has the following form

$$\Lambda^*(\gamma, k) = \begin{cases} D(\gamma||q^k) & \text{if } \gamma \geq q^k \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

where

$$D(p_1||p_2) = p_1 \log \frac{p_1}{p_2} + (1-p_1) \log \frac{1-p_1}{1-p_2} \tag{13}$$

is the Kullback-Leibler distance (also known as the relative entropy in the information theory) between two Bernoulli random variables with means p_1 and p_2 (see e.g., [9]). This gives us an intuitive explanation of Theorem 1. If every packet is transmitted k times by the sender, then the probability that a packet is lost at a receiver is reduced from q to q^k. As a result of Sanov's theorem (see e.g., [9]), the probability that a receiver has more than pn lost packets is roughly $\exp(-nD(p||q^k))$, which is exponentially small when n is large. Since there are r^n receivers, the probability that at least one receiver not receiving enough number of packets is roughly $r^n \exp(-nD(p||q^k))$, which can be made exponentially small as long as k is large enough to ensure $D(p||q^k) > \log r$. Certainly, the formal proof of Theorem 1 requires a more careful treatment when α is not an integer.

III. Multicasting without feedback and with partial feedback

Like most multicasting protocols that use FEC in the literature, we assume in [A2] there is an ideal feedback channel to achieve reliable multicasting. In this section, we show that feedback may not be needed if the sender knows the packet loss probability q and the number of receivers r^n.

To formalize the result, let $Z_{i,j}(m)$, $i = 1, \ldots, r^n$, $j = 1, \ldots, n$, be the indicator random variable for the event that the i^{th} receiver has not received the j^{th} packet after m packets are sent by the sender. Also, let $Y_i(\alpha) = \sum_{j=1}^{n} Z_{i,j}(\lfloor \alpha \rfloor)$ be the number of packets that have not been received by the i^{th} receiver after $\lfloor \alpha \rfloor$ packets are sent by the sender. Then the event $\{\max_{1\leq i\leq r^n}[Y_i(\alpha)] \geq pn\}$ is the event that there is at least one receiver not receiving enough distinct packets after $\lfloor \alpha \rfloor$ packets are sent by the sender.

Theorem 3 *Let S be the asymptotic throughput in Theorem 1. Then*

$$\lim_{n\to\infty}\frac{1}{n}\max_{1\leq i\leq r^n}[Y_i(\frac{1-p}{S}n)] = p, \quad a.s. \tag{14}$$

The proof of Theorem 3 is shown in Appendix C.

Theorem 3 implies that the sender can simply compute the asymptotic throughput S first and then send αn packets with some $\alpha > \frac{1-p}{S}$. This will ensure all the receivers receive enough packets with probability 1 when $n \to \infty$. In fact, as shown in Lemma 8(iv), the probability that there is at least one receiver not receiving enough packets after αn packets (with $\alpha > \frac{1-p}{S}$) approaches 0 exponentially fast when $n \to \infty$. Thus, we can achieve highly reliable multicasting without using any form of feedback. If 100% reliability is required, one still needs to implement feedback. A receiver starts sending a request for retransmission to the sender only after αn packets are sent by the sender. If α is slightly larger than $\frac{1-p}{S}$, then our result implies that the probability that the sender will receive a request for retransmission can be made arbitrarily small while keeping the throughput almost the same as that of the system with feedback after every packet.

The asymptotic throughput formula in Theorem 1 can be used for estimating the loss probability q. Suppose now the sender does not have the information of the packet loss probability q and there is only a subset of receivers with size r_0^n ($r_0 \leq r$) that have the ideal feedback capability. We call this a multicasting problem with partial feedback. Let $T_0(n)$ be the number of packets that the sender sends to ensure all the subset of receivers receive at least $n(1-p)$ distinct packets. As $T_0(n)$ now is available to the sender, the sender can estimate the packet loss probability q by solving

$$\Lambda^*(p, \frac{T_0(n)}{n}) = \log r_0. \tag{15}$$

Once the sender obtains the estimate for q, then the problem becomes the multicasting problem without feedback. The sender can use the estimate for q to solve the asymptotic throughput S and continue sending packets until αn packets are sent for some $\alpha > \frac{1-p}{S}$.

IV. Maximizing the asymptotic throughput

As in the previous section, suppose that the packet loss probability q and the number of receivers r^n is known. How do we find the optimal fraction of redundancy p of the forward error correction code $(n, n(1-p), np)$ such that the asymptotic throughput is maximized? To be precise, we formulate the following maximization problem:

$$\begin{aligned} &\max_p S \\ &\text{s.t.} \quad \Lambda^*(p, \frac{1-p}{S}) = \log r. \end{aligned}$$

The problem of finding p that maximizes throughput can be greatly simplified by the following lemma.

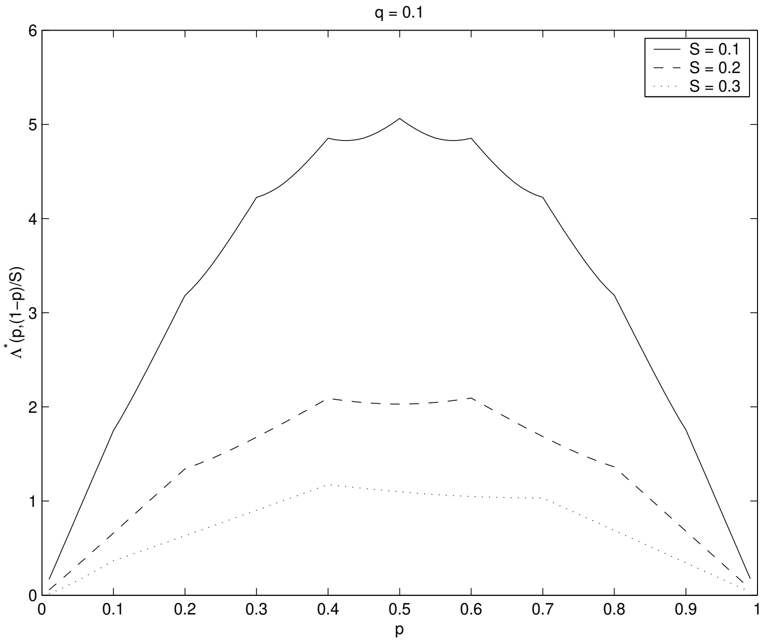

Fig. 3. The function of $\Lambda^*(p, \frac{1-p}{S})$

Lemma 4 *For a fixed S, $\Lambda^*(p, \frac{1-p}{S})$ is maximized when $\frac{1-p}{S}$ is an integer.*

To gain some insight for Lemma 4, in Figure 3 we plot $\Lambda^*(p, \frac{1-p}{S})$ as a function of p for fixed $S = 0.1, 0.2, 0.3$ and $q = 0.1$. As shown in Figure 3, all the non-differentiable points are when $\frac{1-p}{S}$ are integers. Segments between two non-differentiable points are convex. It implies that the maximum occurs only at the boundary points of segments.

Proof. Let $\alpha = \frac{1-p}{S}$. As in Theorem 1, let $k = \lfloor \alpha \rfloor$ and $\delta = \alpha - k$ be the integral part and the fractional part of α. Now

$$\Lambda^*(p, \frac{1-p}{S}) = \Lambda^*(1-\alpha S, \alpha) = \Lambda^*(1-kS-\delta S, k+\delta)$$

is a function of δ. Consider the function

$$g(\theta, \delta, k) = \theta p - \Lambda(\theta, \alpha).$$

From (3), it follows that

$$\Lambda^*(1-kS-\delta S, k+\delta) = \sup_{\theta \geq 0}[g(\theta, \delta, k)]. \tag{16}$$

Note that $\Lambda(\theta, \alpha)$ in (2) is a linear function in δ. Then $g(\theta, \delta, k)$ is also linear in δ as

$$p = 1 - \alpha S = 1 - kS - \delta S$$

is linear in δ. Thus,

$$\begin{aligned}
&\Lambda^*(1-kS-\delta s, k+\delta) = \sup_{\theta \geq 0}[g(\theta, \delta, k)] \\
&= \sup_{\theta \geq 0}[(1-\delta)g(\theta, 0, k) + \delta g(\theta, 1, k)] \\
&\leq (1-\delta)\sup_{\theta \geq 0}[g(\theta, 0, k)] + \delta \sup_{\theta \geq 0}[g(\theta, 1, k)] \\
&= (1-\delta)\Lambda^*(1-kS, k) + \delta\Lambda^*(1-kS-S, k+1) \\
&\leq \max[\Lambda^*(1-kS, k), \Lambda^*(1-(k+1)S, k+1)].
\end{aligned}$$

As a direct consequence, the maximum of $\Lambda^*(p, \frac{1-p}{S})$ occurs when $\frac{1-p}{S}$ is an integer. ■

As $\Lambda^*(p, \frac{1-p}{S})$ is decreasing in S for a fixed p (Lemma 2(vii)), one has from Lemma 4 that

$$\max_p S = \max_{k \geq 1}[S_k],$$

where $S_k < 1-q$, $k = 1, 2, \ldots$, satisfies the equation

$$\Lambda^*(1-kS_k, k) = \log r. \tag{17}$$

We note that there may not be a solution for (17) if k is too small. To find the minimum k, denoted by k_1, that has a nonzero S_k, we observe from (12) that

$$D(1||q^{k_1}) > \log r.$$

Thus,

$$k_1 = \lfloor -\frac{\log r}{\log q} \rfloor + 1. \tag{18}$$

On the other hand, as $1 - kS_k \geq 0$, we have

$$S_k \leq \frac{1}{k}.$$

This implies that we do not need to compute S_k for k that is too large to produce the maximum throughput. The following algorithm outlines the steps to compute the maximum throughput $S_{\max}$ and the associated $p_{\max}$ for the forward error correction code.

Algorithm 1.

1. Let $k_1 = \lfloor -\frac{\log r}{\log q} \rfloor + 1$.

   Solve $S_1 < 1 - q$ from $D(1 - k_1 S||q^{k_1}) = \log r$.

   Let $S_{\max} = S_1$ and $p_{\max} = 1 - k_1 S_{\max}$.

2. Repeat from $k = k_1 + 1$ while $k \leq 1/S_{\max}$.

   Solve $S_k < 1 - q$ from $D(1 - kS||q^k) = \log r$.

   If $S_k \geq S_{\max}$, then let $S_{\max} = S_k$ and $p_{\max} = 1 - kS_{\max}$.

   Increase k by 1.

In Figure 4, we plot the maximum throughput curve versus $\log r$ for a fixed q. An interesting observation is that the maximum throughput approaches $1 - q$ when $r \to 1$. This is consistent with information theory as the capacity of a binary erasure channel is $1 - q$ (see e.g., [9]).

In Figure 5, we plot the maximum throughput curve versus q for a fixed $\log r$ and the approximation in (19) that will be discussed later. In Figure 6, we also plot the fraction of redundancy p that achieves the maximum throughput as a function of q. The curves in Figure 6 are quite interesting as they all start by oscillating and then converge to certain constants. Each oscillation corresponds to the increase of the number of transmissions by 1 for each packet. The intuition behind these

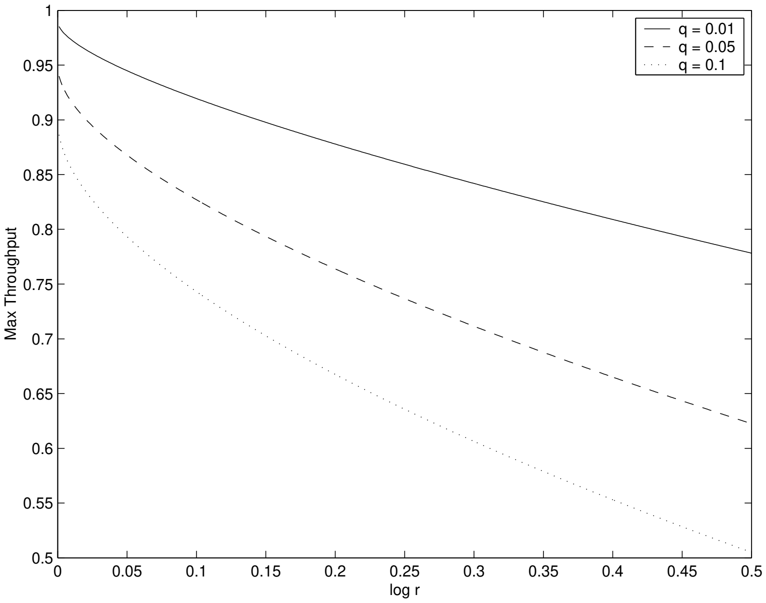

Fig. 4. Maximum throughput as a function of $\log r$

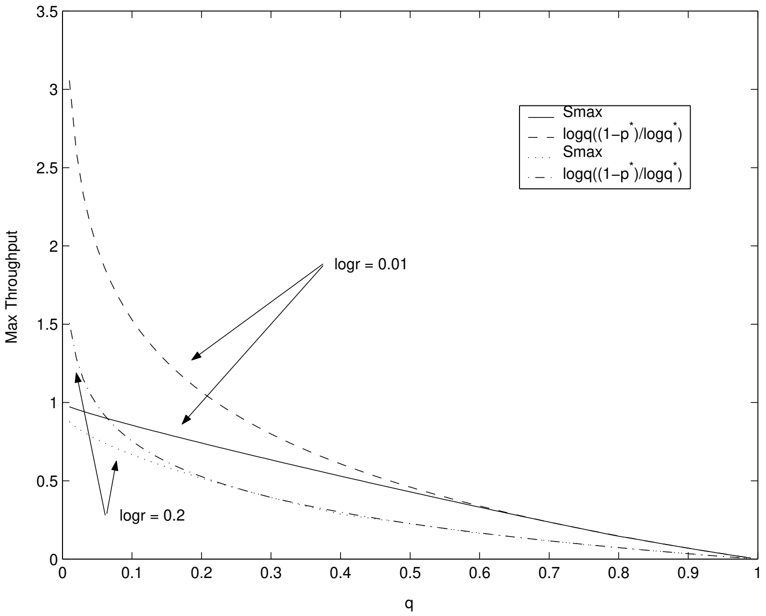

Fig. 5. Maximum throughput and its approximation in (19) as functions of q

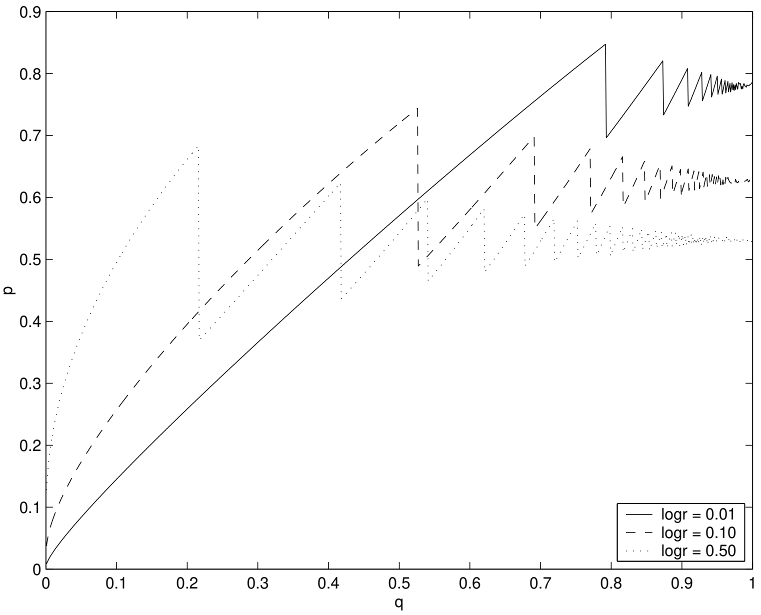

Fig. 6. The optimal fraction of redundancy p as a function of the packet loss probability q

curves is that the fraction of redundancy p increases as the packet loss probability q increases until packets are overprotected (break points). At such break points, increasing the number of transmissions to reduce the packet loss probability for each packet and lowering the fraction of redundancy p is better than continuing to increase the fraction of redundancy p (Readers should keep in mind that our results are asymptotic results and the above argument only holds for a sufficiently large and fixed block length n).

To gain more intuition for the case that $q \to 1$, suppose that the curves in Figure 6 converge as $q \to 1$. From Algorithm 1, we may approximate the converging constants in Figure 6 by letting $p^* \approx 1 - k_{\max} S_{\max}$ and $q^* \approx q^{k_{\max}}$, where $k_{\max}$ is the number of transmissions for each packet in order to maximize the asymptotic throughput. Eliminating $k_{\max}$ yields

$$S_{\max} \approx \frac{1-p^*}{\log q^*} \log q. \tag{19}$$

In order to maximize the throughput, we can solve p^* and q^* from the following constrained maximization problem

$$\begin{aligned} &\max_{p^*} -\frac{1-p^*}{\log q^*} \\ &\text{s.t.} \quad D(p^*||q^*) = \log r, \quad p^* > q^*. \end{aligned}$$

Once we obtain p^* and q^*, we can have (19) as an approximation for the asymptotic throughput when $q \to 1$. In Figure 5, we plot both the original maximum asymptotic throughput and the approximation in (19). They match very well when $q \to 1$.

One interesting observation from (19) is that the maximum asymptotic throughput $S_{\max}$ is proportional to $\log q$ when $q \to 1$. We further verify this by various numerical results in Figure 7. As shown in Figure 7, the ratios of $S_{\max}$ to $\log q$ converge to constants for various values of $\log r$.

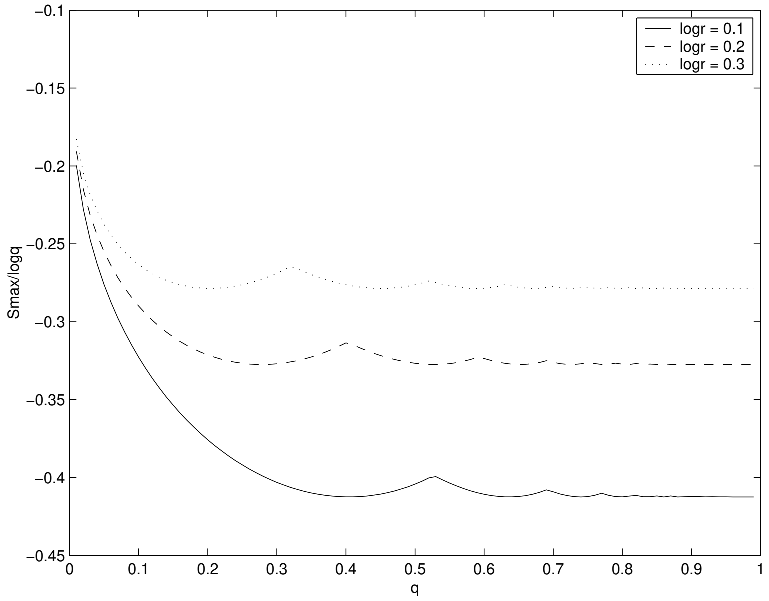

Fig. 7. The ratios of $S_{\max}$ to $\log q$

V. Heterogeneous groups of receivers

In this section, we extend our result to heterogeneous groups of receivers. Instead of having a single group of receivers, we assume there are L groups of receivers. The ℓ^{th} group contains r_ℓ^n receivers for some $r_\ell > 1$, $\ell = 1, \ldots, L$. Packets to the ℓ^{th} group of receivers are lost *independently* with probability $q_\ell < 1$, $\ell = 1, \ldots, L$. The sender still follows the procedure defined in [A2]. As in the previous section, let $T(n)$ be the number of packets sent by the sender until all the receivers receive at least $n(1-p)$ distinct packets, and $S(n) = \frac{n(1-p)}{T(n)}$ be the throughput of the sender.

Theorem 5 *For $\ell = 1, \ldots, L$, let*

$$\Lambda_\ell(\theta, \alpha) = \delta \log(q_\ell^{k+1} e^\theta + (1 - q_\ell^{k+1})) + (1-\delta) \log(q_\ell^k e^\theta + (1 - q_\ell^k)), \quad \alpha > 0, \tag{20}$$

where $k = \lfloor \alpha \rfloor$ is the integral part of α and $\delta = \alpha - k$ is the fractional part of α. Also, let

$$\Lambda_\ell^*(\gamma, \alpha) = \sup_{\theta \geq 0} [\theta\gamma - \Lambda_\ell(\theta, \alpha)] \tag{21}$$

be the one-sided Legendre transform of $\Lambda_\ell(\theta, \alpha)$, $\ell = 1, \ldots, L$. Then

$$\lim_{n \to \infty} S(n) = \min_{1 \leq \ell \leq L} S_\ell, \quad a.s., \tag{22}$$

where S_ℓ is the unique solution of the following equation

$$\Lambda_\ell^*(p, \frac{1-p}{S_\ell}) = \log r_\ell. \tag{23}$$

Theorem 5 indicates that the asymptotic throughput of the multicasting problem with heterogeneous groups is determined by the throughput of the group that has the worst performance. This provides a theoretical justification for the observation made in [22].

Proof. For $\ell = 1, \ldots, L$, let $T_\ell(n)$ be the number of packets that the sender sends to ensure all the receivers in the ℓ^{th} group receive at least $n(1-p)$ distinct packets. Then we have

$$T(n) = \max_{1 \leq \ell \leq L} [T_\ell(n)]. \tag{24}$$

From Theorem 1 and the continuous mapping theorem of the strong law (see e.g., Theorem 5.2 in [6]), we have

$$\lim_{n \to \infty} \frac{T(n)}{n(1-p)} = \max_{1 \leq \ell \leq L} [\frac{1}{S_\ell}], \quad a.s. \tag{25}$$

This in turn implies

$$\lim_{n \to \infty} S(n) = \min_{1 \leq \ell \leq L} S_\ell, \quad a.s.$$

■

VI. Multicasting with one common shared link

In this section, we extend our result to multicasting with one common shared link (see Figure 8). We make the same assumptions as in Section I except that each packet from the sender to a receiver has to go through one shared link and one dedicated link. This model corresponds to mutlicasting via a satellite in [14]. The sender first sends packets to the satellite and the satellite then broadcasts the packets back to the receivers. Assume that packets are lost independently with probability q_1 on the shared link and with the same probability q_2 on all the dedicated links. Also, we assume that the intermediate node is *transparent* in the sense that it simply relays every packet (it receives from the sender) on all the dedicated links. As in Section II, we show that there is a strong law for the asymptotic throughput of the sender. Let $\widetilde{T}(n)$ be the minimum number of packets transmitted by the sender so that all the receivers receive at least $n(1-p)$ distinct packets. Similarly, define $\widetilde{S} = \frac{n(1-p)}{\widetilde{T}(n)}$ to be the throughput of the sender.

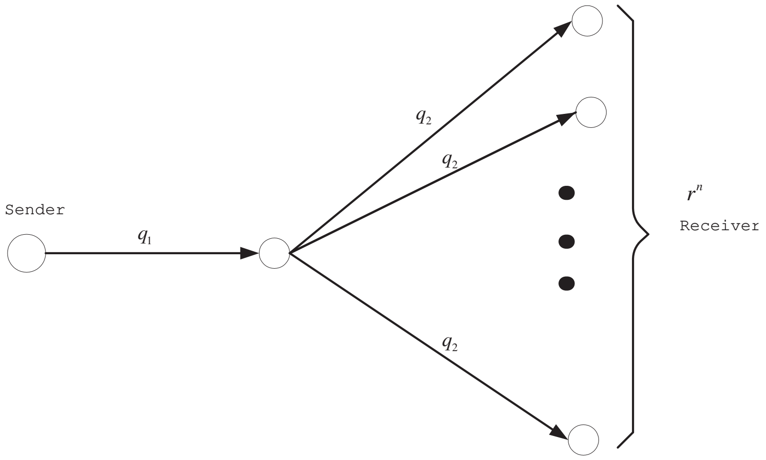

Fig. 8. Multicasting with one common shared link

Theorem 6 *Let*

$$\widetilde{\Lambda}(\theta,\alpha) = \delta \sum_{i=0}^{k+1} b_{k+1,i} \log(q_2^i e^\theta + (1-q_2^i)) + (1-\delta) \sum_{i=0}^{k} b_{k,i} \log(q_2^i e^\theta + (1-q_2^i)), \quad \alpha > 0, \tag{26}$$

where $k = \lfloor \alpha \rfloor$ *is the integral part of* α *,* $\delta = \alpha - k$ *is the fractional part of* α*, and*

$$b_{k,i} = \left(\begin{array}{c} k \\ i \end{array} \right) (1-q_1)^i q_1^{k-i} \tag{27}$$

is a binomial distribution coefficient. Also, let

$$\widetilde{\Lambda}^*(\gamma,\alpha) = \sup_{\theta \geq 0} [\theta\gamma - \widetilde{\Lambda}(\theta,\alpha)] \tag{28}$$

be the one-sided Legendre transform of $\widetilde{\Lambda}(\theta, \alpha)$. *Then*

$$\lim_{n\to\infty} \widetilde{S}(n) = \widetilde{S}, \quad a.s., \tag{29}$$

where $0 < \widetilde{S} < 1 - (q_1 + q_2 - q_1 q_2)$ *is the unique solution of the following equation*

$$\widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}}) = \log r. \tag{30}$$

The proof for Theorem 6 is based on the fact that when conditioning on the transmission result of the shared link, the spatial loss correlation between receivers can be removed. As such, we can then apply the methods used for the proof of Theorem 1.

To ease our presentation, we call the model in Section II the direct multicasting model and the model in this section the satellite model. Let $N_j(\alpha n)$ be the random variable representing *the number of distinct packets that are successfully transmitted* j *times on the shared link after the sender sends* $\lfloor \alpha n \rfloor$ *packets in the satellite model*. Also, let $\underline{N}(\alpha n) = (N_0(\alpha n), N_1(\alpha n), \ldots, N_{\lceil \alpha \rceil}(\alpha n))$. Note that $\sum_{j=0}^{\lceil \alpha \rceil} N_j(\alpha n) = n$ (the total number of distinct packets). Denote by $\widetilde{Y}_i(\alpha n)$ the number of packets that are not received by the i^{th} receiver after the sender sends $\lfloor \alpha n \rfloor$ packets in the satellite model.

Consider a probability vector $\underline{\delta} = (\delta_0, \delta_1, \delta_2, \cdots, \delta_{\lceil \alpha \rceil})$ (with $\sum_{j=0}^{\lceil \alpha \rceil} \delta_j = 1$). Let $Y_i(n\underline{\delta})$ be the number of packets that are not received by the i^{th} receiver after $n\delta_j$ packets, $j = 0, 1, \ldots, \lceil \alpha \rceil$, are sent j times in the direct multicasting model. Then conditioning on the transmission result at the shared link $\underline{N}(\alpha n) = n\underline{\delta}$, we have

$$\widetilde{Y}_i(\alpha n) = Y_i(n\underline{\delta}). \tag{31}$$

Once we establish the connection between these two models, the rest of the proof of Theorem 6 can be argued heuristically in the following two steps. First, observe that $\underline{N}(\alpha n)/n$ is the empirical distribution. From Sanov's theorem (see e.g., [9]), we know that such an empirical distribution converges to the (mixed) binomial distribution, i.e.,

$$\lim_{n\to\infty} \frac{\underline{N}(\alpha n)}{n} = \underline{b}_\alpha, \quad a.s., \tag{32}$$

where $\underline{b}_\alpha = \delta \underline{b}_k + (1-\delta)\underline{b}_{k+1}$, $k = \lfloor \alpha \rfloor$, $\delta = \alpha - k$, $\underline{b}_k = (b_{k,0}, b_{k,1}, ..., b_{k,k}, 0, 0, ...)$, $b_{k,i} = \left(\begin{array}{c} k \\ i \end{array}\right) (1-q_1)^i q_1^{k-i}$. Moreover, the probability that the empirical distribution $\underline{N}(\alpha n)/n$ deviates from the genuine distribution $\underline{b}_\alpha$ is exponentially small. Intuitively, when n is large, the outcome at the shared link is almost deterministic and it is near $\underline{b}_\alpha$. The second step is then to reduce the satellite model to the direct multicasting model by conditioning on $\frac{\underline{N}(\alpha n)}{n} = \underline{b}_\alpha$. Note that

$$\frac{1}{n}\log \mathsf{E}\exp(\theta Y_i(n\underline{\delta})) = \sum_{j=0}^{\lceil\alpha\rceil} \delta_j \log(q_2^j e^\theta + (1-q_2^j)). \tag{33}$$

and that $Y_i(n\underline{\delta}), i = 1, 2, .., r^n$ are independent. Hence we can apply the same method in the proof of Theorem 1 to obtain probabilistic bounds under the condition that $\underline{N}(\alpha n) = n\underline{b}_\alpha$. The formal proof of Theorem 6 is given in Appendix D.

Note that Theorem 1 becomes a special case in Theorem 6 with $q_1 = 0$ and $q_2 = q$. To gain more insights between these two models, we present several numerical results for the asymptotic results in Theorem 1 and Theorem 6. In Figure 9, we choose $q_1 = 0.05$, $q_2 = 1/19$ in Theorem 6 and choose $q = 0.1$ in Theorem 1. By so doing, the (end-to-end) packet loss probabilities are the same as $q = q_1 + q_2 - q_1q_2 = 0.1$. As shown in Figure 9, the asymptotic throughput in the satellite model link is always better than that in the direct multicasting model when p (the fraction of redundancy) and $\log r$ are kept the same.

In Figure 10, we vary the loss probability q_1 on the shared link, but keep the end-to-end packet loss probability fixed by choosing q_2 with $q_1 + q_2 - q_1q_2 = 0.1$. As $q_1 \to 0$, the model becomes the direct multicasting model. On the other hand, as $q_1 \to 0.1$ ($q_2 \to 0$), it becomes a single link model. Thus, increasing q_1 increases the spatial loss dependency. As shown in Figure 10, the asymptotic throughput $\widetilde{S}$ increases as q_1 increases. This shows that the spatial correlation increases the asymptotic throughput of a multicast system.

In Figure 11, we compare the asymptotic throughput under three scenarios: (i) transparent (ii) packet discard (iii) error recovery. In the *transparent* case (the case in Theorem 6), the satellite (the intermediate node) simply broadcasts the packets received from the sender to the receivers (even though there might be an error in a packet). The asymptotic throughput $\widetilde{S}$ then can be computed from Theorem 6 with $q_1 = 0.05$ and $q_2 = 1/19$. In the *packet discard* case, we assume that the satellite has the intelligence to decide whether the packets received from the sender are correct or not, and hence only broadcasts the correct ones. In this case, the asymptotic throughput is

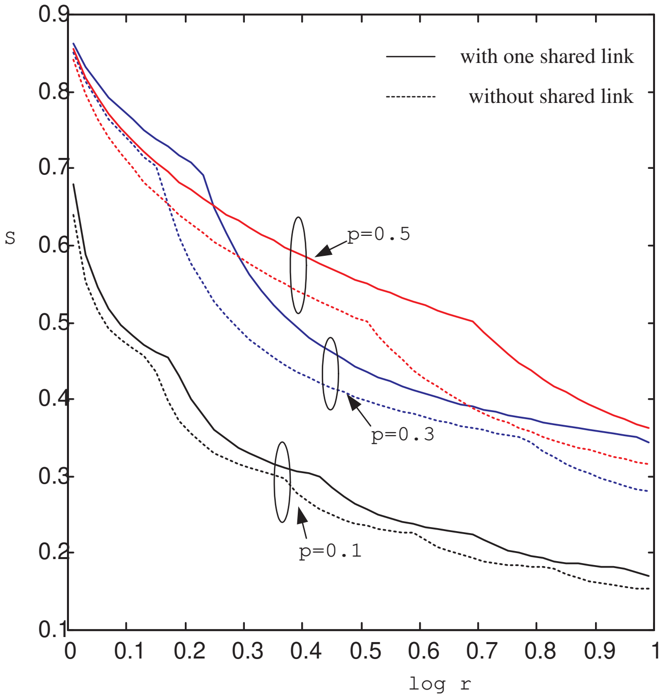

Fig. 9. Comparison of the asymptotic throughput when fixing the fraction of redundancy p.

$\widetilde{S}/(1-q_1)$ (by counting the number of transmissions on the dedicated links). In the third case, we further assume that the satellite is capable of *error recovery*, i.e., it collects enough packets to recover the original data and then broadcasts the packets to the receivers using the same FEC code. Hence, the asymptotic throughput $\widetilde{S}$ is computed with $q_1 = 0$ and $q_2 = 1/19$ (by counting the number of transmissions on the dedicated links). As shown in Figure 11, the asymptotic throughput increases as the intelligence of the satellite increases. Also, note that the packet discard scheme performs as well as the error recovery scheme when the number of receivers is small. This is intuitive as these two schemes are mathematically equivalent if each of the n packets is sent at most once.

VII. Conclusions

In this paper, we proved several strong laws of large numbers for the asymptotic throughput of a multicasting problem with incremental forward error correction. These strong laws not only

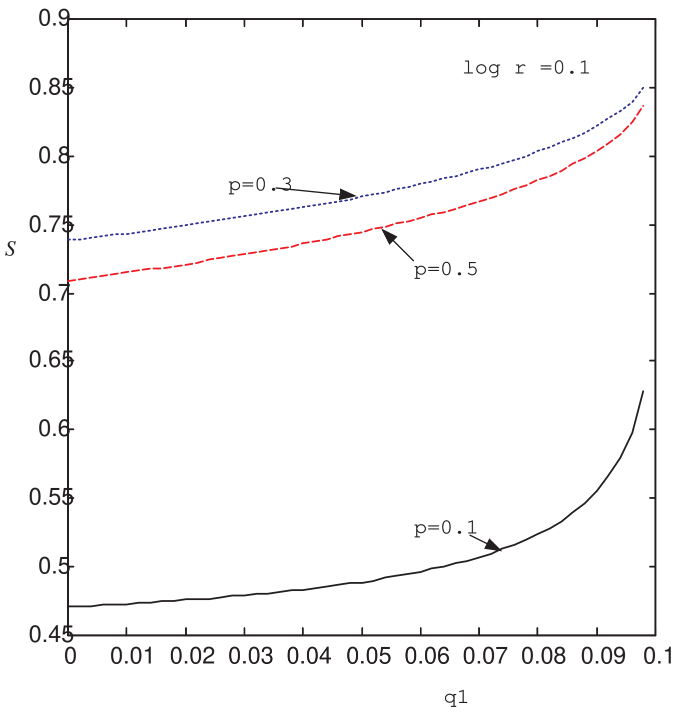

Fig. 10. The asymptotic throughput increases as q_1 increases while keeping the packet loss probability constant.

provide rigorous justification for several important observations reported in the literature, but also provide new insights that might have impact on future design of multicasting protocols.

One possible extension is the tree model in [19]. For instance, consider a two-tier multicasting model in Figure 12. Instead of sending packets directly to the r^n receivers, the sender may send packets to a subset of receivers with size r_1^n and have each of them resend packets to receivers beneath them. As now the numbers of receivers in these two tiers are reduced from r^n to r_1^n and $(r/r_1)^n$, the asymptotic throughput in each tier is better than the direct multicasting model. As a result, the overall bandwidth efficiency in the two-tier model is better than that in the direct multicasting model. This also complies with the numerical results in [19]. However, proving the strong law for the asymptotic throughput may not be as easy as that in Theorem 6. This is due to the fact that the spatial dependence in the tree model cannot be easily decoupled as the satellite model in Theorem 6.

The good thing for the tree model is that we still have the lower bound on the asymptotic through-

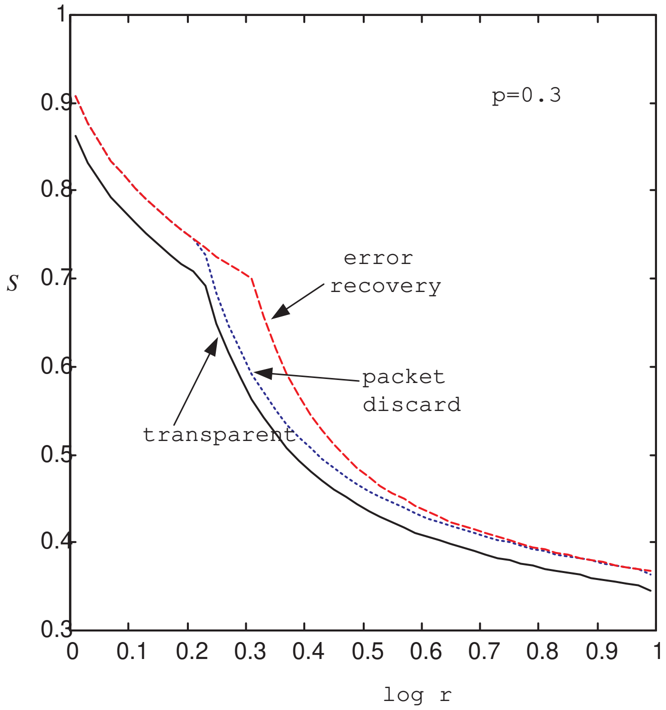

Fig. 11. The asymptotic throughput increases as the intelligence of the intermediate node increases.

put in the direct multicasting model. In fact, such a lower bound holds for any model without temporal loss correlation. Note that our proof for Theorem 1 is still valid for

$$\lim_{n\to\infty} \frac{T(n)}{n} \leq \frac{1-p}{S}, \quad a.s. \tag{34}$$

and S obtained in (5) for the direct multicasting model becomes a lower bound on the asymptotic throughput for the model with spatial loss correlation. (Note that Lemma 8(i)(ii) and Lemma 2 still hold in this case, and these lemmas are used in the proof of Theorem 1 to show $\sum_{n=1}^{\infty} \mathsf{P}(\frac{T(n)}{n} - \frac{1-p}{S} > \epsilon) < \infty$. Hence (34) follows from the Borel-Cantelli lemma.) This is consistent with the observation made in [18] and our numerical results in Section VI.

Acknowledgements:

The first author would like to thank Chi-Chao Chao and Duan-Shin Lee for many helpful discussions on the multicasting problem with incremental FEC.

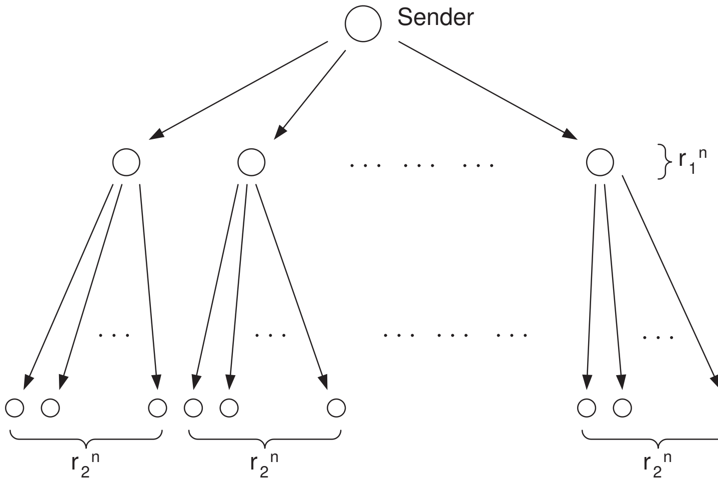

Fig. 12. The two-tier multicasting model

APPENDIX

A. *Proof of Lemma 2*

In this section, we prove Lemma 2.

(i) Obviously, $\Lambda(\theta, \alpha)$ in (2) is continuous in δ. It remains to show that (2) is continuous when α is an integer. Note that there are two representations for an integer α: $\delta = 0, k = \alpha$ and $\delta = 1$, $k = \alpha - 1$. It is straightforward to verify that $\Lambda(\theta, \alpha)$ in (2) yields the same value for both representations and is thus continuous when α is an integer.

(ii) It is easy to see that $\Lambda(\theta, \alpha)$ in (2) is strictly decreasing in both k and δ for $\theta > 0$ and thus strictly decreasing in α. As $q < 1$, letting $k \to \infty$ in (2) yields

$$\lim_{\alpha \to \infty} \Lambda(\theta, \alpha) = 0.$$

(iii) Note that

$$\Lambda'(\theta, \alpha) = \delta \frac{q^{k+1} e^{\theta}}{q^{k+1} e^{\theta} + (1 - q^{k+1})} + (1 - \delta) \frac{q^k e^{\theta}}{q^k e^{\theta} + (1 - q^k)} \tag{35}$$

is strictly increasing in θ. Thus, for $\gamma \leq \Lambda'(0, \alpha)$, the function $\theta\gamma - \Lambda(\theta, \alpha)$ is decreasing in θ for all $\theta \geq 0$ and thus its maximum occurs at $\theta = 0$. This leads to $\Lambda^*(\gamma, \alpha) = 0\gamma - \Lambda(0, \alpha) = 0$ for all $\gamma < \Lambda'(0, \alpha)$ (as $\Lambda(0, \alpha) = 0$).

(iv) For $\Lambda'(0,\alpha) < \gamma < 1$, the maximum of the function $\theta\gamma - \Lambda(\theta,\alpha)$ occurs uniquely at a positive $\theta(\gamma,\alpha)$ in (7), which can be easily obtained from solving θ in the equation

$$\Lambda'(\theta,\alpha) = \gamma.$$

As the maximum is unique,

$$\Lambda^*(\gamma,\alpha) > \Lambda^*(\Lambda'(0,\alpha),\alpha) = 0$$

for $\Lambda'(0,\alpha) < \gamma < 1$.

(v) This is a direct result of (iii) and (iv).

(vi) Observe that $\theta(\gamma,\alpha)$ in (7) is continuous in α. That $\Lambda^*(\gamma,\alpha)$ is continuous follows from (v) and the continuity of $\theta(\gamma,\alpha)$ and $\Lambda(\theta,\alpha)$ in (i).

(vii) From (3), (ii),(iv), it follows that for any $\epsilon > 0$

$$\begin{aligned}
&\Lambda^*(\gamma,\alpha+\epsilon) \geq \theta(\gamma,\alpha)\gamma - \Lambda(\theta(\gamma,\alpha),\alpha+\epsilon) \\
&> \theta(\gamma,\alpha)\gamma - \Lambda(\theta(\gamma,\alpha),\alpha) \\
&= \Lambda^*(\gamma,\alpha).
\end{aligned} \tag{36}$$

It remains to show that

$$\Lambda^*(\gamma,\alpha-\epsilon) < \Lambda^*(\gamma,\alpha). \tag{37}$$

If $\Lambda^*(\gamma,\alpha-\epsilon) = 0$, then (37) holds trivially as $\Lambda^*(\gamma,\alpha) > 0$. On the other hand, if $\Lambda^*(\gamma,\alpha-\epsilon) > 0$, we can use the argument in (36) to show (37).

(viii) Since $\Lambda^*(\gamma,\alpha) \geq \theta\gamma - \Lambda(\theta,\alpha)$ for all $\theta \geq 0$, we have from (i) that for all $\theta \geq 0$

$$\lim_{\alpha\to\infty} \Lambda^*(\gamma,\alpha) \geq \theta\gamma - \lim_{\alpha\to\infty} \Lambda(\theta,\alpha) = \theta\gamma.$$

Letting $\theta \to \infty$ completes the argument.

B. Proof of Theorem 1

In this section, we prove Theorem 1. We first show the solution in (5) is unique in the following lemma.

Lemma 7 *For $r > 1$ and $0 < p < 1$, there is a unique solution S of (5). Moreover, $0 < S < 1 - q$.*

Proof. The existence of a unique solution follows from Lemma 2 (vi), (vii) and (viii). That $S > 0$ is trivial as $\lim_{\alpha\to\infty} \Lambda^*(p,\alpha) = \infty$. To prove $S < 1 - q$, it suffices to show that $\Lambda^*(p, \frac{1-p}{1-q}) = 0$. As shown in Lemma 2(iii), a sufficient condition for $\Lambda^*(p, \frac{1-p}{1-q}) = 0$ is

$$\Lambda'(0, \frac{1-p}{1-q}) \geq p. \tag{38}$$

Let $k = \lfloor \frac{1-p}{1-q} \rfloor$ and $\delta = \frac{1-p}{1-q} - k$. Note from (35) that

$$\begin{aligned}
&\Lambda'(0, \frac{1-p}{1-q}) = \delta q^{k+1} + (1-\delta) q^k \\
&= q^k \Big((\frac{1-p}{1-q} - k) q + 1 - \frac{1-p}{1-q} + k \Big) \\
&= q^k (k(1-q) + p).
\end{aligned}$$

Let $p' = k(1-q) + p$. Then

$$p' = \lfloor \frac{1-p}{1-q} \rfloor (1-q) + p \leq 1.$$

Note that now

$$\Lambda'(0, \frac{1-p}{1-q}) = q^k p'$$

and that

$$p = p' - k(1-q).$$

Thus, it suffices to show that for all $p' < 1$, $q < 1$ and k that

$$q^k p' \geq p' - k(1-q). \tag{39}$$

We show this by induction on k. The above inequality holds trivially for $k = 0$. Assume it holds for $k = m$ as the induction hypothesis. Then

$$\begin{aligned}
&q^{m+1} p' - p' + (m+1)(1-q) \\
&= q(q^m p') - p' + (m+1)(1-q) \\
&\geq q(p' - m(1-q)) - p' + (m+1)(1-q) \\
&= (1-q)\Big(m(1-q) + (1-p') \Big) \geq 0.
\end{aligned}$$

■

Recall that $Z_{i,j}(m)$, $i = 1, \ldots, r^n$, $j = 1, \ldots, n$, is the indicator random variable for the event that the i^{th} receiver has not received the j^{th} packet when m packets are sent by the sender, and that $Y_i(\alpha) = \sum_{j=1}^{n} Z_{i,j}(\lfloor \alpha \rfloor)$ is the number of packets that have not been received by the i^{th} receiver after $\lfloor \alpha \rfloor$ packets are sent by the sender. As $T(n)$ is the number of packets that the sender sends to ensure all the receivers receive at least $n(1-p)$ distinct packets, we then have

$$T(n) = \inf\{m : \max_{1 \le i \le r^n} [Y_i(m)] < pn\}. \tag{40}$$

In the following lemma, we use the large deviation bounds to derive some probabilistic bounds that will be used in the proof of Theorem 1.

Lemma 8 (i) *For all $i = 1, \ldots, r^n$,*

$$\lim_{n \to \infty} \frac{1}{n} \log \mathsf{E} \exp(\theta Y_i(\alpha n)) = \Lambda(\theta, \alpha), \tag{41}$$

where $\Lambda(\theta, \alpha)$ is defined in (2).

(ii)

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathsf{P}(\max_{1 \le i \le r^n} [Y_i(\alpha n)] \ge \gamma n) \le \log r - \Lambda^*(\gamma, \alpha). \tag{42}$$

(iii) *If $\gamma > \Lambda'(0, \alpha)$, then for every $\epsilon > 0$, there exists an $n_0 < \infty$ such that for all $n \ge n_0$*

$$\mathsf{P}(Y_1(\alpha n) > \gamma n) \ge \exp\Big(- n(\Lambda^*(\gamma, \alpha) + \epsilon)\Big). \tag{43}$$

(iv) *If $\gamma > \Lambda'(0, \alpha)$, then for every $\epsilon > 0$, there exists an $n_0 < \infty$ such that for all $n \ge n_0$*

$$\log \mathsf{P}(\max_{1 \le i \le r^n} [Y_i(\alpha n)] \le \gamma n) \le - \exp\Big(n(\log r - \Lambda^*(\gamma, \alpha) - \epsilon)\Big). \tag{44}$$

(v) *For an integer $k \ge \log \gamma / \log q$,*

$$\mathsf{P}(\max_{1 \le i \le r^n} [Y_i(kn)] \ge \gamma n) \le \exp\Big(n(\log r + \mathcal{H}(\gamma) - k\gamma \log \frac{1}{q})\Big), \tag{45}$$

where

$$\mathcal{H}(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma) \tag{46}$$

is the well-known entropy of a Bernoulli random variable with parameter γ (see e.g., [9]).

Proof. (i) Without loss of generality, assume that αn is an integer. As $k = \lfloor \alpha \rfloor$ is the integral part of α and $\delta = \alpha - k$ is the fractional part of α, the packets $j = 1, \ldots, \delta n$, are sent $k+1$ times and the remaining packets are sent k times. Thus,

$$\mathsf{P}(Z_{i,j}(\alpha n) = 1) = \begin{cases} q^{k+1} & \text{if } 1 \leq j \leq \delta n \\ q^k & \text{if } \delta n + 1 \leq j \leq n. \end{cases}$$

As packets are lost independently,

$$Y_i(\alpha n) = \sum_{j=1}^{\delta n} Z_{i,j}(\alpha n) + \sum_{j=\delta n+1}^{n} Z_{i,j}(\alpha n) \tag{47}$$

is simply the sum of independent Bernoulli random variables. Computing the moment generating function of $Y_i(\alpha n)$ yields the desired result.

(ii) Note from Chernoff's bound and the inequality $\max[x_1, x_2] \leq x_1 + x_2, \quad \forall\, x_1, x_2 \geq 0$ that for all $\theta \geq 0$

$$\begin{aligned} \mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(\alpha n)] \geq \gamma n) &\leq e^{-\gamma n\theta}\mathsf{E}\exp(\theta \max_{1\leq i\leq r^n}[Y_i(\alpha n)]) \\ &= e^{-\gamma n\theta}\mathsf{E}\max_{1\leq i\leq r^n}[\exp(\theta Y_i(\alpha n))] \leq e^{-\gamma n\theta}\sum_{1\leq i\leq r^n}\mathsf{E}\exp(\theta Y_i(\alpha n)) \end{aligned}$$

As $Y_i(\alpha n)$, $i = 1, \ldots, r^n$ are identically distributed,

$$\mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(\alpha n)] \geq \gamma n) \leq \exp\Big(n\big(\log r - \theta\gamma + \frac{1}{n}\log \mathsf{E}\exp(\theta Y_1(\alpha n))\big)\Big). \tag{48}$$

We then have from (i) that for all $\theta \geq 0$

$$\limsup_{n\to\infty}\frac{1}{n}\log \mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(\alpha n)] \geq \gamma n) \leq \log r - \gamma\theta + \Lambda(\theta, \alpha).$$

Choosing the best bound among $\theta \geq 0$ yields the desired result.

(iii) Note that $\Lambda(\theta, \alpha)$ is finite and differentiable for all θ. From (i) and the Gärtner-Ellis theorem (see e.g., [4], [10], [23]), it follows that $\{Y_1(\alpha n)/n, n \geq 1\}$ satisfies the large deviation principle with the rate function $\tilde{\Lambda}^*(\gamma, \alpha)$, where $\tilde{\Lambda}^*(\gamma, \alpha) = \sup_\theta[\theta\gamma - \Lambda(\theta, \alpha)]$ is the two-sided Legendre transform. Using the large deviation lower bound, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathsf{P}(Y_1(\alpha n) > \gamma n) \geq -\inf_{\gamma_1 > \gamma} \tilde{\Lambda}^*(\gamma_1, \alpha). \tag{49}$$

Under the condition $\gamma > \Lambda'(0, \alpha)$, it is well-known that the two-sided transform is the same as the one-sided transform (as the maximum occurs at some $\theta > 0$). Moreover,

$$\inf_{\gamma_1 > \gamma} \tilde{\Lambda}^*(\gamma_1, \alpha) = \inf_{\gamma_1 > \gamma} \Lambda^*(\gamma_1, \alpha) = \Lambda^*(\gamma, \alpha), \tag{50}$$

as $\Lambda^*(\gamma, \alpha)$ is increasing and continuous when $0 < \Lambda^*(\gamma, \alpha) < \infty$. The result then follows from (49) and (50).

(iv) Since $Y_i(\alpha n)$, $i = 1, \ldots, r^n$, are independent and identically distributed,

$$\begin{aligned}
&\mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(\alpha n)] \leq \gamma n) = \mathsf{P}(Y_1(\alpha n) \leq \gamma n)^{r^n} \\
&= (1 - \mathsf{P}(Y_1(\alpha n) > \gamma n))^{r^n} = \exp\Big(r^n \log(1 - \mathsf{P}(Y_1(\alpha n) > \gamma n))\Big).
\end{aligned}$$

Using the inequality $\log x \leq x - 1$ and (iii) in this lemma, we have for every $\epsilon > 0$ and $n \geq n_0$ that

$$\log \mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(\alpha n)] \leq \gamma n) \leq -r^n \mathsf{P}(Y_1(\alpha n) > \gamma n) \leq -e^{n(\log r - \Lambda^*(\gamma,\alpha) - \epsilon)}.$$

(v) Note from (47) that

$$\mathsf{E} \exp(\theta Y_1(kn)) = (q^k e^\theta + 1 - q^k)^n.$$

In conjunction with the inequality in (48), we have

$$\mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(kn)] \geq \gamma n) \leq \exp\Big(n(\log r - \theta\gamma + \log(q^k e^\theta + 1 - q^k))\Big).$$

Choosing the best bound for $\theta \geq 0$ yields

$$\mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(kn)] \geq \gamma n) \leq \exp\Big(n(\log r - \sup_{\theta\geq 0}[\theta\gamma - \log(q^k e^\theta + 1 - q^k)])\Big).$$

Since $k \geq \log \gamma / \log q$, we then have

$$\mathsf{P}(\max_{1\leq i\leq r^n}[Y_i(kn)] \geq \gamma n) \leq \exp\Big(n(\log r - D(\gamma || q^k))\Big),$$

where $D(p_1||p_2)$ is the Kullback-Leibler distance in (13). The proof is completed by observing that

$$D(\gamma||q^k) = \gamma \log \frac{\gamma}{q^k} + (1-\gamma) \log \frac{1-\gamma}{1-q^k}$$
$$\geq \gamma \log \frac{\gamma}{q^k} + (1-\gamma) \log (1-\gamma) = -\mathcal{H}(\gamma) + k\gamma \log \frac{1}{q}.$$

■

Proof. (Theorem 1) For the strong law, it is equivalent to showing that

$$\lim_{n\to\infty} \frac{T(n)}{n} = \frac{1-p}{S}, \quad a.s. \tag{51}$$

We will prove (51) by showing that for every $\epsilon > 0$

$$\sum_{n=1}^{\infty} \mathsf{P}(|\frac{T(n)}{n} - \frac{1-p}{S}| > \epsilon) < \infty. \tag{52}$$

The limit in (51) then follows from the well-known Borel-Cantelli lemma (see e.g., [8]).

As $\Lambda^*(p, \frac{1-p}{S}) = \log r > 0$ and $\Lambda^*(p, \alpha)$ is continuous in α, we can choose $\epsilon > 0$ small enough such that $\Lambda^*(p, \frac{1-p}{S} - \epsilon) > 0$. Let

$$\epsilon_1 = \Lambda^*(p, \frac{1-p}{S}) - \Lambda^*(p, \frac{1-p}{S} - \epsilon),$$

and

$$\epsilon_2 = \Lambda^*(p, \frac{1-p}{S} + \epsilon) - \Lambda^*(p, \frac{1-p}{S}).$$

It follows from the strictly increasing property in Lemma 2(vii) that both $\epsilon_1 > 0$ and $\epsilon_2 > 0$.

Note that

$$\mathsf{P}(T(n) < n(\frac{1-p}{S} - \epsilon)) = \mathsf{P}\Big(\max_{1\leq i\leq r^n} [Y_i((\frac{1-p}{S} - \epsilon)n)] < pn\Big).$$

As $\Lambda^*(p, \frac{1-p}{S} - \epsilon) > 0$, we have from Lemma 2(iii) that $p > \Lambda'(0, \frac{1-p}{S} - \epsilon)$. From Lemma 8 (iv), there exists an $n_1 < \infty$ such that for $\epsilon_1/2 > 0$ and for all $n \geq n_1$

$$\mathsf{P}(T(n) < n(\frac{1-p}{S} - \epsilon)) \leq \exp\Big(-\exp\big(n(\log r - \Lambda^*(p, \frac{1-p}{S} - \epsilon) - \frac{\epsilon_1}{2}))\Big)$$

$$= \exp\Big(-\exp(n(\log r - \Lambda^*(p, \frac{1-p}{S}) + \frac{\epsilon_1}{2}))\Big)$$
$$= \exp\Big(-\exp(n\frac{\epsilon_1}{2})\Big).$$

Thus,

$$\sum_{n=1}^{\infty} \mathsf{P}(\frac{T(n)}{n} - \frac{1-p}{S} < -\epsilon) < \infty.$$

On the other hand, we also note that

$$\mathsf{P}(T(n) > n(\frac{1-p}{S} + \epsilon)) = \mathsf{P}\Big(\max_{1 \le i \le r^n} [Y_i((\frac{1-p}{S} + \epsilon)n)] \ge pn\Big).$$

Using Lemma 8(ii), there exists an $n_2 < \infty$ such that for $\epsilon_2/2 > 0$ and for all $n \ge n_2$

$$\begin{aligned} &\mathsf{P}(T(n) > n(\frac{1-p}{S} + \epsilon)) \le \exp\left(n(\log r - \Lambda^*(p, \frac{1-p}{S} + \epsilon) + \frac{\epsilon_2}{2})\right) \\ &= \exp\left(n(\log r - \Lambda^*(p, \frac{1-p}{S}) - \frac{\epsilon_2}{2})\right) \\ &= \exp(-n\frac{\epsilon_2}{2}). \end{aligned} \tag{53}$$

Thus, we also have

$$\sum_{n=1}^{\infty} \mathsf{P}(\frac{T(n)}{n} - \frac{1-p}{S} > \epsilon) < \infty.$$

Now we show that

$$\lim_{n\to\infty} \frac{\mathsf{E}T(n)}{n} = \frac{1-p}{S}. \tag{54}$$

From Fatou's lemma (see e.g., [8]) and the strong law in (51), we know that

$$\lim_{n\to\infty} \frac{\mathsf{E}T(n)}{n} \ge \mathsf{E} \lim_{n\to\infty} \frac{T(n)}{n} = \frac{1-p}{S}.$$

We only need to show the other inequality. Choose an integer k_1 large enough such that

$$k_1 > \max[\frac{1-p}{S}, \frac{\log p}{\log q}, \frac{\log r + \mathcal{H}(p)}{p \log \frac{1}{q}}]. \tag{55}$$

Choose $\epsilon > 0$ small enough such that $\frac{1-p}{S} + \epsilon < k_1$. Thus, we have

$$\mathsf{E}T(n) = \sum_{\ell=1}^{\infty} \mathsf{P}(T(n) \ge \ell)$$

$$= \sum_{\ell=1}^{\lfloor(\frac{1-p}{S}+\epsilon)n\rfloor} \mathsf{P}(T(n) \geq \ell) + \sum_{\ell=\lfloor(\frac{1-p}{S}+\epsilon)n\rfloor+1}^{nk_1} \mathsf{P}(T(n) \geq \ell) + \sum_{\ell=nk_1+1}^{\infty} \mathsf{P}(T(n) \geq \ell)$$

$$\leq (\frac{1-p}{S}+\epsilon)n + nk_1 \mathsf{P}(T(n) \geq (\frac{1-p}{S}+\epsilon)n) + \sum_{\ell=nk_1+1}^{\infty} \mathsf{P}(T(n) \geq \ell) \tag{56}$$

Note from Lemma 8(v) that

$$\begin{aligned}
&\sum_{\ell=nk_1+1}^{\infty} \mathsf{P}(T(n) \geq \ell) \\
&= \sum_{k=k_1}^{\infty} \sum_{i=1}^{n} \mathsf{P}(T(n) \geq nk+i) \\
&\leq \sum_{k=k_1}^{\infty} n\mathsf{P}(T(n) \geq nk). \\
&\leq n \sum_{k=k_1}^{\infty} \exp(n(\log r + \mathcal{H}(p) - kp\log\frac{1}{q})) \\
&= n\exp\Big(n(\log r + \mathcal{H}(p) - k_1 p\log\frac{1}{q})\Big)\Big(1-\exp(-np\log\frac{1}{q})\Big)^{-1}.
\end{aligned}$$

In conjunction with (56), we then have

$$\begin{aligned}
\frac{\mathsf{E}T(n)}{n} \leq \frac{1-p}{S} + \epsilon + k_1\mathsf{P}(T(n) \geq (\frac{1-p}{S}+\epsilon)n) \\
+ \exp\Big(n(\log r + \mathcal{H}(p) - k_1 p\log\frac{1}{q})\Big)\Big(1-\exp(-np\log\frac{1}{q})\Big)^{-1}.
\end{aligned}$$

Note from (53) that

$$\lim_{n\to\infty} \mathsf{P}(T(n) \geq (\frac{1-p}{S}+\epsilon)n) = 0.$$

As k_1 in (55) is large enough to ensure that

$$\log r + \mathcal{H}(p) - k_1 p \log\frac{1}{q} < 0,$$

letting $n \to \infty$ and then $\epsilon \to 0$ yields

$$\lim_{n\to\infty} \frac{\mathsf{E}T(n)}{n} \leq \frac{1-p}{S}.$$

■

C. Proof of Theorem 3

In this section, we prove Theorem 3. The proof is quite similar to that in Theorem 1. From the well-known Borel-Cantelli lemma, it suffices to show that for every $\epsilon > 0$

$$\sum_{n=1}^{\infty} \mathsf{P}(|\frac{1}{n} \max_{1\le i\le r^n} Y_i(\frac{1-p}{S}n) - p| > \epsilon) < \infty. \tag{57}$$

As $\Lambda^*(p, \frac{1-p}{S}) = \log r > 0$ and $\Lambda^*(p,\alpha)$ is continuous in p, we can choose $\epsilon > 0$ small enough such that $\Lambda^*(p-\epsilon, \frac{1-p}{S}) > 0$ and $\Lambda^*(p+\epsilon, \frac{1-p}{S}) < \infty$. Let

$$\epsilon_1 = \Lambda^*(p, \frac{1-p}{S}) - \Lambda^*(p-\epsilon, \frac{1-p}{S}),$$

and

$$\epsilon_2 = \Lambda^*(p+\epsilon, \frac{1-p}{S}) - \Lambda^*(p, \frac{1-p}{S}).$$

Since $\Lambda^*(\gamma,\alpha)$ is strictly increasing in γ (when $0 < \Lambda^*(\gamma,\alpha) < \infty$), we have both $\epsilon_1 > 0$ and $\epsilon_2 > 0$.

As $\Lambda^*(p-\epsilon, \frac{1-p}{S}) > 0$, we have from Lemma 2(iii) that $p-\epsilon > \Lambda'(0, \frac{1-p}{S})$. From Lemma 8 (iv), there exists an $n_1 < \infty$ such that for $\epsilon_1/2 > 0$ and for all $n \ge n_1$

$$\begin{aligned}
&\mathsf{P}\Big(\max_{1\le i\le r^n}[Y_i(\frac{1-p}{S}n)] < (p-\epsilon)n\Big)\\
&\le \exp\Big(-\exp\big(n(\log r - \Lambda^*(p-\epsilon, \frac{1-p}{S}) - \frac{\epsilon_1}{2}))\Big)\\
&= \exp\Big(-\exp(n(\log r - \Lambda^*(p, \frac{1-p}{S}) + \frac{\epsilon_1}{2}))\Big)\\
&= \exp\Big(-\exp(n\frac{\epsilon_1}{2})\Big).
\end{aligned}$$

Thus,

$$\sum_{n=1}^{\infty} \mathsf{P}(\frac{1}{n} \max_{1\le i\le r^n}[Y_i(\frac{1-p}{S}n)] - p < -\epsilon) < \infty.$$

Using Lemma 8(ii), there exists an $n_2 < \infty$ such that for $\epsilon_2/2 > 0$ and for all $n \ge n_2$

$$\begin{aligned}
&\mathsf{P}\Big(\max_{1\le i\le r^n}[Y_i(\frac{1-p}{S}n)] \ge (p+\epsilon)n\Big)\\
&\le \exp(n(\log r - \Lambda^*(p+\epsilon, \frac{1-p}{S}) + \frac{\epsilon_2}{2}))\\
&= \exp(n(\log r - \Lambda^*(p, \frac{1-p}{S}) - \frac{\epsilon_2}{2}))\\
&= \exp(-n\frac{\epsilon_2}{2}).
\end{aligned}$$

Thus, we also have

$$\sum_{n=1}^{\infty} \mathsf{P}(\frac{1}{n} \max_{1 \le i \le r^n} [Y_i(\frac{1-p}{S} n)] - p > \epsilon) < \infty.$$

D. Proof of Theorem 6

In this section, we prove Theorem 6. As in the proof of Theorem 1, we first establish some useful properties of $\widetilde{\Lambda}(\theta, \alpha)$ and $\widetilde{\Lambda}^*(\gamma, \alpha)$.

Lemma 9 (i) *$\widetilde{\Lambda}(\theta, \alpha)$ is continuous in $\alpha > 0$.*

(ii) *For $\theta > 0$, $\widetilde{\Lambda}(\theta, \alpha)$ is strictly decreasing in α to 0.*

(iii) *For $\gamma \le \widetilde{\Lambda}'(0, \alpha)$, $\widetilde{\Lambda}^*(\gamma, \alpha) = 0$.*

(iv) *For $\widetilde{\Lambda}'(0, \alpha) < \gamma < 1$,*

$$\widetilde{\Lambda}^*(\gamma, \alpha) = \theta(\gamma, \alpha)\gamma - \widetilde{\Lambda}(\theta(\gamma, \alpha), \alpha) > 0, \tag{58}$$

where $\theta(\gamma, \alpha) > 0$ is obtained by solving θ in the equation

$$\widetilde{\Lambda}'(\theta, \alpha) = \gamma \tag{59}$$

(v) *For a fixed $0 < \gamma < 1$,*

$$\widetilde{\Lambda}^*(\gamma, \alpha) = \theta(\gamma, \alpha)^+\gamma - \widetilde{\Lambda}(\theta(\gamma, \alpha)^+, \alpha), \tag{60}$$

where $x^+ = \max(0, x)$ and $\theta(\gamma, \alpha)$ is obtained by solving (59).

(vi) *For a fixed $0 < \gamma < 1$, $\widetilde{\Lambda}^*(\gamma, \alpha)$ is continuous in $\alpha > 0$.*

(vii) *For a fixed $0 < \gamma < 1$, if $\widetilde{\Lambda}^*(\gamma, \alpha) > 0$, then $\widetilde{\Lambda}^*(\gamma, \alpha)$ is strictly increasing in $\alpha > 0$.*

(viii) *For a fixed $0 < \gamma < 1$,*

$$\lim_{\alpha \to \infty} \widetilde{\Lambda}^*(\gamma, \alpha) = \infty.$$

Proof. We only prove (ii), (iii) and (vi). The rests of the lemma follow from the same arguments in the proof of Lemma 2.

(ii) Since $\widetilde{\Lambda}(\theta, \alpha) = \delta\widetilde{\Lambda}(\theta, k+1) + (1-\delta)\widetilde{\Lambda}(\theta, k)$, it suffices to show that $\widetilde{\Lambda}(\theta, k+1) < \widetilde{\Lambda}(\theta, k)$ and $\lim_{k\to\infty} \widetilde{\Lambda}(\theta, k) = 0$ for $\theta > 0$. Using the identity $\begin{pmatrix} k+1 \\ i \end{pmatrix} = \begin{pmatrix} k \\ i \end{pmatrix} + \begin{pmatrix} k \\ i-1 \end{pmatrix}$, and the inequality $\log(q_2^{i+1}e^\theta + (1-q_2^{i+1})) < \log(q_2^i e^\theta + (1-q_2^i))$ for $\theta > 0$, we have

$$
\begin{aligned}
\widetilde{\Lambda}(\theta, k+1) &= q_1\widetilde{\Lambda}(\theta,k) + (1-q_1)\sum_{i=0}^{k} b_{k,i}\log(q_2^{i+1}e^\theta + (1-q_2^{i+1})) \\
&< q_1\widetilde{\Lambda}(\theta,k) + (1-q_1)\sum_{i=0}^{k} b_{k,i}\log(q_2^{i}e^\theta + (1-q_2^{i})) \\
&= q_1\widetilde{\Lambda}(\theta,k) + (1-q_1)\widetilde{\Lambda}(\theta,k) \\
&= \widetilde{\Lambda}(\theta,k),
\end{aligned}
$$

where $b_{k,i}$ is the binomial distribution coefficient defined in (27).

Now we write

$$
\widetilde{\Lambda}(\theta,k) = \sum_{i=0}^{N} b_{k,i}\log(q_2^i e^\theta + (1-q_2^i)) + \sum_{i=N+1}^{k} b_{k,i}\log(q_2^i e^\theta + (1-q_2^i)). \qquad (61)
$$

Since $\lim_{i\to\infty}\log(q_2^i e^\theta + (1-q_2^i)) = 0$, for any $\epsilon > 0$ we may choose N large enough so that $\log(q_2^i e^\theta + (1-q_2^i) < \epsilon$ for $i > N$. Thus, the second term on the right hand side of (61) is at most ϵ as $\sum_{i=0}^{k} b_{k,i} = 1$. Also, observe that $\lim_{k\to\infty} b_{k,i} = 0$ for any fixed i and thus the first term approaches 0 as $k \to \infty$ for any fixed N. It then follows that $\lim_{k\to\infty}\widetilde{\Lambda}(\theta,k) \le \epsilon$. Since ϵ is arbitrary, letting $\epsilon \to 0$ yields that $\lim_{k\to\infty}\widetilde{\Lambda}(\theta,k) = 0$.

(iii) Note that

$$
\widetilde{\Lambda}'(\theta,\alpha) = \delta\sum_{i=0}^{k+1} b_{k+1,i}\frac{q_2^i e^\theta}{q_2^i e^\theta + (1-q_2^i)} + (1-\delta)\sum_{i=0}^{k} b_{k,i}\frac{q_2^i e^\theta}{q_2^i e^\theta + (1-q_2^i)} \qquad (62)
$$

is strictly increasing in θ to 1. Thus, for $\gamma \le \widetilde{\Lambda}'(0,\alpha)$, the function $\theta\gamma - \widetilde{\Lambda}(\theta,\alpha)$ is decreasing in θ for all $\theta \ge 0$ and thus its maximum occurs at $\theta = 0$. This leads to $\widetilde{\Lambda}^*(\gamma,\alpha) = 0\gamma - \widetilde{\Lambda}(0,\alpha) = 0$ for all $\gamma < \widetilde{\Lambda}'(0,\alpha)$ (as $\widetilde{\Lambda}(0,\alpha) = 0$).

(vi) Since $\widetilde{\Lambda}'(\theta,\alpha)$ in (62) is continuous in α and in fact is continuously differentiable in θ, it follows from the implicit function theorem (see e.g., Theorem 7.21 in [1]) that $\theta(\gamma,\alpha)$ in (59) is continuous in α . That $\widetilde{\Lambda}^*(\gamma,\alpha)$ is continuous follows from (v) and the continuity of $\theta(\gamma,\alpha)$ and $\widetilde{\Lambda}(\theta,\alpha)$ in (i).

■

We show that the solution in (30) is unique in the following lemma.

Lemma 10 *For $r > 1$ and $0 < p < 1$, there is a unique solution $\widetilde{S}$ of (30). Moreover, $0 < \widetilde{S} < 1 - q$, where $q = q_1 + q_2 - q_1 q_2$ is the end-to-end loss probability for a packet to go through the shared link and the dedicated link.*

Proof. The existence of a unique solution follows from Lemma 9 (vi), (vii) and (viii). That $\widetilde{S} > 0$ is trivial as $\lim_{\alpha \to \infty} \widetilde{\Lambda}^*(p, \alpha) = \infty$. To prove $\widetilde{S} < 1 - q$, it suffices to show that $\widetilde{\Lambda}^*(p, \frac{1-p}{1-q}) = 0$. As shown in Lemma 9(iii), a sufficient condition for $\widetilde{\Lambda}^*(p, \frac{1-p}{1-q}) = 0$ is

$$\widetilde{\Lambda}'(0, \frac{1-p}{1-q}) \geq p. \tag{63}$$

Let $k = \lfloor \frac{1-p}{1-q} \rfloor$ and $\delta = \frac{1-p}{1-q} - k$. Note from (62) that

$$\begin{aligned}
&\widetilde{\Lambda}'(0, \frac{1-p}{1-q}) = \delta \sum_{i=0}^{k+1} b_{k+1,i} q_2^i + (1 - \delta) \sum_{i=0}^{k} b_{k,i} q_2^i \\
&= \delta q^{k+1} + (1 - \delta) q^k
\end{aligned}$$

Hence (63) follows from the proof in Lemma 7.

■

In the following lemma, we use the large deviation bounds to derive some probabilistic bounds that will be used in the proof of Theorem 6.

Lemma 11 (i) *For all $i = 1, \ldots, r^n$,*

$$\lim_{n \to \infty} \frac{1}{n} \log \mathsf{E} \exp(\theta Y_i(n\underline{\delta})) = \Lambda(\theta, \underline{\delta}), \tag{64}$$

where $\Lambda(\theta, \underline{\delta}) = \sum_{i=0}^{\lceil \alpha \rceil} \delta_i \log(q_2^i e^\theta + (1 - q_2^i))$.

(ii)

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathsf{P}(\max_{1 \leq i \leq r^n} [Y_i(n\underline{\delta})] \geq \gamma n) \leq \log r - \Lambda^*(\gamma, \underline{\delta}), \tag{65}$$

where $\Lambda^*(\gamma, \underline{\delta}) = \sup_{\theta \geq 0} [\theta\gamma - \Lambda(\theta, \underline{\delta})]$ is the one-sided Legendre transform of $\Lambda(\theta, \underline{\delta})$.

(iii) *If* $\gamma > \Lambda'(0, \underline{\delta})$, *then for every* $\epsilon > 0$, *there exists an* $n_0 < \infty$ *such that for all* $n \geq n_0$

$$\mathsf{P}(Y_i(n\underline{\delta}) > \gamma n) \geq \exp\Big(- n(\Lambda^*(\gamma, \underline{\delta}) + \epsilon)\Big). \tag{66}$$

(iv) *If* $\gamma > \Lambda'(0, \underline{\delta})$, *then for every* $\epsilon > 0$, *there exists an* $n_0 < \infty$ *such that for all* $n \geq n_0$

$$\log \mathsf{P}(\max_{1 \leq i \leq r^n} [Y_i(n\underline{\delta})] \leq \gamma n) \leq -\exp\Big(n(\log r - \Lambda^*(\gamma, \underline{\delta}) - \epsilon)\Big). \tag{67}$$

Proof. (i) This is a direct consequence of (33). The rests of the lemma follow from the same arguments in the proof of Lemma 8.

■

Note that $\Lambda(\theta, \underline{b}_\alpha)$ is just $\widetilde{\Lambda}(\theta, \alpha)$ defined in (26). Also, $\Lambda^*(\gamma, \underline{b}_\alpha)$ is just $\widetilde{\Lambda}^*(\gamma, \alpha)$ defined in (28). In the following, we show some continuous properties of $\Lambda^*(\gamma, \underline{\delta})$ and $\Lambda'(\theta, \underline{\delta})$. These properties are needed in our proof of Theorem 6.

Lemma 12 (i) *For all* $0 < \gamma < 1$ *and* $\alpha > 0$,

$$\lim_{\|\underline{\delta} - \underline{b}_\alpha\|_1 \to 0} \Lambda^*(\gamma, \underline{\delta}) = \widetilde{\Lambda}^*(\gamma, \alpha), \tag{68}$$

where $\|\underline{\delta} - \underline{b}_\alpha\|_1 = \sum_{i=0}^{\lceil \alpha \rceil} |\delta_i - b_{\alpha,i}|$ *is the* L_1 *norm of* $(\underline{\delta} - \underline{b}_\alpha)$.

(ii) *For any fixed* $\alpha > 0$,

$$\lim_{\|\underline{\delta} - \underline{b}_\alpha\|_1 \to 0} \Lambda'(0, \underline{\delta}) = \widetilde{\Lambda}'(0, \alpha). \tag{69}$$

Proof. (i) Note that

$$\widetilde{\Lambda}^*(\gamma, \alpha) = \sup_{\theta \geq 0} [\theta\gamma - \sum_{i=0}^{\lceil \alpha \rceil} b_{\alpha,i} \log(q_2^i e^\theta + (1 - q_2^i))].$$

Let θ^* be the argument that achieves the supremum in $\widetilde{\Lambda}^*(\gamma, \alpha)$. Similarly,

$$\Lambda^*(\gamma, \underline{\delta}) = \sup_{\theta \geq 0} [\theta\gamma - \sum_{i=0}^{\lceil \alpha \rceil} \delta_i \log(q_2^i e^\theta + (1 - q_2^i))],$$

and let θ' be the argument that achieves the supremum in $\Lambda^*(\gamma, \underline{\delta})$. Then, we have that

$$
\begin{aligned}
&\Lambda^*(\gamma, \underline{\delta}) - \widetilde{\Lambda}^*(\gamma, \alpha) \\
&= (\theta'\gamma - \sum_{i=0}^{\lceil\alpha\rceil} \delta_i \log(q_2^i e^{\theta'} + (1 - q_2^i))) - (\theta^*\gamma - \sum_{i=0}^{\lceil\alpha\rceil} b_{\alpha,i} \log(q_2^i e^{\theta^*} + (1 - q_2^i))) \\
&\leq (\theta'\gamma - \sum_{i=0}^{\lceil\alpha\rceil} \delta_i \log(q_2^i e^{\theta'} + (1 - q_2^i))) - (\theta'\gamma - \sum_{i=0}^{\lceil\alpha\rceil} b_{\alpha,i} \log(q_2^i e^{\theta'} + (1 - q_2^i))) \\
&= \sum_{i=0}^{\lceil\alpha\rceil} (b_{\alpha,i} - \underline{\delta}) \log(q_2^i e^{\theta'} + (1 - q_2^i))) \leq \|\underline{\delta} - \underline{b}_\alpha\|_1 \theta'.
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
&\Lambda^*(\gamma, \underline{\delta}) - \widetilde{\Lambda}^*(\gamma, \alpha) \\
&\geq \sum_{i=0}^{\lceil\alpha\rceil} (b_{\alpha,i} - \underline{\delta}) \log(q_2^i e^{\theta^*} + (1 - q_2^i))) \geq -\|\underline{\delta} - \underline{b}_\alpha\|_1 \theta^*.
\end{aligned}
$$

The proof is complete if we can show that θ' is bounded. From the optimality condition, $\gamma = \sum_{i=0}^{\lceil\alpha\rceil} (\delta_i \frac{q_2^i e^{\theta'}}{q_2^i e^{\theta'} + (1-q_2^i)})$, and $(\sum_{i=0}^{\ulcorner\alpha\urcorner} \delta_i) = 1$, we have

$$
\begin{aligned}
\gamma &= \sum_{i=0}^{\ulcorner\alpha\urcorner} (\delta_i \frac{q_2^i e^{\theta'}}{q_2^i e^{\theta'} + (1 - q_2^i)}) \\
&\geq (\sum_{i=0}^{\ulcorner\alpha\urcorner} \delta_i) \frac{q_2^{\ulcorner\alpha\urcorner} e^{\theta'}}{q_2^{\ulcorner\alpha\urcorner} e^{\theta'} + (1 - q_2^{\ulcorner\alpha\urcorner})} \\
&= \frac{q_2^{\ulcorner\alpha\urcorner} e^{\theta'}}{q_2^{\ulcorner\alpha\urcorner} e^{\theta'} + (1 - q_2^{\ulcorner\alpha\urcorner})}.
\end{aligned}
$$

It follows that $\theta' \leq \log(\frac{1-q_2^{\ulcorner\alpha\urcorner}}{q_2^{\ulcorner\alpha\urcorner}} \frac{\gamma}{1-\gamma})$.

(ii) Note that

$$
|\Lambda'(0, \underline{\delta}) - \widetilde{\Lambda}'(0, \alpha)| = |\sum_{i=0}^{\infty} \delta_i q_2^i - \sum_{i=0}^{\infty} b_{\alpha,i} q_2^i| \leq \|\underline{\delta} - \underline{b}_\alpha\|_1.
$$

■

Proof. (Theorem 6)

For the proof of the strong law (29), it is equivalent to showing that

$$\lim_{n\to\infty} \frac{\widetilde{T}(n)}{n} = \frac{1-p}{\widetilde{S}}, \quad a.s. \tag{70}$$

We will prove (70) by showing that for every $\epsilon > 0$

$$\sum_{n=1}^{\infty} \mathsf{P}(|\frac{\widetilde{T}(n)}{n} - \frac{1-p}{\widetilde{S}}| > \epsilon) < \infty. \tag{71}$$

The limit in (70) then follows from the well-known Borel-Cantelli lemma (see e.g., [8]).

As $\widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}}) = \log r > 0$ and $\widetilde{\Lambda}^*(p,\alpha)$ is continuous in α, we can choose $\epsilon > 0$ small enough such that $\widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}} - \epsilon) > 0$. Let

$$\epsilon_1 = \widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}}) - \widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}} - \epsilon), \tag{72}$$

and

$$\epsilon_2 = \widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}} + \epsilon) - \widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}}). \tag{73}$$

It follows from the strictly increasing property in Lemma 9(vii) that both $\epsilon_1 > 0$ and $\epsilon_2 > 0$.

To ease our presentation, let

$$\beta = \frac{1-p}{\widetilde{S}} - \epsilon.$$

Note that

$$\begin{aligned} \mathsf{P}(\widetilde{T}(n) < \beta n) \quad &= \mathsf{P}\Big(\max_{1\le i\le r^n} [\widetilde{Y}_i(\beta n)] < pn\Big) \\ &= \mathsf{P}\Big(\max_{1\le i\le r^n} [Y_i(\underline{N}(\beta n))] < pn\Big). \end{aligned}$$

Now we can decompose the above probability based on the value of $\frac{\underline{N}(\beta n)}{n}$. Since packet losses on the shared link and the packet losses on the dedicated links are independent, it follows that

$$\begin{aligned} &\mathsf{P}\Big(\max_{1\le i\le r^n} [Y_i(\underline{N}(\beta n))] < pn\Big) = \sum_{\underline{\delta}} \mathsf{P}\Big(\underline{N}(\beta n) = n\underline{\delta}, \max_{1\le i\le r^n} [Y_i(n\underline{\delta})] < pn\Big) \\ &= \sum_{\underline{\delta}} \mathsf{P}\Big(\underline{N}(\beta n) = n\underline{\delta}\Big)\mathsf{P}\Big(\max_{1\le i\le r^n} [Y_i(n\underline{\delta})] < pn\Big) \\ &= \sum_{\|\underline{\delta} - \underline{b}_\beta\|_1 \ge \epsilon_3} \mathsf{P}\Big(\underline{N}(\beta n) = n\underline{\delta}\Big)\mathsf{P}\Big(\max_{1\le i\le r^n} [Y_i(n\underline{\delta})] < pn\Big) \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{||\underline{\delta}-\underline{b}_\beta||_1<\epsilon_3} \mathsf{P}\Big(\underline{N}(\beta n)=n\underline{\delta}\Big)\mathsf{P}\Big(\max_{1\le i\le r^n}[Y_i(n\underline{\delta})]<pn\Big)\\
&\le \sum_{||\underline{\delta}-\underline{b}_\beta||_1\ge\epsilon_3} \mathsf{P}\Big(\underline{N}(\beta n)=n\underline{\delta}\Big) + \sup_{||\underline{\delta}-\underline{b}_\beta||_1<\epsilon_3} \mathsf{P}\Big(\max_{1<i\le r^n}[Y_i(n\underline{\delta})]<pn\Big),
\end{aligned}
$$

for any $\epsilon_3 > 0$. From the Sanov's theorem(see e.g., [9]), we have

$$
\sum_{||\underline{\delta}-\underline{b}_\beta||_1\ge\epsilon_3} \mathsf{P}\Big(\underline{N}(\beta n)=n\underline{\delta}\Big) \le \exp(-n \inf_{\{||\underline{\delta}-\underline{b}_\beta||_1\ge\epsilon_3\}} D(\underline{\delta}||\underline{b}_\beta)),
$$

for n large enough, where $D(\underline{\delta}||\underline{b}_\beta) = \sum_{i=0}^{\infty}\delta_i\log(\frac{\delta_i}{b_{\beta,i}})$ is the Kullback-Leibler distance (also known as the relative entropy in the information theory) between two distributions $\underline{\delta}$ and $\underline{b}_\beta$. Since $D(\underline{\delta}||\underline{b}_\beta)$ has the unique minimum 0 at $\underline{\delta}=\underline{b}_\beta$, it follows that $\inf_{\{||\underline{\delta}-\underline{b}_\beta||_1\ge\epsilon_3\}} D(\underline{\delta}||\underline{b}_\beta) > 0$ for any $\epsilon_3 > 0$. Hence, there exists an $\epsilon_5 > 0$ such that

$$
\sum_{||\underline{\delta}-\underline{b}_\beta||_1\ge\epsilon_3} \mathsf{P}\Big(\underline{N}(\beta n)=n\underline{\delta}\Big) \le \exp(-n\epsilon_5),
$$

for n large enough.

As $\widetilde{\Lambda}^*(p,\beta) > 0$, we have from Lemma 9(iii) that $p > \widetilde{\Lambda}'(0,\beta)$. From the continuous properties in Lemma 12(i) and (ii), we can choose ϵ_3 small enough such that for all $\underline{\delta}$ in $||\underline{\delta}-\underline{b}_\beta||_1 < \epsilon_3$,

$$
p > \Lambda'(0,\underline{\delta}),
$$

and

$$
\widetilde{\Lambda}^*(p,\beta) - \Lambda^*(p,\underline{\delta}) \le \frac{\epsilon_1}{4}. \tag{74}
$$

From Lemma 11(iv), (72) and (74), it follows that there exists an $n_1 < \infty$ such that for $\epsilon_1/2 > 0$ and for all $n \ge n_1$

$$
\begin{aligned}
&\sup_{||\underline{\delta}-\underline{b}_\beta||_1<\epsilon_3} \mathsf{P}\Big(\max_{1\le i\le r^n}[Y_i(n\underline{\delta})]<pn\Big)\\
&\le \sup_{||\underline{\delta}-\underline{b}_\beta||_1<\epsilon_3} \exp\Big(-\exp\big(n(\log r - \Lambda^*(p,\underline{\delta}) - \frac{\epsilon_1}{2})\big)\Big)\\
&= \sup_{||\underline{\delta}-\underline{b}_\beta||_1<\epsilon_3} \exp\Big(-\exp\big(n(\widetilde{\Lambda}^*(p,\beta) - \Lambda^*(p,\underline{\delta}) + \frac{\epsilon_1}{2})\big)\Big)\\
&\le \exp\Big(-\exp(n\frac{\epsilon_1}{4})\Big).
\end{aligned}
$$

Thus, the probability $\mathsf{P}(\frac{T(n)}{n} - \frac{1-p}{\widetilde{S}} < -\epsilon)$ is exponentially bounded and we have

$$\sum_{n=1}^{\infty} \mathsf{P}(\frac{T(n)}{n} - \frac{1-p}{\widetilde{S}} < -\epsilon) < \infty.$$

The proof for the other inequality is similar. Note that we also have

$$\begin{aligned}
&\mathsf{P}(\widetilde{T}(n) > n(\frac{1-p}{\widetilde{S}} + \epsilon)) \\
&\leq \sum_{\|\underline{\delta} - \underline{b}_{(\frac{1-p}{\widetilde{S}}+\epsilon)}\|_1 \geq \epsilon_4} \mathsf{P}\Big(\underline{N}((\frac{1-p}{\widetilde{S}} + \epsilon)n) = n\underline{\delta}\Big) + \sup_{\|\underline{\delta} - \underline{b}_{(\frac{1-p}{\widetilde{S}}+\epsilon)}\|_1 < \epsilon_4} \mathsf{P}\Big(\max_{1<i\leq r^n} [Y_i(n\underline{\delta})] \geq pn\Big)
\end{aligned}$$

for any $\epsilon_4 > 0$. Using Sannov's theorem, there exists an $\epsilon_6 > 0$ such that

$$\sum_{\|\underline{\delta} - \underline{b}_{(\frac{1-p}{\widetilde{S}}+\epsilon)}\|_1 \geq \epsilon_4} \mathsf{P}\Big(\underline{N}((\frac{1-p}{\widetilde{S}} + \epsilon)n) = n\underline{\delta}\Big) < \exp(-n\epsilon_6),$$

for n large enough. Using Lemma 11(ii) and the continuous property in Lemma 12(i), we can choose ϵ_4 small enough such that for n large enough

$$\begin{aligned}
&\sup_{\|\underline{\delta} - \underline{b}_{(\frac{1-p}{\widetilde{S}}+\epsilon)}\|_1 < \epsilon_4} \mathsf{P}\Big(\max_{1\leq i\leq r^n} [Y_i(n\underline{\delta})] \geq pn\Big) \\
&\leq \sup_{\|\underline{\delta} - \underline{b}_{(\frac{1-p}{\widetilde{S}}+\epsilon)}\|_1 < \epsilon_4} \exp\big(n(\log r - \Lambda^*(p, \underline{\delta}) + \frac{\epsilon_2}{2}))\Big) \\
&= \sup_{\|\underline{\delta} - \underline{b}_{(\frac{1-p}{\widetilde{S}}+\epsilon)}\|_1 < \epsilon_4} \exp\big(n(\widetilde{\Lambda}^*(p, \frac{1-p}{\widetilde{S}} + \epsilon) - \Lambda^*(p, \underline{\delta}) - \frac{\epsilon_2}{2}) \\
&\leq \exp(-n\frac{\epsilon_2}{4}).
\end{aligned}$$

Thus, we also have

$$\sum_{n=1}^{\infty} \mathsf{P}(\frac{T(n)}{n} - \frac{1-p}{\widetilde{S}} > \epsilon) < \infty.$$

■

REFERENCES

[1] T. Apostol, *Calculus, Volume II* , John Wiley & Sons, 1969.

[2] M. Avriel, *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, 1976.

[3] J.D. Biggins, “The First and Last Birth Problem for a Multitype Age-Dependent Branching Processes,” *Adv. Appl. Prob.*, Vol. 8, pp. 446-459, 1976.

[4] J. Bucklew, *Large Deviation Techniques in Decision, Simulation and Estimation.* New York, NY: J. Wiley & Sons, Inc., 1990.

[5] J. Byers, M . Luby, M. Mitzenmacher, and A. Rege, “A digital fountain approach to reliable distribution of bulk data,” *Proc. of ACM SIGCOMM’98*, Vancouver, 1998.

[6] H. Chen and D.D. Yao, *Fundamentals of Queueing Networks*, New York: Springer, 2001.

[7] C.S. Chang and R.D. Nelson, ”Bounds on the speedup and efficiency of partial synchronization in parallel processing systems,” *Journal of the ACM*, Vol. 42, pp. 204-231, 1995.

[8] Y.S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag, 1988.

[9] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: Wiley & Sons, 1991.

[10] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Boston: Jones and Barlett Publishers, 1992.

[11] R.S. Ellis, “Large deviations for a general class of random vectors, ” *Ann. Probab.*, Vol. 12, pp. 1-12, 1984.

[12] J. Gärtner, “On large deviations from invariant measure,” *Theory Probab. Appl.*, Vol. 22, pp. 24-39, 1977.

[13] Hiriart-Urruty, J.-B., and Lemaaréchal, C. *Convex Analysis and Minimization Algorithms II.* New York: Springer-Verlag, 1991.

[14] M. Jung, J. Nonnenmacher, and E.W. Biersack, “Reliable multicast via satellite: uni-directional vs. bi-directional communication,” in *Proceedings of KiVS’99*, Darmstadt, Germany.

[15] S.K. Kasera, G. Hjalmtysson, D. Towsley and J. Kurose, “Scalable reliable multicast using multiple multicast channels,” *IEEE/ACM Transactions on Networking*, 2000.

[16] S.K. Kasera, J. Kurose and D. Towsley, “A comparison of server-based and receiver-based local recovery approaches for reliable multicast,” *Proc. IEEE INFOCOM’98*, San Francisco, CA U.S.A., March 1998.

[17] J.F.C. Kingman, “The first birth problem for an age-dependent branching process”, *The Annals of Probability*, Vol 3, No 5, pp. 790-801, 1975.

[18] J. Nonnenmacher, E. Biersack and D. Towsley, “Parity-based loss recovery for reliable multicast transmission,” *Proc. ACM SIGCOMM’97*, Cannes, France, pp. 289-300, Sept. 1997.

[19] J. Nonnenmacher, M. Lacher, M. Jung, E.W. Biersack and G. Carle, “How bad is reliable multicast without local recovery?,” *Proc. IEEE INFOCOM’98*, San Francisco, CA U.S.A., March 1998.

[20] L. Rizzo, “On the feasibility of software fec,” *Tech. rep.,* Univ. di Pisa, Italy, Jan. 1997.

[21] R.T. Rockafeller. *Convex Analysis*. Princeton: Princeton University Press, 1970.

[22] D. Rubenstein, S. Kasera, D. Towsley and J. Kurose, “Improving reliable multicast using active parity encoding services (APES),” *Proc. IEEE INFOCOM’99*, pp. 1248-1255, New York, U.S.A., March 1999.

[23] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis: Queues, Communication and Computing.* 1994.

[24] D. Towsley, “Group (Multicast) Communication in Wide-Area Networks,” ACM Sigmetrics 2000 Tutorial.

[25] L. Vicisano, J. Crowcroft and L. Rizzo, “TCP-like congestion control for layered multicast data transfer,” *Proc. IEEE INFOCOM’98*, San Francisco, CA U.S.A., March 1998.