Impacts of Channel Variability on Link-Level Throughput in Wireless Networks

Can Emre Koksal, Kyle Jamieson*, Emre Telatar, and Patrick Thiran
Department of Communication and Computer Sciences, EPFL
1015 Lausanne, Switzerland
{emre.koksal, emre.telatar, patrick.thiran}@epfl.ch, jamieson@csail.mit.edu

ABSTRACT

We study analytically and experimentally the throughput of the packetized time-varying discrete erasure channel with feedback, which closely captures the behavior of many practical physical layers. We observe that the channel variability at different time scales affects the link-level throughput positively or negatively depending on its time scale. We show that the increased variability in the channel at a time scale smaller than a single packet increases the link-level throughput, whereas the variability at a time scale longer than a single packet reduces it. We express the throughput as a function of the number of transmissions per packet and evaluate it as in terms of the cumulants of the samples of the stochastic processes, which model the channel. We also illustrate our results experimentally using mote radios.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design.

General Terms

Theory, performance, experimentation.

Keywords

Link estimation, channel variability, channel modelling.

1. INTRODUCTION

In this paper[1] we study channel variability and quantify the impact of channel variability on the link-level throughput.

*MIT Computer Science and Artificial Intelligence Laboratory, MIT, 32 Vassar St., Cambridge, MA 02139.

[1]The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

SIGMetrics/Performance'06, June 26–30, 2006, Saint Malo, France.

We use the time-varying discrete binary erasure channel with feedback (described in Section 3) which is simple, insightful and yet it closely captures the behavior of systems that rely on variable rate coding in the form of automatic repeat request (ARQ). ARQ based systems rely on the presence of error detection for each packet received. When an error is detected in a received packet, a request is sent back for the transmitter to retransmit the packet. Hence, each packet may be repeated multiple times until it is decoded correctly.

We describe the number of transmissions per packet as a stochastic process and study the impact of variability of the channel parameters on this stochastic process. Somewhat surprisingly, we show that the effect of channel variability at different time scales can have quite different effects and they can even work in opposite ways. Indeed, in Section 4 we show that the variability at the time scale of a single packet transmission reduces the number of transmissions per packet (probabilistically), whereas the number of transmissions per packet is increased by variability in time scales larger than the time it takes to successfully transmit a packet.

There has been a large number of studies on the throughput of ARQ based systems with channel variability. Most of these studies focus on either calculating the throughput of certain ARQ systems under different models of fading (e.g., [17], [12], [2], [21], [16]) or developing ARQ based codes for possibly variable channels (e.g., [20], [8], [3]). Our work is more general in the sense that, we study the channel variability in terms of fundamental physical layer quantities such as the channel erasure probability. We do not focus on the variability caused by a specific type of fading or interference. Rather, we use a model for which our results hold under all sources of variability. Our expressions for the throughput illustrate a clear separation of time scales over which the variability has a positive or a negative effect on the throughput. Moreover, we support our insights in Section 5 with experimental data.

Although the main purpose of this paper is to build some basic understanding on channel variability and its consequences, our results do have direct practical implications. Indeed in the next section and in Conclusions, we elaborate on how our results can be used to construct and/or improve a wide variety of algorithms in wireless networks.

2. MOTIVATION

Since the quality of wireless communication depends on many different parameters, it can vary dramatically over time and with even slight environmental changes.

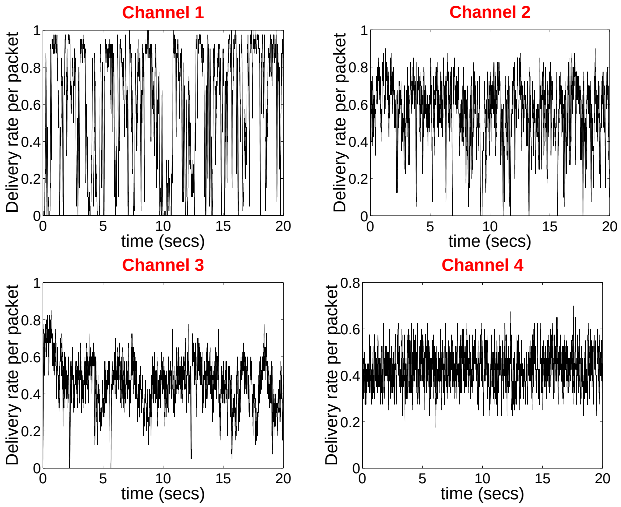


Figure 1: Packet delivery rate for four distinct links of a mesh network.

Examples of sources for channel variability include multipath propagation, mobility and time-varying multiuser interference. Different sources cause variability spanning a large window of time scales from bits to thousands of packets. For instance, relative movement of the transmitter-receiver pair may cause variability at a long time scale since a very large number of packets can be transmitted during the time it takes for the stations to move long enough for the channel to vary significantly. On the other hand, the interference caused by other concurrent transmissions may change significantly from one packet transmission to another. Also, the multipath nature of the propagation medium may cause fast and/or slow fading[2] in the channel.

For instance, consider Fig. 1 which illustrates the packet delivery ratios taken from four distinct links in a mesh network. Each node in the network has an 802.11b wireless card and an antenna. The transmission rate is set to a constant 11 Mbps. The delivery ratios were obtained by sending a sequence of 1500-byte broadcast packets, keeping track of which packets were received successfully. The successful receipt or loss of a packet defines a binary random variable; each sample delivery ratio in the graphs is the average of a window of 40 successive binary random variables. The window advances by 1 for each reported sample.

One can observe significant variability in all the channels in Fig. 1. Moreover, the characteristics of this variability can be quite different from one channel to another. This kind of variability and heterogeneity is common to almost all wireless communications.

In this paper we analyze channel variability and quantify its impact on the channel quality. This enables us to choose between channels, which ultimately makes it possible to enhance the performance of many networking tasks such as routing and multiple access. Now we briefly discuss how this is possible, but we do not address these networking issues in this paper.

In multihop wireless networks, messages are routed from a source to a destination through possibly multiple other terminals acting as relays. A path is the set of links (i.e., channels) between the source and the destination. If we can pick paths with "better" channels, messages can be transmitted much more efficiently consuming much less energy for transmission. This kind of routing is known as *quality aware routing* (see e.g., [4], [18], [19], [1]).

Similarly, multiple access is an important problem in all wireless networks. Suppose multiple terminals try to access to a single terminal. If the receiving terminal can choose between the contending access terminals, it can choose the one with the "best" channel so that the overall network throughput increases and networking can be done more efficiently. This kind of multiple access protocols are known as *opportunistic scheduling* protocols (see e.g., [11], [15]).

The general approach in using the channel state in both the quality aware routing and the opportunistic scheduling literature is illustrated in Fig. 2. A channel parameter such as the delivery rate is measured and a moving average of the measurements is taken. The averaging period is adjusted in such a way that the time scale in which significant variations occur in this moving average is larger than the time scale that the networking algorithms can respond. For instance if a routing algorithm updates its optimal paths between pairs of terminals once every 10 seconds, the averaging period is picked to be comparable to 10 seconds so that the first component in Fig. 2 does not vary significantly in that time scale. While the first component is used in network algorithms, the variations that occur in smaller time scales have been completely disregarded.

One can realize that with such an approach, the networking protocols would not be able to distinguish between the four channels plotted in Fig. 1. Each one of these channels is as good as any other. In this paper, we question this limitation which is a by-product of the long term averaging approach in measuring channels. We quantify the variability so as to be able to compare different channels such as the ones given in Fig. 1 with possibly same average behavior and different variability and analyze the relative importance of variability with respect to the average channel behavior.

The channel model we choose captures the essential characteristics of an ARQ based transmission, which is widely adopted in practice. Indeed, almost all of the coding in most 802.11 systems is based on ARQ. In some systems, the packets that contain errors are completely discarded and the

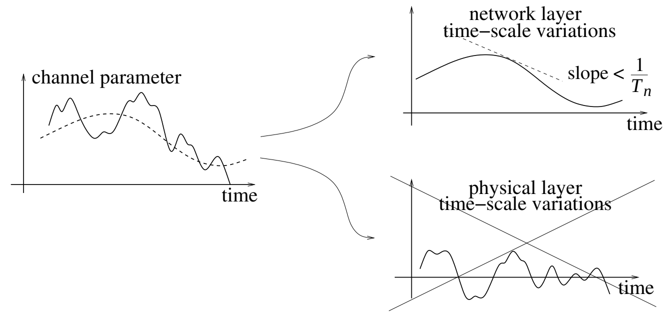


Figure 2: The channel parameter is decomposed into a slowly varying and a fast varying component. Algorithms generally use the former which varies slower than the time scale, T_n, that the networking algorithms respond.

[2]Note that, in the rest of the paper we assume all fading to be frequency flat.

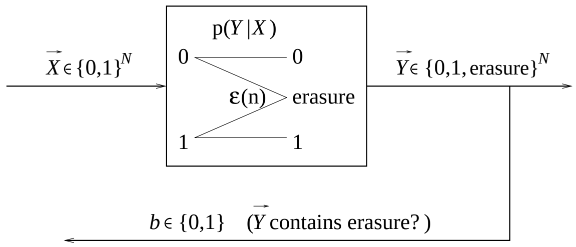


Figure 3: Time varying binary erasure channel with instantaneous erasure probability, $\epsilon(n)$. The feedback, b is a single bit indicator whether $\vec{Y}$ contains erasures or not.

retransmitted copy is treated separately. Such systems are also known as *plain* ARQ systems. Others may choose to keep erroneous packets to be repaired using special procedures that *combine* successive copies of the packet. Such systems are also referred to as *hybrid* ARQ systems (see [10] for instance). In this paper we focus on plain ARQ systems and only briefly mention some results for the hybrid ARQ systems.

Our throughput metric is the number of transmissions per packet, which we believe is a relevant parameter for many networking tasks in the sense that it can be directly translated into network and application level quality constraints. For example, the power consumption is closely connected to the number of transmissions. Indeed (if there is no power control in the system), the energy consumed per bit transmission is directly proportional to the number of transmissions. Another example is the TCP throughput. The retransmission process is not infinitely persistent at the link layer. A packet is dropped and marked lost if it does not get through within a certain number of attempts. For instance in 802.11 based systems, this number is 16. If a packet is lost at a link, then it has to be retransmitted by the higher layer protocol (if it is a reliable protocol such as TCP). Such a retransmission not only reduces the higher layer throughput (since the TCP window size will be cut by half) but the same packet has to travel over the entire path from the origin to the destination. Thus, number of transmissions at each link is closely tied with the performance of the entire network.

3. SYSTEM SETTING AND ASSUMPTIONS

We consider the time-varying discrete binary erasure channel model shown in Fig. 3. Each input symbol is selected from the binary alphabet $\{0, 1\}$ and each output symbol is an element of the set $\{0, 1, \text{erasure}\}$. An erasure indicates that it is ambiguous at the receiver whether the transmitted bit is a 0 or a 1.

We assume that the input symbols are transmitted in groups of N bits. A binary feedback, b, is sent back to the transmitter for every group of N symbols received: $b = 1$ if a group contains erasures and $b = 0$ otherwise. Thus, $b = 0$ means all the bits in the group is decoded successfully. We assume the entire group is retransmitted if $b = 1$.

The above model identifies the behavior of an ARQ based system very closely. Generally, such systems count on the presence of error detection. If a packet contains errors, it is asked to be retransmitted by the transmitter. Note that in most of the systems currently built (such as 802.11), error correction relies mostly on packet retransmission rather than forward error correction. We believe that the feedback channel model we use is suitable to capture such cases.

In the rest of the paper, we use the following terminology. A *packet* is a group of N bits. Each packet can be transmitted multiple times until it is received erasure-free. We call each of these a *transmission*. Let us define $\vec{X}_{ij}$ and $\vec{Y}_{ij}$ as the N-dimensional vectors that represent the transmitted bits and the received sequence of symbols, respectively for the jth transmission of the ith packet. Hence, $X_{ij}(n)$ and $Y_{ij}(n)$ represent symbol n, $1 \leqslant n \leqslant N$, in these groups of N symbols. Also let the erasure probability be denoted by $\epsilon_{ij}(n) = P(Y_{ij}(n) = \text{erasure})$. We use the vector notation $\vec{\epsilon}_{ij}$ for the sequence of erasure probabilities of packet i transmission j. Due to the somewhat unconventional nature of this model, we find it necessary to emphasize that $\vec{\epsilon}_{ij}$ represents two things:

1. It is a sample outcome of a 2-dimensional random process $\{\vec{\epsilon}_{ij},\ i, j \geqslant 1\}$. Each sample of this process takes a value in $[0, 1]^N$.
2. The $\epsilon_{ij}(n)$ is the probability that an event (erasure) occurs for symbol n of packet i transmission j.

The process $\{\vec{\epsilon}_{ij},\ i, j \geqslant 1\}$ will have a central importance in our analysis. We model the characteristics of channel variability (e.g., due to fading, mobility and multiuser interference) using the statistics of this process. We make certain assumptions on the joint distribution of $\vec{\epsilon}_{ij},\ i, j \geqslant 1$ to analyze the impact of different time scales.

We assume that $\vec{\epsilon}_{ij}$ is independent of the channel inputs for all $i, j \geqslant 1$ and the channel is conditionally memoryless, i.e.,

$$\mathrm{P}\left(\vec{Y}_{ij}|\vec{X}_{ij}, \vec{\epsilon}_{ij}\right) = \prod_{t=1}^{N} \mathrm{P}\left(Y_{ij}(n)|X_{ij}(n), \epsilon_{ij}(n)\right).$$

This, by no means, implies that the channel is memoryless. On the contrary, each vector $\vec{\epsilon}_{ij}$ may contain highly correlated entries since it is likely that a channel does not vary significantly from one bit to the next.

Defining the *erasure parameter* $\eta_{ij}(n) = -\log(1-\epsilon_{ij}(n))$, we can write

$$\mathrm{P}\left(\vec{Y}_{ij} \text{ contains no erasures } \middle| \vec{\epsilon}_{ij}\right) = \prod_{t=1}^{N}(1 - \epsilon_{ij}(n)) \tag{1}$$

$$= \exp\left(-\sum_{t=1}^{N}\eta_{ij}(n)\right). \tag{2}$$

Elementary analysis (see the appendix) lets one conclude that

$$\epsilon_{ij}(n) \leqslant \eta_{ij}(n) \leqslant \epsilon_{ij}(n) + \frac{\epsilon_{ij}^2(n)}{1 - \epsilon_{ij}^2(n)}. \tag{3}$$

Thus, if the individual bit erasure probability is small[3] for some n, i.e., $\epsilon_{ij}^2(n) \ll \epsilon_{ij}(n)$ with high probability, then

[3]Unless the individual samples of the erasure probability are small, for large packet sizes it would be very unlikely for a packet to be successfully transmitted even with a large number of attempts. Thus, the expected number of erasures per packet must be reasonably low (e.g., a few) for a channel to be "usable."

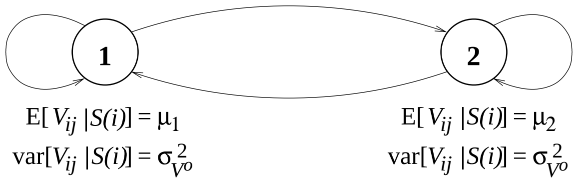


Figure 4: The two-state Markov channel with $S(i) \in \{1,2\}$ for all $i \geqslant 1$. The mean cumulative erasure parameter is different in each state.

$\eta_{ij}(n) \approx \epsilon_{ij}(n)$. For instance, if $\epsilon_{ij}(n) = 10^{-3}$, then $\eta_{ij}(n) = 1.001 \times 10^{-3}$. It may be helpful and more intuitive for the reader to think the two are replaceable. However, despite its mathematical convenience, this approximation may lead to misinterpretations of some of the results later on. We shall make note of the necessary caveats in Appendix B.

Lastly, we define the *cumulative erasure parameter*, $V_{ij} = \sum_{n=1}^{N} \eta_{ij}(n)$. In the rest of the paper, the cumulative erasure parameter will be the only channel parameter we use to derive our results.

In this paper, we analyze the number, Z_i, of transmissions for a given packet i until it is received erasure-free. Therefore, Z_i is the stopping time for observing the infinite sequence of random variables, $V_{i1}, V_{i2}, \ldots$: Given $Z_i = z$, the first $z-1$ transmissions of packet i contain erasures and the zth transmission is erasure-free. Hence, the conditional probability

$$\mathrm{P}\left(Z_i > z \mid V_{i1}, V_{i2}, \ldots, V_{iz} \right) = \prod_{j=1}^{z} \left[1 - \exp(-V_{ij})\right] \qquad (4)$$

follows from (2), and the complementary distribution function for Z_i can be found by simply taking the expectation of (4) over $V_{i1}, V_{i2}, \ldots, V_{iz}$:

$$\mathrm{P}\left(Z_i > z\right) = \mathrm{E}\left[\prod_{j=1}^{z} \left(1 - \exp(-V_{ij})\right)\right]. \qquad (5)$$

To say something interesting about the number of transmissions, we need to make certain assumptions on the joint distribution of these erasure parameters. The assumptions we make shall depend on the time scale we analyze. We study (5) for two different scenarios in which V_{ij} for distinct (i,j) pairs are iid or follow a Markov process[4].

IID $\mathbf{V_{ij}}$: First, we assume that the channel variations are so fast that different transmissions of a packet observe independent erasure probability vector. We show that if $\{V_{ij},\ i,j \geqslant 1\}$ is an iid process (in both i and j), then $\mathrm{E}[Z_1]$ decreases with increasing σ_V^2 for a given μ_V, where μ_V and σ_V^2 are the mean and the variance of V_{ij} for all $i,j \geqslant 1$. Since V_{ij} is approximately the sum of the erasure probabilities over the duration of a packet, we conclude that the variability of the channel at time scales shorter than a single transmission is desirable.

Markov Modulated $\mathbf{V_{ij}}$: We use a Markov analysis to study the impact of variability at time scales longer than a single packet in conjunction with the short time scale variability. We assume that $\{V_{ij},\ i,j \geqslant 1\}$ is modulated by a Markov chain with the set of states $\mathcal{S}$, which contains more than a single recurrent state. For each packet i, the Markov chain makes a state transition (possibly a self transition) to state $S(i) \in \mathcal{S}$, and remains in state $S(i)$ for every transmission of packet i until it is received erasure-free. In a given state, samples of the cumulative erasure parameter are iid[5]. Therefore, conditional on $S(i)$, $V_{i1}, V_{i2}, \ldots$ is an iid process and for a given j, $V_{1j}, V_{2j}, \ldots$ is a Markov modulated process.

We assume that mean of the cumulative erasure probability varies and that all the other moments are constant for different states. We use $\mu_{S(i)} = \mathrm{E}\left[V_{ij}|S(i)\right]$ and $\sigma^2_{V^o} = \mathrm{var}\left(V_{ij}|S(i)\right)$ for all $i,j \geqslant 1$. An example with a two-state (i.e., $S(i) \in \{1,2\}$ for all $i \geqslant 1$) Markov channel is illustrated in Fig. 4. This is also an example of the famous Gilbert-Elliott channel model (see [7] and [5]).

We analyze the steady state behavior of this channel, assuming that the initial state, $S(1)$, of the chain is picked according to the steady state distribution of the chain. Hence, for any given packet i, $\mu_{S(i)}$ is a discrete random variable whose value is determined according to these steady state distribution. Since the only change in V_{ij} is the mean, $\mathrm{E}\left[V_{ij}\right] = \mu_{S(i)}$ from one state to another, we can decompose V_{ij} as

$$V_{ij} = \mu_{S(i)} + V^o_{ij}, \qquad (6)$$

where $\mu_{S(i)}$ is constant for a given packet i and V^o_{ij} is a 0 mean iid process for both $i,j \geqslant 1$ and V^o_{ij} is independent of $S(i)$, and therefore independent of $\mu_{S(i)}$. Thus,

$$\mathrm{E}\left[V_{ij}\right] = \mathrm{E}\left[\mu_{S(i)}\right] \qquad (7)$$

and

$$\mathrm{var}\left(V_{ij}\right) = \sigma_{\mu^2_{S(i)}} + \sigma^2_{V^o} \qquad (8)$$

for all $i,j \geqslant 1$. The two components of the "channel variability" $\mathrm{var}\left(V_{ij}\right)$ in (8) capture the long and the short term variability in the given order. Physically, $\sigma_{\mu^2_{S(i)}}$ can be used to model the variability due to shadowing effects, mobility and other short time scale effects, whereas $\sigma^2_{V^o}$ handles the faster fluctuations due to fast fading, multiuser interference, etc.

The analysis of the impact of $\sigma^2_{V^o}$ is identical to the analysis of the iid case in Section 4.1 and it will not be repeated in Section 4.2. We show that $\mathrm{E}\left[Z_1\right]$ increases with increasing $\sigma^2_{\mu_{S(1)}} = \mathrm{var}\left(\mu_{S(1)}\right)$. Since $\mu_{S(i)}$ remains constant for packet i, and varies only for different packets, we conclude that the variability at time scales longer than a single packet is undesirable. Moreover, our expression for $\mathrm{E}\left[Z_1\right]$ captures the impact of the short and the long term variability component simultaneously, since it involves both $\sigma^2_{\mu_{S(1)}}$ and $\sigma^2_{V^o}$. The *separation of time scales* will be apparent in our results very clearly. The following table summarizes the notation we use for the Markov analysis. Note that $V_{ij}|S(i)$ is the cumulative erasure parameter for packet i conditioned on the state of the chain for that packet.

[4]Note that, in practice V_{ij} for distinct (i,j) pairs may or may not be correlated depending on the time elapsed between these transmissions. In general, this duration between two retransmissions may be several round trip propagation times due to the windowing scheme (e.g., selective ACK) employed.

[5]This is a conditional independence given the current state of the Markov chain. In that sense this independence resembles the conditional memorylessness of the time varying erasure channel.

↓ **statistic - variable** →	$\mathbf{V_{ij}\vert S(i)}$	$\mu_{\mathbf{S(i)}}$	$\mathbf{V^o_{ij}}$
mean	$\mu_{S(i)}$	$\mathrm{E}\left[\mu_{S(i)}\right]$	0
variance	$\sigma^2_{V^o}$	$\sigma^2_{\mu_{S(i)}}$	$\sigma^2_{V^o}$

Note that the described Markov model is a bit unnatural since a transition occurs only when the packets are successfully decoded regardless of how long it takes for a packet to be transmitted. We could employ a more realistic model and assume that a state transition occurs at regularly spaced intervals (e.g., once every certain number of transmissions) and could still arrive at the same conclusion. However, the derivation would be unnecessarily complicated and hence the expressions would be much less insightful. Besides, if the self transition probabilities are close to 1, i.e., multiple packets can be transmitted before significant variations occur in the channel, the constraint that the state transition occurs at the end of successful transmissions become less important.

4. VARIABILITY AND THE NUMBER OF TRANSMISSIONS

In this section we analyze the number of transmissions for the two scenarios described in the previous section.

4.1 Short Time Scale Variations

If $\{V_{ij},\ i,j \geqslant 1\}$ is an iid process, then the number of transmissions, $\{Z_i,\ i \geqslant 1\}$ is also iid. Thus, (5) reduces to

$$\mathrm{P}\left(Z_1 > z\right) = \left(1 - \mathrm{E}\left[\exp(-V_{11})\right]\right)^z . \tag{9}$$

Using (9), we can write

$$\begin{aligned}\mathrm{E}\left[Z_1\right] &= \sum_{z=0}^{\infty} \mathrm{P}\left(Z_1 > z\right) \\ &= \sum_{z=0}^{\infty} \left(1 - \mathrm{E}\left[\exp(-V_{11})\right]\right)^z \\ &= \left(\mathrm{E}\left[\exp\left(-V_{11}\right)\right]\right)^{-1} \\ &= g_V^{-1}(-1),\end{aligned} \tag{10}$$

where $g_V(r)$ is the moment generating function of V_{11}. Note that we assume that there exists an $r_{\max} > 0$ such that $g_V(r) < \infty$ for all $r < r_{\max}$.

Let $\Lambda_V(\cdot) = \log g_V(\cdot)$. Since $0 < g_V(r) < \infty$ for all $r < r_{\max}$, $\Lambda_V(r)$ exists and is finite for all $r < r_{\max}$ as well. The Taylor series expansion for $\Lambda_V(r)$,

$$\Lambda_V(r) = \sum_{k=1}^{\infty} \Lambda_V^{(k)}(0) \frac{r^k}{k!}, \tag{11}$$

is also known as the cumulant expansion (the lower limit of the summation is 1 rather than 0 since $\Lambda_V(0) = 0$) and the kth derivative, $\Lambda_V^{(k)}(0)$ is known as the kth cumulant[6] of V. Thus, the log moment generating function is also known as the cumulant generating function. Cumulants of a random variable can be written as functions of the central moments of the random variable.

From Taylor's theorem, there exists an $r_t \in (-1, 0]$ such that

$$\Lambda_V(-1) = -\mu_V + \frac{1}{2}\Lambda_V''\left(r_t\right), \tag{12}$$

where $\mu_V = \Lambda_V'(0) = \mathrm{E}\left[V_{11}\right]$. The second term on the right hand side is known as the Lagrange remainder. Rewriting (10) using (12) we get

$$\mathrm{E}\left[Z_1\right] = \exp\left(\mu_V - \frac{1}{2}\Lambda_V''\left(r_t\right)\right). \tag{13}$$

As expected, the mean number of transmissions, $\mathrm{E}\left[Z_1\right]$, increases with μ_V. Next, we discuss how the second term in the expansion affects $\mathrm{E}\left[Z_1\right]$.

As shown in [6], $\Lambda_V(r)$ is convex for all $r < r_{\max}$. Hence, $\Lambda_V''\left(r_t\right) \geqslant 0$. The inequality sign "$\geqslant$" can be replaced with "$>$" if V is non-atomic, i.e., does not take on a single value with probability 1. We say that the "variability," i.e., the combined effect of all the moments except the mean, has a positive effect since it reduces the expected number of transmissions. To further illustrate this let us give $\mathrm{E}\left[Z_1\right]$ with the third order Lagrange remainder for the expansion of $\Lambda_V(-1)$. For some $r_t' \in (-1, 0)$,

$$\mathrm{E}\left[Z_1\right] = \exp\left(\mu_V - \frac{1}{2}\sigma_V^2 + \frac{1}{6}\Lambda_V'''\left(r_t'\right)\right). \tag{14}$$

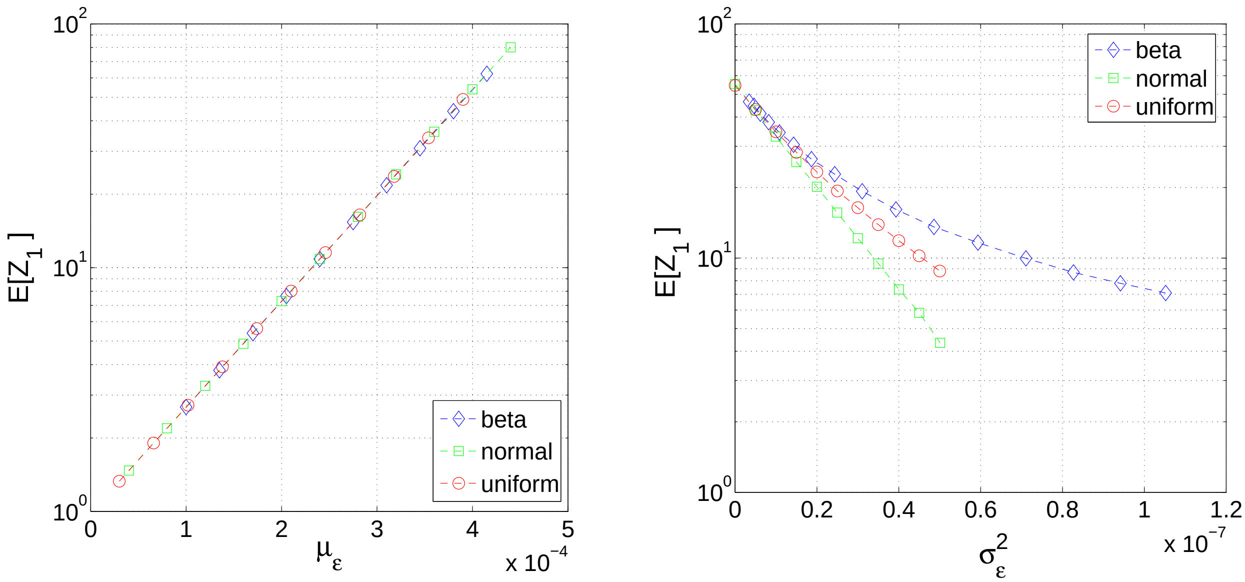
Thus, for a fixed μ_V, $\mathrm{E}\left[Z_1\right]$ decreases with σ_V^2. That is, the more variable the number of erasures per packet, the lower the expected number of transmissions.

In most cases, keeping only the first two terms in the expansion (11) for $\Lambda_V(-1)$ is a fairly good approximation since the variability of the channel is well captured by the variance of V. Hence, $r_t \approx 0$ because $\Lambda_V''(0) = \sigma_V^2$. This can also be called the Gaussian approximation since the higher order cumulants beyond the variance is 0 (i.e., $r_t = 0^-$) if V were $\mathcal{N}(\mu_V, \sigma_V^2)$. Also, if V has a distribution which is symmetric around its mean (e.g., uniform), it has 0 skewness, i.e., $\Lambda_V'''(0) = 0$. The Gaussian approximation works well in such cases as well since the effect of the higher order terms in (11) diminishes quickly.

EXAMPLE 1. *We compute $E\left[Z_1\right]$ under three different distributions for the erasure probability. In each case, we assume that the erasure probability (and hence the erasure parameter) is constant for the duration of each transmission, i.e., $\epsilon_{ij}(n) = \epsilon_{ij}$ where $\{\epsilon_{ij},\ i,j \geqslant 1\}$ is an iid process. Also, $\epsilon_{ij} \ll 1$, so $V_{ij} \approx N\epsilon_{ij}$, where the packet size $N = 10^4$. Based on these assumptions we generate traces of packets, for which ϵ_{ij} is picked according to Gaussian ($\mathcal{N}$), uniform ($\mathcal{U}$) and beta ($\mathcal{B}$) distributions*[7]*. We run two sets of simulations for each distribution to illustrate the mean number of transmissions as a function of $\mu_\epsilon \approx \mu_V/N$ and $\sigma_\epsilon^2 \approx \sigma_V^2/N^2$. $E\left[Z_1\right]$ vs. μ_ϵ: We increase μ_ϵ step by step, keeping every other central moment constant. Hence only the first cumulant changes and all the others remain the same. We can achieve this by just adding a different constant to ϵ_{ij} and hence shifting its distribution at each step by a different amount. In Fig. 5(a), we plot the mean number of transmissions per packet, where a new trace of size 10^6 packets is generated for each value of μ_ϵ. For each distribution, σ_ϵ^2 is*

[6]The first two cumulants are the mean and the variance; and the third and the fourth cumulants are respectively, the normalized versions of the skewness and the kurtosis (excess).

[7]We use $\epsilon_{ij} \sim \mathcal{U}(a, b)$ if ϵ_{ij} is uniform in $[a, b]$. The third (and all the odd numbered) cumulant for the uniform distribution is 0 and if $\epsilon_{ij} \ll 1$, the fourth and the higher order even numbered cumulants are negligible compared to the first and the second cumulants. The beta distribution has two defining parameters, α and β. If $\epsilon_{ij} \sim \mathcal{B}(\alpha, \beta)$, then it takes on values in $[0, 1]$ and $\mu_\epsilon = \frac{\alpha}{\alpha+\beta}$. We use $a + b\mathcal{B}(\alpha, \beta)$ to represent the distribution of $a + b\epsilon_{ij}$



(a) Expected number of transmissions versus the mean

(b) Expected number of transmissions versus the variance

Figure 5: The expected number of transmissions is illustrated as a function of μ_ϵ and σ_ϵ^2 for three different distributions. For the graph 5(a), σ_ϵ^2 is fixed at 3×10^{-10} ($\sigma_V^2 \approx 3 \times 10^{-2}$) for all distributions. Similarly for the graph 5(b), $\mu_\epsilon = 4 \times 10^{-4}$ (4 erasures per transmission) for all three distributions.

fixed at 3×10^{-10}. Thus, $\sigma_V^2 \approx 0.03$, which implies that the channel is not very variable in this set of simulations. The three cases we consider are

- Normal distribution*: $\epsilon_{ij} \sim \mathcal{N}(\mu_\epsilon, 3 \times 10^{-10})$.*
- Uniform distribution*: $\epsilon_{ij} \sim \mathcal{U}(\mu_\epsilon - 3 \times 10^{-5}, \mu_\epsilon + 3 \times 10^{-5}]$.*
- Beta distribution*: $\epsilon_{ij} \sim \mu_\epsilon + 10^{-2}\left(\mathcal{B}(30, 2970) - 10^{-2}\right)$. Note that, $\epsilon_{ij} \in [\mu_\epsilon, \mu_\epsilon + 10^{-2}]$, the third and the fourth cumulants of V are 0.0021 and 2.17×10^{-4} respectively. Thus, in the cumulant expansion for V, μ_V is the dominant term (μ_V varies between 1 and 5).*

As expected, all three curves in Fig. 5(a) grow as $\exp(N\mu_\epsilon) = \exp(\mu_V)$.
$E[Z_1]$ vs. σ_ϵ^2: We keep $\mu_\epsilon \approx \mu_V/N = 4 \times 10^{-4}$ fixed and increase σ_ϵ^2 step by step. In Fig. 5(b), we plot the mean number of transmissions per packet as a function of σ_ϵ^2, where a new trace of size 10^6 packets is generated for each value of σ_ϵ^2. The three cases we consider are

- Normal distribution*: $\epsilon_{ij} \sim \mathcal{N}(4 \times 10^{-4}, \sigma_\epsilon^2)$*
- Uniform distribution*: $\epsilon_{ij} \sim \mathcal{U}[4 \times 10^{-4} - \sqrt{3}\sigma_\epsilon, 4 \times 10^{-4} + \sqrt{3}\sigma_\epsilon]$.*
- Beta distribution*: $\epsilon_{ij} \sim 4 \times 10^{-2}\mathcal{B}(\alpha, \beta)$. We increase α step by step from 1.5 to 46, while keeping $\beta/\alpha = 99$ fixed.*

As expected, the mean number of transmissions is monotonically decreasing with σ_ϵ^2 for all three distributions. However, the rate of change in $E[Z_1]$ is now different among different distributions due to the impact of the higher order cumulants. It is not possible to increase the variance of ϵ_{ij} without increasing the higher order cumulants and they become non-negligible for larger values of σ_ϵ^2. The non-linear increase in these cumulants is the cause for the non-linear decrease in $E[Z_1]$. For instance, the increase in the third cumulant, $\Lambda_V'''(0)$, of V as a function of σ_V^2 is illustrated for the beta distribution in Fig 6.

Lastly, we would like to make some comments about the distribution of Z_1. Since the number, Z_1, of transmissions is geometric, $\mathrm{var}(Z_1) = \mathrm{E}[Z_1] - 1$. Thus, the more variable V_{11} gets, the less variable Z_1 becomes. Also, we can rewrite (9) as

$$\mathrm{P}(Z_1 > z) = \left(\frac{\mathrm{E}[Z_1] - 1}{\mathrm{E}[Z_1]}\right)^z. \qquad (15)$$

For any given z, $\mathrm{P}(Z_1 > z)$ decreases with decreasing $\mathrm{E}[Z_1]$. Therefore, if Z_1' is the number of transmissions with the cumulative erasure parameter V' and $\sigma_{V'}^2 > \sigma_V^2$, then Z_1 stochastically dominates Z_1'. Note that this statement implies that $\mathrm{E}[Z_1]$ increases with decreasing σ_V^2, but the reverse is not necessarily true.

4.2 Long Time Scale Variations

Now we assume that $\{V_{ij},\ i, j \geqslant 1\}$ is the Markov modulated process described in Section 3. Recall that we decom-

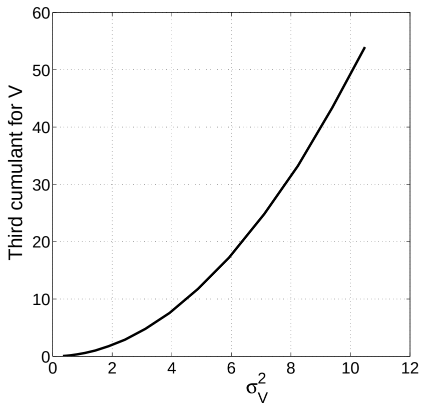


Figure 6: The third cumulant of V is given as a function of σ_V^2 for the beta distribution.

posed V_{ij} into two independent components as

$$V_{ij} = \mu_{S(i)} + V_{ij}^o$$

where $\mu_{S(i)}$ is constant for a given packet i and V_{ij}^o is a 0 mean iid process for $i, j \geqslant 1$. The initial state of the Markov chain is picked at random according to the steady state distribution of the chain. In this section, we show that the variability (combined effect of all the cumulants) of $\mu_{S(1)}$ increases the expected number, $\mathrm{E}[Z_1]$, of transmissions per packet. Then, we express $\mathrm{E}[Z_1]$ in terms of $\sigma^2_{\mu_{S(1)}}$ and $\sigma^2_{V^o}$ and illustrate that these two terms have opposing effects on $\mathrm{E}[Z_1]$. Let

$$\Lambda_{V^o}(r) = \log \mathrm{E}\left[\exp(rV_{ij}^o)\right].$$

Since $\mathrm{E}\left[V_{ij}^o\right] = 0$, $\Lambda'_{V^o}(0) = 0$. Let us define

$$\begin{aligned}\Lambda_{V|S(i)}(r) &= \log \mathrm{E}\left[\exp(rV_{ij})|S(i)\right] \\ &= r\mu_{S(i)} + \Lambda_{V^o}(r). \end{aligned} \tag{16}$$

Note that $\Lambda'_{V|S(i)}(0) = \mu_{S(i)}$ and $\Lambda^{(k)}_{V|S(i)}(r) = \Lambda^{(k)}_{V^o}(r)$ for all $k \geqslant 2$. Since, $\{V_{1j}|S(1),\ j \geqslant 1\}$ is an iid process we can use (13) to evaluate the conditional expected number of transmissions, $\mathrm{E}[Z_1|S(1)]$, for the first packet given the initial state. Hence, there exists an $r_t \in (-1, 0)$ such that

$$\begin{aligned}\mathrm{E}[Z_1] &= \mathrm{E}\left[\mathrm{E}\left[Z_1|S(1)\right]\right] \\ &= \mathrm{E}\left[\exp\left(\mu_{S(1)} - \frac{1}{2}\Lambda''_{V^o}(r_t)\right)\right] \qquad (17) \\ &= \mathrm{E}\left[\exp\left(\mu_{S(1)}\right)\right] \cdot \exp\left(-\frac{1}{2}\Lambda''_{V^o}(r_t)\right). \qquad (18)\end{aligned}$$

Let us define the log moment generating function of $\mu_{S(1)}$,

$$\Lambda_{\mu_{S(1)}}(r) = \log \mathrm{E}\left[\exp\left(r\mu_{S(1)}\right)\right],$$

so that we can write

$$\mathrm{E}\left[\exp\left(\mu_{S(1)}\right)\right] = \exp\left(\Lambda_{\mu_{S(1)}}(1)\right).$$

Expanding $\Lambda_{\mu_{S(1)}}$ up to the second term and using Taylor's theorem, we can show that there exists an $r_w \in (0, 1)$ for which (17) becomes

$$\begin{aligned}\mathrm{E}[Z_1] &= \exp\left(\Lambda_{\mu_{S(1)}}(1) - \frac{1}{2}\Lambda''_{V^o}(r_t)\right) \\ &= \exp\left(\mathrm{E}\left[\mu_{S(1)}\right] + \frac{1}{2}\left[\Lambda''_{\mu_{S(1)}}(r_w) - \Lambda''_{V^o}(r_t)\right]\right). \end{aligned} \tag{19}$$

As the chain has multiple recurrent states, $\Lambda_{\mu_{S(1)}}(r)$ is strictly convex and $\Lambda''_{\mu_{S(1)}}(r_w) > 0$. Since this term represents the impact of $\mu_{S(1)}$ on $\mathrm{E}[Z_1]$ beyond that of $\mathrm{E}\left[\mu_{S(1)}\right]$, we conclude from Eq. (19) that the channel variability in time scales larger than a single packet (including retransmissions) increases the expected number of transmissions.

Next, we express $\mathrm{E}[Z_1]$ using the third order Lagrange remainder for the expansion of both $\Lambda_{V^o}(-1)$ and $\Lambda_{\mu_1}(1)$. For some $r'_t \in (-1, 0)$ and $r'_w \in (0, 1)$,

$$\begin{aligned}\mathrm{E}[Z_1] = \exp\Bigg(&\mathrm{E}\left[\mu_{S(1)}\right] + \frac{1}{2}(\sigma^2_{\mu_{S(1)}} - \sigma^2_{V^o}) \\ &+ \frac{1}{6}\left[\Lambda'''_{\mu_{S(1)}}(r'_w) + \Lambda'''_{V^o}(r'_t)\right]\Bigg), \end{aligned} \tag{20}$$

where $\sigma^2_{\mu_{S(1)}} = \mathrm{var}\left(\mu_{S(1)}\right)$. In Eq. (20), we observe the *separation of time scales* very clearly. For a fixed $\mathrm{E}\left[\mu_{S(1)}\right]$, the expected number of transmissions increases with increasing $\sigma^2_{\mu_{S(1)}}$, and decreasing $\sigma^2_{V^o}$ which represent, respectively, the long term and the short term channel variability. Recall that $\sigma^2_{\mu_{S(1)}} + \sigma^2_{V^o} = \mathrm{var}(V_{ij})$. We just showed how the cumulative erasure parameter can be decomposed into two components with variations at disjoint time scales so that the variability in these time scales have an exact opposite (equal magnitude in the opposite directions) impact on the log of the expected number of transmissions per packet.

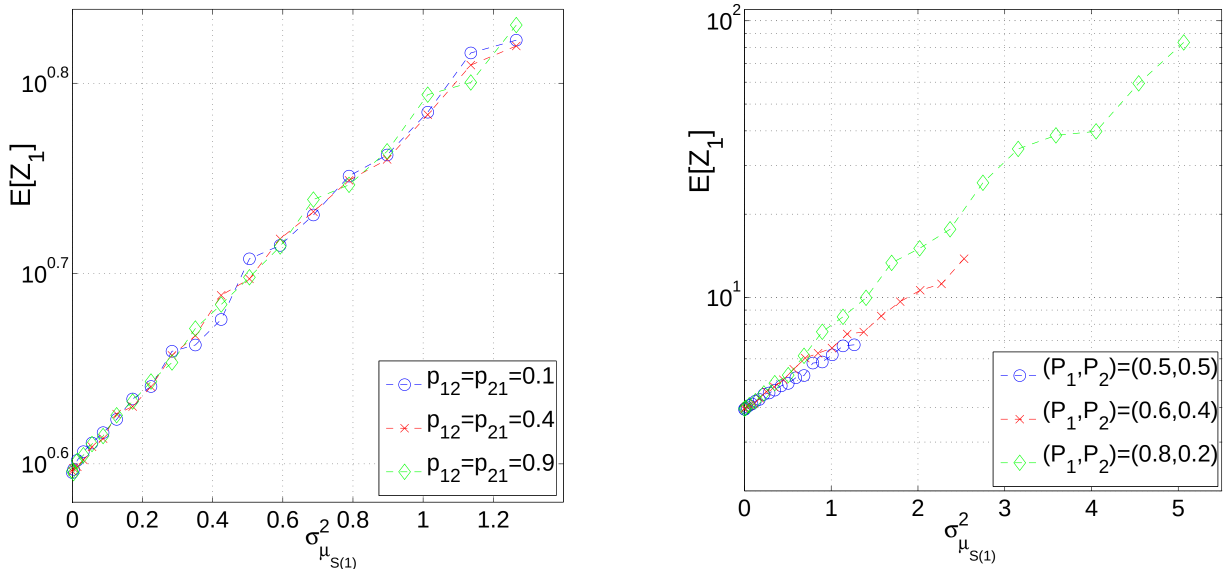
Also note that the expectation, $\mathrm{E}\left[\mu_{S(1)}\right]$, is over the steady state probabilities of the chain. Therefore, $\mathrm{E}[Z_1]$ is independent of the transition probabilities of the chain as long as the steady state probabilities remain fixed.

EXAMPLE 2. *In this example, we illustrate the expected number of transmissions for the Gilbert-Elliott channel. There are two states, i.e., $S(i) \in \{1, 2\}$ for all $i \geqslant 1$. We use p_{ml} to indicate the transition probability from state m to state l. We call a chain symmetric, if $p_{ij} = p_{ji}$ for all $i, j \in \mathcal{S}$. Also let P_k represent the steady state probability of state k.*

In this example we set $\sigma^2_{V^o} = 2.5 \times 10^{-3}$ and the packet size to 5000 bits. We assume that the erasure probability is constant for each bit during a transmission. For any given Markov chain, we generate a trace of length 20,000 packets. ***Symmetric Chain:*** *We run three sets of simulations in which the chain is symmetric: $p_{12} = p_{21} = 0.1$ in the first set, $p_{12} = p_{21} = 0.6$ in the second set and $p_{12} = p_{21} = 0.9$ in the third set. Hence, the steady state distribution, $(P_1, P_2) = (0.5, 0.5)$ in all sets.*

In each set of simulations, we initially set $(\mu_1, \mu_2) = (0.5, 5)$. Then, we increase μ_1 and decrease μ_2 gradually (keeping $E[\mu_1]$ constant at 2.75) all the way to $(\mu_1, \mu_2) = (2.75, 2.75)$, which corresponds to the scenario with iid erasure vectors. With the given parameters, one can compute that, $\sigma^2_{\mu_{S(1)}}$ decreases from 1.27 to 0 as we change the (μ_1, μ_2) pair.

The expected number of transmissions as a function of $\sigma^2_{\mu_{S(1)}}$ is illustrated in Fig. 7(a). As expected, there is no



(a) Symmetric Markov chains

(b) Asymmetric Markov chains

Figure 7: The expected number of transmissions is illustrated as a function of $\sigma^2_{\mu_{S(1)}}$. The distribution $(P_1, P_2) = (0.5, 0.5)$ in Fig. 7(a), while it varies from $(0.5, 0.5)$ to $(0.8, 0.2)$ in Fig. 7(b) where $\mu_1 < \mu_2$. E $\left[\mu_{S(1)}\right] = 2.75$ for all the simulations.

difference between these three curves since $\sigma_{\mu_{S(1)}}$ only depends on the steady state distribution. $E[Z_1]$ varies almost as $\exp(\sigma^2_{\mu_{S(1)}}/2)$ and the impact of the higher order terms (captured by the term $\frac{1}{6}\Lambda'''_{\mu_{S(1)}}(r'_w)$ in (20)) is minor for these symmetric chains.

Asymmetric Chain: *In this case, we vary $\sigma^2_{\mu_{S(1)}}$ in each simulation by changing the transition probabilities of the chain as well as the pair (μ_1, μ_2). The purpose is to illustrate how the expected number of transmissions is affected by the chain asymmetry of the Markov chain as well as by $\sigma^2_{\mu_{S(1)}}$.*

- Simulation 1: *Symmetric chain with $p_{12} = p_{21} = 0.1$. Increase μ_1 and decrease μ_2 from $(\mu_1, \mu_2) = (0.5, 5)$ to $(2.75, 2.75)$ incrementally, keeping $E\left[\mu_{S(1)}\right] = 2.75$.*
- Simulation 2: *Asymmetric chain with $p_{12} = 1 - p11 = 0.1$ and $p_{21} = 1 - p_{22} = 0.2$. Thus, $(P_1, P_2) = (2/3, 1/3)$. Increase μ_1 and decrease μ_2 from $(\mu_1, \mu_2) = (0.5, 7.25)$ to $(2.75, 2.75)$ incrementally, keeping $E\left[\mu_{S(1)}\right] = 2.75$.*
- Simulation 3: *Asymmetric chain with $p_{12} = 1 - p11 = 0.1$ and $p_{21} = 1 - p_{22} = 0.4$. Thus, $(P_1, P_2) = (4/5, 1/5)$. Increase μ_1 and decrease μ_2 from $(\mu_1, \mu_2) = (0.5, 11.75)$ to $(2.75, 2.75)$ incrementally, keeping $E\left[\mu_{S(1)}\right] = 2.75$.*

The results are illustrated in Fig. 7(b). One can observe that the range of the possible variance values increase with increasing asymmetry due to the increasing effect of the higher order cumulants of $\mu_{S(1)}$. Also, $E[Z_1]$ increases with increasing $|P_1 - P_2|$ for a given $\sigma_{\mu_{S(1)}}$. This increase is due to the varying effect of higher order terms ($\frac{1}{6}\Lambda'''_{\mu_{S(1)}}(r'_w)$ in (20)) for different steady state distributions. Indeed, one can show that for a Bernoulli random variable, with a given variance, $\Lambda'''_{\mu_{S(1)}}(r)$ increases for all $r > 0$ as $|\mu_1 - \mu_2|$ increases.

5. EXPERIMENTAL RESULTS

If we truncate the Taylor series expansion in Eq. (20) to the second order, we get

$$\mathrm{E}\left[Z_i\right] \approx \exp\left(\mathrm{E}\left[\mu_{S(i)}\right] + \frac{1}{2}(\sigma^2_{\mu_{S(i)}} - \sigma^2_{V^o})\right) \qquad (21)$$

for all $i \geqslant 1$. In this section we present experimental measurements taken from mote radios to support (21). For this purpose, we generate traces of packets and record the V_{ij} values for each packet. Then, we calculate the sample means, $\bar{Z}_i$ and $\bar{\mu}_{S(i)}$ of the number of transmissions and the cumulative erasure probability, respectively. We also estimate $\hat{\sigma}^2_{\mu_{S(i)}}$ and $\hat{\sigma}^2_{V^o}$, the sample variances of the two components of the cumulative erasure parameter given in (6). Then, we estimate the coefficients, c_μ, c_{lt} and c_{st} in

$$\bar{Z}_i = \exp\left(c_\mu \bar{\mu}_{S(i)} + c_{lt}\hat{\sigma}^2_{\mu_{S(i)}} + c_{st}\hat{\sigma}^2_{V^o}\right), \qquad (22)$$

and compare these estimates with the theoretically derived values of 1, 1/2 and −1/2, respectively.

The experimental setup is as follows. We placed two Berkeley Mica2 motes to act as a transmitter-receiver pair. The radio on the Mica2 mote is the Chipcon CC1000, an FSK radio operating at 433 MHz. The RF output power of the radio is programmable on a per-packet basis, and can vary from -20 dBm to 10 dBm in 1 dBm increments. We placed the motes at a distance of 20 feet apart. The reason for using a short distance between the pair is to suppress the uncontrollable sources of variability (e.g., fading) as much as possible and acquire measurements as we introduce variability in a controlled manner. We use the variable transmit power level to emulate the desired channel variability.

We increased the TinyOS packet size to 960 bits by modifying file `AM.h` in the standard TinyOS 1.x distribution, retaining the ActiveMessage layer unmodified. We also verified that this change did not break the MAC layer.

Recall that we define a *packet* as a group of N bits. Each packet can be transmitted multiple times until it is received erasure-free. We call each of these a *transmission*. Each experiment consists of sending 120 test packets from one node to the other, each at a power level chosen at random, from a discrete uniform distribution. We change the power level sufficiently frequently depending on the time scale of the variability we emulate. To emulate the short time scale variability, we change the power level for every transmission and to emulate the long time scale variability, we change it for every packet and keep it constant during all its possible retransmissions. The new power level is chosen independently and identically distributed as the previous power levels in each experiment. The nodes perform ARQ, and the receiving node sends its acknowledgments back at the maximum power level of 10 dBm. Hence we rarely observe losses in the reverse link. Each experimental data point in the graphs in Fig. 8(a) and 8(b) averages five experimental runs (i.e., a total of 600 packets).

We control the variance of the bit error rate (BER can be viewed as the erasure probability in our model) by varying the range of values that the uniform distribution for the transmit power level takes on. However, due to the convex relationship[8] between the power level and the BER, increasing the variance of the transmit power reduces the average BER at the same time as it increases the variance of BER. For the experiment involving the short term variability, our purpose is to analyze the impact of the change in variance alone, we reduce back the average BER by decreasing the mean transmit power, while keeping its variance unchanged. After this process, we still end up having slight differences in the average BER for different experimental traces. In both scenarios, instead of plotting the mean number, $\bar{Z}_i$, of transmissions as a function of the variance, $\text{var}\,(V_ij)$, we normalize it and plot $\bar{Z}_i/\exp(\bar{\mu}_{S(i)})$ as (21) suggests.

Short time scale variability: The transmitter picks the power level of each transmission (initial transmission and each subsequent retransmission) independently and at random, from a uniform distribution. We keep the transmit power level (and thus the BER) constant during one transmission of $N = 960$ bits. Since the power level is selected from the same distribution for all packets, there is just a single channel state, i.e., $\sigma^2_{\mu_{S(i)}} = 0$ and $\text{var}\,(V_{ij}) = \sigma^2_{V^o}$.

We generate each trace (a total of 10 traces) using a different uniform distribution for the transmit power. We try to keep $\bar{\mu}_{S(i)}$ around $0.3 - 0.5$ per packet and vary $\hat{\sigma}^2_{V^o}$ between 0.2 and 1. We calculate the sample mean number, $\bar{Z}_i$ of transmissions for each trace. The normalized mean number, $\bar{Z}_i/\exp(\bar{\mu}_{S(i)})$, of transmissions per packet for these 10 traces is plotted as a function of the sample variance, $\hat{\sigma}^2_{V^o}$, in Fig. 8(a).

It is clear that this number tends to decrease with increased $\hat{\sigma}^2_{V^o}$. Indeed the slope of the line fitted to the data points is -0.71, i.e., $c_{st} = 0.71$ in (22). This is reasonably close to the theoretical value, $-1/2$, given in (21). Also note that the normalized mean number of transmissions < 1 for all data points. Hence, the combined effect of the higher order cumulants of V^o_{ij} is negative.

Long time scale variability: The transmitter picks the power level of each transmission of $N = 320$ bits independently and at random, from a discrete uniform distribution. Subsequent retransmissions of the same packet are at the same power level of the original transmission. Thus, V_{ij} follows a symmetric Markov chain with a number of states identical to the number of power levels that the uniform distribution takes. Also, $\sigma^2_{V^o} \approx 0$ since the transmit power level is constant over all transmissions of a packet, so $\text{var}\,(V_{ij}) = \sigma_{\mu_{S(i)}}$.

We generate each trace (a total of 22 traces) using a different uniform distribution for the transmit power. For these traces, $\bar{\mu}_{S(i)}$ varies between $0.5 - 2.2$ and $\hat{\sigma}^2_{\mu_{S(i)}}$ takes on values from 0.8 up to 7. We calculate the sample mean number, $\bar{Z}_i$ of transmissions for each trace. The normalized mean number, $\bar{Z}_i/\exp(\bar{\mu}_{S(i)})$ of transmissions per packet is plotted as a function of the sample variance, $\hat{\sigma}^2_{\mu_{S(i)}}$ in Fig. 8(b).

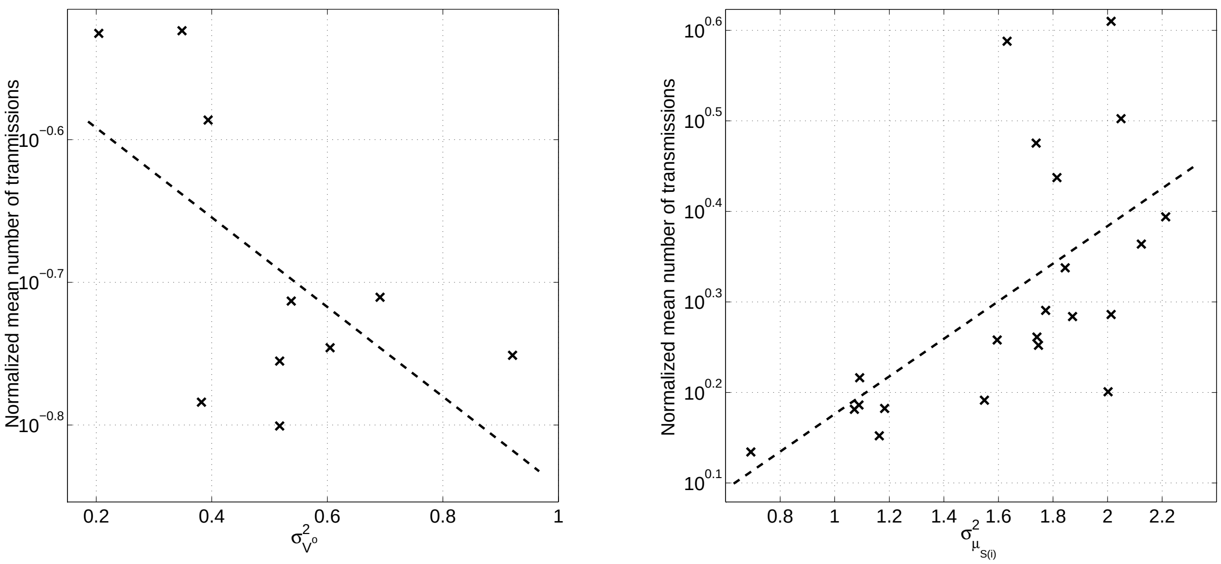
We get an almost perfect match with the theoretical result given in (21). Indeed, the slope of the line fitted to the data points is 0.478 and if we interpolate this line, the ordinate of the 0 variance point is 0.93. Thus, $c_{lt} = 0.478$ and $c_\mu = 0.93^{-1} = 1.075$ which are almost identical to the theoretically derived values of 1/2 and 1 respectively.

As a side note, the correlation coefficient between the measured mean and the variance of the number of bit errors per transmission is 0.58 over all the traces we generate. Thus, for a given wireless channel the average link behavior and the variability may be fairly dependent. This is not very surprising due to a very fundamental reason: If the average value of a non-negative stochastic process is very close to 0, it will not have much room to vary around (e.g., if the mean BER is 0, than the variance of the BER cannot be anything but 0).

6. CONCLUSIONS AND PRACTICAL IMPLICATIONS

Time diversity has a major impact on the throughput of a communication channel and thus, it must be taken into consideration when designing a communication system. In this paper, we studied the packet-level throughput of a wireless link with variable quality. We analytically evaluated the expected number of transmissions per packet using a discrete time-varying erasure channel with feedback. We expressed this quantity as a function of the cumulants of the samples of the stochastic process, which we used to model the erasure probability. We also illustrated our results experimentally using mote radios.

[8]The bit error rate is exponentially decreasing with the SNR level at the receiver. Since the distance between the transmitter and the receiver is small, we can assume that the BER decreases roughly exponentially with the transmitted power level as well. We verified this, but due to space constraints, we do not give the BER vs. transmit power curve. Hence, if the transmit power level is picked from a uniform distribution, the bit error probability is log uniform, i.e., log of this probability is uniform. The log uniform distribution has all its cumulants positive.



(a) Short term variability

(b) Long term variability

Figure 8: The normalized expected number of transmissions is illustrated as a function of the variance of the measured cumulative erasure parameter. The dashed lines illustrate the best fit to the log of the measured mean number of transmissions.

We showed that the channel variability at different time scales affects the link-level throughput positively or negatively depending on its time scale. For the time varying discrete binary erasure channel, we discovered that the variability in samples of the erasure probability (which can be also viewed as the bit error probability) separated as closely as a single packet increases the link-level throughput. Inversely, an increase in the variability of samples separated longer than a single packet reduces the link-level throughput.

In this paper, we did not consider the effect of possible losses in the reverse link where the ACKs are sent. Our results can be easily extended to take the reverse link into consideration as well. We also assumed that the packets containing erasures are dropped. Such packets could be kept and combined with other transmissions of the same packet to improve reliability. Indeed, hybrid ARQ systems are based on this idea. A natural extension of this work is to consider a system with packet combining and to study the impact of channel variability. Our initial analysis of a system with packet combining shows that, even though the channel variability affects the throughput similarly at exactly the same time scales as in the system without combining, the magnitude of the impact is much smaller in the system with combining. Therefore, the systems with packet combining tend to be much more robust with respect to channel variability.

Also, there are a number of implications of the findings of this paper on the quality aware networking protocols. Indeed, in [9] we expand the already existing routing protocols that are based on metrics for average channel behavior. Using the derivation developed in this paper, we showed that the widely used ETX metric in [4] disregards the variability at all time scales longer than a packet. We describe a new metric, the mETX, which uses the formulations of this paper to quantify the variability and combines it with ETX. Based on Roofnet link measurements, we show that packet losses in Roofnet are more correlated with channel variability than ETX. We illustrate that an improvement of up to 60% is possible in TCP throughput in randomly constructed mesh network topologies that use the Roofnet links. We believe the mETX metric can also be used to improve the throughput of MAC protocols based on opportunistic scheduling.

We can use the insights of this paper in wireless system design as well. For instance, at a transmitter, a packet scheduler can be designed to take advantage of the variability if the channel has a periodically (stochastically or deterministically) varying quality. An example of such a scenario can be a mobile sensor network, such as the one in [13], where a sink node mechanically cycles in the area to collect information. Packets can be interleaved in such a way that the channel quality is as negatively correlated as possible during distinct retransmissions. Also, since a high correlation is desirable at short time scales, the quality can be kept as constant as possible during the transmission of a packet, e.g., by reduced mobility (given that the average quality does not decrease by doing so).

Part of our analyses also apply to multi-radio based systems such as the one described in [14]. In such systems, multiple receivers (i.e., multiple channels) are used. Multiple channels for the transmission of the same packet can be viewed as a single channel in which the same packet is transmitted at different instances. With this analogy, we can show that in the two-radio system described in [14], the antenna placement with the greatest negative correlation among the two channels should be preferred.

7. REFERENCES

[1] Adya, A., Bahl, P., Padhye, J., Wolman, A., and Zhou, L. A multi-radio unification protocol for ieee 802.11 wireless networks. No. MSR-TR-2003-44.

[2] Chang, L. F. Throughput estimation of arq protocols for a rayleigh fading channel using fade and interfade duration statistics. *IEEE Trans. on Veh. Tech. 40* (1991), 223—229.

[3] Chase, D. Code combining–a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets. *IEEE Trans. on Communications* (May 1985), 385—393.

[4] DeCouto, D. S. J., Aguayo, D., Bicket, J., and Morris, R. A high-throughput path metric for multi-hop wireless routing. In *Proc. of ACM MOBICOM'03* (San Diego, CA, Sept. 2003).

[5] Elliott, E. O. Estimates of Error Rates for Codes on Burst-Noise Channels. *Bell Systems Technical Journal 42* (1963), 1977–1997.

[6] Gallager, R. *Discrete Stochastic Processes*, first ed. Kluwer Academic Publishers, 1996.

[7] Gilbert, E. N. Capacity of a Burst Noise Channel. *Bell Systems Technical Journal 39* (1960), 1253–1265.

[8] Hagenauer, J. Rate-compatible punctured convolutional codes (rcpc codes) and their applications. *IEEE Trans. on Communications 36* (Apr 1988), 389—400.

[9] Koksal, C. E., and Balakrishnan, H. Quality aware routing in time-varying wireless networks. *To Appear in IEEE Journal on Selected Areas of Communication, Special Issue on Multihop Wireless Mesh Networks* (2006).

[10] Lin, S., and Costello, D. J. *Error Control Coding: Fundamentals and Applications*, second ed. Prentice Hall, 2004.

[11] Liu, Y., and Knightly, E. Opportunistic fair scheduling over multiple wireless channels. In *Proc. of IEEE INFOCOM'03* (San Francisco, CA, Mar. 2003).

[12] Lu, D. L., and Chang, J. F. Performance of arq protocols in nonindependent channel errors. *IEEE Trans. on Communications 41* (1993), 721—730.

[13] Luo, J., and Hubaux, J.-P. Joint mobility and routing for lifetime elongation in wireless sensor networks. In *Proc. of IEEE INFOCOM'05* (Mar 2005).

[14] Miu, A., Balakrishnan, H., and Koksal, C. E. Improving loss resilience with multi-radio diversity in wireless networks. In *Proc. of ACM MOBICOM'05* (Cologne, Germany, Sept. 2005).

[15] Qin, X., and Berry, R. Opportunistic splitting algorithms for wireless networks. In *Proc. of IEEE INFOCOM'04* (Hong Kong, Mar. 2004).

[16] Rasmussen, L. K., and Wicker, S. B. A comparison of trellis-coded and reed-solomon coded hybrid-arq protocols over slowly fading rayleigh channels. *Wireless Pers. Commun. 2* (1996), 393—413.

[17] Towsley, D. A statistical analysis of arq protocols operating in a non-independent error environment. *IEEE Trans. on Communications 29* (1981), 971—981.

[18] Woo, A., Tong, T., and Culler, D. Taming the underlying challenges of reliable multihop routing in sensor networks. In *Proc. of ACM SenSys'03* (2003).

[19] Yarvis, M., Conner, W., Krishnamurthy, L., Chhabra, J., Elliott, B., and Mainwaring, A. Real world experiences with an interactive ad-hoc sensor network. In *Proceedings of the International Workshop on Ad Hoc Networking* (Vancouver, BC, Canada, Aug. 2002).

[20] Zhang, Q., and Kassam, S. A. Hybrid ARQ with Selective Combining for Fading Channels. *IEEE Journal on Selected Areas in Communications 17*, 5 (May 1999), 867–880.

[21] Zorzi, M., and Rao, R. R. On the use of renewal theory in the analysis of arq protocols. *Mobile Networks and Applications 2* (1997), 183–193.

APPENDIX

A. BOUNDS ON THE ERASURE PARAMETER

In this section we prove the following inequality.

Lemma 1. *Let $\eta = -\log(1-\epsilon)$. Then,*

$$\epsilon \leqslant \eta \leqslant \epsilon + \frac{\epsilon^2}{1-\epsilon^2}.$$

Proof: The inequality $\log x \leqslant x - 1$ is a simple consequence of Jensen's inequality and the concavity of the log function. The lower bound simply follows by setting $x = 1-\epsilon$. The upper bound can be derived as follows.

$$\begin{aligned} -\log(1-\epsilon) &= \sum_{k=1}^{\infty} \frac{\epsilon^k}{k} &(23)\\ &\leqslant \epsilon + \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} + \frac{\epsilon^4}{4} + \frac{\epsilon^4}{4} + \frac{\epsilon^4}{4} + \frac{\epsilon^4}{4} \\ &\qquad + \frac{\epsilon^8}{8} + \cdots &(24)\\ &= \epsilon + \sum_{j=1}^{\infty} \epsilon^{2^j} \\ &\leqslant \epsilon + \sum_{j=1}^{\infty} \epsilon^{2j} \\ &= \epsilon + \frac{\epsilon^2}{1-\epsilon^2}, &(25) \end{aligned}$$

where (23) follows from the Taylor series expansion of $-\log(1-\epsilon)$ and (24) follows since the sum in (23) increases if we replace the all the terms $2^j + 1 \leqslant k \leqslant 2^{j+1} - 1$ by the term 2^j completing the proof.

B. DISCUSSION: ERASURE PROBABILITY AND THE ERASURE PARAMETER

In our entire analysis of variability, we used the erasure parameter, $\eta_{ij}(n)$, instead of the actual erasure probability $\epsilon_{ij}(n)$. In the analysis of the short time scale variability, we

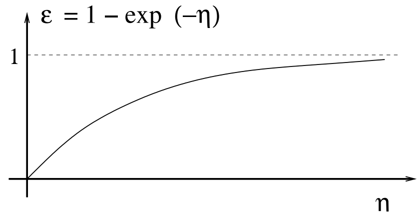


Figure 9: The erasure parameter versus the erasure probability.

found the relation between the variance, σ_V^2, of the sum of the samples of the erasure parameter over a packet time and the number of transmissions. Moreover, we used "channel variability" to refer to σ_V^2.

Even though it was noted in Eq. (3) that for small erasure probabilities, $\eta_{ij}(n) \approx \epsilon_{ij}(n)$, further analysis of the relation between the two is insightful. For notational simplicity, we drop all the subscripts and the time parameters.

First consider the memoryless channel, i.e., when $\epsilon(1), \epsilon(2), \ldots$ form an iid sequence. Let $\mu_\epsilon = \mathrm{E}[\epsilon]$. For a simple illustration of the insights to be given, we assume that each sample, $\eta(n) \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$. Then

$$\begin{aligned} \mu_\epsilon &= 1 - \mathrm{E}[\exp(-\eta)] \\ &= 1 - \exp\left(-\mu_\eta + \frac{1}{2}\sigma_\eta^2\right), \qquad (26) \end{aligned}$$

where (26) follows since η is Gaussian. Hence,

$$\begin{aligned} \mathrm{E}[Z] &= (1 - \mu_\epsilon)^{-N} \qquad (27) \\ &= \exp\left(N\mu_\eta - \frac{1}{2}N\sigma_\eta^2\right), \qquad (28) \end{aligned}$$

where (27) follows since the channel is memoryless and (28) follows from (14). Eq. (27) implies that for the memoryless erasure channel, the expected number of transmissions is not affected by the sample variance of ϵ. However, we have the σ_η^2 term on the exponent on the right side of (28). Here, we gave an example in which $\mathrm{E}[Z]$ decreases with increasing σ_η^2, but the channel variability has no effect on $\mathrm{E}[Z]$. Thus, we may not be able to claim that the variability of the erasure probability impacts the channel solely based on the presence of the σ_η^2 term in the expression of $\mathrm{E}[Z]$. Indeed, this term is a correction factor for the increased value of μ_η due to σ_ϵ^2.

The above analysis suggests the need for further understanding of the relation between the channel variability and the number of transmissions. For that purpose, first consider the other extreme case where the erasure probability is constant over each packet time, i.e., $\epsilon(n) = \epsilon$ for all $n \geqslant 1$. Note that this was the case in the examples we illustrated earlier on. Also, let us assume $\eta \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$. Hence,

$$\mathrm{E}[Z] = \exp\left(N\mu_\eta - \frac{1}{2}N^2\sigma_\eta^2\right). \qquad (29)$$

We showed earlier that a factor $\exp\left(-\frac{1}{2}N\sigma_\eta^2\right)$ in (29) is due to the concavity of $1 - \exp(-\eta)$. Thus, the remaining $\exp\left(\frac{1}{2}(N^2 - N)\sigma_\eta^2\right)$ reflects the actual impact of the variance of the erasure probability on the expected number of transmissions. Indeed the slope of the $\mathrm{E}[Z_1]$ versus σ_V^2 curve in Fig. (5(b)) for the Gaussian case is $\frac{1}{2}\left(1 - \frac{1}{N}\right)$, which is very close to $\frac{1}{2}$ since the packet size $N \gg 1$.

Next, we generalize the above analysis for any given possibly non-iid $\epsilon(n)$. Again, let us assume $V \sim \mathcal{N}(\mu_V, \sigma_V^2)$. Thus, as shown in (14),

$$\mathrm{E}[Z] = \exp\left(\mu_V - \frac{1}{2}\sigma_V^2\right).$$

The variance of V can be decomposed as

$$\sigma_V^2 = N\sigma_\eta^2 + 2\sum_{k=1}^{N}(N-k)K_\eta(k), \qquad (30)$$

where $K_\eta(k) = \mathrm{cov}(\eta(n), \eta(n+k))$ for $n \geqslant 1,\ k \geqslant 0$. For the memoryless channel, $\sigma_V^2 = N\sigma_\eta^2$ which is the first term in (30). We also showed that in the memoryless channel, the sample variance of the erasure probability has no effect in the number of transmissions. Consequently, the decrease in the expected number of transmissions due to variability is captured by the second term of (30), which is a function of the autocovariance of the erasure parameter. The part of the variability that affects the number of transmissions is therefore due to the covariances of different samples of the erasure probability, rather than the variances of each sample. Depending on whether this weighted sum of covariances is negative or positive, i.e., whether σ_V^2 is less than or greater than $N\sigma_\eta^2$, the variability of the erasure probability may have a negative or positive impact on the expected number of transmissions. Since the samples of erasure probability are close to each other in time, they are almost always highly positively correlated in practice.