Path Diversity over Packet Switched Networks: Performance Analysis and Rate Allocation

Shervan Fashandi, Shahab Oveis Gharan and Amir K. Khandani, *Member, IEEE*

***Abstract*—Path diversity works by setting up multiple parallel connections between the end points using the topological path redundancy of the network. In this paper, *Forward Error Correction* (FEC) is applied across multiple independent paths to enhance the end-to-end reliability. We prove that the probability of irrecoverable loss (P_E) decays exponentially with the number of paths. Furthermore, the *rate allocation* (RA) problem across independent paths is studied. Our objective is to find the optimal RA, i.e. the allocation which minimizes P_E. The RA problem is solved for a large number of paths. Moreover, it is shown that in such asymptotically optimal RA, each path is assigned a positive rate *iff* its *quality* is above a certain threshold. Finally, using memoization technique, a heuristic suboptimal algorithm with polynomial runtime is proposed for RA over a finite number of paths. This algorithm converges to the asymptotically optimal RA when the number of paths is large. For practical number of paths, the simulation results demonstrate the close-to-optimal performance of the proposed algorithm.** [1]

***Index Terms*—Path diversity, Wireless Mesh Networks, Internet, MDS codes, erasure, forward error correction, rate allocation, complexity.**

I. Introduction

A. Motivation

In recent years, *path diversity* over packet switched networks has received significant attention. This idea is applied over different types of networks like wireless mesh networks [2]–[4], the Internet [5]–[7], and Peer-to-peer networks [8]. Many studies have shown that path diversity has the ability to simultaneously improve the end-to-end rate and reliability [1], [5], [6], [9]–[11]. In order to apply path diversity over any packet switched network, two problems need to be addressed: i) setting up multiple independent paths between the end-nodes (multipath routing) ii) utilizing the given independent paths to improve the end-to-end throughput and/or reliability. In this paper, we focus on the second problem and try to develop a mathematical analysis of path diversity which is valid for any type of underlying network. Due to the inherent flexibility of wireless mesh networks, many routing protocols can be modified to support multipath routing over such networks [12]–[18]. Thus, we consider a wireless network as the underlying network. However, it should be noted that the results of this work stay valid for any other underlying network (e.g. path diversity over the Internet) as long as multiple independent paths are given. Assuming a set of independent paths, we utilize *Forward Error Correction* (FEC) across the given paths and analyze the reliability gain achieved by path diversity mathematically. Furthermore, the *rate allocation* (RA) problem across the given paths is addressed, and a polynomial suboptimal algorithm is introduced for this purpose.

B. Multipath Routing over Wireless Mesh Networks

In order to exploit path diversity, it is desirable to set multiple independent paths between the end nodes. This problem is addressed throughput the literature [12]–[20]. A set of paths are defined to be independent if their corresponding packet loss patterns are independent. According to the definition, any set of disjoint paths are independent. Even when the paths are not completely disjoint, their loss and delay patterns show a high degree of independence as long as they do not share any congestion points or bottlenecks [6], [21]–[26]. Many techniques are proposed to detect the shared congestion points, such as cross-correlation-based approach [27], entropy-based approach [28], and wavelet-based approach [29]. Hence, the independence of a set of paths can be verified by the mentioned bottleneck detection algorithms.

Many well-known mesh network routing protocols like AODV [30] and DSR [31] can be modified to support multipath routing. Indeed, DSR can find multiple paths naturally by its flooding behavior [31]. However, it does guarantee that the found paths are disjoint. The Split Multipath Routing (SMR) [12] solves this problem as it avoids dropping duplicate *Route Request* (RREQ) packets by the intermediate nodes. Of course, this is achieved at the cost of more RREQs and higher routing overhead. Similarly, the Multipath Source Routing (MSR) [18] introduces a multiple path routing protocol extended from DSR. Based on the measurement of Round-Trip Times (RTT), MSR also proposes a scheme to distribute the load among multiple paths. Leung *et al.* [17] propose the MP-DSR protocol which focuses on a newly defined metric for the QoS called the *end-to-end reliability*. MP-DSR is an algorithm which selects multiple paths with low fail probability associated by stable radio links. [16] addresses the problem of transmitting video with double description in the case where non of the paths to the destination is significantly more reliable than the others. The problem is turned into an optimization which is too complex to have a closed-form solution. Thus, the authors apply the metaheuristic *genetic algorithm* to find a suboptimal solution. Then, it is shown that this method can be incorporated into many existing on-demand routing protocols like DSR [16]. Finally, the Robust Multipath Source Routing Protocol (RMPSR) is another extension to DSR to support multipath video communication over wireless networks.

AMODV [14] is an Ad-hoc On-demand Multipath Distance Vector routing protocol based on the concept of link

[1]This manuscript is an extended version of the conference paper published in Globecom 2007 [1].

reversal extending from AODV. In contrast with the DSR-based multipath routing protocols, AMODV discovers multiple link-disjoint loopfree paths. AODVM [15] is another extension to AODV which finds multiple reliable routing paths. Similarly, AODV-BR [13] introduces an algorithm to find back up routing paths over Ad hoc networks. [3] proposes a novel multipath hybrid routing protocol, Multipath Mesh (MMESH), which effectively discovers multiple paths over wireless mesh networks. Simulation results show that MMESH is able to balance the traffic by avoiding hot paths, i.e., the paths with higher traffic load. AMTP [19], an ad hoc multipath streaming protocol for multimedia delivery which selects multiple maximally disjointed paths with best QoS to maximize the aggregate end-to-end throughput. AMTP is able to accurately differentiate between packet losses due to different network conditions. In case of a path being broken, it seamlessly switches to a proper path and therefore maintains high streaming quality. When there are multiple channels between the wireless mesh nodes, it is easier to find multiple independent paths across the network. Reference [2] applies the idea of multipath routing in such a scenario to increase the end-to-end throughput. Wei *et al.* [10] address the problem of path selection over a wireless network by taking into account the interference between the wireless links. Their goal is to minimize the *packet drop probability* (PDP). The problem of optimal multipath selection is shown to be NP-hard. Therefore, they introduce a heuristic algorithm to find a close-to-optimal set of paths. A previous work by the same authors [32] studies video multicast over wireless ad hoc networks. To take advantage of network path diversity in the multicast case, an algorithm to find multiple disjoint and near-disjoint trees is proposed.

C. Path Diversity over the Internet

In the Internet, the end-points have no control over the path selection process. Indeed, letting the end nodes set the paths requires modification of the IP routing protocol and extra signaling between the routers which are extremely costly. To avoid such an expense, *overlay networks* are introduced [24], [25], [33]. The basic idea of the overlay network is to equip very few nodes (smart nodes) with the desired new functionalities while the rest remain unchanged. The smart nodes form a virtual network connected through virtual or logical links on top of the physical network. Thus, overlay nodes can be used as relays to set up independent paths between the end nodes [7], [34]–[36].

Topology of the underlying physical network is an important factor in the design of the overlay network. Indeed, improper design of the overlay network can result in shared bottlenecks between different virtual links [37]. In such cases, even if two paths are disjoint in the virtual level, a large degree of dependency may be observed between them. Hence, a class of *topology-aware* overlay networks are proposed to maximize independence between the virtual links [37]–[43]. For instance, the overlay nodes can utilize latency [38], [39] or the underlying IP topological information [37], [40]–[43] to select the neighbors and form the overlay graph. It is shown that the topology-aware overlay networks can provide a satisfactory degree of independence between disjoint paths (disjoint in the virtual level) [37]. Moreover, distributed algorithms can be utilized to construct and/or maintain overlay networks. Reference [44] addresses the problem of distributed overlay network design based on a game theoretical approach, while [45] studies overlay networks failure detection and recovery through dynamic probing.

Another issue which may degrade path diversity in overlay networks is having bottlenecks in the links connecting the end-nodes to the network. To address this problem, the idea of *multihoming* is proposed [7], [46]. In this technique, the end users are connected to more than one *Internet Service Providers* (ISP's) simultaneously. It is shown that multihoming assists overlay networks to set up extra independent paths between the end-points, i.e. improves the end-to-end reliability considerably [7].

D. Applications of Path Diversity

Recently, path diversity is utilized in many applications (see [47]–[52]). Reference [49] combines multiple description coding and path diversity to improve the quality of service (QoS) in video streaming. In [9], multiple descriptions of video are routed throughput different paths across a wireless mesh network. It is assumed that coding is *non-hierarchical* in the sense that none of the descriptions is the main description. Instead, the distortion decreases gradually as the receiver receives more descriptions of the video. Moreover, non of the paths has significantly better quality than the others, and each link is modeled by a 2-state Markov model called the Gilbert channel. [9] concludes that in this setup, utilizing multiple paths improves both the rate and reliability.

Packet scheduling over multiple paths is addressed in [53] to optimize the rate-distortion function of a video stream. Reference [52] utilizes path diversity to improve the quality of Voice over IP streams. According to [52], sending some redundant voice packets through an extra path helps the receiver buffer and the scheduler optimize the trade-off between the maximum tolerable delay and the packet loss ratio [52].

In [5], multipath routing of TCP packets is applied to control the congestion with minimum signaling overhead. When the underlying network is an ad hoc wireless network, a similar result is reported [54]. In other words, transmitting video over multiple paths is shown to decrease the average congestion and end-to-end distortion. [55] proposes a multiflow reatime transport protocol for wireless networks. Through both mathematical analysis and comprehensive simulation, it is shown that partitioning the video packets across multiple paths improves queuing performance of the multimedia data, resulting is less congestion, smaller delay, and higher utilization of the bottleneck link bandwidth [55].

Content Distribution Networks (CDN's) can also take advantage of path diversity for performance improvement. CDN's are a special type of overlay networks consisting of *Edge Servers* (nodes) responsible for delivery of the contents from an original server to the end users [33], [56]. Current commercial CDN's like *Akamai* use path diversity based techniques like *SureRoute* to ensure that the edge servers maintain reliable connections to the original server. Video server selection schemes are discussed in [34] to maximize path diversity in CDN's.

E. Contribution and Relation to Previous Works

References [11], [6], and [57] study the RA problem over multiple independent paths. Assuming each path follows the leaky bucket model, reference [11] shows that a water-filling scheme provides the minimum end-to-end delay. On the other hand, reference [6] considers a scenario of multiple senders and a single receiver, assuming all the senders share the same source of data. The connection between each sender and the receiver is assumed to be independent from others and follow the *Gilbert model*. In order to benefit from path diversity, the authors apply FEC across independent paths. A *Maximum Distance Separable* (MDS) block code, like Reed-Solomon code, is used for FEC. [6] proposes a receiver-driven protocol for packet partitioning and rate allocation. The packet partitioning algorithm ensures no sender sends the same packet, while the RA algorithm minimizes the probability of irrecoverable loss in the FEC scheme [6]. They only address the RA problem for the case of two paths. A brute-force search algorithm is proposed in [6] to solve the problem. Generalization of this algorithm over multiple paths results in an exponential complexity in terms of the number of paths. Moreover, it should be noted that the scenario of [6] is equivalent, without any loss of generality, to the case in which multiple independent paths connect a pair of end-nodes as they assume the senders share the same data.

Djukic and Valaee utilize path diversification to provide low probability of packet loss (PPL) in wireless networks [4]. Similar to our work, they consider each path as an erasure channel following the multi-state Markov model. Moreover, it is assumed that the feedback is not fast enough to acknowledge the receipt of each packet. Thus, an MDS code is applied across multiple independent paths as a FEC method. The authors of [4] compare two RA schemes: blind allocation and optimal allocation. The blind RA is used when the source has no information about the quality of the paths. Hence, it distributes the traffic across the paths uniformly. It is shown that even blind RA outperforms single-path transmission. When a feedback mechanism periodically provides the source with information about the quality of each path, the transmitter has the chance to find the RA which minimizes PPL (optimal allocation). The authors propose a greedy algorithm for this purpose.

Most recently, in an independent work, Li *et al.* have addressed the RA problem [57]. Same as [4], [6] and our work, the authors of [57] apply an MDS code for FEC across multiple independent paths. However, unlike [6], the authors study the problem for any general number of paths, denoted by L. Using the *discrete to continuous* approximation, the authors approximate the total number of lost packets over all paths with a continuous random variable. Furthermore, assuming a large number of paths with a large number of packets over each path, they apply the Central-Limit Theorem (CLT) [58] to approximate the distribution of the number of lost packets with the *Normal Distribution*. Using this distribution, the authors propose a pseudo-polynomial algorithm, based on *Dynamic Programming*, to estimate the optimal RA for a large number of paths. However, CLT can not be applied to solve this problem. The reason is that in this case, the variance of the fraction of lost packets scales as $O(\frac{1}{L})$ to zero. Instead, as we show in this paper, the distribution of lost packets can be computed using *Large Deviation Principle* (LDP) which results in a distribution totally different from the normal distribution. Hence, the pseudo-polynomial algorithm proposed in [57] can not necessarily approximate the optimal RA even for large number of paths.

In this work, we utilize path diversity to improve the performance of FEC between two end-nodes over a general packet switched network. The details of path setup process is not discussed here. Similar to [4], [6], [11], [57], it is assumed that L independent paths are set up by a smart multipath routing scheme or overlay network. Moreover, as in [4], [57], [59], [60], each path is assumed to be an erasure channel modeled as a continuous M-state extended Gilbert model. It should be noted that the well-known 2-state Gilbert channel used in [6], [50], [61]–[63] is a special case of the extended Gilbert model studied here. Probability of irrecoverable loss (P_E) is defined as the measure of FEC performance. In another work, we have shown that MDS block codes have the minimum probability of error over any erasure channel with or without memory [64]. Hence, as in [4], [6], [57], MDS codes are applied for FEC throughout this paper. The contributions of this paper can be listed as follows:

- Path diversity is shown to simultaneously achieve an exponential decay in P_E and a linear increase in the end-to-end rate with respect to L, while the delay stays fixed. Furthermore, the decaying exponent is analyzed mathematically based on LDP.
- The RA problem is solved for the asymptotic case (large values of L).
- It is proved in the asymptotically optimal RA, each path is assigned a positive rate *iff* its *quality* is above a certain threshold. Quality of a path is defined as the percentage of the time it spends in the bad state. This result is important since for the first time in the literature, an analytical criterion is proposed to predict whether adding an extra path improves reliability.
- A heuristic suboptimal polynomial algorithm, based on the memoization technique, is introduced to solve the RA problem for any arbitrary number of paths. Unlike the brute-force search in [6], this algorithm has a polynomial complexity, in terms of L.
- The proposed algorithm is proved to converge to the asymptotically optimal RA as L grows.
- Through the simulation results, the proposed algorithm is shown to achieve a near-optimal performance for practical number of paths.

The rest of this paper is organized as follows. Section II describes the system model. Performance of FEC in two cases of multiple identical paths, and non-identical paths are

TABLE I
IMPORTANT PARAMETERS

Notation	Refers to	Section
L	number of the paths	I-E
N	length of an FEC block (in packets)	II-B
K	number of information packets in an FEC block	II-B
$\alpha = (N-K)/N$	FEC overhead	II-B
T	transmission time of an FEC block	II-C
S_{req}	required end-to-end rate (pkt/sec)	II-C
N_i	number of packets transmitted on path i in each FEC block	II-C
S_i, W_i	rate and max. rate of path i (pkt/sec)	II-C
P_E	probability of irrecoverable loss	I-E
$x_i = B_i/T$	fraction of bad bursts on path i during T	III
ρ_i	fraction of end-to-end rate assigned to path i	III-A
J	number of path types	III-B
$\gamma_j = L_j/L$	fraction of paths of type j	III-B
η_j	fraction of the end-to-end rate allocated to paths of type j, see (7)	III-B
$\boldsymbol{\eta}^\star$	asymptotically optimal rate allocation vector	III-B
$\boldsymbol{\eta}^{opt} = \mathbf{N}^{opt}/N$	optimal rate allocation vector	III-B
N_j	number of packets transmitted on paths of type j in each FEC block	IV
$P_e^{\mathbf{N}}(k,j)$	probability of having more than k errors over paths of types 1 to j for the allocation vector $\mathbf{N}$	IV
$Q_j(n,k)$	probability of having exactly k errors out of the n packets sent over paths of type j	IV
$\mathbf{N}^{opt}$	optimum allocation vector	IV
$P_e^{opt}(n,k,j)$	$P_e^{\mathbf{N}^{opt}}(k,j)$, i.e., min P_E	IV
$\hat{P}_e(n,k,j)$	lowerbound of $P_e^{opt}(n,k,j)$, see (16)	IV
$\hat{\mathbf{N}} = (N_1, \ldots, N_J)$	suboptimum allocation vector	IV
$\mathbf{K} = (K_1, \ldots, K_J)$	typical error event	IV

analyzed in section III. Section IV studies the RA problem, and proposes a suboptimal RA algorithm. Finally, section V concludes the paper.

II. SYSTEM MODELING AND FORMULATION

A. End-to-End Channel Model

From an end to end protocol's perspective, performance of the lower layers in the protocol stack can be modeled as a random *channel* called the *end-to-end channel*. Since each packet usually includes an internal error detection coding (for instance a Cyclic Redundancy Check), the end-to-end channel is modeled as an erasure channel.

Numerous measurements studies have suggested that bursty loss behavior is the most dominant characteristic of the end-to-end channel over different underlying networks, including wireless mesh networks and the Internet [4], [60], [65]–[67]. Hence, a variety of models have been proposed to capture this bursty behavior, including the 2-state Gilbert model, the M-state Extended Gilbert model, and the Hidden Markov model [59], [60], [65], [68], [69]. This paper assumes the continuous time M-state extended Gilbert model for the end-to-end channel, see Fig. 1. This model achieves a good balance between model accuracy and simplicity [57], [59], [60]; it is much more accurate than the 2-state Gilbert Model, while only requires $2(M-1)$ parameters to be estimated (as opposed to M^2 parameters in the General Markov Model). It should be noted that the well-known 2-state Gilbert channel used in [6], [50], [61]–[63] is a special case of the extended Gilbert model studied here.

It is worth mentioning that the main results of this paper remain valid for any end-to-end channel model. More precisely,

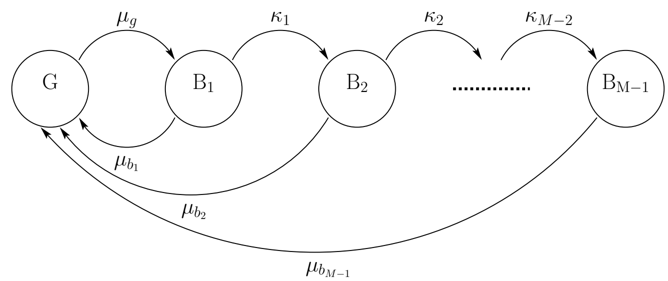


Fig. 1. Continuous-time M-state Extended Gilbert model of the end-to-end channel

P_E still decays exponentially versus L and the asymptotically optimal RA follows the same formula. However, the decaying exponent of P_E is a function of the bad burst probability distribution which should be recomputed according to the new end-to-end channel model. Moreover, in the proposed suboptimal RA algorithm, no assumption is made regarding the end-to-end channel model and/or the bad burst probability distribution. In other words, the input parameters to the proposed algorithm consist of the probability mass function (pmf) associated with the number of erasures over different paths. These input parameters are computed in polynomial time in appendix H for any general Markov model which obviously includes the extended Gilbert model as a special case.

The behavior of the continuous time extended Gilbert model can be described as follows. The channel spends an exponentially distributed random amount of time with the mean $\frac{1}{\mu_g}$ in the *Good* state. Then, it alternates to the first *Bad* state, B_1, and stays in that state for another random duration exponentially distributed with the mean $\frac{1}{\mu_{b_1}+\kappa_1}$. Then, the channel either goes back to the good state or transits to a deeper bad state, denoted by B_2. Similarly, the channel can move to deeper bad states consecutively before going back to the good state. The steady state probability of being in the good or any of the bad states are denoted by π_g and π_{b_i}. It is easy to observe that $\pi_g = \frac{1}{\mu_g \Xi}$ and $\pi_{b_i} = \frac{1}{(\mu_{b_i}+\kappa_i)\Xi}$ where $\Xi \triangleq \frac{1}{\mu_g} + \sum_{i=1}^{M-1} \frac{1}{\mu_{b_i}+\kappa_i}$. The packets transmitted during the good state are received correctly, while they are lost if transmitted during any of the bad states (B_1 to B_{M-1}). Therefore, the average probability of error, π_b, is equal to the steady state probability of being in any of the bad states, $\pi_b = \sum_{i=1}^{M-1} \pi_{b_i}$.

B. FEC Model

In real-time applications like video and audio over wireless mesh networks or IP, due to the delay requirement, conventional retransmission based schemes such as automatic repeat request (ARQ) are impractical. On the other hand, FEC is shown to be favorable for such real-time scenarios with tight QoS requirement [4], [61], [62], [70]–[72]. However, FEC could be ineffective when bursty packet loss occurs and such loss exceeds the recovery capability of the FEC codes. To mitigate this problem via path diversity, this work applies FEC across multiple paths.

Each packet is provided with an internal coding such as the Cyclic Redundancy Check (CRC) which enables the receiver to detect an error inside each packet. Hence, the receiver can consider the end-to-end channel as an erasure channel. Assuming the length of each packet is r bits, the alphabet size of the end-to-end channel would be $q = 2^r$. Other than the coding inside each packet, a FEC scheme is applied between packets. Every K packets are encoded to a *Block* of N packets where $N > K$ to create some redundancy. The N packets of each block are distributed across the L available independent paths, and are received at the destination with some loss (erasure). The ratio of $\alpha \triangleq \frac{N-K}{N}$ defines the FEC overhead. It is proved that among all block codes of the same size, any *Maximum Distance Separable* (MDS) code, such as the Reed-Solomon code, provides the minimum probability of error over an erasure channel (either memoryless or with memory) [64]. Moreover, MDS codes can reconstruct the original K data packets at the receiver side if K or more of the N packets are received correctly [73]. This property makes MDS codes favorable FEC schemes over the erasure channels [57], [74]–[76].

Since MDS codes are used for FEC, the probability of irrecoverable loss, P_E, is adopted as the reliability metric. An irrecoverable loss occurs when more than $N - K$ packets are lost in a block of N packets. It is shown in [64] that P_E is almost equal to the error probability of the maximum likelihood decoder for an MDS code, $P_{\mathcal{E}}$. More precisely, P_E can be bounded as

$$P_{\mathcal{E}} \leq P_E \leq \left(1 + \frac{1}{q-1}\right) P_{\mathcal{E}}$$

where q denotes the alphabet size of the MDS code which is very large in our application. The reason P_E is used as the measure of system performance is that while many practical low-complexity decoders for MDS codes work perfectly if the number of correctly received symbols is at least K, their probability of correct decoding is much less than that of maximum likelihood decoders when the number of correctly received symbols is less than K [73]. Thus, in the rest of this paper, P_E is used as a close approximation of decoding error.

C. Rate Allocation Problem

The RA problem is formulated as follows. L independent paths, $1, 2, \ldots, L$, connect the source to the destination, as indicated in Fig. 2(a). Information bits are transmitted as packets, each of a constant length r. Each path has a rate constraint of W_i packets per second. This constraint can be considered as an upperbound imposed by the physical characteristics of the path. For a specific application and FEC scheme, we require the rate of S_{req} packets per second from the source to the destination. Obviously, we should have $S_{req} \leq \sum_{i=1}^{L} W_i$ to have a feasible solution. As mentioned in the previous subsection, the information packets are coded in blocks of length N packets. Hence, it takes $T = \frac{N}{S_{req}}$ seconds to transmit one block.

The RA vector $\mathbf{N} = (N_1, \ldots, N_L)$ is defined as the number of packets in one block sent through each path. The objective of the RA problem is to find the optimal RA vector, i.e. the RA vector minimizing the probability of irrecoverable loss, P_E, defined in the previous subsection. The RA vector should

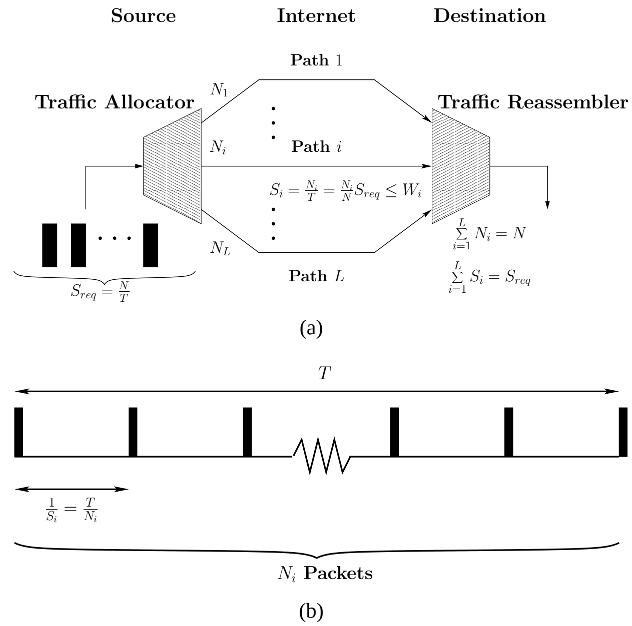


Fig. 2. RA problem: a block of N packets is being sent from the source to the destination through L independent paths over the network during the time interval T with the required rate $S_{req} = \frac{N}{T}$. The block is distributed over the paths according to the vector $\mathbf{N} = (N_1, \ldots, N_L)$ which corresponds to the RA vector $\mathbf{S} = (S_1, \ldots, S_L)$

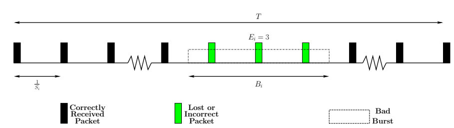


Fig. 3. A bad burst of duration B_i happens in a block of length T. $E_i = 3$ packets are corrupted or lost during the interval B_i. Packets are transmitted every $\frac{1}{S_i}$ seconds, where S_i is the rate of path i in pkt/sec.

satisfy the constraints $\sum_{i=1}^{L} N_i = N$ and $\frac{N_i}{T} \leq W_i, \forall\, 1 \leq i \leq L$. The latter constraint follows from the bandwidth constraint, $S_i = \frac{N_i}{T} \leq W_i$.

The above formulation of RA problem is valid for any finite number of paths and any chosen values of N and T. However, in section III where the performance of path diversity is studied for a large number of paths, and also in Theorem 2 where the optimality of the proposed suboptimal algorithm is proved for the asymptotic case, we assume that N grows linearly in terms of the number of paths, i.e. $N = n_0 L$, for a fixed n_0. The reason behind this assumption is that when L grows asymptotically large, the number of paths eventually exceeds the block length, if N stays fixed. Thus, $L - N$ paths become useless for the values of N larger than N. At the same time, it is assumed that the delay imposed by FEC, T, stays fixed with respect to L. This model results in a linearly increasing rate as the number of paths grows.

D. Discrete to Continuous Approximation

To compute P_E, we have to find the probability of k_i packets being lost out of the N_i packets transmitted through path i, for all $1 \leq i \leq L$, $0 \leq k_i \leq N_i$. Let us denote the number of erroneous or lost packets over the path i with the random variable E_i. Any two subsequent packets transmitted

TABLE II
MAIN ASSUMPTIONS

Assumption	Comments
L independent paths	justified in subsection I-B and I-C used in sections III and IV
discrete to continuous approximation	justified in subsection II-D used in section III
Extended Gilbert Model	justified in subsection II-A used in section III results valid without this assumption see subsections II-A and III-A for details

over the path i are $\frac{1}{S_i}$ seconds apart in time, where S_i is the transmission rate over the i'th path. Now, we define the continuous random variable B_i as the duration of time that path i spends in the bad state in a block duration, T. It is easily observed that the probability $\mathbb{P}\{E_i \geq k_i\}$ can be approximated with the continuous counterpart $\mathbb{P}\{B_i \geq \frac{k_i}{S_i}\}$ when the inter-packet interval is much shorter than the average bad burst duration. According to the extended Gilbert model, the average bad burst duration can be lower-bounded by $\frac{1}{\mu_{b_1}+\kappa_1}$. Therefore, as long as we have $\dfrac{1}{S_i} \ll \dfrac{1}{\mu_{b_1}+\kappa_1}$, the discrete to continuous approximation is valid (see Fig. 3).

The necessity of this condition can be justified as follows. In case this condition does not hold, any two consecutive packets have to be transmitted on two independent states of the channel. Thus, no gain would be achieved by applying diversity over multiple independent paths. The continuous approximation is just used in section III. On the other hand, section IV studies the RA problem in the original discrete format.

E. Notation and System Parameters

Table II summarizes the main assumptions made in our network model and problem formulation. The important parameters which are used throughout the paper are summarized in Table I. Moreover, the following mathematical notations are used in the rest of the paper. $\mathbb{P}\{.\}$ and $\mathbb{E}\{.\}$ are defined as the probability and expected value operators, respectively. The notation $P_E \doteq e^{-u(\alpha)L}$ means $\lim_{L\to\infty} -\frac{\log P_E}{L} = u(\alpha)$. $f(L) = o(g(L))$ is equivalent to $\lim_{L\to\infty} \frac{f(L)}{g(L)} = 0$, and $f(L) = O(g(L))$ means that $\exists L_0, M > 0 : \forall L > L_0, |f(L)| < M\,|g(L)|$.

III. PERFORMANCE ANALYSIS OF FEC ON MULTIPLE PATHS

According to the discrete to continuous approximation in subsection II-D, when the N_i packets of the FEC block are sent over path i, the loss count can be written as $\frac{B_i}{T}N_i$. Hence, the total ratio of lost packets is equal to

$$\sum_{i=1}^{L} \frac{B_i N_i}{TN} = \sum_{i=1}^{L} \frac{B_i \rho_i}{T}$$

where $\rho_i \triangleq \frac{S_i}{S_{req}}$, $0 \leq \rho_i \leq 1$, denotes the portion of the bandwidth assigned to path i. $x_i \triangleq \frac{B_i}{T}$ is defined as the portion of time that path i has been in the bad state ($0 \leq x_i \leq 1$). Hence, the probability of irrecoverable loss for an MDS code is equal to

$$P_E = \mathbb{P}\left\{\sum_{i=1}^{L} \rho_i x_i > \alpha\right\}. \qquad (1)$$

In order to find the optimum rate allocation, P_E has to be minimized with respect to the allocation vector (ρ_i's), subject to the following constraints:

$$0 \leq \rho_i \leq \min\left\{1, \frac{W_i}{S_{req}}\right\}, \quad \sum_{i=1}^{L} \rho_i = 1 \qquad (2)$$

where W_i is the bandwidth constraint on path i defined in subsection II-C.

A. Identical Paths

When the paths are identical and have equal bandwidth constraints[2] ($W_i = W$ for $\forall\ 1 \leq i \leq L$), due to the symmetry of the problem, the uniform RA ($\rho_i = \frac{1}{L}$) is obviously the optimum solution. Of course, the solution is feasible only when we have $\frac{1}{L} \leq \frac{W}{S_{req}}$. Then, the probability of irrecoverable loss can be simplified as

$$P_E = \mathbb{P}\left\{\frac{1}{L}\sum_{i=1}^{L} x_i > \alpha\right\}. \qquad (3)$$

Let us define $Q(x)$ as the probability density function of x. Since x is defined as $x = \frac{B}{T}$, clearly we have $Q(x) = Tf_B(xT)$, where $f_B(t)$ is the probability density function (pdf) of B. Defining $\mathbb{E}\{\}$ as the expected value operator throughout this paper, $\mathbb{E}\{x\}$ can be computed based on $Q(x)$. We observe that in (3), the random variable x_i's are bounded and independent. Hence, the following well-known upperbound in large deviation theory [77] can be applied

$$P_E \leq e^{-u(\alpha)L}$$
$$u(\alpha) = \begin{cases} 0 & \text{for } \alpha \leq \mathbb{E}\{x\} \\ \lambda\alpha - \log(\mathbb{E}\{e^{\lambda x}\}) & \text{otherwise} \end{cases} \qquad (4)$$

where the $\log$ function is computed in Neperian base, and λ is the solution of the following non-linear equation, which is shown to be unique by Lemma 1.

$$\alpha = \frac{\mathbb{E}\{xe^{\lambda x}\}}{\mathbb{E}\{e^{\lambda x}\}}. \qquad (5)$$

Since λ is unique, we can define $l(\alpha) = \lambda$. Even though being an upperbound, inequality (4) is exponentially tight for large values of L [77]. More precisely

$$P_E \doteq e^{-u(\alpha)L} \qquad (6)$$

where the notation $\doteq$ means $\lim\limits_{L\to\infty} -\dfrac{\log P_E}{L} = u(\alpha)$. Note that $u(\alpha)$ depends on the pdf of B, $f_B(t)$, which is computed in appendix A. Of course, equation (6) is valid regardless of the pdf of B.

Next, we state the following lemmas which are required for the analysis of the next subsection. The proofs can be found in the appendices B and C, respectively.

[2]The case where W_i's are different is discussed in Remark 4 of subsection III-B

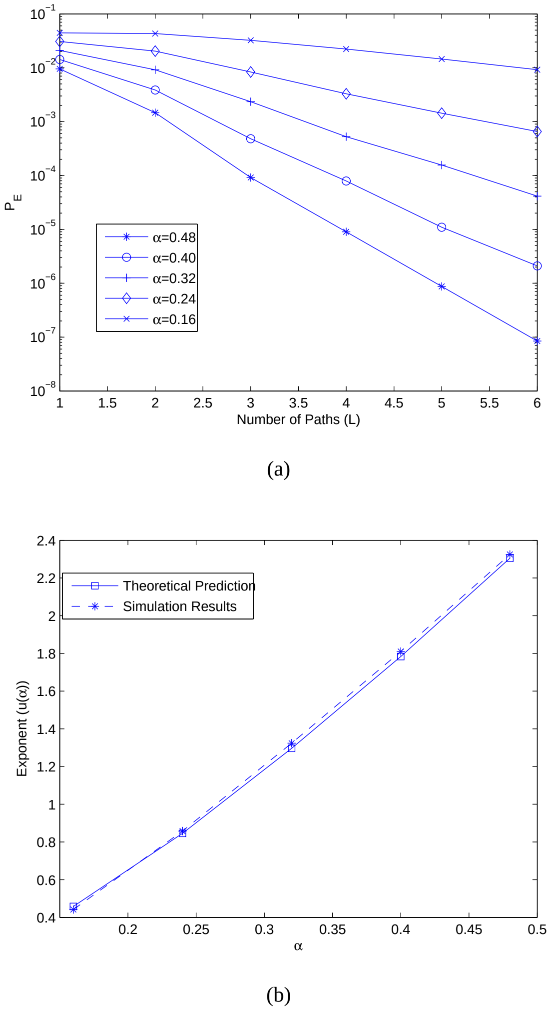


Fig. 4. (a) P_E vs. L for different values of α. (b) The exponent (slope) of plot (a) for different values of α: experimental versus theoretical values.

Lemma 1. *$u(\alpha)$ and $l(\alpha)$ have the following properties:*

1) $\frac{\partial}{\partial\alpha}l(\alpha) > 0$
2) $l(\alpha = 0) = -\infty$
3) $l(\alpha = \mathbb{E}\{x\}) = 0$
4) $l(\alpha = 1) = +\infty$
5) $\frac{\partial}{\partial\alpha}u(\alpha) = l(\alpha) > 0$ *for* $\alpha > \mathbb{E}\{x\}$

Lemma 2. *Defining $y = \frac{1}{L}\sum_{i=1}^{L} x_i$, where x_i's are i.i.d. random variables as already defined, the probability density function of y satisfies $f_y(\alpha) \doteq e^{-u(\alpha)L}$, for all $\alpha > \mathbb{E}\{x\}$.*

Remark 1. A special case is when the block code uses all the bandwidth of the paths. In this case, we have $N = LWT$, where W is the maximum bandwidth of each path, and T is the block duration. Assuming $\alpha > \mathbb{E}\{x\}$ is a constant independent of L, we observe that the information packet rate is equal to $\frac{K}{T} = (1-\alpha)WL$, and the error probability is $P_E \doteq e^{-u(\alpha)L}$. This shows using MDS codes over multiple independent paths provides an exponential decay in the irrecoverable loss probability and a linearly growing end-to-end rate in terms of the number of paths, simultaneously.

Example 1. Consider the scenario of transmitting a video stream with the DVD quality (using either MPEG-2 or MPEG-4) over multiple identical paths. The bitrate per path is selected to be 1 Mbps. The number of paths varies from $L = 1$ to $L = 6$. Hence, the end-to-end video bitrate varies in the range of $1 - 6$ Mbps, in accordance with [78]–[82]. The block transmission time is $T = 200$ ms which imposes an acceptable end-to-end delay for the video stream. The payload of each video packet is assumed to be 4 kb. Accordingly, the block length equals to $N = n_0 L$ where n_0 can be written as $n_0 = \dfrac{1 \text{ Mbps}}{4 \text{ kb}}T = 50$. The end-to-end channel follows a 2-state Gilbert model with $\frac{1}{\mu_g} = 2500$ ms and $\frac{1}{\mu_b} = 52$ ms, in accordance with [6], [50]. Coding overhead is changed from $\alpha = 0.16$ to $\alpha = 0.48$. Figure 4 compares the result of (6) with the simulation results. P_E is plotted versus L in semilogarithmic scale in Fig. 4(a) for different values of α. We observe that as L increases, $\log P_E$ decays linearly which is expected noting equation (6). Also, Fig. 4(b) compares the slope of each plot in Fig. 4(a) with $u(\alpha)$. Figure 4 shows a good agreement between the theory and the simulation results for practical number of paths. Moreover, it verifies the fact that the stronger the FEC code is (larger α), the higher is the gain we achieve through path diversity (larger exponent).

B. Non-Identical Paths

Now, let us assume there are J types of paths between the source and the destination, consisting of L_j identical paths of type j ($\sum_{j=1}^{J} L_j = L$). Without loss of generality, we assume that the paths are ordered according to their associated type, i.e. the paths from $1 + \sum_{k=1}^{j-1} L_k$ to $\sum_{k=1}^{j} L_k$ are of type j. We denote $\gamma_j = \frac{L_j}{L}$. According to the i.i.d. assumption, it is obvious that ρ_i has to be the same for all paths of the same type. η_j and y_j are defined as

$$\begin{aligned}
\eta_j &= \sum_{\sum_{k=1}^{j-1} L_k < i \le \sum_{k=1}^{j} L_k} \rho_i \\
y_j &= \frac{\eta_j}{L\gamma_j} \sum_{\sum_{k=1}^{j-1} L_k < i \le \sum_{k=1}^{j} L_k} x_i.
\end{aligned} \tag{7}$$

Following Lemma 2, we observe that $f_{y_j}(\beta_j) \doteq e^{-\gamma_j u_j(\frac{\beta_j}{\eta_j})L}$. We define the sets $\mathcal{S}_I$, $\mathcal{S}_O$ and $\mathcal{S}_T$ as

$$\begin{aligned}
\mathcal{S}_I &= \left\{ (\beta_1, \beta_2, \cdots, \beta_J) \,|0 \le \beta_j \le 1, \; \sum_{j=1}^{J} \beta_j > \alpha \right\} \\
\mathcal{S}_O &= \left\{ (\beta_1, \beta_2, \cdots, \beta_J) \,|0 \le \beta_j \le 1, \; \sum_{j=1}^{J} \beta_j = \alpha \right\} \\
\mathcal{S}_T &= \left\{ (\beta_1, \beta_2, \cdots, \beta_J) \,|\eta_j \mathbb{E}\{x_j\} \le \beta_j, \sum_{j=1}^{J} \beta_j = \alpha \right\}
\end{aligned}$$

respectively. Hence, P_E can be written as

$$\begin{aligned} P_E &= \mathbb{P}\left\{\sum_{j=1}^{J} y_j > \alpha\right\} \\ &= \int_{\mathcal{S}_I} \prod_{j=1}^{J} f_{y_j}(\beta_j) d\beta_j \\ &\doteq \int_{\mathcal{S}_I} e^{-L\sum_{j=1}^{J}\gamma_j u_j(\frac{\beta_j}{\eta_j})} \prod_{j=1}^{J} d\beta_j \\ &\overset{(a)}{\doteq} e^{-L \min_{\boldsymbol{\beta}\in\mathcal{S}_I\cup\mathcal{S}_O} \sum_{j=1}^{J}\gamma_j u_j\left(\frac{\beta_j}{\eta_j}\right)} \\ &\overset{(b)}{\doteq} e^{-L \min_{\boldsymbol{\beta}\in\mathcal{S}_O} \sum_{j=1}^{J}\gamma_j u_j\left(\frac{\beta_j}{\eta_j}\right)} \\ &\overset{(c)}{\doteq} e^{-L \min_{\boldsymbol{\beta}\in\mathcal{S}_T} \sum_{j=1}^{J}\gamma_j u_j\left(\frac{\beta_j}{\eta_j}\right)} \\ &\overset{(d)}{\doteq} e^{-L \sum_{j=1}^{J}\gamma_j u_j\left(\frac{\beta_j^\star}{\eta_j}\right)} \end{aligned} \tag{8}$$

where (a) follows from Lemma 3, (b) follows from the fact that $u_j(\alpha)$ is a strictly increasing function of α, for $\alpha > \mathbb{E}\{x_j\}$, and (c) can be proved as follows. Let us denote the vector which minimizes the exponent over the set $\mathcal{S}_O$ as $\hat{\boldsymbol{\beta}}^\star$. Since $\mathcal{S}_T$ is a subset of $\mathcal{S}_O$, $\hat{\boldsymbol{\beta}}^\star$ is either in $\mathcal{S}_T$ or in $\mathcal{S}_O - \mathcal{S}_T$. In the former case, (c) is obviously valid. When $\hat{\boldsymbol{\beta}}^\star \in \mathcal{S}_O - \mathcal{S}_T$, we can prove that $0 \leq \hat{\beta}_j^\star \leq \eta_j\mathbb{E}\{x_j\}$, for all $1 \leq j \leq J$, by contradiction. Let us assume the opposite is true, i.e., there is at least one index $1 \leq j \leq J$ such that $0 \leq \hat{\beta}_j^\star \leq \eta_j\mathbb{E}\{x_j\}$, and at least one other index $1 \leq k \leq J$ such that $\eta_k\mathbb{E}\{x_k\} < \hat{\beta}_k^\star$. Then, knowing that the derivative of of $u_j(\alpha)$ is zero for $\alpha = \mathbb{E}\{x_j\}$ and strictly positive for $\alpha > \mathbb{E}\{x_j\}$, a small increase in $\hat{\beta}_j^\star$ and an equal decrease in $\hat{\beta}_k^\star$ reduces the objective function, $\sum_{j=1}^{J}\gamma_j u_j\left(\frac{\beta_j}{\eta_j}\right)$, which contradicts the assumption that $\hat{\boldsymbol{\beta}}^\star$ is a minimum point. Knowing that $0 \leq \hat{\beta}_j^\star < \eta_j\mathbb{E}\{x_j\}$, for all $1 \leq j \leq J$, it is easy to show that the minimum value of the objective function is zero over $\mathcal{S}_O$, and $\mathcal{S}_T$ has to be an empty set. Defining the minimum value of the positive objective function as zero over an empty set ($\mathcal{S}_T$) makes (c) valid for the latter case where $\hat{\boldsymbol{\beta}}^\star \in \mathcal{S}_O - \mathcal{S}_T$. Finally, applying Lemma 4 results in (d) where $\boldsymbol{\beta}^\star$ is defined in the lemma.

Lemma 3. *For any continuous positive function $h(\mathbf{x})$ over a convex set $\mathcal{S}$, and defining $H(L)$ as*

$$H(L) = \int_{\mathcal{S}} e^{-h(\mathbf{x})L} d\mathbf{x}$$

we have

$$\lim_{L\to\infty} -\frac{\log(H(L))}{L} = \inf_{\mathcal{S}} h(\mathbf{x}) = \min_{cl(\mathcal{S})} h(\mathbf{x})$$

where $cl(\mathcal{S})$ denotes the closure of $\mathcal{S}$ (refer to [83] for the definition of the closure operator).

Proof of Lemma 3 can be found in appendix D.

Lemma 4. *There exists a unique vector $\boldsymbol{\beta}^\star$ with the elements $\beta_j^\star = \eta_j l_j^{-1}\left(\frac{\nu\eta_j}{\gamma_j}\right)$ which minimizes the convex function $\sum_{j=1}^{J}\gamma_j u_j(\frac{\beta_j}{\eta_j})$ over the convex set $\mathcal{S}_T$, where ν satisfies the following condition*

$$\sum_{j=1}^{J}\eta_j l_j^{-1}\left(\frac{\nu\eta_j}{\gamma_j}\right) = \alpha. \tag{9}$$

$l^{-1}()$ denotes the inverse of the function $l()$ defined in subsection III-A.

Proof of Lemma 4 can be found in appendix E.

Equation (8) is valid for any fixed value of $\boldsymbol{\eta}$. To achieve the most rapid decay of P_E, the exponent must be maximized over $\boldsymbol{\eta}$.

$$\lim_{L\to\infty} -\frac{\log P_E}{L} = \max_{0\leq\eta_j\leq 1}\sum_{j=1}^{J}\gamma_j u_j\left(\frac{\beta_j^\star}{\eta_j}\right) \tag{10}$$

where $\boldsymbol{\beta}^\star$ is defined for any value of the vector $\boldsymbol{\eta}$ in Lemma 4. Theorem 1 solves the maximization problem in (10) and identifies the asymptotically optimum RA. The proof can be found in appendix F.

Theorem 1. *Consider a point-to-point connection over the network with L independent paths from the source to the destination, with a large enough bandwidth constraint*[3]*. The paths are from J different types, L_j paths from the type j. Assume a block FEC of size $[N, K]$ is sent during a time interval T. Let N_j denote the number of packets in a block of size N assigned to the paths of type j, such that $\sum_{j=1}^{J} N_j = N$. The RA vector $\boldsymbol{\eta}$ is defined as $\eta_j = \frac{N_j}{N}$. For fixed values of $\gamma_j = \frac{L_j}{L}$, $n_0 = \frac{N}{L}$, $k_0 = \frac{K}{L}$, T and asymptotically large number of paths L, the optimum rate allocation vector $\boldsymbol{\eta}^\star$ equals to*

$$\eta_j^\star = \begin{cases} 0 & \textit{if } \alpha \leq \mathbb{E}\{x_j\} \\ \dfrac{\gamma_j l_j(\alpha)}{\displaystyle\sum_{i=1,\ \alpha>\mathbb{E}\{x_i\}}^{J} \gamma_i l_i(\alpha)} & \textit{otherwise} \end{cases} \tag{11}$$

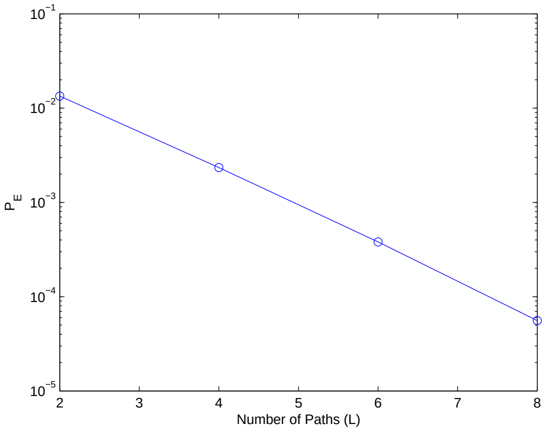
if there is at least one $1 \leq j \leq J$ for which $\alpha > \mathbb{E}\{x_j\}$. Furthermore, the probability of irrecoverable loss corresponding to $\boldsymbol{\eta}^\star$ decays as

$$P_E \doteq e^{-L\sum_{j=1}^{J}\gamma_j u_j(\alpha)}. \tag{12}$$

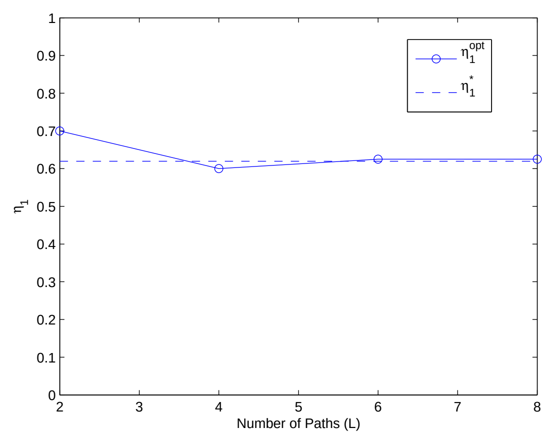
In the case where $\alpha \leq \mathbb{E}\{x_j\}$ for $1 \leq j \leq J$, $P_E \doteq 1$ independent of the allocation vector $\boldsymbol{\eta}$.

Remark 2. Theorem 1 can be interpreted as follows. For large values of L, adding a new type of path contributes to

[3]By the term 'large enough', we mean the bandwidth constraint on a path of type j, W_j, satisfies the condition $\frac{\eta_j n_0}{T\gamma_j} \leq W_j$. The reason is that η_j must satisfy both conditions of $0 \leq \eta_j \leq 1$ and $\frac{N_j}{TL_j} = \frac{\eta_j n_0 L}{T\gamma_j L} \leq W_j$, simultaneously. When W_j is large enough such that $\frac{\eta_j n_0}{T\gamma_j} \leq W_j$, the latter condition is automatically satisfied, and the optimization problem can be solved.



(a)



(b)

Fig. 5. (a) P_E versus L for the combination of two path types, half from type I and half from type II. (b) The normalized aggregated weight of type I paths in the optimal rate allocation (η_1^{opt}), compared with the value of η_1 which maximizes the exponent of equation (10) ($\eta_1^{\star}$).

the path diversity *iff* the path satisfies the quality constraint $\alpha > \mathbb{E}\{x\}$, where x is the percentage of time that the path spends in the bad state during the time interval $[0, T]$. Only in this case, adding the new type of path exponentially improves the performance of the system in terms of the probability of irrecoverable loss.

Remark 3. Observing the exponent coefficient corresponding to the optimum allocation vector $\boldsymbol{\eta}^{\star}$, we can see that the typical error event occurs when the ratio of the lost packets on all types of paths is the same as the total fraction of the lost packets, α. However, this is not the case for any arbitrary RA vector $\boldsymbol{\eta}$.

Remark 4. An interesting extension of Theorem 1 is the case where all types have identical erasure patterns ($u_j(x) = u_k(x)$ for $\forall\ 1 \leq j, k \leq J$ and $\forall x$), but different bandwidth constraints. Adopting the notation of Theorem 1, the bandwidth constraint on η_j can be written as $\frac{\eta_j n_0 L}{T \gamma_j L} \leq W_j$, where W_j is the maximum bandwidth for a path of type j. Let us define $\tilde{\boldsymbol{\eta}}^{\star}$ as the allocation vector which maximizes the objective function of equation (10), and satisfies the bandwidth constraints too. $\boldsymbol{\eta}^{\star}$ is the maximizing vector for the unconstrained problem, defined in Theorem 1. According to equation (11), we have $\eta_j^{\star} = \gamma_j$ for $\forall 1 \leq j \leq J$. It is obvious that $\tilde{\boldsymbol{\eta}}^{\star} = \boldsymbol{\eta}^{\star}$ if $\eta_j^{\star} \leq \frac{\gamma_j W_j T}{n_0}$ for all j. In case $\eta_j^{\star}$ does not satisfy the bandwidth constraint for some j, $\tilde{\boldsymbol{\eta}}^{\star}$ can be found by the water-filling algorithm. More accurately, we have

$$\tilde{\eta}_j^{\star} = \begin{cases} \dfrac{\gamma_j W_j T}{n_0} & \text{if } \tilde{\eta}_j^{\star} \leq \gamma_j \Upsilon \\ \gamma_j \Upsilon & \text{if } \tilde{\eta}_j^{\star} < \dfrac{\gamma_j W_j T}{n_0} \end{cases} \tag{13}$$

where Υ can be found by imposing the condition $\sum_{j=1}^{J} \tilde{\eta}_j^{\star} = 1$. Figure 6 depicts water-filling among identical paths with four different bandwidth constraints. Proof of equation (13) can be found in appendix G.

Example 2. Consider the scenario of transmitting a video stream with the DVD quality (using either MPEG-2 or MPEG-4) over multiple paths of two types. The number of paths for each type are equal, i.e. $\gamma_1 = \gamma_2 = 0.5$. The total number of paths varies from $L = 2$ to $L = 8$. Both type of paths are modeled as 2-state Gilbert channels with $\frac{1}{\mu_g} = 2500$ ms, in accordance with [6], [50]. Furthermore, the average bad burst duration are equal to $\frac{1}{\mu_{b_1}} = 50$ ms for the first type and $\frac{1}{\mu_{b_2}} = 100$ ms for the second type. The block transmission time is $T = 200$ ms which imposes an acceptable end-to-end delay for the video stream. The payload of each video packet is assumed to be 5 kb. The end-to-end rate increases linearly with L such that $\frac{S_{req}}{L} = 1$ Mbps. Hence, the block length equals to $N = 40L$. The coding overhead is $\alpha = 0.3$. Figure 5(a) shows P_E of the optimum RA versus L. The optimal RA, $\boldsymbol{\eta}^{opt}$, is found by exhaustive search among all possible allocation vectors. The figure depicts a linear behavior in semi-logarithmic scale with the exponent of 0.9137, which is comparable to 0.9256 predicted by (11).

In this scenario, let us denote $\eta_1^{\star}$ as the value of the first element of $\boldsymbol{\eta}^{\star}$, given in equation (11). Obviously, $\eta_1^{\star}$ does not depend on L. Moreover, η_1^{opt} is defined as the normalized aggregated weight of type I paths in the optimal RA. Figure 5(b) compares η_1^{opt} with $\eta_1^{\star}$ for different number of paths. It is observed that η_1^{opt} converges rapidly to $\eta_1^{\star}$ as L grows.

IV. Suboptimal Rate Allocation

In order to compute the complexity of the RA problem, we focus our attention on the original discrete formulation in subsection II-C. According to the model of subsection III-B, we assume the available paths are from J types, L_j paths from type j, such that $\sum_{j=1}^{J} L_j = L$. Obviously, all the paths from the same type should have equal rate. Therefore, the RA problem is turned into finding the vector $\mathbf{N} = (N_1, \ldots, N_J)$ such that $\sum_{j=1}^{J} N_j = N$, and $0 \leq N_j \leq L_j W_j T$ for all j. N_j denotes the number of packets assigned to all the paths of type j. Let us temporarily assume that all paths have enough bandwidth such that N_j can vary from 0 to N for all j. There are $\binom{N+J-1}{J-1}$ L-dimensional non-negative vectors of the form

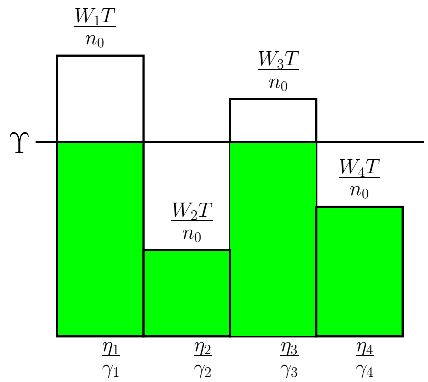


Fig. 6. WaterFilling algorithm over identical paths with four different bandwidth constraints.

$(N_1, \ldots, N_J)$ which satisfy the equation $\sum_{j=1}^{J} N_j = N$ each representing a distinct RA. Hence, the number of candidates is exponential in terms of J.

First, we prove the RA problem is NP [84] in the sense that P_E can be computed in polynomial time for any candidate vector $\mathbf{N} = (N_1, \ldots, N_J)$. Let us define $P_e^{\mathbf{N}}(k, j)$ as the probability of having more than k errors over the paths of types 1 to j for a specific allocation vector $\mathbf{N}$. We also define $Q_j(n, k)$ as the probability of having exactly k errors out of the n packets sent over the paths of type j. In appendices H and I, $Q_j(n, k)$'s are computed for any general M-state Markov channel model with polynomial complexity. Hence, we can assume that $Q_j(n, k)$'s are precomputed and stored for all n and k and path types. Then, the following recursive formula holds for $P_e^{\mathbf{N}}(k, j)$

$$P_e^{\mathbf{N}}(k,j) = \begin{cases} \sum_{i=0}^{N_j} Q_j(N_j, i) P_e^{\mathbf{N}}(k-i, j-1) & \text{if } k \geq 0 \\ 1 & \text{if } k < 0 \end{cases}$$
$$P_e^{\mathbf{N}}(k,1) = \sum_{i=k+1}^{N_1} Q_1(N_1, i). \tag{14}$$

To compute $P_e^{\mathbf{N}}(K, J)$ by the above recursive formula, we apply a well-known technique in the theory of algorithms called *memoization* [85]. Memoization works by storing the computed values of a recursive function in an array. By keeping this array in the memory, memoization avoids recomputing the function for the same arguments when it is called later. To compute $P_e^{\mathbf{N}}(K, J)$, an array of size $O(KJ)$ is required. This array should be filled with the values of $P_e^{\mathbf{N}}(k, j)$ for $0 < k \leq K$, and $1 \leq j \leq J$. Computing $P_e^{\mathbf{N}}(k, j)$ requires $O(K)$ operations assuming the values of $P_e^{\mathbf{N}}(i, j-1)$ and $Q_j(N_j, i)$ and $\sum_{i=k+1}^{N_j} Q_j(N_j, i)$ are already computed for $0 \leq i \leq k$. Thus, $P_e^{\mathbf{N}}(K, J)$ can be computed with the complexity of $O(K^2 J)$ if the values of $Q_j(N_j, k)$ are given for all N_j and $0 \leq k \leq K$. Following appendix I, we note that for each j, $Q_j(N_j, k)$ for $0 \leq k \leq K$ is computed offline with the complexity of $O(K^2 L_j) + O\left(M^2 \frac{N_j}{L_j} K\right)$. Hence, the total complexity of computing $P_e^{\mathbf{N}}(K, J)$ adds up to

$$\begin{aligned} & O(K^2 J) + \sum_{j=1}^{J} O\left(K^2 L_j + M^2 \frac{N_j}{L_j} K\right) \\ \stackrel{(a)}{=} \; & O(K^2 J) + \sum_{j=1}^{J} O\left(K^2 L_j + M^2 N_j K\right) \\ \stackrel{(b)}{=} \; & O\left(K^2 L + M^2 K N\right) \end{aligned} \tag{15}$$

where (a) follows from the fact that $\frac{N_j}{L_j} < N_j$, and the term $O(K^2 J)$ is omitted in (b) since we know that $J < L$.

Now, we propose a suboptimal polynomial time algorithm to estimate the best path allocation vector, $\mathbf{N}^{opt}$. Let us define $P_e^{opt}(n, k, j)$ as the probability of having more than k errors for a block of length n over the paths of types 1 to j minimized over all possible RA's ($\mathbf{N} = \mathbf{N}^{opt}$). First, we find a lowerbound $\hat{P}_e(n, k, j)$ for $P_e^{opt}(n, k, j)$ from the following recursive formula

$$\hat{P}_e(n,k,j) = \begin{cases} \min_{0 \leq n_j \leq \min\{n, \lfloor L_j W_j T \rfloor\}} \sum_{i=0}^{n_j} Q_j(n_j, i) \cdot \hat{P}_e(n - n_j, k - i, j - 1) & \text{if } k > 0 \\ 1 & \text{if } k \leq 0 \end{cases}$$
$$\hat{P}_e(n,k,1) = \sum_{i=k+1}^{n} Q_1(n, i). \tag{16}$$

Using memoization technique, we need an array of size $O(NKJ)$ to store the values of $\hat{P}_e(n, k, j)$ for $0 < n \leq N$, $0 < k \leq K$, and $1 \leq j \leq J$. According to the recursive definition above, computing $\hat{P}_e(n, k, j)$ requires $O(NK)$ operations assuming the values of $Q_j(n_j, i)$ and $\hat{P}_e(n - n_j, k - i, j - 1)$ and $\sum_{i=k+1}^{n_j} Q_j(n_j, i)$ are already computed for all i and n_j. Thus, it is easy to verify that $\hat{P}_e(N, K, J)$ can be computed with the complexity of $O(N^2 K^2 J)$ when the values of $Q_j(n_j, i)$ are given for all $0 < n_j \leq N$ and $0 \leq i \leq n_j$. According to appendix I, for each $1 \leq j \leq J$, $Q_j(n_j, i)$ can be computed for all $0 < n_j \leq N$ and $0 \leq i \leq n_j$ with the complexity of $O(N^3 L_j) + O(M^2 \frac{N^2}{L_j})$. Thus, computing $Q_j(n_j, i)$ for all $1 \leq j \leq J$, and $0 < n_j \leq N$, and $0 \leq i \leq n_j$, has the complexity of $\sum_{j=1}^{J} O(N^3 L_j) + O(M^2 \frac{N^2}{L_j}) = O(N^3 L + M^2 N^2 J)$. Finally, $\hat{P}_e(N, K, J)$ can be computed with the total complexity of $O(N^2 K^2 J + N^3 L + M^2 N^2 J)$.

The following lemma guarantees that $\hat{P}_e(n, k, j)$ is in fact a lowerbound for $P_e^{opt}(n, k, j)$. The proof is given in appendix J.

Lemma 5. $P_e^{opt}(n, k, j) \geq \hat{P}_e(n, k, j)$.

Algorithm 1 recursively finds a suboptimum allocation vector $\hat{\mathbf{N}}$ based on the lowerbound of Lemma 5.

Intuitively speaking, the proposed suboptimal algorithm recursively finds the typical error event (K_j's) which has the maximum contribution to the error probability, and assigns the RA ($\hat{N}_j$'s) such that the estimated typical error probability ($\hat{P}_e$) is minimized. Indeed, Lemma 5 shows that the estimate used in the algorithm ($\hat{P}_e$) is a lower-bound for the minimum achievable error probability (P_e^{opt}). Comparing (16) and the **while** loop in Algorithm 1, we observe that the values of

Algorithm 1 Proposed Suboptimal RA Algorithm

Input: $N,\ K,\ J,\ (L_1,\ldots,L_J),\ Q_j(\cdot,\cdot),\ \hat{P}_e(\cdot,\cdot,\cdot)$
Output: $(\hat{N}_1,\ldots,\hat{N}_J)$
  Initialize $j \leftarrow J$; $n \leftarrow N$; $k \leftarrow K$;
  while $j > 1$ and $k \geq 0$ **do**
    $\hat{N}_j \leftarrow \underset{0\leq n_j\leq \min\{n,\lfloor L_j W_j T\rfloor\}}{\operatorname{argmin}} \sum_{i=0}^{n_j} \hat{P}_e(n-n_j,k-i,j-1)\cdot Q_j(n_j,i)$;
    $K_j \leftarrow \underset{0\leq i\leq \hat{N}_j}{\operatorname{argmax}} Q_j(\hat{N}_j,i)\hat{P}_e(n-\hat{N}_j,k-i,j-1)$;
    Update $n \leftarrow n-\hat{N}_j$; $k \leftarrow k-K_j$; $j \leftarrow j-1$;
  end while
  for $m = 1$ to j **do**
    $\hat{N}_m \leftarrow \lfloor \frac{n}{j} \rfloor$;
  end for
  for $m = 1$ to $(n$ **mod** $j)$ **do**
    $\hat{N}_m \leftarrow \hat{N}_m + 1$;
  end for
  return $(\hat{N}_1,\ldots,\hat{N}_J)$;

$\hat{N}_j$ and K_j can be found in $O(1)$ during the computation of $\hat{P}_e(N,K,J)$. Hence, complexity of the proposed algorithm is the same as that of computing $\hat{P}_e(N,K,J)$ which is $O(N^2K^2J+N^3L+M^2N^2J)$.

The following theorem guarantees that the output of the above algorithm converges to the asymptotically optimal RA introduced in Theorem 1 of section III-B, and accordingly, it performs optimally for large number of paths. The proof can be found in appendix K.

Theorem 2. *Consider a point-to-point connection over the network with L independent paths from the source to the destination, each with a large enough bandwidth constraint. The paths are from J different types, L_j paths from the type j. Assume a block FEC of the size $[N,K]$ is sent during an interval time T. For fixed values of $\gamma_j = \frac{L_j}{L}$, $n_0 = \frac{N}{L}$, $k_0 = \frac{K}{L}$, T and asymptotically large number of paths (L) we have*

1) $\hat{P}_e(N,K,J) \doteq P_e^{opt}(N,K,J) \doteq e^{-L\sum_{j=1}^{J}\gamma_j u_j(\alpha)}$
2) $\frac{\hat{N}_j}{N} = \eta_j^{\star} + o(1)$
3) $\frac{K_j}{\hat{N}_j} = \alpha + o(1)$ *for* $\alpha > \mathbb{E}\{x_j\}$.

where $\alpha = \frac{k_0}{n_0}$ and $u_j()$ are defined in subsections III-A and III-B. $\hat{P}_e(N,K,J)$ is the lowerbound for $P_e^{opt}(n,k,j)$ defined in equation (16)*. $\hat{N}_j$ is the total number of packets assigned to the paths of type j by the suboptimal rate allocation algorithm. $\eta_j^{\star}$ is the asymptotically optimal RA given in equation* (11)*. K_j is also defined in Algorithm 1.*

Example 3. The proposed algorithm is compared with four other allocation schemes over $L = 4$ and $L = 3$ paths in Fig. 7. The optimal method uses exhaustive search over all possible allocations. '*Best Path Allocation*' assigns everything to the best path only, ignoring the rest. '*Equal Distribution*' scheme distributes the packets among all paths equally. Finally, the '*Asymptotically Optimal*' allocation assigns the rates based on equation (11). A DVD-quality video stream with the end-to-end rate of $S_{req} = 3.2$ Mbps is studied in both scenarios of Fig. 7. The block transmission time is $T = 250$ ms which imposes an acceptable end-to-end delay for the video stream. The payload of each packet is adopted to be 4 kb. Accordingly, the block length would be equal to $N = S_{req}T = 200$ packets. The FEC coding overhead is fixed at $\alpha = 0.2$. The paths follow the 2-state Gilbert model with $\frac{1}{\mu_g} = 2500$ ms. However, quality of the paths are different as they have different average bad burst durations: (a) In the case of 3 paths, the average bad burst of the paths ($\frac{1}{\mu_b}$'s) are listed as $[75\text{ ms}, 75\text{ ms} \pm \Delta]$; (b) In the case of 4 paths, the average bad burst of the paths ($\frac{1}{\mu_b}$'s) are listed as $[75\text{ ms} \pm \frac{\Delta}{2}, 75\text{ ms} \pm \frac{3\Delta}{2}]$; As observed, the median of $\frac{1}{\mu_b}$ of paths is fixed at 75 ms in both scenarios. Δ represents a measure of deviation from this median. $\Delta = 0$ describes the case where all the paths are identical. The larger is Δ, the more variety we have among the paths and the more diversity gain might be achieved using a judicious RA.

As seen, our suboptimal algorithm tracks the optimal algorithm so closely that the corresponding curves are not easily distinguishable in most cases. However, the '*Asymptotically Optimal*' RA results in lower performance since L is relatively small which makes the asymptotic analysis assumptions invalid. Comparing Fig. 7(a) and Fig. 7(b), it is observed that increasing L from 3 to 4 paths reduces the gap between the '*Asymptotically Optimal*' RA and the optimal RA considerably.

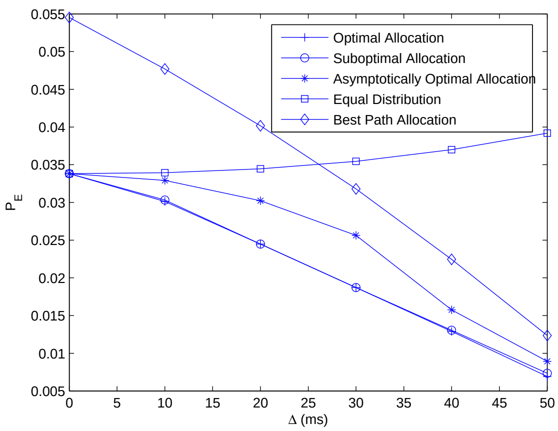
When $\Delta = 0$, the '*Equal Distribution*' scheme obviously coincides with the optimal allocation. This scheme eventually diverges from the optimal algorithm as Δ grows. However, it still outperforms the best path allocation method as long as Δ is not too large. For very large values of Δ, the best path dominates all the other ones, and we can ignore the rest of the paths. Hence, the best path allocation eventually converges to the optimal scheme when Δ increases.

V. Conclusion

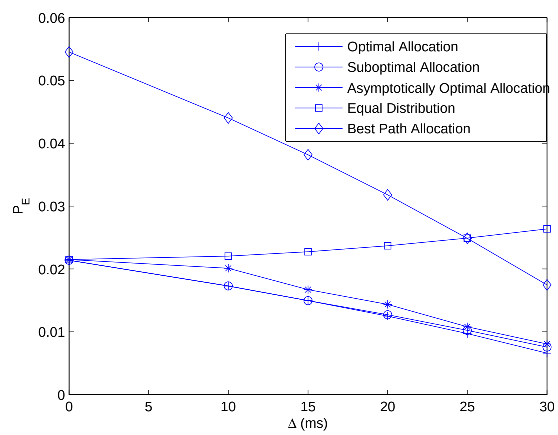
In this work, we have studied the performance of *Forward Error Correction* over a block of packets sent through multiple independent paths. Adopting MDS codes, the probability of irrecoverable loss (P_E) is shown to decay exponentially with the number of paths. Furthermore, the *rate allocation* (RA) problem across independent paths is studied. It is shown that in the asymptotically optimal RA, each path is assigned a positive rate *iff* its *quality* is above a certain threshold. Finally, the RA problem is studied for any arbitrary number of paths. A heuristic suboptimal algorithm is proposed which computes a near-optimal allocation in polynomial time. For large values of L, the result of this algorithm is shown to converge to the optimal RA. Simulation results verify the validity of the theoretical analyses in several practical scenarios and also show the near-optimal performance of the proposed suboptimal algorithm.

Appendix A
Probability Distribution of B_i

First, we compute the distribution of B_i for the 2-state Gilbert model. We denote the values of B_i with the parameter t to emphasize that they are expressed in the unit of time.



(a)



(b)

Fig. 7. Optimal and suboptimal RA's are compared with equal distribution and best path allocation schemes for different values of Δ: (a) $L=3$, (b) $L=4$.

Here, we focus on one path, for example path 1. Therefore, the index i can be temporarily dropped in analyzing the probability density function (pdf) of B_i.

We define the events g and b, respectively, as the channel being in the good or bad states at the start of a block. Then, the pdf of B can be written as

$$f_B(t) = f_{B|b}(t)\pi_b + f_{B|g}\pi_g. \tag{17}$$

Let $\mathcal{N}_s^T$ denote the number of consecutive states the channel experiences during the interval T. For instance, $\mathcal{N}_s^T = 3$ means that the channel switches its state twice in a block transmission time. Now, we define $f_{B|b}^m(t)$ as

$$f_{B|b}^m(t) = \lim_{\delta\to 0} \frac{\mathbb{P}\left\{t \le B < t+\delta \;\&\; \mathcal{N}_s^T = m|b\right\}}{\delta}. \tag{18}$$

$f_{B|g}^m(t)$ can be defined similarly.

For $m=1$, due to the memoryless nature of the exponential distribution, we have

$$\begin{aligned} f_{B|b}^1(t) &= \delta(t-T)e^{-\mu_b T} \\ f_{B|g}^1(t) &= \delta(t)e^{-\mu_g T}. \end{aligned} \tag{19}$$

For odd values of $m>1$, let τ_1 to τ_m denote the times the channel spends in different states. If the channel starts from the bad state, we have $\sum_{i=1}^{\frac{m+1}{2}} \tau_{2i-1} = t$ and $\sum_{i=1}^{\frac{m-1}{2}} \tau_{2i} = T-t$. Thus $f_{B|b}^m(t)$ can be written as

$$\begin{aligned} f_{B|b}^m(t) &= \int_{\mathcal{D}} \mu_b e^{-\mu_b\tau_1}\mu_g e^{-\mu_g\tau_2}\dots\mu_g e^{-\mu_g\tau_{m-1}} e^{-\mu_b\tau_m} \prod_{i=1}^{m-2}\tau_i \\ &= \mu_b^{\frac{m-1}{2}}\mu_b^{\frac{m-1}{2}} e^{-\mu_b t} e^{-\mu_g(T-t)} \triangle_{\frac{m-1}{2}}(t)\triangle_{\frac{m-3}{2}}(T-t) \end{aligned} \tag{20}$$

where $\mathcal{D}$ and $\triangle_k(t)$ are defined as

$$\begin{aligned} \mathcal{D} &\triangleq \left\{(\tau_1,\dots,\tau_m)\middle| \forall i: \tau_i > 0, \sum_{i=1}^{\frac{m+1}{2}} \tau_{2i-1} = t, \right. \\ &\qquad \left. \sum_{i=1}^{\frac{m-1}{2}} \tau_{2i} = T-t \right\}, \\ \triangle_k(t) &\triangleq \int_{z_i>0}\dots\int_{\sum_{i=1}^k z_i \le t} dz_1\dots dz_k. \end{aligned}$$

It is easy to observe that $\triangle_k(t)$ is the volume of a k-dimensional simplex with the edge of length t. By mathematical induction on k, it can be shown that $\triangle_k(t) = \frac{t^k}{k!}$. Therefore, making similar arguments for the even values of m, we have

$$f_{B|b}^m(t) = \begin{cases} \mu_g \dfrac{(\mu_b t)^{\frac{m-1}{2}} (\mu_g(T-t))^{\frac{m-3}{2}}}{\left(\frac{m-1}{2}\right)!\left(\frac{m-3}{2}\right)!} e^{-\mu_b t} e^{-\mu_g(T-t)} & \text{for } m \text{ odd} \\ \mu_b \dfrac{(\mu_b t \mu_g(T-t))^{\frac{m}{2}-1}}{\left(\frac{m}{2}-1\right)!\left(\frac{m}{2}-1\right)!} e^{-\mu_b t} e^{-\mu_g(T-t)} & \text{for } m \text{ even} \end{cases}$$

Based on similar arguments, $f_{B|g}^m(t)$ can be written as

$$f_{B|g}^m(t) = \begin{cases} \mu_b \dfrac{(\mu_b t)^{\frac{m-3}{2}} (\mu_g(T-t))^{\frac{m-1}{2}}}{\left(\frac{m-3}{2}\right)!\left(\frac{m-1}{2}\right)!} e^{-\mu_b t} e^{-\mu_g(T-t)} & \text{for } m \text{ odd} \\ \mu_g \dfrac{(\mu_b t \mu_g(T-t))^{\frac{m}{2}-1}}{\left(\frac{m}{2}-1\right)!\left(\frac{m}{2}-1\right)!} e^{-\mu_b t} e^{-\mu_g(T-t)} & \text{for } m \text{ even} \end{cases}$$

Having $f_{B|b}^m(t)$ and $f_{B|g}^m(t)$ for all m, we can write

$$\begin{aligned} f_{B|b}(t) &= \sum_{m=1}^{\infty} f_{B|b}^m(t) \\ f_{B|g}(t) &= \sum_{m=1}^{\infty} f_{B|g}^m(t). \end{aligned} \tag{21}$$

Combining the above equations with (17), $f_B(t)$ can be computed. Noting the factorial terms in the denominator of $f_{B|b}^m(t)$ and $f_{B|g}^m(t)$ and the fact that $\max\{t, T-t\} = T$ for $0 \le t \le T$, it can be verified that both $f_{B|b}^m(t)$ and $f_{B|g}^m(t)$ decrease very rapidly for $\frac{m-3}{2} > \max\{\mu_b T, \mu_g T\}$. Therefore, in the practical cases, we do not need to compute an infinte summation to get a close approximation of $f_B(t)$.

For the Extended Gilbert model, the pdf of B can be computed as follows. Here, equation (17) should be replaced with $f_B(t) = f_{B|g}(t)\pi_g + \sum_{i=1}^{M-1} f_{B|b_i}(t)\pi_{b_i}$. Moreover, for any specific sequence of state transitions $(\tau_1, \ldots, \tau_m)$ of length m, similar to the argument of equation (20), it can be shown that $f_{B|b_i}(\tau_1, \ldots, \tau_m)$ only depends on the summation of τ_i's which belong to the same state. Accordingly, similar to (21), $f_{B|b_i}(t)$ and $f_{B|g}(t)$ can be recomputed by summing over all lengths m and all state transition sequences of length m.

Appendix B
Proof of Lemma 1

1) We define the function $v(\lambda)$ as

$$v(\lambda) = \frac{\mathbb{E}\{xe^{\lambda x}\}}{\mathbb{E}\{e^{\lambda x}\}}. \tag{22}$$

Then, the first derivative of $v(\lambda)$ will be

$$\frac{\partial}{\partial \lambda} v(\lambda) = \frac{\mathbb{E}\{x^2 e^{\lambda x}\}\mathbb{E}\{e^{\lambda x}\} - [\mathbb{E}\{xe^{\lambda x}\}]^2}{[\mathbb{E}\{e^{\lambda x}\}]^2}. \tag{23}$$

According to Cauchy-Schwarz inequality, the following statement is always true for any two functions of $f()$ and $g()$

$$\left(\int_x f(x)g(x)dx\right)^2 < \int_x f^2(x)dx \int_x g^2(x)dx \tag{24}$$

unless $f(x) = Kg(x)$ for a constant K and all values of x. If we choose $f(x) = \sqrt{x^2 Q(x) e^{x\lambda}}$ and $g(x) = \sqrt{Q(x)e^{x\lambda}}$, they can not be proportional to each other for all values of x. Therefore, the numerator of equation (23) has to be strictly positive for all λ. Since the function $v(\lambda)$ is strictly increasing, it has an inverse $v^{-1}(\alpha)$ which is also strictly increasing. Moreover, the non-linear equation $v(\lambda) = \alpha$ has a unique solution of the form $\lambda = v^{-1}(\alpha) = l(\alpha)$.

2) To show that $l(\alpha = 0) = -\infty$, we prove an equivalent statement of the form $\lim_{\lambda\to-\infty} v(\lambda) = 0$. Since x is a random variable in the range $[0, 1]$ with the probability density function $Q(x)$, for any $0 < \epsilon < 1$, we can write

$$\begin{aligned}
\lim_{\lambda\to-\infty} v(\lambda) &= \lim_{\lambda\to-\infty} \frac{\int_0^\epsilon xQ(x)e^{x\lambda}dx + \int_\epsilon^1 xQ(x)e^{x\lambda}dx}{\int_0^1 Q(x)e^{x\lambda}dx} \\
&\leq \lim_{\lambda\to-\infty} \frac{\int_0^\epsilon xQ(x)e^{x\lambda}dx}{\int_0^\epsilon Q(x)e^{x\lambda}dx} + \frac{\int_\epsilon^1 xQ(x)dx}{\int_0^\epsilon Q(x)e^{(x-\epsilon)\lambda}dx} \\
&\overset{(a)}{=} \lim_{\lambda\to-\infty} \frac{\int_0^\epsilon xQ(x)e^{x\lambda}dx}{\int_0^\epsilon Q(x)e^{x\lambda}dx} \\
&\overset{(b)}{=} \lim_{\lambda\to-\infty} \frac{x_1 Q(x_1)e^{\lambda x_1}}{Q(x_2)e^{\lambda x_2}}
\end{aligned} \tag{25}$$

for some $x_1, x_2 \in [0, \epsilon]$. (a) follows from the fact that for $x \in [0, \epsilon]$, $(x-\epsilon)\lambda \to +\infty$ when $\lambda \to -\infty$, and (b) is a result of the mean value theorem for integration [86]. This theorem states that for every continuous function $f(x)$ in the interval $[a, b]$, we have

$$\exists\ x_0 \in [a, b] \quad s.t. \quad \int_a^b f(x)dx = f(x_0)[b-a]. \tag{26}$$

Equation (25) is valid for any arbitrary $0 < \epsilon < 1$. If we choose $\epsilon \to 0$, x_1 and x_2 are both squeezed in the interval $[0, \epsilon]$. Thus, we have

$$\lim_{\lambda\to-\infty} v(\lambda) \leq \lim_{\lambda\to-\infty} \lim_{\epsilon\to 0} \frac{x_1 Q(x_1)e^{\lambda x_1}}{Q(x_2)e^{\lambda x_2}} = \lim_{\epsilon\to 0} x_1 = 0 \tag{27}$$

Based on the distribution of x, $v(\lambda)$ is obviously non-negative for any λ. Hence, the inequality in (27) can be replaced by equality.

3) By observing that $v(\lambda = 0) = \mathbb{E}\{x\}$, it is obvious that $l(\alpha = \mathbb{E}\{x\}) = 0$.

4) To show that $l(\alpha = 1) = +\infty$, we prove the equivalent statement of the form $\lim_{\lambda\to+\infty} v(\lambda) = 1$. For any $0 < \epsilon < 1$ and $x \in [1-\epsilon, 1]$, $(x - 1 + \epsilon)\lambda \to +\infty$ when $\lambda \to +\infty$. Then, defining $\zeta = 1 - \epsilon$, we have

$$\lim_{\lambda\to+\infty} \frac{\int_0^\zeta xQ(x)e^{x\lambda}dx}{\int_0^1 Q(x)e^{x\lambda}dx} \leq \lim_{\lambda\to+\infty} \frac{\int_0^\zeta xQ(x)dx}{\int_\zeta^1 Q(x)e^{(x-\zeta)\lambda}dx} = 0. \tag{28}$$

Since the fraction in (28) is obviously non-negative for all λ, this inequality can be replaced by an equality. Similarly, we have

$$\lim_{\lambda\to+\infty} \frac{\int_0^\zeta Q(x)e^{x\lambda}dx}{\int_\zeta^1 xQ(x)e^{x\lambda}dx} \leq \lim_{\lambda\to+\infty} \frac{\int_0^\zeta Q(x)dx}{\int_\zeta^1 xQ(x)e^{(x-\zeta)\lambda}dx} = 0. \tag{29}$$

which can also be replaced by equality. Now, the limit of $v(\lambda)$ is written as

$$\begin{aligned}
\lim_{\lambda\to+\infty} v(\lambda) &= \lim_{\lambda\to+\infty} \frac{\int_0^\zeta xQ(x)e^{x\lambda}dx + \int_\zeta^1 xQ(x)e^{x\lambda}dx}{\int_0^1 Q(x)e^{x\lambda}dx} \\
&\overset{(a)}{=} \lim_{\lambda\to+\infty} \frac{\int_\zeta^1 xQ(x)e^{x\lambda}dx}{\int_0^1 Q(x)e^{x\lambda}dx} \\
&\overset{(b)}{=} \left(\lim_{\lambda\to+\infty} \frac{\int_0^\zeta Q(x)e^{x\lambda}dx + \int_\zeta^1 Q(x)e^{x\lambda}dx}{\int_\zeta^1 xQ(x)e^{x\lambda}dx}\right)^{-1} \\
&\overset{(c)}{=} \left(\lim_{\lambda\to+\infty} \frac{\int_\zeta^1 Q(x)e^{x\lambda}dx}{\int_\zeta^1 xQ(x)e^{x\lambda}dx}\right)^{-1} \\
&\overset{(d)}{=} \left(\lim_{\lambda\to+\infty} \frac{Q(x_1)e^{x_1\lambda}}{x_2 Q(x_2)e^{x_2\lambda}}\right)^{-1}
\end{aligned} \tag{30}$$

for some $x_1, x_2 \in [1-\epsilon, 1]$. (a) follows from equation (28), and (b) is valid since the final result shows that $\lim_{\lambda\to+\infty} v(\lambda)$ is finite and non-zero [86]. (c) follows from equation (29), and (d) is a result of the mean value theorem for integration. If we choose $\epsilon \to 0$, x_1 and x_2 are both squeezed in the interval $[1-\epsilon, 1]$. Then, equation (30) turns into

$$\lim_{\lambda\to+\infty} v(\lambda) = \left(\lim_{\lambda\to+\infty} \lim_{\epsilon\to 0} \frac{Q(x_1)e^{x_1\lambda}}{x_2 Q(x_2)e^{x_2\lambda}}\right)^{-1} = \left(\lim_{\epsilon\to 0} \frac{1}{x_2}\right)^{-1} = 1.$$

5) According to equations (4) and (5), the first derivative of

$u(\alpha)$ is

$$\frac{\partial u(\alpha)}{\partial \alpha} = l(\alpha) + \alpha \frac{\partial l(\alpha)}{\partial \alpha} - \frac{\mathbb{E}\{x e^{\lambda x}\}}{\mathbb{E}\{e^{\lambda x}\}} \frac{\partial l(\alpha)}{\partial \alpha} = l(\alpha).$$

APPENDIX C
PROOF OF LEMMA 2

Based on the definition of probability density function, we have

$$\begin{aligned}
&\lim_{L\to\infty} -\frac{1}{L}\log\left(f_y(\alpha)\right) \\
&= \lim_{L\to\infty} -\frac{1}{L}\log\left(\lim_{\delta\to 0}\frac{\mathbb{P}\{y>\alpha\}-\mathbb{P}\{y>\alpha+\delta\}}{\delta}\right) \\
&\stackrel{(a)}{=} \lim_{\delta\to 0}\lim_{L\to\infty} -\frac{1}{L}\log\left(\frac{\mathbb{P}\{y>\alpha\}-\mathbb{P}\{y>\alpha+\delta\}}{\delta}\right) \\
&\geq \lim_{\delta\to 0}\lim_{L\to\infty} \frac{1}{L}\left(-\log\left(\mathbb{P}\{y>\alpha\}\right)+\log\delta\right) \\
&\stackrel{(b)}{=} u(\alpha) \qquad (31)
\end{aligned}$$

where (a) is valid since $\log$ is a continuous function, and both limitations do exist and are interchangeable. (b) follows from equation (6). The exponent of $f_y(\alpha)$ can be upper-bounded as

$$\begin{aligned}
&\lim_{L\to\infty} -\frac{1}{L}\log\left(f_y(\alpha)\right) \\
&\stackrel{(a)}{=} \lim_{\delta\to 0}\lim_{L\to\infty} \frac{-\log\left(\mathbb{P}\{y>\alpha\}-\mathbb{P}\{y>\alpha+\delta\}\right)+\log\delta}{L} \\
&\stackrel{(b)}{\leq} \lim_{\delta\to 0}\lim_{L\to\infty} \frac{-\log\left(e^{-L(u(\alpha)+\epsilon)}-e^{-L(u(\alpha+\delta)-\epsilon)}\right)+\log\delta}{L} \\
&= \lim_{\delta\to 0}\lim_{L\to\infty} u(\alpha)+\epsilon-\frac{\log\left(1-e^{-L\chi}\right)}{L} \\
&\stackrel{(c)}{=} u(\alpha)+\epsilon \qquad (32)
\end{aligned}$$

where $\chi = u(\alpha+\delta) - u(\alpha) - 2\epsilon$. Since $u(\alpha)$ is a strictly increasing function (Lemma 1), we can make χ positive by choosing ϵ small enough. (a) is valid since $\log$ is a continuous function, and both limits do exist and are interchangeable. (b) follows from the definition of limit if L is sufficiently large, and (c) is a result of χ being positive. Selecting ϵ arbitrarily small, results (31) and (32) prove the lemma.

APPENDIX D
PROOF OF LEMMA 3

According to the definition of infimum, we have

$$\begin{aligned}
&\lim_{L\to\infty} -\frac{\log(H(L))}{L} \\
&\geq \lim_{L\to\infty} -\frac{1}{L}\log\left(e^{-L\inf_{\mathcal{S}} h(\mathbf{x})}\int_{\mathcal{S}} d\mathbf{x}\right) \\
&\stackrel{(a)}{=} \inf_{\mathcal{S}} h(\mathbf{x}). \qquad (33)
\end{aligned}$$

where (a) follows from the fact that $\mathcal{S}$ is a bounded region. Since $h(\mathbf{x})$ is a continuous function, it has a minimum in the bounded closed set $cl(\mathcal{S})$ which is denoted by $\mathbf{x}^\star$. Due to the continuity of $h(\mathbf{x})$ at $\mathbf{x}^\star$, for any $\epsilon > 0$, there is a neighborhood $\mathcal{B}(\epsilon)$ centered at $\mathbf{x}^\star$ such that any $\mathbf{x} \in \mathcal{B}(\epsilon)$ has the property of $|h(\mathbf{x}) - h(\mathbf{x}^\star)| < \epsilon$. Moreover, since $\mathcal{S}$ is a convex set, we have vol $(\mathcal{B}(\epsilon)\cap\mathcal{S}) > 0$. Now, we can write

$$\begin{aligned}
&\lim_{L\to\infty} -\frac{\log(H(L))}{L} \\
&\leq \lim_{L\to\infty} -\frac{1}{L}\log\left(\int_{\mathcal{S}\cap\mathcal{B}(\epsilon)} e^{-Lh(\mathbf{x})}d\mathbf{x}\right) \\
&\leq \lim_{L\to\infty} -\frac{1}{L}\log\left(e^{-L(h(\mathbf{x}^\star)+\epsilon)}\int_{\mathcal{S}\cap\mathcal{B}(\epsilon)} d\mathbf{x}\right) \\
&= h(\mathbf{x}^\star)+\epsilon. \qquad (34)
\end{aligned}$$

Selecting ϵ to be arbitrarily small, (33) and (34) prove the lemma.

APPENDIX E
PROOF OF LEMMA 4

According to Lemma 1, $u_j(x)$ is increasing and convex for $\forall 1 \leq j \leq J$. Thus, the objective function $f(\boldsymbol{\beta}) = \sum_{j=1}^{J}\gamma_j u_j(\frac{\beta_j}{\eta_j})$ is also convex, and the region $\mathcal{S}_T$ is determined by J convex inequality constraints and one affine equality constraint. Hence, in this case, KKT conditions are both necessary and sufficient for optimality [87]. In other words, if there exist constants ϕ_j and ν such that

$$\frac{\gamma_j}{\eta_j} l_j(\frac{\beta_j^\star}{\eta_j}) - \phi_j - \nu = 0 \quad \forall 1 \leq j \leq J \qquad (35)$$

$$\phi_j\left[\eta\mathbb{E}\{x_j\} - \beta_j^\star\right] = 0 \quad \forall 1 \leq j \leq J \qquad (36)$$

then the point $\boldsymbol{\beta}^\star$ is a global minimum.

Now, we prove that either $\beta_j^\star = \eta_j\mathbb{E}\{x_j\}$ for all $1 \leq j \leq J$, or $\beta_j^\star > \eta_j\mathbb{E}\{x_j\}$ for all $1 \leq j \leq J$. Let us assume the opposite is true, and there are at least two elements of the vector $\boldsymbol{\beta}^\star$, indexed with k and m, which have the values of $\beta_k^\star = \eta_k\mathbb{E}\{x_k\}$ and $\beta_m^\star > \eta_m\mathbb{E}\{x_m\}$, respectively. For any arbitrary $\epsilon > 0$, the vector $\boldsymbol{\beta}^{\star\star}$ can be defined as below

$$\beta_j^{\star\star} = \begin{cases} \beta_j^\star + \epsilon & \text{if } j = k \\ \beta_j^\star - \epsilon & \text{if } j = m \\ \beta_j^\star & \text{otherwise.} \end{cases} \qquad (37)$$

Then, we have

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{f(\boldsymbol{\beta}^{\star\star}) - f(\boldsymbol{\beta}^\star)}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\gamma_k u_k\left(\frac{\beta_k^\star+\epsilon}{\eta_k}\right) + \gamma_m u_m\left(\frac{\beta_m^\star-\epsilon}{\eta_m}\right)\right. \\
&\qquad \left. -\gamma_m u_m\left(\frac{\beta_m^\star}{\eta_k}\right)\right\} \\
&\stackrel{(a)}{=} \lim_{\epsilon\to 0}\frac{\gamma_k}{\eta_k} l_k\left(\frac{\beta_k^\star+\epsilon'}{\eta_k}\right) - \frac{\gamma_m}{\eta_m} l_m\left(\frac{\beta_m^\star+\epsilon''}{\eta_m}\right) \\
&= -\frac{\gamma_m}{\eta_m} l_m\left(\frac{\beta_m^\star}{\eta_m}\right) < 0 \qquad (38)
\end{aligned}$$

where $\epsilon', \epsilon'' \in [0, \epsilon]$, and (a) follows from the Taylor's theorem. Thus, moving from $\boldsymbol{\beta}^\star$ to $\boldsymbol{\beta}^{\star\star}$ decreases the function which contradicts the assumption of $\boldsymbol{\beta}^\star$ being the global minimum.

Out of the remaining possibilities, the case where $\beta_j^\star = \eta_j\mathbb{E}\{x_j\}$ $(\forall 1 \leq j \leq J)$ obviously agrees with Lemma 4 for the special case of $\nu = 0$. Therefore, the lemma can be proved assuming $\beta_j^\star > \eta_j \mathbb{E}\{x_j\}$ ($\forall 1 \leq j \leq J$). Then, equation (36) turns into $\phi_j = 0$ ($\forall 1 \leq j \leq J$). By rearranging equation (35) and using the condition $\sum_{j=1}^{J} \beta_j = \alpha$, Lemma 4 is proved.

APPENDIX F
PROOF OF THEOREM 1

Sketch of the proof: First, it is proved that $\eta_j^\star > 0$ if $\mathbb{E}\{x_j\} < \alpha$. At the second step, we prove that $\eta_j^\star = 0$, if $\mathbb{E}\{x_j\} \geq \alpha$. Then, KKT conditions [87] are applied for the indices $1 \leq k \leq J$ where $\mathbb{E}\{x_k\} < \alpha$ to find the maximizing allocation vector, $\boldsymbol{\eta}^\star$.

Proof: The parameter ν is obviously a function of the vector $\boldsymbol{\eta}$. Differentiating equation (9) with respect to η_k results in

$$\frac{\partial \nu}{\partial \eta_k} = -\frac{v_k\left(\frac{\nu\eta_k}{\gamma_k}\right) + \frac{\nu\eta_k}{\gamma_k} v_k'\left(\frac{\nu\eta_k}{\gamma_k}\right)}{\sum_{j=1}^{J} \frac{\eta_j^2}{\gamma_j} v_j'\left(\frac{\nu\eta_j}{\gamma_j}\right)} \tag{39}$$

where $v_j(x) = l_j^{-1}(x)$, and $v_j'(x)$ denotes its derivative with respect to its argument. The objective function can be simplified as

$$g(\boldsymbol{\eta}) \triangleq \sum_{j=1}^{J} \gamma_j u_j\left(\frac{\beta_j^\star}{\eta_j}\right) = \sum_{j=1}^{J} \gamma_j u_j\left(v_j\left(\frac{\nu\eta_j}{\gamma_j}\right)\right). \tag{40}$$

$\nu^\star$ is defined as the value of ν corresponding to $\boldsymbol{\eta}^\star$. Next, we show that $\nu^\star > 0$. Let us assume the opposite is true, i.e., $\nu^\star \leq 0$. Then, according to Lemma 1, we have $v_j(\frac{\nu^\star\eta_j}{\gamma_j}) \leq \mathbb{E}\{x_j\}$ for all j which results in $g(\boldsymbol{\eta}^\star) = 0$. However, it is possible to achieve a positive value of $g(\boldsymbol{\eta})$ by setting $\eta_j = 1$ for the one vector which has the property of $\mathbb{E}\{x_j\} < \alpha$, and setting $\eta_j = 0$ for the rest. Thus, $\boldsymbol{\eta}^\star$ can not be the maximal point. This contradiction proves the fact that $\nu^\star > 0$.

At the first step, we prove that $\eta_j^\star > 0$ if $\mathbb{E}\{x_j\} < \alpha$. Assume the opposite is true for an index $1 \leq k \leq J$. Since $\sum_{j=1}^{J} \eta_j^\star = 1$, there should be at least one index m such that $\eta_m^\star > 0$. For any arbitrary $\epsilon > 0$, the vector $\boldsymbol{\eta}^{\star\star}$ can be defined as below

$$\eta_j^{\star\star} = \begin{cases} \epsilon & \text{if } j = k \\ \eta_j^\star - \epsilon & \text{if } j = m \\ \eta_j^\star & \text{otherwise.} \end{cases} \tag{41}$$

$\nu^{\star\star}$ is defined as the corresponding value of ν for the vector $\boldsymbol{\eta}^{\star\star}$. Based on equation (39), we can write

$$\begin{aligned} &\Delta\nu = \\ &\nu^{\star\star} - \nu^\star = \\ &\frac{v_m\left(\frac{\nu^\star\eta_m^\star}{\gamma_m}\right) + \frac{\nu^\star\eta_m^\star}{\gamma_m} v_m'\left(\frac{\nu^\star\eta_m^\star}{\gamma_m}\right) - \mathbb{E}\{x_k\}}{\sum_{j=1}^{J} \frac{\eta_j^{\star 2}}{\gamma_j} v_j'\left(\frac{\nu^\star\eta_j^\star}{\gamma_j}\right)} \epsilon + O(\epsilon^2). \end{aligned} \tag{42}$$

Then, we have

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{g(\boldsymbol{\eta}^{\star\star}) - g(\boldsymbol{\eta}^\star)}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left\{ \frac{\nu^{\star 2}\eta_k^\star}{\gamma_k} v_k'\left(\frac{\nu^\star\eta_k^\star}{\gamma_k}\right)\epsilon - \frac{\nu^{\star 2}\eta_m^\star}{\gamma_m} v_m'\left(\frac{\nu^\star\eta_m^\star}{\gamma_m}\right)\epsilon \right. \\ &\qquad \left. + \nu^\star \Delta\nu \sum_{j=1}^{J} \frac{\eta_j^{\star 2}}{\gamma_j} v_j'\left(\frac{\nu^\star\eta_j^\star}{\gamma_j}\right) + O(\epsilon^2) \right\} \\ &\overset{(a)}{=} \nu^\star \left\{ v_m\left(\frac{\nu^\star\eta_m^\star}{\gamma_m}\right) - \mathbb{E}\{x_k\} \right\} \end{aligned} \tag{43}$$

where (a) follows from (42). If the value of (43) is positive for an index m, moving in that direction increases the objective function which contradicts with the assumption of $\boldsymbol{\eta}^\star$ being a maximal point. If the value of (43) is non-positive for all indexes m whose $\eta_m^\star > 0$, we can write

$$\mathbb{E}\{x_k\} \geq \sum_{m=1}^{J} \eta_m^\star v_m\left(\frac{\nu^\star\eta_m^\star}{\gamma_m}\right) = \alpha \tag{44}$$

which obviously contradicts the assumption of $\mathbb{E}\{x_k\} < \alpha$.

At the second step, we prove that $\eta_j^\star = 0$ if $\mathbb{E}\{x_j\} \geq \alpha$. Assume the opposite is true for an index $1 \leq r \leq J$. Since $\sum_{j=1}^{J} \eta_j^\star = 1$, we should have $\eta_s^\star < 1$ for all other indices s. For any arbitrary $\epsilon > 0$, the vector $\boldsymbol{\eta}^{\star\star\star}$ can be defined as

$$\eta_j^{\star\star\star} = \begin{cases} \eta_j^\star - \epsilon & \text{if } j = r \\ \eta_j^\star + \epsilon & \text{if } j = s \\ \eta_j^\star & \text{otherwise.} \end{cases} \tag{45}$$

$\nu^{\star\star\star}$ is defined as the corresponding value of ν for the vector $\boldsymbol{\eta}^{\star\star\star}$. Based on equation (39), we can write

$$\begin{aligned} \Delta\nu &= \nu^{\star\star\star} - \nu^\star \\ &= \frac{\epsilon}{\sum_{j=1}^{J} \frac{\eta_j^{\star 2}}{\gamma_j} v_j'\left(\frac{\nu^\star\eta_j^\star}{\gamma_j}\right)} \left\{ v_r\left(\frac{\nu^\star\eta_r^\star}{\gamma_r}\right) + \frac{\nu^\star\eta_r^\star}{\gamma_r} v_r'\left(\frac{\nu^\star\eta_r^\star}{\gamma_r}\right) \right. \\ &\quad \left. - v_s\left(\frac{\nu^\star\eta_s^\star}{\gamma_s}\right) - \frac{\nu^\star\eta_s^\star}{\gamma_s} v_s'\left(\frac{\nu^\star\eta_s^\star}{\gamma_s}\right) \right\} + O(\epsilon^2). \end{aligned} \tag{46}$$

Then, we have

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{g(\boldsymbol{\eta}^{\star\star\star}) - g(\boldsymbol{\eta}^\star)}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left\{ \frac{\nu^{\star 2}\eta_s^\star}{\gamma_s} v_s'\left(\frac{\nu^\star\eta_s^\star}{\gamma_s}\right)\epsilon - \frac{\nu^{\star 2}\eta_r^\star}{\gamma_r} v_r'\left(\frac{\nu^\star\eta_r^\star}{\gamma_r}\right)\epsilon \right. \\ &\qquad \left. + \nu^\star \Delta\nu \sum_{j=1}^{J} \frac{\eta_j^{\star 2}}{\gamma_j} v_j'\left(\frac{\nu^\star\eta_j^\star}{\gamma_j}\right) + O(\epsilon^2) \right\} \\ &\overset{(a)}{=} \nu^\star \left\{ v_r\left(\frac{\nu^\star\eta_r^\star}{\gamma_r}\right) - v_s\left(\frac{\nu^\star\eta_s^\star}{\gamma_s}\right) \right\} \end{aligned} \tag{47}$$

where (a) follows from (46). If the value of (47) is positive for an index s, moving in that direction increases the objective function which contradicts with the assumption of $\boldsymbol{\eta}^\star$ being a maximal point . If the value of (47) is non-positive for all indices s whose $\eta_s^\star > 0$, we can write

$$\mathbb{E}\{x_r\} < v_r\left(\frac{\nu^\star\eta_r^\star}{\gamma_r}\right) \leq \sum_{s=1}^{J} \eta_s^\star v_s\left(\frac{\nu^\star\eta_s^\star}{\gamma_s}\right) = \alpha \tag{48}$$

which obviously contradicts the assumption of $\mathbb{E}\{x_r\} \geq \alpha$.

Now that the boundary points are checked, we can safely use the KKT conditions [87] for all $1 \leq k \leq J$, where $\mathbb{E}\{x_k\} < \alpha$, to find the maximizing allocation vector, $\boldsymbol{\eta}^\star$.

$$\zeta = \frac{\nu^{\star 2}\eta_k^\star}{\gamma_k} v_k' \left(\frac{\nu^\star \eta_k^\star}{\gamma_k}\right) + \nu^\star \sum_{j=1}^{J} \frac{\eta_j^{\star 2}}{\gamma_j} v_j' \left(\frac{\nu^\star \eta_j^{\star 2}}{\gamma_j}\right) \frac{\partial \nu}{\partial \eta_k}|_{\nu=\nu^\star}$$
$$\overset{(a)}{=} -\nu^\star v_k \left(\frac{\nu^\star \eta_k^\star}{\gamma_k}\right) \quad (49)$$

where ζ is a constant independent of k, and (a) follows from (39). Using the fact that $\sum_{j=1}^{J} \eta_j = 1$ together with equations (9) and (49) results in

$$\zeta = -\alpha\nu^\star$$
$$\nu^\star = \sum_{\mathbb{E}\{x_j\}<\alpha} \gamma_j l_j(\alpha). \quad (50)$$

Combining equations (49) and (50) results in equation (11) and $g(\boldsymbol{\eta}^\star) = \sum_{j=1}^{J} \gamma_j u_j(\alpha)$.

APPENDIX G
PROOF OF REMARK 4

Based on the arguments similar to the ones in appendix F, it can be shown that $\tilde{\eta}_j^\star = 0$ *iff* $\mathbb{E}\{x_j\} \geq \alpha$. Since all the types are identical here, this means $\tilde{\eta}_j^\star > 0$ for all j. Similar to equation (49), applying KKT conditions [87], gives us

$$v_j \left(\frac{\tilde{\nu}^\star \tilde{\eta}_j^\star}{\gamma_j}\right) = \begin{cases} -\zeta & \text{if } \tilde{\eta}_j^\star < \dfrac{\gamma_j W_j T}{n_0} \\ \\ -\zeta - \sigma_j & \text{if } \tilde{\eta}_j^\star = \dfrac{\gamma_j W_j T}{n_0} \end{cases} \quad (51)$$

where σ_j's are non-negative parameters [87]. Putting $\Upsilon = \frac{l_j(-\zeta)}{\tilde{\nu}^\star}$ proves equation (13).

APPENDIX H
DISCRETE ANALYSIS OF ONE PATH

$Q(n,k,l)$ is defined as the probability of having exactly k errors out of the n packets sent over the path l. To compute $Q(n,k,l)$ for any general M-state Markov model, the following parameters are required: 1) a $M \times M$ matrix $\mathbf{\Pi}$ with the elements $\pi_{s'|s}$ which represents the channel transition behavior. $\pi_{s'|s}$ is the probability of the channel being in the state s' provided that it has been in the state s when the last packet was transmitted; 2) a vector $\mathbf{q} = (q_1, \ldots, q_M)$ where q_s denotes the probability of having erasure conditioned on being in the state s.

For $\forall\ s \in \{1, \ldots, M\}$, π_s is defined as the steady state probability of being in the state s. Obviously, the steady state probability vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$ can be computed using the equation set $\boldsymbol{\pi} = \mathbf{\Pi}\boldsymbol{\pi}$ and $\sum_{s=1}^{M} \pi_s = 1$.

Depending on the initial state of the path l, $P_s(n,k,l)$ is defined as the probability of having k errors out of the n packets sent over this path when we start the transmission in the state s. It is easy to see that

$$Q(n,k,l) = \sum_{s=1}^{M} \pi_s P_s(n,k,l). \quad (52)$$

$P_s(n,k,l)$ can be computed from the following recursive equation

$$P_s(n,k,l) = \sum_{s'=1}^{M} q_s \pi_{s'|s} P_{s'}(n-1,k-1,l) + \sum_{s'=1}^{M} (1-q_s)\, \pi_{s'|s} P_{s'}(n-1,k,l) \quad (53)$$

with the initial conditions

$$P_s(n,k,l) = 0 \qquad \text{for } k > n$$
$$P_s(n,k,l) = 0 \qquad \text{for } k < 0$$
$$P_s(n,k,l) = 1 \qquad \text{for } k = n = 0. \quad (54)$$

According to the recursive equations in (53), to compute $P_s(n,k,l)$ by memoization technique, the functions $P_s()$ should be calculated at the following set of points denoted as $\mathcal{S}(n,k)$

$$\mathcal{S}(n,k) = \{(n',k') \mid 0 \leq k' \leq k,\ n'-n+k \leq k' \leq n'\}.$$

Cardinality of the set $\mathcal{S}(n,k)$ is of the order $|\mathcal{S}(n,k)| = O(k(n-k))$. Since $O(M)$ operations are needed to compute the recursive functions $P_s()$ at each point and M functions $P_s(n,k,l)$ $(s = 1, \ldots, M)$ have to be computed, $P_s(n,k,l)$ is computable with the complexity of $O\left(M^2 k(n-k)\right)$ which give us $Q(n,k,l)$ according to equation (52). It is worth mentioning that if the M-state extended Gilbert model is adopted, the computational complexity of obtaining $Q(n,k,l)$ would be reduced to $O(Mk(n-k))$.

APPENDIX I
DISCRETE ANALYSIS OF ONE TYPE

When there are n packets to be distributed over L_j identical paths of type j, uniform distribution is obviously the optimum. However, since the integer n may be indivisible by L_j, the L_j dimensional vector $\mathbf{N}$ is selected as

$$N_l = \begin{cases} \lfloor \dfrac{n}{L_j} \rfloor + 1 & \text{for } 1 \leq l \leq \text{Rem}(n, L_j) \\ \\ \lfloor \dfrac{n}{L_j} \rfloor & \text{for Rem}(n, L_j) < l \leq L_j \end{cases} \quad (55)$$

where Rem(a,b) denotes the remainder of dividing a by b. $\mathbf{N}$ represents the closest integer vector to a uniform distribution.

$E^{\mathbf{N}}(k,l)$ is defined as the probability of having exactly k erasures among the n packets transmitted over the identical paths 1 to l with the allocation vector $\mathbf{N}$. According to the definitions of $Q_j(n,k)$ and $E^{\mathbf{N}}(k,l)$, it is obvious that $Q_j(n,k) = E^{\mathbf{N}}(k, L_j)$. $E^{\mathbf{N}}(k,l)$ can be computed recursively as

$$E^{\mathbf{N}}(k,l) = \sum_{i=0}^{k} E^{\mathbf{N}}(k-i, l-1) Q(N_l, i, l)$$
$$E^{\mathbf{N}}(k,1) = Q(N_1, k, 1) \quad (56)$$

where $Q(N_l, i, l)$ is given in appendix H. Since all the paths are assumed to be identical here, $Q(N_l, k, l)$ is the same for all path indices, l. According to the recursive equations in (53), the values of $Q(N_l,i,l)$ for all $0 \le i \le k$ and $1 \le l \le L_j$ can be calculated with the complexity of $O(M^2 N_l k) = O\left(M^2 \frac{n}{L_j} k\right)$. According to the recursive equations in (56), computing $E^{\mathbf{N}}(k,l)$ requires memoization over an array of size $O(kl)$ whose entries can be calculated with $O(k)$ operations each. Thus, $E^{\mathbf{N}}(k,l)$ is computable with the complexity of $O(k^2 l)$ if $Q(N_l,i,l)$'s are already given. Finally, noting that $Q_j(n,k) = E^{\mathbf{N}}(k,L_j)$, we can compute $Q_j(n,k)$ with the overall complexity of $O(k^2 L_j) + O\left(M^2 \frac{n}{L_j} k\right)$.

APPENDIX J
PROOF OF LEMMA 5

The lemma is proved by induction on j. The case of $j=1$ is obviously true as $\hat{P}_e(n,k,1) = P_e^{opt}(n,k,1)$. Let us assume this statement is true for $j=1$ to $J-1$. Then, for $j=J$, we have

$$\begin{aligned} & \hat{P}_e(n,k,J) \\ \overset{(a)}{\le} & \sum_{i=0}^{N_J} Q_J(N_J^{opt},i)\hat{P}_e(n-N_J^{opt},k-i,J-1) \\ \overset{(b)}{\le} & \sum_{i=0}^{N_J} Q_J(N_J^{opt},i)P_e^{opt}(n-N_J^{opt},k-i,J-1) \\ \overset{(c)}{\le} & \sum_{i=0}^{N_J} Q_J(N_J^{opt},i)P_e^{\mathbf{N}^{opt}}(k-i,J-1) \\ \overset{(d)}{=} & P_e^{\mathbf{N}^{opt}}(k,J) = P_e^{opt}(n,k,J) \end{aligned}$$

where $\mathbf{N}^{opt}$ denotes the optimum allocation of n packets among the J types of paths such that the probability of having more than k lost packets is minimized. (a) follows from the recursive equation (14), and (b) is the induction assumption. (c) comes from the definition of $P_e^{opt}(n,k,l)$, and (d) is a result of equation (16).

APPENDIX K
PROOF OF THEOREM 2

Sketch of the proof: First, the asymptotic behavior of $Q_j(n,k)$ is analyzed, and it is shown that for large values of L_j (or equivalently L), equation (60) computes the exponent of $Q_j(n,k)$ versus L. Next, we prove the first part of the theorem by induction on J. The proof of this part is divided to two different cases, depending on whether $\frac{K}{N}$ is larger than $\mathbb{E}\{x_J\}$ or vice versa. Finally, the second and the third parts of the theorem are proved by induction on j while the total number of path types, J, is fixed. Again, the proof is divided into two different cases, depending on whether $\frac{K}{N}$ is larger than $\mathbb{E}\{x_j\}$ or vice versa.

Proof: First, we compute the asymptotic behavior of $Q_j(n,k)$ for $k > n\mathbb{E}\{x_j\}$, and n growing proportionally to L_j, i.e. $n = n' L_j$. Here, we can apply Sanov's Theorem [77], [88] as n and k are discrete variables and n' is a constant.

Sanov's Theorem. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. discrete random variables from an alphabet set $\mathcal{X}$ with the size $|\mathcal{X}|$ and probability mass function (pmf) $Q(x)$. Let $\mathcal{P}$ denote the set of pmf's in $\mathbb{R}^{|\mathcal{X}|}$, i.e. $\mathcal{P} = \left\{\mathbf{P} \in \mathbb{R}^{|\mathcal{X}|} \mid P(i) \ge 0,\ \sum_{i=1}^{|\mathcal{X}|} P(i) = 1\right\}$. Also, let $\mathcal{P}_L$ denote the subset of $\mathcal{P}$ corresponding to all possible empirical distributions of $\mathcal{X}$ in L observations [88], i.e. $\mathcal{P}_L = \{\mathbf{P} \in \mathcal{P} \mid \forall i,\ LP(i) \in \mathbb{Z}\}$. For any dense and closed set [83] of pmf's $E \subseteq \mathcal{P}$, the probability that the empirical distribution of L observations belongs to the set E is equal to*

$$\mathbb{P}\{E\} = \mathbb{P}\{E \cap \mathcal{P}_L\} \doteq e^{-LD(\mathbf{P}^\star || \mathbf{Q})} \tag{57}$$

where $\mathbf{P}^\star = \underset{\mathbf{P}\in E}{\operatorname{argmin}} D(\mathbf{P}||\mathbf{Q})$ and $D(\mathbf{P}||\mathbf{Q}) = \sum_{i=1}^{|\mathcal{X}|} P(i)\log\frac{P(i)}{Q(i)}$.

Focusing our attention on the main problem, assume that $\mathbf{P}$ is defined as the empirical distribution of the number of errors in each path, i.e. for $\forall i,\ 1 \le i \le n'$, $P(i)$ shows the ratio of the total paths which contain exactly i lost packets. Similarly, for $\forall i,\ 1 \le i \le n'$, $Q(i)$ denotes the probability of exactly i packets being lost out of the n' packets transmitted on a path of type j. The sets E and E_{out} are defined as follows

$$\begin{aligned} E &= \{\mathbf{P} \in \mathcal{P} \mid \sum_{i=0}^{n'} iP(i) \ge \beta\} \qquad (58) \\ E_{out} &= \{\mathbf{P} \in \mathcal{P} \mid \sum_{i=0}^{n'} iP(i) = \beta\} \end{aligned}$$

where $\beta = \dfrac{k}{n}$. Noting E and E_{out} are dense sets, we can compute $Q_j(n,k)$ as

$$Q_j(n,k) \overset{(a)}{=} \mathbb{P}\{E_{out}\} \overset{(b)}{\doteq} e^{-L_j \underset{\mathbf{P}\in E_{out}}{\min} D(\mathbf{P}||\mathbf{Q})} \tag{59}$$

where (a) follows from the definition of $Q_j(n,k)$ as the probability of having exactly k errors out of the n packets sent over the paths of type j given in section IV, and (b) results from Sanov's Theorem.

Knowing the fact that the Kullback Leibler distance, $D(\mathbf{P}||\mathbf{Q})$, is a convex function of $\mathbf{P}$ and $\mathbf{Q}$ [89], we conclude that its minimum over the convex set E either lies on an interior point which is a global minimum of the function over the whole set $\mathcal{P}$ or is located on the boundary of E. However, we know that the global minimum of Kullback Leibler distance occurs at $\mathbf{P} = \mathbf{Q} \notin E$. Thus, the minimum of $D(\mathbf{P}||\mathbf{Q})$ is located on the boundary of E. This results in

$$\begin{aligned} Q_j(n,k) &\overset{(a)}{\doteq} e^{-L_j \underset{\mathbf{P}\in E_{out}}{\min} D(\mathbf{P}||\mathbf{Q})} \\ &= e^{-L_j \underset{\mathbf{P}\in E}{\min} D(\mathbf{P}||\mathbf{Q})} \overset{(b)}{\doteq} e^{-\gamma_j L u_j(\frac{k}{n})} \end{aligned} \tag{60}$$

where (a) and (b) follow from equations (59) and (6), respectively.

1) We prove the first part of the theorem by induction on J. When $J=1$, the statement is correct for both cases of $\frac{K}{N} > \mathbb{E}\{x_1\}$ and $\frac{K}{N} \le \mathbb{E}\{x_1\}$, recalling the fact that $\hat{P}_e(n,k,1) = P_e^{opt}(n,k,1)$ and $u_1(x) = 0$ for $x \le \mathbb{E}\{x_1\}$. Now, let us assume the first part of the theorem is true for $j=1$ to $J-1$. We prove the same statement for J as well. The proof can be divided into two different cases, depending on whether $\frac{K}{N}$ is larger than $\mathbb{E}\{x_J\}$ or vice versa.

1.1) $\dfrac{K}{N} > \mathbb{E}\{x_J\}$

According to the definition, the value of $\hat{P}_e(N,K,J)$ is computed by minimizing $\sum_{i=0}^{n_J} Q_J(n_J,i)\hat{P}_e(N-n_J,K-i,J-1)$ over n_J (see equation (16)). Now, we show that for any value of n_J, the corresponding term in the minimization is asymptotically at least equal to $P_e^{opt}(N,K,J)$. n_J can take integer values in the range $0 \le n_J \le N$. We split this range into three non-overlapping intervals of $0 \le n_J \le \epsilon L$, $\epsilon L \le n_J \le N(1-\epsilon)$, and $N(1-\epsilon) < n_J \le N$ for any arbitrary constant $\epsilon \le \min\left\{\gamma_j, 1-\frac{K}{N}\right\}$. The reason is that equation (60) is valid in the second interval only, and we need separate analyses for the first and last intervals.

First, we show the statement for $\epsilon L \le n_J \le N(1-\epsilon)$. Defining $i_J = \lfloor n_J \frac{K}{N} \rfloor$, we have

$$\begin{aligned} \frac{i_J}{n_J} &= \frac{K}{N} + O(\frac{1}{L}), \\ \frac{K-i_J}{N-n_J} &= \frac{K}{N} + O(\frac{1}{L}) \end{aligned} \tag{61}$$

as ϵ is constant, and $K = O(L)$, $N = O(L)$. Hence, we have

$$\begin{aligned} & \sum_{i=0}^{n_J} Q_J(n_J,i)\hat{P}_e(N-n_J,K-i,J-1) \\ \ge\ & Q_J(n_J,i_J)\hat{P}_e(N-n_J,K-i_J,J-1) \\ \overset{(a)}{\doteq}\ & e^{-L\sum_{j=1}^{J}\gamma_j u_j\left(\frac{K}{N}+O\left(\frac{1}{L}\right)\right)} \\ \overset{(b)}{\doteq}\ & e^{-L\sum_{j=1}^{J}\gamma_j u_j\left(\frac{K}{N}\right)} \end{aligned} \tag{62}$$

where (a) follows from (60) and the induction assumption, and (b) follows from the fact that $u_j()$'s are differentiable functions according to Lemma 1 in subsection III-A.

For $0 \le n_J \le \epsilon L$, since $\epsilon < \gamma_j$, the number of packets assigned to the paths of type J is less than the number of such paths. Thus, one packet is allocated to n_J of the paths, and the rest of the paths of type J are not used. Defining $\pi_{b,J}$ as the probability of a path of type J being in the bad state, we can write

$$Q_J(n_J,n_J) = \pi_{b,J}^{n_J} = e^{-n_J\log\left(\frac{1}{\pi_{b,J}}\right)}. \tag{63}$$

Therefore, for $0 \le n_J \le \epsilon L$, we have

$$\begin{aligned} & \sum_{i=0}^{n_J} Q_J(n_J,i)\hat{P}_e(N-n_J,K-i,J-1) \\ \ge\ & Q_J(n_J,n_J)\hat{P}_e(N-n_J,K-n_J,J-1) \\ \doteq\ & e^{-L\sum_{j=1}^{J-1}\gamma_j u_j\left(\frac{K-n_J}{N-n_J}\right)-n_J\log\left(\frac{1}{\pi_{b,J}}\right)} \\ \overset{(a)}{\ge}\ & e^{-L\sum_{j=1}^{J-1}\gamma_j u_j\left(\frac{K}{N}\right)-L\epsilon\log\left(\frac{1}{\pi_{b,J}}\right)} \\ \overset{(b)}{\doteq}\ & e^{-L\sum_{j=1}^{J-1}\gamma_j u_j\left(\frac{K}{N}\right)} \ge e^{-L\sum_{j=1}^{J}\gamma_j u_j\left(\frac{K}{N}\right)} \end{aligned} \tag{64}$$

where (a) follows from the fact that $\frac{K-n_J}{N-n_J} \le \frac{K}{N}$, and (b) results from the fact that we can select ϵ arbitrarily small.

Finally, we prove the statement for the case $n_J > N(1-\epsilon)$. In this case, we have

$$\begin{aligned} & \sum_{i=0}^{n_J} Q_J(n_J,i)\hat{P}_e(N-n_J,K-i,J-1) \\ \ge\ & Q_J(n_J,K)\hat{P}_e(N-n_J,0,J-1) \\ \overset{(a)}{\ge}\ & e^{-L\gamma_J u_J\left(\frac{K}{N(1-\epsilon)}\right)} \\ \overset{(b)}{\dot{\ge}}\ & e^{-L\sum_{j=1}^{J}\gamma_j u_j\left(\frac{K}{N}\right)} \end{aligned} \tag{65}$$

where (a) follows from the fact that $\epsilon < 1-\frac{K}{N}$ and $\hat{P}_e(n,0,j) = 1$, for all n and j. Setting ϵ small enough results in (b).

Inequalities (62), (64), and (65) result in

$$\hat{P}_e(N,K,J) \ \dot{\ge}\ e^{-L\sum_{j=1}^{J}\gamma_j u_j(\alpha)} \tag{66}$$

Combining (66) with Lemma 5 proves the first part of Theorem 2 for the case when $\frac{K}{N} > \mathbb{E}\{x_J\}$.

1.2) $\dfrac{K}{N} \le \mathbb{E}\{x_J\}$

Similar to the case of $\dfrac{K}{N} > \mathbb{E}\{x_J\}$ in subsection 1.1, we show that for any value of $0 \le n_J \le N$, the corresponding term of the minimization in equation (16) is asymptotically at least equal to $P_e^{opt}(N,K,J)$. Again, the range of n_J is partitioned into three non-overlapping intervals.

For any arbitrary $0 < \epsilon < \min\left\{\gamma_J, 1-\frac{K}{N}, \frac{1}{K}\right\}$, and for all n_J in the range of $\epsilon L < n_J \le N(1-\epsilon)$, we define i_J as $i_J = \lceil n_J \mathbb{E}\{x_J\} \rceil$. We have

$$\begin{aligned} \frac{i_J}{n_J} &= \mathbb{E}\{x_J\} + O\left(\frac{1}{L}\right) \ge \mathbb{E}\{x_J\} \\ \frac{K-i_J}{N-n_J} &< \frac{K}{N} + O\left(\frac{1}{L}\right) \end{aligned} \tag{67}$$

Hence,

$$\begin{aligned}
&\sum_{i=0}^{n_J} Q_J(n_J,i)\hat{P}_e(N-n_J,K-i,J-1)\\
\geq\ & Q_J(n_J,i_J)\hat{P}_e(N-n_J,K-i_J,J-1)\\
\stackrel{(a)}{\doteq}\ & e^{-L\gamma_J u_J\left(\frac{i_J}{n_J}\right)-L\sum_{j=1}^{J-1}\gamma_j u_j\left(\frac{K-i_J}{N-n_J}\right)}\\
\stackrel{(b)}{\geq}\ & e^{-L\gamma_J u_J\left(\mathbb{E}\{x_J\}+O\left(\frac{1}{L}\right)\right)}\cdot\\
& e^{-L\sum_{j=1}^{J-1}\gamma_j u_j\left(\frac{K}{N}+O\left(\frac{1}{L}\right)\right)}\\
\stackrel{(c)}{\doteq}\ & e^{-L\sum_{j=1}^{J}\gamma_j u_j\left(\frac{K}{N}\right)} \qquad (68)
\end{aligned}$$

where (a) follows from (60) and the induction assumption, and (b) is based on (67). (c) results from the facts that $u_j()$'s are differentiable functions, and we have $u_J\left(\mathbb{E}\{x_J\}\right)=0$, both according to Lemma 1 in subsection III-A.

For $0\leq n_J\leq \epsilon L$, the analysis of section 1.1 and inequality (64) are still valid. For $n_J>(1-\epsilon)N$, we set $i_J=\lceil\mathbb{E}\left\{x_J\right\}n_J\rceil$. Now, we have

$$i_J\geq n_J\mathbb{E}\{x_J\}>(1-\epsilon)N\mathbb{E}\{x_J\}\geq(1-\epsilon)K. \qquad (69)$$

The above inequality can be written as

$$K-i_J<\epsilon K<1 \qquad (70)$$

since $\epsilon<\frac{1}{K}$. Noting that K and i_J are integer values, it is concluded that $K\leq i_J$. Now, we can write

$$\begin{aligned}
&\sum_{i=0}^{n_J} Q_J(n_J,i)\hat{P}_e(N-n_J,K-i,J-1)\\
\geq\ & Q_J(n_J,i_J)\hat{P}_e(N-n_J,K-i_J,J-1)\\
\stackrel{(a)}{=}\ & Q_J(n_J,i_J)\\
\dot{\geq}\ & e^{-L\gamma_J u_J\left(\mathbb{E}\left\{x_J\right\}+\frac{1}{n_J}\right)}\\
\stackrel{(b)}{\dot{\geq}}\ & e^{-L\gamma_J u_J\left(\mathbb{E}\left\{x_J\right\}+\frac{1}{(1-\epsilon)N}\right)}\\
\doteq\ & e^{-L\gamma_J u_J\left(\mathbb{E}\left\{x_J\right\}+O\left(\frac{1}{L}\right)\right)}\stackrel{(c)}{\doteq}1 \qquad (71)
\end{aligned}$$

where (a) follows from the fact that $K\leq i_J$, and $\hat{P}_e(n,k,j)=1$, for $k\leq 0$. (b) and (c) result from $n_J>(1-\epsilon)N$ and $u_J\left(\mathbb{E}\{x_J\}\right)=0$, respectively.

Hence, inequalities (64), (68), and (71) result in

$$\hat{P}_e(N,K,J)\ \dot{\geq}\ e^{-L\sum_{j=1}^{J}\gamma_j u_j(\alpha)} \qquad (72)$$

which proves the first part of Theorem 2 for the case of $\frac{K}{N}\leq\mathbb{E}\{x_J\}$ when combined with Lemma 5.

2) We prove the second and the third parts of the theorem by induction on j while the total number of types, J, is fixed. The proof of the statements for the base of the induction, $j=J$, is similar to the proof of the induction step, from $j+1$ to j. Hence, we just give the proof for the induction step. Assume the second and the third parts of the theorem are true for $m=J$ to $j+1$. We prove the same statements for j. The proof is divided into two different cases, depending on whether $\frac{K}{N}$ is larger than $\mathbb{E}\{x_j\}$ or vice versa.

Before we proceed further, it is helpful to introduce two new parameters N' and K' as

$$\begin{aligned}
N' &= N-\sum_{m=j+1}^{J}\hat{N}_j\\
K' &= K-\sum_{m=j+1}^{J}K_j.
\end{aligned}$$

According to the above definitions and the induction assumptions, it is obvious that

$$\frac{K'}{N'}=\frac{K}{N}+o(1)=\alpha+o(1). \qquad (73)$$

2.1) $\frac{K}{N}>\mathbb{E}\{x_j\}$

First, by contradiction, it will be shown that for small enough values of $\epsilon>0$, we have $\hat{N}_j>\epsilon N'$. Let us assume the opposite is true, i.e. $\hat{N}_j\leq\epsilon N'$. Then, we can write

$$\begin{aligned}
&\hat{P}_e(N',K',j)\\
\stackrel{(a)}{=}\ & \sum_{i=0}^{\hat{N}_j}\hat{P}_e(N'-\hat{N}_j,K'-i,j-1)Q_j(\hat{N}_j,i)\\
\geq\ & \hat{P}_e(N'-\hat{N}_j,K'-\hat{N}_j,j-1)Q_j(\hat{N}_j,\hat{N}_j)\\
\stackrel{(b)}{\doteq}\ & Q_j(\hat{N}_j,\hat{N}_j)e^{-L\sum_{r=1}^{j-1}\gamma_r u_r\left(\frac{K'-\hat{N}_j}{N'-\hat{N}_j}\right)}\\
\stackrel{(c)}{\dot{\geq}}\ & e^{-Ln_0\left(1-\sum_{r=j+1}^{J}\eta_r\right)\epsilon\log\left(\frac{1}{\pi_{b,j}}\right)}\cdot\\
& e^{-L\sum_{r=1}^{j-1}\gamma_r u_r\left(\frac{K'}{N'}\right)}\\
\stackrel{(d)}{\dot{>}}\ & e^{-L\sum_{r=1}^{j}\gamma_r u_r(\alpha)} \qquad (74)
\end{aligned}$$

where (a) follows from equation (16) and step (2) of our sub-optimal algorithm, (b) results from the first part of Theorem 2, and (c) can be justified using arguments similar to those of inequality (64). (d) is obtained assuming ϵ is small enough such that the corresponding term in the exponent is strictly less than $L\gamma_j u_j\left(\frac{K'}{N'}\right)$ and also the fact that $\frac{K'}{N'}=\alpha+o(1)$. The result in (74) is obviously in contradiction with the first part of Theorem 2, proving that $\hat{N}_j>\epsilon N'$.

Now, we show that if $\hat{N}_j>(1-\epsilon)N'$ for arbitrarily small values of ϵ, we should have $\mathbb{E}\left\{x_r\right\}>\alpha$ for all $1\leq r\leq j-1$. In such a case, we observe $\frac{\hat{N}_j}{N'}=1+o(1)$, proving the second statement of Theorem 2. To show this, let us assume $\hat{N}_j > (1-\epsilon)N'$. Hence,

$$\begin{aligned}\hat{P}_e(N',K',j) &= \sum_{i=0}^{\hat{N}_j} \hat{P}_e(N'-\hat{N}_j, K'-i, j-1)Q_j(\hat{N}_j, i)\\ &\dot{\geq} \hat{P}_e(N'-\hat{N}_j, 0, j-1)Q_j(\hat{N}_j, K')\\ &\overset{(a)}{\dot{\geq}} e^{-L\gamma_j u_j\left(\frac{K'}{(1-\epsilon)N'}\right)} \overset{(b)}{\doteq} e^{-L\gamma_j u_j(\alpha+o(1))} \quad (75)\end{aligned}$$

where (a) follows from the fact that $\hat{P}_e(n,0,j)=1$, for all values of n and j, and the fact that $\hat{N}_j \geq (1-\epsilon)N'$. (b) is obtained by making ϵ arbitrarily small and using equation (73). Applying (75) and knowing the fact that $\hat{P}_e(N',K',j) \doteq e^{-L\sum_{r=1}^{j}\gamma_r u_r(\alpha)}$, we conclude that $\mathbb{E}\{x_r\} > \alpha$, for all values of $1 \leq r \leq j-1$.

$\hat{P}_e(N',K',j)$ can be written as

$$\begin{aligned}&\hat{P}_e(N',K',j)\\ =& \min_{0\leq N_j \leq N'} \sum_{i=0}^{N_j} \hat{P}_e(N'-N_j, K'-i, j-1)Q_j(N_j,i)\\ \overset{(a)}{\doteq}& \min_{\epsilon N' \leq N_j \leq (1-\epsilon)N'} \max_{0\leq i\leq N_j}\\ & \hat{P}_e(N'-N_j, K'-i, j-1)Q_j(N_j,i)\\ \overset{(b)}{\doteq}& \min_{\epsilon N' \leq N_j \leq (1-\epsilon)N'} \max_{\mathbb{E}\{x_j\}N_j < i \leq N_j}\\ & e^{-L\gamma_j u_j\left(\frac{i}{N_j}\right) - L\sum_{r=1}^{j-1}\gamma_r u_r\left(\frac{K'-i}{N'-N_j}\right)}\\ \doteq& e^{-L \max_{\epsilon N' \leq N_j \leq (1-\epsilon)N'} \min_{\mathbb{E}\{x_j\}N_j < i \leq N_j} M_d(i,N_j)}\\ \overset{(c)}{\doteq}& e^{-L \max_{\epsilon \leq \lambda_j \leq (1-\epsilon)} \min_{\mathbb{E}\{x_j\}\lambda_j < \beta_j \leq \lambda_j} M_c(\beta_j,\lambda_j)}. \quad (76)\end{aligned}$$

where $M_d(i,N_j)$ and $M_c(\beta_j,\lambda_j)$ are defined as

$$\begin{aligned}M_d(i,N_j) &= \gamma_j u_j\left(\frac{i}{N_j}\right) + \sum_{r=1}^{j-1}\gamma_r u_r\left(\frac{K'-i}{N'-N_j}\right)\\ M_c(\beta_j,\lambda_j) &= \gamma_j u_j\left(\frac{\beta_j}{\lambda_j}\right) + \sum_{r=1}^{j-1}\gamma_r u_r\left(\frac{\alpha-\beta_j}{1-\lambda_j}\right).\end{aligned}$$

In (76), (a) follows from the fact that $\hat{N}_j$ is bounded as $\epsilon N' \leq \hat{N}_j \leq (1-\epsilon)N'$. (b) results from equation (60), $\hat{P}_e(n,k,j)$ being a decreasing function of k, and the fact that we have $Q_j(N_j,i) \leq 1 \doteq Q_j(N_j, \mathbb{E}\{x_j\}N_j)$ for $i < \mathbb{E}\{x_j\}N_j$. β_j and λ_j are defined as $\beta_j = \frac{i}{N'}$ and $\lambda_j = \frac{N_j}{N'}$. (c) is a result of having $M_c(\beta_j,\lambda_j) = M_d(i,N_j) + O\left(\frac{1}{L}\right)$. Hence, the discrete to continuous relaxation is valid.

Let us define $\left(\beta_j^*, \lambda_j^*\right)$ as the values of (β_j,λ_j) which solve the max-min problem in (76). Differentiating $M_c(\beta_j,\lambda_j)$ with respect to β_j and λ_j results in

$$0 = \frac{\gamma_j}{\lambda_j^*} l_j\left(\frac{\beta_j^*}{\lambda_j^*}\right) - \sum_{\substack{r=1,\\ \mathbb{E}\{x_r\}<\zeta}}^{j-1} \frac{\gamma_r}{1-\lambda_j^*} l_r(\zeta)$$

$$0 = \left\{ -\frac{\gamma_j\beta_j^*}{\lambda_j^{*2}} l_j\left(\frac{\beta_j^*}{\lambda_j^*}\right) + \sum_{\substack{r=1,\\ \mathbb{E}\{x_r\}<\zeta}}^{j-1} \frac{\gamma_r(\alpha-\beta_j^*)}{(1-\lambda_j^*)^2} l_r(\zeta) + \left(\frac{\gamma_j}{\lambda_j^*} l_j\left(\frac{\beta_j^*}{\lambda_j^*}\right) - \sum_{\substack{r=1,\\ \mathbb{E}\{x_r\}<\zeta}}^{j-1} \frac{\gamma_r}{1-\lambda_j^*} l_r(\zeta) \right) \frac{\partial \beta_j^*}{\partial \lambda_j}|_{\lambda_j=\lambda_j^*} \right\}$$

where $\zeta = \dfrac{\alpha-\beta_j^*}{1-\lambda_j^*}$. Solving the above equations gives the unique optimum solution (β_j^*, λ_j^*) as

$$\begin{aligned}\beta_j^* &= \alpha\lambda_j^*\\ \lambda_j^* &= \frac{\gamma_j l_j(\alpha)}{\sum_{r=1,\alpha>\mathbb{E}\{x_r\}}^{j} l_r(\alpha)} \quad (77)\end{aligned}$$

Hence, the integer parameters $K_j, \hat{N}_j$ defined in the suboptimal algorithm have to satisfy $\frac{K_j}{N'} = \beta_j^* + o(1)$ and $\frac{\hat{N}_j}{N'} = \lambda_j^* + o(1)$, respectively. Based on the induction assumption, it is easy to show that

$$\frac{N'}{N} = \frac{\sum_{r=1,\mathbb{E}\{x_r\}<\alpha}^{j} \gamma_r u_r(\alpha)}{\sum_{r=1,\mathbb{E}\{x_r\}<\alpha}^{J} \gamma_r u_r(\alpha)} \quad (78)$$

which completes the proof for the case of $\mathbb{E}\{x_j\} < \frac{K}{N}$.

2.2) $\dfrac{K}{N} \leq \mathbb{E}\{x_j\}$

In this case, we show that $\frac{\hat{N}_j}{N} = o(1)$. Defining $i_j = \lceil \mathbb{E}\{x_j\}\hat{N}_j \rceil$, we have

$$\frac{K'-i_j}{N'-\hat{N}_j} = \alpha - (\mathbb{E}\{x_j\}-\alpha)\frac{\hat{N}_j}{N'-\hat{N}_j} + o(1) \quad (79)$$

using equation (73). Now, we have

$$\begin{aligned}&\hat{P}_e(N',K',j)\\ =& \sum_{i=0}^{\hat{N}_j} \hat{P}_e(N'-\hat{N}_j, K'-i, j-1)Q_j(\hat{N}_j, i)\\ \geq& \hat{P}_e(N'-\hat{N}_j, K'-i_j, j-1)Q_j(\hat{N}_j, i_j)\\ \overset{(a)}{\doteq}& e^{-L\gamma_j u_j(\mathbb{E}\{x_j\}+o(1))}.\\ & e^{-L\sum_{r=1}^{j-1}\gamma_r u_r\left(\alpha-(\mathbb{E}\{x_j\}-\alpha)\frac{\hat{N}_j}{N'-\hat{N}_j}\right)}\\ \doteq& e^{-L\sum_{r=1}^{j-1}\gamma_r u_r\left(\alpha-(\mathbb{E}\{x_j\}-\alpha)\frac{\hat{N}_j}{N'-\hat{N}_j}\right)} \quad (80)\end{aligned}$$

where (a) follows from the first part of Theorem 2 and (60). On the other hand, according to the result of the first part of Theorem 2, we know that

$$\hat{P}_e(N', K', j) \doteq e^{-L \sum_{r=1}^{j-1} \gamma_r u_r(\alpha)}. \quad (81)$$

According to Lemma 1, $u_r(\beta)$ is an increasing function of β for all $1 \leq r \leq j-1$. Thus, $\sum_{r=1}^{j-1} \gamma_r u_r(\beta)$ is also a one-to-one increasing function of β. Noting this fact and comparing (80) and (81), we conclude that $\frac{\hat{N}_j}{N'} = o(1)$ as $\mathbb{E}\{x_j\} - \alpha$ is strictly positive. Noting (78), we have $\frac{\hat{N}_j}{N} = o(1)$ which proves the second part of Theorem 2 for the case of $\frac{K}{N} \leq \mathbb{E}\{x_j\}$.

REFERENCES

[1] S. Fashandi, S. Oveisgharan, and A.K. Khandani, "Path Diversity in Packet Switched Networks: Performance Analysis and Rate Allocation," in *IEEE Global Telecommunications Conference, GLOBECOM '07*, 2007, pp. 1840–1844.
[2] W.H. Tarn and Y.C. Tseng, "Joint Multi-Channel Link Layer and Multi-Path Routing Design for Wireless Mesh Networks," in *INFOCOM*, 2007, pp. 2081–2089.
[3] N.S. Nandiraju, D.S. Nandiraju, and D.P. Agrawal, "Multipath Routing in Wireless Mesh Networks," in *IEEE International Conference on Mobile Adhoc and Sensor Systems*, 2006, pp. 741–746.
[4] P. Djukic and S. Valaee, "Reliable Packet Transmissions in Multipath Routed Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 5, pp. 548– 559, 2006.
[5] H. Han, S. Shakkottai, C.V. Hollot, R. Srikant, and D. Towsley, "Multi-Path TCP: A Joint Congestion Control and Routing Scheme to Exploit Path Diversity in the Internet," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1260 – 1271, 2006.
[6] T. Nguyen and A. Zakhor , "Path diversity with forward error correction (pdf) system for packet switched networks," in *IEEE INFOCOM Proc. IEEE Vol. 1*, 2003, pp. 663– 672.
[7] A. Akella, J. Pang, B. Maggs, S. Seshan, and A. Shaikh, "A Comparison of Overlay Routing and Multihoming Route Control," in *ACM SIGCOMM*, 2004, pp. 93 – 106.
[8] V.N. Padmanabhan, H.J. Wang, P.A. Chou, "Resilient Peer-to-peer Streaming," in *IEEE International Conference on Network Protocols*, 2003, pp. 16–27.
[9] N. Gogate, D.M. Chung, S.S. Panwar, Yao Wang, "Supporting Image and Video Applications in a Multihop Radio Environment Using Path Diversity and Multiple Description Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 777–792, 2002.
[10] Wei Wei and A. Zakhor, "Interference Aware Multipath Selection for Video Streaming in Wireless Ad Hoc Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, pp. 165 – 178, 2009.
[11] S. Mao, S.S. Panwar, and Y.T. Hou, "On optimal partitioning of realtime traffic over multiple paths," in *INFOCOM 2005, Proc. IEEE Vol. 4*, 2005, pp. 2325–2336.
[12] S.J. Lee and M. Gerla, "SMR: Split Multipath Routing with Maximally Disjoint Paths in Ad hoc Networks," in *IEEE Interneational Conference on Communications*, 2001, pp. 3201–3205.
[13] ——, "AODV-BR: Backup Routing in Ad Hoc Networks," in *Wireless Communications and Networking Conference*, 2000, pp. 1311–1316.
[14] M.K. Marina and S.R. Das, "On-Demand Multipath Distance Vector Routing for Ad Hoc Networks," in *IEEE International Conference on Network Protocols*, 2001, pp. 14–23.
[15] Z. Ye, S. V. Krishnamurthy, and S. K. Tripathi, "A Framework for Reliable Routing in Mobile Ad Hoc Networks," in *INFOCOM*, 2003, pp. 270– 280.
[16] S. Mao, Y.T. Hou, X. Cheng, H.D. Sherali, S.F. Midkiff, and Y.Q. Zhang, "On Routing for Multiple Description Video Over Wireless Ad Hoc Networks," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 1063–1074, 2006.
[17] R. Leung, J. Liu, E. Poon, A.L.C. Chan, and Baochun Li, "MP-DSR: A QoS-aware Multi-path Dynamic Source Routing Protocol For Wireless Ad-hoc Networks," in *IEEE International Conference on Local Computer Networks*, 2001, pp. 132–141.
[18] L. Wang, Y. Shu, O. Yang, M. Dong, and L. Zhang, "Adaptive Multipath Source Routing in Ad Hoc Networks,," in *IEEE International Conference on Communications*, 2001, pp. 867–871.
[19] K. Rojviboonchai, F. Yang, Q. Zhang, H. Aida1, and W. Zhu, "AMTP: A Multipath Multimedia Streaming Protocol for Mobile Ad Hoc Networks," in *IEEE International Conference on Communications*, 2005, pp. 1246–1250.
[20] Wei Wei and A. Zakhor, "Robust Multipath Source Routing Protocol (RMPSR) for Video Communication over Wireless Ad Hoc Networks," in *International Conference on Multimedia*, 2004, p. 13791382.
[21] Y. J. Liang, E. G. Steinbach, and B. Girod , "Multi-stream Voice over IP using Packet Path Diversity," in *IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, pp. 555–560.
[22] S. Nelakuditi, Z. Zhang, and D. H. C. Du, "On Selection of Candidate Paths for Proportional Routing," *Elsevier Computer Networks*, vol. 44, no. 1, pp. 79–102, 2004.
[23] J. Han and F. Jahanian, "Impact of Path Diversity on Multi-homed and Overlay Networks," in *International Conference on Dependable Systems and Networks*, 2004, pp. 29–38.
[24] David G. Andersen, *Resilient Overlay Networks*. Master's Thesis, Massachusetts Institute of Technology, 2001.
[25] D. G. Andersen, A. C. Snoeren, and H. Balakrishnan, "Best-path vs. Multi-path Overlay Routing," in *Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement*, 2003, pp. 91 – 100.
[26] R. Teixeira, K. Marzullo, S. Savage, and G. M. Voelker, "In Search of Path Diversity in ISP Networks," in *Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement*, 2003, pp. 313 – 318.
[27] D. Rubenstein, J. Kurose, and D. Towsley, "Detecting Shared Congestion of Flows via End-to-end Measurement," *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, p. 381395, 2002.
[28] D. Katabi, I. Bazzi, and Xiaowei Yang, "A Passive Approach for Detecting Shared Bottlenecks," in *IEEE International Conference on Computer Communications and Networks*, 2001, pp. 174–181.
[29] M. S. Kim, T. Kim, Y. J. Shin, S. S. Lam, E.J. Powers, "A Wavelet-Based Approach to Detect Shared Congestion," *IEEE/ACM Transactions on Networking*, vol. 16, no. 4, pp. 763–776, 2008.
[30] C.E. Perkins, E.M. Royer, S.R. Das, "Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks," *IEEE Personal Communications*, vol. 8, no. 1, pp. 16–28, 2001.
[31] D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing*, p. 153181, 1996.
[32] Wei Wei and A. Zakhor, "Multiple Tree Video Multicast over Wireless Ad Hoc Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 1, pp. 2–15, 2007.
[33] D. Clark, W. Lehr, S. Bauer, P. Faratin, R. Sami, and J. Wroclawski, "Overlay Networks and Future of the Internet," *Journal of Communications and Strategies*, vol. 3, no. 63, pp. 1–21, 2006.
[34] M. Guo, Q. Zhang, and W. Zhu , "Selecting Path-diversified Servers in Content Distribution Networks," in *IEEE Global Telecommunications Conference*, vol. 6, 2003, pp. 3181–3185.
[35] M. Cha, S. Moon, C. D. Park, and A. Shaikh, "Placing Relay Nodes for Intra-Domain Path Diversity," in *IEEE INFOCOM*, 2006, pp. 1–12.
[36] S. Srinivasan, *Design and Use of Managed Overlay Networks*. PhD Dissertation, Georgia Institute of Technology, 2007.
[37] J. Han, D. Watson, and F. Jahanian, "Enhancing End-to-End Availability and Performance via Topology-aware Overlay Networks," *Elsevier Computer Networks*, vol. 52, no. 16, pp. 3029–3046, 2008.
[38] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "Topologically-aware Overlay Construction and Server Selection," in *IEEE INFOCOM*, 2002, pp. 1190–1199.
[39] M. Castro, P. Druschel, Y. C. Hu, and A. Rowstron, "Topology-aware Routing in Structured Peer-to-peer Overlay Networks."
[40] Zhi Li and P. Mohapatra, "QRON: QoS-aware Routing in Overlay Networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 1, pp. 29–40, 2004.
[41] ——, "The Impact of Topology on Overlay Routing Service," in *IEEE INFOCOM*, 2004, pp. 408–418.
[42] A. Nakao, L. Peterson, and A. Bavier, "A Routing Underlay for Overlay Networks," in *Proceedings of ACM SIGCOMM*, 2003, pp. 11–18.
[43] J. Touch, Y. Wang, V. Pingali, L. Eggert, R. Zhou, and G. Finn, "A Routing Underlay for Overlay Networks," in *Proc. IEEE Tridentcom*, 2005, pp. 194–203.
[44] B-G Chun, R. Fonseca, I. Stoica, and J. Kubiatowicz, "Characterizing Selfishly Constructed Overlay Routing Networks," in *IEEE INFOCOM*, 2004, pp. 1329–1339.

[45] Z. Li, L. Yuan, P. Mohapatra, and C-N. Chuah, “On the Analysis of Overlay Failure Detection and Recovery,” *Computer Networks Journal*, vol. 51, no. 13, pp. 3838–3843, 2007.

[46] A. Akella, B. Maggs, S. Seshan, and A. Shaikh, “On the Performance Benefits of Multihoming Route Control,” *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 91–104, 2008.

[47] Y.H. Wang and H.Z. Lin, “Multipath QoS Routing with Interference Provision in Ad Hoc Wireless Network,” *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1325–1338, 2006.

[48] Roger Karrer, and Thomas Gross, “Multipath Streaming in Best-Effort Networks,” in *Proc. of the IEEE International Conference on Communications (ICC’03)*, 2003.

[49] J.G. Apostolopoulos, T. Wong, W. Tan, and S.J. Wee, “On Multiple Description Streaming with Content Delivery Networks,” in *IEEE INFOCOM, Proc. IEEE Vol. 3*, 2002, pp. 1736 – 1745.

[50] T. Nguyen and A. Zakhor, “Multiple Sender Distributed Video Streaming,” *IEEE transactions on multimedia*, vol. 6, no. 2, pp. 315– 326, 2004.

[51] “Akamai SureRoute,” http://www.akamai.com/dl/feature_sheets/fs_edgesuite_sureroute.pdf.

[52] M. Ghanassi and P. Kabal, “Optimizing Voice-over-IP Speech Quality Using Path Diversity,” in *IEEE 8th Workshop on Multimedia Signal Processing*, 2006, pp. 155–160.

[53] J. Chakareski and B. Girod, “Rate-distortion optimized packet scheduling and routing for media streaming with path diversity,” in *Proc. IEEE Data Compression Conference*, 2003, pp. 203– 212.

[54] E. Setton, X. Zhu, and B. Girod, “Congestion-optimized Multi-path Streaming of Video over Ad Hoc Wireless Networks,” in *IEEE International Conference on Multimedia*, 2004, pp. 1619– 1622.

[55] S. Mao, D. Bushmitch, S. Narayanan, and S.S. Panwar, “MRTP: A Multiflow Real-Time Transport Protocol for Ad Hoc Networks,” *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 356–369, 2006.

[56] M. Afergan, J. Wein, and A. LaMeyer, “Experience with some Principles for Building an Internet-Scale Reliable System,” in *Proceedings of the Fifth IEEE International Symposium on Network Computing and Applications (NCA’06)*, 2006, p. 3.

[57] Yi Li, Yin Zhang, Li Li Qiu, and S. Lam, “SmartTunnel: Achieving Reliability in the Internet,” in *IEEE INFOCOM*, 2007, pp. 830–838.

[58] A. Populis and S. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

[59] H. Sanneck, G. Carle, “A Framework Model for Packet Loss Metrics Based on Loss Runlengths,” in *ACM SIGMM Multimedia Computing and Networking Conference*, 2000.

[60] J. C. Bolot, “End-to-end Packet Delay and Loss Behavior in the Internet,” in *ACM SIGCOMM Conference on Communication Architectures, Protocols and Applications*, 1993, pp. 289 – 298.

[61] J.C. Bolot, S. Fosse-Parisis, and D. Towsley, “Adaptive FEC-based error control for Internet telephony,” in *IEEE INFOCOM, Proc. IEEE Vol. 3*, 1999, pp. 1453–1460.

[62] J.C. Bolot and T. Turletti, “Adaptive Error Control For Packet Video In The Internet,” in *Proc. IEEE International Conference on Image Processing*, 1996, pp. 25 – 28.

[63] F. L. Leannec, F. Toutain, and C. Guillemot, “Packet Loss Resilient MPEG-4 Compliant Video Coding for the Internet,” *Journal of Image Communication, Special Issue on Real-time video over the Internet*, no. 15, pp. 35–56, 1999.

[64] S. Fashandi, S. Oveisgharan, and A.K. Khandani, “Coding over an Erasure Channel with a Large Alphabet Size,” in *IEEE International Symposium on Information Theory*, 2008, pp. 1053–1057.

[65] M. Yajnik, S.B. Moon, J.F. Kurose, and D.F. Towsley , “Measurement and Modeling of the Temporal Dependence in Packet Loss,” in *IEEE INFOCOM Proc. IEEE Vol. 1*, 1999, pp. 345–352.

[66] Wenyu Jiang and Henning Schulzrinne, “Modeling of Packet Loss and Delay and their Effect on Real-Time Multimedia Service Quality,” in *Proc. Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, 2000.

[67] Yin Zhang and Nick Duffield, “On the Constancy of Internet Path Properties,” in *ACM SIGCOMM Workshop on Internet Measurement*, 2001, pp. 197 – 211.

[68] P. Rossi, G. Romano, F. Palmieri, and G. Iannello , “A Hidden Markov Model for Internet Channels,” in *IEEE International Symposium on Signal Processing and Information Technology*, 2003.

[69] K. Salamatian and Vaton, “Hidden Markov Modeling for Network Communication Channels,” in *Proc. ACM SIGMETRICS*, 2001, pp. 92 – 101.

[70] D. Wu, Y.T. Hou, and Y-Q Zhang , “Transporting Real-time Video over the Internet: Challenges and Approaches,” *Proceedings of the IEEE*, vol. 88, no. 12, pp. 1855–1877, 2000.

[71] M. Podolsky, C. Romer, and S. McCanne, “Simulation of FEC-based Error Control for Packet Audio on the Internet,” in *IEEE INFOCOM*, 1998, pp. 505–515.

[72] R. Puri, K. Ramchandran, K.W. Lee, and V. Bharghavan, “Forward Error Correction (FEC) Codes Based Multiple Description Coding for Internet Video Streaming and Multicast,” *Elsevier Signal Processing: Image Communication*, vol. 16, no. 8, pp. 745–762, 2001.

[73] Ron M. Roth, *Introduction to Coding Theory*, 1st ed. Cambridge University Press, 2006, pp. 183–204.

[74] W.T. Tan and A. Zakhor, “Video Multicast Using Layered FEC and Scalable Compression,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 373–386, 2001.

[75] L. Dairaine, L. Lancrica, J. Lacan, and J. Fimes, “Content-Access QoS in Peer-to-Peer Networks Using a Fast MDS Erasure Code,” *Elsevier Computer Communications*, vol. 28, no. 15, pp. 1778–1790, 2005.

[76] X. H. Peng, “Erasure-control Coding for Distributed Networks,” *IEE Proceedings on Communications*, vol. 152, pp. 1075 – 1080, 2005.

[77] Amir Dembo and Ofer Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer, 1998, pp. 11–43.

[78] ISO/IEC JTC 1/SC 29, *Information technology – Generic Coding of Moving Pictures and Associated Audio Information: Systems*, ISO/IEC 13818-1, International Standard, October 2007.

[79] ——, *Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding*, ISO/IEC 14496-10, International Standard, November 2008.

[80] K.H. Goh, D.J. Wu, J.Y. Tham, T.K. Chiew and W.S. Lee, “Real-time Software MPEG-2 TO H.264 Video Transcoding,” in *IEEE International Conference on Multimedia and Expo*, 2008, pp. 165–168.

[81] B. Bing, “Real-time Software MPEG-2 TO H.264 Video Transcoding,” in *Communication Networks and Services Research Conference*, 2008, pp. 13–17.

[82] T. Yamaguchi, T. Kanekiyo, M. Horii, K. Kawazoe and F. Kishino, “Highly Efficient Transmission System for Digital Broadcasting Redistribution Services over IP Multicast Networks,” *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 920–924, 2008.

[83] J. L. Kelley, *General Topology*. Springer, 1975, pp. 40–43.

[84] C. H. Papadimitriou, *Computational Complexity*, 1st ed. New York: Addison Wesley, 1994.

[85] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. MIT Press, 2001, pp. 347–349.

[86] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.

[87] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge, UK: Cambridge University Press, 2004, pp. 243–245.

[88] T. Cover and J. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991, pp. 291–294.

[89] ——, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991, pp. 30–31.