Amplify-and-Forward Cooperative Diversity Wireless Networks: Model, Analysis, and Monotonicity Properties

Teerawat Issariyakul, *Member, IEEE*, and Vikram Krishnamurthy, *Fellow, IEEE*

***Abstract*—This paper models and analyzes the performance of an amplify-and-forward cooperative diversity wireless network. We propose a Markov-based model, which encompasses the following aspects: 1) the transmission using amplify-and-forward cooperative diversity at the physical layer; 2) a flow control protocol, finite and infinite transmitting buffers, and an ARQ-based error recovery mechanism at the radio link layer; and 3) a bursty traffic pattern at the application layer. We derive expressions for packet delivery probability and distribution of packet delivery delay. We numerically quantify improvement in terms of packet delivery probability and packet delivery delay for increasing SNR and/or cooperative nodes. For an additional cooperative node, we quantify the amount of SNR which can be reduced (i.e., SNR saving) without degrading the system performance. Also, the minimum SNR and cooperative nodes which satisfy a probabilistic delay bound are computed. We then derive a sufficient condition that ensures an increase in packet delivery probability. Unlike numerical evaluation of the model, this sufficient condition does not require computation of stationary distribution of the Markov chain. It only involves parameter adjustment at physical, radio link, and application layers, hence substantially reducing the computation effort. Based on the developed model, we design a power allocation algorithm, which computes the minimum transmission power under a packet delivery probability constraint. We then use the derived sufficient condition to reduce complexity of the power allocation algorithm.**

***Index Terms*—Amplify-and-forward (AF), cooperative diversity (CD), Markov chain, monotonicity, stochastic dominance.**

I. Introduction

In general, any type of *diversity* (e.g., space, time, or frequency) is beneficial in enhancing data transmission performance of a wireless network. Space diversity employs an array of spatially separated antennae to exploit location dependent fading characteristics in a wireless channel. This concept has been utilized in multi-input multi-output (MIMO) [1] systems. However, as the size of a mobile terminal becomes smaller, it is more difficult to design antennae with large spatial separation. Closer antennae lead to increased correlated fading characteristics, and therefore, may not provide significant performance improvement. *Cooperative diversity* [2] is a promising technique to overcome this limitation.

Cooperative diversity is a form of space-time diversity, which makes use of multiple antennae to improve received signal quality. Unlike MIMO systems, cooperative diversity relies on data transmission by several nodes. Each node acts as a *virtual antenna* and cooperatively transmits data to a particular destination. Since each node tends to be at different places, cooperative diversity benefits from the tendency to find multiple antennae with independent fading.

Cooperative diversity can be classified into two main categories [3]. One is *amplify-and-forward* cooperative diversity (CD-AF). With this type of cooperative diversity, each *cooperative* node simply amplifies and forwards the received signal towards the destination node. Another is *decode-and-forward* cooperative diversity (CD-DF). Each cooperative node in this type decodes and re-encodes the received signal before forwarding the signal to the destination. While the CD-DF helps avoid error propagation, the CD-AF maintains its simplicity and cost-effectiveness. Furthermore, the CD-AF is transparent under adaptive modulation, and is able to maintain the maximum diversity order as the number of cooperative nodes increases. This paper focuses on the CD-AF type.

This paper models and analyzes the *cross-layer* performance of an AF (*amplify-and-forward*) system (Fig. 1). The model encompasses the following three layers. First, the application layer models traffic generated by a mobile user. Second, the radio link layer stores packets received from the application layer in a transmitting buffer, and transmits each of them in a first-in-first-out (FIFO) manner. We insert a link-layer flow control protocol to block some packets from the application layer, preventing the transmitting buffer from being overloaded. The radio link layer is also responsible for retransmitting unsuccessfully transmitted packets using an automatic repeat request (ARQ) protocol. Finally, the physical layer models the symbol error probability of data transmission between a source node (S) and a destination node (D) with the aid of a cooperative node (R). In addition to its own parameters (i.e., flow control parameters, buffer size, and ARQ persistency), the radio link layer takes traffic parameters from the application layer and symbol error probability from the physical layer, and models the following performance measures at the radio link layer: packet dropping probability, packet blocking probability, packet delivery probability, and packet delivery delay.

In practice, a CD-AF scheme can be used to improve link reliability and energy efficiency for various applications such as

Manuscript received June 07, 2006; revised April 11, 2007 and October 31, 2007; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor R. Mazumdar. First published June 20, 2008; current version published February 19, 2009.

T. Issariyakul is with TOT Public Company Limited, Bangkok, Thailand 10250 (e-mail: teerawat@ece.ubc.ca; iteerawat@hotmail.com)

V. Krishnamurthy is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4 (e-mail: vikramk@ece.ubc.ca).

Digital Object Identifier 10.1109/TNET.2008.925090

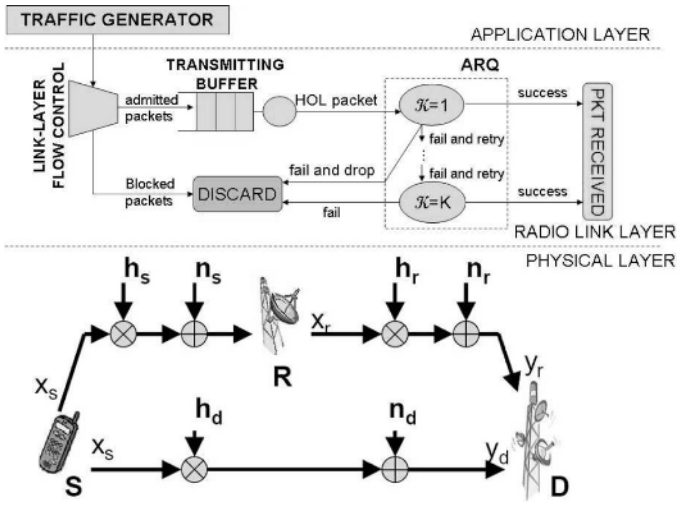

Fig. 1. Diagram of an AF system which consists of: 1) a traffic generator from the application layer; 2) link-layer flow control and ARQ protocols as well as a transmitting buffer at the radio link layer; and 3) transmission using a CD-AF scheme at the physical layer.

wireless sensor networks [4] or multi-hop cellular wireless networks (MCNs) [5]. A wireless sensor network deploys sensor nodes to collect information and help each other deliver the collected information to a processing center. Since sensor nodes tend to be at different location, a CD-AF scheme can be naturally applied to a wireless sensor network. In this case, a sensor node can act as a source node or a cooperative node, while the data processing center can act as a destination node.

An MCN (multi-hop cellular wireless network) employs relay nodes to increase the service coverage, transmission rate, link reliability, and/or energy efficiency. In a fixed MCN, a service provider installs relay nodes strategically to avoid line-of-sight obstruction (e.g., from buildings or trees). Unlike a base station, these relay (i.e., cooperative) nodes only amplify and forward the received signal from the air interface, and need no data cable. Therefore, the cost for implementing a fixed MCN is relatively inexpensive. Since each relay node is fixed, it may be equipped with power supply to alleviate energy constraints. When serving a customer, the service provider may simply lookup the predefined route computed during the network planning phase to setup a connection between a mobile and a base station. In this paper, we model and analyze the performance of a particular route consisting of few relay (i.e., cooperative) nodes in an AF system.

A. Main Results

1) Markov-Based Radio Link Layer Performance Model: This paper presents a Markov-based performance model, and derives two main radio link layer performance measures: packet delivery probability and packet delivery delay (see the definitions in Section V). The model takes into account physical layer parameters (e.g., signal-to-noise ratio (SNR) or the number of cooperative nodes) of a transmission using CD-AF, and captures dynamic burstiness of incoming traffic, link-layer flow control mechanism, the transmission counter of an underlying ARQ protocol, and the buffer occupancy.

We derive expressions for packet delivery probability and probability mass function (*pmf*) of packet delivery delay. The derived *pmf* directly translates into a probabilistic delay bound (i.e., a packet is delivered before t seconds with probability p) and could be useful in determining the amount of quality of service (QoS) which can be provided to customers. By numerical evaluation, we quantify an increase in packet delivery probability and/or a decrease in packet delivery delay, for increasing SNR and/or the number of cooperative nodes. For an additional cooperative node, we quantify the amount of SNR which can be reduced (i.e., SNR saving) without degrading the system performance. For a certain probabilistic delay bound, we determine the minimum required number of cooperative nodes (or minimum required SNR) under a constraint on power budget (or cooperative nodes).

2) Monotonicity of Packet Delivery Probability: For a class of link-layer flow control protocols, we derive a sufficient condition (**A3** in Result 2) which ensures an increase in packet delivery probability. This condition includes adjustment of physical, ARQ, and traffic parameters. The packet delivery probability is a function of blocking probability [p_b in (14)], which is a function of stationary probabilities of a Markov chain representing an AF system. Due to the lack of a closed-form expression of the blocking probability, it is extremely difficult to derive a necessary condition which ensures an increase in packet delivery probability. In this paper, we derive the above sufficient condition using the concept of *stochastic dominance* of Markov chains. This monotonicity result provides an interesting insight into the structural behavior of an AF system. Based on this result, we can adjust system parameters (e.g., SNR, ARQ persistency, or traffic load) to increase the packet delivery probability, without having to compute stationary probabilities of the Markov chain.

3) Power Allocation Algorithm: As a direct application of the first two contributions, we design two power allocation algorithms. The objective is to compute the minimum transmission power which satisfies a constraint on packet delivery probability. The first algorithm, a direct power computation algorithm, iteratively guesses the value of transmission power and computes the corresponding packet delivery probability using (6). The second algorithm, a power update algorithm, aims at reducing the complexity of the first algorithm. Reacting to a change in system parameters (e.g., number of cooperative nodes), the power update algorithm employs Result 2 to update transmission power and keep packet delivery probability unchanged. Since the power update algorithm does not involve in computation of stationary distribution of the Markov chain, its complexity reduces from $O(n^3)$ to $O(n)$.[1]

B. Related Works

Analytical results for CD-AF in the literature mainly focus on physical layer performance, but not *cross-layer* performance. For example, [6] derives the outage probability (i.e., the probability that the received SNR falls below a certain threshold) in AF and DF cooperative diversity systems. For a two-hop AF system, the symbol error probability at the destination node is derived in [7] under a Rayleigh fading channel. A relatively accurate approximation for symbol error probability

[1]In the literature, protocol complexity is measured in terms of $O(f(n))$ where $f(n)$ is a function of n, and n is the size of the problem. For example, in case of a matrix inversion problem, n is the matrix dimension.

TABLE I
LIST OF KEY NOTATIONS

Notation	Meaning	Defined in Section
$\overline{\gamma}_d$	Average SNR of the direct signal	III
$\overline{\gamma}_{ij}$	Average SNR received by j^{th} cooperative node on i^{th} cooperative branch	III
R	Number of cooperative branches	III
H_i	Number of nodes on the i^{th} cooperative branch	III
ϵ_s	Symbol error probability	III
ϵ_p	Packet error probability	II
B	Buffer size	IV
L	Number of symbols per packet	II
$\mathbf{B}$	D-BMAP arrival state transition probability matrix	IV-A
M	Number of D-BMAP arrival states	IV-A
$\mathbf{A}_n, \mathbf{A}_n^+$	D-BMAP arrival matrices	IV-A
N	Maximum batch size	IV-A
$\overline{\lambda}$	Average arrival rate	IV-A
ξ_x	Flow control admission probability when the system state is x	IV-B
ν_x	Flow control marking probability when the system state is x	IV-B
ψ_i	ARQ retransmission probability when the transmission counter is i	IV-C
K	ARQ retry limit	IV-C
$\mathbf{P}, \mathbf{P}_{sm}$	Transition probability matrices of the DTMC representing an AF system	V
$\boldsymbol{\pi}, \boldsymbol{\pi}_{sm}$	Stationary probability vectors of the DTMC representing an AF system	V
p_s	Packet delivery probability	V
p_d	Packet dropping probability	V
p_b	Packet blocking probability	V
$f_{\mathscr{D}}(d)$ and $E[\mathscr{D}]$	Probability mass function and expectation of packet delivery delay	V
$\Theta \triangleq \{\overline{\gamma}_d, \overline{\gamma}_{ij}, R, \sim H_i\}$	Physical layer parameters	VII
$\Psi \triangleq \{\psi_i; K\}$	ARQ persistency	VII
$\Xi \triangleq \{B; \xi_x\}$	Buffer management parameters	VII
$\mathbf{a}, \mathbf{a}^+$	Packet arrival row vectors	V-C,VII-B

in a multi-hop multi-branch AF system is presented in [8]. However, the statistics of the radio link layer performance measures are not analyzed in these works.

A link-level flow control protocols is responsible for deciding whether to block or admit packets from the application to the transmitting buffer. Examples of link-layer flow control protocols include a *droptail* queue, which blocks incoming packets only when the buffer is full, and a *Random Early Detection* (RED) queue [9], which probabilistically blocks (i.e., marks[2]) incoming packets when the average buffer occupancy becomes too large. A RED queue provides better interaction to higher layer protocols such as Transmission Control Protocol (TCP) than a droptail does. When multiple consecutive packets are discarded (e.g., when the buffer of a droptail queue is full), TCP waits for timeout and resets its transmission window to the initial value. By blocking packets earlier, a RED queue implicitly notifies TCP to slow down its transmission rate (i.e., decrease its transmission window), hence avoiding timeouts. Although a flow control protocol helps prevent congestion, it may lead to unnecessary packet blocking. Optimizing a flow control protocol requires a more precise model of the traffic source which incorporates many technical details (e.g., window adjustment, timeout mechanism), and is considered as our future study. In this paper, we model and analyze an AF system under a given link layer flow control protocol.

To quantify the radio link layer performance, several papers propose mathematical models and evaluate the performance numerically. For example, [10] studies the performance of video transmission over a wireless local area network (WLAN), and proposes an algorithm to change ARQ persistency adaptively. Taking into account a finite transmitting buffer and ARQ persistency, [11] minimizes packet loss rate by adjusting the modulation index adaptively. Both of these works (and many other similar works) propose a mathematical model for the system under consideration, and obtain the results by numerical evaluation. Despite their significant contributions, we argue that numerically solving an optimization problem does not provide much insight. In this paper, we *analytically* examine the trend of packet delivery probability with respect to system parameters (e.g., SNR, ARQ persistency, traffic load).

Notation: Regular and boldface letters represent scalar values and matrices, respectively. We denote an all-one column vector, a zero matrix, and an identity matrix with size n by $\mathbf{1}_n$, $\mathbf{0}_n$, and $\mathbf{I}_n$, respectively. The subscript is dropped when the dimension is explicit from the context. $[\mathbf{X}]_{i,j}$ denotes the (i,j) element of the matrix $\mathbf{X}$. $(x_n)_{n=a}^b$ is $(x_a, \ldots, x_b)$, where (x_a, x_b) formulates a row vector (or matrix) by concatenating x_b to the right of x_a. A diagonal matrix whose diagonal entries is $\{a_1, a_2, \ldots, a_N\}$ is denoted by $diag\{a_1, a_2, \ldots, a_N\}$. A Kronecker product is denoted by $\otimes$. The list of key notations is given in Table I.

The organization of this paper is as follows. Section II presents an overall view of the AF system under consideration. Section III gives an overview on an CD-AF scheme. Section IV discusses the models for an arrival process, a link-layer flow control protocol, and a service process. Section V formulates three following queueing models: a simplified model, a generalized model, and a semi-Markov model. Section VI shows numerical and simulation results for the model developed in Section V. Section VII investigates the model developed in Section V and derives a sufficient condition for a monotone increase in packet delivery probability. As an example application of the Markov-based model in Section V and the structural result in Section VII, we design two power allocation algorithms which compute the minimum transmission power satisfying a constraint on packet delivery probability in Section VIII. The paper summary and conclusion are stated in Section IX. Finally, relevant proofs are given in Appendices I–IV.

[2]In [9], "mark" has the same implication as "discard".

II. Description of an AF System Under Consideration

This paper focuses on a time-division multiple access (TDMA)-based AF system. In each time slot, only one active mobile (e.g., the source node S in Fig. 1) is allowed to transmit a packet, and we are interested in the link-layer performance of this mobile only. We assume that each packet consists of L symbols, and the mobile transmits each symbol using a CD-AF scheme to improve link reliability and energy efficiency. Based on the model proposed in [8], each symbol is subject to error probability ϵ_s in (2). Without error correction codes, the error probability for a packet transmission ϵ_p is given by

$$\epsilon_p = 1 - (1 - \epsilon_s)^L. \tag{1}$$

After a packet transmission, the destination node (e.g., the destination node D in Fig. 1) determines whether the packet is successfully received. If not, the ARQ module at the source node (defined in Section IV-C) will decide whether to retransmit or drop[3] the lost packet. After the head-of-line packet is successfully transmitted or dropped by the ARQ, the next packet is fetched to the head of the buffer, and the packet transmission process repeats.

III. Physical Layer Model: Amplify-and-Forward Cooperative Diversity

We adopt an CD-AF scheme from [8] to transmit each symbol from Node S (a source node) to Node D (a destination node) in Fig. 1. The transmission begins when Node S transmits data to Node D over a broadcast channel. Node R (a cooperative node) intercepts the transmitted signal, amplifies the intercepted signal, and forwards the amplified signal to Node D over another orthogonal channel. Node D waits until it receives signals from Node S and Node R, and estimates the transmitted signal by using a maximal ratio combiner (MRC).[4]

We assume that the source node uses m-PSK (m-ary phase shift keying) with fixed modulation index m. Data transmission takes place over orthogonal flat-fading complex channels with the additive white Gaussian noise (AWGN). Communication channels between S-D, S-R, and R-D are characterized by fading coefficients h_d, h_s, and h_r, and by AWGN terms n_d, n_s, and n_r, with variance N_0, respectively. In accordance with [8], we assume that all fading coefficients are available via training at the receiver. Using an MRC, the SNR at the destination is $\gamma_z = \gamma_d + \gamma_r$, where $\gamma_d = |h_d|^2 \frac{X_s}{N_0}$ is the SNR at the destination in case of direct transmission, $\gamma_r = |Ah_sh_r|^2 \frac{X_s}{N_0(1+|Ah_r|^2)}$ is the SNR at the destination when the signal is received from the cooperative node, X_s is the transmission power of the source node, and $A = \sqrt{\frac{X_s}{X_s|h_s|^2+N_0}}$ is the amplifying gain of the cooperative node[5].

In general, there can be R cooperative nodes, which amplify the signal intercepted from the same source node and forward the amplified signal to the same destination node. We refer to a signal path with cooperative node $i = \{1, \ldots, R\}$ as a "cooperative branch i." Also, the signal transmitted on branch i can be relayed (in a multi-hop manner) among H_i nodes before reaching the destination. Here, we define H_i as the "number of cooperative nodes on the cooperative branch i".

Given the SNR (γ_z) of the signal combined from all cooperative branches, the symbol error probability (ϵ_s) can be calculated as follows (see also [8]):

$$\begin{aligned} \epsilon_s &= Q(\sqrt{C_m\gamma_z}) \\ &\approx \frac{C(R)}{(C_m)^{R+1}} \cdot f_{\gamma_d}(0) \prod_{i=1}^{R} \left(\sum_{j=0}^{H_i} f_{\gamma_{ij}}(0) \right) \\ C(R) &= \prod_{i=1}^{R+1} \frac{2i-1}{(R+1)! \cdot 2i^{m+1}}. \end{aligned} \tag{2}$$

Here, $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-u^2/2} du$, C_m is a constant depending on the modulation type (e.g., $C_m = 2$ for BPSK), $f_{\gamma_{ij}}(\gamma)$ is the probability density function of γ_{ij}, γ_d is a random variable representing the SNR of the direct signal received at the destination, and γ_{ij} is a random variable representing the SNR of the signal received at the jth cooperative node on the ith cooperative branch.

Similar to [8], we assume that the CD-AF topology is known prior to data transmission. Also, since the source node and the cooperative nodes transmit on orthogonal channels, their do not interfere each other. Therefore, we do not consider the problems of routing and scheduling in this paper.

IV. Arrival Process, Link-Layer Flow Control, and Service Process

Here, we discuss the three main components necessary to formulate a radio link-layer model. First, the traffic generator generates packets according to the arrival process defined in Section IV-A. Second, a link-layer flow control protocol, defined in Section IV-B, decides whether to block or to admit a packet to the transmitting buffer. Served in an FIFO manner, each admitted packets is transmitted using a CD-AF scheme defined in Section III and is subject to the packet error probability ϵ_p defined in (1). Finally, the service process is characterized by the packet error probability ϵ_p as well as the ARQ protocol defined in Section IV-C.

A. Arrival Process: Discrete Batch Markovian Arrival Process (D-BMAP)

Due to its generality, a D-BMAP is used widely to model traffic arrival processes[6] [13]. A D-BMAP comprising of M arrival states is characterized by a transition probability matrix $\mathbf{B}$ and arrival matrices $\mathbf{A}_n$, where batch size $n \in \{0, \ldots, N\}$ is the number of packets which arrive during one transition, and N is the maximum batch size. Each entry $[\mathbf{B}]_{ij}$ $(i, j \in \{1, \ldots, M\})$ represents the probability that the arrival process changes its state from i to j. Let $\mathbf{A}_n$ be $diag\{a_{1n}, a_{2n}, \ldots, a_{Mn}\}$, where a_{jn} $(j \in \{1, \ldots, M\}, n \in \{0, \ldots, N\})$ is the probability that

[3]There are two reasons for a packet to be *discarded*. A packet could be blocked by the link-layer flow control protocol. A packet admitted to the transmitting buffer can also be *dropped* by the ARQ module.

[4]An MRC is a diversity combiner whose output is a weighted sum of the input signals from each channel. It is used in a conventional wireless system to reduce the probability of finding a deep fade in a single channel [12].

[5]The amplifying gain A is chosen such that the source node and the cooperative node will use the same transmission power.

[6]As an example, an ON–OFF traffic model constantly generates packets in an ON state, and does not generate any packet in an OFF state. Correspondingly, $\mathbf{A}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{A}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. A transition between ON and OFF states is characterized by a transition probability matrix $\mathbf{B}$.

n packets arrive when the arrival state is j. Then, $\mathbf{BA}_n$ represents an arrival of n packets after a change in the arrival state. The stationary probability vector $(\mathbf{b})$ of the arrival process can be obtained by solving $\mathbf{bB} = \mathbf{b}$ and $\mathbf{b1} = 1$. Also, the average arrival rate $(\overline{\lambda})$ can be computed from

$$\overline{\lambda} = \sum_{n=1}^{N} n \cdot \mathbf{bA}_n \mathbf{1}. \tag{3}$$

When the arrival state is j, a_{jn} measures the fraction of *time* where n packets arrive within one transition (i.e., batch size is n). Another interesting variable is the fraction of *packets* corresponding to a batch size of n when the arrival state is j, $a_{jn}^{+} = na_{jn} / \sum_{k=1}^{N} ka_{kn}, n \in \{1, \ldots, N\}$. As we shall see, diagonal matrices $\mathbf{A}_n^{+} = diag\left\{a_{1n}^{+}, \ldots, a_{Mn}^{+}\right\}$ would be useful in calculating the blocking probability, since the sample space of a_{jn}^{+} is in terms of "packets" rather than "time."

B. Link-Layer Flow Control Protocol

The main function of a link-layer flow control protocol is to prevent the transmitting buffer from being overloaded. The dynamics of the transmitting buffer depend on the following three factors: 1) the instantaneous buffer occupancy (b); 2) the arrival state (a) which determines how much traffic will arrive at the transmitting buffer; and 3) the transmission counter (k) which indicates the time until the head-of-line packet leaves the transmitting buffer. We keep track of a system state x which is (a, b) when the buffer is empty and is (a, b, k) when the buffer is non-empty. Given a system state x, we assume that the link-layer flow control protocol admits or blocks an incoming packet with probabilities ξ_x or $\nu_x = 1 - \xi_x$, respectively, where ν_x is called the "marking probability" for a system state x.

Suppose a batch of n packets arrives when the system state is (a, b, k). The first admitted packet changes the system state to $(a, b+1, k)$ and the next packet will be admitted to the buffer with probability $\xi_{a,b+1,k}$. To capture the link-layer flow control dynamics, we define an admission probability matrix $\boldsymbol{\xi}$ for a finite buffer as

$$\boldsymbol{\xi} = \begin{pmatrix} \mathbf{I} - \boldsymbol{\xi}_0 & \boldsymbol{\xi}_0 & & \\ & \ddots & \ddots & \\ & & \mathbf{I} - \boldsymbol{\xi}_{B-1} & \boldsymbol{\xi}_{B-1} \\ & & & \mathbf{I} \end{pmatrix}. \tag{4}$$

Here, $\boldsymbol{\xi}_b = diag\{\xi_{1,b,K}, \ldots, \xi_{1,b,1}, \xi_{2,b,K}, \ldots, \xi_{M,b,1}\}$, $b = \{1, \ldots, B-1\}$, $\boldsymbol{\xi}_0 = diag\{\xi_{1,0}, \ldots, \xi_{M,0}\}$, M is the number of arrival states, B is the buffer size, and K is the ARQ retry limit. Block (i, j) of $\boldsymbol{\xi}$ represents the probability that the buffer occupancy changes from i to j due to an incoming packet. Similarly, block (i, j) of $\boldsymbol{\xi}^n$ represents the collective change for a batch of n packets. Note that $\boldsymbol{\xi}$ in (4) is defined for a finite transmitting buffer. For an infinite transmitting buffer, we remove the last (boundary) row and let each row repeat for a countably infinite number of times. In this case, $[\boldsymbol{\xi}]_{ij}$ is $\mathbf{I} - \boldsymbol{\xi}_{i-1}$ for $i = j$, is $\boldsymbol{\xi}_{i-1}$ for $j = i + 1$, and is $\mathbf{0}$ otherwise.

C. Service Process: CD-AF Transmission and ARQ Mechanism

The service time is the random interval during which the packet is at the head of the buffer. There are two main factors contributing to the service process: transmission using CD-AF and ARQ mechanism. We abstract the physical layer CD-AF operation by using the packet error probability ϵ_p in (1). We model ARQ mechanism using an absorbing discrete time Markov chain (DTMC). Finally, we characterize the service time as having phase type (PH) distribution.[7]

In this paper, we assume a stop-and-wait ARQ mechanism for each transmitted packet. In the rest of this section, we describe this protocol and its modeling in terms of a DTMC. At the source node, the ARQ module maintains a transmission counter, which is reset to one, if a packet is dropped or successfully transmitted. After each packet transmission, an acknowledgement is relayed via an immediate error-free out-of-band feedback channel to inform the source node of the transmission result. If the transmission fails, the ARQ will retransmit the packet with probability ψ_i and increase the transmission counter by 1, where $i = \{1, \ldots, K\}$ is the transmission counter. With probability $1 - \psi_i$, the ARQ drops the head-of-line packet and resets the transmission counter to 1. Without loss of generality, we assume $\psi_K = 0$, where K is the retry limit. After the counter is reset, a new packet, if available, is fetched to the head of the buffer, and the process repeats.

We formulate the above ARQ mechanism as a DTMC with one absorbing state. Consider the ARQ block in Fig. 1. Let the ARQ transmission counter, $\mathcal{K} \in \{K, \ldots, 1\}$, be transient states of an absorbing DTMC and combine the "DISCARD" and "PKT RECEIVED" blocks as an absorbing state. Then, the transition probability matrix $\mathbf{S}$ of the above absorbing DTMC is formulated in

$$\begin{aligned} \mathbf{S} &= \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \boldsymbol{\omega} & \boldsymbol{\Omega} \end{array}\right) \\ (\boldsymbol{\omega}|\boldsymbol{\Omega}) &= \left(\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline \omega_{K-1} & \Omega_{K-1} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \omega_1 & 0 & \cdots & \Omega_1 & 0 \end{array}\right) \\ \boldsymbol{\alpha} &= (\mathbf{0}, 1) \\ \omega_i &= 1 - \Omega_i \\ \Omega_i &= \epsilon_p \psi_i \\ i &= \{1, \ldots, K-1\} \end{aligned} \tag{5}$$

where ψ_i is the probability that the ARQ retransmits the packet when the transmission counter is i, given that the transmission fails, and $\boldsymbol{\alpha}$ is the initial probability row vector. The above DTMC starts when a new packet is fetched to the head of the buffer (i.e., $\mathcal{K} = 1$) and stops when the head-of-line packet leaves the buffer (i.e., either in "DISCARD" or "PKT RECEIVED" states). Its absorption time is equivalent to the service time defined earlier in this section.

V. Queueing Models and Radio Link-Layer Performance Measures

This section formulates a DTMC which captures dynamics of the arrival state $(\mathcal{A} \in \{1, \ldots, M\})$, the buffer occupancy $(\mathcal{B} \in \{0, \ldots, B\}$ for a finite buffer or $\mathcal{B} \in \{0, 1, \ldots\}$ for an infinite buffer), and the transmission counter $(\mathcal{K} \in \{K, \ldots, 1\})$. We

[7]PH distribution refers to absorption time distribution of an absorbing DTMC with one state.

define $\mathcal{X}$ as $(\mathcal{A}, \mathcal{B})$ when the buffer is empty, and as $(\mathcal{A}, \mathcal{B}, \mathcal{K})$ when the buffer is not empty. A transition of a DTMC $\mathcal{X}$ represents a change in a fixed *transmission interval*.[8] Consisting of multiple time slots, a *transmission interval* is an interval from the point where the source node transmits a packet to the point where it is allowed to transmit another packet (possibly the same packet). In each transmission interval, the following four events occur chronologically.

1) Arrival process: We assume that n packets arrive after a change in the arrival state with probability $\mathbf{BA}_n$.
2) Link-layer flow control: Each generated packet is admitted to the transmitting buffer with probability ξ_x where $x \in \mathcal{X}$.
3) Service process: The head-of-line packet stays in (for retransmission) or leaves (a transmission success or a packet drop) the transmitting buffer with transition probability matrices $\mathbf{\Omega}$ or $\boldsymbol{\omega}$, respectively.
4) Observation: The system state $\mathcal{X}$ is observed immediately after the service process.

We shall develop the models for three following cases. First, Section V-A models a droptail queue with a Bernoulli arrival process. Second, Section V-B generalizes the model in Section V-A for a D-BMAP arrival process with an arbitrary link layer flow control protocol. Finally, Section V-C modifies the generalized model in Section V-B to be a semi-Markov chain with an independent arrival process.

For each case, we shall formulate a transition probability matrix $(\mathbf{P})$ and determine its stationary probability vector $(\boldsymbol{\pi})$ by solving $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. Based on the derived stationary distribution, we shall compute the following two performance measures.

1) Packet Delivery Probability (p_s) is the probability that a packet is not blocked (with probability $1 - p_b$) from the transmitting buffer and is not dropped (with probability $1 - p_d$) by the ARQ module. It is the ratio of successfully delivered (not discarded) packets to overall packets generated by the application. Mathematically

$$p_s = (1 - p_b)(1 - p_d) \tag{6}$$

$$p_d = 1 - (1 - \epsilon_p) \sum_{i=1}^{K} r_i \tag{7}$$

$$r_i = \begin{cases} 1 & i = 1 \\ \prod_{j=1}^{i-1} \Omega_j, & i = \{2, \ldots, K\}. \end{cases} \tag{8}$$

Here, dropping probability (p_d) is the probability that a packet is dropped by the ARQ module, $\Omega_j = \epsilon_p \psi_j$ is the probability that the jth transmission is in error and the ARQ module decides to retransmit the lost packet, ψ_j is the retransmission probability when the transmission counter is j, and packet error probability ϵ_p is computed from (1). The derivation of blocking probability (p_b) for each of the above three cases will be discussed separately.

2) Packet Delivery Delay $(\mathcal{D})$ is the time interval between when a packet enters the transmitting buffer and when it is successfully received by the destination. If the packet is blocked, the corresponding delay will not be included.

A. Droptail Queue With a Bernoulli Arrival Process

This section formulates a model for a droptail queue assuming a Bernoulli arrival process. With a droptail queue, the packet admission probability $\xi_{a,b,k}$ is 0 for $b = B$ and is 1 otherwise, where B is the buffer size. For a Bernoulli arrival process, zero and one packet arrive in a transmission interval with probability $1 - \lambda$ and λ, respectively. Based on the arrival model in Section IV-A, $\mathbf{B} = 1$, $\mathbf{V}_0 = 1 - \lambda$, $\mathbf{V}_1 = \lambda$, and the transition probability matrix $\mathbf{P}$ is formulated as

$$\mathbf{P} = \begin{pmatrix} 1-\lambda & \lambda & & & & \\ (1-\lambda)\boldsymbol{\omega}\mathbf{1} & \mathbf{V}_1 & \mathbf{V}_0 & & & \\ & \mathbf{V}_2 & \mathbf{V}_1 & \mathbf{V}_0 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mathbf{V}_2 & \mathbf{V}_1 & \mathbf{V}_0 \\ & & & & \boldsymbol{\omega}\boldsymbol{\alpha} & \mathbf{\Omega} \end{pmatrix} \tag{9}$$

where $\mathbf{V}_2 = (1 - \lambda)\boldsymbol{\omega}\boldsymbol{\alpha}$, $\mathbf{V}_1 = (1 - \lambda)\mathbf{\Omega} + \lambda\boldsymbol{\omega}\boldsymbol{\alpha}$, and $\mathbf{V}_0 = \lambda\mathbf{\Omega}$. Again, the stationary probability vector $\boldsymbol{\pi}$ is computed by solving $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. Based on the computed $\boldsymbol{\pi}$, the blocking probability (p_b), the expected packet delivery delay $(E[\mathcal{D}])$, and the probability that the packet delivery delay is d $(f_{\mathcal{D}}(d))$ can be computed as follows:

$$\begin{aligned} p_b &= \sum_{x \in \mathcal{X}:\mathcal{B}=B} \pi_x \\ E[\mathcal{D}] &= \sum_{d=0}^{\infty} d \times f_{\mathcal{D}}(d) \\ f_{\mathcal{D}}(d) &= \frac{\sum_{x \in \mathcal{X}:\mathcal{B}<B} (\pi_x \times f_{\mathcal{D}|\mathcal{X}}(d|x))}{\sum_{x \in \mathcal{X}:\mathcal{B}<B} \pi_x} \end{aligned} \tag{10}$$

where π_x is the stationary probability that the DTMC is in state x and $f_{\mathcal{D}|\mathcal{X}}(d|x)$ is the probability that packet delivery delay is d conditioned on that the DTMC is in state x. The formulation of $f_{\mathcal{D}|\mathcal{X}}(d|x)$ is given in Appendix I.

B. Generalized Queueing Model

This section generalizes the DTMC defined in Section V-A for a D-BMAP arrival process, an arbitrary link-layer flow control protocol, and an ARQ mechanism defined in Sections IV-A, IV-B, and IV-C, respectively. We formulate the transition probability matrix $\mathbf{P}$ and compute p_b, $E[\mathcal{D}]$, and $f_{\mathcal{D}}(d)$ for two cases—a finite buffer and an infinite buffer—as follows.

1) Finite Transmitting Buffer ($\mathcal{B} = \{0, \ldots, b\}$): Consider arrival and link-layer flow control processes together. These two processes cause a change in the arrival state and an increase in buffer occupancy, which can be expressed by the transition probability matrix $\mathbf{C}$ in

$$\mathbf{C} = \sum_{n=0}^{N} [\mathbf{I}_B \otimes \mathbf{BA}_n \otimes \mathbf{I}_K] \boldsymbol{\xi}^n. \tag{11}$$

[8]Under a round robin scheduler, "fixed transmission interval" is a typical assumption in a TDMA system (e.g., cellular or IEEE 802.16 wireless systems [14]). With n active mobiles, a transmission interval is simply $n \times \tau$, where τ is the length of a time slot.

Here, $[\mathbf{I}_B \otimes \mathbf{BA}_n \otimes \mathbf{I}_K]$ contains probabilities of having no change in buffer occupancy $(\mathbf{I}_B)$, a change in the arrival state $(\mathbf{B})$, a generation of n packets $(\mathbf{A}_n)$, and no change in the transmission counter $(\mathbf{I}_K)$. These n packets are admitted to the transmitting buffer according to $\boldsymbol{\xi}^n$. Block $(b+1, b+n+1)$ of $\mathbf{C}$, $\mathbf{c}_n(b)$ contains probabilities that the buffer occupancy changes from b to $b+n$.

We now incorporate the service process $(\boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\omega})$ and construct the transition probability matrix $(\mathbf{P})$ for the DTMC $\mathcal{X}$ in (12), shown at the bottom of the page, where N is the maximum batch size. $\mathbf{U}_n = \mathbf{c}_n(0)[\mathbf{I}_M \otimes \boldsymbol{\alpha}]$, $\mathbf{c}_n(b) = [\mathbf{C}]_{b+1,b+n+1}$, $\mathbf{C}$ is defined in (11), $\mathbf{V}'_0(b) = \mathbf{c}_0(b)[\mathbf{I}_M \otimes \boldsymbol{\omega}]$, $\mathbf{V}_0(b) = \mathbf{c}_0(b)[\mathbf{I}_M \otimes \boldsymbol{\omega}\boldsymbol{\alpha}]$, $\mathbf{V}_{N+1}(b) = \mathbf{c}_N(b)[\mathbf{I}_M \otimes \boldsymbol{\Omega}]$, $\mathbf{V}_n(b) = \mathbf{c}_{n-1}(b)[\mathbf{I}_M \otimes \boldsymbol{\Omega}] + \mathbf{c}_n\boldsymbol{\omega}\boldsymbol{\alpha}$, $\mathbf{W}_n(b) = \sum_{i=n-1}^{N} \mathbf{c}_i(b)[\mathbf{I}_M \otimes \boldsymbol{\Omega}]$, and $\mathbf{W}'_n(b) = \sum_{i=n}^{N} \mathbf{c}_i(b)[\mathbf{I}_M \otimes \boldsymbol{\omega}\boldsymbol{\alpha}]$. Based on $\mathbf{P}$, we compute the stationary probability vector $\boldsymbol{\pi}$ by solving $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$.

2) Infinite Transmitting Buffer ($\mathcal{B} = \{0, 1, \ldots\}$): The construction of transition probability matrix in this case is very similar to that in case of finite buffer. We only need to remove the boundary blocks, and let each row repeat infinitely. In particular, we have

$$\boldsymbol{\xi} = \begin{pmatrix} \mathbf{I} - \boldsymbol{\xi}_0 & \boldsymbol{\xi}_0 & & \\ & \mathbf{I} - \boldsymbol{\xi}_1 & \boldsymbol{\xi}_1 & \\ & & \ddots & \ddots \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} \mathbf{U}_0 & \mathbf{U}_1 & \cdots & \mathbf{U}_N & & \\ \mathbf{V}'_0(1) & \mathbf{V}_1(1) & \cdots & \mathbf{V}_N(1) & \mathbf{V}_{N+1}(1) & \\ & \ddots & & & \ddots & \ddots \end{pmatrix}. \tag{13}$$

Unfortunately, (13) does not contain repeating rows. To compute stationary probability vector $\boldsymbol{\pi}$ corresponding to $\mathbf{P}$ in (13), we have to approximate the infinite buffer system with a finite buffer model [using (12)] with $B = B_{th}$, assuming that the probability that the buffer occupancy exceeds B_{th} is negligible. In a special case, where the admission probability does not depends on buffer occupancy (i.e., $\xi_{a,b,k} = k_{a,k}$, $\forall\, b$), each row i of $\mathbf{P}$ repeats itself for $i = \{3, 4, \ldots\}$. With this structure, we can use a *Matrix Geometric* method to compute $\boldsymbol{\pi}$. The detail of an algorithm to compute $\boldsymbol{\pi}$ as well as the stability condition are given in [15].

3) Computation of p_b, $E[\mathcal{D}]$, and $f_{\mathcal{D}}(d)$: After obtaining stationary probability vector $(\boldsymbol{\pi})$, we now compute the blocking probability (p_b), expected packet delivery delay $(E[\mathcal{D}])$, and probability that the packet delivery delay is d $(f_{\mathcal{D}}(d))$ as follows.

The blocking probability in case of an infinite buffer is clearly zero. For a finite buffer, it can be computed as follows:

$$p_b = \boldsymbol{\pi}[(\mathbf{p}_b(0), \mathbf{p}_b(1), \ldots)^\top]\boldsymbol{\nu} = \boldsymbol{\pi}\mathbf{p}_b\boldsymbol{\nu}, \text{ where} \tag{14}$$

$$\mathbf{p}_b = \frac{\sum_{n=1}^{N}\sum_{m=n}^{N} [\mathbf{I}_B \otimes \mathbf{BA}_m \otimes \mathbf{I}_K]\left(\boldsymbol{\xi}^{n-1}\right)}{\sum_{n=1}^{N}\sum_{m=n}^{N} [\mathbf{I}_B \otimes \mathbf{BA}_m \otimes \mathbf{I}_K]} = \sum_{n=1}^{N} [\mathbf{I}_B \otimes \mathbf{BA}_n^+ \otimes \mathbf{I}_K]\left(\frac{1}{n}\sum_{m=1}^{n} \boldsymbol{\xi}^{m-1}\right). \tag{15}$$

Here, $\boldsymbol{\nu} = ((\mathbf{1} - \boldsymbol{\xi}_0\mathbf{1})^\top, \ldots, (\mathbf{1} - \boldsymbol{\xi}_{B-1}\mathbf{1})^\top, \mathbf{1}^\top)^\top$ is the marking probability column vector. Again, for an infinite buffer, we remove the last row of $\boldsymbol{\nu}$ and let each row repeat for an infinite number of times. The derivation of p_b is given in Appendix II.

Next, define $f_{\mathcal{D}|\mathcal{X}}(d|x)$ as the conditional *pmf* of the packet delivery delay given that a packet enters the buffer when the system state is $x \in \mathcal{X}$, $[\mathbf{f}_{\mathcal{D}}]_{1,d} = f_{\mathcal{D}}(d)$, and $[\mathbf{f}_{\mathcal{D}|\mathcal{X}}]_{x,d} = f_{\mathcal{D}|\mathcal{X}}(d|x)$. Then

$$f_{\mathcal{D}}(d) = [\mathbf{f}_{\mathcal{D}}]_{1,d}, \quad \mathbf{f}_{\mathcal{D}} = \frac{\boldsymbol{\pi}\cdot\mathbf{D}\cdot\mathbf{f}_{\mathcal{D}|\mathcal{X}}}{\boldsymbol{\pi}\cdot\mathbf{D}\cdot\mathbf{1}} \tag{16}$$

$$[\mathbf{D}]_{ij} = \begin{cases} \sum_{n=j-i+1}^{N} \mathbf{c}_n(i+1), & j = \{i, \ldots, i+N\} \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{17}$$

$$E[\mathcal{D}] = \sum_{d=0}^{\infty} d \times f_{\mathcal{D}}(d). \tag{18}$$

The derivation of $\mathbf{f}_{\mathcal{D}}$ as well as $f_{\mathcal{D}|\mathcal{X}}(d|x)$ is given in Appendix I. Each row of $\mathbf{D}$ is the tail (i.e., *complementary cumulative distribution function*) of each row of $\mathbf{C}$ defined in (11). Again, for an infinite buffer, we remove the boundary block of $\mathbf{D}$ and let each row repeat infinitely.

C. Semi-Markov Chain Formulation and an Independent Arrival Process

To facilitate the analysis of blocking probability p_b in Section VII, we simplify the generalized model developed in Section V-B, by assuming an independent arrival process. Our

$$\mathbf{P} = \left(\begin{array}{cccccc} \mathbf{U}_0 & \mathbf{U}_1 & \ldots & \mathbf{U}_N & & \\ \mathbf{V}'_0(1) & \mathbf{V}_1(1) & \ldots & \mathbf{V}_N(1) & \mathbf{V}_{N+1}(1) & \\ \ddots & & & \ddots & \ddots & \\ & \mathbf{V}_0(B-N-1) & \ldots & \mathbf{V}_{N-}(B-N-1) & \mathbf{V}_N(B-N-1) & \mathbf{V}_{N+1}(B-N-1) \\ \hline & & \ldots & \mathbf{V}_{N-2}(B-N) & \mathbf{W}'_{N-1}(B-N) & \mathbf{W}_N(B-N) \\ & & \ddots & \vdots & \vdots & \vdots \\ & & & \mathbf{V}_0(B-1) & \mathbf{W}'_1(B-1) & \mathbf{W}_2(B-1) \\ & & & & \mathbf{W}'_0(B) & \mathbf{W}_1(B) \end{array}\right) \tag{12}$$

first objective is to reformulate the model using a semi-Markov chain for a finite transmitting buffer.[9] The second objective is to derive the blocking probability from the developed semi-Markov chain.

1) Formulation of a Transition Probability Matrix: We formulate a semi-Markov chain as well as its transition probability matrix as follows. First, since the arrival process is independent, we set the number of arrival states (M) to be 1. The arrival process in this case is represented by $\mathbf{a} = (a_0, \ldots, a_N)$, where $a_n, n \in \{0, \ldots, N\}$, is the probability of having n incoming packets in a transmission interval. Second, since $M = 1$, the matrices $\boldsymbol{\xi}_b$ and $\mathbf{I}$ defined in (4) become scalar values ξ_b and 1, respectively, where ξ_b is the admission probability when the buffer occupancy is b. Third, we keep track of the buffer occupancy only. Finally, we formulate a semi-Markov chain by observing the buffer occupancy as follows. When the buffer is empty, we observe the buffer occupancy at the end of every transmission interval. When the buffer is nonempty, we observe the buffer occupancy at the end of each transmission interval where the head-of-line packet leaves the buffer.

When the buffer is empty, the model is similar to that in Section V-B, since an epoch length of the semi-Markov chain is a single transmission interval. In each transmission interval, the buffer occupancy changes according to a transition probability matrix $\mathbf{C}_z = \sum_{i=0}^{N} a_i \boldsymbol{\xi}^i$ [simplified from $\mathbf{C}$ in (11)], which takes into account the arrival process and the link layer flow control protocol.

When the buffer is nonempty, the epoch length is defined as the time interval until the head-of-line packet leaves the transmitting buffer. Consider a service process defined in Section IV-C. The probability that an epoch lasts for j transmission intervals (t_j) is

$$t_j = \begin{cases} (1-\Omega_j) r_j, & j = \{1, \ldots, K-1\} \\ r_j, & j = K. \end{cases} \tag{19}$$

Here, $\Omega_j = \epsilon_p \psi_j$, K is the retry limit, and r_j is defined in (8). In an epoch with j transmission intervals (occurring with probability t_j), the buffer occupancy changes for j times, each with transition probability $\mathbf{C}_z$ defined above. In an entire epoch, the buffer occupancy changes according to $\mathbf{C}'_{nz}$ formulated by removing the first row of $\mathbf{C}_{nz} = \sum_{j=1}^{K} t_j \mathbf{C}_z^j$. The first row is removed since the semi-Markov chain starts when the buffer is not empty.

The transition probability matrix of the above semi-Markov chain $(\mathbf{P}_{sm})$ is formulated as

$$\mathbf{P}_{sm} = \begin{pmatrix} \text{the first row of } \mathbf{C}_z \\ \mathbf{C}'_{nz} \end{pmatrix}. \tag{20}$$

Again, the stationary probability vector $(\boldsymbol{\pi}_{sm})$ of the semi-Markov process is computed by solving $\boldsymbol{\pi}_{sm}\mathbf{P}_{sm} = \boldsymbol{\pi}_{sm}$ and $\boldsymbol{\pi}_{sm}\mathbf{1} = 1$.

2) Derivation of Blocking Probability: Similar to the transition probability matrix, the blocking probability is derived by considering empty-buffer and nonempty-buffer cases separately. When the buffer is empty, the blocking probability $p_b(0)$ in the following equation is a simplified version of that in (14)–(15):

$$p_b(0) = [\boldsymbol{\pi}_{sm}]_{1,1} \cdot \mathbf{p}_{b_z} \boldsymbol{\nu}, \qquad \mathbf{p}_{b_z} = \text{the first row of } \bar{\boldsymbol{\xi}}_z$$
$$\bar{\boldsymbol{\xi}}_z = \sum_{n=1}^{N} a_n^+ \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\xi}^{i-1} \right) \tag{21}$$

where $a_n^+ = n a_n / \sum_{i=1}^{N} i a_i$, and $\boldsymbol{\nu} = (1 - \xi_0, \ldots, 1 - \xi_{B-1}, 1)^\top$ is the marking probability vector.

For a nonempty buffer, the blocking probability is derived differently. Suppose that the epoch length is at least i transmission intervals. The blocking probability in this case is $\mathbf{C}_z^{i-1} \bar{\boldsymbol{\xi}}_z \boldsymbol{\nu}$. Also, this event occurs with probability $T_i \triangleq \frac{\sum_{j=i}^{K} t_j}{\sum_{m=1}^{K} \sum_{l=m}^{K} t_l}$ (i.e., for all events with epoch length not less than i). Therefore, the blocking probability for the buffer size $i = \{1, \ldots, B\}$ is

$$p_b(i) = [\boldsymbol{\pi}_{sm}]_{1,i+1} \cdot [\mathbf{p}_{b_{nz}} \boldsymbol{\nu}]_{i,1}, \quad i = \{1, \ldots, B\}$$

where

$$\mathbf{p}_{b_{nz}} = \bar{\boldsymbol{\xi}}_{nz} \text{ with the first row removed}$$
$$\bar{\boldsymbol{\xi}}_{nz} = \sum_{i=1}^{K} T_i \mathbf{C}_z^{i-1} \bar{\boldsymbol{\xi}}_z. \tag{22}$$

Combining (21) for the empty buffer case and (22) for the nonempty buffer case, we obtain the blocking probability (p_b) as

$$p_b = \boldsymbol{\pi}_{sm} \mathbf{p}_b \boldsymbol{\nu}, \quad \mathbf{p}_b = \begin{pmatrix} \mathbf{p}_{b_z} \\ \mathbf{p}_{b_{nz}} \end{pmatrix}. \tag{23}$$

TABLE II
DEFAULT PARAMETER SETTINGS

Parameters	Values
SNR of the direct signal (γ_d)	10 dB
Buffer size (B)	20 packets
Number of symbols per packet (L)	1080 bits
Number of cooperative branches (R)	2
Number of cooperative nodes in each branch (H)	1
Retry limit (K)	3
Packet arrival probability (λ)	0.9

VI. NUMERICAL RESULTS

We validate the model in Section V via Monte Carlo simulations. Each simulation is run for 5×10^4 transmission intervals. Unless otherwise specified, we use "lines" and "symbols" to plot the results obtained from the model and from the simulation, respectively. We limit our experiments only to a *Bernoulli* arrival process with packet arrival probability λ, a droptail queue, a limited-persistent ARQ protocol ($\psi_i = 1, i = \{1, \ldots, K-1\}$ and $\psi_K = 0$), and binary phase-shift keying (BPSK) modulation over a two-hop (as shown in Fig. 1) Rayleigh-fading AF system.[10] The mathematical model for this experiment is given in Section V-A. The default parameters are specified in Table II. For all the results, we shall see that the simulation results are very similar to those obtained from the model in Section V-A.

[9]The model for an infinite transmitting buffer can be derived by removing boundary blocks of the related matrices (e.g., $\boldsymbol{\xi}, \boldsymbol{\nu}$) and letting each row repeat infinitely.

[10]This physical-layer parameter setting is the same as in [8].

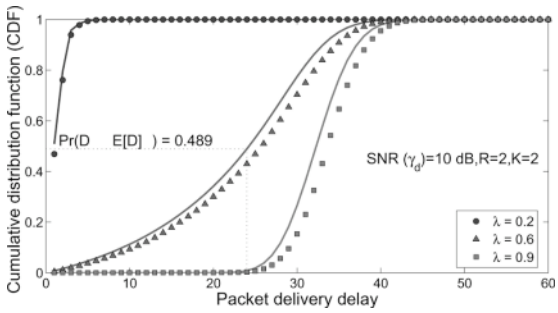

Fig. 2. CDF of packet delivery delay: the dotted line represents the probability that the delay would not be less than the expected delay.

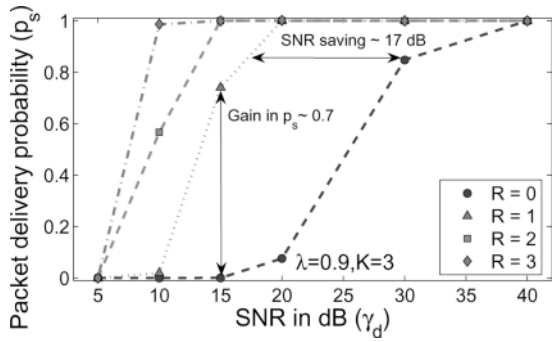

Fig. 3. Typical variation of the packet delivery probability (p_s) with respect to SNR: vertical and horizontal lines represents gain in p_s and SNR saving, respectively.

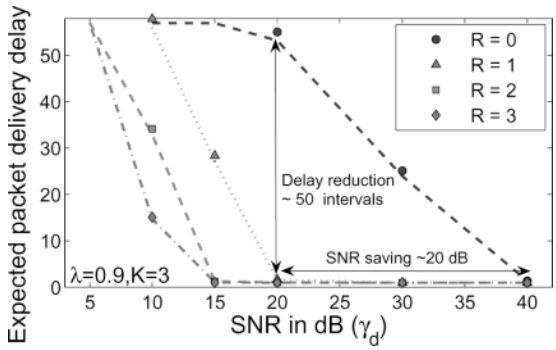

Fig. 4. Typical variation of the expected packet delivery delay with respect to SNR: vertical and horizontal lines represents delay reduction and SNR saving, respectively.

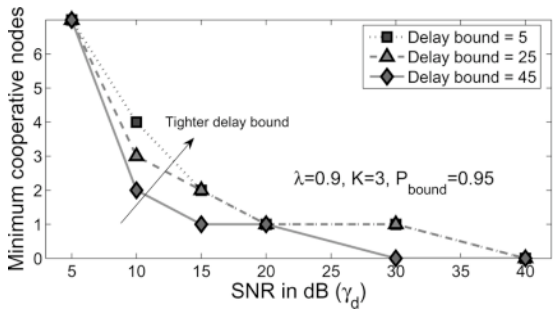

Fig. 5. The minimum required number of cooperative nodes under probabilistically bounded packet delivery delay.

1) Cumulative Distribution Function (cdf) of Packet Delivery Delay: Fig. 2 plots the cdf of the packet delivery delay for packet arrival probabilities $\lambda = \{0.2, 0.6, 0.9\}$. The dotted line shows the point on the cdf where the delay is $\lceil E[\mathcal{D}] \rceil$. We observe that the expectation is not an informative performance indicator, since almost 50% of delivered packets experienced delays greater than $E[\mathcal{D}]$.

2) Effects of SNR and the Number of Cooperative Nodes: Figs. 3 and 4 depict the packet delivery probability and the expected packet delivery delay as functions of SNR and the number of cooperative nodes, respectively. As discussed in Section III, the source node and the cooperative node transmit with the same power. The term "SNR" here refers to the signal transmitted from the source node and detected at the destination node. With a fixed noise level, increasing SNR and decreasing SNR correspond to an increase and a decrease (respectively) in transmission power of the source node and the cooperative node. With fixed transmission power, on the other hand, increasing SNR and decreasing SNR imply more and less (respectively) tolerance to noise fluctuation at the destination node.

Intuitively, increasing the SNR and/or the number of cooperative nodes increases the packet delivery probability and decreases the packet delivery delay. The horizontal distance between two lines in Figs. 3 and 4 represents *SNR saving*—the amount of SNR which can be reduced without degrading the packet delivery probability and/or the packet delivery delay, if the number of cooperative nodes is increased. Similarly, the vertical distance between two lines indicates the improvement in terms of the packet delivery probability and/or the packet delivery delay for an increasing number of cooperative nodes. Clearly, the model quantifies the improvement as a function of network resources and provides a basis for optimizing a cooperative diversity wireless system.

3) Minimum Required Number of Cooperative Nodes: Fig. 5 plots the minimum number of cooperative nodes, which guarantees that the packet delivery delay is less than a delay bound with probability $P_{bound} = 0.95$. For higher SNR, the minimum required number of cooperative nodes can be reduced while maintaining the probabilistic delay bound. As the delay bound becomes tighter, each line moves away from the origin, implying that more cooperative nodes are required to meet the delay requirement.

VII. Structural Results for Packet Delivery Probability of an AF System

This section examines how the packet delivery probability [p_s in (6)] changes with respect to system parameters (e.g., SNR, ARQ persistency). The main result in this section is the sufficient condition (**A3** in Result 2), which ensures an increase in packet delivery probability. This result will be useful in many aspects. For example, it is used to reduce complexity of the power allocation algorithm in Section VIII.

The main challenge in deriving the sufficient condition is the unavailability of a closed-form expression of the blocking probability. The blocking probability [p_b in (14)] is a function of the stationary probability vector $(\boldsymbol{\pi})$ of the DTMC developed in Section V. In general, the stationary probability vector

is computed by numerical evaluation. Involving matrix inversion, such the evaluation requires a prohibitive computational effort. Therefore, our objective is to understand how the packet delivery probability (p_s) behaves structurally with respect to system parameters (defined below) through the monotone results developed in this section.

Define physical-layer parameters as $\Theta \triangleq \{\overline{\gamma}_d, \overline{\gamma}_{ij}, R, -H_i\}$, $i = \{1, \ldots, R\}$, $j = \{1, \ldots, H_i\}$, ARQ persistency as $\Psi \triangleq \{\psi_i; K\}$, $i = \{1, \ldots, K\}$, buffer management parameters as $\Xi \triangleq \{B; \nu_x\}$, $x \in \mathcal{X}$, and traffic load as $\Lambda \triangleq \{\mathbf{B}; \mathbf{A}_n\}$, $n \in \{0, \ldots, N\}$ where all the above variables are defined in Table I. Also, define a parameter adjustment policy $\mathbf{A1}$ as follows:

$$\mathbf{A1}: \begin{cases} a) \text{ Adjust physical parameters } (\Theta) \text{ to decrease } \epsilon_p \\ b) \text{ Decrease ARQ persistency } (\Psi) \\ c) \text{ Decrease average arrival rate } (\overline{\lambda}) \\ d) \text{ Fixed buffer management parameters } (\Xi). \end{cases}$$

To decrease the packet error probability ϵ_p, we may increase SNR ($\overline{\gamma}_d$ and $\overline{\gamma}_{ij}$), increase the number of cooperative branches (R), and/or decrease the number of cooperative nodes on the cooperative branch i (H_i). To decrease ARQ persistency Ψ, we decrease retransmission probability when the transmission counter is i (ψ_i) and/or the retry limit K. We adjust D-BMAP parameters $\mathbf{B}$ and $\mathbf{A}_n$ to decrease $\overline{\lambda}$ defined in (3). Finally, fixed Ξ implies that the buffer size (B) and the marking probability when the DTMC state is x (ν_x) are given and do not change.

Intuitively, it may be thought that $\mathbf{A1}$ leads to a decrease in packet blocking probability (p_b). However, a numerical example in Result 1 shows that this intuition is not always true. A *decrease* in average arrival rate $(\overline{\lambda})$ does not necessarily result in a *decrease* in p_b.

Since a decrease in ARQ persistency leads to an increase in packet dropping probability (p_d), we cannot derive a sufficient condition for an increase in packet delivery probability $(p_s = (1 - p_d)(1 - p_b))$ from $\mathbf{A1}.b$. Next, we introduce the concept of *stochastic dominance*. Then, we use the stochastic dominance concept to derive a parameter adjustment policy $\mathbf{A2}$, which leads to a decrease in p_b in Section VII-B, and a policy $\mathbf{A3}$, which ensures an increase in p_s in Section VII-C.

A. Stochastic Dominance

Definition 1: Let $\mathcal{X}$ and $\mathcal{Y}$ be two discrete random variables with probability mass functions $f_{\mathcal{X}}(u)$ and $f_{\mathcal{Y}}(u)$, respectively, where $u \in \{1, \ldots, U\}$ and U is a positive integer. Suppose $\sum_{i=u}^{U} f_{\mathcal{X}}(i) \leqslant \sum_{i=u}^{U} f_{\mathcal{Y}}(i)$, $\forall\, u$. Then, $\mathcal{X}$ is said to be stochastically smaller than $\mathcal{Y}$, denoted by $\mathcal{X} <_{st} \mathcal{Y}$ or by $(f_{\mathcal{X}}(u))_{u=1}^{U} <_{st} (f_{\mathcal{Y}}(u))_{u=1}^{U}$. Furthermore, $\mathcal{X} <_{st} \mathcal{Y}$ if and only if $\sum_{i=1}^{U} g(i) f_{\mathcal{X}}(i) \leqslant \sum_{i=1}^{U} g(i) f_{\mathcal{Y}}(i)$ for all increasing function $g(i)$ [16]. □

Definition 2: Let $\mathbf{P}$ be an $N \times N$ stochastic matrix, where N is a positive integer. Denote row i of $\mathbf{P}$ by $\mathbf{p}_i$. Suppose, $\mathbf{p}_i <_{st} \mathbf{p}_{i+1}$, $i = \{1, \ldots, N-1\}$. Then, $\mathbf{P}$ is said to be order-keeping [16]. □

Definition 3: Let $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ be two $N \times N$ stochastic matrices, where N is a positive integer. Suppose $\mathbf{p}_i^{(1)} <_{st} \mathbf{p}_i^{(2)}$, $i = \{1, \ldots, N\}$. Then, $\mathbf{P}^{(1)}$ is said to be stochastically smaller than $\mathbf{P}^{(2)}$ (denoted by $\mathbf{P}^{(1)} <_{st} \mathbf{P}^{(2)}$) [16]. □

Proposition 1: Let $\mathbf{P}^{(1)}$, $\mathbf{P}^{(2)}$, $\mathbf{Q}^{(1)}$, and $\mathbf{Q}^{(2)}$ be $N \times N$ stochastic matrices, where N is a positive integer.

1) If $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ are order-keeping (see Definition 2), $\mathbf{P}^{(1)}\mathbf{P}^{(2)}$ is also order-keeping.
2) If $\mathbf{P}^{(1)} <_{st} \mathbf{P}^{(2)}$, $\mathbf{Q}^{(1)} <_{st} \mathbf{Q}^{(2)}$, and at least one of $\mathbf{Q}^{(1)}$ or $\mathbf{Q}^{(2)}$ is order-keeping, then $\mathbf{P}^{(1)}\mathbf{Q}^{(1)} <_{st} \mathbf{P}^{(2)}\mathbf{Q}^{(2)}$.

□

Proof: See [16]. ■

Proposition 2: Consider two DTMCs with transition probability matrices $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ and with stationary probability vectors $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$. Suppose at least one of $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ is order-keeping. Then, $\mathbf{P}^{(1)} <_{st} \mathbf{P}^{(2)} \Rightarrow \boldsymbol{\pi}^{(1)} <_{st} \boldsymbol{\pi}^{(2)}$. □

Proof: By Proposition 1, $\boldsymbol{\pi}^{(1)} = \lim_{k\to\infty}(\mathbf{P}^{(1)})^k <_{st} \lim_{k\to\infty}(\mathbf{P}^{(2)})^k = \boldsymbol{\pi}^{(2)}$. ■

B. Sufficient Condition $(\mathbf{A2})$ for a Decrease in Blocking Probability

Here, we analyze the blocking probability of an AF system assuming an independent arrival process, whose corresponding model was presented in Section V-C. The main result for the blocking probability is the following.

Result 1: Consider an AF system with an independent arrival process, where, in each transmission interval, a batch of n packets arrives with probability $a_n, n \in \{0, \ldots, N\}$. Then:

- $\mathbf{A1}$ does not necessarily lead to a decrease in blocking probability (p_b);
- $\mathbf{A2}$, given in the following, is sufficient to ensure that the blocking probability decreases for both finite transmission buffer and infinite transmission buffer cases:

$$\mathbf{A2}: \begin{cases} a) \text{ Adjust } (\Theta; \Psi) \text{ to decrease } r_i, r_i^+, i = \{1, \ldots, K\} \\ b) \text{ Stochastically decrease } \mathbf{a} \text{ and } \mathbf{a}^+ \\ c) \text{ Fixed } (\Xi) \text{ with } \nu_i \text{ increasing in buffer occupancy } i \end{cases}$$

where r_i, defined in (8), is the probability that a packet is transmitted for at least i times, and $r_i^+ = \sum_{j=i}^{K} \frac{r_j}{\sum_{n=1}^{K} r_n}$. Arrival matrices are $\mathbf{a} = (a_0, \ldots, a_n)$, $\mathbf{a}^+ = (a_0^+, \ldots, a_n^+)$, and $a_n^+ = n a_n / \sum_{i=1}^{N} i a_i$. Θ, Ψ, and Ξ are physical parameters, ARQ persistency, and buffer management parameters, respectively.

□

Proof: To demonstrate the first claim, we construct an example where $\mathbf{A1}$ does not lead to a decrease in blocking probability. Consider an example with two following cases. Both cases have the same physical (Θ), ARQ (Ψ), and buffer management (Ξ) parameters, but they have different arrival probabilities: Case 1 with high average arrival rate $\overline{\lambda} = 0.65$ and Case 2 with low average arrival rate $\overline{\lambda} = 0.6$. Arrival probability vectors $(\mathbf{a})$ for these two cases, and the marking probability vector $(\boldsymbol{\nu})$ for buffer size $B = 6$ are shown in Table III. All other parameters are set to be the same as in Table II.

In this particular example, a decrease in the average arrival rate $(\overline{\lambda} = 0.65 \rightarrow 0.6)$ leads to an increase in the blocking probability $(p_b = 0.2353 \rightarrow 0.3208)$. This statement proves the first claim of Result 1. This counter-intuitive behavior occurs due to the tail distribution of the arrival processes of Case 1 and Case 2, which are plotted in Fig. 6. Although Case 1 has higher arrival rate $(\overline{\lambda})$, its arrival probability is not stochastically larger

TABLE III
COUNTER EXAMPLE WHICH CORROBORATES RESULT 1: DECREASE IN ARRIVAL RATE CAUSES AN INCREASE IN BLOCKING PROBABILITY

	Case 1	Case 2
$\bar{\lambda}$	0.65	0.6
p_b	0.2353	0.3208
$\mathbf{a}$	$(0.65, 0.15, 0.1, 0.1, 0)$	$(0.85, 0, 0, 0, 0.15)$
$\boldsymbol{\nu}$	$(0, 0, 0, 0.1, 0.3, 0.5, 1)^\top$	

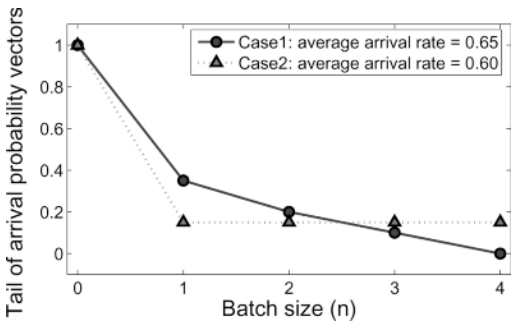

Fig. 6. Counter example: although the average arrival rate of Case 1 is larger, its tail is not larger. In this particular example, the blocking probability corresponding to Case 1 is less than that corresponding to Case 2 (see Table III).

than that of Case 2. Hence, the blocking probability for Case 1 is less than that for Case 2.

To prove the sufficiency, of $\mathbf{A2}$, consider $p_b = (\boldsymbol{\pi}_{sm}\mathbf{p}_b)\boldsymbol{\nu}$ in (23). By Proposition 2, p_b decreases (or $-p_b$ increases) if: 1) $\boldsymbol{\nu}$ is an increasing function of the buffer occupancy and 2) $-\boldsymbol{\pi}_s\mathbf{p}_b$ increases stochastically or $\boldsymbol{\pi}_s\mathbf{p}_b$ decreases stochastically. By $\mathbf{A2}.c$, $\boldsymbol{\nu}$ increases as the buffer occupancy increases. We show in Appendix III that $\boldsymbol{\pi}_s\mathbf{p}_b$ is stochastically decreasing, if $\mathbf{A2}.a$ and $\mathbf{A2}.b$ hold. Therefore, we conclude that p_b decreases under $\mathbf{A2}$. Note that the proof is valid for both finite and infinite buffers, since the definitions and propositions in Section VII-A are based on matrices whose dimension are denumerable. ■

Remark 1:

- Suppose $\mathbf{x}$ is a stochastic matrix. To stochastically decrease $\mathbf{x}$ is to adjust $\mathbf{x}$ to $\mathbf{x}'$ such that $\mathbf{x}' <_{st} \mathbf{x}$.
- Compared to $\mathbf{A1}.a$ and $\mathbf{A1}.b$, $\mathbf{A2}.a$ is less restrictive and potentially more useful. It can be shown that $\{\mathbf{A1}.a \text{ and } \mathbf{A1}.b\} \Rightarrow \{\Omega_i \text{ is decreasing } \forall i\} \Rightarrow \mathbf{A2}.a$, but the converse is not true. By $\mathbf{A1}$, we cannot characterize the trend of blocking probability when ARQ persistency and transmission power increase simultaneously. $\mathbf{A2}.a$, on the other hand, suggests how much power should be increased such that an increase in ARQ persistency does not lead to an increase in packet blocking probability.
- Ensuring a decrease in the packet blocking probability p_b, $\mathbf{A2}$ involves no matrix inversion. With $\mathbf{A2}$, we can adjust the system parameters to decrease p_b with protocol complexity $O(n)$ rather than $O(n^3)$, where n is the size of a transition probability matrix.
- The structural results such as how a policy depends on the buffer state are of significant interest [17]. They give insight into the behavior of the system rather than merely brute force computation to obtain a number.

C. Sufficient Condition ($\mathbf{A3}$) for an Increase in Packet Delivery Probability

We conclude this section with the main result of packet delivery probability in Result 2 below.

Result 2: Consider an AF system with a finite or an infinite transmission buffer in Fig. 1. Then, assuming an independent arrival process, $\mathbf{A3}$ as follows is sufficient to ensure an increase in packet delivery probability (p_s):

$$\mathbf{A3}: \begin{cases} a) \text{ Adjust } \Theta \text{ and } \Psi \text{ to decrease } p_d, r_i, r_i^+, j = \{1, \ldots, K\} \\ b) \text{ Stochastically decrease } \mathbf{a} \text{ and } \mathbf{a}^+ \\ c) \text{ Fixed } \Xi \text{ with } \nu_i \text{ increasing in buffer occupancy.} \end{cases}$$

Here, r_i, r_i^+, $\mathbf{a}$, and $\mathbf{a}^+$ are defined in Result 1. Θ, Ψ, and Ξ are physical parameters, ARQ persistency, and buffer management parameters, respectively. □

Proof: It is simple to see that $\mathbf{A2} \Rightarrow \mathbf{A3}$, and $\mathbf{A3}.a$ also implies a decrease in p_d. Therefore, packet delivery probability increases under $\mathbf{A3}$. ■

VIII. POWER ALLOCATION PROBLEM: MARKOVIAN MODEL AND STRUCTURAL RESULTS

In Section VI, we showed that a CD-AF scheme allows us to decrease the transmission power without compromising the link layer performance (e.g., SNR saving in Figs. 3 and 4). However, we did not quantify how much power can be reduced for an additional cooperative node. In this section, we are interested in "computing minimum power which satisfies a constraint on packet delivery probability p_s." This is formalized as follows:

Problem 1: Given the following:

- all physical parameters in Θ except for transmission power, ARQ persistency Ψ, traffic load Λ, and buffer management paramters Ξ;
- minimum transmission power ($X_{\min}$) and maximum transmission power ($X_{\max}$);
- target packet delivery probability P_{th}, we have

$$\text{Compute } X^* = \min\{X \in [X_{\min}, X_{\max}] : p_s(X) = P_{\text{th}}\} \tag{24}$$

where $p_s(X)$ denotes the packet delivery probability corresponding to transmission power X, and X^* is the minimum transmission power whose corresponding packet delivery probability is P_{th}. □

A. Direct Power Computation Algorithm

Since $p_s(\cdot)$ in (24) does not have a closed-form expression, we resort to a numerical method described in Algorithm 1 to solve Problem 1.

Algorithm 1:

(i) Set $k = 0$ and Initialize x_0.

(ii) Compute $p_k = p_s(X_k)$.

(iii) If $p_k = P_{th}$, store X_k in X^*, terminate the algorithm, and return X^* as the solution. Otherwise, set $k = k + 1$, adjust X_k, and go back to Step (ii). □

Algorithm 1 is a deterministic search algorithm for computing X^*. There are numerous variants of Algorithm 1 in the literature.

B. Power Update Algorithm

This section shows how the structural results in Section VII can help reduce complexity of Algorithm 1. The expensive step of Algorithm 1 is the packet delivery probability computation process [i.e., computing $p_s(X_k)$ in Step (ii)], where the matrix inversion requires $O(n^3)$ complexity. Several iterations may be required to get close to the solution X^*. Furthermore, since system parameters (e.g., number of cooperating nodes) change over time, we may need to re-compute the transmission power repeatedly (by using Algorithm 1). Below, we design a power update algorithm, which avoids the necessity to recompute the packet delivery probability using Algorithm 1.

The main idea of the power update algorithm is as follows: React to a change in system parameters by adjusting the transmission power in order to keep the packet delivery probability unchanged. This task is fairly straightforward at the physical layer, since we only have to keep the symbol error probability ϵ_s in (2) unchanged. The problem becomes more complicated at the link layer, when ARQ persistency changes. A decrease in ARQ persistency decreases packet blocking probability p_b but increases packet dropping probability p_d. We need to use the sufficient condition **A3** in Result 2 for the power update algorithm. The description of the power update algorithm is given in Algorithm 2 here.

Algorithm 2:

(i) Set $k = 0$. Let the Current Parameter Setting Be $(\theta_k; \psi_k; \xi_k; \mathbf{A}_k)$.

(ii) Compute transmission power X_k^* using Algorithm 1.

(iii) Wait for a parameter change and update parameters $(\Theta_k; \Psi_k; \Xi_k; \mathbf{a}_k)$.

(iv) If the parameter change is in Θ_k, adjust the transmission power such that

$$\epsilon_s\left(X_{k+1}^*\right) = \epsilon_s\left(X_k^*\right). \tag{25}$$

In other words, set the transmission power to keep ϵ_s unchanged. Set $k = k + 1$ and go back to Step (iii).

(v) If the parameter change is in Ψ_k or both Ψ_k and Θ_k, adjust the transmission power such that

$$\begin{aligned} p_d\left(X_{k+1}^*\right) &= p_d\left(X_k^*\right), \quad r_i\left(X_{k+1}^*\right) = r_i\left(X_k^*\right), \\ r_i^+\left(X_{k+1}^*\right) &= r_i^+\left(X_k^*\right), \quad i = \{1, \ldots, K\} \end{aligned} \tag{26}$$

that is, set the transmission power to keep p_d, r_i, and r_i^+ unchanged. Set $k = k + 1$ and go back to Step (iii).

(vi) Set $k = k + 1$ and go back to Step (ii). □

The objective of Algorithm 1 is to directly search for a solution, while that of Algorithm 2 is to update the solution as the system parameters change. Therefore, Algorithm 1 is a part [Step (ii)] of Algorithm 2.

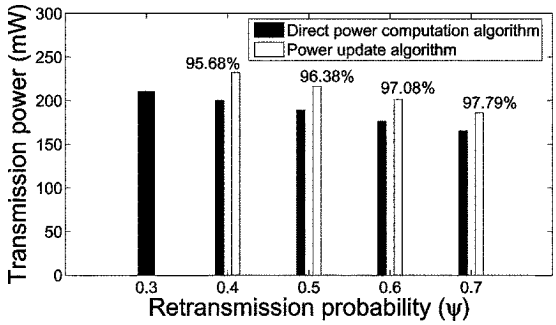

Fig. 7. Comparison of direct power computation and power update algorithms.

Algorithm 2 reduces the number of times Algorithm 1 needs to be invoked in each iteration (i.e., each parameter change). We only need to update (not to recompute) transmission power, when the parameters in Θ and/or Ψ change. Note that Steps (iii) and (iv) in Algorithm 2 have complexity of $O(n)$, which is less than $O(n^3)$ in case of Algorithm 1.

C. Example Numerical Result

We now present examples to illustrate the power allocation algorithms. Fig. 7 shows the minimum transmission power which satisfies the target packet delivery probability P_{th} of 95% in presence of 1 mW noise power. We modified slightly the experiment setup in Section VI, by setting the ARQ parameters to be $\psi_i = \psi$, $i = \{1, \ldots, K-1\}$ and $\psi_K = 0$. Other system parameters are the same as in Table II.

Fig. 7 shows the solutions of both the direct power computation algorithm (Algorithm 1; black bars) and the power update algorithm (Algorithm 2; white bars). We compute the transmission power using Algorithm 1 for $\psi = \{0.3, 0.4, 0.5, 0.6, 0.7\}$. The computed power at $\psi = 0.3$ is used as a reference to update transmission power using Algorithm 2. The numbers on the top of the white bars are the actual packet delivery probabilities corresponding to the transmission power obtained from Algorithm 2.

Since Algorithm 2 is developed from the sufficient condition on packet delivery probability, the transmission power obtained from Algorithm 2 is slightly higher than that obtained from Algorithm 1. Correspondingly, the packet delivery probability obtained from Algorithm 2 is slightly higher than that obtained from Algorithm 1.

IX. Conclusion

We developed a Markovian model to evaluate the performance of an amplify-and-forward cooperative diversity (CD-AF) wireless system. The model takes into account transmission using a CD-AF scheme at the physical layer, a link-layer flow control protocol, an ARQ-based error recovery mechanism, and the dynamics of the transmitting buffer at the radio link layer, and a traffic generation pattern from the application layer. The numerical evaluation and analysis of the model yield the following interesting results.

- The proposed model quantifies the equivalent SNR saving per additional cooperative node to maintain the packet delivery probability and packet deliver delay.

- The minimum transmission power and/or the number of cooperative nodes required for a given probabilistic delay bound can be computed by using the proposed model.
- The packet delivery probability monotonically increases under $\mathbf{A3}$ in Result 2.

We also presented two power allocation algorithms. The direct power computation algorithm utilizes the Markovian model to compute the minimum transmission power which serves as a bound for the packet delivery probability. The second algorithm proposed is a power update algorithm that exploits the sufficient condition $\mathbf{A3}$ in Result 2 to reduce the complexity of the former algorithm. Numerical results show that a mobile using the power update algorithm uses higher transmission power, and achieves higher packet delivery probability compared to operating under the direct power computation algorithm.

Appendix I
Derivation of Conditional Distribution $(f_{\mathcal{D}|\mathcal{X}}(d|x))$ and Distribution $(\mathbf{f}_{\mathcal{D}}(d))$ of Packet Delivery Delay

To Compute $f_{\mathcal{D}|\mathcal{X}}(d|X)$: First, let $M=1$, $\mathbf{B}=1$, $\mathbf{A}_0=1$, and $\mathbf{A}_n=0$ for $n=\{1,\ldots,N\}$. Secondly, formulate a transient transition probability matrix $(\mathbf{Q})$ as $\mathbf{P}$. Then, remove the rows and columns corresponding to states $\{x=(a,b,k)\in\mathcal{X}: b=0 \text{ or } b>\mathcal{B}(x)\}$, where $\mathcal{B}(x)$ returns the buffer occupancy corresponding to the state $x\in\mathcal{X}$. Next, formulate an absorbing transition probability matrix $(\mathbf{Q}_0)$ as $\begin{pmatrix}\boldsymbol{\omega}\\ \mathbf{0}\end{pmatrix}$. Finally, compute $f_{\mathcal{D}|\mathcal{X}}(d|x)$ from

$$f_{\mathcal{D}|\mathcal{X}}(d|x)=\begin{cases} f_z, & \mathcal{B}(x)=0\\ f_{sd}, & \mathcal{K}(x)=K\\ (1-\epsilon_p)\times f_{sd}+\epsilon_p\times f_f, & \text{otherwise,}\end{cases}$$

$$f_\star=\frac{\boldsymbol{\beta}_\star\mathbf{Q}^{d-1}\mathbf{Q}_0}{\boldsymbol{\beta}_\star(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{Q}_0},\quad \star\in\{z,sd,f\},\tag{27}$$

where (27) is the absorption time distribution normalized by successful transmission probability [15], $\mathcal{K}(x)$ is the transmission counter when the DTMC state is x. The entries $(1,z)$, $(1,sd)$, and $(1,f)$ of the initial probability vectors $\boldsymbol{\beta}_z$, $\boldsymbol{\beta}_{sd}$, and $\boldsymbol{\beta}_f$, are one and their other entries are zero, where $z=\{(a,b,k)\in\mathcal{X}: b=1,k=0\}$, $sd=\{(a,b,k)\in\mathcal{X}: b=\mathcal{B}(x),k=0\}$, and $f=\{(a,b,k)\in\mathcal{X}: b=\mathcal{B}(x)+1,k=\mathcal{K}(x)+1\}$.

To Formulate $\mathbf{f}_{\mathcal{D}}(d)$: For each $x\in\mathcal{X}$, we weigh $f_{\mathcal{D}|\mathcal{X}}(d|x)$ with appropriate probabilities and a normalization factor as follows. Suppose $n\geqslant m$ packets are admitted to the transmitting buffer in a transmission interval. The delay of the first admitted packet for $x=(a,b,k)$ is the same as that of the mth admitted packet for $x=(a,b-m+1,k)$, $m=\{1,\ldots,N\}$. Since the probability of admitting the mth packet when the buffer occupancy is $b-m+1$ is $\sum_{m=n}^{N}\mathbf{c}_m(b-m+1)$, we formulate weighting matrix $\mathbf{D}$ as in (17). Since $\mathbf{D}$ is not a stochastic matrix, we normalize (17) with $\boldsymbol{\pi}\mathbf{D}\mathbf{1}$.

Appendix II
Derivation of Blocking Probability p_b in (15)

Consider $\mathbf{p}_b$ in (15). The conditional blocking probability of the nth packet given the batch of $n,n+1,\ldots,N$ packets is $\boldsymbol{\xi}^{n-1}\boldsymbol{\nu}$. Therefore, we weigh the blocking probability $\boldsymbol{\xi}^{n-1}\boldsymbol{\nu}$ with $\sum_{i=n}^{N}[\mathbf{I}_B\otimes\mathbf{B}A_i\otimes\mathbf{I}_K]$ and normalize the first term with the denominator in (15). This formulates the first term of $\mathbf{p}_b$ in (15). Equivalently, the blocking probability is defined on the space of "packets" not "transmission intervals." We take into account the number of packets generated in each interval, by using $\mathbf{A}_n^+$ instead of $\mathbf{A}_n$. $\left[\mathbf{I}_B\otimes\mathbf{B}A_n^+\otimes\mathbf{I}_K\right]$ represents the probability that an incoming packet belongs a batch of n packets. Given that the batch size is n, the average blocking probability is $\frac{1}{n}\sum_{m=1}^{n}\boldsymbol{\xi}^{m-1}\boldsymbol{\nu}$. This yields the second term of $\mathbf{p}_b$ in (15).

Appendix III
Part of the Proof of Result 1

First, we claim that, under $\mathbf{A2}.a$ and $\mathbf{A2}.b$, both $\boldsymbol{\pi}_s$ and $\mathbf{p}_b$ are stochastically decreasing and $\mathbf{p}_b$ is order-keeping. By Proposition 1, this claim implies that $\boldsymbol{\pi}_s\mathbf{p}_b$ is stochastically decreasing, as required. The proofs of the above claims are given below.

$\boldsymbol{\pi}_s$ is Stochastically Decreasing: Since $\boldsymbol{\xi}$ is order-keeping, so is $\boldsymbol{\xi}^i$. Since $(a_0,\ldots,a_N)$ and $\mathbf{T}=(T_1,\ldots,T_K)$ are probability vectors, $\mathbf{C}_z=\sum_{i=0}^{N}a_i\boldsymbol{\xi}^i$ and $\sum_{j=1}^{K}T_j\mathbf{C}_z^j$ are order-keeping. The bottom part of $\mathbf{P}_{sm}$ (i.e., $\mathbf{C}_{nz}$) is now order-keeping. Since $\mathbf{I}<_{st}\mathbf{C}_z$, we have $\mathbf{C}_z<_{st}\mathbf{C}_z^j, j=\{1,\ldots\}$ and $\mathbf{C}_z<_{st}\sum_{j=1}^{K}T_j\mathbf{C}_z^j$ (by Proposition 3). Therefore, every row of $\mathbf{C}_z$ is stochastically smaller than every row of $\mathbf{C}_{nz}$, and $\mathbf{P}_{sm}$ is order-keeping.

Again, $\boldsymbol{\xi}<_{st}\cdots<_{st}\boldsymbol{\xi}^N$. By $\mathbf{A2}.b$, $\mathbf{a}$ is stochastically decreasing, and so is $\mathbf{C}_z=\sum_{i=0}^{N}a_i\boldsymbol{\xi}^i$ (by Proposition 4). Similarly, $\mathbf{C}_z<_{st}\cdots<_{st}\mathbf{C}_z^K$. By $\mathbf{A2}.a$, $\mathbf{t}=(t_1,\ldots,t_K)$ is stochastically decreasing, and so is $\sum_{i=0}^{N}t_i\mathbf{C}_z^j$ (by Proposition 4). Since $\mathbf{P}_{sm}$ is order-keeping and every row of $\mathbf{P}_{sm}$ is stochastically decreasing, $\boldsymbol{\pi}_s$ is stochastically decreasing by Proposition 2.

$\mathbf{p}_b$ is Order-Keeping and Stochastically Decreasing: Again, $\boldsymbol{\xi}$ and $\mathbf{C}_z$ are order-keeping, and so are $\overline{\boldsymbol{\xi}}_z=\sum_{n=1}^{N}a_n^+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}^{i-1}$ and $\overline{\boldsymbol{\xi}}_{nz}=\sum_{i=1}^{K}T_i\mathbf{C}_z^{i-1}\overline{\boldsymbol{\xi}}_z$. Clearly, $\overline{\boldsymbol{\xi}}_z<_{st}\overline{\boldsymbol{\xi}}_{nz}$. Therefore, $\mathbf{p}_b$ is order-keeping. By $\mathbf{A2}.b$, $\mathbf{a}^+$ is stochastically decreasing, and so is $\overline{\boldsymbol{\xi}}_z$ (by Proposition 4). By $\mathbf{A2}.a$, $r_j^+=\frac{\sum_{n-j}^{K}\prod_{i-m}^{K}\Omega_m}{\sum_{n-1}^{K}\prod_{i-m}^{K}\Omega_m}$ is decreasing for all j, impling that $\mathbf{T}=(T_1,\ldots,T_K)$ is stochastically decreasing. Coupled with that $\mathbf{C}_z$ is stochastically decreasing, $\overline{\boldsymbol{\xi}}_{nz}$ are stochastically decreasing (by Proposition 4). Therefore, we conclude that under $\mathbf{A2}.a$ and $\mathbf{A2}.b$, $\mathbf{p}_b$ is order-keeping and stochastically decreasing.

Appendix IV
Useful Propositions

Define a tail function $\mathcal{T}:\mathbb{R}^{N\times N}\to\mathbb{R}^{N\times N}$ as $[\mathcal{T}(\mathbf{X})]_{ij}=\sum_{k=j}^{N}[\mathbf{X}]_{ik}$.

Proposition 3: Consider a set of stochastic matrices $\mathbf{X}_0<_{st}\cdots<_{st}\mathbf{X}_N$. Suppose $\mathbf{a}=(a_0,\ldots,a_N)$ is a probability vector. Then

$$\mathbf{X}_0<_{st}\sum_{i=0}^{N}a_i\mathbf{X}_i<_{st}\mathbf{X}_N.\tag{28}$$

□

Proof: [Contradiction] Let $\mathbf{T}_i = \mathcal{T}(\mathbf{X}_i)$. Suppose

$$a_0\mathbf{T}_0 + \sum_{i=1}^{N} a_i\mathbf{T}_i \prec \mathbf{T}_0$$

$$\sum_{i=1}^{N} a_i(\mathbf{T}_i - \mathbf{T}_0) \prec \mathbf{0}.$$

Since $a_i \geq 0$ and $\mathbf{T}_i - \mathbf{T}_0 \succ \mathbf{0}$, we establish a contradiction, which proves $\mathbf{X}_0 <_{st} \sum_{i=0}^{N} a_i\mathbf{X}_i$. We use the same approach to prove $\sum_{i=0}^{N} a_i\mathbf{X}_i <_{st} \mathbf{X}_N$.

Proposition 4: Consider a set of stochastic row vectors $\mathbf{X}_0 <_{st} \cdots <_{st} \mathbf{X}_N$. Suppose $\mathbf{a} = (a_0, \ldots, a_N) <_{st} (b_0, \ldots, b_N) = \mathbf{b}$. Then

$$\sum_{i=0}^{N} a_i\mathbf{X}_i <_{st} \sum_{i=0}^{N} b_i\mathbf{X}_i. \tag{29}$$

□

Proof: Let $\mathbf{T}_i = \mathcal{T}(\mathbf{X}_i)$. With simple manipulation, we have

$$\sum_{i=0}^{N} a_i\mathbf{T}_i = \sum_{i=0}^{N} \left\{ \left(\sum_{j=i}^{N} a_j \right) (\mathbf{T}_i - \mathbf{T}_{i-1}) \right\}. \tag{30}$$

Similarly, $\sum_{i=0}^{N} b_i\mathbf{T}_i = \sum_{i=0}^{N} \left\{ \left(\sum_{j=i}^{N} b_j \right) (\mathbf{T}_i - \mathbf{T}_{i-1}) \right\}$. Since $\mathbf{a} <_{st} \mathbf{b}$, $\mathcal{T}(\mathbf{a}) \prec \mathcal{T}(\mathbf{b})$. Together with $\mathbf{X}_i >_{st} \mathbf{X}_{i-1}$, the proposition follows.

REFERENCES

[1] D. Gesbert, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Mar. 2003.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[3] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[4] S. Cui, R. Madan, A. J. Goldsmith, and S. Lall, "Cross-layer energy and delay optimization in small-scale sensor networks," *IEEE/ACM Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3688–3699, Oct. 2007.

[5] R. Pabst *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sep. 2004.

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

[7] P. A. Anghel and M. Kaveh, "Exact symbol error probability of a cooperative network in a Rayleigh-fading environment," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1416–1421, Sep. 2004.

[8] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1264–1272, May 2005.

[9] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Netw.*, vol. 1, no. 4, pp. 397–413, 1993.

[10] Q. Li and M. V. D. Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 278–290, Apr. 2004.

[11] Q. Liu, S. Zhou, and G. B. Giannakis, "Queueing with adaptive modulation and coding over wireless links: Cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, May 2005.

[12] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[13] A. E. Kamal, "Discrete-time modeling of TCP reno under background traffic interference with extension to RED-based routers," *Elsevier Performance Eval.*, vol. 58, no. 2+3, pp. 109–142, 2004.

[14] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE standard 802.16 Working Group Std., 2002.

[15] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*. Baltimore, MD: The John Hopkins Univ. Press, 1981.

[16] S. Cheng and J. Zhou, "On some applications of stochastic orders to actuarial science," in *Proc. Int. Conf. Appl. Statistics, Actuarial Sci. Financial Math.*, Dec. 2002, Available online.

[17] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ: Princeton Univ. Press, 1998.

Teerawat Issariyakul (S'02–M'06) received the B.Eng. degree from Thammasat University in 1997, the M.Eng. degree from the Asian Institute of Technology (AIT), Pathumthani, Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering, University of Manitoba, Winnipeg, MB, Canada, in 2005.

From 2001 to 2005, he was affiliated with the Telecommunication Research Laboratory (TRLabs). In 2005 and 2008, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of British Columbia and the University of Manitoba, respectively. Currently, he is a System Engineer with TOT Public Company Limited, Bangkok, Thailand. He also holds an adjunct faculty position with the Asian Institute of Technology. His current research area is modeling and optimization of cognitive radio networks and intelligent transportation networks.

Vikram Krishnamurthy (M'91–SM'99–F'05) was born in 1966. He received the B.S. degree from the University of Auckland, Auckland, New Zealand, in 1988 and the Ph.D. degree from the Australian National University, Canberra, in 1992.

Since 2002, he has been a Professor and Canada Research Chair with the Department of Electrical Engineering, University of British Columbia, Vancouver, BC, Canada. Prior to this, he was a chaired Professor with the Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, Australia. His research interests span several areas including ion channels and nanobiology, stochastic scheduling and control, statistical signal processing and wireless telecommunications.

Dr. Krishnamurthy has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS AEROSPACE AND ELECTRONIC SYSTEMS, the IEEE TRANSACTIONS ON NANOBIOSCIENCE, the IEEE TRANSACTIONS CIRCUITS AND SYSTEMS II, *Systems and Control Letters,* and *European Journal of Applied Signal Processing*. He was a Guest Editor of a special issue of the IEEE TRANSACTIONS ON NANOBIOSCIENCE, March 2005, on bio-nanotubes.