Cross-Layer Resource Allocation Over Wireless Relay Networks for Quality of Service Provisioning

Jia Tang, *Student Member, IEEE,* and Xi Zhang, *Senior Member, IEEE*

***Abstract*— We propose a physical-datalink cross-layer resource allocation scheme over wireless relay networks for quality-of-service (QoS) guarantees. By integrating information theory with the concept of effective capacity, our proposed scheme aims at maximizing the relay network throughput subject to a given delay QoS constraint. This delay constraint is characterized by the so-called QoS exponent θ, which is the only requested information exchanged between the physical layer and the datalink layer in our cross-layer design based scheme. Over both amplify-and-forward (AF) and decode-and-forward (DF) relay networks, we develop the associated dynamic resource allocation algorithms for wireless multimedia communications. Over DF relay network, we also study a fixed power allocation scheme to provide QoS guarantees. The simulations and numerical results verify that our proposed cross-layer resource allocation can efficiently support diverse QoS requirements over wireless relay networks. Both AF and DF relays show significant superiorities over direct transmissions when the delay QoS constraints are stringent. On the other hand, our results demonstrate the importance of deploying the dynamic resource allocation for stringent delay QoS guarantees.**

***Index Terms*— Cross-layer design and optimization, wireless relay networks, wireless multimedia communications, quality-of-service (QoS), effective capacity, convex optimization, resource allocation.**

I. Introduction

WITH THE explosive developments of wireless communications, quality-of-service (QoS) provisioning has become a critically important performance metric for the future wireless networks. Unlike wireline networks, in which QoS can be guaranteed by independent optimization within each layer in the open system interconnection (OSI) model, over wireless networks there is a strong interconnection between layers, which makes the layered design and optimization approach less efficient. For example, at the physical layer, a great deal of research focuses on techniques that can enhance the spectral efficiency of wireless systems. The framework used to evaluate these techniques is mainly based on information theory, using the concept of Shannon capacity [1]. However, it is well known that Shannon capacity does not place any restrictions on complexity and delay [2]. As a result, the optimization merely at the physical layer may not lead to the desired delay QoS requested by the services at upper-protocol layers.

To deal with this problem, there have been increasing interests in design for wireless networks that rely on interactions between various layers of the protocol stack. This approach, called *cross-layer design and optimization*, has been widely recognized as a promising solution to provide diverse QoS provisioning in wireless multimedia communications [3]. The cross-layer approach relaxes the layering architecture of the conventional network model, which can result in a significant performance enhancement. However, such a design principle across different layers usually involves high complexity, which may cause the optimization problem intractable [4]. Consequently, how to develop efficient cross-layer approaches while minimizing the additional requested information exchanged between layers is an important issue from both theoretical and practical point-of-views.

On the other hand, relay communications have recently emerged as a powerful spatial diversity technique that can improve the performance over conventional point-to-point transmissions. The original work on relay communications was initialed by Cover and Gamal [5]. Since then, it has been extensively studied using different performance metrics [6]–[13], especially when the concept of *user cooperation* was proposed [6][7]. Clearly, combining the idea of cross-layer design with the relay network architecture, it is possible to significantly improve the system QoS provisioning performance. However, the research on how to efficiently employ the unique nature of relay architecture for designing the cross-layer protocols, and what is the impact of cross-layer resource allocation on supporting diverse QoS requirements over wireless relay networks, are still quite scarce [14].

To remedy the above deficiency, in this paper we propose a cross-layer resource allocation scheme for relay networks with the target at delay QoS guarantees for wireless multimedia communications. Our proposed scheme aims at maximizing the relay network throughput subject to a given delay QoS constraint. Our work builds on the integration of information theoretic results with the theory of statistical QoS guarantees, in particular, the recently developed powerful concept termed *effective capacity* [15]–[18]. The theory of statistical QoS guarantees has been extensively studied in the early 90's with the emphasis on wired asynchronous transfer mode (ATM) networks [19]–[23]. This theory enables us to analyze network statistics such as queue distributions, buffer overflow probabilities, and delay-bound violation probabilities, which are all important delay QoS metrics. As a part of the statistical QoS theory, effective capacity is particularly convenient for analyzing the statistical QoS performance of wireless multimedia transmissions where the service process is driven by the time-varying wireless channel.

Specifically, our resource allocation scheme is across the physical and the datalink layers. Applying the effective-

Manuscript received May 15th, 2006, revised November 24, 2006. The research reported in this paper was supported in part by the National Science Foundation CAREER Award under Grant ECS-0348694.

The authors are with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mails: {jtang,xizhang}@ece.tamu.edu).

Digital Object Identifier 10.1109/JSAC.2007.070502.

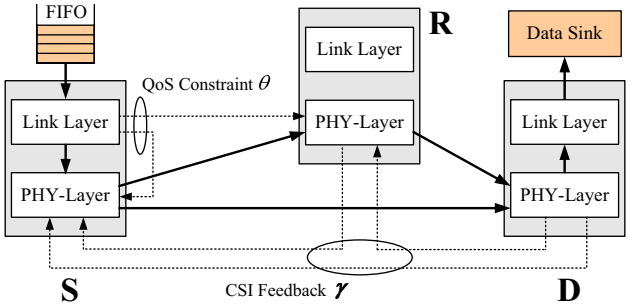


Fig. 1. The cross-layer relay network model.

capacity based approach, we convert the original throughput maximization to effective capacity maximization, and characterize the delay constraint by the so-called *QoS exponent* θ, which is the only requested information exchanged between the physical layer and the datalink layer in our cross-layer scheme. In particular, the dynamics of θ corresponds to different delay QoS constraints. For instance, non-real-time services such as data disseminations aim at maximizing the throughput with a loose delay constraint ($\theta \to 0$). In contrast, the key QoS requirement for real-time multimedia services is the timely delivery with stringently upper-bounded delay ($\theta \to \infty$). There also exist some services falling in between, like paging and interactive web surfing, which are delay sensitive but the delay QoS requirements are not as stringent as those of real-time multimedia applications ($0 < \theta < \infty$).

We focus on simple half-duplex relay protocols proposed in [9], namely, amplify-and-forward (AF) and decode-and-forward (DF), and develop the associated dynamic resource-allocation algorithms, where the resource allocation policies are functions of both the network channel state information (CSI) and the QoS constraint θ. The resulting resource allocation policy in turn provides a guideline on how to design the relay protocol that can efficiently support stringent QoS constraints. For DF relay networks, we also study a fixed power allocation scheme and investigate its performance. The simulations and numerical results verify that our proposed cross-layer resource allocation can efficiently support diverse QoS requirements over wireless relay networks. Moreover, both AF and DF relays show significant superiorities over direct transmissions when the delay constraints are stringent. On the other hand, our results demonstrate the importance of deploying the dynamic resource allocation for stringent delay QoS guarantees.

The rest of the paper is organized as follows. Section II describes our cross-layer relay network model. Sections III introduces the concept of statistical QoS guarantees. Sections IV and V develop the cross-layer resource allocation policies for AF and DF relay networks, respectively. Section VI investigates a fixed power allocation policy for DF relay networks. Section VII presents simulations and numerical results to evaluate the performance of our proposed cross-layer resource allocation. The paper concludes with Section VIII.

II. System Descriptions

A. Network Model

The cross-layer relay network model is shown in Fig. 1. We concentrate on a discrete-time system with a source node (S), a destination node (D), and a relay node (R), where the relay assists communications between the source and the destination without having its own data to send. As illustrated by Fig. 1, a first-in-first-out (FIFO) queue is implemented at the source node, which comprises the datalink-layer *packets* to be transmitted to the destination. At the physical (PHY)-layer, the datalink-layer packets are divided into *frames*, which form the data units through wireless transmissions. The frame duration is denoted by T_f, which is assumed to be less than the fading coherence time, but sufficiently long so that the information-theoretic assumption of infinite code-block length is meaningful. Based on a given QoS constraint θ requested by the service (which will be detailed in Section III) and CSI fed back from the corresponding receivers, the source and the relay need to find an optimal resource allocation strategy that can maximize the throughput subject to the QoS constraint θ. At the relay node, the transmission only involves the physical layer, as shown by Fig. 1. In this paper, we also make the following assumptions:

A1: The discrete-time channel is assumed to be block fading. The path gains are invariant within a frame's duration T_f, but vary independently from one frame to another. The block fading channel model is commonly used in literatures, which can also greatly simplify our analyses, as will be explained in Section III. Moreover, through the study of [24] we observe that there exists a simple and efficient approach to convert the resource allocation policy obtained in block fading channels to that over correlated fading channels, making the investigation of block fading channel more applicable.

A2: We assume that CSI is perfectly estimated at the corresponding receivers and reliably fed back to the source and the relay without delay. The assumption that the feedback is reliable can be (at least approximately) satisfied by using heavily coding feedback channels. On the other hand, the feedback delay can be compensated by channel prediction [25].

A3: We further assume that for a given instantaneous channel gain, the physical-layer codewords adaptively operates at the instantaneous achievable rate of the relay protocol. This assumption implies that an ideal adaptive modulation and coding scheme is implemented.

B. Channel Model

The relay channel model is shown in Fig. 2. We assume a flat fading channel model. The instantaneous channel coefficient between sender i and receiver j is denoted by $\{h_{i,j}\}$, where $i \in \{s, r\}$ and $j \in \{r, d\}$ with $i \neq j$, and s, r, d represent the source, relay, and destination, respectively. The magnitudes of these channel coefficients are assumed to follow an independent Rayleigh distribution, with the mean determined by the large-scale path loss. At each receiver, the additive noise is modeled as independent zero-mean, circularly symmetric complex white Gaussian with *unit* variance.

In the following discussions, we denote the channel gain $\gamma_1 = |h_{s,d}|^2$, $\gamma_2 = |h_{s,r}|^2$, and $\gamma_3 = |h_{r,d}|^2$, where γ_i follows

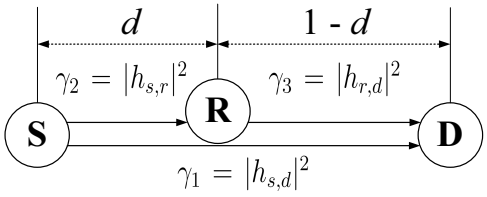


Fig. 2. The relay channel model.

an exponential distribution with parameter λ_i, $i \in \{1, 2, 3\}$. To study the impact of relay location on network performance, we normalize the distance between the source and the destination to one, and let the relay be located in a line between the source and the destination. The source-relay distance and the relay-destination distance are denoted by d and $1-d$, respectively, where $0 < d < 1$. Based on the channel model and network topology described above, the network CSI is determined by a 3-tuple $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$, which follows independent exponential distribution with parameter $\lambda_1 = 1$, $\lambda_2 = d^{\alpha}$, and $\lambda_3 = (1-d)^{\alpha}$, respectively, where α denotes the path loss exponent. A typical value of α lies in the range of $(2, 5)$. In the simulations and numerical results presented in Section VII, we will assume $\alpha = 4$.

C. Relay Protocols

Different relay protocols have been investigated in literatures [8]–[13]. In this paper, we mainly focus on the simple half-duplex relay protocols proposed in [9], namely, amplify-and-forward (AF) and decode-and-forward (DF). As compared to full-duplex relay, half-duplex relay restricts that terminals cannot transmit and receive simultaneously at the same frequency band, which enjoys much lower implementation complexity than the full-duplex relay. Moreover, the orthogonal transmission strategies of AF and DF protocols in [9] eliminates the potential interference between the source and the relay.

Specifically, for both AF and DF relays, each frame duration T_f is divided into two equal portions. During the first half period of frame duration, the source transmits to both the relay and the destination. In the second half period, the relay forwards the message to the destination, where the forward strategy depends on specific relay protocol used.

1) AF Protocol: In AF mode, the relay simply amplifies and then forwards what it receives during the first half to the destination. This strategy is also called non-regenerative relay or analog relay protocol [12][13]. Let P_s and P_r denote the average transmit power assigned to the source and the relay, respectively. Then, the achievable rate of AF protocol, denoted by R_{AF}, can be expressed as [9][1]

$$R_{AF} = \left(\frac{T_f B}{2}\right) \log_2 \left(1 + 2\gamma_1 P_s + \frac{4\gamma_2 P_s \gamma_3 P_r}{1 + 2\gamma_2 P_s + 2\gamma_3 P_r}\right) \tag{1}$$

where B denotes the system spectral bandwidth. Note that in Eq. (1), since each transmitter sends data only for half of the frame duration, the source uses power $2P_s$ during the first half and the relay uses power $2P_r$ during the second half, which results in a total average power of $P_s + P_r$ per frame.

2) DF Protocol: In DF mode, the relay forwards the message to the destination if it decodes successfully. Correspondingly, this strategy is also called regenerative relay or digital relay [12][13]. The achievable rate of DF protocol, denoted by R_{DF}, can be expressed as [9]

$$R_{DF} = \left(\frac{T_f B}{2}\right) \min\Big\{\log_2\Big(1 + 2\gamma_2 P_s\Big), \log_2\Big(1 + 2\gamma_1 P_s + 2\gamma_3 P_r\Big)\Big\}. \tag{2}$$

Throughout this paper, we further assume that the relay network has a mean total transmit power constraint, denoted by $\overline{P}$. Thus, the transmit power P_s and P_r need to satisfy

$$\mathbb{E}[P_s + P_r] \leq \overline{P} \tag{3}$$

where $\mathbb{E}[\cdot]$ denotes the expectation.

III. Preliminaries on Statistical QoS Guarantees

A. The QoS Exponent

During the early 90's, statistical QoS guarantees have been extensively studied in the contexts of effective bandwidth theory [19]–[23]. The literature on effective bandwidth is abundant. The readers are referred to Chang [19] and Kelly *et. al.* [20] for a comprehensive review.

Based on large deviation principle (LDP), Chang in [19] showed that for a dynamic queueing system with stationary ergodic arrival and service processes, under sufficient conditions, the queue length process $Q(t)$ converges in distribution to a random variable $Q(\infty)$ such that

$$-\lim_{x \to \infty} \frac{\log\left(\Pr\{Q(\infty) > x\}\right)}{x} = \theta. \tag{4}$$

To be more specific, the above theorem states that the probability of the queue length exceeding a certain threshold x decays exponentially fast as the threshold x increases. Note that in Eq. (4), the parameter θ ($\theta > 0$) plays a critically important role for statistical QoS guarantees, which indicates the exponential decay rate of the QoS violation probabilities. A smaller θ corresponds to a slower decay rate, which implies that the system can only provide a *looser* QoS guarantee, while a larger θ leads to a faster decay rate, which means that a more *stringent* QoS requirement can be supported. In particular, when $\theta \to 0$, the system can tolerate an arbitrarily long delay. On the other hand, when $\theta \to \infty$, the system cannot tolerate any delay. Due to its close relationship with statistical QoS provisioning, θ is called the *QoS exponent* [15]–[18].

B. The Effective Capacity

Inspired by the effective bandwidth theory, Wu and Negi proposed a powerful concept termed *effective capacity* [15]–[18]. The effective capacity is defined as the maximum constant arrival rate that a given service process can support in order to guarantee a QoS requirement specified by θ. Analytically, the effective capacity can be formally defined as follows.

[1]Note that in our model, the unit for the rate is "bits per frame".

Let the sequence $\{R[i],\ i = 1, 2, ...\}$ denote a discrete-time stationary and ergodic stochastic service process and $S[t] \triangleq \sum_{i=1}^{t} R[i]$ be the partial sum of the service process. Assume that the Gärtner-Ellis limit of $S[t]$, expressed as $\Lambda_{\rm C}(\theta) = \lim_{t\to\infty}(1/t)\log\left(\mathbb{E}\left\{e^{\theta S[t]}\right\}\right)$ exists and is a convex function differentiable for all real θ [19, pp. 921]. Then, the effective capacity of the service process, denoted by $\mathrm{E_C}(\theta)$, where $\theta > 0$, is defined as [16, eq. (12)]

$$\mathrm{E_C}(\theta) \triangleq -\frac{\Lambda_{\rm C}(-\theta)}{\theta} = -\lim_{t\to\infty}\frac{1}{\theta t}\log\Big(\mathbb{E}\left[e^{-\theta S[t]}\right]\Big). \quad (5)$$

When the sequence $\{R[i],\ i = 1, 2, ...\}$ is uncorrelated, it is clear that the effective capacity $\mathrm{E_C}(\theta)$ reduces to

$$\mathrm{E_C}(\theta) = -\frac{1}{\theta}\log\Big(\mathbb{E}\left[e^{-\theta R[i]}\right]\Big) = -\frac{1}{\theta}\log\Big(\mathbb{E}\left[e^{-\theta R[1]}\right]\Big). \quad (6)$$

The effective capacity expression Eq. (6) in uncorrelated case only depends on marginal statistics of a service process, which is much simpler than the general expression given by Eq. (5), where the higher order statistics of the service process are required. Since the block fading channel model generates an independent identically distributed (i.i.d.) service process, it can greatly simplify the effective capacity derivations.

In this paper, our original problem is maximizing the throughput subject to a given delay-QoS constraint. Notice that the effective capacity can be considered as the maximum throughput under the constraint of QoS exponent θ. Therefore, by interpreting θ as the QoS constraint in our original problem, we can formulate an equivalent new problem, which is to maximize the effective capacity for a given θ. In the following sections, we will focus on this new problem and design the corresponding resource-allocation algorithms.

IV. Dynamic Resource Allocation for AF Relay Networks

A. Problem Formulation

Conventionally, the resource allocation policy can be expressed as a function of the instantaneous network CSI $\boldsymbol{\gamma}$. In contrast, our resource allocation policy is a function of not only the instantaneous CSI $\boldsymbol{\gamma}$, but also the QoS exponent θ. Correspondingly, let us define $\boldsymbol{\nu} \triangleq (\theta, \boldsymbol{\gamma})$ as *network state information* (NSI). Then, we can rewrite Eq. (1) as

$$R_{AF}(\boldsymbol{\nu}) = \left(\frac{T_f B}{2}\right)\log_2\left(1 + 2\gamma_1 P_s(\boldsymbol{\nu}) + \frac{4\gamma_2 P_s(\boldsymbol{\nu})\gamma_3 P_r(\boldsymbol{\nu})}{1 + 2\gamma_2 P_s(\boldsymbol{\nu}) + 2\gamma_3 P_r(\boldsymbol{\nu})}\right) \quad (7)$$

where both the achievable rate $R_{AF}(\boldsymbol{\nu})$ and the average transmit power $P_s(\boldsymbol{\nu})$ and $P_r(\boldsymbol{\nu})$ are functions of the NSI $\boldsymbol{\nu}$. For a given QoS constraint specified by θ, in order to find the optimal resource allocation policy that maximizes the effective capacity of Eq. (6), we can formulate our maximization problem as follows:

$$(P1)\quad \arg\max_{\mathbf{P}(\boldsymbol{\nu})}\left\{-\frac{1}{\theta}\log\left(\mathbb{E}_{\boldsymbol{\gamma}}\left[\exp\Big(-\theta R_{AF}(\boldsymbol{\nu})\Big)\right]\right)\right\} \quad (8)$$

subject to the following power constraints:

$$\begin{cases} \mathbb{E}_{\boldsymbol{\gamma}}\left[P_s(\boldsymbol{\nu}) + P_r(\boldsymbol{\nu})\right] \le \overline{P} \\ P_s(\boldsymbol{\nu}) \ge 0 \text{ and } P_r(\boldsymbol{\nu}) \ge 0 \end{cases} \quad (9)$$

where we define $\mathbf{P}(\boldsymbol{\nu}) \triangleq (P_s(\boldsymbol{\nu}), P_r(\boldsymbol{\nu}))$ as network power allocation policy, and $\mathbb{E}_{\boldsymbol{\gamma}}[\cdot]$ emphasizes that the expectation is with respect to $\boldsymbol{\gamma}$. Since $\log(\cdot)$ is a monotonically increasing function, for each given QoS constraint $\theta > 0$, the maximization problem $(P1)$ is equivalent to the following minimization problem:

$$(P1')\quad \arg\min_{\mathbf{P}(\boldsymbol{\nu})}\left\{\mathbb{E}_{\boldsymbol{\gamma}}\left[\exp\Big(-\theta R_{AF}(\boldsymbol{\nu})\Big)\right]\right\} \quad (10)$$

subject to the same constraints given by Eq. (9). Problem $(P1')$ is still not easy to solve since the objective is not convex. To simplify the problem, we can make the following approximation at the high signal-to-noise ratio (SNR) regime:

$$1 + 2\gamma_2 P_s(\boldsymbol{\nu}) + 2\gamma_3 P_r(\boldsymbol{\nu}) \approx 2\gamma_2 P_s(\boldsymbol{\nu}) + 2\gamma_3 P_r(\boldsymbol{\nu}). \quad (11)$$

The high SNR approximation made by Eq. (11) is commonly used to simplify the derivations and obtain insight of the problem [26]. The difference between the actual value and the approximate value becomes negligible as the SNR increases. Substituting Eq. (11) into Eq. (7), we approximate rate $R_{AF}(\boldsymbol{\nu})$ by $\widetilde{R}_{AF}(\boldsymbol{\nu})$ as follows [13]:

$$\widetilde{R}_{AF}(\boldsymbol{\nu}) = \left(\frac{T_f B}{2}\right)\log_2\left(1 + 2\gamma_1 P_s(\boldsymbol{\nu}) + \frac{2\gamma_2 P_s(\boldsymbol{\nu})\gamma_3 P_r(\boldsymbol{\nu})}{\gamma_2 P_s(\boldsymbol{\nu}) + \gamma_3 P_r(\boldsymbol{\nu})}\right). \quad (12)$$

The approximation $\widetilde{R}_{AF}(\boldsymbol{\nu})$ in Eq. (12) takes the advantages of mathematical tractability over $R_{AF}(\boldsymbol{\nu})$ in Eq. (7). Specifically, $\widetilde{R}_{AF}(\boldsymbol{\nu})$ is strictly concave on the space spanned by $(P_s(\boldsymbol{\nu}), P_r(\boldsymbol{\nu}))$, which makes the related optimization much easier than the original problem $(P1')$. Furthermore, $\widetilde{R}_{AF}(\boldsymbol{\nu})$ serves as a tight upper-bound for $R_{AF}(\boldsymbol{\nu})$ at the high SNR regime. We will observe from our numerical results later that the approximation is also accurate even at moderate-low SNR regime.

Replacing $R_{AF}(\boldsymbol{\nu})$ by $\widetilde{R}_{AF}(\boldsymbol{\nu})$ in Eq. (10), we get the following optimization problem ready to be solved:

$$\begin{aligned}(P1'')\quad &\arg\min_{\mathbf{P}(\boldsymbol{\nu})}\left\{\mathbb{E}_{\boldsymbol{\gamma}}\left[\exp\Big(-\theta\widetilde{R}_{AF}(\boldsymbol{\nu})\Big)\right]\right\} \\ &= \arg\min_{\mathbf{P}(\boldsymbol{\nu})}\left\{\mathbb{E}_{\boldsymbol{\gamma}}\left[\left(1 + 2\gamma_1 P_s(\boldsymbol{\nu}) + \frac{2\gamma_2 P_s(\boldsymbol{\nu})\gamma_3 P_r(\boldsymbol{\nu})}{\gamma_2 P_s(\boldsymbol{\nu}) + \gamma_3 P_r(\boldsymbol{\nu})}\right)^{-\frac{\beta}{2}}\right]\right\}\end{aligned} \quad (13)$$

subject to the constraints given by Eq. (9), where we define

$$\beta \triangleq \frac{\theta T_f B}{\log 2}$$

as the *normalized* QoS exponent.

B. Resource Allocation Policy

The following theorem solves the optimization problem $(P1'')$ derived in the above:

Theorem 1: The optimal resource allocation policy $\mathbf{P}(\boldsymbol{\nu})$ that solves $(P1'')$ is determined by

$$\begin{cases} P_s(\boldsymbol{\nu}) = uP_r(\boldsymbol{\nu}) \\ P_r(\boldsymbol{\nu}) = \dfrac{1}{v}\left(\left[\left(\dfrac{\gamma_0}{\gamma_3}\right)\left(\dfrac{\gamma_3+c}{\gamma_1+c}\right)^2\right]^{-\frac{2}{\beta+2}} - 1\right), \end{cases} \tag{14}$$

if both $P_s(\boldsymbol{\nu}) > 0$ and $P_r(\boldsymbol{\nu}) > 0$ in Eq. (14), where the parameters u and v are defined by

$$\begin{cases} u = \frac{\gamma_3(\gamma_1+c)}{(\gamma_3-\gamma_1)\gamma_2} \\ v = \frac{2c\gamma_3(\gamma_1+c)^2}{(\gamma_3-\gamma_1)\gamma_2(\gamma_3+c)}, \end{cases} \tag{15}$$

with $c = \sqrt{\gamma_1\gamma_3 + \gamma_2\gamma_3 - \gamma_1\gamma_2}$, and γ_0 is a cutoff threshold determined by the mean total network power constraint.

Otherwise, the policy reduces to direct transmission and $\mathbf{P}(\boldsymbol{\nu})$ is determined by

$$\begin{cases} P_s(\boldsymbol{\nu}) = \dfrac{1}{2}\left[\left(\gamma_0^{\frac{2}{\beta+2}}\gamma_1^{\frac{\beta}{\beta+2}}\right)^{-1} - {\gamma_1}^{-1}\right]^+ \\ P_r(\boldsymbol{\nu}) = 0 \end{cases} \tag{16}$$

where $[x]^+ \triangleq \max\{x, 0\}$.

Proof: The proof is provided in Appendix I. ■

As mentioned in Section IV-A, the above solution for the problem $(P1'')$ serves as a tight upper-bound for the optimal effective capacity. On the other hand, by applying the above solution directly to the original problem $(P1)$, we can obtain a lower-bound for the optimal effective capacity (since it is an achievable effective capacity). We will see by the numerical examples later that the upper-bound and lower-bound are very close, especially at the high SNR regime.

Since a necessary condition for the resource allocation to take the form of Eq. (14) is $P_s(\boldsymbol{\nu}) > 0$, which in turn requests $u > 0$, implying $\gamma_3 > \gamma_1$, we therefore have the following corollary.

Corollary 1: For AF protocol, if $\gamma_1 \geq \gamma_3$ (the direct S-D link is better than the relay R-D link), then the optimal resource allocation reduces to the direct transmission, no matter what the QoS constraint θ is. ■

C. Limiting Resource Allocation Policies

As explained in Section III, the dynamics of the QoS exponent θ characterizes how stringent the QoS requirement is. We showed in [24] that as the QoS exponent $\theta \to 0$, the optimal effective capacity approaches the ergodic capacity of the system. On the other hand, as the QoS exponent $\theta \to \infty$, the optimal effective capacity approaches the zero-outage capacity of the system. Making use of these properties, we can obtain the resource allocation policy that characterizes the upper- and lower-bounds for the ergodic and zero-outage capacity of the AF relay protocol.

Proposition 1: The resource allocation policy that can upper- and lower-bound the ergodic capacity of the AF relay protocol is determined by

$$\begin{cases} P_s(\boldsymbol{\nu}) = uP_r(\boldsymbol{\nu}) \\ P_r(\boldsymbol{\nu}) = \dfrac{1}{v}\left[\dfrac{\gamma_3}{\gamma_0}\left(\dfrac{\gamma_1+c}{\gamma_3+c}\right)^2 - 1\right], \end{cases} \tag{17}$$

if both $P_s(\boldsymbol{\nu}) > 0$ and $P_r(\boldsymbol{\nu}) > 0$ in Eq. (17); otherwise, it reduces to the direct transmission (water-filling) as

$$\begin{cases} P_s(\boldsymbol{\nu}) = \dfrac{1}{2}\left[\dfrac{1}{\gamma_0} - \dfrac{1}{\gamma_1}\right]^+ \\ P_r(\boldsymbol{\nu}) = 0. \end{cases} \tag{18}$$

Proof: Letting $\theta \to 0$ in Eqs. (14) and (16), we can obtain the desired results. ■

Proposition 2: The resource allocation policy that can upper- and lower-bound the zero-outage capacity of the AF relay protocol is given by

$$\begin{cases} P_s(\boldsymbol{\nu}) = uP_r(\boldsymbol{\nu})\mathcal{I}\{\gamma_3 > \gamma_1\} + \frac{\sigma}{2\gamma_1}\mathcal{I}\{\gamma_3 \leq \gamma_1\} \\ P_r(\boldsymbol{\nu}) = \dfrac{\sigma}{v}\mathcal{I}\{\gamma_3 > \gamma_1\} \end{cases} \tag{19}$$

where $\mathcal{I}\{\cdot\}$ denotes the indicator function, and σ is a constant such that the mean total network power constraint is satisfied. Under such a policy, the transmission maintains a constant service rate $(T_fB/2)\log_2(1+\sigma)$, no matter what the channel realization $\boldsymbol{\gamma}$ is.

Proof: Letting $\theta \to \infty$ in Eqs. (14) and (16), we can obtain the desired results, where the constant σ is determined by $\sigma = \lim_{\theta\to\infty} \gamma_0^{-\frac{2}{\beta+2}} - 1$. ■

V. Dynamic Resource Allocation for DF Relay Networks

A. Resource Allocation for Original DF Protocol

Similar to the AF case, we first re-write Eq. (2) as a function of the NSI $\boldsymbol{\nu}$ as follows:

$$R_{DF}(\boldsymbol{\nu}) = \left(\frac{T_fB}{2}\right)\min\Big\{\log_2\Big(1+2\gamma_2P_s(\boldsymbol{\nu})\Big), \log_2\Big(1+2\gamma_1P_s(\boldsymbol{\nu})+2\gamma_3P_r(\boldsymbol{\nu})\Big)\Big\}. \tag{20}$$

Then, the optimization problem can be formulated as

$$(P2) \quad \arg\max_{\mathbf{P}(\boldsymbol{\nu})}\left\{-\frac{1}{\theta}\log\Big(\mathbb{E}_{\boldsymbol{\gamma}}\left[\exp\Big(-\theta R_{DF}(\boldsymbol{\nu})\Big)\right]\Big)\right\} \tag{21}$$

subject to the constraint given by Eq. (9). Again, the above maximization problem $(P2)$ is equivalent to the following minimization problem:

$$(P2') \quad \arg\min_{\mathbf{P}(\boldsymbol{\nu})}\Big\{\mathbb{E}_{\boldsymbol{\gamma}}\Big[\max\big\{\mathcal{F}_1(\boldsymbol{\nu}), \mathcal{F}_2(\boldsymbol{\nu})\big\}\Big]\Big\} \tag{22}$$

where

$$\mathcal{F}_1(\boldsymbol{\nu}) = \Big(1+2\gamma_2P_s(\boldsymbol{\nu})\Big)^{-\frac{\beta}{2}} \tag{23}$$

and

$$\mathcal{F}_2(\boldsymbol{\nu}) = \Big(1+2\gamma_1P_s(\boldsymbol{\nu})+2\gamma_3P_r(\boldsymbol{\nu})\Big)^{-\frac{\beta}{2}}. \tag{24}$$

It is easy to show that $(P2')$ is a strictly convex optimization problem and thus has the unique optimal solution. To solve $(P2')$, we consider the following two scenarios.

Scenario-1: $\gamma_2 < \gamma_1$.

If $\gamma_2 < \gamma_1$, then $\mathcal{F}_1(\boldsymbol{\nu}) > \mathcal{F}_2(\boldsymbol{\nu})$ always holds, no matter what the value of $P_r(\boldsymbol{\nu})$ is. To save the transmit power, the

optimal resource allocation strategy must satisfy $P_r(\boldsymbol{\nu}) = 0$. As a result, problem $(P2')$ becomes:

$$\arg \min_{\mathbf{P}(\boldsymbol{\nu})} \Big\{ \mathbb{E}_{\boldsymbol{\gamma}} \left[\mathcal{F}_1(\boldsymbol{\nu}) \right] \Big\} \tag{25}$$

subject to $E_{\boldsymbol{\gamma}}[P_s(\boldsymbol{\nu})] = \overline{P}$. This is equivalent to a direct transmission problem, where the transmission link is from the source to the relay. The above problem has been solved by our previous work [24]. The optimal resource allocation policy is determined by:

$$\begin{cases} P_s(\boldsymbol{\nu}) = \dfrac{1}{2} \left[\left(\gamma_0^{\frac{2}{\beta+2}} \gamma_2^{\frac{\beta}{\beta+2}} \right)^{-1} - \gamma_2{}^{-1} \right]^+ \\ P_r(\boldsymbol{\nu}) = 0. \end{cases} \tag{26}$$

Scenario-2: $\gamma_2 \geq \gamma_1$.

If $\gamma_2 \geq \gamma_1$, then we can find appropriate $P_s(\boldsymbol{\nu})$ and $P_r(\boldsymbol{\nu})$ such that

$$\mathcal{F}_1(\boldsymbol{\nu}) = \mathcal{F}_2(\boldsymbol{\nu}), \tag{27}$$

which in turn gives

$$P_r(\boldsymbol{\nu}) = \tilde{u} P_s(\boldsymbol{\nu}) \tag{28}$$

where $\tilde{u} = (\gamma_2 - \gamma_1)/\gamma_3$. In this case, the objective function of the problem $(P2')$ is the same as that given in Eq. (25), but subject to the constraints given by Eqs. (9) and (28). Thus, we can construct the following Lagrangian problem as

$$\begin{aligned} \mathcal{J}_2 &= \mathbb{E}_{\boldsymbol{\gamma}} \left[\mathcal{F}_1(\boldsymbol{\nu}) \right] + \lambda \mathbb{E}_{\boldsymbol{\gamma}} \left[P_s(\boldsymbol{\nu}) + P_r(\boldsymbol{\nu}) \right] \\ &= \mathbb{E}_{\boldsymbol{\gamma}} \left[\Big(1 + 2\gamma_2 P_s(\boldsymbol{\nu}) \Big)^{-\frac{\beta}{2}} \right] + \lambda \mathbb{E}_{\boldsymbol{\gamma}} \left[(1+\tilde{u}) P_s(\boldsymbol{\nu}) \right]. \end{aligned}$$

Solving the above Lagrangian problem, we obtain the resource allocation policy under the condition of $\gamma_2 \geq \gamma_1$ as

$$\begin{cases} P_s(\boldsymbol{\nu}) = \dfrac{1}{2} \left[\left(\left[(1+\tilde{u}) \gamma_0 \right]^{\frac{2}{\beta+2}} \gamma_2^{\frac{\beta}{\beta+2}} \right)^{-1} - \gamma_2^{-1} \right]^+ \\ P_r(\boldsymbol{\nu}) = \tilde{u} P_s(\boldsymbol{\nu}). \end{cases} \tag{29}$$

In summary, the optimal resource allocation policy for the original DF protocol is given by either Eq. (26) or Eq. (29), depending on whether $\gamma_2 < \gamma_1$ or not.

To study the zero-outage capacity of DF relay networks, we let $\theta \to \infty$, and the corresponding resource allocation policy can be expressed as

$$\begin{cases} P_s(\boldsymbol{\nu}) = \dfrac{\sigma}{2\gamma_2} \\ P_r(\boldsymbol{\nu}) = \tilde{u} P_s(\boldsymbol{\nu}) \mathcal{I}\{\gamma_2 \geq \gamma_1\}. \end{cases} \tag{30}$$

Similar to the AF case, under such a policy, the transmission maintains a constant service rate $(T_f B/2) \log_2(1+\sigma)$, no matter what the channel realization $\boldsymbol{\gamma}$ is.

It is important to notice that when $\theta \to 0$, the corresponding resource allocation policy does not lead to the ergodic capacity of DF relay networks, since the ergodic capacity of DF relay networks is determined by

$$\begin{aligned} \mathcal{C} &= \left(\frac{T_f B}{2} \right) \max_{\mathbf{P}(\boldsymbol{\nu})} \min \Big\{ \mathbb{E}_{\boldsymbol{\gamma}} \left[\log_2 \Big(1 + 2\gamma_2 P_s \Big) \right], \\ &\qquad \mathbb{E}_{\boldsymbol{\gamma}} \left[\log_2 \Big(1 + 2\gamma_1 P_s + 2\gamma_3 P_r \Big) \right] \Big\}, \end{aligned} \tag{31}$$

but the optimal effective capacity of our scheme at $\theta \to 0$ is given by

$$\begin{aligned} \mathcal{C}' &= \left(\frac{T_f B}{2} \right) \max_{\mathbf{P}(\boldsymbol{\nu})} \mathbb{E}_{\boldsymbol{\gamma}} \Bigg[\min \Big\{ \log_2 \Big(1 + 2\gamma_2 P_s \Big), \\ &\qquad \log_2 \Big(1 + 2\gamma_1 P_s + 2\gamma_3 P_r \Big) \Big\} \Bigg]. \end{aligned} \tag{32}$$

By Jensen's inequality, $\mathcal{C} \geq \mathcal{C}'$ always holds. From an implementation perspective, Eqs. (31) and (32) correspond to two different transmission strategies for the relay node [11]. On one hand, it can immediately transmit a received and decoded frame to the destination whenever it receives the frame from the source, which corresponds to Eq. (32). On the other hand, it can also queue data and then transmit the queued contents when the channel is favorable, which corresponds to Eq. (31). Since in our system model, the relay strategy falls into the first category, our obtained effective capacity is no greater than the ergodic capacity. However, this strategy is more practical because it results in shorter delay. The readers are referred to [11] for detailed discussions about resource allocation to achieve the ergodic capacity of relay networks.

B. Improved DF Protocol for Stringent QoS Provisioning

One major drawback of the original DF relay protocol, which we will call the protocol $(R0)$ hereafter, is that it cannot support stringent QoS requirement for any non-zero arrival process. We have the following proposition to formally characterize this problem.

Proposition 3: As the QoS exponent $\theta \to \infty$, the optimal effective capacity for protocol $(R0)$ approaches zero, no matter how much spectral bandwidth and power resources are assigned for the transmission.

Proof: The proof is provided in Appendix II-A. ■

To be more specific, Proposition 3 states that the zero-outage capacity of protocol $(R0)$ is zero. Intuitively, from Eq. (2) we can observe that the performance of the protocol $(R0)$ is upper-bounded by the direct transmission from the source to the relay. However, it is well known that when each terminal has a single antenna, direct transmission cannot achieve zero outage probability with a finite average power limitation. Therefore, in order to improve the performance of DF relay for stringent QoS guarantees, we need modify the original protocol $(R0)$.

1) Protocol $(R1)$: A straight-forward idea of modification is to improve the performance of DF relay under the case of $\gamma_1 > \gamma_2$ (i.e., **Scenario-1**). Since in this case, the performance of DF relay is always worse than the direct transmission, we can use direct transmission instead of relay.

The optimal resource allocation policy for protocol $(R1)$ can be described as follows:

- If $\gamma_1 > \gamma_2$, then the resource allocation is determined by Eq. (16).
- Otherwise, the resource allocation is determined by Eq. (29).

Unfortunately, even under this revision, the resulting DF protocol still cannot support stringent QoS requirement as $\theta \to \infty$, as proved by Appendix II-B.

2) Protocol $(R2)$: The major reason that protocol $(R1)$ still cannot support stringent QoS is that it does not provide *diversity* for the link from relay to the destination. To overcome this problem, we need to use the direct transmission instead of relay when either γ_2 or γ_3 is less than γ_1. This strategy in fact provide *selection diversity* to the R-D link. The optimal resource allocation policy for protocol $(R2)$ can be described as follows:

- If $\gamma_1 > \gamma_2$ or $\gamma_1 > \gamma_3$, then the resource allocation is determined by Eq. (16).
- Otherwise, the resource allocation is determined by Eq. (29).

Proposition 4: As the QoS exponent $\theta \to \infty$, the optimal effective capacity for protocol $(R2)$ approaches a non-zero constant rate with a finite mean total power constraint given by Eq. (3).

Proof: The proof is provided in Appendix II-C. ■

A similar protocol is called "opportunistic cooperative" in [10], where the focus is mainly on the outage probability minimization.

VI. Fixed Power Allocation for DF Relay Networks

In previous sections, we assume that the source and the relay nodes can dynamically allocate the transmit power under a mean total power constraint. In this section, we study the case where they do not have the ability to perform temporal power allocation. Let $P_s = \kappa\overline{P}$ and $P_r = (1-\kappa)\overline{P}$ be the power assigned to source and relay, respectively, where $\kappa \in (0,1)$. Then, our goal is to find an optimal κ that can maximize the effective capacity under the constraint of a given QoS exponent θ.

Rewrite the rate in Eq. (2) for the DF protocol as follows:

$$\begin{aligned} R_{DF}(\kappa,\boldsymbol{\gamma}) &= \left(\frac{T_f B}{2}\right)\min\Big\{\log_2\Big(1+2\gamma_2\overline{P}\kappa\Big), \\ &\qquad \log_2\Big(1+2\gamma_1\overline{P}\kappa+2\gamma_3\overline{P}(1-\kappa)\Big)\Big\} \\ &= \left(\frac{T_f B}{2}\right)\log_2\Big(1+\widetilde{\gamma}\Big) \end{aligned} \tag{33}$$

where we define $\widetilde{\gamma} = 2\overline{P}\min\{\gamma_2\kappa, \gamma_1\kappa+\gamma_3(1-\kappa)\}$. Following the similar idea to the previous sections, we can formulate the optimization problem as

$$\begin{aligned} (P3) \qquad &\max_{\kappa\in(0,1)}\left\{-\frac{1}{\theta}\log\Big(\mathbb{E}_{\boldsymbol{\gamma}}\left[\exp\Big(-\theta R_{DF}(\kappa,\boldsymbol{\gamma})\Big)\right]\Big)\right\} \\ &= \max_{\kappa\in(0,1)}\left\{-\frac{1}{\theta}\log\Big(\mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right)\Big)\right\}. \end{aligned} \tag{34}$$

In Eq. (34), we define

$$\mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right) = \int_0^{+\infty}\Big(1+\widetilde{\gamma}\Big)^{-\frac{\beta}{2}} p_{\widetilde{\Gamma}}(\widetilde{\gamma})d\widetilde{\gamma} \tag{35}$$

where $\boldsymbol{\lambda} = (\lambda_1,\lambda_2,\lambda_3)$ with λ_i denoting the parameter of the exponential distribution for the channel gain γ_i, and $p_{\widetilde{\Gamma}}(\widetilde{\gamma})$ denotes the probability density function (pdf) of $\widetilde{\gamma}$, which can be derived by using the following proposition.

Proposition 5: Let $X \triangleq \min\{X_2, X_1+X_3\}$, where X_i is an independent exponential distributed random variable with parameter μ_i for $i \in \{1,2,3\}$. Then the pdf of X, denoted by $p_X(x)$, can be expressed as follows:

$$p_X(x) = \begin{cases} \frac{\mu_3(\mu_1+\mu_2)}{\mu_3-\mu_1}e^{-(\mu_1+\mu_2)x} + \frac{\mu_1(\mu_2+\mu_3)}{\mu_1-\mu_3}e^{-(\mu_2+\mu_3)x}, \\ \qquad\qquad \text{if } \mu_1 \neq \mu_3 \\ \left[\mu_2+\mu_1(\mu_1+\mu_2)x\right]e^{-(\mu_1+\mu_2)x}, \\ \qquad\qquad \text{if } \mu_1 = \mu_3. \end{cases} \tag{36}$$

Proof: The proof follows by the direct derivations. ■

Corollary 2: The pdf of $\widetilde{\gamma}$, denoted by $p_{\widetilde{\Gamma}}(\widetilde{\gamma})$, can be obtained directly by letting $\mu_1 = \lambda_1/(2\overline{P}\kappa)$, $\mu_2 = \lambda_2/(2\overline{P}\kappa)$, and $\mu_3 = \lambda_3/[2\overline{P}(1-\kappa)]$ in Eq. (36). ■

The integral in Eq. (35) can be calculated by using the results given in [27, Sec. 3.383.5]. After some algebraic manipulations, we can obtain the closed-form expression for $\mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right)$ given in Eq. (35) as follows:

$$\begin{aligned} \mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right) &= -\frac{\lambda_1[\lambda_2(1-\kappa)+\lambda_3\kappa]}{2\overline{P}\kappa[\lambda_3\kappa-\lambda_1(1-\kappa)]} \\ &\quad\cdot\exp\left(\frac{\lambda_2(1-\kappa)+\lambda_3\kappa}{2\overline{P}\kappa(1-\kappa)}\right)E_{\frac{\beta}{2}}\left(\frac{\lambda_2(1-\kappa)+\lambda_3\kappa}{2\overline{P}\kappa(1-\kappa)}\right) \\ &\quad+\frac{(\lambda_1+\lambda_2)\lambda_3}{2\overline{P}[\lambda_3\kappa-\lambda_1(1-\kappa)]}\exp\left(\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right)E_{\frac{\beta}{2}}\left(\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right) \end{aligned} \tag{37}$$

if $(1-\kappa)\lambda_1 \neq \kappa\lambda_3$. Otherwise, if $(1-\kappa)\lambda_1 = \kappa\lambda_3$, we get

$$\begin{aligned} \mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right) &= \frac{\lambda_2}{2\overline{P}\kappa}\exp\left(\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right)E_{\frac{\beta}{2}}\left(\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right) \\ &\quad+\frac{\lambda_1(\lambda_1+\lambda_2)}{4\overline{P}^2\kappa^2}\left[\frac{\Gamma\left(\frac{\beta}{2}-2\right)}{\Gamma\left(\frac{\beta}{2}\right)}{}_1F_1\left(2,3-\frac{\beta}{2},\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right)\right. \\ &\quad\left.+\Gamma\left(2-\frac{\beta}{2}\right)\left(\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right)^{\frac{\beta}{2}-2}{}_1F_1\left(\frac{\beta}{2},\frac{\beta}{2}-1,\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right)\right] \end{aligned} \tag{38}$$

where $\Gamma(\cdot)$ denotes the Gamma function, $E_\nu(\cdot)$ denotes the νth order exponential integral function, and ${}_1F_1(\cdot,\cdot,\cdot)$ denotes the confluent hypergeometric function.

At the high SNR regime, using the Taylor expansion $e^x = 1+x+x^2/2+\cdots$, and the asymptotic property of the exponential integral function [28]

$$\lim_{z\to 0}E_\nu(z) = \Gamma(1-\nu)z^{\nu-1} - \frac{1}{1-\nu} \tag{39}$$

it turns out that Eq. (37) can be simplified as

$$\begin{aligned} \lim_{\overline{P}\to\infty}\mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right) &= \frac{\lambda_2}{\overline{P}\kappa(\beta-2)} + \frac{\kappa(1-\kappa)\Gamma\left(1-\frac{\beta}{2}\right)}{\lambda_3\kappa-\lambda_1(1-\kappa)} \\ &\quad\cdot\left[\frac{\lambda_3}{1-\kappa}\left(\frac{\lambda_1+\lambda_2}{2\overline{P}\kappa}\right)^{\frac{\beta}{2}} - \frac{\lambda_1}{\kappa}\left(\frac{\lambda_2(1-\kappa)+\lambda_3\kappa}{2\overline{P}\kappa(1-\kappa)}\right)^{\frac{\beta}{2}}\right]. \end{aligned} \tag{40}$$

However, it is still difficult to solve the optimal κ^* explicitly, due to the intractability of $\mathcal{G}\left(\kappa,\theta,\boldsymbol{\lambda},\overline{P}\right)$. Therefore, we will use the numerical search to obtain the optimal $\kappa^* \in (0,1)$.

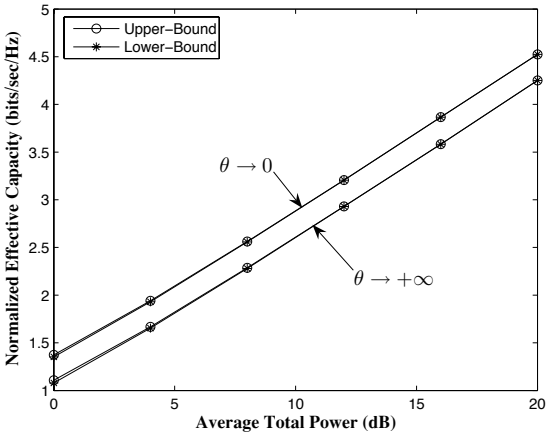


Fig. 3. The effective capacity upper-bound and lower-bound for AF protocol under optimal resource allocation policies. The S-R distance is set to $d = 0.5$

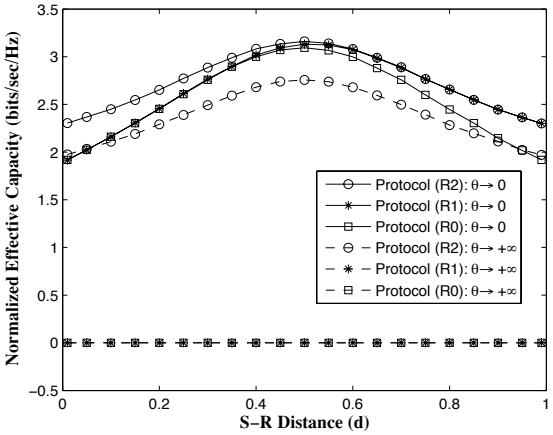


Fig. 4. The effective capacity of different DF protocols under optimal resource allocation. The average total power is 10 dB.

VII. Simulations and Numerical Evaluations

We evaluate the performance of our proposed cross-layer resource allocation scheme over wireless relay networks by simulations and numerical analyses. In the following, we set the product of frame duration and spectral bandwidth $T_f B/\log(2) = 1$ such that $\beta = \theta$. The other system parameters are detailed respectively in each of the figures.

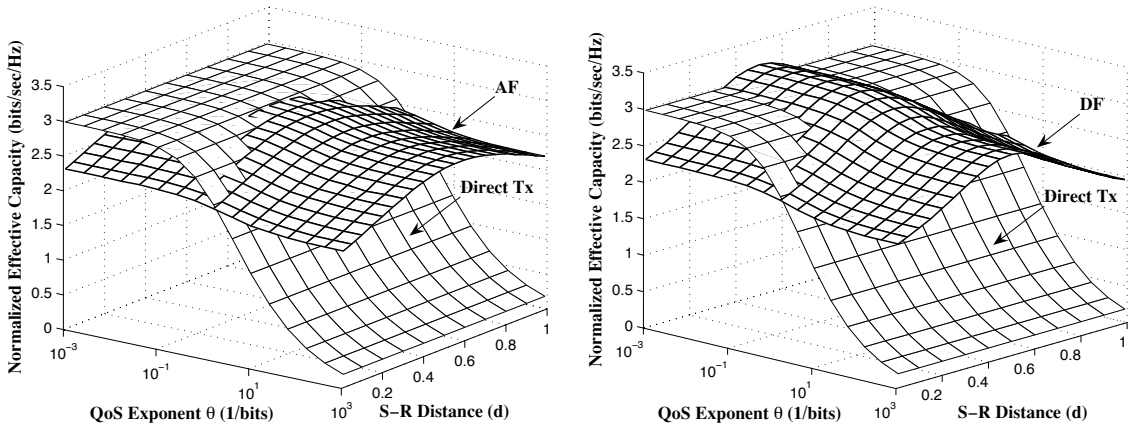
Using Propositions 1 and 2, Fig. 3 plots the effective capacity upper- and lower-bounds for the AF protocol by using our proposed resource allocation policy. We can observe from Fig. 3 that for both the loose QoS constraint ($\theta \to 0$) and stringent QoS constraint ($\theta \to \infty$), the upper- and lower-bounds are very close to each other. In particular, at the high SNR regime, the upper- and lower-bounds are indistinguishable, indicating that the deployed resource allocation policy essentially achieves the optimality. Thus, in the following, we will only plot the lower-bound of the effective capacity for AF relay protocol for simplicity.

Fig. 4 plots the optimal effective capacities for different DF relay protocols. As verified by Fig. 4, when the QoS is loose ($\theta \to 0$), protocol $(R2)$ achieves the best effective capacity and the original protocol $(R0)$ attains the worst performance among the three protocols. The performance of protocol $(R1)$ approaches protocol $(R0)$ when the relay is close to the source and approaches protocol $(R2)$ when the relay is close to the destination. On the other hand, as the QoS constraint becomes stringent ($\theta \to \infty$), only protocol $(R2)$ can achieve a nonzero effective capacity. For both protocols $(R0)$ and $(R1)$, the optimal effective capacity approaches zero, as pointed out by our analytical analyses. Thus, in the following, we will only plot the effective capacity by using protocol $(R2)$.

Fig. 5 shows the optimal effective capacities for AF and DF protocols as a function of the QoS exponent θ and the S-R distance d. For comparison purpose, we also plot the optimal effective capacity obtained by using direct transmission [24]. Since there is no relay node for the direct transmission scheme, the performance of direct transmission is independent of parameter d. We can observe from Fig. 5 that when the QoS constraint is loose, the effective capacities of AF and DF are close to that of direct transmission. However, as the QoS constraint become more stringent, the two relay protocols both show significant advantages over direct transmission. The performance comparisons of AF and DF relay are plotted in Fig. 6 in terms of the ratio of the effective capacity for DF relay to that for AF relay. We can observe that the two protocols show the similar performances. DF relay performs relatively better when the relay is close to the source, while AF relay performs better when the relay is close to the destination. Note that when the QoS constraint is loose, the direct transmission may outperform relay transmission, which is due to the fact that both relay protocols operate in half-duplex mode and only utilize half of the degree of freedom. When more powerful relay protocols are employed, e.g., the relay protocols proposed in [10][11], the performance of relay transmission is expected to show significant gain over direct transmission for both loose and stringent QoS provisioning.

Fig. 7(a) numerically plots the optimal power assignment κ^* for DF relay protocol under fixed power allocation policy. By applying the optimal power assignment, the resulting effective capacity is shown in Fig. 7(b). We can observe from Fig. 7(b) that even using the optimal power assignment, the effective capacity also converges to zero as the QoS exponent $\theta \to \infty$. However, this is the best that the fixed power allocation can do to maximize the effective capacity. This implies that no matter how much power and spectral bandwidth resource are assigned and no matter how elegant coding/modulation is employed, the fixed power allocation cannot support stringent QoS over Rayleigh flat-fading channels.

To compare the performance of dynamic and static resource allocations, the effective capacity gain of protocol $(R2)$ over fixed power assignment is shown in Fig. 8. We can see from Fig. 8 that for loose QoS constraint, the performance of the two strategies are similar. The effective capacity by using dynamic resource allocation is only slightly better than that by using fixed power assignment. However, as the QoS constraint becomes stringent, the dynamic resource allocation significantly outperforms the fixed power allocation, which confirms the importance of employing the dynamic resource allocation for stringent QoS provisioning.



(a) AF protocol. (b) DF protocol.

Fig. 5. The effective capacity of AF and DF schemes under optimal resource allocation policies. The average total power is 10 dB.

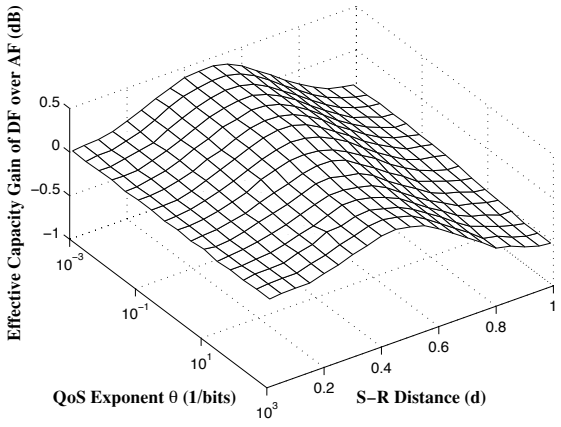


Fig. 6. The effective capacity gain ratio of DF protocol over AF protocol under optimal resource allocation. The average total power is 10 dB.

VIII. Conclusions

We proposed and analyzed the cross-layer resource allocation scheme for relay networks to guarantee diverse QoS requirements. By integrating information theory with the effective capacity, our proposed cross-layer scheme characterizes the delay QoS constraint by the QoS exponent, which turns out to be a simple and efficient approach in the cross-layer design and optimization. Over both AF and DF wireless relay networks, we developed the associated resource allocation algorithms. The simulation and numerical results verified that our proposed cross-layer resource allocation can efficiently support diverse QoS requirements. On the other hand, even for simple AF and DF protocols, the relay transmission shows significant advantages over direct transmissions when the delay constraints become stringent.

In this paper, our focus is mainly on how to apply the effective-capacity-based approach to wireless relay networks as an efficient cross-layer design strategy, whereas the problem of what is the optimal relay protocol is beyond the scope of this paper. It is worth noting that the performances of different relay protocols are significantly different. When employing more powerful relay protocols, e.g., those proposed and studied in [10][11], the network performance can be much better. However, the cross-layer resource allocation scheme developed in this paper can be readily extended to the other scenarios using the more powerful relay protocols.

Appendix I
Proof of Theorem 1

Proof: Based on the concavity of $\widetilde{R}_{AF}(\boldsymbol{\nu})$, it is easy to show that $(P1'')$ is a strictly convex optimization problem and therefore has the unique optimal solution. Construct the Lagrange as follows[2]:

$$\mathcal{J}_1 = \mathbb{E}_{\boldsymbol{\gamma}}\left[\left(1+2\gamma_1 P_s(\boldsymbol{\nu})+\frac{2\gamma_2 P_s(\boldsymbol{\nu})\gamma_3 P_r(\boldsymbol{\nu})}{\gamma_2 P_s(\boldsymbol{\nu})+\gamma_3 P_r(\boldsymbol{\nu})}\right)^{-\frac{\beta}{2}}\right] + \lambda\mathbb{E}_{\boldsymbol{\gamma}}\left[P_s(\boldsymbol{\nu})+P_r(\boldsymbol{\nu})\right]. \quad (41)$$

If there exists the solution $\mathbf{P}(\boldsymbol{\nu})$ such that both $P_s(\boldsymbol{\nu})>0$ and $P_r(\boldsymbol{\nu})>0$, then according to Karush-Kuhn-Tucker (KKT) condition we get

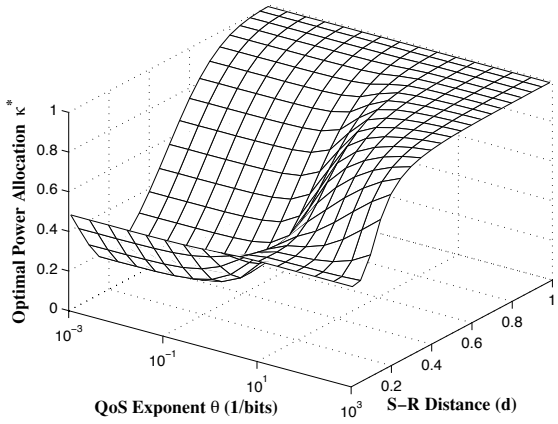
$$\frac{\partial \mathcal{J}_1}{\partial P_i(\boldsymbol{\nu})}=0, \text{ for } i\in\{s,r\} \quad (42)$$

which yields

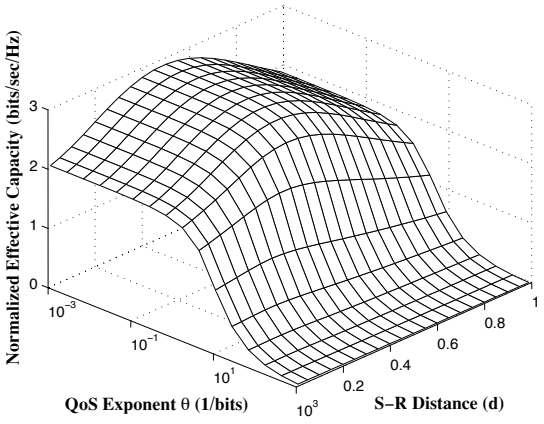
$$\begin{cases} \left(\gamma_1+\frac{\gamma_2\gamma_3^2P_r^2(\boldsymbol{\nu})}{[\gamma_2P_s(\boldsymbol{\nu})+\gamma_3P_r(\boldsymbol{\nu})]^2}\right) \\ \quad\cdot\left(1+2\gamma_1P_s(\boldsymbol{\nu})+\frac{2\gamma_2P_s(\boldsymbol{\nu})\gamma_3P_r(\boldsymbol{\nu})}{\gamma_2P_s(\boldsymbol{\nu})+\gamma_3P_r(\boldsymbol{\nu})}\right)^{-1-\frac{\beta}{2}}=\gamma_0 \\ \frac{\gamma_2^2\gamma_3P_s^2(\boldsymbol{\nu})}{[\gamma_2P_s(\boldsymbol{\nu})+\gamma_3P_r(\boldsymbol{\nu})]^2} \\ \quad\cdot\left(1+2\gamma_1P_s(\boldsymbol{\nu})+\frac{2\gamma_2P_s(\boldsymbol{\nu})\gamma_3P_r(\boldsymbol{\nu})}{\gamma_2P_s(\boldsymbol{\nu})+\gamma_3P_r(\boldsymbol{\nu})}\right)^{-1-\frac{\beta}{2}}=\gamma_0 \end{cases} \quad (43)$$

where $\gamma_0 \triangleq \lambda/\beta$. Solving Eq. (43), we can obtain Eq. (14). Note that Eq. (14) is a feasible solution when $u>0$ and $P_r(\boldsymbol{\nu})>0$. Otherwise, the AF protocol reduces to direct transmission, and thus the problem can be solved by the similar approach used in [24], which leads to Eq. (16). Finally,

[2]The explicit Lagrangian multipliers corresponding to the constraints $P_i(\boldsymbol{\nu})\geq 0$ are omitted.



(a) Optimal power allocation κ^*.



(b) Optimal effective capacity.

Fig. 7. The effective capacity of DF relay with optimal fixed power allocation. The average total power is equal to 10 dB.

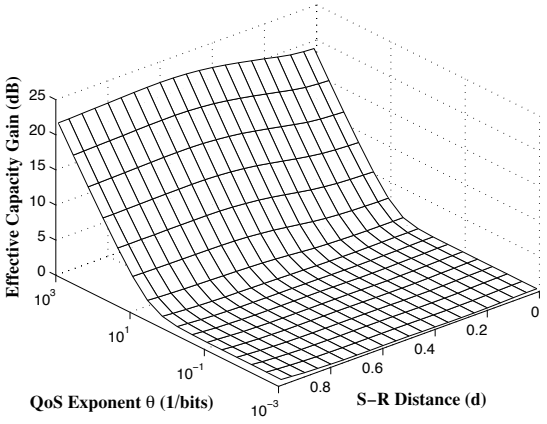


Fig. 8. The effective capacity gain of dynamic resource allocation over fixed power allocation for DF protocol. The average total power is 10 dB.

the parameter γ_0 is determined by the mean total network power constraint. The proof follows. ■

APPENDIX II
POWER LIMIT FOR DIFFERENT DF RELAY PROTOCOLS

A. Protocol $(R0)$ (Proof of Proposition 3)

Proof: As $\theta \to \infty$, the optimal resource allocation policy for the original DF relay protocol $(R0)$ becomes Eq. (30). To prove Proposition 3, it is equivalent to show that σ in Eq. (30) is always equal to zero for any finite power constraint $\overline{P}$.

Using Eq. (30), it is easy to show that the total transmit power can be expressed as

$$\begin{aligned} P_s(\boldsymbol{\nu}) + P_r(\boldsymbol{\nu}) &= \frac{(\gamma_2+\gamma_3-\gamma_1)\sigma}{2\gamma_2\gamma_3}\mathcal{I}\{\gamma_2 \geq \gamma_1\} \\ &\quad + \frac{\sigma}{2\gamma_2}\mathcal{I}\{\gamma_2 < \gamma_1\} \end{aligned} \tag{44}$$

Therefore, the constant σ is determined by the following equation:

$$\begin{aligned} \frac{2\overline{P}}{\sigma} &= \underbrace{\mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{\gamma_2+\gamma_3-\gamma_1}{\gamma_2\gamma_3}\middle|\gamma_2 \geq \gamma_1\right]}_{A}\Pr\{\gamma_2 \geq \gamma_1\} \\ &\quad + \underbrace{\mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_2}\middle|\gamma_1 > \gamma_2\right]}_{B}\Pr\{\gamma_1 > \gamma_2\} \end{aligned} \tag{45}$$

where we can show, but omit the details, that

$$\begin{cases} A = \lambda_3\Gamma(0^+)\left[\frac{\lambda_2}{\lambda_1}\log\left(1+\frac{\lambda_1}{\lambda_2}\right)+\frac{\lambda_1}{\lambda_1+\lambda_2}\right] \\ B = \lambda_2\left[\Gamma(0^+)-\log\left(1+\frac{\lambda_1}{\lambda_2}\right)\right] \end{cases} \tag{46}$$

where $\Gamma(0^+) = \lim_{x\to 0^+}\Gamma(x) = +\infty$. Therefore, we have $A = B = +\infty$, which results in $\sigma = 0$ for any finite power constraint $\overline{P}$. The proof follows. ■

B. Protocol $(R1)$

Proof: As $\theta \to \infty$, the optimal resource allocation policy for protocol $(R1)$ becomes

$$\begin{cases} P_s(\boldsymbol{\nu}) = \frac{\sigma}{2\gamma_2}\mathcal{I}\{\gamma_2 \geq \gamma_1\} + \frac{\sigma}{2\gamma_1}\mathcal{I}\{\gamma_2 < \gamma_1\} \\ P_r(\boldsymbol{\nu}) = \tilde{u}P_s(\boldsymbol{\nu})\mathcal{I}\{\gamma_2 \geq \gamma_1\}. \end{cases} \tag{47}$$

Similarly to Appendix II-A, we prove that for protocol $(R1)$, σ in Eq. (47) is always equal to zero for any finite power constraint $\overline{P}$. Omitting the details, we can show that the constant σ is determined by the following equation:

$$\begin{aligned} \frac{2\overline{P}}{\sigma} &= \underbrace{\mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{\gamma_2+\gamma_3-\gamma_1}{\gamma_2\gamma_3}\middle|\gamma_2 \geq \gamma_1\right]}_{A}\Pr\{\gamma_2 \geq \gamma_1\} \\ &\quad + \underbrace{\mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 > \gamma_2\right]}_{B'}\Pr\{\gamma_1 > \gamma_2\} \end{aligned} \tag{48}$$

where $A = +\infty$ from Eq. (46) and

$$\begin{aligned} B' &= \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 \geq \gamma_2\right] \\ &= \int_0^\infty \left(\int_{\gamma_2}^\infty \frac{1}{\gamma_1}\lambda_1 e^{-\lambda_1\gamma_1}d\gamma_1\right)\lambda_2 e^{-\lambda_2\gamma_2}d\gamma_2 \\ &= \int_0^\infty \left[-\lambda_1 E_i(-\lambda_1\gamma_2)\right]\lambda_2 e^{-\lambda_2\gamma_2}d\gamma_2 \\ &\overset{(a)}{=} \lambda_1 \log\left(1+\frac{\lambda_2}{\lambda_1}\right) \end{aligned} \tag{49}$$

where $E_i(x)$ denotes the exponential integral function, and the equation of (a) holds due to the results given in [27, Sec. 6.224.1]. Again, we have $\sigma = 0$ for any finite power constraint $\overline{P}$. The proof follows. ■

C. Protocol $(R2)$ (Proof of Proposition 4)

Proof: As $\theta \to \infty$, the optimal resource allocation policy for protocol $(R2)$ becomes

$$\begin{cases} P_s(\boldsymbol{\nu}) = \dfrac{\sigma}{2\gamma_2}\mathcal{I}\{\gamma_2 \geq \gamma_1 \text{ and } \gamma_3 \geq \gamma_1\} \\ \qquad\qquad + \frac{\sigma}{2\gamma_1}\mathcal{I}\{\gamma_2 < \gamma_1 \text{ or } \gamma_3 < \gamma_1\} \\ P_r(\boldsymbol{\nu}) = \tilde{u}P_s(\boldsymbol{\nu})\mathcal{I}\{\gamma_2 \geq \gamma_1 \text{ and } \gamma_3 \geq \gamma_1\}. \end{cases} \tag{50}$$

To prove Proposition 4, it is equivalent to show that for protocol $(R2)$, σ in Eq. (50) is bounded away from zero for any finite power constraint $\overline{P}$. Likewise, the following equations hold for the constant σ:

$$\begin{aligned} \frac{2\overline{P}}{\sigma} &= \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{\gamma_2+\gamma_3-\gamma_1}{\gamma_2\gamma_3}\middle|\gamma_2 \geq \gamma_1 \text{ and } \gamma_3 \geq \gamma_1\right] \\ &\quad \cdot \Pr\{\gamma_2 \geq \gamma_1 \text{ and } \gamma_3 \geq \gamma_1\} \\ &\quad + \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 > \gamma_2 \text{ or } \gamma_1 > \gamma_3\right] \\ &\quad \cdot \Pr\{\gamma_1 > \gamma_2 \text{ or } \gamma_1 > \gamma_3\} \\ &< \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{\gamma_2+\gamma_3-\gamma_1}{\gamma_2\gamma_3}\middle|\gamma_2 \geq \gamma_1 \text{ and } \gamma_3 \geq \gamma_1\right] \\ &\quad + \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 > \gamma_2 \text{ or } \gamma_1 > \gamma_3\right] \\ &< \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_2}+\frac{1}{\gamma_3}\middle|\gamma_2 \geq \gamma_1 \text{ and } \gamma_3 \geq \gamma_1\right] \\ &\quad + \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 > \gamma_2 \text{ or } \gamma_1 > \gamma_3\right] \\ &< \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_2}\middle|\gamma_2 \geq \gamma_1\right] + \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_3}\middle|\gamma_3 \geq \gamma_1\right] \\ &\quad + \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 > \gamma_2\right] + \mathbb{E}_{\boldsymbol{\gamma}}\left[\frac{1}{\gamma_1}\middle|\gamma_1 > \gamma_3\right] \\ &= \lambda_2 \log\left(1+\frac{\lambda_1}{\lambda_2}\right) + \lambda_3 \log\left(1+\frac{\lambda_1}{\lambda_3}\right) \\ &\quad + \lambda_1 \log\left(\left(1+\frac{\lambda_2}{\lambda_1}\right)\left(1+\frac{\lambda_3}{\lambda_1}\right)\right). \end{aligned} \tag{51}$$

Therefore, σ is bounded away from zero for any finite power constraint $\overline{P}$. The proof follows. ■

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[2] A. J. Goldsmith and S. Chua, "Vairable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[3] V. Kawadia and P. R. Kumar, "A cautionary perspective on cross-layer design," *IEEE Wireless Commun.*, pp. 3–11, Feb. 2005.

[4] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Commun.*, pp. 59–65, Aug. 2005.

[5] T. M. Cover and A. A. El Gamal, "Capacity theorem for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572–584, Sept. 1979.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, Part I: System description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003.

[7] ——, "User cooperation diversity, Part II: Implementation aspectes and performance analysis," *IEEE Trans. Commun.*, vol. 51, pp. 1939–1948, Nov. 2003.

[8] N. Ahmed, M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Ourage minimization with limited feedback for the fading relay channel," *IEEE Trans. Commun.*, vol. 54, no. 3, March 2006.

[9] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficienct protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.

[10] D. Gunduz and E. Erkip, "Opportunistic cooperation by dynamic resource allocation," submitted for publication. [on-line]: http://eeweb.poly.edu/∼elza/Publications/opp05.pdf.

[11] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[12] M. O. Hasna and M.-S. Alouini, "Optimal power allocation for relayed transmissions over Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1999–2004, Nov 2004.

[13] M. O. Hasna and M.-S. Alouini, "Performance analysis of two-hop relayed transmissions over Rayleigh fading channels," in *Proc. IEEE VTC*, 2002.

[14] S. Cui and A. Goldsmith, "Cross-layer design in energy-constrained networks using cooperative MIMO techniques," *EURASIP J. Applied Signal Processing*, to appear in 2006.

[15] D. Wu, "Providing quality-of-service guarantees in wireless networks," Ph.D. Dissertation, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, August, 2003.

[16] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, Jul. 2003.

[17] D. Wu and R. Negi, "Downlink scheduling in a cellular network for quality-of-service assurance," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1547–1557, Sep. 2004.

[18] D. Wu and R. Negi, "Utilizing multiuser diversity for efficient support of quality of service over a fading channel," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 1198–1206, May 2005.

[19] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Automat. Contr.*, vol. 39, no. 5, pp. 913–931, May 1994.

[20] F. Kelly, S. Zachary, and I. Ziedins, "Stochastic Networks: Theory and Applications", *Royal Statistical Society Lecture Notes Series*, vol. 4, Oxford University Press, pp. 141–168, 1996.

[21] J. G. Kim and M. Krunz, "Bandwidth allocation in wireless networks with gauranteed packet-loss performance," *IEEE/ACM Trans. Networking*, vol. 8, pp. 337–349, June 2000.

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

[23] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 329–343, June 1993.

[24] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *IEEE Trans. Wireless Commun.*, to appear.

[25] S. Falahati, A. Svensson, T. Ekman, and M. Sternad, "Adaptive modulation systems for predicted wireless channels," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 307–316, Feb. 2004.

[26] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[27] I.S. Gradshteyn and I.M. Ryzhik, *Table of integral, series, and products*, Academic Press, 1992.

[28] `http://functions.wolfram.com/06.34.06.0004.01`

Jia Tang (S'03) received the B.S. degree in Electrical Engineering from Xi'an Jiaotong University, Xi'an, China, in 2001. He is currently a research assistant working toward the Ph.D. degree in Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA.

His research interests include mobile wireless communications and networks, with emphasis on cross-layer design and optimizations, wireless quality-of-service (QoS) provisioning for mobile multimedia networks, wireless diversity techniques, and wireless resource allocation.

Mr. Tang received Fouraker Graduate Research Fellowship Award from Department of Electrical and Computer Engineering, Texas A&M University in 2005.

Xi Zhang (S'89-SM'98) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, the M.S. degree from Lehigh University, Bethlehem, PA, all in electrical engineering and computer science, and the Ph.D. degree in electrical engineering and computer science (Electrical Engineering — Systems) from The University of Michigan, Ann Arbor.

He is currently an Assistant Professor and the Founding Director of the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station. He was an Assistant Professor and the Founding Director of the Division of Computer Systems Engineering, Department of Electrical Engineering and Computer Science, Beijing Information Technology Engineering Institute, Beijing, China, from 1984 to 1989. He was a Research Fellow with the School of Electrical Engineering, University of Technology, Sydney, Australia, and the Department of Electrical and Computer Engineering, James Cook University, Queensland, Australia, under a Fellowship from the Chinese National Commission of Education. He worked as a Summer Intern with the Networks and Distributed Systems Research Department, AT&T Bell Laboratories, Murray Hills, NJ, and with AT&T Laboratories Research, Florham Park, NJ, in 1997. He has published more than 100 research papers in the areas of wireless networks/communications, mobile computing, cross-layer optimizations for QoS guarantees over mobile wireless networks, effective capacity and effective bandwidth theories for wireless networks, DS-CDMA, MIMO-OFDM and space-time coding, adaptive modulations and coding (AMC), wireless diversity techniques and resource allocations, wireless sensor and Ad Hoc networks, cognitive radio and cooperative communications/relay networks, vehicular Ad Hoc networks, multi-channel MAC protocols, wireless and wired network security, wireless and wired multicast networks, network protocols design and modeling, statistical communications theory, information theory, random signal processing, and control theory and systems.

Prof. Zhang received the U.S. National Science Foundation CAREER Award in 2004 for his research in the areas of mobile wireless and multicast networking and systems. He also received the TEES Select Young Faculty Award for Excellence in Research Performance from the Dwight Look College of Engineering at Texas A&M University, College Station, in 2006. He is currently serving as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, an Associate Editor for the IEEE COMMUNICATIONS LETTERS, and an Editor for the WILEY'S JOURNAL ON WIRELESS COMMUNICATIONS AND MOBILE COMPUTING, and is also serving as the Guest Editor for the *IEEE Wireless Communications Magazine* for the Special Issues on "Next Generation of CDMA versus OFDMA for 4G Wireless Applications". He has frequently served as the Panelist on the U.S. National Science Foundation Research-Proposal Review Panels. He is serving or has served as the Co-Chair for the IEEE ICC 2008 — Information and Network Security Symposium and the Co-Chair for the IEEE Globecom 2008 — Wireless Communications Symposium, respectively, the Symposium Chair for the IEEE International Cross-Layer Optimized Wireless Networks Symposium 2006 and 2007, respectively, the TPC Chair for the IEEE IWCMC 2006 and 2007, respectively, the Student Travel Grants Co-Chair for the IEEE INFOCOM 2007, the Panel Co-Chair for the IEEE ICCCN 2007, the Poster Chair for the IEEE QShine 2006 and the IEEE/ACM MSWiM 2007, and the Publicity Chair for the IEEE WirelessCom 2005 and QShine 2007, and the Panelist on WiFi-Hotspots/WLAN QoS Panel at the IEEE QShine 2004. He has served as the Executive-Committee/Technical Program Committee Members for more than 40 IEEE/ACM conferences, including the IEEE INFOCOM, IEEE Globecom, IEEE ICC, IEEE WCNC, IEEE VTC, IEEE/ACM QShine, IEEE WoWMoM, IEEE ICCCN, etc.

Prof. Zhang is a Senior Member of the IEEE and a Member of the Association for Computing Machinery (ACM).