Delay Asymptotics with Retransmissions and Fixed Rate Codes over Erasure Channels

Jian Tan*, Yang Yang†, Ness B. Shroff†, Hesham El Gamal†
†Department of Electrical and Computer Engineering
The Ohio State University, Columbus 43210, OH
*IBM T.J. Watson Research, Hawthorne 10532, NY

***Abstract*—Recent work has shown that retransmissions can cause heavy-tailed transmission delays even when packet sizes are light-tailed. Moreover, the impact of heavy tailed delays persist even when packets are of finite size. The key question we study in this paper is how the use of coding techniques to transmit information could mitigate delays. To investigate this problem, we consider an important communication channel called the Binary Erasure Channel, where transmitted bits are either received successfully or lost (called an *erasure*). This model is a good abstraction of not only the wireless channel but also the higher layer link, where erasure errors can happen. Many coding schemes, known as erasure codes, have been designed for this channel. Specifically, we focus on the fixed rate coding scheme, where decoding is said to be successful if a certain fraction β of the codeword is received correctly. We study two different scenarios: (I) A codeword of length L_c is retransmitted as a unit until the receiver successfully receives more than βL_c bits in the last transmission. (II) All successfully received bits from every (re)transmissions are buffered at the receiver according to their positions in the codeword, and the transmission completes once the received bits become decodable for the first time.**

Our studies reveal that complicated and surprising relationships exist between the coding complexity and the transmission delay/throughput. From a theoretical perspective, our results provide a benchmark to quantify the tradeoffs between coding complexity and transmission throughput for receivers that use memory to buffer (re)transmissions until success and those that do not buffer intermediate transmissions.

I. Introduction

The use of retransmissions is a fundamental mechanism to ensure reliable transfer of data over communication channels and networks [1]. Recent studies [2], [3], [4] have revealed that all retransmission-based protocols could cause heavy-tailed transmission delays, resulting in very long delays and possibly zero throughput. Moreover, the distribution of the delay could have a large heavy tailed component, even when packets are bounded. In this paper, we investigate the use of coding techniques to transmit information, which could substantially reduce the delay and improve the throughput.

This work has been supported in part by NSF project CNS-0721236 and ARO MURI project W911NF-08-1-0238.

In our analysis, we consider an important communication channel called the *Binary Erasure Channel*. The Binary erasure channel, first introduced by Elias [5] in 1954, has been found to be invaluable in characterizing a number of typical communication channels. Erasures occur when the bits are not received successfully during the transmission while the positions of the erroneous bits are known in the received stream. This model describes the situation when information may get lost due to a variety of factors (e.g., signal fading, interference, obstructions, contention with other nodes and node mobility). The Binary erasure channel captures these errors in a direct way: the binary bit transmitted is either received successfully or lost.

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 the buffer overflow at intermediate nodes and therefore never reach the destination [6]. All these errors can result in erasures in the received bit stream.

It is well known that the capacity of a binary erasure channel with erasure probability of $1-\gamma$ is equal to γ [7], which is difficult to achieve using only feedback and retransmissions. Therefore, many coding schemes, known as erasure codes, have been designed for this channel. Using erasure codes, even when some portions of the codeword are lost, it is still possible for the receiver to recover the corresponding message using the rest of successfully received bits. Roughly speaking, the encoder transforms a message of L symbols into a longer codeword of L_c symbols, where the ratio $\beta_\epsilon = L/L_c$ is called the code rate. An erasure code is said to be near optimal if it requires slightly more than L symbols, say $(1+\varepsilon)L$ ones ($\varepsilon > 0$), to recover the message, where ε can be made arbitrary small at the cost of increased encoding and decoding complexity. Until now, many elegant low complexity erasure codes have been designed for erasure channels. They can be divided into two broad categories: fixed rate and rateless codes. Fixed rate erasure code, e.g. Tornado Code [8], is named so because the code rate β_ε is fixed during the transmission. On the other hand, rateless erasure codes, also known as fountain codes, e.g. LT-code [9] and Raptor code [6], takes a different approach. It generates infinite output blocks for an input message of length L, which results in a variable code rate. Yet it can still guarantee that any successfully received $(1+\varepsilon)L$ number of bits result in successful decoding.

In terms of the time complexity for encoding and decoding, the best erasure code is of the order $O\left((\log 1/\varepsilon)L/\beta_\epsilon\right)$ [8][6]. Throughout this paper, we only focus on fixed rate codes, where decoding is said to be successful if a fixed fraction $\beta \triangleq \beta_\varepsilon(1+\varepsilon)$ of the codeword is received correctly.

Specifically, we study two different scenarios in this paper. (I) **Without memory**: A codeword of length L_c is retransmitted as a unit until the receiver successfully receives more than βL_c number of bits in the last transmission. This is the typical scenario in most current communication paradigms, where the receiver does not keep track of which bits were received successfully. This scenario occurs because receivers may not have the requisite computation/storage power to keep track of all the erasure positions and the bits that have been successfully received, especially when the receiver is responsible for handling a large number of flows simultaneously. (II) **With memory**: All successfully received bits from every (re)transmissions are buffered at the receiver according to their positions in the codeword, and the transmission completes once the received bits become decodable for the first time. With increasing processor and memory speeds, this scenario is likely to become standard in future communications.

The main contributions of our work can be summarized as follows: Our studies reveal complicated relationships between the coding complexity (determined by β) and the transmission delay/throughput. For example, although a smaller β implies a higher coding/decoding complexity and a longer codeword, a longer codeword does not necessarily cause a longer transmission delay. However, we note that if a codeword is too long it degrades the throughput as well, since the throughput goes to zero when $\beta \to 0$. Under a general Markov erasure channel (correlated channel) with a codeword length having an exponential (light) tail, we show that, when the receiver cannot utilize the successfully received bits from previous transmissions of the same codeword (memoryless case), the system exhibits an intriguing phase transition phenomenon: the transmission delay follows power law distribution if $\beta > \gamma$ (recalling that γ is the erasure channel capacity) and exponential distribution if $\beta < \gamma$. This phase transition phenomenon may have an important impact on the channel throughput: the system will experience a zero throughput when the transmission delay follows a power law with index less than one; the delay will have less variability if the delay distribution is light-tailed, which is more desirable if the decay rate of this distribution is large. For both cases, we characterize the distribution of the delay in Theorem 1, and show how they are related to the channel dynamics and codeword length variability.[1] On the other hand, if the receiver can combine the received bits of the same codeword from all (re)transmissions, the delay distribution is always exponential, with a complicated decay rate that depends on both codewords and channel dynamics. The computation of this decay rate involves an optimization problem, which can be solved using numerical methods. From a theoretical perspective, our results provide benchmarks to quantify the tradeoffs between coding complexity and transmission throughput for receivers with and without memory.

The remainder of this paper is structured as follows: after the model description in Section II, we provide the results for the situation without memory in Section III, where a phase transition phenomenon for the delay distribution is presented. Then, in Section IV, we investigate the situation with memory, and show that the delay distribution is light tailed. All the proofs are presented in Section VI. Finally, in Section V, we provide numerical studies to verify our main results.

II. System Model

We denote the number of bits of the codeword by L_c, which has a lower bound $l_{\min}$ with $l_{\min} \triangleq \inf\{x : \mathbb{P}[L_c > x] < 1\}$. We model the channel dynamics as a slotted system such that within each time slot only one bit can be transmitted. 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 lost (called an erasure). We focus on the fixed rate codes, where decoding is successful when a fixed fraction $0 < \beta < 1$ of the codeword is received correctly.

In practice, the channel dynamics are often temporarily correlated. To this end, we investigate the situation where the current channel status distribution only depends on the preceding $k \geq 0$ time slots. More precisely, for $\mathcal{F}_n = \{X_i\}_{i\leq n}, n \geq 1$ and fixed $k \geq 0$, we define $\mathcal{H}_n = \{X_n, \cdots, X_{n-k+1}\}$ for $n \geq k \geq 1$ with $\mathcal{H}_n = \{\emptyset, \Omega\}$ for $k = 0$, and assume throughout this paper that $\mathbb{P}[X_n = 1|\mathcal{F}_{n-1}] = \mathbb{P}[X_n = 1|\mathcal{H}_{n-1}]$ for all $n \geq k$. In other words, the augmented state $\bar{Y}_n \triangleq (X_n, \cdots, X_{n-k}), n \geq k$ form a Markov chain. Denote by Π the transition matrix (2^k by 2^k) of this Markov chain $\{Y_n\}_{n\geq k+1}$, where

$$\Pi = \left(\pi(s,u)\right)_{s,u\in\{0,1\}^k} \tag{1}$$

[1]As an aside, it should be noted that although bounded packet sizes result in the tail of the delay distribution being eventually exponential, the main body of the distribution, as shown in [10], can still follow a power law (i.e., heavy tailed). Moreover, as shown in [10], the heavy tailed main body could dominate even for relatively small values of the maximum packet size. This implies that the impact of retransmissions on delays needs to be carefully examined and controlled.

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 [11]. Therefore, for any initial value $\mathcal{H}_k$, the parameter γ is well defined

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

and, from the ergodic theorem (see Theorem 1.10.2 in [11]),

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \gamma\right] = 1.$$

Note that the value of the current channel state X_n is equal to the first element of the vector $\bar{Y}_n$. Thus, we define the function $f(\cdot)$ that returns this element for a vector drawn from the set $\{0,1\}^k$, i.e.,

$$f([X_n, \cdots, X_{n-k+1}]) = X_n.$$

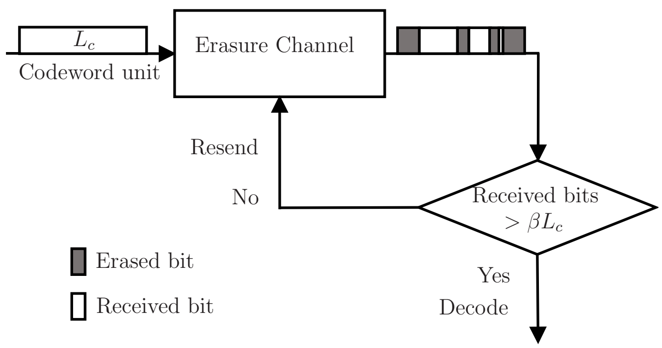

Fig. 1. Codewords sent over erasure channel

We investigate two scenarios: with and without memory, as discussed in the Introduction.

Definition 1 (Without memory)**.** *The total number of transmissions for a codeword of length L_c is defined as*

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

and, the total transmission time is $T_f \triangleq N_f L_c$.

Definition 2 (With memory)**.** *The total number of transmissions for a codeword of length L_c is defined as*

$$N_m \triangleq \inf\left\{n : \sum_{i=1}^{L_c} \mathbf{1}\left(\sum_{j=1}^{n} X_{(j-1)L_c+i} \geq 1\right) > \beta L_c\right\},$$

and, the total transmission time is $T_m \triangleq N_m L_c$.

In Sections III and IV, we will investigate the delay asymptotics of the retransmission system for each of the above cases. However, we first study the failure probability of just a single transmission.

A. Failure probability of a single transmission

We first investigate the failure probability of a single transmission $\mathbb{P}[N_f > 1]$ when L_c is a fixed value, i.e., $L_c \equiv l_c$, and β is very close to γ (more general relationships between β and γ are studied in later sections). The following result characterizes the relationship between β (which determines the code rate and complexity) and the failure probability for one transmission $\mathbb{P}[N_f > 1]$.

Proposition 1. *If $\{X_i\}_{i\geq 1}$ is an i.i.d. sequence and $\beta = \gamma(1+\alpha\sqrt{\gamma/(\beta l_c)})^{-1}$ for some fixed α, then, for $0<\gamma<1$,*

$$\lim_{l_c\to\infty} \mathbb{P}[N_f > 1] = \mathbb{Q}\left(\alpha\sqrt{\gamma(1-\gamma)^{-1}}\right).$$

where $\mathbb{Q}(x) = \int_x^\infty 1/\sqrt{2\pi} e^{-u^2/2} du$.

Proof: For the Bernoulli channel model,

$$\begin{aligned}
\lim_{l_c\to\infty} \mathbb{P}[N_f > 1] &= \lim_{l_c\to\infty} \mathbb{P}\left[\sum_{i=1}^{l_c} X_i \leq \beta l_c\right] \\
&= \lim_{l_c\to\infty} \mathbb{P}\left[\frac{\sum_{i=1}^{l_c} X_i - \gamma l_c}{\sqrt{l_c \mathrm{Var}(X)}} \leq \frac{(\beta-\gamma) l_c}{\sqrt{l_c \mathrm{Var}(X)}}\right] \\
&= \lim_{l_c\to\infty} \mathbb{Q}\left(\frac{(\gamma-\beta)\sqrt{l_c}}{\sqrt{\mathrm{Var}(X)}}\right) = \mathbb{Q}\left(\alpha\sqrt{\gamma(1-\gamma)^{-1}}\right).
\end{aligned}$$

■

We can prove a similar result for the more general Markov channel. Due to limited space, we present it in the technical report [12]. Since $\mathbb{Q}(x)$ decreases very fast, choosing β close to γ, according to Proposition 1, gives a good balance between complexity and failure probability. On the other hand, the preceding result shows that the failure probability $\mathbb{P}[N_f > 1]$ is very sensitive to α, which implies that the error from estimating γ can change the failure probaility of a single transmission dramatically if choosing $\beta \approx \gamma$. Note that $\lim_{l_c\to\infty} \beta/\gamma = 1$ in Proposition 1; see Example 1 in Section V for more discussions.

III. Receiver Without Memory

For receivers that do not have the required computation/storage power, it is difficult to keep track of all the erasure positions and the bits that have been successfully received. In this section we study the situation when the transmission only completes when the number of successfully received bits in the last transmission exceeds β fraction of the codeword.

Interestingly, we observe an intriguing phase transmission phenomenon for this situation. We show that, under a general Markov channel model, when the length of the codeword has an exponential tail, the transmission delay is light-tailed (exponential) only if $\gamma > \beta$, and heavy-tailed (power law) if $\gamma < \beta$.

In order to present the results, we first introduce some necessary definitions. Recalling Equation (1) and the function $f(\cdot)$ defined afterwards, for a real number θ and $k \geq 1$, we define a matrix $\Pi_\theta = (\pi_\theta(s,u))$ by

$$\pi_\theta(s,u) = \pi(s,u)e^{\theta f(u)}, s, u \in \{0,1\}^k, k \geq 1,$$

and for $k=0$ (when $\{X_i\}_{i\geq 1}$ is an i.i.d. sequence),

$$\pi_\theta(s,u) = \mathbb{P}[X_1 = u]e^{\theta f(u)}, s, u \in \{0,1\}, k = 0.$$

Definition 3. *Let $\rho(\pi_\theta)$ denote the Perron-Frobenius eigenvalue (see Theorem 3.1.1 in [13]) of the matrix Π_θ, which is the largest eigenvalue of Π_θ.*

Theorem 1 (Phase Transition Phenomenon). *If there exists $\lambda > 0$ and $z > 0$, such that,*

$$\lim_{x\to\infty} \frac{\log \mathbb{P}[x < L_c \leq x+z]}{x} = -\lambda, \tag{2}$$

we obtain:

1) *If $\beta > \gamma$, then*

$$\lim_{n\to\infty} \frac{\log \mathbb{P}[N_f > n]}{\log n} = \lim_{t\to\infty} \frac{\log \mathbb{P}[T_f > t]}{\log t} = -\frac{\lambda}{\Lambda(\beta)}, \tag{3}$$

where $\Lambda(\beta) \triangleq \sup_\theta\{\theta\beta - \log\rho(\pi_\theta)\}$.

2) *If $\beta < \gamma$, then,*

$$\lim_{l_{\min}\to\infty} \lim_{t\to\infty} \frac{\log \mathbb{P}[T_f > t]}{t} = -\min\{\Lambda(\beta), \lambda\}. \tag{4}$$

Remark 1.1. The tail distribution of the transmission delay changes from power law (3) to exponential (4), depending on the relationship between the parameters β and γ. If $\lambda/\Lambda(\beta) < 1$, the system even has a zero throughput. The study of the critical case $\beta = \gamma$ requires a more refined scaling, which is related to Proposition 1. Equation (4) assumes that $l_{\min}$ is large, which is a fact in many real communication systems. Condition (2) can be relaxed to $z = o(\log x)$; we avoid this generalization due to limited space.

Proof: See Section VI. ■

For the special case when $\{X_i\}$ is an i.i.d. sequence ($k=0$), we can compute $\Lambda(\beta)$ explicitly, as shown in the following corollary.

Corollary 1.1. *Under the assumptions of Theorem 1 and that $\{X_i\}_{i\geq 1}$ are i.i.d., we obtain*

$$\Lambda(\beta) = \beta \log\frac{\beta}{\gamma} + (1-\beta)\log\frac{1-\beta}{1-\gamma}.$$

Proof: Since $\pi(1,0) = 1-\gamma, \pi(1,1) = \gamma, \pi(0,1) = \gamma, \pi(0,0) = 1-\gamma$, we obtain

$$\Pi_\theta = \begin{pmatrix} 1-\gamma & \gamma e^\theta \\ 1-\gamma & \gamma e^\theta \end{pmatrix}.$$

Then, to compute $\Lambda(\beta)$ is straightforward. ■

IV. Receiver With Memory

For some applications, the receiver is equipped with powerful devices that have the ability to keep track of all the erasure positions and the bits that have been successfully received from all the (re)transmissions of the same codeword. Different from the situation without memory, we show that the transmission delays will be light-tailed and have no phase transition under the general Markov channel model. To compute the decay rate is complicated for general Markov channels, and therefore, we study the problem under the more restricted condition when $\{X_i\}_{i\geq 1}$ is an i.i.d. sequence. The decay rate of the delay distribution involves a complicated optimization problem. We use numerical methods to solve it in Example 4 of Section V.

When $\{X_i\}_{i\geq 1}$ are i.i.d., using large deviation results (e.g., Exercise 1.17 in [14], we know that, for $x \in (0,1)$ and $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n (1-X_i) \in (x(1-\epsilon), x(1+\epsilon))\right] = -\inf_{x(1-\epsilon)<y<x(1+\epsilon)} \Lambda^c(y),$$

where

$$\Lambda^c(x) = x\log\frac{x}{1-\gamma} + (1-x)\log\frac{1-x}{\gamma}. \tag{5}$$

To present our result, we introduce some necessary notation. For a sequence $\{\beta_i\}_{j\geq 1}, 0 < \beta_j < 1$, we let $\nu_j \triangleq \prod_{j=1}^i \beta_j, i \geq 1$ with $\nu_0 = 1$.

Theorem 2. *Under condition (2), there exists $h, \delta > 0$ such that*

$$\mathbb{P}[T_m > t] < he^{-\delta t}. \tag{6}$$

In addition, if $\{X_i\}_{i\geq 1}$ are i.i.d., then

$$\lim_{t\to\infty} \frac{\log \mathbb{P}[T_m > t]}{t} = -\min\{\Lambda^\circ, \lambda\}, \tag{7}$$

where

$$\Lambda^\circ = \inf_{\substack{\prod_{j=1}^{\lfloor y \rfloor} \beta_j \geq 1-\beta \\ y \geq 1}} y^{-1}\left(\sum_{j=1}^{\lfloor y \rfloor} \Lambda^c(\beta_j)\nu_{j-1} + \lambda\right).$$

Remark 2.1. Equation (6) shows that, when the receiver can combine all the bits from each (re)transmission of the codeword, the delay is light-tailed (the distribution is upper bounded by an exponential function) for general Markov channel. From the result (7), we see that if $\Lambda^\circ > \lambda$, then, combining all the bits from each (re)transmissions can dramatically improve the system performance, since it results in a delay distribution that is of the same order of the distribution of L_c, which is optimal in view of $T_m \geq L_c$.

Proof: See Section VI. ■

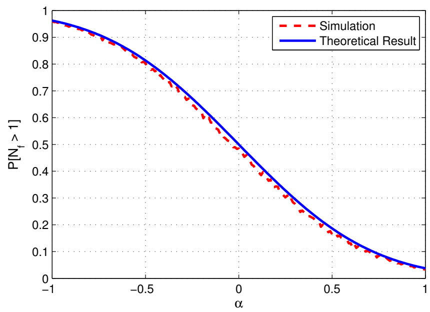

Fig. 2. Illustration for Example 1

V. Numerical And Simulation Examples

In this section, we conduct simulation results to verify our main results. As is evident from the following figures, the simulations match theoretical results quite well. In addition, when an explicit expression of the analytical result is not possible, e.g., for Theorem 2, we use numerical methods to solve the optimization problem for computing the asymptote.

Example 1: This example verifies Proposition 1, where the codeword length l_c is fixed. We choose a Markov channel with $k = 3$, $\gamma = 0.4706$ and $\beta l_c = 1000$. According to the result, if $\beta = \gamma(1+\alpha\sqrt{\gamma/(\beta l_c)})^{-1}$, then $\lim_{l_c \to \infty} \mathbb{P}[N_f > 1] = \mathbb{Q}(\xi\alpha)$, where ξ has an complicated explicit expression (due to the limited space, we present the details in the technical report [12]). Here we directly present the numerical result $\xi = 1.782$ that is computed using the transition matrix (8×8, not shown here) of the Markov channel. We plot $\mathbb{Q}(\xi\alpha)$ and the simulated result for $\mathbb{P}[N_f > 1]$ in Figure 2. From the figure, it is clear that the $\mathbb{Q}(\xi\alpha)$ function approximates $\mathbb{P}[N_f > 1]$ closely. Note that $\mathbb{P}[N_f > 1]$ is sensitive with respect to α. On the other hand, the corresponding code rate $\beta = 1/(2.125 + 0.04610\alpha)$ is not sensitive to α. Therefore, carefully choosing β compared to γ is very important in practice, especially when it may be difficult to obtain an accurate estimate of γ

Example 2: In this example, we illustrate the interesting phase transition phenomenon that occurs when receivers do not combine previously received bits to decode (memoryless case). We choose a Bernoulli channel where $\{X_i\}$ is an i.i.d. sequence with $\gamma = E[X_1] = 0.2$ and assume that the codeword length L_c is geometrically distributed with mean 100.

First, in Figure 3, we show that the delay distribution follows a power law distribution when $\beta < \gamma$. This experiment takes 3 sets of code rate: $\beta_1 = 0.24, \beta_2 = 0.25, \beta_3 = 0.26$. By Theorem 1 (or Corollary 1), $\mathbb{P}[N_f > n]$ follows power law distribution with exponent equal to -2.095, -1.355 and -0.9503, respectively. We plot the simulation results for 10^6 samples and the corresponding asymptotes on the log-log scale in Figure 3. As you can see, they match very well for large n.

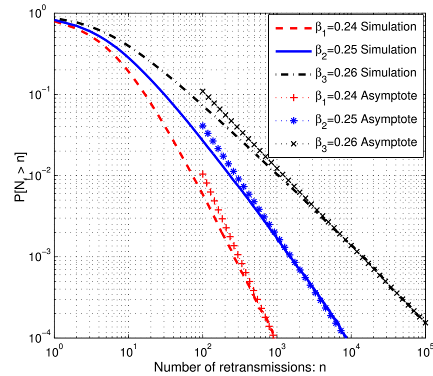

Fig. 3. Illustration for Example 2

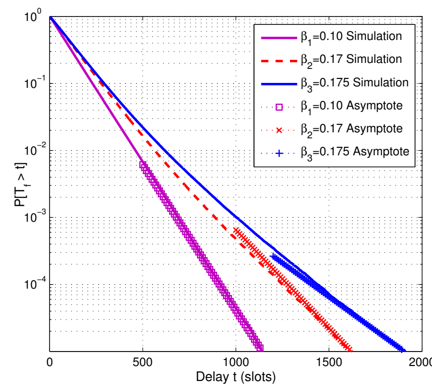

Fig. 4. Illustration for Example 2

Next, we show that the delay distribution is exponential when $\beta > \gamma$. We take three sets of code rate: $\beta_1 = 0.100, \beta_2 = 0.170, \beta_3 = 0.175$. According to Theorem 1, for these three settings, $-\min\{\Lambda(\beta), \lambda\}$ is equal to -0.0043, -0.0029 and -0.0020, respectively. Notice that $\lambda > \Lambda(\beta_2)$, $\lambda > \Lambda(\beta_3)$ and $\lambda < \Lambda(\beta_1)$. We plot the results in Figure 4. Again, they match very well.

Example 3: In this experiment we verify Theorem 1 when channel dynamics are correlated with $k = 1$. Assume that the codeword length L_c is geometrically distributed with mean 100. Let $\mathbb{P}[X_{i+1} = 1 | X_i = 0] = p_{01}, \mathbb{P}[X_{i+1} = 0 | X_i = 1] = p_{10}$, and we choose two sets of $p_{01}^j, p_{10}^j, j = 1, 2$. For $p_{01}^{(1)} = 0.2, p_{10}^{(1)} = 0.8$ and $p_{01}^{(2)} = 0.1, p_{10}^{(2)} = 0.4$, it is clear that $\gamma = p_{01}^{(j)}/(p_{01}^{(j)} + p_{10}^{(j)}) = 0.2$ for $j = 1, 2$. Assuming code rate $\beta = 0.25$, we know, by Theorem 1, that the distribution of the number of retransmisions and delay will both follow power laws. Using numerical method, we can compute the power law decay rates -1.3546 and -4.1661, repectively. We plot them in Figure 5.

Example 4: For receivers that can combine all the received bits from all (re)transmissions, the delay distribution is always light-tailed, as shown in Theorem 2. However, the computation of the decay rate involves

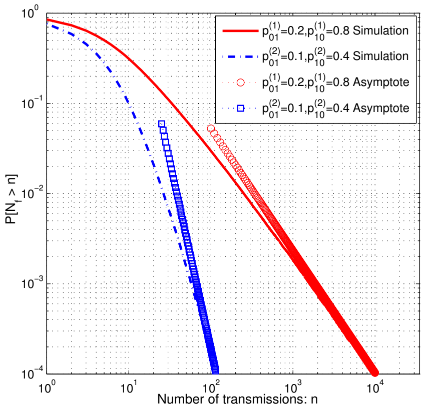

Fig. 5. Illustration for Example 3

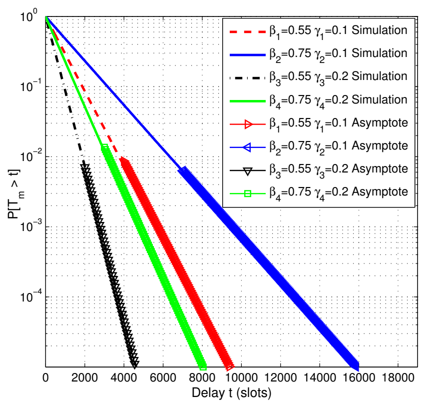

Fig. 6. Illustration for Example 4

a complicated optimization problem. In order to verify this result, we assume that the codeword length L_c is geometrically distributed with mean 100, and choose four different sets of parameters: $\beta_1 = 0.55, \gamma_1 = 0.1, \beta_2 = 0.75, \gamma_2 = 0.1, \beta_3 = 0.55, \gamma_3 = 0.2$, and $\beta_4 = 0.75, \gamma_4 = 0.2$. The corresponding decay rates $-\min\{\Lambda^\circ, \lambda\}$ can be computed by solving the optimization problem numerically, -5.4287×10^{-4}, -3.1599×10^{-4}, -1.1000×10^{-3}, -6.2042×10^{-4}. From Figure 6, we can see that the asymptotic results are quite accurate.

VI. Proofs

A. Proof of Theorem 1

In order to the prove Theorem 1, we need the following lemma that covers the case $\beta > \gamma$ and $\beta < \gamma$, respectively.

Lemma 2.1. *For $\epsilon > 0$, $j \geq 1$ and any values of $\{X_i\}_{(j-1)l+1 \leq i \leq (j-1)l+k}$, there exists $l_\epsilon > 0$ such that, for all $l > l_\epsilon$,*

1) *If $\beta > \gamma$, then*

$$\mathbb{P}\left[\sum_{i=(j-1)l+k+1}^{jl} X_i > \beta l\right] \geq e^{-\Lambda(\beta)(1+\epsilon)l}, \qquad (8)$$

$$\mathbb{P}\left[\sum_{i=(j-1)l+k+1}^{jl} X_i > \beta l - k\right] \leq e^{-\Lambda(\beta)(1-\epsilon)l}. \qquad (9)$$

2) *If $\beta < \gamma$, then*

$$\mathbb{P}\left[\sum_{i=(j-1)l+k+1}^{jl} X_i \leq \beta l\right] \geq e^{-\Lambda(\beta)(1+\epsilon)l}, \qquad (10)$$

$$\mathbb{P}\left[\sum_{i=(j-1)l+k+1}^{jl} X_i \leq \beta l - k\right] \leq e^{-\Lambda(\beta)(1-\epsilon)l}. \qquad (11)$$

Proof: This lemma is a direct application of Theorem 3.1.2 in [13]. ■

Proof of Theorem 1: 1) Observe the event that the transmission of L_c fails in the first $n \geq 1$ times is equivalent to

$$\sum_{i=1}^{L_c} X_i \leq \beta L_c, \sum_{i=L_c+1}^{2L_c} X_i \leq \beta L_c, \cdots, \sum_{i=(n-1)L_c+1}^{nL_c} X_i \leq \beta L_c,$$

implying that

$$\mathbb{P}[N_f > n \mid L_c] = \mathbb{P}\left[\bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c\right\}\middle| L_c\right].$$

Due to the dependencies along the sequence $\{X_i\}$, the events $\left\{\sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c\right\}$, $1 \leq j \leq n$ are not independent. Now, we will construct upper and lower bounds where we can decouple these events.

First, we prove the lower bound. Note that, for $L_c > k$,

$$\bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+1}^{jL_c} X_i \leq \beta L_c\right\} \supseteq \bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+k+1}^{jL_c} X_i \leq \beta L_c - k\right\}, \qquad (12)$$

therefore, if we ignore the first k bits for each transmission of L_c, we get a lower bound to $\mathbb{P}[N > n|L_c]$,

$$\mathbb{P}[N_f > n \mid L_c] \geq \mathbb{P}[N_f > n, L_c > k \mid L_c] \geq \mathbb{P}\left[L_c > k, \bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k\right\}\middle| L_c\right].$$

Let $\mathcal{E}_j = \{X_{(j-1)L_c+1}, \cdots, X_{(j-1)L_c+k}\}$, $1 \leq j \leq n$. Due to the memoryless property of Markov chain, we know that, conditional on $\mathcal{E}_n$, the events $\bigcap_{1\leq j\leq n-1}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k\right\}$ and

$\left\{\sum_{i=(n-1)L_c+1+k}^{nL_c} X_i \leq \beta L_c - k\right\}$ are independent. Therefore,

$$\mathbb{P}\left[L_c > k, \bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k\right\}\middle| L_c\right]$$
$$= \mathbb{E}\Bigg[\mathbb{P}\Bigg[L_c > k, \bigcap_{1\leq j\leq n-1}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k\right\}\middle| L_c, \mathcal{E}_n\Bigg]$$
$$\times \mathbb{P}\left[L_c > k, \sum_{i=(n-1)L_c+1+k}^{nL_c} X_i \leq \beta L_c - k \middle| L_c, \mathcal{E}_n\right]\Bigg| L_c\Bigg]. \tag{13}$$

Now, in view of (9), for $\epsilon > 0$ and l_ϵ chosen in Lemma 2.1, we have, using (9) and the independence of L_c and $\{X_i\}_{i\geq 1}$,

$$\mathbb{P}\left[L_c > k, \sum_{i=(n-1)L_c+1+k}^{nL_c} X_i \leq \beta L_c - k \middle| L_c, \mathcal{E}_n\right]$$
$$\geq \mathbb{P}\left[L_c > l_\epsilon, \sum_{i=(n-1)L_c+1+k}^{nL_c} X_i \leq \beta L_c - k \middle| L_c, \mathcal{E}_n\right]$$
$$\geq \mathbf{1}(L_c > l_\epsilon)\left(1 - e^{-\Lambda(\beta)(1-\epsilon)L_c}\right),$$

which, in combination with (13), implies that

$$\mathbb{P}\left[L_c > k, \bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k\right\}\middle| L_c\right]$$
$$\geq \mathbb{P}\left[L_c > l_\epsilon, \bigcap_{1\leq j\leq n-1}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c - k\right\}\middle| L_c\right]$$
$$\times \left(1 - e^{-\Lambda(\beta)(1-\epsilon)L_c}\right). \tag{14}$$

By the same approach to condition on $\mathcal{E}_{n-1}$, we can repeat the preceding argument to prove that

$$\mathbb{P}\left[L_c > l_\epsilon, \bigcap_{1\leq j\leq n-1}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c\right\}\middle| L_c\right]$$
$$\geq \mathbb{P}\left[L_c > l_\epsilon, \bigcap_{1\leq j\leq n-2}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c\right\}\middle| L_c\right]$$
$$\times \left(1 - e^{-\Lambda(\beta)(1-\epsilon)L_c}\right)^2,$$

which, by further conditioning on $\mathcal{E}_{n-2}, \cdots$, results in

$$\mathbb{P}[N_f > n \mid L_c] \geq \mathbf{1}\left[L_c > l_\epsilon\right]\left(1 - e^{-\Lambda(\beta)(1-\epsilon)L_c}\right)^n.$$

Therefore, recalling condition (2) and unconditioning on L_c, we obtain, for n large enough,

$$\mathbb{P}[N_f > n] = \mathbb{E}\left[\mathbb{P}[N_f > n \mid L_c]\right]$$
$$\geq \mathbb{E}\left[L_c > l_\epsilon, \left(1 - e^{-\Lambda(\beta)(1-\epsilon)L_c}\right)^n\right]$$
$$\geq \mathbb{E}\Big[\frac{\log n}{\Lambda(\beta)(1-\epsilon)} < L_c < \frac{\log n}{\Lambda(\beta)(1-\epsilon)} + z, \left(1 - e^{-\Lambda(\beta)(1-\epsilon)L_c}\right)^n\Big]$$
$$\geq \mathbb{E}\Big[\frac{\log n}{\Lambda(\beta)(1-\epsilon)} < L_c < \frac{\log n}{\Lambda(\beta)(1-\epsilon)} + z, \left(1 - e^{-(\log n)}\right)^n\Big]$$
$$\geq e^{-\lambda(1+\epsilon)\frac{\log n}{\Lambda(\beta)(1-\epsilon)}}\left(1 - e^{-(\log n)}\right)^n.$$

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

$$\limsup_{n\to\infty}\frac{\log \mathbb{P}[N_f > n]}{\log n} \geq -\frac{\lambda(1+\epsilon)}{\Lambda(\beta)(1-\epsilon)},$$

which, when $\epsilon \to 0$, results in the lower bound.

Next, we prove the upper bound. Note that

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

which implies that

$$\mathbb{P}[N_f > n] \leq \mathbb{P}\left[\bigcap_{1\leq j\leq n}\left\{\sum_{i=(j-1)L_c+1+k}^{jL_c} X_i \leq \beta L_c\right\}\right].$$

Using the same technique as in the proof of the lower bound and Equation (8), we can prove

$$\mathbb{P}[N_f > n] \leq \mathbb{P}\left[L_c > l_\epsilon, \left(1 - e^{-\Lambda(\beta)(1+\epsilon)L_c}\right)^n\right]$$
$$+ \mathbb{P}[N_f > n, L_c \leq l_\epsilon]$$
$$\leq \sum_{l=l_\epsilon}^{\infty}\left(1 - e^{-\Lambda(\beta)(1+\epsilon)l}\right)^n \mathbb{P}[L_c = l] + O(e^{-\xi n}),$$

since $\mathbb{P}[N_f > n, L_c \leq l_\epsilon] = O(e^{-\xi n})$ for some $\xi > 0$.

Condition (2) implies that $\mathbb{P}[L_c = l] \leq e^{-\lambda(1-\epsilon)l}$ for $l > l_\epsilon$, and thus

$$\mathbb{P}[N_f > n] \leq O\left(\int_0^\infty \left(1 - e^{-\Lambda(\beta)(1+\epsilon)x}\right)^n e^{-\lambda(1-\epsilon)x}dx\right)$$
$$+ O(e^{-\xi n}).$$

Computing the integrated in the preceding inequality, we obtain

$$\liminf_{n\to\infty}\frac{\log \mathbb{P}[N_f > n]}{\log n} \leq -\frac{\lambda(1-\epsilon)}{\Lambda(\beta)(1+\epsilon)},$$

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

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

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

where $\lim_{t\to\infty} \log \mathbb{P}[N_f > t/(h \log t)]/\log t = \lambda/\Lambda(\beta)$, and $\mathbb{P}[L_c > h \log t] = o\left(\mathbb{P}[N_f > t/(h \log t)]\right)$ for h large enough.

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

$$\mathbb{P}[T_f > t] \geq \mathbb{P}[N_f L_c > t, l_1 < L_c < l_2] \\ \geq \mathbb{P}[N_f > t/l_1]\mathbb{P}[l_1 < L_c < l_2],$$

since this part is standard, we present the details in the technical report due to the limited space.

2) In this part, we prove the equation (4). Define $N(l)$ to be the number of retransmissions for a packet of length l over the channel $\{X_i\}$, and we obtain

$$\mathbb{P}[T_f > t] = \sum_{l=\lfloor l_{\min}\rfloor}^{\infty} \mathbb{P}[T > t, L_c = l] \\ \leq \sum_{l=\lfloor l_{\min}\rfloor}^{\lceil t\rceil} \mathbb{P}[T > t, L_c = l] + \mathbb{P}[L_c > t] \\ \leq \sum_{l=\lfloor l_{\min}\rfloor}^{\lceil t\rceil} \mathbb{P}\left[N(l) > \frac{t}{l}\right]\mathbb{P}[L_c \geq l] + \mathbb{P}[L_c > t]. \quad (15)$$

Noting

$$\left\{N(l) > \frac{t}{l}\right\} = \bigcap_{j=1}^{\lceil t/l\rceil}\left\{\sum_{i=1}^{l} X_{(j-1)l+i} \leq \beta l\right\}$$

and using the same approach as the proof of Equation (3) to decouple the dependencies, we obtain, by (10), for $\min(\lambda, \Lambda(\beta)) > \epsilon > 0$, $l > l_\epsilon$,

$$\mathbb{P}\left[N(l) > \frac{t}{l}\right] \leq \prod_{j=1}^{\lceil t/l\rceil - 1} e^{-(\Lambda(\beta)-\epsilon)l} \leq e^{-(\Lambda(\beta)-\epsilon)t},$$

implying

$$P_1 \leq \sum_{l=\lfloor l_{\min}\rfloor}^{\lceil t\rceil} e^{-(\Lambda(\beta)-\epsilon)t} e^{-(\lambda-\epsilon)l} \leq O\left(e^{-(\Lambda(\beta)-\epsilon)t}\right).$$

Combining the preceding inequality, the fact that $\lim_{t\to\infty} \log \mathbb{P}[T > t]/t = -\lambda$, and recalling (15), we finish the proof of the upper bound for (4). The lower bound follows by noting that

$$\mathbb{P}[T_f > t] \geq \max\left\{\max_l \left(\mathbb{P}\left[N(l) > \frac{t}{l}\right]\mathbb{P}[L_c = l]\right), \\ \mathbb{P}[L_c > t]\right\}.$$

■

B. Proof of Theorem 2

Since the proof of Equation (6) is relatively easy, we present it in [12] due to limited space. Now, we focus on the proof of (7). Let E_j denote the number of erasure bits immediately after the j'th retransmission; let $E_0 = L_c$. To ease our presentation, we let $Y_{ji} \in \{0, 1\}$ be sequences of i.i.d. Bernoulli processes with parameter $1 - \gamma$. Let $N(k)$ be the number of transmissions to successfully transmit a codeword of length k.

We begin with the proof of the upper bound. For $c > l_{\min}$, we obtain

$$\mathbb{P}[T > t] \leq \sum_{l=\lfloor t/c\rfloor}^{\lfloor t\rfloor} \mathbb{P}[T_m > t, L_c = l] \\ + \mathbb{P}[T_m > t, c \leq L_c \leq t/c] \\ + \sum_{l=l_{\min}}^{\lfloor c\rfloor} \mathbb{P}[T_m > t, L_c = l] + \mathbb{P}[L_c > t] \\ \triangleq P_1(t) + P_2(t) + P_3(t) + O\left(e^{-(1-\epsilon)\lambda t}\right). \quad (16)$$

Now, for $\epsilon > 0$ and a sequence $\{\beta_j\}_{j\geq 1}, 0 < \beta_j < 1$, we obtain, by the i.i.d. assumption of $\{X_i\}_{i\geq 1}$,

$$P_1(t) \leq \sum_{l=\lfloor t/c\rfloor}^{\lfloor t\rfloor} \mathbb{P}\left[N(l) > \frac{t}{l}\right] e^{-\lambda(1-\epsilon)l} \\ \leq \sum_{l=\lfloor t/c\rfloor}^{\lfloor t\rfloor} \max_{\prod_{j=1}^{\lfloor t/l\rfloor}\beta_j \geq 1-\beta} \prod_{j=1}^{\lfloor t/l\rfloor} \mathbb{P}\Bigg[\beta_j(1-\epsilon)E_{j-1} < \\ \sum_{i=1}^{E_{j-1}} Y_{ji} < \beta_j(1+\epsilon)E_{j-1}\Bigg] e^{-\lambda(1-\epsilon)l} \\ \leq \int_1^{c+\epsilon} \max_{\prod_{j=1}^{\lfloor y\rfloor}\beta_j \geq 1-\beta} \prod_{j=1}^{\lfloor y\rfloor} \mathbb{P}\Bigg[\beta_j(1-\epsilon)E_{j-1} < \\ \sum_{i=1}^{E_{j-1}} Y_{ji} < \beta_j(1+\epsilon)E_{j-1}\Bigg] e^{-\lambda(1-\epsilon)t/y} dy. \quad (17)$$

For $\Lambda^c(x)$ in (5), it is easy to check that, there exists $0 < \zeta < \infty$ such that

$$\Lambda^c(x) - \zeta\epsilon \leq \inf_{x(1-\epsilon)<y<x(1+\epsilon)} \Lambda^c(y) \leq \Lambda^c(x).$$

Recalling $\nu_j = \prod_{j=1}^{i} \beta_j, i \geq 1$ with $\nu_0 = 1$. The inequality (17), by Cramér's Theorem [13], implies

$$P_1(t) \leq (c+\epsilon) \max_{1\leq y\leq c+\epsilon} \\ \max_{\prod_{j=1}^{\lfloor y\rfloor}\beta_j \geq 1-\beta} \prod_{j=1}^{\lfloor y\rfloor} e^{-(\Lambda^c(\beta_j)-\zeta\epsilon)\nu_{j-1}(1-\epsilon)^{j-1}\frac{t}{y}} e^{-\frac{\lambda(1-\epsilon)t}{y}} \\ \leq (c+\epsilon) \max_{1\leq y\leq c+\epsilon} \max_{\prod_{j=1}^{\lfloor y\rfloor}\beta_j \geq 1-\beta} \\ e^{-y^{-1}\left(\sum_{j=1}^{\lfloor y\rfloor}((\Lambda^c(\beta_j)-\zeta\epsilon)\nu_{j-1}(1-\epsilon)^c)+\lambda(1-\epsilon)\right)t}. \quad (18)$$

Next, we evaluate $P_2(t)$. Let the number of j's ($j \geq 1$) with $E_j \leq E_{j-1}(1-\epsilon)$ be equal to n_ϵ. Note that before the codeword can be decoded, at least $1-\beta$ fraction of bits have erasure errors. Therefore, $(1-\epsilon)^{n_\epsilon} > 1-\beta$, implying $n_\epsilon \leq \log(1-\beta)/\log(1-\epsilon)$. Thus, choosing $c = \log(1-\beta)/(\epsilon \log(1-\epsilon))$, we know that $n_\epsilon < \epsilon t/c$ conditional on $\{L_c < t/c\} \cup \{T_m > t\}$. Therefore, for $t > c$ and using Cramér's theorem, we obtain

$$\begin{aligned} P_2(t) &\leq \sum_{l=\lfloor c \rfloor}^{\lfloor t/c \rfloor} \mathbb{P}\left[N(l) > \frac{t}{l}\right] e^{-\lambda(1-\epsilon)l} \\ &\leq \sum_{l=\lfloor c \rfloor}^{\lfloor t/c \rfloor} \mathbb{P}\left[\sum_{i=1}^{(1-\beta)l} (1 - Y_{1i}) < \epsilon l\right]^{\frac{t}{l} - n_\epsilon} e^{-\lambda(1-\epsilon)l} \\ &\leq \sum_{l=\lfloor c \rfloor}^{\lfloor t/c \rfloor} \left(e^{-\Lambda^c\left(1-\frac{\epsilon}{1-\beta}\right)(1-\beta)l}\right)^{(1-\epsilon)\frac{t}{l}} e^{-\lambda(1-\epsilon)l} \\ &= O\left(e^{-\Lambda^c\left(1-\frac{\epsilon}{1-\beta}\right)(1-\epsilon)(1-\beta)t}\right). \end{aligned} \quad (19)$$

Using a similar approach as in proving (19), we can show that

$$P_3(t) \leq O\left(e^{(1-\epsilon)(1-\lfloor \beta l_{\min} \rfloor / l_{\min}) \log(1-\gamma) t}\right). \quad (20)$$

Combining (16), (18), (19), (20), noting that

$$\lim_{\epsilon \to 0} \Lambda^c \left(1 - \epsilon/(1-\beta)\right) = -\log(1-\gamma),$$

using the continuity, and passing $\epsilon \to 0$, yields

$$\begin{aligned} &\limsup_{t\to\infty} \frac{\log P_1(t)}{t} \leq \\ &\quad -\min\Bigg\{ \inf_{\substack{\prod_{j=1}^{\lfloor y \rfloor} \beta_j \geq 1-\beta \\ y \geq 1}} y^{-1} \left(\sum_{j=1}^{\lfloor y \rfloor} \Lambda^c(\beta_j)\nu_{j-1} + \lambda \right), \\ &\qquad -(1-\beta)\log(1-\gamma), \lambda \Bigg\}, \end{aligned}$$

which, by verifying that

$$\lim_{y\to\infty} y^{-1} \left(\sum_{j=1}^{\lfloor y \rfloor} \Lambda^c(\beta_j)\nu_{j-1} + \lambda \right) = -(1-\beta)\log(1-\gamma),$$

implies

$$\limsup_{t\to\infty} \frac{\log \mathbb{P}[T_m > t]}{t} \leq -\min\{\Lambda^\circ, \lambda\}.$$

Due to limited space, the proof of the lower bound, which follows similar arguments as in the proof of the upper bound, is presented in the technical report [12]. □

VII. Conclusion

In this paper, we characterize the performance of coding schemes on mitigating delays in a communication system with retransmissions. We consider an important communication channel called the Binary Erasure Channel, where transmitted bits are either received successfully or lost and focus on the fixed rate coding scheme, where decoding is said to be successful if a certain fraction of the codeword is received correctly. We study two different scenarios: (I) A codeword of length L_c is retransmitted as a unit until the receiver successfully receives more than βL_c bits in the last transmission. (II) All successfully received bits from every (re)transmissions are buffered at the receiver according to their positions in the codeword, and the transmission completes once the received bits become decodable for the first time.

Our studies reveal that there is a clear cost benefit tradeoff between delay and control complexity. We find that either by using a powerful codeword or by designing a system that keeps track of all received information in prior transmissions, the delay due to retransmissions can be shown to decay exponentially fast rather than have a slow power-law decay. These results provide a benchmark to quantify the tradeoffs between coding complexity and transmission throughput for receivers that use memory to buffer (re)transmissions until success and those that do not buffer intermediate transmissions, and could be used as guidelines for network designers.

References

[1] D. P. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.

[2] P. R. Jelenković and J. Tan, "Can retransmissions of superexponential documents cause subexponential delays?" *In Proceedings of IEEE INFOCOM'07*, pp. 892–900, 2007.

[3] ——, "Are end-to-end acknowlegements causing power law delays in large multi-hop networks?" in *14th Informs Applied Probability Conference*, Eindhoven, July 9-11 2007.

[4] ——, "Is ALOHA causing power law delays?" in *Proceedings of the 20th International Teletraffic Congress,* Ottawa, Canada, June 2007*; Lecture Notes in Computer Science,* No 4516, pp. 1149-1160, Springer-Verlag, 2007.

[5] P. Elias, "Coding for two noisy channels," *Information Theory, Third London Symposium*, pp. 61–76, 1955.

[6] A. Shokrollahi, "Raptor codes," in *IEEE Transactions on Information Theory*, 2006, pp. 2551–2567.

[7] T. M. Cover and J. A. Thomas, *Element of Information Theory*. New York: Wiley-Interscience, 1991.

[8] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Transactions on Information Theory*, vol. 47, pp. 569–584, 2001.

[9] M. G. Luby, "Lt codes," 2002.

[10] J. Tan and N. B. Shroff, "Transition from heavy to light tails in retransmission durations," in *INFOCOM'10: Proceedings of the 29th conference on Information communications*. Piscataway, NJ, USA: IEEE Press, 2010, pp. 1334–1342.

[11] J. R. Norris, *Markov Chain*. Cambridge, UK: Cambridge University Press, 1998.

[12] J. Tan, Y. Yang, N. B. Shroff, and H. El-Gamal, "Delay asymptotics with retransmissions and fixed rate codes over erasure channels," Department of Electrical and Computer Engineering, The Ohio State University, Tech. Rep., July 2010. [Online]. Available: http://ece.osu.edu/~tan/

[13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.

[14] A. Schwartz and A. Weiss, *Large Deviations For Performance Analysis: QUEUES, Communication and Computing*. CRC Press, 1995.