The Queued-Code in Finite-State Markov Fading Channels with Large Delay Bounds

Satashu Goel, Rohit Negi
{satashug,negi}@ece.cmu.edu, ECE department, Carnegie Mellon University

***Abstract*— A 'queued-code' is a novel channel code, whose input rate is instantaneously adapted to the channel state information, assumed to be known to the transmitter. Previously, a queued-code was demonstrated for i.i.d. block fading channels, demonstrating significant gains, in terms of error exponent, over a fixed rate code. This paper demonstrates the performance gain of the queued-code in the more realistic case of a Finite-State Markov fading channel. Similar to the block-fading result, the analysis required combines the concept of random coding bounds from information theory, and the concept of effective capacity, arising from large deviation theory analysis of queues. However, a result on the limiting moment generating function of Markov chains is the additional tool required for the analysis.**

I. Introduction

In networking, Quality of Service (QoS) guarantees of data rate, error rate and delay bound for delay-sensitive applications have typically been provided by the analysis of corresponding queuing systems. In [1], a pure queuing model was considered, assuming ideal rate-adaptive channel codes achieving the instantaneous Shannon capacity, where the effect of channel variations on link performance was captured by a single function called 'effective capacity'. While the paper successfully showed that the effective capacity function provides an efficient and compact representation of link performance, the results there will typically be too optimistic for a real system, which must deal with channel noise, using channel codes operating below capacity. Thus, in reality, one needs to consider not only the ideal queuing model (as is common in networking), but also an explicit physical layer model for channel coding (requiring information theory).

Therefore, we introduced the concept of a 'queued-code' (Figure 1) in [3]. The queued-code is fed by a source of *constant rate* μ. The channel is assumed to be time-varying. The key feature of the queued-code is that its instantaneous code rate R_k is equal to the instantaneous rate of a server at its input. The server's rate is decided by the channel state information (CSI), assumed known by the transmitter. Now, if the server has a high R_k, the encoder must choose channel codes with large rates, thus clearing the queue quickly, but resulting in a high decoding error probability. On the other hand, if R_k is small, the decoding error probability will be small, but only at the expense of large (potentially unstable) queuing delays. Thus, there must exist a queued-code that balances the queuing and the channel coding operations optimally, resulting in the best possible QoS. In [3], ideas from the fields of queuing theory and communication/information theory were combined to design such an optimal system. A block-fading channel model [2] was assumed. Assuming that the server rate in block k, $R_k = R(g_k)$ is a function of the instantaneous channel gain g_k, it was shown that a queued-code exists, which achieves an error exponent (with respect to delay D) that is the minimum of, a) the delay bound violation exponent of a certain queuing system, and b) the random coding error exponent of a certain channel coding system. Since these two exponents represent the large-delay asymptotic limit of "QoS" in queuing theory and information theory respectively, the result of that paper had a satisfying interpretation. A criticism of that paper is the use of the block-fading channel model, which is not altogether realistic. The contribution of the present paper is to show that a similar result as in [3] holds for *Finite-State Markov fading channels*. i.e., the queued-code error exponent is equal to the minimum of appropriately defined queuing and random coding exponents.

Note that other researchers have considered either the pure queuing problem [4] or the pure channel coding problem (see references in [5]), for wireless time-varying channels. Past research that has considered queuing and channel coding jointly assumes that the channel code is restricted to a large block of constant channel gain (so that efficient rate-adaptive codes can be used) [6]. But the channel codes in this approach are sub-optimal, in that they span a block whose length is determined by the coherence time T_c of the channel, rather than the entire delay D allowed by the application, which could be substantially larger than T_c. In contrast, our queued-code spans the entire delay D allowed by the application, as well as allows rate adaptation within this delay. Interesting combinations of information theory and queuing for multiple access was presented in [7]. Several papers have explored trade-offs between power/throughput and delay [8], [9], [10], [11]. However, each of these papers assumes that capacity achieving channel codes can be used; in effect, allowing coding only within each block of a block-fading channel.

II. Problem Formulation

Denote $\{a_0, a_1, \ldots, a_k\}$ by a_0^k and expectation over $\mathbf{g}$ by $\mathop{\mathbf{E}}_{\mathbf{g}}[\,\cdot\,]$. We begin by formally describing the queued-code, shown in Figure 1. The discrete-time memoryless channel (DMC) model [12] is specified by the conditional probability distribution $p(y_k|g_k, x_k),\ k = 0, 1, \ldots$, where x_k, y_k are the channel input symbol and output sample respectively, at time k, while g_k is a channel parameter (e.g., gain) at time k. g_k is assumed known to both, the transmitter and receiver. In this paper, as opposed to the block-fading model in [3], we consider the realistic case where $\{g_k\}$ form a Hidden Markov chain with a finite number of states. i.e., $P(g_0^k|S_0^k) = \prod_{i=0}^{k} P(g_i|S_i)$, where the sequence $\{S_k\}$ is an *irreducible* Markov chain with states in $\mathcal{S} = \{1, 2, \ldots, |\mathcal{S}|\}$.

Notice that the transmitter, which consists of a queue+server followed by an encoder, combines the canonical models of queuing and information theory. We assume that the server chooses an 'instantaneous server capacity' $R_k = R(g_k)$ nats, where $R(g)$ is some function. (In the rest of the paper, 'optimality' will refer to this memoryless-server assumption.) The encoder receives bits from the server at rate R_k, encodes them into a sequence of symbols and transmits the symbols. We assume a 'streaming-code' encoder [13] , since a block-encoder [12], although simpler, requires additional delay to buffer the data until a block of data is ready for encoding. Thus, we assume that bits that depart the server at time k are immediately encoded into symbols $x_k, x_{k+1}, \ldots$. The decoder decodes the channel code after buffering the received y_k sufficiently. The system needs to be designed so that the QoS demanded by the source application is met. We formally define QoS as follows. The source has a constant rate μ and a maximum delay bound of D. Given the channel statistics and the code, it is required that the probability of error be as small as possible. The probability of error metric must capture both; delay bound violations due to queuing, as well as decoding errors due to channel noise. This will be specified in the following.

Consider the source bit that arrives at the queue from the source at time $k - D_k$. After spending time D_k in the queue, it departs the server at time k. The encoder encodes the bit(s) into the code symbols $x_k, x_{k+1}, \ldots$. Since the source demands a delay bound of D, the decoder must decode this bit at time $k - D_k + D - 1$. Let P_{err} be the probability of bit decoding error. Then, a bit decoding error can occur in two cases; a) the bit spent the entire time D (or more) in the queue, and therefore, never got encoded, or b) the bit was sent within a delay of D, but was decoded incorrectly. Thus, P_{err} captures both, the queuing-theoretic notion of error (delay-bound violation) and the information-theoretic notion of decoding error. Therefore, the problem is to design a queued-code which minimizes P_{err} for the given μ, D and channel statistics. Our formulation encompasses a pure coding framework, which does not allow queuing, by setting $R(g) \equiv \mu$, and thus, is expected to perform better than the latter. The main result of this paper appears next.

III. Joint Queuing and Coding

A. Background

The key insight of this paper is as follows. If the encoder is assumed to use 'ideal channel' codes that achieve the instantaneous capacity, then the recently developed effective capacity [1] result can be used to calculate the P_{err} (which is simply the delay bound violation probability now) in the asymptotic regime of $D \to \infty$. Essentially, effective capacity is a pure queuing (i.e., no coding) large-deviations framework, which is the channel dual of the notion of effective bandwidth [4]. To elaborate, for a channel whose instantaneous capacity at time t is $r(t)$ (recall that capacity-achieving codes are assumed in a pure queuing approach), it was shown [1] that if $P_{err} \doteq Pr\{D(\infty) \geq D\}$ (i.e., delay bound D violation probability at steady state), then

$$\lim_{D\to\infty} \frac{-1}{D} \log(P_{err}) = \theta_q(\mu) \tag{1}$$

where $\theta_q(\mu) = \mu\alpha^{-1}(\mu)$ and $\alpha^{-1}(\cdot)$ is the inverse of the effective capacity function (if it exists),

$$\alpha(u) = -\lim_{t\to\infty} \frac{1}{ut} \log \mathbf{E}[e^{-u\sum_{\tau=0}^{t-1} r(\tau)}], \ \forall\ u \geq 0. \tag{2}$$

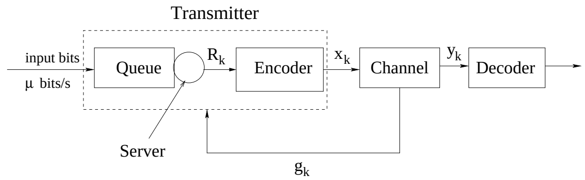

Fig. 1. A 'queued-code'

The effective capacity function exists for a wide range of channel processes; in particular for Markov chains. Interestingly, we will show shortly that the effective capacity function appears as one of the components of QoS in the queued-code.

On the other hand, if a pure coding approach is considered (i.e., no queuing, or server rate $R_k \equiv \mu, \forall\ k$), then the random coding exponent [12] provides a bound on P_{err} (which is simply the decoding error probability now). To elaborate, for a *block code* of rate μ and code length $D/2$, the bit decoding error probability can be bounded as,

$$P_{err} \quad \leq \quad \exp(-(D/2)E_c(\mu)) \tag{3}$$

$$E_c(\mu) = \max_{0\leq\rho\leq 1} \ E_0(\rho,\mu) \quad \text{'Gallager error exponent'} \tag{4}$$

For example, for Gaussian noise and codebooks, and with the assumption of i.i.d. channel gains g_k,

$$E_0(\rho,\mu) = -\log \mathbf{E}_g \exp\left\{-\rho\left(\log\left(1+\frac{gP_0}{1+\rho}\right)-\mu\right)\right\} \tag{5}$$

where P_0 is the ratio of transmit power to noise power. (In Section III-B, we will modify this result to account for the Finite-State Markov fading channel, and the use of a 'streaming code' instead of the pessimistic block code. Notice how $\theta_q(\mu)$ and $E_c(\mu)$ are both exponents that bound P_{err}, using two completely different approaches, which apply respectively, under two different assumptions. Our main contribution is to show that the optimal queued-code must consider both these exponents, to obtain the best possible error exponent for P_{err}.

B. A queued-code exponent

Since we have limited our design to queued-codes which have a server 'instantaneous capacity' $R_k = R(g_k)$ (i.e., memoryless), we need to choose the function $R(g) \geq 0$ correctly. This choice must be made so that P_{err} is minimized for the given μ, D. We work in the regime of asymptotically large D, so that large-deviations queuing and information theoretic coding results can be applied. Thus, we must choose an $R(g)$ that maximizes the 'queued-code exponent' $\theta(\mu)$, where

$$\liminf_{D\to\infty} \frac{-1}{D} \log(P_{err}) = \theta(\mu). \tag{6}$$

In order to calculate $\theta(\mu)$, we need to analyze the combined queuing and coding system. This analysis is done in two stages. In stage 1, for a *fixed decoding delay*, analyze the decoding error probability of the streaming code, which encodes bits received by the encoder at time k, into symbols $x_k, x_{k+1}, \ldots$, possibly encoding a different number of new bits in different symbols (as specified by the server capacity $R_k = R(g_k)$). Note that we use a streaming code, rather than a block code, to avoid the extra delay in the latter, whose encoding delay equals decoding delay. If the specified decoding delay is zero or less, we can upper bound the error probability by one. In stage 2, we analyze the queuing delay D_k caused by the queuing system (due to the chosen $R(g)$),

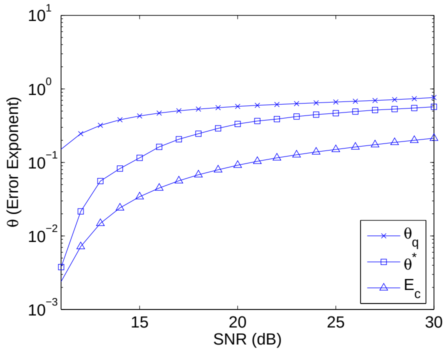

Fig. 2. Error exponents versus SNR

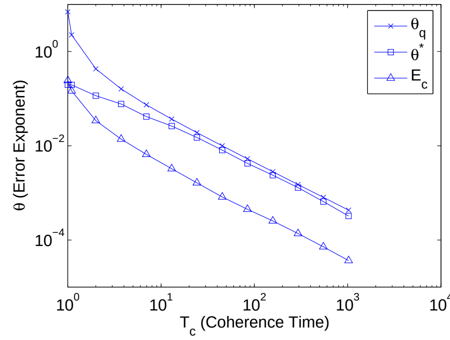

Fig. 3. Error exponents versus coherence time T_c

and thus, obtain the decoding delay $D - D_k$ available for the decoder, for the QoS delay bound to be met. For bits that have $D - D_k \leq 0$, we bound the error probability by one, since these bits do not even get encoded within the allowed delay bound. Thus, the *bit error probability* P_{err} of the code, fed by the queue, can be calculated by combining stages 1 and 2. The analysis results in the following two lemmas.

Lemma 1: For a streaming code with a delay-bound of D, and source rate $R_k = R(g_k)$, the error exponent is given by,

$$\liminf_{D\to\infty} \frac{-1}{D} \log(P_{err}) \geq E_c \tag{7}$$

where E_c is given by (17). Note that this is the same exponent as achieved by a block code, whose block length is equal to D, but without the extra delay of D required to buffer the source data prior to encoding.

Proof: The proof follows [12], adapted to the case of a streaming code, and assuming the specific time-varying source rate. See Appendix A. The difference from [3] is that a result on the asymptotic moment generating function (mgf) of Markov chains is required to handle the Finite-State Markov fading channel. This is the analysis for stage 1. ■

Lemma 2: Consider a queued-code (Figure 1), which has a constant source rate μ, a delay-bound of D, and which employs a streaming code. For a Finite-State Markov fading channel, the error exponent is given by,

$$\liminf_{D\to\infty} \frac{-1}{D} \log(P_{err}) \geq \theta^* \tag{8}$$

where θ^* is given by (29).

Proof: The proof analyzes the queuing delay (stage 2), and then uses Lemma 1 to analyze the queued-code. See Appendix B. The difference from [3] is the use of the effective capacity argument to analyze the queuing delay. ■

Lemma 2 shows that the optimum error exponent for the queued-code is given by the minimum of the respective error exponents $\theta_q(\mu)$ and $E_c(\mu)$ (see (29)). Thus, the queued-code must balance the requirements of queuing (choosing large $R(g)$ so that the queue clears quickly) with the requirements of the encoder (choosing small $R(g)$ so that the code is highly redundant, and therefore, not susceptible to errors due to noise). Thus, the lemma provides a satisfyingly symmetric result that clearly shows the trade-off between queuing and coding. In the next section, we show that the queued-code approach provides a significantly larger error exponent θ^*, compared to a fixed rate code.

IV. Illustrative Results

We compare the queued-code with a pure queuing system (which assumes that the instantaneous capacity can be achieved) and the pure coding system (which sets $R(g) \equiv \mu$, indicating the absence of a queue). Since the queuing system does not use channel coding, if noise is present in the channel, it will be unable to achieve arbitrarily small P_{err}, even at large D. Therefore, we assume that the only source of error for the queuing system is delay-bound violation. Consequently, the error exponent $\theta_q(\mu)$ that we obtain for the queuing system should only be interpreted as an unattainable upper bound on the maximum possible queued-code exponent $\theta^*(\mu)$.

Assume a simple Gaussian channel model, $y_k = \sqrt{g_k} x_k + n_k$, where the noise n_k is Gaussian with variance one, and Gaussian codebooks, transmitted at fixed power P_0 per-symbol, are used. A Gilbert-Elliot channel model [14] is assumed. This model assumes a 2-state Markov chain $\{S_k\}$, with the states labeled $\{G, B\}$. We assume that the probability of state transition is a for both states. Further, we assume that $g = g_G$ and $g = g_B$ for the G, B states respectively. We define coherence time $T_c = 1/a$, which is the average time before a state change. A range of (constant) source rate μ, average signal-to-noise-ratio $SNR = P_0\mathbf{E}[g]$ and T_c are simulated. $\theta_q(\mu), E_c(\mu), \theta^*(\mu)$ are calculated and shown below. Note that $P_{err} \approx \exp(-\theta \cdot D)$ holds for each exponent, respectively.

Figures 2 and 3 show that the pure queuing exponent is an upper bound on both, the queued-code and pure coding exponents. These figures show that the queued-code provides a substantial gain over the pure coding system for a wide range of average SNR and T_c. Figure 2 shows the variation of error exponents with average SNR at $\mu = 1.5$ nats/symbol and transition probability $a = 0.5$. The plot shows that the pure coding and queued-code systems perform much worse than the pure queuing system at low SNR. However, a practical queuing system will be hard hit by noise at low SNR. This fact is not reflected in the θ_q plot, since we (naively) assume that the pure queuing system achieves the instantaneous capacity. Figure 3 shows the variation of error exponents with T_c for $\mu = 1.5$ nats/symbol and $SNR = 15$ dB. The performance of the queued-code is close to that of the pure queuing system for large values of T_c, since large T_c allows the code to achieve the instantaneous Shannon capacity. The performance gain of the queued-code over pure coding increases with T_c, showing that rate adaptation is useful in slow fading channels.

V. Conclusion

This paper presented a joint approach to queuing and channel coding, by combining ideas from large deviation theory (effective capacity) and information theory (random coding bounds), which are applicable in the regime of large delay-bounds ($D \to \infty$). This approach aims to find the ultimate (asymptotic) limit of delay-bounded communication, in contrast to other approaches, which attempt such a combination using specific queuing and practical channel coding models, such as rate-adaptive convolutional coding. A queued-code was assumed, where the instantaneous channel code rate is adapted as $R(g)$, a function of current CSI, by the mechanism of a queue and server. This paper extended the analysis of such a queued-code in [3], to Finite-State Markov fading channels. A lower bound on the decay exponent of decoding error probability was calculated for the queued-code, assuming a streaming code. Illustrative results show that the queued-code can provide significantly larger error exponent, over a range of situations, as compared to the pure coding approach.

Acknowledgment

This work was supported by the National Science Foundation under Career award 0347455.

Appendix

A. Proof of Lemma 1

Encoding: The source begins transmitting bits at time 0. We denote the symbol transmitted at time k by x_k and the set of bits that arrive at the encoder over time $0, 1, \ldots, k$ as b_0^k. We assume that $R_k = |b_0^k| - |b_0^{k-1}|$ is a time-varying source rate, which depends solely on the current CSI g_k as $R_k = R(g_k)$. We consider a streaming code where x_k is generated based on b_0^k. Specifically, for each sequence b_0^k, a symbol x_k is chosen according to an arbitrary probability distribution $Q(x)$, independent of $x_{k-1}, x_{k-2}, \ldots, x_0$. Notice that two bit streams will have the same code symbols prior to the time at which *they first differ* from each other, and will have independent symbols from that time onwards. The probability of choosing a particular codeword x_0^{k+D-1} is,

$$Q(x_0^{k+D-1}) = \prod_{i=0}^{k+D-1} Q(x_i). \tag{9}$$

Decoding: We consider ML decoding, assuming that the decoder knows g_0^k at time k. The k^{th} symbol is decoded at time $k + D - 1$ (since a delay constraint of D symbols is assumed). To decode the symbol x_k, all samples y_0^{k+D-1} are utilized. The bit error probability P_{err} is upper-bounded by the probability that the transmitted message b_0^k (i.e., any of the bits up to time k) is decoded incorrectly. We now compute this message error probability, given that message $m = b_0^{k+D-1}$ is transmitted up to time $k+D-1$. Notice that m is assumed to consist of bits up to time $k+D-1$, since those time samples are used for decoding x_k. The initial part of the proof broadly follows [12], except that it is applied to a streaming code rather than to a block code. Defining $\mathbf{g} \doteq g_0^{k+D-1}$ and state sequence $\mathbf{S} \doteq S_0^{k+D-1}$, the message error probability is bounded as [3],

$$P_{e,m}(\mathbf{g}, \mathbf{S}) \leq \sum_{p=0}^{k} \exp\{-\sum_{i=p}^{k+D-1} (E(g_i) - \rho R(g_i))\} \tag{10}$$

where $E(g_i)$ depends on the current $P(y|x, g_i)$ as,

$$E(g_i) = -\log\Big[\sum_y \Big(\sum_x Q(x) P(y|x, g_i)^{1/(1+\rho)}\Big)^{1+\rho}\Big] \tag{11}$$

Unlike [3], the g_i's are not i.i.d. and hence, the expectation calculation for (10) is not straightforward. To compute the expected probability of error, we average over $\mathbf{g}$ and $\mathbf{S}$.

$$\begin{aligned} P_{e,m} &= \mathop{\mathbf{E}}_{\mathbf{S}}[\mathop{\mathbf{E}}_{\mathbf{g}}[P_{e,m}(\mathbf{g}, \mathbf{S})|\mathbf{S}]] & (12)\\ &\stackrel{a}{=} \sum_{p=0}^{k} \mathop{\mathbf{E}}_{\mathbf{S}}[\prod_{i=p}^{k+D-1} \mathop{\mathbf{E}}_{g_i}[\exp\{-(E(g_i) - \rho R(g_i))\}|S_i]] & (13)\\ &= \sum_{p=0}^{k} \mathop{\mathbf{E}}_{\mathbf{S}}[\prod_{i=p}^{k+D-1} t(S_i)], & (14) \end{aligned}$$

where $t(S_i)$ is the expectation over g_i in (13). (a) holds because of (10) and because g_i's are independent, given $\mathbf{S}$. The expectation in (14) can be computed using result (36.13) on characteristic functions in [16], which can be shown to hold for moment generating functions as well. Let P be the state transition matrix of a Markov chain $\{S_k\}$ with $P_{ij} = P(S_{k+1} = j|S_k = i)$, $t(j)$ be an arbitrary function of the state, and the matrix P^* has elements $P^*_{ij} \doteq P_{ij} t(j)$. Then, the expectation in (14) can be written as $\sum_{h=0}^{\beta} A_h \lambda_h^{k+D-p}$, where $\lambda_0, \lambda_1, \ldots, \lambda_\beta$ are the eigenvalues of matrix P^* with multiplicities $m_0, m_1, \ldots, m_\beta$. A_h is a polynomial in $k+D-p$ of degree $m_h - 1$. Therefore, the expectation in (14) can be upper-bounded by $c_1(k+D-1)^{|\mathcal{S}|-1}(\max_h |\lambda_h|)^{k+D-p}$. Thus, computing the sum in (14), we bound P_{err} as,

$$\begin{aligned} P_{err} &\leq c_2 D^{|\mathcal{S}|-1} \exp\{-D \cdot E_c\} & (15)\\ \liminf_{D\to\infty} \frac{-1}{D} \log(P_{err}) &\geq E_c, & (16)\\ \text{where} \quad E_c &= -\log(\max_h |\lambda_h|) & (17) \end{aligned}$$

where c_2 is a constant and assuming that $E_c > 0$.

B. Proof of Lemma 2

Let q_k be the queue length at time k and D_k be the queuing delay of the bit that departs the server at time k. Since the source rate is assumed constant, $D_k = q_k/\mu$ [1]. Since $R_k = R(g_k)$, q_{k+1} can be expanded as [15],

$$\begin{aligned} q_{k+1} &= \max(0, \mu - R_{k+1}, 2\mu - R_{k+1} - R_k,\\ &\qquad 3\mu - R_{k+1} - R_k - R_{k-1}, \ldots) & (18)\\ D_{k+1} &= \max(0, a_{k+1}, a_{k+1} + a_k, \ldots) & (19)\\ \text{where,} &\quad a_k \doteq 1 - R_k/\mu = 1 - R(g_k)/\mu. & (20) \end{aligned}$$

The bit error probability P_{err} can be upper-bounded by the message error probability, which is bounded as (12),

$$\begin{aligned} P_{err} &\leq \mathop{\mathbf{E}}_{\mathbf{S}}[\mathop{\mathbf{E}}_{\mathbf{g}}[\sum_{p=0}^{k} \exp\{-\sum_{i=p}^{N}(E(g_i) - \rho R(g_i))\}|\mathbf{S}]]\\ &\stackrel{a}{=} \sum_{p=0}^{k} \mathop{\mathbf{E}}_{\mathbf{S}}[\mathop{\mathbf{E}}_{g_0^k}[\mathop{\mathbf{E}}_{g_{k+1}^N}[\prod_{i=k+1}^{N} \exp\{-(E(g_i) - \rho R(g_i))\}|g_0^k, \mathbf{S}]\\ &\qquad \prod_{i=p}^{k} \exp\{-(E(g_i) - \rho R(g_i))\}|S_0^k]], \end{aligned} \tag{21}$$

where $N \doteq k + D - D_k - 1$. (a) holds because expectation is a linear operator, and because conditional expectation was used to separate expectation calculation over g_0^k and g_{k+1}^N. Note that given g_0^k, D_k has a fixed value. Also, given $\mathbf{S}$, g_i's are independent random variables. We now separate the expectation calculation over S_0^k and S_{k+1}^N in similar manner,

$$P_{err} \leq \sum_{p=0}^{k} \mathop{\mathbf{E}}_{S_0^k}[\mathop{\mathbf{E}}_{g_0^k}[\mathop{\mathbf{E}}_{S_{k+1}^N}[\prod_{i=k+1}^{N} \mathop{\mathbf{E}}_{g_i}[\exp\{-(E(g_i) - \rho R(g_i))\}|S_i] \\ |g_0^k, S_0^k] \prod_{i=p}^{k} \exp\{-(E(g_i) - \rho R(g_i))\}|S_0^k]]. \quad (22)$$

Note that given S_0^k, g_0^k and S_{k+1}^N are independent, and hence, in (22), the order of expectations over g_0^k and S_{k+1}^N could be reversed. The expectation over S_{k+1}^N in (22) is similar to the one in (14) and it can be represented as [16],

$$\mathop{\mathbf{E}}_{S_{k+1}^N}[\prod_{i=k+1}^{N} t(S_i)|g_0^k, S_0^k] = \sum_h B_h \lambda_h^{D-D_k-1}, \quad (23)$$

where B_h is a polynomial in $D - D_k - 1$ of degree at most $|\mathcal{S}| - 1$. Using the same arguments as in Appendix A, the sum in (23) can be upper bounded by $\tilde{c}_2 D^{|\mathcal{S}|-1} \exp\{-E_c \cdot (D - D_k)\}$, where E_c is defined by (17). We now expand D_k using (19),

$$\exp(D_k E_c) = \exp(E_c \cdot \max(0, a_k, a_k + a_{k-1}, \ldots)) \\ \leq 1 + \sum_{j=0}^{k} \exp(E_c \sum_{i=0}^{j} a_{k-i})$$

Now, substituting for $\exp(D_k E_c)$ in (22), we obtain

$$P_{err} \leq \tilde{c}_2 D^{|\mathcal{S}|-1} \exp\{-E_c D\}\Big\{\Big[\sum_{p=0}^{k} \mathop{\mathbf{E}}_{S_0^k}[\mathop{\mathbf{E}}_{g_0^k}[\prod_{i=p}^{k} \exp\{-(E(g_i) \\ -\rho R(g_i))\}|S_0^k]]\Big] + \Big[\sum_{p=0}^{k}\sum_{j=0}^{k} \mathop{\mathbf{E}}_{S_0^k}[\mathop{\mathbf{E}}_{g_0^k}[\prod_{i=p}^{k} \exp\{-(E(g_i) \\ -\rho R(g_i))\} \prod_{i=k-j}^{k} \exp\{E_c a_i\}|S_0^k]]\Big]\Big\}. \quad (24)$$

The first summation in (24) converges as long as $E_c > 0$. The second summation can be broken into three parts, namely $\sum_{j=0}^{k}\sum_{p=k-j}(\cdot)$, $\sum_{j=0}^{k}\sum_{p>k-j}(\cdot)$ and $\sum_{j=0}^{k}\sum_{p<k-j}(\cdot)$. Of these, the first part can be upper-bounded by

$$\sum_{j=0}^{k} \mathop{\mathbf{E}}_{S_0^k}[\prod_{i=k-j}^{k} \mathop{\mathbf{E}}_{g_i}[\exp\{E_c a_i\}|S_i]] = \sum_{j=0}^{k} \mathop{\mathbf{E}}_{S_0^k}[\prod_{i=k-j}^{k} \tilde{t}(S_i)] \quad (25)$$

The expectation in (25) is of the same form as the one in (14), and so, can be represented as $\sum_h \tilde{A}_h \tilde{\lambda}_h^{j+1}$, where $\{\tilde{\lambda}_h\}$ are the eigenvalues of matrix $\tilde{P}^*$, where the element $\tilde{P}^*_{ij} = P_{ij}\tilde{t}(j)$. The $\tilde{A}_h$ are polynomials of degree at most $|\mathcal{S}| - 1$. Therefore, the sum in (25) converges, as long as $\max_h |\tilde{\lambda}_h| < 1$. It can be shown that this condition is equivalent to $\alpha_R(E_c/\mu) > \mu$, where $\alpha_R(u)$ is the effective capacity function corresponding to $\{R(g_k)\}$ (i.e., $r(\tau) = R(g_\tau)$ in (2)). The summation $\sum_{j=0}^{k}\sum_{p>k-j}(\cdot)$ converges under the same condition. The third part can be shown to converge, using an additional result from [17], which states that the largest eigenvalue of a non-negative irreducible matrix is a strictly increasing function of the elements of the matrix. Therefore, P_{err} in (24) can be bounded as,

$$P_{err} \leq c_5 D^{|\mathcal{S}|-1} \exp\{-E_c D\}, \quad (26)$$

The 'queued-code exponent' $\theta(\mu) \doteq \liminf_{D\to\infty} -\frac{1}{D}\log P_{err} > E_c$ (using (26)). Therefore, a lower bound on the optimum $\theta(\mu)$ is obtained by solving the following optimization problem,

$$\theta^*(\mu) = \max_{R(g)\geq 0} E_c \quad (27)$$

$$\text{s.t.} \quad \alpha_R(E_c/\mu) > \mu. \quad (28)$$

where the max is over $\rho, R(g), Q(x)$. Therefore, we require $E_c < \mu\alpha_R^{-1}(\mu) \doteq \theta_q(\mu)$, since $\alpha_R(u)$ is a decreasing function [1]. Since scaling $R(g)$ up strictly increases θ_q (to an arbitrarily large value) while strictly decreasing E_c (up to the lower bound of zero), the optimization problem (27) can be stated compactly as,

$$\theta^*(\mu) = \max_{R(g)\geq 0} \quad \min\{E_c(\mu), \theta_q(\mu)\} \quad (29)$$

where, $E_c(\mu)$ is given by (17), while $\theta_q(\mu) = \mu\alpha_R^{-1}(\mu)$.

References

[1] Dapeng Wu, R. Negi, "Effective capacity: a wireless link model for support of QoS," *IEEE Trans. Wireless Comm.*, pp. 630-643, July 2003.

[2] R. J. McEliece, W. E. Stark, "Channel with block interference," *IEEE Trans. Info. Theory*, vol. 30, pp. 44-53, Jan. 1984.

[3] R. Negi, and S. Goel, "An information-theoretic approach to queuing in wireless channels with large delay bounds," *Proc. IEEE Globecom*, pp. 116-122, Dallas, USA, Dec. 2004.

[4] G. Kesidis, J. Walrand, and C. Chang, "Effective bandwidths for multiclass markov fluids and other ATM sources," *IEEE/ACM Trans. Networking*, vol. 1, no. 4, pp. 424–428, Aug. 1993.

[5] R. Negi, "Power adaptation strategies for delay constrained channels," *Ph.D. thesis*, Stanford University, 2000.

[6] J. Hagenauer, N. Seshadri, and C. Sundberg, "The performance of rate-compatible punctured convolutional codes for digital mobile radio," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 966–980, July 1990.

[7] I. E. Telatar and R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, pp. 963–969, Aug. 1995.

[8] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Info. Theory*, vol. 48, pp. 1135-1149, May 2002.

[9] D. Rajan, A. Sabharwal and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Info. Theory*, vol. 50, pp. 125-144, January 2004.

[10] E. M. Yeh and A. S. Cohen, "Information theory, queueing, and resource allocation in multi-user fading communications," *Proc. CISS*, 2004.

[11] R. Negi, and J. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Trans. Info. Theory*, vol. 48, pp. 2478-2494, Sept. 2002.

[12] R. G. Gallager, "Information Theory and Reliable Communication," John Wiley and Sons, 1968.

[13] E. Martinian, G. W. Wornell, "Universal Codes For Minimizing Per-User Delay on Streaming Broadcast Channels," *Proc. Allerton Conf.* , 2003.

[14] E. Gilbert, "Capacity of a burst noise channel," *Bell Labs Tech. Journal*, vol. 39, pp. 1253-1265, 1960.

[15] L. Kleinrock, "Queuing systems: Vol. I," John Wiley, New York, 1976.

[16] V. I. Romanovsky, "Discrete Markov Chains," Wolters-Noordhoff Publishing, Groningen, Netherlands, 1970.

[17] R. G. Gallager, "Discrete Stochastic Processes," Kluwer Academic Publishers, 1996.