

Delay Asymptotics with Retransmissions and Incremental Redundancy Codes over Erasure Channels

Yang Yang, Jian Tan, Ness B. Shroff *Fellow, IEEE*, and Hesham El Gamal *Fellow, IEEE*

Abstract—Recent studies have shown that retransmissions can cause heavy-tailed transmission delays even when packet sizes are light-tailed. Moreover, the impact of heavy-tailed delays persists even when packets size are upper bounded. The key question we study in this paper is how the use of coding techniques to transmit information, together with different system configurations, would affect the distribution of delay. To investigate this problem, we model the underlying channel as a Markov modulated binary erasure channel, where transmitted bits are either received successfully or erased. Erasure codes are used to encode information prior to transmission, which ensures that a fixed fraction of the bits in the codeword can lead to successful decoding. We use incremental redundancy codes, where the codeword is divided into codeword trunks and these trunks are transmitted one at a time to provide incremental redundancies to the receiver until the information is recovered. We characterize the distribution of delay under two different scenarios: (I) Decoder uses memory to cache all previously successfully received bits. (II) Decoder does not use memory, where received bits are discarded if the corresponding information cannot be decoded. In both cases, we consider codeword length with infinite and finite support. From a theoretical perspective, our results provide a benchmark to quantify the tradeoff between system complexity and the distribution of delay.

I. INTRODUCTION

Retransmission is the basic component used in most medium access control protocols and it is used to ensure reliable transfer of data over communication channels with failures [2]. Recent studies [3][4][5] have revealed the surprising result that retransmission-based protocols could cause heavy-tailed transmission delays even if the packet length is light tail distributed, resulting in *very long delays* and possibly *zero throughput*. Moreover, [6] shows that even when the packet sizes are upper bounded, the distribution of delay, although eventually light-tailed, may still have a heavy-tailed main body, and that the heavy-tailed main body could dominate even for relatively small values of the maximum packet size. In this paper we investigate the use of coding techniques to transmit information in order to alleviate the impact of heavy tails, and substantially reduce the incurred transmission delay.

In our analysis, we focus on the *Binary Erasure Channel*. Erasures in communication systems can arise in different layers. At the physical layer, if the received signal falls

outside acceptable bounds, it is declared as an erasure. At the data link layer, some packets may be dropped because of checksum errors. At the network layer, packets that traverse through the network may be dropped because of buffer overflow at intermediate nodes and therefore never reach the destination. All these errors can result in erasures in the received bit stream.

In order to investigate how different transmission techniques would affect the delay distribution, we use a general coding framework called *incremental redundancy codes*. In this framework, each codeword is split into several pieces with equal size, which are called codeword trunks. The sender sends only one codeword trunk at a time. If the receiver cannot decode the information, it will request the sender to send another piece of the codeword trunk. Therefore, at every transmission, the receiver gains extra information, which is called incremental redundancy.

In order to combat channel erasures, we use erasure codes as channel coding to encode the information. Erasure codes represent a group of coding schemes which ensure that even when some portions of the codeword are lost, it is still possible for the receiver to recover the corresponding information. Roughly speaking, the encoder transforms a data packet of l symbols into a longer codeword of l_c symbols, where the ratio $\beta = l/l_c$ is called the code-rate. A lower β indicates a larger redundancy in the codeword. An erasure code is said to be near optimal if it requires slightly more than l symbols, say $(1 + \varepsilon)l$ symbols, to recover the information, where ε can be made arbitrary small at the cost of increased encoding and decoding complexity. Many elegant low complexity erasure codes have been designed for erasure channels, e.g., Tornado Code [7], LT code [8], and Raptor code [9]. For the sake of simplicity, throughout the paper, we assume $\varepsilon = 0$. In other words, any β fraction of the codeword can recover the corresponding information.

We specify two different scenarios in this paper. In the first scenario, as shown in Fig. 1, the entire codeword is transmitted as a unit, and received bits are simply discarded if the corresponding information cannot be recovered. Note that in this scenario, the decoder memory is not exploited for caching received bits across different transmissions. This scenario occurs in some random access protocols such as 802.11a/b [13][14], where the base station cannot control who can transmit

The preliminary version of this paper has appeared in [1]

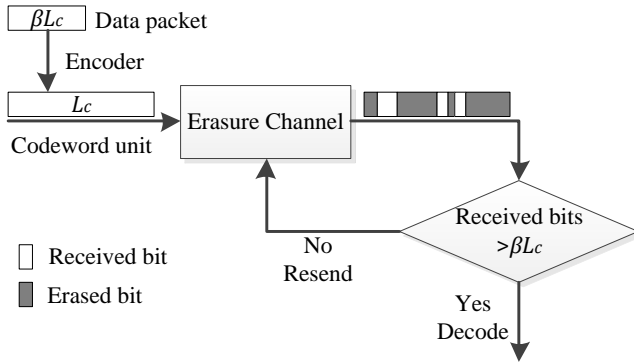


Fig. 1. Decoder that does not use memory scenario

or retransmit a packet and when a transmission can be initiated, and may not have the requisite computation power/storage to keep track of all the undecodable packets from many different users, especially when the base station is responsible for handling a very large number of flows simultaneously. In the second scenario, we assume that the receiver has enough memory space and computational power to cache received bits from different (re)transmissions, which enables the use of incremental redundancy codes, where a codeword of length l_c is split into r codeword trunks with equal size, and these codeword trunks are transmitted one at a time. At the receiver, all successfully received bits from every transmission are buffered at the receiver memory according to their positions in the codeword. If the receiver cannot decode the corresponding information, it will request the sender to send another piece of codeword trunk. At the sender, these codeword trunks are transmitted in a round-robin manner. We call these two scenarios **decoder that does not use memory** and **decoder that uses memory**, respectively.

Given the above two different types of decoder, there are two more factors that can affect the distribution of delay. (I) **Channel Dynamics**: In order to capture the time correlation nature of the wireless channels, we assume that the channel is Markovian modulated. More specifically, we assume a time slotted system where one bit can be transmitted per time slot, and the current channel state distribution depends on channel states in the previous k time slots. When $k = 0$, it corresponds to the *i.i.d.* channel model. (II) **Codeword length distribution**: We assume throughout the paper that the codeword length is light tail distributed, which implies that the system works in a benign environment. We consider two different codeword length distributions, namely, codeword length with *infinite support* and codeword length with *finite support*, respectively. For the former, the codeword length distribution has an exponentially decaying tail with decay rate λ , for the latter, the codeword length has an upper bound b .

Contribution

The main results of this work are the following:

- When decoder memory is not exploited, the tail of the delay distribution depends on the code rate. Specifically, we show that when the coding rate is above a certain threshold, the delay distribution is heavy tailed, otherwise it is light tailed. This shows that substantial gains in delay can be achieved over the standard retransmission case (no forward error correction) by adding a certain amount of redundancy in the codeword. As mentioned earlier, prior work has shown that plain ARQ without FEC results in heavy tailed delays even when the packet size are light tailed.
- When decoder memory is exploited, the tail of the delay distribution is always light-tailed. This implies that the use of receiver memory results in a substantial reduction in the transmission delay.
- The aforementioned results are for the case when the codeword size can have infinite support. We also characterize the transmission delay for each of the above cases when the codeword size has finite support (zero-tailed), and show similar impact of the coding rate and the use of receiver memory in terms of the main body of the delay distribution (rather than the eventual tail).

The remainder of this paper is structured as follows: In Section II, we describe the system model. In Section III we consider the scenario where the decoder memory is exploited. Then, in Section IV we investigate the situation where the decoder does not use memory. Finally, in Section V, we provide numerical studies to verify our main results.

II. SYSTEM MODEL

The channel dynamics are modeled as a slotted system where one bit can be transmitted per slot. Furthermore, we assume that the slotted channel is characterized by a binary stochastic process $\{X_n\}_{n \geq 1}$, where $X_n = 1$ corresponds to the situation when the bit transmitted at time slot n is successfully received, and $X_n = 0$ when the bit is erased (called an erasure).

Since, in practice, the channel dynamics are often temporarily correlated, we investigate the situation in which the current channel state distribution depends on the channel states in the preceding k time slots. More precisely, for $\mathcal{F}_n = \{X_i\}_{i \leq n}$ and fixed k , we define $\mathcal{H}_n = \{X_n, \dots, X_{n-k+1}\}$ for $n \geq k \geq 1$ with $\mathcal{H}_n = \{\emptyset, \Omega\}$ for $k = 0$, and assume that $\mathbb{P}[X_n = 1 | \mathcal{F}_{n-1}] = \mathbb{P}[X_n = 1 | \mathcal{H}_{n-1}]$ for all $n \geq k$. To put it another way, the augmented state $Y_n \triangleq [X_n, \dots, X_{n-k}]$, $n \geq k$ forms a Markov chain. Let Π denote the transition matrix of the Markov chain $\{Y_n\}_{n \geq k+1}$, where

$$\Pi = [\pi(s, u)]_{s, u \in \{0, 1\}^k},$$

with $\pi(s, u)$ being the one-step transition probability from state s to state u . Throughout this paper, we assume that

Π is irreducible and aperiodic, which ensures that this Markov chain is ergodic [10]. Therefore, for any initial value \mathcal{H}_k , the parameter γ is well defined and given by

$$\gamma = \lim_{n \rightarrow \infty} \mathbb{P}[X_n = 1],$$

and, from ergodic theorem (see Theorem 1.10.2 in [10])

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_i}{n} = \gamma \right] = 1,$$

which means the long-term fraction of the bits that can be successfully received is equal to γ . Therefore, we call γ the *channel capacity*.

In the degenerated case when $k = 0$, we have a memoryless binary erasure channel (i.i.d. binary erasure channel). Correspondingly, $\mathcal{H}_n = \{\emptyset, \Omega\}$ and $\Pi = [\gamma]$.

We focus on erasure codes with code rate β . Furthermore, we assume that the erasure codes considered are optimal, in the sense that any β fraction of bits in the codeword can lead to a successful decoding. In other words, for any code rate β that is less than the channel capacity γ , an arbitrarily small error probability can be achieved given a sufficiently large codeword size.

We let L_c denote the number of bits in the codeword with infinite support, and assume that there exist $\lambda > 0$ and $z > 0$ such that

$$\lim_{x \rightarrow \infty} \frac{\log \mathbb{P}[x < L_c < x + z]}{x} = -\lambda. \quad (1)$$

The above equation implies that the distribution of the codeword length L_c has an exponentially decaying tail with decay rate λ . This assumption is quite general since it captures a group of distributions such as gamma distribution, hyperbolic distribution, Laplace distribution, etc.

Note that in reality, transmitted packet sizes are often bounded by the maximum transmission unit (MTU). This means that the distribution of the codeword size often has finite support. We let $L_c(b)$ denote codeword size with finite support, with b being the maximum codeword length, and let $\mathbb{P}[L_c(b) > x] = \mathbb{P}[L_c > x | L_c < b]$ for any $x > 0$.

As mentioned in the introduction, we study two different scenarios in this paper, namely *decoder that uses memory* and *decoder that does not use memory*. In the first scenario, the sender splits a codeword into r codeword trunks with equal size and transmits them one at a time in a round-robin manner, while the receiver uses memory to cache all previously successfully received bits according to their positions in the codeword. In the second scenario, the sender transmits the entire codeword as a unit, and the receiver discards a received codeword if it cannot recover the original information.

Formal definitions of the number of retransmissions and the delays are given as follows:

Definition 1 (Decoder that uses memory). *The total number of transmissions for a codeword with variable length L_c and number of codeword trunks r when the decoder uses*

memory is defined as

$$N_m^{(r)} \triangleq \inf \left\{ n : \sum_{l=1}^r \sum_{i=(L_c/r)(l-1)+1}^{(L_c/r)l} \mathbf{1} \left(\sum_{j=1}^{\lfloor (n-l)/r \rfloor + 1} X_{(j-1)L_c+i} \geq 1 \right) > \beta L_c \right\}.$$

The transmission delay is defined as $T_m^{(r)} = N_m^{(r)} L_c / r$.

Definition 2 (Decoder that does not use memory). *The total number of transmissions for a codeword with variable length L_c when the decoder does not use memory is defined as*

$$N_f \triangleq \inf \left\{ n : \sum_{i=1}^{L_c} X_{(n-1)L_c+i} > \beta L_c \right\}.$$

The transmission delay is defined as $T_f = N_f L_c$.

For a codeword with variable length $L_c(b)$, the corresponding numbers of transmissions and delays are denoted as $N_m^{(r)}(b)$, $T_m^{(r)}(b)$, $N_f(b)$, and $T_f(b)$, respectively.

Notations

In order to present the main results, we introduce some necessary notations here.

Notation 1. Let $\rho(M)$ denote the Perron-Frobenius eigenvalue (see Theorem 3.11 in [11]) of the matrix M , which is the largest eigenvalue of M .

Notation 2. For $k \geq 1$, let $\{s_i\}_{1 \leq i \leq 2^k} = \{0, 1\}^k$ denote the state space of $\{Y_n\}_{n \geq k+1}$, where $s_i = [s_{i1}, s_{i2}, \dots, s_{ik}]$ and $s_{ij} \in \{0, 1\} \forall i, j$. Then, we define a mapping f from $\{s_i\}_{1 \leq i \leq 2^k}$ to $\{0, 1\}$ as

$$f(s_i) = 1 - s_{ik}.$$

Notation 3. Let $\Lambda_n(\beta, \Pi)$ denote the large deviation rate function, which is given by

$$\Lambda_n(\beta, \Pi) = \sup_{\theta} \{ \theta(1 - \beta) - \log \rho_n(\theta, \Pi) \},$$

where¹

$$\rho_n(\theta, \Pi) = \begin{cases} \rho \left(e^{\theta D^{\otimes n}} \Pi^{\otimes n} \right) & k \geq 1 \\ (1 - \gamma)^n + (1 - (1 - \gamma)^n) e^{\theta} & k = 0 \end{cases},$$

$$D = \text{diag} [f(s_1), f(s_2), \dots, f(s_{2^k})] \text{ for } k \geq 1.$$

Notation 4. Let μ_n denote the root of the rate function $\Lambda_n(\beta, \Pi)$. More precisely,

$$\Lambda_n(\mu_n, \Pi) = 0.$$

¹For a matrix A , $A^{\otimes n}$ is the n -fold Kronecker product of A with itself, or we can call it the n^{th} Kronecker power of A .

Notation 5.

$$\begin{aligned} \alpha &= \inf \{n : \mu_n \geq \beta\} \\ \Lambda_1^o &= \inf_{n \in \mathbb{N}} \frac{\lambda + \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha)}{n + 1} \\ \Lambda_2^o &= \inf_{n \in \mathbb{N}} \frac{\lambda + \Lambda_{n+1}(\beta, \Pi) \mathbf{1}(n \geq \alpha - 1)}{n + 1} \\ \tilde{\Lambda}_3^o &= \min_{\lceil r\beta/\gamma \rceil \leq m \leq r-1} \frac{\lambda + \Lambda_1\left(\frac{r\beta}{m}, \Pi\right)}{(m + 1)/r} \\ \Lambda_3^o &= \begin{cases} \lambda & \text{if } \beta > \gamma \\ \min \left\{ \frac{\lambda r}{\lceil r\beta/\gamma \rceil}, \tilde{\Lambda}_3^o \right\} & \text{if } \beta \leq \gamma \end{cases} \end{aligned}$$

III. DECODER THAT USES MEMORY

When the decoder uses memory to cache all previously successfully received bits, we can apply incremental redundancy codes, where the sender splits a codeword into r codeword trunks and transmits one codeword trunk at a time. If the receiver, after receiving a codeword trunk, is not able to decode the corresponding information, it will use memory to cache this codeword trunk and request the sender to send another codeword trunk. In this way, at every transmission, the receiver gains extra information, which we call *incremental redundancy*. The sender will send these codeword trunks in a round-robin manner, meaning that if all of the codeword trunks have been requested, it will start over again with the first codeword trunk.

Similar to the definition of the error exponent in [12], here we define the error exponent after n transmissions as

$$-\lim_{l_c \rightarrow \infty} \frac{\mathbb{P} \left[N_m^{(r)} > n | L_c = l_c \right]}{l_c}.$$

This error exponent represents the decay rate of the decoding error probability after n (re)transmissions, as we increase the codeword length l_c . In the special case when $r = 1$, for every transmission, the whole codeword is transmitted as a unit. Since the decoder retains all the previously received codewords, the retransmission under $r = 1$ can be seen as additional repetition coding. The following lemma shows that after n (re)transmissions of a codeword with rate β , the error exponent is $\Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha)$, where $\Lambda_n(\beta, \Pi)$ and α are defined in Notation 3 and 5.

Lemma 1. For any $n \in \mathbb{N}$, we have,

$$\mathbb{P}[N_m^{(1)} > n | L_c = l_c] = e^{-l_c \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha) + g_n(l_c)},$$

where

$$g_n(l_c) \in \begin{cases} o(l_c) & \text{if } n \geq \alpha \\ o(1) & \text{otherwise} \end{cases}.$$

Proof: see Appendix A. ■

For more general cases when $r > 1$, the error exponents are characterized in the following lemma.

Lemma 2. For any $n, B \in \mathbb{N}$ with $0 < B < r$, we have,

$$\begin{aligned} -\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(r)} > nr | L_c = l_c \right]}{l_c} &= \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha), \\ -\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(r)} > B | L_c = l_c \right]}{l_c} &= \Lambda_1 \left(\frac{\beta r}{B}, \Pi \right) \mathbf{1} \left(\frac{\beta r}{B} < \gamma \right), \\ -\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(r)} > nr + B | L_c = l_c \right]}{l_c} &= \min_{\substack{z_1 > 0, z_2 > 0 \\ z_1 + z_2 < \beta}} \left\{ \Lambda_{n+1} \left(\frac{z_1 r}{B}, \Pi \right) \mathbf{1} \left(\frac{z_1 r}{B} < \mu_{n+1} \right) + \Lambda_n \left(\frac{z_2 r}{r - B}, \Pi \right) \mathbf{1} \left(\frac{z_2 r}{r - B} < \mu_n \right) \right\}. \end{aligned} \quad (2)$$

Proof: see Appendix B. ■

A. Codeword with infinite support

When the distribution of the codeword length L_c has an exponentially decaying tail with decay rate λ , as indicated by Equation (1), we find that the delay will always be light-tailed, and we characterize the decay rate in Theorem 1.

Theorem 1. In the case when the decoder uses memory, when we apply an incremental redundancy code with parameter r to transmit a codeword with variable length L_c , we obtain a lower and upper bound on the decay rate of the delay,

$$\begin{aligned} -\liminf_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(r)} > t \right]}{t} &\leq \min \{ \Lambda_2^o, \Lambda_3^o \}, \\ -\limsup_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(r)} > t \right]}{t} &\geq \min \{ \Lambda_1^o, \Lambda_3^o \}. \end{aligned}$$

In the special case when $r = 1$,

$$-\lim_{n \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(1)} > t \right]}{t} = \min \{ \Lambda_1^o, \lambda \}.$$

The definitions of Λ_1^o, Λ_2^o and Λ_3^o can be found in Notation 5.

Proof: see Appendix E. ■

Remark 1.1. Although it is possible to characterize the exact decay rate of the delay for any r , from Lemma 2 we know that the expression of that decay rate contains a combinatorial optimization problem (as can be seen from Equation (2)) which is hard to evaluate and does not give useful insights. Therefore, in Theorem 1, instead of giving the exact decay rate of the delay, we obtain an upper bound and an lower bound on the delay rate, which are much easier to evaluate.

Remark 1.2. From the definitions of Λ_1^o, Λ_2^o and Λ_3^o in Notation 5 we observe that firstly, the decay rate of the delay when $r = 1$ is no greater than the decay rate

of the delay when $r > 1$ ($\min\{\Lambda_1^o, \lambda\} \leq \min\{\Lambda_1^o, \Lambda_3^o\}$), which means that incremental redundancy codes ($r > 1$) outperform fixed rate erasure codes ($r = 1$); secondly, the decay rate of the delay increases with the increase of r , which means we can reduce delay by increasing the number of codeword trunks r ; thirdly, when $\Lambda_1^o > \Lambda_3^o$ and $\Lambda_2^o > \Lambda_3^o$, the two bounds on the decay rate of the delay coincide. These observations are verified through Example 1 in Section V.

B. Codeword with finite support

In practice, codeword length is bounded by the maximum transmission unit (MTU). Therefore, we investigate the case when the codeword has variable length $L_c(b)$, with b being the maximum codeword length, and characterize the corresponding delay distribution in Theorem 2.

Theorem 2. *In the case when the decoder uses memory, when we apply an incremental redundancy code with parameter r to transmit a codeword with variable length $L_c(b)$, for any $\eta, \delta > 0$, we can find $b(\eta, \delta) > 0$ such that for any $b > b(\eta, \delta)$, we have*

1) $\forall t \in [n_2^o(b - \delta), n_2^o b]$,

$$(1 - \eta)\Lambda_1^b \leq -\frac{\log \mathbb{P} [T_m^{(r)}(b) > t]}{t} \leq (1 + \eta)\Lambda_2^b.$$

2) in the special case when $r = 1$, $\forall t \in [n_1^o(b - \delta), n_1^o b]$,

$$1 - \eta \leq -\frac{\log \mathbb{P} [T_m^{(1)}(b) > t]}{t} \frac{1}{\Lambda^b} \leq 1 + \eta,$$

where

$$n_1^o = \arg \inf_{n \in \mathbb{N}} (\lambda + \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha)) / (n + 1),$$

$$n_2^o = \arg \inf_{n \in \mathbb{N}} (\lambda + \Lambda_{n+1}(\beta, \Pi) \mathbf{1}(n \geq \alpha - 1)) / (n + 1),$$

$$\Lambda_1^b = \Lambda_1^o + \min\{0, \Lambda_3^o - \Lambda_1^o\} \mathbf{1}(n_2^o = 1),$$

$$\Lambda_2^b = \Lambda_2^o + \min\{0, \Lambda_3^o - \Lambda_2^o\} \mathbf{1}(n_2^o = 1),$$

$$\Lambda^b = \Lambda_1^o + \min\{0, \lambda - \Lambda_1^o\} \mathbf{1}(n_1^o = 1).$$

Proof: see Appendix F. ■

Remark 2.1. This theorem shows that even if the codeword length has an upper bound b , the distribution of the delay still has a light-tailed main body whose decay rate is similar to the decay rate under the infinite support scenario. The waist of this main body is at least $n_2^o b$ when $r > 1$ and $n_1^o b$ when $r = 1$. Since both n_2^o and n_1^o are independent of b , we know that the waist of this light-tailed main body scales linearly with respect to the maximum codeword length b . This theorem is verified through Example 2 in Section V.

IV. DECODER THAT DOES NOT USE MEMORY

There are a number of systems that use hybrid ARQ and perform soft combing to improve the probability of successful decoding. However, there are also several systems that use plain ARQ. For example, in the IEEE

802.11a/b [13][14] protocol family, whenever an undecodable packet arrives, the receiver simply discards it and asks for a retransmission by sending a feedback message. The reason is that 802.11 uses a random access MAC layer protocol and thus the base station has no control of exactly who can use the channel to transmit or retransmit a packet and when a transmission can be initiated. In other words, if the decoder at the 802.11 base station were to use hybrid ARQ to combine different transmissions, it would have to keep track of which packets came from which users in the physical layer all the time in order to combine the right packets. Hence, a lot of decoder memory would have to be allocated to store packets from many different users, especially when the number of users is large.

Therefore, in this section, we study the case when the decoder does not use memory, as illustrated in Fig. 1. Since the decoder simply discards a received packet after a failed decoding attempt, we do not need to consider the case when a codeword is divided into several codeword trunks before transmission, because a codeword trunk will be discarded if itself cannot recover the original information. For example, if we divide a codeword with rate β into two codeword trunks and transmit them one after the other, then the transmission is successful only when either one of codeword trunks itself can be decoded. Therefore, it will be equivalent to transmitting an entire codeword with rate 2β .

The following lemma shows that when the code rate β is less than the channel capacity γ , after n transmissions, the error exponent is $n\Lambda_1(\beta, \Pi)$.

Lemma 3. 1) if $\beta > \gamma$, then

$$\mathbb{P} [N_f > n | L_c = l_c] = \left(1 - e^{-l_c \Lambda_1(\beta, \Pi)(1+g(l_c))}\right)^n,$$

where $g(l_c) \in o(1)$ as $l_c \rightarrow \infty$.

2) if $\beta < \gamma$, then

$$\mathbb{P} [N_f > n | L_c = l_c] = e^{-nl_c \Lambda_1(\beta, \Pi)(1+s(l_c))},$$

where $s(l_c) \in o(1)$ as $l_c \rightarrow \infty$.

Proof: see Appendix G. ■

A. Codeword with infinite support

Interestingly, we observe an intriguing threshold phenomenon. We show that when the codeword length distribution is light-tailed and has an infinite support, the transmission delay is light-tailed (exponential) only if $\gamma > \beta$, and heavy-tailed (power law) if $\gamma < \beta$.

Theorem 3 (Threshold phenomenon). *In the case when the decoder does not use memory and the codeword has variable length L_c , we get*

1) if $\beta > \gamma$, then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\log \mathbb{P} [N_f > n]}{\log n} \\ &= \lim_{t \rightarrow \infty} \frac{\log \mathbb{P} [T_f > t]}{\log t} = -\frac{\lambda}{\Lambda_1(\beta, \Pi)}. \end{aligned}$$

2) if $\beta < \gamma$, then

$$\lim_{t \rightarrow \infty} \frac{\log \mathbb{P}[T_f > t]}{t} = -\min\{\lambda, \Lambda_1(\beta, \Pi)\}.$$

The definition of $\Lambda_1(\beta, \Pi)$ can be found in Notation 3.

Proof: see Appendix G. ■

Remark 3.1. The tail distribution of the transmission delay can follow either power law or exponential distribution, depending on the relationship between code-rate β and channel capacity γ . When $\beta > \gamma$ and $\lambda/\Lambda_1(\beta, \Pi) < 1$, the system even has a zero throughput.

B. Codeword with finite support

Under the heavy-tailed delay case when $\beta > \gamma$, we can further show that if the codeword length is upper bounded, the delay distribution still has a heavy-tailed main body, although it eventually becomes light-tailed.

Theorem 4. In the case when the decoder does not use memory and the codeword has variable length $L_c(b)$, if $\beta > \gamma$, for any $\eta > 0$, we can find $n(\eta) > 0$ and $b(\eta) > 0$ such that for any $b > b(\eta)$ we have,

1) $\forall n \in [n(\eta), n_b]$ and $\forall t \in [n(\eta)b, n_b b]$,

$$1 - \eta \leq -\frac{\log \mathbb{P}[N_f(b) > n] \Lambda_1(\beta, \Pi)}{\log n \lambda} \leq 1 + \eta,$$

$$1 - \eta \leq -\frac{\log \mathbb{P}[T_f(b) > t] \Lambda_1(\beta, \Pi)}{\log t \lambda} \leq 1 + \eta.$$

where

$$n_b = (\mathbb{P}[N_f = 1 | L_c = b])^{-1}, \quad (3)$$

and the definition of $\Lambda_1(\beta, \Pi)$ can be found in Notation 3.

2)

$$\lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f(b) > n]}{n} = \log(\mathbb{P}[N_f > 1 | L_c = b]),$$

$$\lim_{t \rightarrow \infty} \frac{\log \mathbb{P}[T_f(b) > t]}{t} = \frac{1}{b} \log(\mathbb{P}[T_f > 1 | L_c = b]).$$

Proof: see Appendix H. ■

Remark 4.1. From Equation (3) and by Lemma 3, we can obtain

$$\lim_{b \rightarrow \infty} \frac{\log n_b}{b} = \Lambda_1(\beta, \Pi), \quad (4)$$

which implies that n_b increases exponentially fast with the increase of the maximum codeword length b . Since the waist of the heavy-tailed main body of the delay distribution is $n_b b$, we know that the waist also increases exponentially fast as we increase the maximum codeword length b .

From Theorem 4 we know that even if the codeword length is bounded, the heavy-tailed main body could still play a dominant role. From Theorem 3 we know that when $\lambda < \Lambda_1(\beta, \Pi)$ and $\beta > \gamma$, the throughput will vanish to zero as b approaches infinity. Now we explore how fast the throughput vanishes to zero as b increases.

Let $\{L_i\}_{i \geq 1}$ be the i.i.d. sequence of codeword lengths with distribution $L_c(b)$. Denote T_i as the transmission delay of L_i . The throughput of this system is defined as $\Delta(b) = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \beta L_i}{\sum_{i=1}^n T_i}$.

Theorem 5 (Throughput). In the case when the decoder does not use memory and the codeword has variable length $L_c(b)$, if $\beta > \gamma$ and $\lambda < \Lambda_1(\beta, \Pi)$, we have

$$-\limsup_{b \rightarrow \infty} \frac{\log \Delta(b)}{b} \geq \Lambda_1(\beta, \Pi) - \lambda.$$

The definition of $\Lambda_1(\beta, \Pi)$ can be found in Notation 3.

Proof: see Appendix I. ■

Remark 5.1. Theorem 5 indicates that when code-rate β is greater than channel capacity γ and $\lambda < \Lambda_1(\beta, \Pi)$, as the maximum codeword length b increases, the throughput vanishes to 0 at least exponentially fast with rate $\Lambda_1(\beta, \Pi) - \lambda$.

V. SIMULATIONS

In this section, we conduct simulations to verify our main results. As is evident from the following figures, the simulations match theoretical results well.

Example 1. In this example, we study the case when the decoder uses memory and the codeword length has infinite support. We assume that the channel is i.i.d. ($k = 0$). As shown in Theorem 1, under the above assumptions, the delay distribution is always light-tailed. In order to verify this result, we assume that L_c is geometrically distributed with mean 100 ($\lambda = 0.01$), and choose code-rate $\beta = 0.25$ and channel capacity $\gamma = 0.50$. By Theorem 1 we know that when $r = 1$, the decay rate of the delay is $\min\{\Lambda_1^o, \lambda\} = 0.0025$; when $r = 3$, the decay rate of the delay is $\min\{\Lambda_1^o, \Lambda_3^o\} = \min\{\Lambda_2^o, \Lambda_3^o\} = 0.0037$; when $r = 5$, the decay rate of the delay is $\min\{\Lambda_1^o, \Lambda_3^o\} = \min\{\Lambda_2^o, \Lambda_3^o\} = 0.0042$. From Fig. 2 we can see that the decay rate of the delay increases when r increases from 1 to 5, and the theoretical result is quite accurate.

This example shows that when the code rate is less than the channel capacity, the delay distribution can benefit from an increased rate of feedback. In other words, by increasing the number of codeword trunks r , the tail of delay distribution can become lighter.

Example 2. In this simulation, we study the case when the decoder uses memory and the codeword length has a finite support. We assume that the channel is i.i.d. ($k = 0$), code-rate $\beta = 0.75$, $\lambda = 0.01$, $r = 1$, and channel capacity $\gamma = 0.1$. From these system parameters we can calculate $n_1^o = 14$ and $\Lambda^b = \min\{\Lambda_1^o, \lambda\} = 7.1429 \times 10^{-4}$. We choose four sets of maximum codeword length b as 200, 400, 600, 800. Theorem 2 indicates that the delay distribution has a light-tailed main body with decay rate $\Lambda^b = 7.1429 \times 10^{-4}$ and waist $n_b b = 14 \times b$. In Fig. 3 we plot the delay distributions when $b = 200, 400, 600, 800$ together with the infinite support case when $b = \infty$,

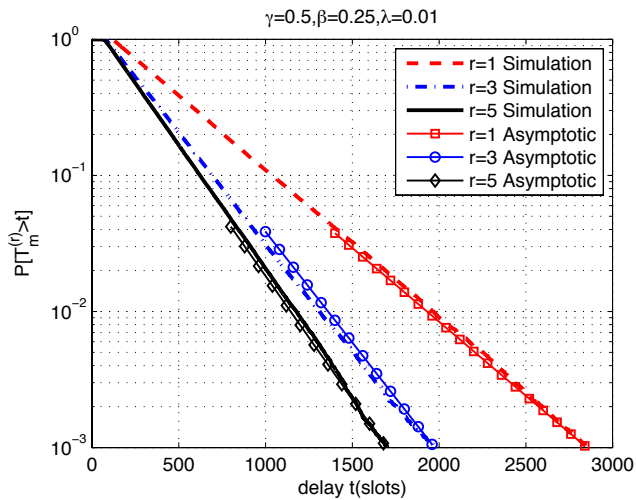


Fig. 2. Illustration for Example 1

and we use a short solid line to indicate the waist of the light-tailed main body. As we can see from Fig. 3, the theoretical waists of the main bodies, which are $n_b b = 14 \times b = 2800, 5600, 8400, 11200$, are close to the simulation results.

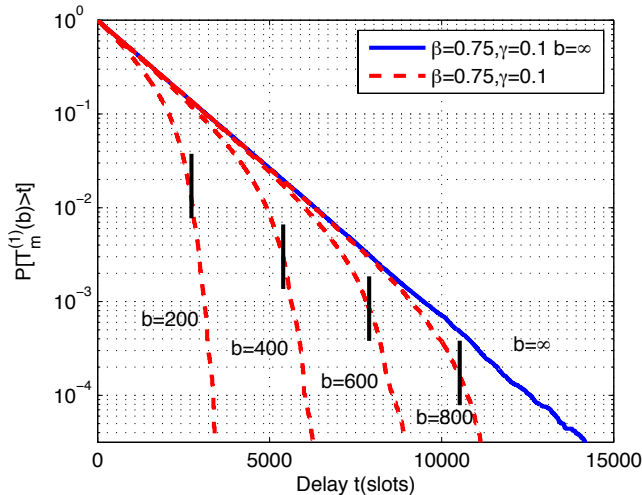


Fig. 3. Illustration for Example 2

Example 3. Now we use simulations to verify Theorem 4. Theorem 4 says that when the decoder does not use memory, if code rate β is greater than channel capacity γ and the codeword length has a finite support, the distribution of delay as well as the distribution of number of retransmissions have a heavy-tailed main body and an exponential tail. The waist of the main body increases exponentially fast with the increase of maximum codeword length b . In this experiment, we set code-rate $\beta = 0.25$, channel capacity $\gamma = 0.20$, $k = 0$, and $\lambda = 0.01$. From these parameters we can get $\Lambda_1(\beta, \Pi) = 0.0074$. We choose four sets of maximum codeword length b as 200, 400, 600, 800.

As Equation (4) indicates, the waist of the heavy-tailed main bodies of the number of retransmissions is $n_b \approx e^{b\Lambda_1(\beta, \Pi)} = 4.3772, 19.1595, 83.8641, 367.0865$. In Fig. 4, we plot the distribution of the number of retransmissions when $b = 200, 400, 600, 800$ together with the infinite support case when $b = \infty$, and we use a short solid line to indicate the waist of the heavy-tailed main body. As can be seen from Fig. 4, the simulation matches with our theoretical result.

The above two examples show that when the decoder does not use memory, the choice of the code rate has a dramatic effect on the delay distribution, even if the codeword length is upper-bounded. More specifically, as we increase the code rate, the main body of the tail of the distribution of the delay could change from an exponential decay function which spans just proportionally to the maximum codeword size, to a much slower power-law decay function which spans exponentially to the maximum codeword size. This emphasizes the importance of rate adaptation when the decoder does not use memory.

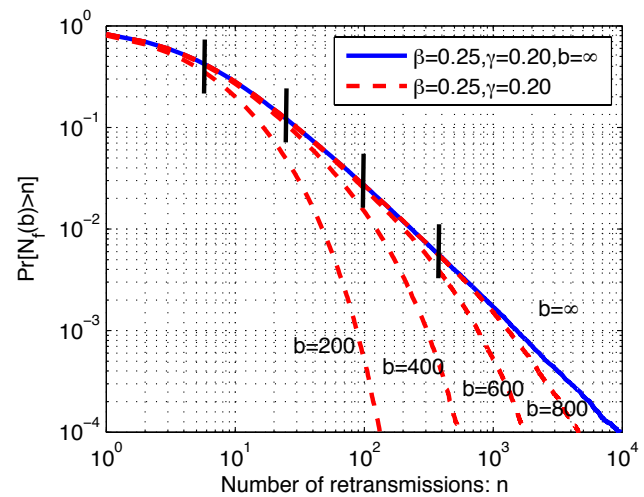


Fig. 4. Illustration for Example 3

VI. CONCLUSION

In this paper, we characterize the delay distribution in a point-to-point Markovian modulated binary erasure channel with variable codeword length. Erasure codes are used to encode information such that a fixed fraction of bits in the codeword can recover the information. We use a general coding framework called incremental redundancy codes. In this framework, the codeword is divided into several codeword trunks and these codeword trunks are transmitted one at a time to the receiver. Therefore, after each transmission, the receiver gains extra information, which is called incremental redundancy. At the receiver end, we investigate two different scenarios, namely *decoder that uses memory* and *decoder that does not use memory*. In the decoder that uses memory case, the decoder caches all previously successfully transmitted

bits. In the decoder that does not use memory case, received bits are discarded if the corresponding information cannot be decoded. In both cases, we first assume that the distribution of the codeword length is light-tailed and has an infinite support. Then, we consider a more realistic case when the codeword length is upper bounded.

Our results show the following. The transmission delay can be dramatically reduced by allowing the decoder to use memory. This is true because the delay is always light-tailed when the decoder uses memory, while the delay can be heavy-tailed when the decoder does not use memory. Secondly, analogously to the non-coding case, the tail effect of the delay distribution persists even if the codeword length has a finite support. When the codeword length is upper bounded, the light-tailed delay distribution will turn into a delay distribution with light-tailed main body whose decay rate is similar to that in the infinite support scenario. Further, we show that the waist of this main body scales linearly with respect to the increase of the maximum codeword length; the heavy-tailed delay distribution will turn into a delay distribution with heavy-tailed main body, whose waist scales exponentially with the increase of the maximum codeword length. Our results provide a benchmark for quantifying the tradeoff between system complexity (which is determined by code-rate β , number of codeword trunks r , maximum codeword length b and whether to use memory at the receiver or not) and the distribution of the delay.

APPENDIX A PROOF OF LEMMA 1

First we consider the case when $k \geq 1$. By Definition 1, we have

$$\begin{aligned}
 & \mathbb{P}[N_m^{(1)} > n | L_c = l_c] \\
 &= \mathbb{P}\left[\sum_{i=1}^{L_c} \mathbf{1}\left(\sum_{j=1}^n X_{(j-1)L_c+i} \geq 1\right) \leq \beta L_c \middle| L_c = l_c\right] \\
 &= \mathbb{P}\left[\sum_{i=1}^{L_c} \mathbf{1}\left(\sum_{j=1}^n X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c \middle| L_c = l_c\right] \\
 &= \mathbb{E}\left[\mathbb{P}\left[\sum_{i=1}^{L_c} \prod_{j=1}^n \mathbf{1}\left(X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c \middle| L_c, \bigcup_{j=1}^n \mathcal{E}_j\right] \middle| L_c = l_c\right], \quad (5)
 \end{aligned}$$

where $\mathcal{E}_j \triangleq \{X_{(j-1)L_c+1}, \dots, X_{(j-1)L_c+k}\}$, $1 \leq j \leq n$. Let $Y_{in} = [Y_i, Y_{L_c+i}, \dots, Y_{(n-1)L_c+i}]$ and

$$f_n(Y_{in}) = \prod_{j=1}^n f(Y_{(j-1)L_c+i}).$$

If $L_c > k$, then given $\bigcup_{j=1}^n \mathcal{E}_j$, $\{Y_{in}\}_{k < i \leq L_c}$ forms a Markov chain with state space $\{0, 1\}^k$ and probability transition matrix $\Pi^{\otimes n}$. We further observe that if $L_c > k$,

we have the following relationship

$$\begin{aligned}
 & \left\{ \sum_{i=1+k}^{L_c} \prod_{j=1}^n \mathbf{1}\left(X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c \right\} \\
 & \subseteq \left\{ \sum_{i=1}^{L_c} \prod_{j=1}^n \mathbf{1}\left(X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c \right\} \\
 & \subseteq \left\{ \sum_{i=1+k}^{L_c} \prod_{j=1}^n \mathbf{1}\left(X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c - k \right\}.
 \end{aligned}$$

Using the above observation, we can construct upper and lower bounds as follows.

$$\begin{aligned}
 & \mathbb{P}\left[\sum_{i=1}^{L_c} \prod_{j=1}^n \mathbf{1}\left(X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c \middle| L_c, \bigcup_{j=1}^n \mathcal{E}_j\right] \\
 & \geq \mathbb{P}\left[\sum_{i=1+k}^{L_c} f_n(Y_{in}) > (1-\beta)L_c \middle| L_c, \bigcup_{j=1}^n \mathcal{E}_j\right], \\
 & \mathbb{P}\left[\sum_{i=1}^{L_c} \prod_{j=1}^n \mathbf{1}\left(X_{(j-1)L_c+i} = 0\right) > (1-\beta)L_c \middle| L_c, \bigcup_{j=1}^n \mathcal{E}_j\right] \\
 & \leq \mathbb{P}\left[\sum_{i=1+k}^{L_c} f_n(Y_{in}) > (1-\beta)L_c - k \middle| L_c, \bigcup_{j=1}^n \mathcal{E}_j\right]. \quad (6)
 \end{aligned}$$

By a direct application of Theorem 3.1.2 in [11], we know that for a given $\varepsilon > 0$ and any values of $\bigcup_{j=1}^n \mathcal{E}_j$, we can find l_ε such that

$$\begin{aligned}
 & \mathbb{P}\left[\sum_{i=1+k}^{L_c} f_n(Y_{in}) > (1-\beta)L_c \middle| L_c = l_c\right] \\
 & \geq e^{-\inf_{1-\omega > 1-\beta} \Lambda_n(\omega, \Pi)(1+\varepsilon)l_c}, \quad (7) \\
 & \mathbb{P}\left[\sum_{i=1+k}^{L_c} f_n(Y_{in}) > (1-\beta)L_c - k \middle| L_c = l_c\right] \\
 & \leq e^{-\inf_{1-\omega > 1-\beta} \Lambda_n(\omega, \Pi)(1-\varepsilon)l_c}, \quad (8)
 \end{aligned}$$

whenever $l_c > l_\varepsilon$. Since $\Lambda_n(\omega, \Pi)$ is a large deviation rate function, from [11] we know that

$$\begin{aligned}
 \inf_{1-\omega > 1-\beta} \Lambda_n(\omega, \Pi) &= \begin{cases} \Lambda_n(\beta, \Pi) & \text{if } \mu_n < \beta \\ 0 & \text{otherwise} \end{cases} \\
 &= \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha). \quad (9)
 \end{aligned}$$

The upper and lower bounds in Equation (8) and (7), together with Equation (5), (6) and (9), imply that

$$\begin{aligned}
 & -\liminf_{l_c \rightarrow \infty} \frac{\log \mathbb{P}\left[N_m^{(1)} > n | L_c = l_c\right]}{l_c} \\
 & \leq (1+\varepsilon)\Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha), \\
 & -\limsup_{l_c \rightarrow \infty} \frac{\log \mathbb{P}\left[N_m^{(1)} > n | L_c = l_c\right]}{l_c} \\
 & \geq (1-\varepsilon)\Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha),
 \end{aligned}$$

which, with $\varepsilon \rightarrow 0$, completes the proof when $k \geq 1$.

Next, let us consider the case when $k = 0$. In this memoryless channel case, for any single bit in the codeword, after the n^{th} transmission, the probability that the bit is

successfully received is $1 - (1 - \gamma)^n$. Therefore, for a certain packet, the probability of the number of transmissions being greater than n is equal to the probability of the number of transmissions being greater than 1 under a memoryless erasure channel with erasure probability $(1 - \gamma)^n$.

Then, by a direct application of Gärtner-Ellis theorem (Theorem 2.3.6 in [11]), we have, for any $n \geq 1$,

$$\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P}[N_m^{(1)} > n | L_c = l_c]}{l_c} = -\Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha),$$

where $\Lambda_n(\beta, \Pi)$ and α can be written in closed-forms, which are

$$\begin{aligned} \Lambda_n(\beta, \Pi) &= \sup_{\theta} \{ \theta(1 - \beta) - \log(\mathbb{E}[X_i e^{\theta X_i}]) \} \\ &= \sup_{\theta} \{ \theta(1 - \beta) - \log(1 - (1 - \gamma)^n + (1 - \gamma)^n e^{\theta} \} \\ &= \beta \log \frac{\beta}{1 - (1 - \gamma)^n} + (1 - \beta) \log \frac{1 - \beta}{(1 - \gamma)^n}, \end{aligned}$$

$$\text{and } \alpha = \left\lceil \frac{\log(1 - \beta)}{\log(1 - \gamma)} \right\rceil.$$

APPENDIX B PROOF OF LEMMA 2

The first equation in Lemma 2 follows directly from the fact that

$$\mathbb{P}[N_m^{(r)} > nr | L_c = l_c] = \mathbb{P}[N_m^{(1)} > n | L_c = l_c].$$

Next, let us focus on the last two equations in Lemma 2. Denote

$$\begin{aligned} Z_1(n, l_c, B) &= \sum_{l=1}^B \sum_{i=(l_c/r)(l-1)+1}^{(l_c/r)l} \mathbf{1} \left(\sum_{j=1}^{n+1} X_{(j-1)l_c+i} \geq 1 \right), \\ Z_2(n, l_c, B) &= \sum_{l=B+1}^r \sum_{i=(l_c/r)(l-1)+1}^{(l_c/r)l} \mathbf{1} \left(\sum_{j=1}^n X_{(j-1)l_c+i} \geq 1 \right). \end{aligned}$$

By Definition 1, it is easy to see that

$$\mathbb{P}[N_m^{(r)} > B | L_c = l_c] = \mathbb{P} \left[\frac{Z_1(0, l_c, B)}{l_c} < \beta \right],$$

and

$$\begin{aligned} &\mathbb{P}[N_m^{(r)} > nr + B | L_c = l_c] \\ &= \sum_{\substack{z_1 \geq 0, z_2 \geq 0 \\ z_1 + z_2 < \beta \\ z_1 l_c, z_2 l_c \in \mathbb{N}}} \mathbb{P} \left[\frac{Z_1(n, l_c, B)}{l_c} = z_1, \frac{Z_2(n, l_c, B)}{l_c} = z_2 \right]. \end{aligned}$$

Then based on the above two equations, the last two equations in Lemma 2 can be proven using the same technique as that used in the proof of Lemma 1.

APPENDIX C PROOF OF LEMMA 3

From Definition 2 we know

$$\begin{aligned} &\mathbb{P}[N_f > n | L_c] \\ &= \mathbb{P} \left[\bigcap_{1 \leq j \leq n} \left\{ \sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c \right\} \middle| L_c \right] \\ &= \mathbb{E} \left[\mathbb{P} \left[\bigcap_{1 \leq j \leq n} \left\{ \sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c \right\} \middle| \bigcup_{j=1}^n \mathcal{E}_j, L_c \right] \middle| L_c \right] \\ &= \mathbb{E} \left[\prod_{j=1}^n \mathbb{P} \left[\sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c \middle| \bigcup_{i=1}^n \mathcal{E}_i, L_c \right] \middle| L_c \right], \quad (10) \end{aligned}$$

where $\mathcal{E}_j \triangleq \{X_{(j-1)L_c+1}, \dots, X_{(j-1)L_c+k}\}$, $1 \leq j \leq n$. The last equation is due to the Markov property of the channel states. Observe that if $L_c > k$, for any $1 \leq j \leq n$,

$$\begin{aligned} &\left\{ \sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c \right\} \\ &\subseteq \left\{ \sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c \right\} \\ &\subseteq \left\{ \sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k \right\}, \end{aligned}$$

which further yields

$$\begin{aligned} &\mathbb{P} \left[\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k \middle| \bigcup_{i=1}^n \mathcal{E}_i, L_c \right] \\ &\geq \mathbb{P} \left[\sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c \middle| \bigcup_{i=1}^n \mathcal{E}_i, L_c \right] \\ &\geq \mathbb{P} \left[\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c \middle| \bigcup_{i=1}^n \mathcal{E}_i, L_c \right]. \end{aligned}$$

Similarly as the proof of Lemma 1, by Theorem 3.1.2 in [11], we obtain, for any $1 \leq j \leq n$

$$\begin{aligned} &\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P} \left[\sum_{i=(j-1)L_c+1}^{jL_c} X_i > \beta L_c \middle| \bigcup_{i=1}^n \mathcal{E}_i, L_c = l_c \right]}{l_c} \\ &= -\Lambda_1(\beta, \Pi) \mathbf{1}(\beta > \gamma), \\ &\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P} \left[\sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c \middle| \bigcup_{i=1}^n \mathcal{E}_i, L_c = l_c \right]}{l_c} \\ &= -\Lambda_1(\beta, \Pi) \mathbf{1}(\beta < \gamma), \end{aligned}$$

which, by combining Equation (10), completes the proof.

APPENDIX D LEMMA 4

In order to prove the theorems, we need the following lemma.

Lemma 4. Assume b is a function of t , which satisfies $b \triangleq b(t) > \frac{t}{y}$. Then, for any $x, y \in \mathbb{N}$ we have

$$\lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(1)} > x, \frac{t}{y+1} < L_c(b) < \frac{t}{y} \right]}{t} = -\frac{\lambda + \Lambda_x(\beta, \Pi) \mathbf{1}(x \geq \alpha)}{y+1}.$$

Proof: Observe that

$$\begin{aligned} & \mathbb{P} \left[N_m^{(1)} > x, \frac{t}{y+1} < L_c(b) < \frac{t}{y} \right] \\ &= \sum_{l_c = \lceil t/(y+1) \rceil}^{\lfloor t/y \rfloor} \mathbb{P} \left[N_m^{(1)} > x \mid L_c(b) = l_c \right] \mathbb{P} [L_c(b) = l_c]. \end{aligned} \quad (11)$$

Let us first consider the case when $x \geq \alpha$. From Lemma 1 we know that when $x \geq \alpha$, for l_c large enough, $\mathbb{P} \left[N_m^{(1)} > x \mid L_c(b) = l_c \right]$ is a decreasing function of l_c . Therefore, for t large enough,

$$\begin{aligned} & \mathbb{P} \left[N_m^{(1)} > x, \frac{t}{y+1} < L_c(b) < \frac{t}{y} \right] \\ & \leq \mathbb{P} \left[N_m^{(1)} > x \mid L_c(b) = \left\lceil \frac{t}{y+1} \right\rceil \right] \mathbb{P} \left[\frac{t}{y+1} < L_c(b) < \frac{t}{y} \right]. \end{aligned} \quad (12)$$

Since $b = b(t) > \frac{t}{y}$, by the definition of $L_c(b)$, we can easily obtain

$$\lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[\frac{t}{y+1} < L_c(b) < \frac{t}{y} \right]}{t} = -\frac{\lambda}{y+1}. \quad (13)$$

Combining Equation (13) and (12), and by Lemma 1, we get, for $x \geq \alpha$

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(1)} > x, \frac{t}{y+1} < L_c(b) < \frac{t}{y} \right]}{t} \\ & \leq -(\lambda + \Lambda_x(\beta, \Pi))/(y+1). \end{aligned}$$

The lower bound can be constructed in a similar manner by noting that

$$\begin{aligned} & \mathbb{P} \left[N_m^{(1)} > x, \frac{t}{y+1} < L_c(b) < \frac{t}{y} \right] \\ & \geq \mathbb{P} \left[N_m^{(1)} > x \mid L_c(b) = \left\lceil \frac{t}{y+1} \right\rceil \right] \mathbb{P} \left[L_c(b) = \left\lceil \frac{t}{y+1} \right\rceil \right]. \end{aligned}$$

Next, let us consider the case when $x < \alpha$. From Lemma 1 we know that for any $\varepsilon > 0$, we can find a large enough l_ε such that for all $l_c > l_\varepsilon$,

$$1 - \varepsilon \leq \mathbb{P} \left[N_m^{(1)} > x \mid L_c(b) = l_c \right] \leq 1,$$

which, by combining Equation (11) and (13), yields, for $x < \alpha$,

$$\lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(1)} > x, \frac{t}{y+1} < L_c(b) < \frac{t}{y} \right]}{t} = -\lambda/(y+1). \quad \blacksquare$$

APPENDIX E PROOF OF THEOREM 1

Observe that

$$\begin{aligned} & \mathbb{P} \left[T_m^{(r)} > t \right] \\ &= \sum_{h=r}^{\infty} \mathbb{P} \left[T_m^{(r)} > t, \frac{tr}{h+1} < L_c \leq \frac{tr}{h} \right] + \mathbb{P} \left[T_m^{(r)} > t, L_c > t \right] \\ &= \sum_{n=1}^{\infty} \sum_{h=nr}^{(n+1)r-1} \mathbb{P} \left[N_m^{(r)} > h, \frac{tr}{h+1} < L_c \leq \frac{tr}{h} \right] \\ & \quad + \mathbb{P} \left[T_m^{(r)} > t, L_c > t \right]. \end{aligned} \quad (14)$$

Let us first focus on the first part of Equation (14). Denote $\mathbb{P}_{ntr} = \sum_{h=nr}^{(n+1)r-1} \mathbb{P} \left[N_m^{(r)} > h, \frac{tr}{h+1} < L_c \leq \frac{tr}{h} \right]$, then it is easy to check that

$$\begin{aligned} \mathbb{P}_{ntr} & \leq \mathbb{P} \left[N_m^{(1)} > n, \frac{t}{n+1} < L_c \leq \frac{t}{n} \right], \\ \mathbb{P}_{ntr} & \geq \mathbb{P} \left[N_m^{(1)} > n+1, \frac{t}{n+1} < L_c \leq \frac{t}{n} \right], \end{aligned}$$

which, by Lemma 4 in Appendix D, yields

$$\begin{aligned} \limsup_{t \rightarrow \infty} \frac{\log \mathbb{P}_{ntr}}{t} & \leq -\frac{\lambda + \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha)}{n+1}, \\ \liminf_{t \rightarrow \infty} \frac{\log \mathbb{P}_{ntr}}{t} & \geq -\frac{\lambda + \Lambda_{n+1}(\beta, \Pi) \mathbf{1}(n \geq \alpha - 1)}{n+1}. \end{aligned} \quad (15)$$

Now let us focus on the second part of Equation (14). Observe that

$$\begin{aligned} & \mathbb{P} \left[T_m^{(r)} > t, L_c > t \right] \\ &= \sum_{n=1}^{r-1} \mathbb{P} \left[T_m^{(r)} > t, \frac{rt}{n+1} < L_c \leq \frac{rt}{n} \right] + \mathbb{P} \left[T_m^{(r)} > t, rt < L_c \right] \\ &= \sum_{n=1}^{r-1} \mathbb{P} \left[N_m^{(r)} > n, \frac{rt}{n+1} < L_c \leq \frac{rt}{n} \right] \\ & \quad + \mathbb{P} \left[N_m^{(r)} > 0, rt < L_c \right]. \end{aligned}$$

From Lemma 2 we know that, for $1 \leq n \leq r-1$,

$$\begin{aligned} & -\lim_{l_c \rightarrow \infty} \frac{\log \mathbb{P} \left[N_m^{(r)} > n \mid L_c = l_c \right]}{l_c} \\ & = \Lambda_1 \left(\frac{\beta r}{n}, \Pi \right) \mathbf{1} \left(\frac{\beta r}{n} < \gamma \right). \end{aligned}$$

By combining the above two equations, we get

$$\begin{aligned} & -\lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(r)} > t, L_c > t \right]}{t} \\ & = -\lim_{t \rightarrow \infty} \frac{\log \mathbb{P} [L_c > t]}{t} = \lambda, \end{aligned} \quad (16)$$

when $\beta > \gamma$, and

$$\begin{aligned} & - \lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(r)} > t, L_c > t \right]}{t} \\ &= \min \left\{ - \lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[L_c > tr / \left\lceil \frac{r\beta}{\gamma} \right\rceil \right]}{t}, \right. \\ & \quad \left. - \lim_{t \rightarrow \infty} \frac{\log \left(\sum_{n=\lceil r\beta/\gamma \rceil}^{r-1} \mathbb{P} \left[N_m^{(r)} > n, \frac{rt}{n+1} < L_c \leq \frac{rt}{n} \right] \right)}{t} \right\} \\ &= \min \left\{ \frac{\lambda r}{\lceil r\beta/\gamma \rceil}, \tilde{\Lambda}_3^o \right\} = \Lambda_3^o, \end{aligned} \quad (17)$$

when $\beta < \gamma$. Combining Equation (14), (15), (16), and (17), we get

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(r)} > t \right]}{t} \\ &= \max \left\{ \limsup_{t \rightarrow \infty} \frac{\log \sum_{n=1}^{\infty} \mathbb{P}_{ntr}}, \right. \\ & \quad \left. \lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(r)} > t, L_c > t \right]}{t} \right\} \\ &\stackrel{(a)}{\leq} \max \left\{ - \inf_n \frac{\lambda + \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha)}{n+1}, -\Lambda_3^o \right\} \\ &= - \min \{ \Lambda_1^o, \Lambda_3^o \}. \end{aligned} \quad (18)$$

The lower bound can be found in a similar manner. Notice that inequality (a) in the preceding equation is true because \mathbb{P}_{ntr} is nonzero only for a finite number of n , due to the fact that L_c cannot be less than 1.

In the special case when $r = 1$, by Lemma 4 and the definition of L_c , we have

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\log \mathbb{P}_{ntr}}{t} &= - \frac{\lambda + \Lambda_n(\beta, \Pi) \mathbf{1}(n \geq \alpha)}{n+1}, \\ \lim_{t \rightarrow \infty} \frac{\log \mathbb{P} \left[T_m^{(1)} > t, L_c > t \right]}{t} &= -\lambda, \end{aligned} \quad (19)$$

which, by combining Equation (14), completes the proof.

APPENDIX F PROOF OF THEOREM 2

From the definition of n_2^o , Λ_1^o and Λ_2^o in Notation 5 and by Lemma 4, we can obtain, for any $\delta > 0$,

$$\begin{aligned} \limsup_{t \rightarrow \infty} - \frac{\log \mathbb{P} \left[T_m^{(r)} > t, \frac{t}{n_2^o+1} < L_c \left(\frac{t}{n_2^o} + \delta \right) \leq \frac{t}{n_2^o} \right]}{t} &\leq \Lambda_2^o, \\ \liminf_{t \rightarrow \infty} - \frac{\log \mathbb{P} \left[T_m^{(r)} > t, \frac{t}{n_2^o+1} < L_c \left(\frac{t}{n_2^o} + \delta \right) \leq \frac{t}{n_2^o} \right]}{t} &\geq \Lambda_1^o. \end{aligned}$$

Then, for any $\eta > 0$ and for any $\delta > 0$, we can find $t(\eta)$ such that

$$\begin{aligned} - \frac{\log \mathbb{P} \left[T_m^{(r)} > t, \frac{t}{n_2^o+1} < L_c \left(\frac{t}{n_2^o} + \delta \right) \leq \frac{t}{n_2^o} \right]}{t} &\leq (1+\eta)\Lambda_2^o, \\ - \frac{\log \mathbb{P} \left[T_m^{(r)} > t, \frac{t}{n_2^o+1} < L_c \left(\frac{t}{n_2^o} + \delta \right) \leq \frac{t}{n_2^o} \right]}{t} &\geq (1-\eta)\Lambda_1^o, \end{aligned}$$

whenever $t > t(\eta)$. We denote $b(\eta) \triangleq \frac{t(\eta)}{n_2^o} + \delta$. In other words, for any $b > b(\eta)$, whenever $t \in [(b-\delta)n_2^o, bn_2^o]$,

$$\begin{aligned} - \frac{\log \mathbb{P} \left[T_m^{(r)} > t, \frac{t}{n_2^o+1} < L_c(b) \leq \frac{t}{n_2^o} \right]}{t} &\leq (1+\eta)\Lambda_2^o, \\ - \frac{\log \mathbb{P} \left[T_m^{(r)} > t, \frac{t}{n_2^o+1} < L_c(b) \leq \frac{t}{n_2^o} \right]}{t} &\geq (1-\eta)\Lambda_1^o, \end{aligned}$$

which, by using the same technique as in Equation (18), completes the proof of the first part. The second part of Theorem 2 follows by noting that

$$\lim_{t \rightarrow \infty} - \frac{\log \mathbb{P} \left[T_m^{(1)} > t, \frac{t}{n_1^o+1} < L_c \left(\frac{t}{n_1^o} + \delta \right) \leq \frac{t}{n_1^o} \right]}{t} = \Lambda_1^o,$$

where the definition of n_1^o can be found in Notation 5.

APPENDIX G PROOF OF THEOREM 3

1) If $\beta > \gamma$, by Lemma 3, for any $\varepsilon > 0$, we can find l_ε such that

$$\mathbb{P}[N_f > n | L_c = l_c] \geq \left(1 - e^{-l_\varepsilon \Lambda_1(\beta, \Pi)(1-\varepsilon)} \right)^n,$$

whenever $l_c > l_\varepsilon$. Then we have, for n large enough,

$$\begin{aligned} \mathbb{P}[N_f > n] &= \mathbb{E}[\mathbb{P}[N_f > n | L_c]] \\ &\geq \mathbb{E} \left[L_c > l_\varepsilon, \left(1 - e^{-\Lambda_1(\beta, \Pi)(1-\varepsilon)L_c} \right)^n \right] \\ &\geq \mathbb{E} \left[\frac{\log n}{\Lambda_1(\beta, \Pi)(1-\varepsilon)} < L_c < \frac{\log n}{\Lambda_1(\beta, \Pi)(1-\varepsilon)} + z, \right. \\ & \quad \left. \left(1 - e^{-\Lambda_1(\beta, \Pi)(1-\varepsilon)L_c} \right)^n \right] \\ &\geq \mathbb{E} \left[\frac{\log n}{\Lambda_1(\beta, \Pi)(1-\varepsilon)} < L_c < \frac{\log n}{\Lambda_1(\beta, \Pi)(1-\varepsilon)} + z, \right. \\ & \quad \left. \left(1 - e^{-(\log n)} \right)^n \right] \\ &\geq e^{-\lambda(1+\varepsilon) \frac{\log n}{\Lambda_1(\beta, \Pi)(1-\varepsilon)}} \left(1 - e^{-(\log n)} \right)^n. \end{aligned}$$

Taking logarithms on both sides of the preceding inequality, we get

$$\liminf_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n]}{\log n} \geq - \frac{\lambda(1+\varepsilon)}{\Lambda_1(\beta, \Pi)(1-\varepsilon)},$$

which, when $\varepsilon \rightarrow 0$, results in the lower bound.

Next, we prove the upper bound. Using the same technique as in the proof of the lower bound, and by the definition of L_c , we can find l_ε such that

$$\begin{aligned} \mathbb{P}[N_f > n] &\leq \mathbb{P} \left[L_c > l_\varepsilon, \left(1 - e^{-\Lambda_1(\beta, \Pi)(1+\varepsilon)L_c} \right)^n \right] \\ &\quad + \mathbb{P}[N_f > n, L_c \leq l_\varepsilon] \\ &\leq \sum_{l=l_\varepsilon}^{\infty} \left(1 - e^{-\Lambda_1(\beta, \Pi)(1+\varepsilon)l} \right)^n \mathbb{P}[L_c = l] + O(e^{-\xi n}) \\ &\leq O \left(\int_0^{\infty} \left(1 - e^{-\Lambda_1(\beta, \Pi)(1+\varepsilon)x} \right)^n e^{-\lambda(1-\varepsilon)x} dx \right) \\ &\quad + O(e^{-\xi n}), \end{aligned}$$

for some $\xi > 0$. Computing the integral in the preceding inequality, we obtain

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n]}{\log n} \leq -\frac{\lambda(1-\epsilon)}{\Lambda_1(\beta, \Pi)(1+\epsilon)},$$

which, with $\epsilon \rightarrow 0$, proves the upper bound.

Now, we prove the result for $\mathbb{P}[T_f > t]$. The upper bound follows by noting that, for any $h > 0$,

$$\begin{aligned} \mathbb{P}[T_f > t] &\leq \mathbb{P}[N_f L_c > t, L_c \leq h \log t] + \mathbb{P}[L_c > h \log t] \\ &\leq \mathbb{P}[N_f > t/(h \log t)] + \mathbb{P}[L_c > h \log t], \end{aligned}$$

where $\lim_{t \rightarrow \infty} \log \mathbb{P}[N_f > t/(h \log t)] / \log t = -\lambda/\Lambda_1(\beta, \Pi)$, and $\mathbb{P}[L_c > h \log t] = o(\mathbb{P}[N_f > t/(h \log t)])$ for h large enough.

The lower bound follows by noting that, for some $l_2 > l_1 > 0$ with $\mathbb{P}[l_1 < L_c < l_2] > 0$,

$$\begin{aligned} \mathbb{P}[T_f > t] &\geq \mathbb{P}[N_f L_c > t, l_1 < L_c < l_2] \\ &\geq \mathbb{P}[N_f > t/l_1] \mathbb{P}[l_1 < L_c < l_2]. \end{aligned}$$

2) Observe that

$$\begin{aligned} &\mathbb{P}[T_f > t] \\ &= \mathbb{P}[L_c > t] + \sum_{n=1}^{\infty} \mathbb{P}\left[T_f > t, \frac{t}{n+1} < L_c \leq \frac{t}{n}\right] \\ &= \mathbb{P}[L_c > t] + \sum_{n=1}^{\infty} \mathbb{P}\left[N_f > n, \frac{t}{n+1} < L_c \leq \frac{t}{n}\right]. \quad (20) \end{aligned}$$

By Lemma 3 we know that $\mathbb{P}[N_f > n | L_c = l_c]$ has a rate function of $n\Lambda_1(\beta, \Pi)$ with respect to the increase of l_c when $\beta > \gamma$. By using the same technique as that used in the proof of Lemma 4, we can easily get, when $\beta < \gamma$,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\log \mathbb{P}\left[N_f > n, \frac{t}{n+1} < L_c \leq \frac{t}{n}\right]}{t} &= -\frac{n\Lambda_1(\beta, \Pi) + \lambda}{n+1}, \\ \lim_{t \rightarrow \infty} \frac{\log \mathbb{P}[L_c > t]}{t} &= -\lambda, \end{aligned}$$

which, by combining Equation (20) and using the same technique as in Equation (18), yields

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\log \mathbb{P}[T_f > t]}{t} &= -\min\left\{\inf_{n \in \mathbb{N}} \left\{\frac{n\Lambda_1(\beta, \Pi) + \lambda}{n+1}\right\}, \lambda\right\} \\ &= -\min\{\Lambda_1(\beta, \Pi), \lambda\}. \end{aligned}$$

APPENDIX H

PROOF OF THEOREM 4

1) By Lemma 3 we know that when $\beta > \gamma$

$$\mathbb{P}[N_f > n | L_c = l_c] = \left(1 - e^{-l_c \Lambda_1(\beta, \Pi)(1+g(l_c))}\right)^n,$$

with $g(l_c) \in o(1)$ as $l_c \rightarrow \infty$. Let us denote l_n as the root of the function $l_c(1+g(l_c)) - \frac{\log n}{\Lambda_1(\beta, \Pi)}$. In other words,

$$l_n(1+g(l_n)) = \frac{\log n}{\Lambda_1(\beta, \Pi)}.$$

For any $b_0 > 0$, we have

$$\begin{aligned} &\mathbb{P}[N_f > n, l_n - z < L_c(l_n + b_0) \leq l_n] \\ &\geq \mathbb{P}[N_f > n | L_c = l_n] \mathbb{P}[l_n - z < L_c(l_n + b_0) \leq l_n]. \quad (21) \end{aligned}$$

Note that, by Lemma 3,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n | L_c = l_n]}{\log n} \\ &= \lim_{n \rightarrow \infty} \frac{\log \left(1 - e^{-l_n \Lambda_1(\beta, \Pi)(1+g(l_n))}\right)^n}{\log n} \\ &= \lim_{n \rightarrow \infty} \frac{n \log(1 + \frac{1}{n})}{\log n} = 0. \quad (22) \end{aligned}$$

Also, by the definition of $L_c(b)$,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[l_n - z < L_c(l_n + b_0) \leq l_n]}{\log n} \\ &= \lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[l_n - z < L_c(l_n + b_0) \leq l_n]}{l_n} \frac{l_n}{\log n} \\ &= -\lambda \frac{1}{\Lambda_1(\beta, \Pi)}. \quad (23) \end{aligned}$$

Combining Equation (21), (22) and (23), we get

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n, l_n - z < L_c(l_n + b_0) \leq l_n]}{\log n} \\ &= -\frac{\lambda}{\Lambda_1(\beta, \Pi)}. \end{aligned}$$

Therefore, for any $\eta > 0$, we can find a $n_1(\eta)$ such that

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n, l_n - z < L_c(l_n + b_0) \leq l_n]}{\log n} \\ &\geq -\frac{\lambda}{\Lambda_1(\beta, \Pi)}(1+\eta), \quad (24) \end{aligned}$$

whenever $n > n_1(\eta)$. Also, by Theorem 3, we can find $n_2(\eta)$ such that

$$\lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n]}{\log n} \leq -\frac{\lambda}{\Lambda_1(\beta, \Pi)}(1-\eta). \quad (25)$$

Let $n(\eta) \triangleq \max\{n_1(\eta), n_2(\eta)\}$ and $b(\eta) \triangleq l_{n(\eta)} + b_0$. By combining Equation (24) and (25), we know that for any $\eta > 0$, we can find $b(\eta)$ such that for any $b > b(\eta)$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f(b) > n]}{\log n} &\leq \lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n]}{\log n} \\ &\leq -\frac{\lambda}{\Lambda_1(\beta, \Pi)}(1-\eta), \end{aligned}$$

and

$$\begin{aligned} &\liminf_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f(b) > n]}{\log n} \\ &\geq \lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f > n, l_n - z < L_c(l_n + b_0) \leq l_n]}{\log n} \\ &\geq -\frac{\lambda}{\Lambda_1(\beta, \Pi)}(1+\eta), \end{aligned}$$

whenever $n \in [n(\eta), n_b]$, where n_b satisfies $b(1+g(b)) = \frac{\log n_b}{\Lambda_1(\beta, \Pi)}$. From Lemma 3, we know that

$$n_b = e^{\Lambda_1(\beta, \Pi)b(1+g(b))} = (\mathbb{P}[N_f = 1 | L_c = b])^{-1}. \quad (26)$$

2) Note that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[N_f(b) > n]}{n} \\ &= \max_{l_c} \left\{ \lim_{n \rightarrow \infty} \frac{\log(\mathbb{P}[L_c(b) = l_c, N_f(b) > n])}{n} \right\} \\ &= \max_{l_c} \left\{ \lim_{n \rightarrow \infty} \frac{\log(\mathbb{P}[L_c(b) = l_c])}{n}, \right. \\ & \quad \left. \log(\mathbb{P}[N_f(b) > 1 | L_c(b) = l_c]) \right\} \\ &= \log(\mathbb{P}[N_f(b) > 1 | L_c(b) = b]). \end{aligned}$$

The characterization of $\mathbb{P}[T_f(b) > t]$ follows by noting that $T_f(b) = N_f(b)L_c(b)$.

APPENDIX I PROOF OF THEOREM 5

Observe that

$$\Delta(b) = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \beta L_i}{n} \frac{n}{\sum_{i=1}^n T_i} = \beta \frac{\mathbb{E}[L_i]}{\mathbb{E}[T_i]}.$$

From Theorem 4 we know that for a given $\eta > 0$, we can find $n(\eta)$ and b large enough such that

$$\begin{aligned} \mathbb{E}[T_i] &\geq \int_{n(\eta)b}^{n_b b} \mathbb{P}[T_f(b) > t] dt \\ &\geq \int_{n(\eta)b}^{n_b b} t^{-\frac{\lambda}{\Lambda_1(\beta, \Pi)}(1+\eta)} dt, \end{aligned}$$

which, by combing the definition of n_b in Equation (26), yields

$$-\limsup_{b \rightarrow \infty} \frac{\log \Delta(b)}{b} = \liminf_{b \rightarrow \infty} \frac{\log \mathbb{E}[T_i]}{b} \geq \Lambda_1(\beta, \Pi) - \lambda.$$

REFERENCES

- [1] J. Tan, Y. Yang, N. Shroff, and H. El Gamal, "Delay asymptotics with retransmissions and fixed rate codes over erasure channels," in *INFOCOM, 2011 Proceedings IEEE*, April 2011, pp. 1260–1268.
- [2] D. P. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.
- [3] P. Jelenkovic and J. Tan, "Can retransmissions of superexponential documents cause subexponential delays?" in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, May 2007, pp. 892–900.
- [4] P. R. Jelenkovic and J. Tan, "Is aloha causing power law delays?" in *Proceedings of the 20th international teletraffic conference on Managing traffic performance in converged networks*, 2007, pp. 1149–1160.
- [5] P. R. Jelenković and J. Tan, "Are end-to-end acknowledgements causing power law delays in large multi-hop networks?" in *14th Inform Applied Probability Conference*, Eindhoven, July 9-11 2007.
- [6] J. Tan and N. Shroff, "Transition from heavy to light tails in retransmission durations," in *INFOCOM, 2010 Proceedings IEEE*, March 2010, pp. 1–9.
- [7] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman, "Efficient erasure correcting codes," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 569–584, Feb 2001.
- [8] M. Luby, "Lt codes," in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, 2002, pp. 271–280.
- [9] A. Shokrollahi, "Raptor codes," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2551–2567, June 2006.
- [10] J. R. Norris, *Markov Chains*. Cambridge, UK: Cambridge University Press, 1998.

- [11] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.
- [12] J. Forney, G., "Exponential error bounds for erasure, list, and decision feedback schemes," *Information Theory, IEEE Transactions on*, vol. 14, no. 2, pp. 206–220, March 1968.
- [13] "IEEE 802.11a. High Speed Physical Layer in the 5GHz band." *Draft Supplement to Standard IEEE 802.11, IEEE*, New York, January 1999.
- [14] "IEEE 802.11b. Higher Speed Physical Layer in the 2.4GHz band." *Draft Supplement to IEEE Standard for Information technology, IEEE*, New York, September 1999.

Yang Yang received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, China, in 2009. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, the Ohio State University. He interned with Qualcomm Corporate R&D, San Diego, during summer of 2013. Mr. Yang is a recipient of the Roberto Padovani Scholarship Award (2013).

Jian Tan received the B.S. degree from University of Science and Technology of China, Hefei, China in 2002, and the M.S. and Ph.D. degrees from Columbia University, New York, NY, USA in 2004 and 2008, respectively, all in Electrical Engineering. His Ph.D. thesis work won the Eliahu Jury Award from Columbia University. From 2010 to now, he is a Research Staff Member at IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. He was a postdoctoral researcher with the Networking and Communications Research Lab, The Ohio State University, Columbus, OH, USA, from 2009 to 2010. He interned with Lucent Bell Laboratories, Murray Hill, NJ, during the summers of 2005 and 2006, and with Microsoft Research, Cambridge, UK, in the winter of 2007. His current research interests focus on systems and theoretical modeling for large-scale distributed computing.

Ness B. Shroff (S'91-M'93-SM'01-F'07) received his Ph.D. degree from Columbia University, NY in 1994 and joined Purdue university immediately thereafter as an Assistant Professor. At Purdue, he became Professor of the school of Electrical and Computer Engineering in 2003 and director of CWSA in 2004, a university-wide center on wireless systems and applications. In July 2007, he joined Ohio State University as the Ohio Eminent Scholar of Networking and Communications. His research interests span the areas of wireless and wireline communication networks. He is especially interested in fundamental problems in the design, performance, control, and security of these networks. Dr. Shroff is a past editor for IEEE/ACM Trans. on Networking and the IEEE Communications Letters. He currently serves on the editorial board of the Computer Networks Journal. He has served on the technical and executive committees of several major conferences and workshops. He was the TPC co-chair of IEEE INFOCOM'03, ACM Mobihoc'08, and general chair of IEEE CCW'99 and WICON'08. Dr. Shroff is a fellow of the IEEE. He has received numerous awards for his work, including two best paper awards at IEEE INFOCOM (in 2006 and 2008), the flagship conference of the field and two of his papers were selected for the runner-up award at the same conference (in 2005 and 2013). He also received the 2005 best paper of the year award for the Journal of Communications and Networking, the 2003 best paper of the year award for Computer Networks, and the NSF CAREER award in 1996. His papers have also received the IEEE WiOPT 2013, IEEE WiOpt 2012, and the IWQoS best student paper award, the 2005 best paper of the year award.

Hesham El Gamal (M'99-SM'03-F'10) received the B.S. and M.S. degrees in electrical engineering from Cairo University, Cairo, Egypt, in 1993 and 1996, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Maryland at College Park, in 1999. From 1993 to 1996, he served as a Project Manager with the Middle East Regional Office of Alcatel Telecom. From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Maryland at College Park. From February 1999 to December 2000, he was with the Advanced Development Group, Hughes Network Systems (HNS), Germantown, MD, as a Senior Member of Technical Staff. Since January 2001, he has been with the Electrical and Computer Engineering Department, Ohio State University, where he is now a Professor. He held visiting appointments at UCLA, Institut Eurecom, and served as a Founding Director for the Wireless Intelligent Networks Center (WINC) at Nile University (2007-2009). He holds 12 patents. Dr. El Gamal is a recipient of the HNS Annual Achievement Award (2000), the OSU College of Engineering Lumley Research Award (2003, 2008), the OSU Electrical Engineering Department FARMER Young Faculty Development Fund (2003-2008), the OSU Stanley E. Harrison Award (2008), and the National Science Foundation CAREER Award (2004). He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS (2001-2005), an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING (2003-2007), a Guest Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY Special Issue on Cooperative Communications (2007), a member of the SP4COM Technical Committee (2002-2005), a Co-Chair of the Globecom'08 Communication Theory Symposium, and a Co-Chair of the 2010 IEEE Information Theory Workshop. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and a Guest Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY.