Threshold Based Relay Selection for Buffer-Aided Cooperative Relaying Systems

Sawsan El-Zahr and Chadi Abou-Rjeily, *Senior Member IEEE*

***Abstract*—In this paper, we propose a novel relay selection strategy for half-duplex decode-and-forward cooperative networks with an arbitrary number of buffer-aided (BA) relays. Unlike most of the existing predetermined relaying protocols, the proposed strategy is adjustable in the sense that it is controlled by K threshold levels. A Markov chain (MC) analysis is adopted for evaluating the outage probability (OP) and average packet delay (APD) of the proposed scheme. Through an asymptotic analysis, we highlight on the impact of the K controlling parameters on the triad of OP, APD and diversity order that can be contemplated. While most of the existing schemes are designed to achieve fixed APD and diversity order values, the proposed scheme can achieve all diversity orders ranging from K to $2K$ while compromising the asymptotic APD that will range from 2 to $2K+2$. We also target the optimization of the buffer size and we prove that a buffer size of three is sufficient for extracting the full capabilities of the BA network. Simulations over Rayleigh fading channels demonstrate the performance gains and the OP-APD tradeoffs that can be attained.**

***Index Terms*—Relaying, cooperative networks, relay selection, buffer, data queue, performance analysis, Markov chain, outage probability, queuing delay, diversity order, half-duplex, buffer size, decode-and-forward.**

I. Introduction

The recent unparalleled increase of mobile data traffic necessitates the development of spectrally-efficient physical-layer techniques [1]. Among the fifth generation (5G) technologies, buffer-aided (BA) cooperative relaying has the potential of improving the reliability as long as the introduced queuing delays can be tolerated [2], [3]. BA cooperative relaying achieves diversity gains by taking advantage of the presence of a set of relays (R's) between the source node (S) and the destination node (D). Equipping the relays with buffers constitutes an additional degree of freedom since the information packets can be temporarily stored until the underlying channel conditions are more favorable. In half-duplex (HD) BA cooperative networks, a key factor that influences the system performance is the relay selection strategy. This strategy determines which relay is to transmit or receive within a time slot, thus predominantly affecting the levels of outage probability (OP) and average packet delay (APD) that can be achieved. This paper tackles relay selection in HD BA networks with an arbitrary number of relays. We introduce a novel relay selection protocol, and we study its impact on the OP, APD and diversity order.

The ubiquitous influence of the BA relaying technology manifests in the large number of studies that covered a broad range of research topics including HD amplify-and-forward (AF) relaying [4], HD decode-and-forward (DF) relaying [5]–[19], full-duplex (FD) BA relaying [20], BA relaying with non-orthogonal multiple access (NOMA) [21], physical layer security in BA relaying networks [22] and BA relaying for optical wireless communications [23]. Single-relay BA DF HD relaying was considered in [5]–[7]. In [5], [6], the objective was to maximize the throughput over a communication session that extends over an infinite number of time slots while ignoring the probability of buffer overflow. Finite buffers were considered in [7] where the performance evaluation revolved around the OP and APD.

The problem of relay selection in multi-relay BA DF HD networks was considered in [8]–[19]. In what follows, we denote by K the number of relays and by L the buffer size at each relay. The *max-link* scheme was suggested in [8] and consists of selecting the strongest link among all available S-R and R-D links. The *max-link* scheme achieves the full diversity order of $2K$ for infinitely large buffer sizes while suffering from a high APD of $KL+1$. In an attempt to reduce the APD of the *max-link* protocol, the scheme of [9] selects a relay based on the strongest link as well while giving preference for the R-D links. Compared with [8], this led to a smaller asymptotic APD value of 2 which was realized at the expense of a reduction in the diversity order. A similar relay selection approach was presented in [10] attempting to equalize the buffer lengths at the relays. While the delay was improved compared to the *max-link* scheme, the diversity order is equal to K as in buffer-free (BF) systems. A channel state information (CSI) based relay selection protocol was proposed in [11] where the priority was given to the S-R and R-D hops in odd and even time slots, respectively. For finite buffer sizes, the scheme in [11] slightly improves the diversity order compared to the *max-link* protocol.

Unlike [8]–[11] where the relay selection policy is based solely on the CSI, the schemes in [12]–[17] include the buffer state in the selection process. A balancing BA scheme was analyzed in [12] targeting to keep the number of packets at each buffer the closest possible to $L/2$ in symmetrical networks. A priority-based *max-link* scheme was proposed in [13] where three classes of priority were considered; namely relays with full, empty and neither full nor empty buffers. The diversity order was also proven to be equal to $2K$ for large values of L in the case of quasi-symmetrical networks. The scheme proposed in [14] classifies the relays as in the transmission mode or the reception mode based on their actual queue length. After this classification, the protocol generates the decision on whether to transmit or receive based on the maximum and

The authors are with the Department of Electrical and Computer Engineering of the Lebanese American University (LAU), Byblos, Lebanon. (e-mails: sawsan.elzahr@lau.edu and chadi.abourjeily@lau.edu.lb).

minimum number of stored packets for the transmission and reception modes, respectively. For symmetrical networks, the scheme in [14] achieves a diversity order of $2K$ along with an improved asymptotic APD of $2K+2$ with finite buffer sizes. A maximum-weight selection protocol was proposed in [15] where a weight is assigned to each link while differentiating between S-R and R-D links. Similar to [14], the scheme in [15] achieves a full diversity order of $2K$ in symmetrical networks for all buffer sizes exceeding two. In order to better tackle the situation where multiple links have the same weight, a relay selection factor was introduced in [16] including the weight of the link as the first metric and the link quality as the second metric. This resulted in two schemes prioritizing either the OP or the APD. Two delay-aware relay selection policies were proposed in [17] based on the availability of the links and buffer sizes. While one policy achieves an asymptotic APD of $4K-1$, the second policy reduces the delay to $2K+1$ at the expense of reducing the diversity order.

While the BA relaying schemes in [8]–[17] are deterministic, probabilistic relay selection was considered in [18], [19]. In [18], after selecting the strongest available S-R link and the strongest available R-D link, the system randomly chooses one of the two links. For quasi-symmetrical networks, this probabilistic scheme achieves the full diversity order of $2K$ with infinite buffer sizes and allows to achieve different levels of tradeoff between OP and APD. In [19], the protocol first selects the S-R and R-D links with the smallest and largest numbers of packets in the corresponding buffers, respectively. Then, a random selection is made among these links according to a probability distribution that takes into consideration the delay constraints.

The main contributions of this work are as follows:

- Proposing a novel threshold based relay selection strategy that combines the advantages of both deterministic relaying and probabilistic relaying with an arbitrary number of relays.
- Introducing the innovative idea of threshold based relaying. In this context, the relay selection decision is based on the relative values of the buffer sizes with respect to K reference threshold levels and not on the implicit values of these buffer sizes as in [8]–[19].
- Analyzing the performance of the proposed scheme for any network setup and deriving closed-form expressions of the OP, APD and diversity order for large values of the signal-to-noise ratio (SNR).
- Proposing adequate choices of the threshold levels for controlling the diversity order and asymptotic APD in an efficient manner.

The proposed strategy constitutes the first known scheme that is capable of achieving different levels of tradeoff between OP and APD while avoiding any uncontrolled randomness in the relay selection process. Unlike the deterministic relay selection schemes in [8]–[17] that can each achieve only one pair of diversity order and APD values, the proposed scheme can be adjusted to achieve $K+1$ such pairs of values. Unlike the probabilistic schemes in [18], [19], the proposed scheme can be fully controlled and geared towards improving the performance of the system in a completely predictable manner. We prove that the proposed relaying scheme is capable of achieving a diversity order of $K+N$ and an asymptotic APD of $2N+2$ where the integer N depends on the K threshold levels and can assume all values between 0 and K. The main challenge in implementing the proposed BA relaying scheme resides in acquiring the CSI and the states of the buffers. In fact, the relaying protocol determines the relay that needs to receive or transmit based on the availability of all S-R and R-D links as well as the numbers of packets stored in the relays' buffers. As such, a central node needs to coordinate the cooperation efforts in the network by gathering the above information, deciding about the relay that must be activated (whether in reception or transmission modes) and then sharing this decision will all relays. This role can be played by any node in the network; in particular, by S or D for example.

The proposed BA solution is practical for 5G networks where user coordination in dense environments is pivotal for meeting the capacity demands. In fact, 5G endorsed a methodical shift from base-station centric to user centric architectures where users are expected to participate in storage, relaying, content delivery and computation within the network [1]. In particular, relay selection plays a key role in device-to-device (D2D) communications that have been recently proposed to increase the spectrum efficiency and network coverage. Moreover, as the node density increases in Internet-of-Things (IoT) networks, various devices can act as relays to forward traffic from the end-nodes to the core network and vice versa. On the other hand, 5G systems are conceived as highly flexible infrastructures that provide enhanced performance in terms of latency, reliability and throughput while meeting diverse requirements from multiple services. In this context, the adjustability of the proposed relaying strategy renders it suitable for supporting multi-services since it can handle the interplay between delay and diversity. For applications like video streaming, web browsing and file sharing, the threshold levels can be adjusted to achieve maximum reliability while sacrificing the latency. For applications like gaming, robotics automation and industrial private networks, the threshold levels can be adjusted to achieve minimum delay rendering the proposed scheme suitable for such applications as long as the predetermined amount of delay can be tolerated.

II. System Model and Relaying Strategy

A. Basic Parameters

Consider a cooperative network comprising $K+2$ nodes including a source node S, a destination node D and K relay nodes denoted by $\mathrm{R}_1, \ldots, \mathrm{R}_K$. It is assumed that there is no direct connection between S and D and, hence, S communicates with D only through the K neighboring relays. We assume that the nodes are equipped with a single antenna each. We also assume that all nodes are half-duplex and, hence, cannot transmit and receive simultaneously.

Denote by h_k and h'_k the channel coefficients of the S-R_k and R_k-D links, respectively, for $k = 1, \ldots, K$. A Rayleigh block fading channel model is assumed where the channel coefficients are assumed to be circularly symmetric complex

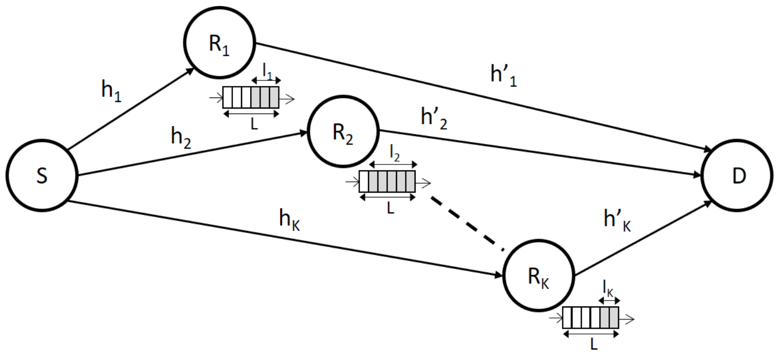

Fig. 1. BA cooperative network with K relays. The relays are numbered in an ascending order of their distances from S.

Gaussian distributed random variables with zero mean and average channel gains Ω_k and Ω'_k for the S-R$_k$ and R$_k$-D links, respectively. We assume that the relays are numbered in an increasing order according to their distances from S with R$_1$ (resp. R$_K$) being the closest (resp. farthest) relay from S resulting in: $\Omega_1 \geq \cdots \geq \Omega_K$. Finally, all S-R and R-D links are corrupted by an additive white Gaussian noise (AWGN) with zero mean and unit variance. The system model is better depicted in Fig. 1.

A communication link is in outage if the corresponding channel capacity falls below the target rate r_0 (in bits per channel use (BPCU)). As such, the outage probabilities along the S-R$_k$ and R$_k$-D links are given by:

$$\mathfrak{p}_k = \Pr\left\{\frac{1}{2}\log_2(1+\bar{\gamma}|h_k|^2) \leq r_0\right\} = 1 - e^{-\frac{2^{2r_0}-1}{\Omega_k \bar{\gamma}}} \quad (1)$$

$$\mathfrak{q}_k = \Pr\left\{\frac{1}{2}\log_2(1+\bar{\gamma}|h'_k|^2) \leq r_0\right\} = 1 - e^{-\frac{2^{2r_0}-1}{\Omega'_k \bar{\gamma}}}, \quad (2)$$

where $\bar{\gamma}$ stands for the average transmit signal-to-noise ratio (SNR).

We assume that the relays are equipped with buffers (data queues), of finite size L, in which the information packets can be temporarily stored so that they can be retransmitted when the channel conditions are more favorable. We denote by $l_k \in \{0,\ldots,L\}$ the number of packets stored in the buffer B$_k$ at R$_k$ for $k = 1,\ldots,K$. While the same buffer size is assumed for all relays, it is worth noting that the parameter L is a variable. As will be highlighted in Sections IV and V, one of the objectives of this work is to suggest adequate values for L. In particular, we prove that there is no need to increase L beyond 3.

We denote by p_k and q_k the unavailability probabilities along the links S-R$_k$ and R$_k$-D, respectively. The link S-R$_k$ is unavailable if either the channel between S and R$_k$ is in outage (with probability $\mathfrak{p}_k$) or B$_k$ is full (since the incoming packet cannot be accommodated). Consequently:

$$p_k(l_k) = \mathfrak{p}_k + \delta_{l_k=L} - \mathfrak{p}_k\delta_{l_k=L}, \quad (3)$$

where $\delta_{\mathcal{S}} = 1$ if the statement S is true while $\delta_{\mathcal{S}} = 0$ otherwise.

Similarly, the link R$_k$-D is unavailable with the following probability:

$$q_k(l_k) = \mathfrak{q}_k + \delta_{l_k=0} - \mathfrak{q}_k\delta_{l_k=0}, \quad (4)$$

since no packet can be transmitted from R$_k$ to D if the buffer B$_k$ is empty.

B. Threshold-Based Relaying Strategy

The relaying strategy is parameterized by K threshold values $\{l_{\text{th},k}\}_{k=1}^{K}$ where the threshold level $l_{\text{th},k}$ determines the operation mode of R$_k$ for $k = 1,\ldots,K$. If $l_k > l_{\text{th},k}$, relay R$_k$ is deemed to have a *large enough* number of packets and, intuitively, transmission (Tx) is given preference over reception (Rx) in an attempt to decrease the number of stored packets at B$_k$ and, hence, avoid the congestion of this buffer. On the other hand, if $l_k \leq l_{\text{th},k}$, the number of packets stored in B$_k$ is judged to be small and, hence, R$_k$ enters the Rx mode since it has *enough room* to accommodate for the incoming packet. We define the parameter Δ_k as follows:

$$\Delta_k \triangleq l_k - l_{\text{th},k} \quad ; \quad k = 1,\ldots,K. \quad (5)$$

Following from (5), the operation mode of each relay can be determined as follows:

$$\begin{cases} \Delta_k > 0 : & \text{R}_k \text{ in Tx mode} \\ \Delta_k \leq 0 : & \text{R}_k \text{ in Rx mode} \end{cases}. \quad (6)$$

Evidently, the threshold levels satisfy the following relation:

$$l_{\text{th},k} \in \{0,\ldots,L-1\} \quad ; \quad k = 1,\ldots,K, \quad (7)$$

since $l_{\text{th},k} = L$ implies that $\Delta_k \leq 0$ and, hence, R$_k$ can never be in the Tx mode.

We define six subsets of the K relays as follows. (i): $\mathcal{T}_c = \{k \mid \text{R}_k\text{-D not in outage}\}$, the set of relays for which the R$_k$-D link is not in outage. (ii): $\mathcal{T}_b = \{k \mid \Delta_k > 0\}$, the set of relays in the Tx mode. (iii): $\mathcal{T}'_b = \{k \mid l_k \neq 0\}$, the set of relays that have packets to transmit to D. (iv): $\mathcal{R}_c = \{k \mid \text{S-R}_k \text{ not in outage}\}$, the set of relays for which the S-R$_k$ link is not in outage. (v): $\mathcal{R}_b = \{k \mid \Delta_k \leq 0\}$, the set of relays in the Rx mode. (vi): $\mathcal{R}'_b = \{k \mid l_k \neq L\}$, the set of relays that have enough space to store a packet from S.

The relaying strategy consists of choosing one relay in each time slot to either receive from S or transmit to D in order to avoid interference. A plausible relaying strategy must put a preference on the transmission to D since the excessive accumulation of the packets at a buffer with an infrequent liberation of these packets will result in excessive and unjustified delays. Based on this observation, the proposed relaying scheme can be implemented based on the following steps for selecting the relay $R_{\hat{k}}$ to transmit or receive:

1) Choose among the relays in Tx mode with available R-D link the relay with the highest Δ_k to transmit: $\hat{k} = \arg\max_{k\in\mathcal{T}_c\cap\mathcal{T}_b}\{\Delta_k\}$.
2) If step 1 returns no available relay ($\mathcal{T}_c \cap \mathcal{T}_b = \phi$), choose among the relays in Rx mode with available S-R link the relay with the smallest Δ_k to receive: $\hat{k} = \arg\min_{k\in\mathcal{R}_c\cap\mathcal{R}_b}\{\Delta_k\}$.
3) While step 1 and step 2 ensure that each relay is operating in its corresponding operation Tx or Rx mode, these steps might yield no available relay. Therefore, if step 2 returns no available relay ($\mathcal{R}_c\cap\mathcal{R}_b = \phi$), relay $R_{\hat{k}}$ is selected for transmission among all relays with non-empty buffers (and not necessarily in Tx mode) with available R-D link: $\hat{k} = \arg\max_{k\in\mathcal{T}_c\cap\mathcal{T}'_b}\{\Delta_k\}$.
4) If step 3 returns no available relay ($\mathcal{T}_c \cap \mathcal{T}'_b = \phi$), the selection of a relay to receive is broadened from the relays in Rx mode to relays with non-full buffers: $\hat{k} = \arg\min_{k\in\mathcal{R}_c\cap\mathcal{R}'_b}\{\Delta_k\}$.

If step 4 returns no available relay, then all S-R and R-D links are not available and the system will be in outage. The relay selection protocol is better described by algorithm 1.

```
Data: T_c, T_b, T'_b, R_c, R_b, R'_b and {Δ_1, ..., Δ_K};
Result: A relay to Tx or Rx;
initialization: No relay is selected;
for n = 1 : 2 do
    if n = 2 then
        Let T_b = T'_b and R_b = R'_b
    end
    if T_c ∩ T_b ≠ φ then
        R_k̂ Tx with k̂ = arg max_{k∈T_c∩T_b} {Δ_k}
        break
    end
    if R_c ∩ R_b ≠ φ then
        R_k̂ Rx with k̂ = arg min_{k∈R_c∩R_b} {Δ_k}
        break
    end
end
```

Algorithm 1: Threshold Based Relay Selection Protocol

Finally, if the comparisons in step 1 or step 3 result in a tie, the smallest value of $\hat{k}$ is selected; i.e. the closest relay to S is chosen to transmit based on the adopted numbering of the relays. In fact, relays that are closer to S have a high rate of arrival of packets because of the good quality of the S-R link and, hence, it is preferable to release a packet from these congested relays. Similarly, if the comparisons in step 2 or step 4 result in a tie, the largest value of $\hat{k}$ is selected. In this case, it is preferred to accommodate the received packet in the buffer that is the least congested (on average) and that corresponds to the relay that is the farthest from S.

As an illustrative example, denote the subsets of the relays in the Tx and Rx modes by $\mathcal{S}_t$ and $\mathcal{S}_r$, respectively. Following from algorithm 1, the priority order of the proposed strategy is as follows. (i): The highest priority is for a relay in $\mathcal{S}_t$ to transmit. (ii): The second priority is for a relay in $\mathcal{S}_r$ to receive. (iii): The third priority is for a relay in $\mathcal{S}_r$ to transmit. (iv): The least priority is for a relay in $\mathcal{S}_t$ to receive. For (i) and (iii), the relay with the largest value of Δ_k with an available R-D link is selected while, for (ii) and (iv), the relay with the smallest value of Δ_k with an available S-R link is selected. From (6), $\mathcal{S}_t$ and $\mathcal{S}_r$ are determined from $\{\Delta_k\}_{k=1}^K$. For example, if $\{\Delta_k\}_{k=1}^K = \{+1, -1, +2, +3, 0\}$, then $\mathcal{S}_t = \{1, 3, 4\}$ while $\mathcal{S}_r = \{2, 5\}$.

III. Generalities and Basic Parameters

In this paper, a Markov Chain (MC) analysis is adopted to evaluate the steady state distribution, the outage probability (OP) and the average packet delay (APD) of the proposed scheme. A state of the MC represents the combination of the number of packets in the buffer of each relay and is defined as $(l_1, \dots, l_K)$. Since $l_k \in \{0, \dots, L\}$ for $k = 1, \dots, K$, then the total number of states involved is $(L+1)^K$.

We denote by $t_{(l_1,\dots,l_K),(l'_1,\dots,l'_K)}$ the transition probability of going from the state $(l_1, \dots, l_K)$ to the state $(l'_1, \dots, l'_K)$. The transition probabilities are stacked in the $(L+1)^K \times (L+1)^K$ state transition matrix $\mathbf{T}$ whose (i, j)-th element is given by:

$$\mathbf{T}_{i,j} = t_{(l_1,\dots,l_K),(l'_1,\dots,l'_K)} \quad ; \quad i = \mathfrak{N}(l'_1, \dots, l'_K) \ , \ j = \mathfrak{N}(l_1, \dots, l_K), \tag{8}$$

where the function $j = \mathfrak{N}(l_1, \dots, l_K) = 1 + \sum_{k=1}^K l_k(L+1)^{K-k}$ defines a one-to-one relation between the integer $j \in \{1, \dots, (L+1)^K\}$ and the state $(l_1, \dots, l_K) \in \{0, \dots, L\}^K$.

We denote by $\pi_{l_1,\dots,l_K}$ the steady-state probability of being in the state $(l_1, \dots, l_K)$. The steady-state probabilities can be determined from [8]:

$$\pi = (\mathbf{T} - \mathbf{I} + \mathbf{B})^{-1}\,\mathbf{b}, \tag{9}$$

where the j-th element of the $(L+1)^K$-dimensional vector π is equal to $\pi_{l_1,\dots,l_K}$ with $j = \mathfrak{N}(l_1, \dots, l_K)$. In (9), $\mathbf{I}$ and $\mathbf{B}$ are the $(L+1)^K \times (L+1)^K$ matrices denoting the identity matrix and the all-one matrix, respectively. $\mathbf{b}$ is the vector whose elements are all equal to 1.

The cooperative network is said to be in outage when no packets can be communicated along any of its $2K$ S-R or R-D constituent links. When the MC is in the state $(l_1, \dots, l_K)$, all S-R and R-D links will be unavailable with the probability $\prod_{k=1}^K p_k(l_k)q_k(l_k)$ following from (3) and (4). Consequently, the outage probability can be determined from:

$$OP = \sum_{l_1=0}^{L} \cdots \sum_{l_K=0}^{L} \pi_{l_1,\dots,l_K} \prod_{k=1}^{K} p_k(l_k)q_k(l_k). \tag{10}$$

Following from the storage of the packets in the relays' buffers, these packets will reach D with a certain queuing delay. Following from [14] and Little's law [24], the average packet delay can be determined from:

$$APD = 1 + \frac{2\bar{L}}{1 - OP}, \tag{11}$$

where $\bar{L}$ denotes the average queue length that can be determined from:

$$\bar{L} = \sum_{l_1=0}^{L} \cdots \sum_{l_K=0}^{L} \pi_{l_1,\dots,l_K} \left[\sum_{k=1}^{K} l_k\right]. \tag{12}$$

Determining the steady-state probabilities from (9) constitutes the major hindrance behind evaluating the performance of K-relay BA systems for any SNR value. In fact, the state transition matrix $\mathbf{T}$ is a $(L+1)^K \times (L+1)^K$ matrix while the steady-state vector π comprises $(L+1)^K$ elements. As such, deriving the exact OP and APD expressions in closed-form for any SNR value is intractable since it is impractical to solve for the very large number of elements of π especially for large values of K. As such, [8]–[19] related the OP and APD to $\mathbf{T}$ and $\pi = (\mathbf{T}-\mathbf{I}+\mathbf{B})^{-1}\mathbf{b}$ without explicitly solving the last matrix equation and, hence, without reaching closed-form expressions for all SNR values. In an attempt to circumvent this challenge, [8]–[11], [13]–[15], [17] provided a complementary asymptotic analysis that yielded intuitive expressions of the OP and APD for large values of the SNR. Therefore, even though the large SNR assumption might not hold in practice, we carry out an asymptotic analysis in Sections IV-B and V-B as in [8]–[11], [13]–[15], [17]. The purpose of this analysis is to derive closed-form expressions of the OP and APD that offer clear and intuitive insights on the system performance. Moreover, the asymptotic analysis is indispensable for deriving the diversity order that constitutes a major performance metric that captures the performance of fading mitigation techniques whose performance gains increase with the SNR. As will be highlighted in Section VI, the derived asymptotic expressions yield accurate results not only for large SNRs but also in the mid-SNR range.

IV. Performance Analysis: 2 Relays

A. Transition Probabilities

For simplicity, in this section the unavailability probabilities $p_k(l_k)$ and $q_k(l_k)$ will be expressed as p_k and q_k, respectively.

The self transition of going from the state (l_1, l_2) to this same state occurs when all S-R and R-D links are unavailable:

$$t_{(l_1,l_2),(l_1,l_2)} = p_1 p_2 q_1 q_2. \quad (13)$$

The remaining transitions depend of the relative values that l_1 and l_2 assume with respect to $l_{\text{th},1}$ and $l_{\text{th},2}$, respectively, resulting in the four following cases.

1) Case 1: $l_1 \leq l_{\text{th},1}$ and $l_2 \leq l_{\text{th},2}$: R_1 and R_2 are in the Rx mode. In this case, four possible transitions can take place as follows. (i): $(l_1, l_2) \to (l_1+1, l_2)$. In this case, the number of packets stored in B_1 will increase by one only if the S-R_1 link is available (with probability $1-p_1$). In this context, R_1 is selected for reception (rather than R_2) if either the S-R_2 link is unavailable or if this link is available with $\Delta_1 < \Delta_2$ since the relay with smaller value of Δ_k is selected for Rx following from step 2 of the proposed relaying protocol. Therefore:

$$t_{(l_1,l_2),(l_1+1,l_2)} = (1-p_1)\left[p_2 + (1-p_2)\delta_{\Delta_1<\Delta_2}\right]. \quad (14)$$

(ii): $(l_1, l_2) \to (l_1, l_2+1)$. Similar to (14), this transition occurs if R_2 is selected for reception:

$$t_{(l_1,l_2),(l_1,l_2+1)} = (1-p_2)[p_1 + (1-p_1)\delta_{\Delta_1 \geq \Delta_2}], \quad (15)$$

since, for $\Delta_1 = \Delta_2$, the preference is to send the packet to the relay that is farther from S which is the relay R_2 following from the adopted numbering convention.

(iii): $(l_1, l_2) \to (l_1-1, l_2)$. Knowing that R_1 and R_2 are in the Rx mode, then in order to have R_1 or R_2 transmitting a packet, both links S-R_1 and S-R_2 should be unavailable. In this case, a relay will be selected for transmission based on the R_1-D and R_2-D linkss status and the values of Δ_1 and Δ_2. R_1 is chosen to transmit a packet if either the link R_1-D is available and the link R_2-D is unavailable or both links R_1-D and R_2-D are available with $\Delta_1 \geq \Delta_2$:

$$t_{(l_1,l_2),(l_1-1,l_2)} = p_1 p_2 (1-q_1)[q_2 + (1-q_2)\delta_{\Delta_1 \geq \Delta_2}], \quad (16)$$

where, for $\Delta_1 = \Delta_2$, the preference is to send from the relay that is closer to S which is the relay R_1.

(iv): $(l_1, l_2) \to (l_1, l_2-1)$. As in (16), R_2 is chosen to transmit a packet with the following probability:

$$t_{(l_1,l_2),(l_1,l_2-1)} = p_1 p_2 (1-q_2)[q_1 + (1-q_1)\delta_{\Delta_1 < \Delta_2}], \quad (17)$$

following from step 3 of the proposed relaying scheme.

Finally, it can be easily proven that the probabilities in (13)-(17) add up to one.

2) Case 2: $l_1 \leq l_{\text{th},1}$ and $l_2 > l_{\text{th},2}$: R_1 in Rx mode and R_2 in Tx mode. In this case, the transitions (excluding the self transition) and their corresponding probabilities are given by:

$$(l_1, l_2) \to \begin{cases} (l_1, l_2-1), & 1-q_2; \\ (l_1+1, l_2), & q_2(1-p_1); \\ (l_1-1, l_2), & q_2 p_1 (1-q_1); \\ (l_1, l_2+1), & q_1 p_1 q_2 (1-p_2). \end{cases} \quad (18)$$

(i): $(l_1, l_2) \to (l_1, l_2-1)$. Since R_2 is the only relay in the Tx mode, then R_2 is chosen to transmit whenever the link R_2-D is available as this scheme puts preference on transmission following from step 1 of the proposed scheme. (ii): $(l_1, l_2) \to (l_1+1, l_2)$. R_1 is chosen to receive if the link R_2-D is unavailable (step 1 did not yield an available relay for Tx) and the link S-R_1 is available (step 2 of the relaying strategy). (iii): $(l_1, l_2) \to (l_1-1, l_2)$. If both relays are unable to operate in their corresponding operation mode (both R_2-D and S-R_1 links are unavailable), according to step 3, R_1 will be chosen to transmit if the link R_1-D is available. (iv): $(l_1, l_2) \to (l_1, l_2+1)$. R_2 is chosen to receive if the links R_2-D, S-R_1 and R_1-D are unavailable and the link S-R_2 is available following from step 4.

3) Case 3: $l_1 > l_{\text{th},1}$ and $l_2 \leq l_{\text{th},2}$: R_1 in Tx mode and R_2 in Rx mode. The analysis in this case is similar to that of case 2. Exchanging the roles of R_1 and R_2 in (18) results in the following transition probabilities:

$$(l_1, l_2) \to \begin{cases} (l_1-1, l_2), & 1-q_1; \\ (l_1, l_2+1), & q_1(1-p_2); \\ (l_1, l_2-1), & q_1 p_2 (1-q_2); \\ (l_1+1, l_2), & q_1 p_2 q_2 (1-p_1). \end{cases} \quad (19)$$

4) Case 4: $l_1 > l_{\text{th},1}$ and $l_2 > l_{\text{th},2}$: R_1 and R_2 are both in the Tx mode. If the relays operate in their designated Tx mode (step 1), the corresponding transition probabilities are given by:

$$(l_1, l_2) \to \begin{cases} (l_1-1, l_2), & (1-q_1)[q_2 + (1-q_2)\delta_{\Delta_1 \geq \Delta_2}]; \\ (l_1, l_2-1), & (1-q_2)[q_1 + (1-q_1)\delta_{\Delta_1 < \Delta_2}]. \end{cases}, \quad (20)$$

where a necessary condition for R_k to be selected for transmission is the availability of the link R_k-D. Regarding the remaining R-D link, it can be either unavailable or available with $\Delta_1 \geq \Delta_2$ (for R_1 to be selected) and $\Delta_1 < \Delta_2$ (for R_2 to be selected).

On the other hand, a necessary condition for the relays to switch to reception is the unavailability of the R-D links. In this case, following from step 4, the corresponding transition probabilities are given by:

$$(l_1, l_2) \rightarrow \begin{cases} (l_1+1, l_2), & q_1 q_2 (1-p_1)[p_2 + (1-p_2)\delta_{\Delta_1 < \Delta_2}]; \\ (l_1, l_2+1), & q_1 q_2 (1-p_2)[p_1 + (1-p_1)\delta_{\Delta_1 \geq \Delta_2}]. \end{cases} \quad (21)$$

B. Asymptotic Analysis

Replacing the transition probabilities in (8) and inverting the matrix $\mathbf{T} - \mathbf{I} + \mathbf{B}$ in (9) will yield intractable results especially for large values of the buffer size L. Therefore, we resort to an asymptotic analysis that is useful for offering clear insights on the system performance and for reaching tangible conclusions regarding the selection of the threshold levels $l_{\text{th},1}$ and $l_{\text{th},2}$. The asymptotic analysis revolves around the following proposition:

Proposition 1: For asymptotically large values of the SNR, the following set $\mathcal{S}$ forms a closed subset of the states:

$$\mathcal{S} = \{(l_{\text{th},1}, l_{\text{th},2}), (l_{\text{th},1}, l_{\text{th},2}+1), (l_{\text{th},1}+1, l_{\text{th},2}), (l_{\text{th},1}+1, l_{\text{th},2}+1)\}, \quad (22)$$

where $\pi_{l_1,l_2} \rightarrow 0$ for $(l_1, l_2) \notin \mathcal{S}$ while the steady-state probabilities of the states in $\mathcal{S}$ tend to the following asymptotic values:

$$\begin{cases} \pi_{l_{\text{th},1}, l_{\text{th},2}} = \frac{1-\mathfrak{q}_2}{2} \\ \pi_{l_{\text{th},1}, l_{\text{th},2}+1} = \frac{1-\mathfrak{p}_2}{2} \end{cases} ; \quad \begin{cases} \pi_{l_{\text{th},1}+1, l_{\text{th},2}} = \frac{\mathfrak{p}_2}{2} \\ \pi_{l_{\text{th},1}+1, l_{\text{th},2}+1} = \frac{\mathfrak{q}_2}{2} \end{cases} . \quad (23)$$

Proof: The proof revolves around ignoring the product of two or more terms in $\{\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{q}_1, \mathfrak{q}_2\}$. The detailed proof is provided in Appendix A. ∎

Following from (23), the OP in (10) assumes the following asymptotic expression:

$$OP_{\text{Asym}} = \sum_{l_1=l_{\text{th},1}}^{l_{\text{th},1}+1} \sum_{l_2=l_{\text{th},2}}^{l_{\text{th},2}+1} \pi_{l_1,l_2} p_1(l_1) p_2(l_2) q_1(l_1) q_2(l_2). \quad (24)$$

Consequently, the choice of the threshold levels $l_{\text{th},1}$ and $l_{\text{th},2}$ will affect the value of OP as well as the diversity order following from the dependence of the unavailability probabilities $\{p_k(l_{\text{th},k}), p_k(l_{\text{th},k}+1), q_k(l_{\text{th},k}), q_k(l_{\text{th},k}+1)\}_{k=1}^{2}$ in (3)-(4) on the threshold levels. Following from (3) and (4):

$$(p_k, q_k) = \begin{cases} (\mathfrak{p}_k, 1), & l_k = 0; \\ (\mathfrak{p}_k, \mathfrak{q}_k), & l_k = 1, \ldots, L-1; \\ (1, \mathfrak{q}_k), & l_k = L. \end{cases} . \quad (25)$$

As such, the asymptotic OP and diversity order depend on whether $l_k = 0$, $l_k \in \{1, \ldots, L-2\}$ or $l_k = L-1$ for $k = 1, 2$ resulting in the nine possible values of OP_{Asym} summarized in Table I. In Table I, OP_{Asym} was approximated by the summation of terms comprising the smallest number of multiplicands among $\{\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{q}_1, \mathfrak{q}_2\}$; i.e., $\sum_n \alpha_n \prod_{k=1}^{2} \mathfrak{p}_k^{i_{k,n}} \mathfrak{q}_k^{j_{k,n}} \approx \sum_{n=\hat{n}} \alpha_n \prod_{k=1}^{2} \mathfrak{p}_k^{i_{k,n}} \mathfrak{q}_k^{j_{k,n}}$ where α_n is a constant and $\hat{n} = \arg\min_n\{d_n\} \triangleq \arg\min_n\{i_{1,n} + i_{2,n} + j_{1,n} + j_{2,n}\}$ resulting in a diversity order of $d = \hat{n}$. The diversity order is defined as the negative slope of the $OP(\bar{\gamma})$ curve on a log-log scale where the product of n terms among $\{\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{q}_1, \mathfrak{q}_2\}$ scales asymptotically as $\bar{\gamma}^{-n}$ (since each outage probability in (1) scales as $\bar{\gamma}^{-1}$) resulting in a diversity order of n.

Table I holds for all $L \geq 2$ where the asymptotic OP values in the case $L = 2$ can be obtained by removing the second row and second column of Table I since $\{1, \ldots, L-2\} = \phi$ in this case. As such, the following conclusions can be reached:

- The threshold levels have a direct impact on the achievable diversity order that varies from 2 to 3 for $L = 2$ and from 2 to 4 for $L > 2$.
- Among the L^2 possible values of $(l_{\text{th},1}, l_{\text{th},2})$ satisfying (7), the value $(l_{\text{th},1}, l_{\text{th},2}) = (0, 0)$ results in the smallest diversity order of 2 which is the same as the diversity order achieved by 2-relay buffer-free systems.
- For $L = 2$, the maximum achievable diversity order is 3. This value can be attained for $(l_{\text{th},1}, l_{\text{th},2}) = (1, 0)$ or $(l_{\text{th},1}, l_{\text{th},2}) = (1, 1)$.
- The proposed BA relaying scheme is capable of achieving the maximum diversity order of 4 for $(l_{\text{th},1}, l_{\text{th},2}) \in \{1, \ldots, L-1\} \times \{1, \ldots, L-2\}$ while confining the threshold levels within $\{1, \ldots, L-2\} \times \{1, \ldots, L-2\}$ results in the smallest asymptotic OP value of $\mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_1\mathfrak{q}_2$ and, hence, in the largest coding gain.

Replacing (23) in (12) shows that the average queue length assumes the following asymptotic expression:

$$\begin{aligned} \bar{L}_{\text{Asym}} &= \pi_{l_{\text{th},1}, l_{\text{th},2}}(l_{\text{th},1} + l_{\text{th},2}) + \pi_{l_{\text{th},1}, l_{\text{th},2}+1}(l_{\text{th},1} + l_{\text{th},2} + 1) \\ &\quad + \pi_{l_{\text{th},1}+1, l_{\text{th},2}}(l_{\text{th},1} + l_{\text{th},2} + 1) \\ &\quad + \pi_{l_{\text{th},1}+1, l_{\text{th},2}+1}(l_{\text{th},1} + l_{\text{th},2} + 2) \\ &= (l_{\text{th},1} + l_{\text{th},2}) + 0.5 + \mathfrak{q}_2. \end{aligned} \quad (26)$$

Replacing (26) in (11) results in:

$$\begin{aligned} APD_{\text{Asym}} &= 1 + \frac{2(l_{\text{th},1} + l_{\text{th},2}) + 1 + 2\mathfrak{q}_2}{1 - OP_{\text{Asym}}} \\ &\approx 2(1 + l_{\text{th},1} + l_{\text{th},2}), \end{aligned} \quad (27)$$

since $OP_{\text{Asym}} \ll 1$ and $\mathfrak{q}_2 \rightarrow 0$ for asymptotically large values of the SNR.

Equation (27) shows that increasing the threshold levels will result in an inconvenient increase in the APD. In this context, the smallest APD value of 2 can be achieved by fixing $l_{\text{th},1} = l_{\text{th},2} = 0$.

C. Conclusions regarding the System Design

Following from Table I and (27), the following conclusions can be reached regarding the selection of the threshold levels $l_{\text{th},1}$ and $l_{\text{th},2}$ as well as the selection of the buffer size L.

Regarding the buffer size, the OP values in Table I show that there is no interest in increasing L beyond 3 in the asymptotic SNR regime. In fact, the smallest possible OP value

TABLE I
ASYMPTOTIC OUTAGE PROBABILITIES FOR $K=2$.

$l_{\text{th},1}$ \ $l_{\text{th},2}$	0	$\{1,\dots,L-2\}$	$L-1$
0	$\mathfrak{p}_1\mathfrak{p}_2/2$	$\mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_2$	$\mathfrak{p}_1\mathfrak{q}_2/2$
$\{1,\dots,L-2\}$	$\mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_1/2$	$\mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_1\mathfrak{q}_2$	$\mathfrak{p}_1\mathfrak{q}_1\mathfrak{q}_2/2$
$L-1$	$\mathfrak{p}_2\mathfrak{q}_1(\mathfrak{p}_1+\mathfrak{p}_2)/2$	$\mathfrak{p}_2\mathfrak{q}_1\mathfrak{q}_2(\mathfrak{p}_1+\frac{\mathfrak{q}_2}{2}+\frac{\mathfrak{p}_2}{2})$	$\mathfrak{q}_1\mathfrak{q}_2(\mathfrak{p}_1+\mathfrak{q}_2)/2$

of $\mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_1\mathfrak{q}_2$ can be achieved with $L=3$ and the choice $L>3$ does not reduce this minimum OP any further. On the other hand, the choice $L=2$ is feasible but it penalizes the diversity order rendering the full diversity order of 4 unachievable. As such, the buffer size $L=3$ is recommended when the SNR is large enough.

Regarding the threshold levels, Table I and (27) show that there is no interest in selecting values of $(l_{\text{th},1}, l_{\text{th},2})$ outside the set $\varphi=\{(0,0),(0,1),(1,0),(1,1)\}$. In fact, increasing $l_{\text{th},1}$ and/or $l_{\text{th},2}$ beyond the value of 1 will increase the APD following from (27) whereas this increase does not present any advantage in terms of the OP performance following from Table I. In fact, from Table I, increasing $l_{\text{th},1}$ and/or $l_{\text{th},2}$ beyond entails one of the following implications that are not beneficial. (i): decreasing the diversity order (for example, increasing $l_{\text{th},2}$ from $L-2$ to $L-1$), (ii): maintaining the same diversity order but increasing the OP (for example, increasing $l_{\text{th},1}$ from $L-2$ to $L-1$) or (iii): keeping the same diversity order and same OP (for example, increasing $l_{\text{th},1}$ or $l_{\text{th},2}$ from 1 to $L-2$). Among the choices within the set φ with $L\geq 3$:

- The choice $(l_{\text{th},1}, l_{\text{th},2})=(0,0)$ achieves the smallest asymptotic APD value of 2 at the expense of the smallest diversity order of 2.
- The choice $(l_{\text{th},1}, l_{\text{th},2})=(1,1)$ achieves the highest diversity order of 4 at the expense of the highest asymptotic APD value of 6.
- The choices $(l_{\text{th},1}, l_{\text{th},2})=(0,1)$ and $(l_{\text{th},1}, l_{\text{th},2})=(1,0)$ achieve a tradeoff between the APD and diversity order with an asymptotic APD value of 4 and a diversity order of 3. Compared to $(l_{\text{th},1}, l_{\text{th},2})=(0,0)$, these choices increase the diversity order from 2 to 3 while, compared to $(l_{\text{th},1}, l_{\text{th},2})=(1,1)$, these choices reduce the asymptotic APD from 6 to 4. From Table I, the choice $(0,1)$ is preferable over the choice $(1,0)$ if $\mathfrak{q}_2<\frac{\mathfrak{q}_1}{2}$.

V. PERFORMANCE ANALYSIS: K RELAYS

After considering the special case of two relays, we next consider the general case of an arbitrary number of relays. As will be highlighted later, the evaluation of the transition probabilities will entail implementing involved recursive functions while the asymptotic OP and APD expressions will be comparable to those obtained in Section IV.

A. Transition Probabilities

In what follows, $p_k(l_k)$ and $q_k(l_k)$ will be expressed as p_k and q_k, respectively, for simplicity. The state will be denoted by $\mathbf{l}=(l_1,\dots,l_K)$, the set of all relays will be denoted by $\mathcal{A}=\{1,\dots,K\}$ and the k-th row of the $K\times K$ identity matrix will be denoted by $\mathbf{e}_k$.

For any state of the MC, a self transition occurs when all links are unavailable:

$$t_{\mathbf{l},\mathbf{l}}=\prod_{k=1}^{K} p_k q_k. \tag{28}$$

At each time slot, the K relays will be classified according to their operation modes resulting in the three following cases.

1) Case 1: $\mathcal{R}_b=\mathcal{A}$ and $\mathcal{T}_b=\phi$ (all relays are in the Rx mode) resulting in the two following possibilities.

(i): A relay R_k will be selected for Rx if the link S-R_k is available and if Δ_k is the *smallest* among all $\{\Delta_i\}$ of the relays whose S-R links are available:

$$t_{\mathbf{l},\mathbf{l}+\mathbf{e}_k}=(1-p_k)\sum_{\mathcal{K}\subset\mathcal{A}\setminus\{k\}}\left[\prod_{i\in\mathcal{K}}(1-p_i)\right]\left[\prod_{j\in\mathcal{A}\setminus\{k\}\cup\mathcal{K}} p_j\right]P_{k,\mathcal{K}}, \tag{29}$$

where $P_{k,\mathcal{K}}$ denotes the probability that Δ_k is *smaller* than $\Delta_{k'}$ for all $k'\in\mathcal{K}$. The notion of *smaller* or *larger* Δ_i must take into consideration the tie breaking rule according to numbering the relays according to their distances from S. As such, $P_{k,\mathcal{K}}=\prod_{k'\in\mathcal{K}}P_{k,k'}$ where $P_{k,k'}$ denotes the probability that Δ_k is *smaller* than $\Delta_{k'}$:

$$P_{k,k'}=\delta_{k'<k}\delta_{\Delta_k\leq\Delta_{k'}}+\delta_{k'>k}\delta_{\Delta_k<\Delta_{k'}} \quad ; \quad k'\neq k, \tag{30}$$

since, for $\Delta_k=\Delta_{k'}$, the preference is to transmit to the relay that is farther from S; i.e, to the relay with higher index.

Therefore, (29) can be written as equation (31) on the top of the next page. This equation can be can be written in a more convenient form as shown in (32) on the top of the next page where this relation can be implemented recursively resulting in the following expression of the transition probability:

$$t_{\mathbf{l},\mathbf{l}+\mathbf{e}_k}=(1-p_k)\left[\prod_{i=1,i\neq k}^{K} p_i\right][1+f_r(\mathcal{A},k,0)], \tag{33}$$

where the function $f_r(\cdot,\cdot,\cdot)$ can be determined in a recursive manner according to algorithm 2.

(ii) A relay R_k (in the Rx mode) is selected to transmit if all S-R links are unavailable, the link R_k-D is available and Δ_k is the *largest* among all $\{\Delta_i\}$ of the relays whose R-D links are available:

$$t_{\mathbf{l},\mathbf{l}-\mathbf{e}_k}=\left[\prod_{j=1}^{K} p_j\right](1-q_k)\times \sum_{\mathcal{K}\subset\mathcal{A}\setminus\{k\}}\left[\prod_{i\in\mathcal{K}}(1-q_i)\right]\left[\prod_{j\in\mathcal{A}\setminus\{k\}\cup\mathcal{K}} q_j\right]Q_{k,\mathcal{K}}, \tag{34}$$

$$t_{\mathbf{1},\mathbf{1}+\mathbf{e}_k} = (1-p_k)\left[\prod_{i=1,i\neq k}^{K} p_i + \sum_{k_1=1,k_1\neq k}^{K}(1-p_{k_1})\left[\prod_{j=1,j\neq k,j\neq k_1}^{K} p_j\right]P_{k,k_1} + \sum_{k_1=1,k_1\neq k}^{K}\sum_{k_2=k_1+1,k_2\neq k}^{K}(1-p_{k_1})(1-p_{k_2})\left[\prod_{j=1,j\neq k,j\neq k_1,j\neq k_2}^{K} p_j\right]P_{k,k_1}P_{k,k_2}+\cdots\right]. \quad (31)$$

$$t_{\mathbf{1},\mathbf{1}+\mathbf{e}_k} = (1-p_k)\left[\prod_{i=1,i\neq k}^{K} p_i\right]\left[1+\sum_{k_1=1,k_1\neq k}^{K}\frac{(1-p_{k_1})}{p_{k_1}}P_{k,k_1}\left[1+\sum_{k_2=k_1+1,k_2\neq k}^{K}\frac{(1-p_{k_2})}{p_{k_2}}P_{k,k_2}\left[1+\dots\left[1+\sum_{k_K=k_{K-1}+1,k_K\neq k}^{K}\frac{(1-p_{k_K})}{p_{k_K}}P_{k,k_K}\right]\right]\right]\right], \quad (32)$$

Function: $f_m(\mathcal{X},k,c)$ for $m=$ "r" or $m=$ "t"
Data: $\mathcal{X}\subset\mathcal{A}$, $k\in\{1,\dots,K\}$ and $c\in\{0,\dots,K\}$;
Result: S;
initialization: $S=0$;
if $c+1>|\mathcal{X}|$ **then**
  return 0
end
for $n=c+1:|\mathcal{X}|$ **do**
  $k'=\mathcal{X}_n$ (n-th element of $\mathcal{X}$)
  if $k'\neq k$ **then**
    if $m=$ "r" **then**
      $S=S+\frac{(1-p_{k'})}{p_{k'}}P_{k,k'}[1+f_r(\mathcal{X},k,n)]$
    end
    if $m=$ "t" **then**
      $S=S+\frac{(1-q_{k'})}{q_{k'}}Q_{k,k'}[1+f_t(\mathcal{X},k,n)]$
    end
  end
end

Algorithm 2: Recursive functions $f_r(\mathcal{X},k,c)$ and $f_t(\mathcal{X},k,c)$

where $Q_{k,\mathcal{K}}$ denotes the probability that Δ_k is *larger* than $\Delta_{k'}$ for all $k'\in\mathcal{K}$. Considering the tie-breaking rule, this probability can be written as $Q_{k,\mathcal{K}}=\prod_{k'\in\mathcal{K}}Q_{k,k'}$ where

$$Q_{k,k'}=\delta_{k'<k}\delta_{\Delta_k>\Delta_{k'}}+\delta_{k'>k}\delta_{\Delta_k\geq\Delta_{k'}} \quad ; \quad k'\neq k, \quad (35)$$

since, for $\Delta_k=\Delta_{k'}$, the preference is to transmit from the relay that is closest to S; i.e, to the relay with lower index.

Given the similarity between (29) and (34), this latter equation can be written in terms of the recursive function $f_t(\cdot,\cdot,\cdot)$ (defined in algorithm 2) as follows:

$$t_{\mathbf{1},\mathbf{1}-\mathbf{e}_k} = \left[\prod_{j=1}^{K} p_j\right](1-q_k)\left[\prod_{i=1,i\neq k}^{K} q_i\right][1+f_t(\mathcal{A},k,0)]. \quad (36)$$

For $K=2$, it can be easily proven that (33) simplifies to the expressions provided in (14) and (15) for $k=1$ and $k=2$, respectively. Similarly, (16) and (17) follow from (36). While (14)-(17) explicitly relate the transition probabilities to the parameters $\{p_k,q_k,\Delta_k\}_{k=1}^2$ in the simple special case of $K=2$, the functions $f_r(\cdot)$ and $f_t(\cdot)$ in (33) and (36) need to be evaluated recursively in the general case of an arbitrary number of relays.

2) Case 2: $\mathcal{R}_b=\phi$ and $\mathcal{T}_b=\mathcal{A}$ (all relays are in the Tx mode). Similar to the analysis provided in the previous subsection, a relay R_k is chosen to transmit a packet with the following probability:

$$t_{\mathbf{1},\mathbf{1}-\mathbf{e}_k} = (1-q_k)\left[\prod_{i=1,i\neq k}^{K} q_i\right][1+f_t(\mathcal{A},k,0)], \quad (37)$$

while a relay R_k is chosen to receive a packet with the following probability:

$$t_{\mathbf{1},\mathbf{1}+\mathbf{e}_k} = \left[\prod_{j=1}^{K} q_j\right](1-p_k)\left[\prod_{i=1,i\neq k}^{K} p_i\right][1+f_r(\mathcal{A},k,0)], \quad (38)$$

where (37) and (38) can be obtained by interchanging the unavailability probabilities $p_k \leftrightharpoons q_k$ and the functions $f_t(\cdot,\cdot,\cdot)\leftrightharpoons f_r(\cdot,\cdot,\cdot)$ in (33) and (36), respectively. As in case 1, (37) and (38) simplify to (20) and (21), respectively, in the special case of $K=2$.

3) Case 3: $\mathcal{R}_b\neq\phi$ and $\mathcal{T}_b\neq\phi$. The four following cases need to be considered.

(i): Following from step 1 of the proposed relaying protocol, the highest priority is given for the transmission from the relay R_k with $k\in\mathcal{T}_b$ having the *largest* Δ_k. This transmission incurs the following transition probability:

$$t_{\mathbf{1},\mathbf{1}-\mathbf{e}_k} = (1-q_k)\left[\prod_{i\in\mathcal{T}_b,i\neq k} q_i\right][1+f_t(\mathcal{T}_b,k,0)] \ ; \ k\in\mathcal{T}_b, \quad (39)$$

in a way similar to (37).

(ii): The second priority is given to relay R_k with $k\in\mathcal{R}_b$ to receive. This occurs if all R_j-D links for $j\in\mathcal{T}_b$ are unavailable (step 1 did not yield a valid relay), the link S-R_k is available and Δ_k is the *smallest* among all $\{\Delta_i\}$ of the relays in $\mathcal{R}_b$ whose S-R links are available. Therefore, similar to (38):

$$t_{\mathbf{l},\mathbf{l}+\mathbf{e}_k} = \left[\prod_{j\in\mathcal{T}_b} q_j\right](1-p_k)\left[\prod_{i\in\mathcal{R}_b, i\neq k} p_i\right]\times \\ [1+f_r(\mathcal{R}_b,k,0)] \quad ; \quad k\in\mathcal{R}_b. \quad (40)$$

(iii): If none of the relays can operate in its corresponding Tx or Rx mode, the third priority is given for a relay R$_k$ not in the Tx mode ($k\notin\mathcal{T}_b$) to transmit. Since $\mathcal{T}_b\cup\mathcal{R}_b=\mathcal{A}$, then $k\in\mathcal{R}_b$. The corresponding transition probability is given by:

$$t_{\mathbf{l},\mathbf{l}-\mathbf{e}_k} = \left[\prod_{j\in\mathcal{T}_b} q_j\right]\left[\prod_{j'\in\mathcal{R}_b} p_{j'}\right](1-q_k)\left[\prod_{i\in\mathcal{R}_b, i\neq k}^{K} q_i\right]\times \\ [1+f_t(\mathcal{R}_b,k,0)] \quad ; \quad k\in\mathcal{R}_b, \quad (41)$$

since for step 3 to occur, (1): step 1 must yield an invalid relay (all R$_j$-D links are unavailable for $j\in\mathcal{T}_b$), (2): step 2 must yield an invalid relay (all S-R$_{j'}$ links are unavailable for $j'\in\mathcal{R}_b$), (3): the link R$_k$-D is available and (4): Δ_k is the *largest* among the relays in $\mathcal{R}_b$ whose R-D links are available.

(iv): Finally, a relay R$_k$ with $k\in\mathcal{T}_b$ is selected to receive if all R-D links are unavailable and all S-R$_j$ links for $j\in\mathcal{R}_b$ are unavailable. R$_k$ will be chosen based on the comparison of Δ_k with the corresponding values of the relays in $\mathcal{T}_b$:

$$t_{\mathbf{l},\mathbf{l}+\mathbf{e}_k} = \left[\prod_{j'=1}^{K} q_{j'}\right]\left[\prod_{j\in\mathcal{R}_b} p_j\right](1-p_k)\left[\prod_{i\in\mathcal{T}_b, i\neq k} p_i\right]\times \\ [1+f_r(\mathcal{T}_b,k,0)] \quad ; \quad k\in\mathcal{T}_b. \quad (42)$$

B. Asymptotic Analysis

The cumbersomeness of the transition probabilities in (33)-(42) and the large number of states ($(L+1)^K$) that grows exponentially with the number of relays motivate the need for an asymptotic analysis that sheds more light on the design parameters of the relaying system. The key observation behind this type of analysis is formulated in the following proposition.

Proposition *2:* For asymptotically large values of the SNR, the following closed subset of 4 states can be identified for all values of K

$$\mathcal{S} = \{(l_1,\dots,l_K)\mid l_k=l_{\text{th},k} \text{ for } k=1,\dots,K-2 \\ \text{and } l_k\in\{l_{\text{th},k}, l_{\text{th},k}+1\} \text{ for } k=K-1,K\}, \quad (43)$$

where $\pi_{l_1,\dots,l_K}\to 0$ for $(l_1,\dots,l_K)\notin\mathcal{S}$ and:

$$\begin{cases}\pi_{l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1},l_{\text{th},K}} = \frac{1-\mathfrak{q}_K}{2} \\ \pi_{l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1},l_{\text{th},K}+1} = \frac{1-\mathfrak{p}_K}{2}\end{cases}; \\ \begin{cases}\pi_{l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1}+1,l_{\text{th},K}} = \frac{\mathfrak{p}_K}{2} \\ \pi_{l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1}+1,l_{\text{th},K}+1} = \frac{\mathfrak{q}_K}{2}\end{cases}. \quad (44)$$

Proof: The proof is provided in Appendix B. Setting $K=2$ in (44) results in (23). ■

Replacing (44) in (10) results in the following expression of the asymptotic OP:

$$OP_{\text{Asymp}} = \left[\prod_{k=1}^{K-2} p_k(l_{\text{th},k})q_k(l_{\text{th},k})\right]\times \\ \sum_{l_{K-1}=l_{\text{th},K-1}}^{l_{\text{th},K-1}+1}\ \sum_{l_K=l_{\text{th},K}}^{l_{\text{th},K}+1} \pi_{l_{\text{th},1},\dots,l_{\text{th},K-2},l_{K-1},l_K}\times \\ p_{K-1}(l_{K-1})q_{K-1}(l_{K-1})p_K(l_K)q_K(l_K). \quad (45)$$

Similar to (27), from (11), (12) and (44), the asymptotic APD can be determined:

$$APD_{\text{Asymp}} = 2\left(1+\sum_{k=1}^{K} l_{\text{th},k}\right). \quad (46)$$

C. Observations and Conclusion

The following observations follow from inspecting the expressions in (45) and (46).

Observation *1:* Selecting the threshold levels $(l_{\text{th},1},\dots,l_{\text{th},K})$ from the set $\{1,\dots,L-2\}^K$ allows to achieve the smallest possible asymptotic OP value of: $OP_{\text{Asymp}}=\prod_{k=1}^{K}\mathfrak{p}_k\mathfrak{q}_k$.

Proof: For $l_{\text{th},k}\in\{1,\dots,L-2\}$, (3)-(4) imply that $p_k(l_{\text{th},k})=p_k(l_{\text{th},k}+1)=\mathfrak{p}_k$ and $q_k(l_{\text{th},k})=q_k(l_{\text{th},k}+1)=\mathfrak{q}_k$ for $k=1,\dots,K$. In this case, (45) simplifies to $OP_{\text{Asymp}}=[\prod_{k=1}^{K}\mathfrak{p}_k\mathfrak{q}_k][\sum_{\mathbf{l}\in\mathcal{S}}\pi_{\mathbf{l}}]=\prod_{k=1}^{K}\mathfrak{p}_k\mathfrak{q}_k$. Any value of $l_{\text{th},k}$ outside the designated set will either increase $p_k(l_{\text{th},k})$ or $p_k(l_{\text{th},k}+1)$ from $\mathfrak{p}_k$ to 1 or increase $q_k(l_{\text{th},k})$ or $q_k(l_{\text{th},k}+1)$ from $\mathfrak{q}_k$ to 1, thus, incurring an increase in the OP. ■

Observation *2:* There is no interest in increasing any of the threshold levels beyond 1.

Proof: Since all elements of $l_{\text{th},k}\in\{1,\dots,L-2\}$ contribute equally to the OP following from observation 1, the choice $l_{\text{th},k}=1$ is the most adequate among all elements of the designated set since it results in the smallest APD value following from (46). ■

Observation *3:* Following from observation 1 and observation 2, the threshold levels must be confined to the set $\{0,1\}^K$.

Observation *4:* For $(l_{\text{th},1},\dots,l_{\text{th},K})\in\{0,1\}^K$, the proposed BA relaying scheme allows to the achieve the following diversity order (d) and asymptotic APD:

$$d=K+N \quad ; \quad APD_{\text{Asymp}}=2\,(N+1)\,, \quad (47)$$

where N is the number of threshold levels that are equal to 1: $N=\sum_{k=1}^{K}\delta_{l_{\text{th},k}=1}$.

Proof: Let $\mathcal{K}=\{k\mid l_{\text{th},k}=1 \text{ for } k=1,\dots,K-2\}$. For the threshold levels in $\{0,1\}^K$, following from (3)-(4), $p_k(l_{\text{th},k})=p_k(l_{\text{th},k}+1)=\mathfrak{p}_k$ for $k\in\{1,\dots,K\}$, $q_k(l_{\text{th},k})=q_k(l_{\text{th},k}+1)=\mathfrak{q}_k$ for $k\in\mathcal{K}$ and $q_k(0)=1$. Consequently, (45) can be written as:

$$OP_{\text{Asymp}} = \left[\prod_{k=1}^{K}\mathfrak{p}_k\right]\left[\prod_{k\in\mathcal{K}}\mathfrak{q}_k\right]S, \quad (48)$$

where the asymptotic values of the summation S and its diversity order d_S are equal to $(S,d_s)=(1/2,0),(\mathfrak{q}_K,1),(\mathfrak{q}_{K-1}/2,1)$ and $(\mathfrak{q}_{K-1}\mathfrak{q}_K,2)$ for

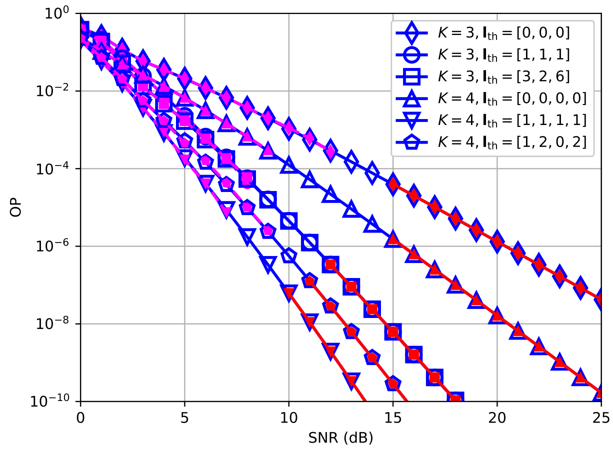

Fig. 2. OP for the asymmetrical case with $L = 8$. Solid lines with hollow marker, dashed pink lines and solid red lines correspond to the theoretical, simulation and asymptotic values, respectively.

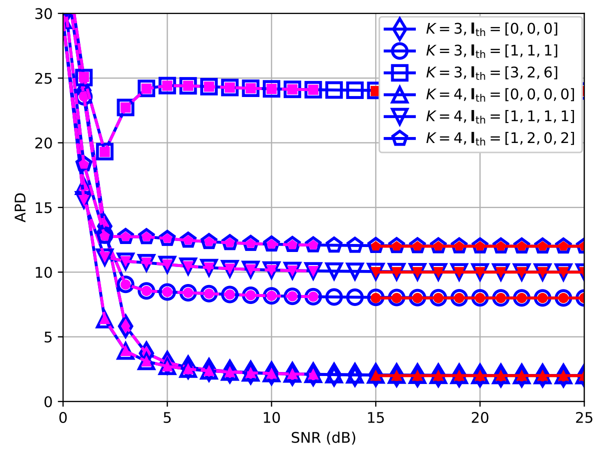

Fig. 3. APD for the asymmetrical case with $L = 8$. Solid lines with hollow marker, dashed pink lines and solid red lines correspond to the theoretical, simulation and asymptotic values, respectively.

$(l_{\text{th},K-1}, l_{\text{th},K})$ equal to $(0,0),(0,1),(1,0)$ and $(1,1)$, respectively. Therefore, from (48), $\delta = K+|\mathcal{K}|+d_S = K+N$ since d_S is equal to the number of terms among $(l_{\text{th},K-1}, l_{\text{th},K})$ that are equal to 1. Finally, the asymptotic value of the APD in (47) follows directly from (46). ∎

Following from the above observations, the following conclusions can be reached:

- All buffer sizes $L \geq 3$ achieve the same OP, APD and diversity order (this follows from $L-2 \geq 1$ so that $\{1,\dots,L-2\}$ is not empty following from observation 1). Therefore, for practical systems, the buffer size of 3 is sufficient for reaping the totality of the performance gains in the asymptotic regime.
- Setting all threshold levels to 0 constitutes the most adequate choice when the delay is considered the most critical performance metric. This allows to achieve the smallest asymptotic APD value of 2 at the expense of the lowest diversity order of K.
- Setting all threshold levels to 1 constitutes the most adequate choice when the outage is considered the most critical performance metric. This allows to achieve the highest diversity order of $2K$ at the expense of the largest asymptotic APD value of $2(K+1)$.
- Other values of the threshold levels in $\{0,1\}^K$ allow to achieve different levels of tradeoff between outage and delay. From (47), each threshold level of 0 will favor a lower delay at the expense of a higher outage.

VI. Numerical Results

In what follows, r_0 is fixed to 1 BPCU in (1). We define $\mathbf{l}_{\text{th}} \triangleq [l_{\text{th},1},\dots,l_{\text{th},K}]$. We also define the K-dimensional vectors $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}'$ as $\boldsymbol{\Omega} = [\Omega_1,\dots,\Omega_K]$ and $\boldsymbol{\Omega}' = [\Omega'_1,\dots,\Omega'_K]$, respectively. From (1)-(2), it can be observed that the knowledge of the parameters r_0, Ω_k and Ω'_k is sufficient for determining the outage probabilities $\mathfrak{p}_k$ and $\mathfrak{q}_k$ (for $k=1,\dots,K$). These outage probabilities will further determine the values of the OP and APD. Assuming a path loss exponent of 2 and a loss of 30 dB at a reference distance of 1 km, the average channel gains can be related to the link distances by $10\log_{10}(\Omega_k) = 30 - 20\log_{10}(d_k)$ and $10\log_{10}(\Omega'_k) = 30 - 20\log_{10}(d'_k)$ where d_k and d'_k stand for the lengths of the links S-R$_k$ and R$_k$-D, respectively. We assume that the middle relay R$_{\lceil K/2 \rceil}$ is aligned with S and D so that the distance between these nodes is equal to $d_{\lceil K/2 \rceil} + d'_{\lceil K/2 \rceil}$. In the simulations, we distinguish between (i): asymmetrical networks with $[d_1,\dots,d_K] = [d'_K,\dots,d'_1]$, (ii): quasi-symmetrical networks with $d_1 = \cdots = d_K$ and $d'_1 = \cdots = d'_K$ and (iii): symmetrical networks with $d_1 = \cdots = d_K = d'_1 = \cdots = d'_K$.

Fig. 2 and Fig. 3 show the OP and APD, respectively, for 3-relay and 4-relay networks with $L=8$. Asymmetrical networks are considered with $\boldsymbol{\Omega} = [4,2.5,1]$ and $\boldsymbol{\Omega}' = [1,2.5,4]$ for $K=3$ whereas $\boldsymbol{\Omega} = [4,3,2,1]$ and $\boldsymbol{\Omega}' = [1,2,3,4]$ for $K=4$. In Fig. 2 and Fig. 3, even though the asymptotic curves are not plotted over the entire SNR range for the sake of clarity, yet results show that (45) and (46) yield very accurate results for average-to-large values of the SNR. In fact, a perfect match is observed between the exact and asymptotic OP and APD curves for average-to-large values of the SNR for all values of the vector $\mathbf{l}_{\text{th}}$. Results also demonstrate the validity of the performed theoretical analysis where the theoretical OP and APD curves, from (10) and (11) respectively, match their numerical counterparts that were obtained through Monte Carlo simulations. The following observations can be made by comparing the different OP and APD curves corresponding to the same number of relays K. (i): The choice $\mathbf{l}_{\text{th}} = [0,\dots,0]$ results in the highest OP, lowest diversity order and lowest APD. In this case, once a relay receives a packet, it will give preference for transmission in the next time slot which leads to small queuing delays. (ii): Selecting all components of $\mathbf{l}_{\text{th}}$ to be different from zero results in the lowest OP and highest diversity order regardless of the specific values of these nonzero components (as long as they are different

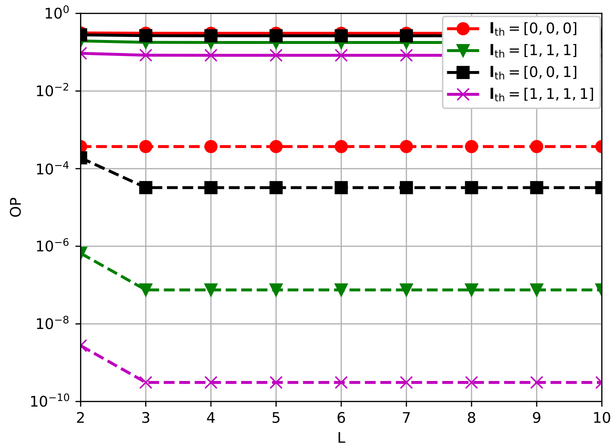

Fig. 4. OP for the quasi-symmetrical case. Solid and dashed lines correspond to the SNR values of 3 dB and 15 dB, respectively.

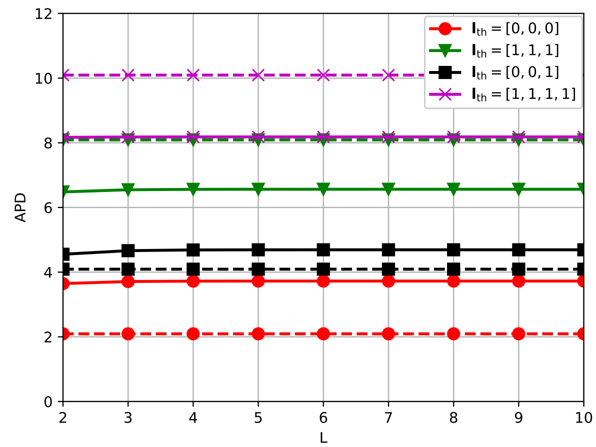

Fig. 5. APD for the quasi-symmetrical case. Solid and dashed lines correspond to the SNR values of 3 dB and 15 dB, respectively.

from $L-1$). For example, for the case $K=3$, the choices $\mathbf{l}_{\rm th}=[1,1,1]$ and $\mathbf{l}_{\rm th}=[3,2,6]$ result in exactly the same OP performance for large SNR values while the latter choice suffers from excessively large values of the APD in coherence with (46). (iii): Selecting one or more components of $\mathbf{l}_{\rm th}$ to be zero will penalize the OP and diversity order. For example, from Fig. 2 for $K=4$, the choice $\mathbf{l}_{\rm th}=[1,2,0,2]$ results in a higher OP compared to the choice $\mathbf{l}_{\rm th}=[1,1,1,1]$ where the corresponding diversity orders are 7 and 8, respectively. Unlike all APD curves in Fig. 3 that are decreasing, the APD in the scenario $\mathbf{l}_{\rm th}=[3,2,6]$ is decreasing for SNRs below 2 dB and increasing for SNRs above 2 dB. This behavior is justified by the large value $l_{{\rm th},3}=6$ at the third relay and by the fact that the link S-R_3 is the weakest among all S-R links. As such, for low SNRs, the link S-R_3 is highly unavailable and the number of stored packets at R_3 will rarely exceed the hight threshold level of 6 where, for example, $\bar{l}_3=1.92$ at 2 dB. Consequently, in this low SNR regime, R_3 is almost excluded from the relaying effort where almost all of the traffic is flowing through R_1 and R_2. As such, increasing the SNR in this regime will improve the data flow through R_1 and R_2 thus reducing the APD. On the other hand, for SNR values exceeding 2 dB, the link S-R_3 becomes more available implying that more packets will reach R_3 since the link S-R_3 is given preference over the other S-R links where, for example, $\bar{l}_3=5.91$ at 4 dB. However, since $l_{{\rm th},3}$ is large, transmissions from R_3 are less frequent (compared to the other relays) implying that the incoming packets will be queued in the buffer of R_3 for a longer time thus increasing the APD. While threshold values exceeding two were considered in Fig. 2 and Fig. 3 for the sake of demonstrating the accuracy of the theoretical analysis for any value of $\mathbf{l}_{\rm th}$, threshold values below two will be considered in the next simulation setups following from the conclusions reached in Section V-C.

Fig. 4 and Fig. 5 show the theoretical OP and APD, respectively, for a quasi-symmetrical network with $\Omega_k=1$ and $\Omega'_k=2$ for $k=1,\dots,K$. These figures target the impact of the buffer size L on the performance for different number of relays K and for different values of $\mathbf{l}_{\rm th}\in\{0,1\}^K$. Results in Fig. 4 and Fig. 5 show that the OP and APD do not practically vary with L for $L\geq 3$. Such observation holds not only for large SNR values as predicted by (45) and (46), but it also holds for small SNR values in the order of 3 dB. Therefore, increasing the buffer size beyond three does not improve the system performance in coherence with the conclusion reached in Section V-C. This demonstrates the capability of the proposed scheme in extracting the full capabilities of the network with small buffer sizes.

Figures 6-9 present a comparison between the proposed scheme and the schemes in [8], [9], [14] and [15] denoted by "*Max-Link*", "*Pref Tx*", "*Buffer state*" and "*Largest weight*", respectively. In Fig. 6 and Fig. 7, we consider a symmetrical 4-relay network with $L=5$ and $\Omega_k=\Omega'_k=1$ for $k=1,\dots,K$. In Fig. 8 and Fig. 9, we consider an asymmetrical 6-relay network with $L=3$, $\mathbf{\Omega}=[4,3.4,2.8,2.2,1.6,1]$ and $\mathbf{\Omega}'=[1,1.6,2.2,2.8,3.4,4]$. As in (47), we denote by N the number of components of $\mathbf{l}_{\rm th}$ that are equal to one. Figures 7 and 9 show that "*Max-Link*" and "*Largest weight*" achieve the same asymptotic APD value as well as comparable APD values for small SNRs. However, from figures 6 and 8, "*Largest weight*" outperforms "*Max-Link*" in terms of the OP performance for all SNR values. From figures 6 and 8, it can be observed that the best OP performance is shared by the "*Buffer state*", "*Largest weight*" and the proposed scheme for $N=K$.

The OP results in Fig. 6 and Fig. 8 show that, by controlling the values of the threshold levels, the proposed scheme is capable of achieving a broad range of OP levels ranging from the best OP performance for $N=K$ and the worst OP performance for $N=0$. This constitutes a distinctive feature that differentiates the proposed scheme from the benchmark schemes. The best OP performance is shared with the "*Buffer state*" and "*Largest weight*" schemes while the worst OP performance is the same as that of the "*Max-Link*" scheme. The APD results in Fig. 7 and Fig. 9 show that, by controlling

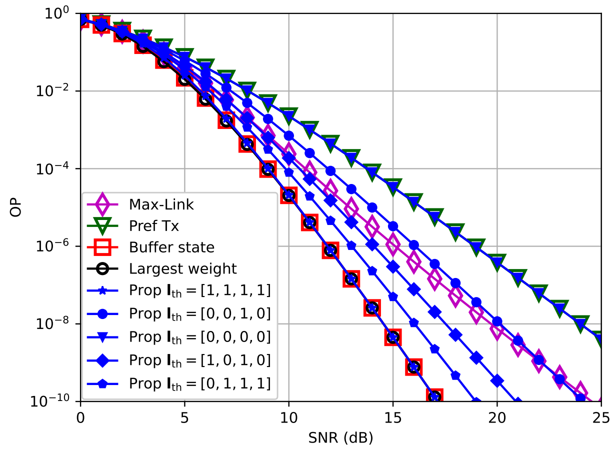

Fig. 6. OP for the symmetrical case with $K = 4$ and $L = 5$.

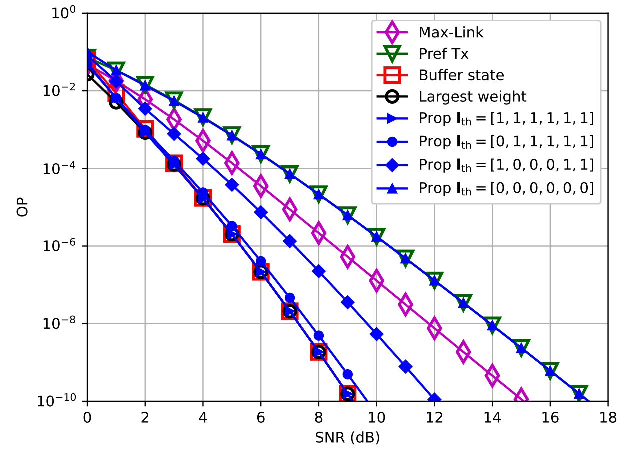

Fig. 8. OP for the asymmetrical case with $K = 6$ and $L = 3$.

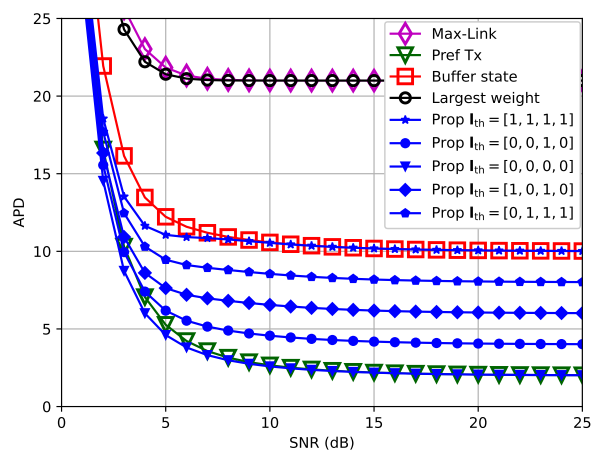

Fig. 7. APD for the symmetrical case with $K = 4$ and $L = 5$.

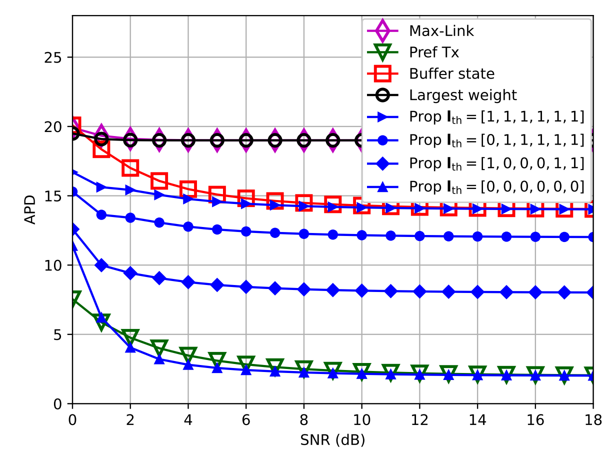

Fig. 9. APD for the asymmetrical case with $K = 6$ and $L = 3$.

the values of the threshold levels, the proposed scheme is unique in its capability of achieving a broad range of APD levels ranging from the best APD performance for $N = 0$ and the worst APD performance for $N = K$. In this context, the proposed scheme achieves smaller APD levels compared to the "*Max-Link*", "*Buffer state*" and "*Largest weight*" schemes for all values of N. Fig. 7 shows that the proposed scheme with $N = 0$ slightly outperforms "*Pref Tx*" in terms of APD for all values of the SNR. Fig. 9 shows that the APD gains with respect to the "*Pref Tx*" scheme are more pronounced but only for average-to-large values of the SNR. As such the following conclusions can be reached. (i): Compared to the "*Pref Tx*" scheme, the proposed scheme with $N = 0$ achieves the same OP performance with slightly smaller asymptotic APD values. On the other hand, the proposed scheme with $N > 0$ allows to decrease the OP at the expense of increasing the APD as expected from (47). (ii): Compared to the "*Max-Link*" scheme, the proposed scheme allows to achieve smaller APD values for all values of N ranging from 0 to K. These APD gains are associated with OP gains for large enough values of N. For example, from Fig. 6, the proposed scheme outperforms the "*Max-Link*" scheme in terms of both outage and delay for $N = 2, 3, 4$. (iii): Compared to the "*Buffer state*" and "*Largest weight*" schemes, the proposed scheme with $N = K$ allows to achieve the same minimal OP performance (whose asymptotic value is $\prod_{k=1}^{K} \mathfrak{p}_k \mathfrak{q}_k$) while profiting from reduced APD levels. In this context, the achievable APD levels are much smaller than those of the "*Largest weight*" scheme as highlighted in Fig. 7 and Fig. 9 for all values of the SNR. Moreover, the achievable asymptotic APD values are the same as those of the "*Buffer state*" scheme while the APD gains are more prominent for smaller values of the SNR especially with large number of relays. On the other hand, setting $N < K$ allows the proposed scheme to further reduce the delay compared to the "*Buffer state*" and "*Largest weight*" schemes at the expense of increasing the OP.

VII. Conclusion

We proposed a novel threshold based BA relaying scheme for cooperative networks with an arbitrary number of relays.

Through an asymptotic Markov chain analysis, we highlighted how the selection of the threshold levels impacts the achievable outage probability, queuing delay and diversity order. The analysis highlighted on the different levels of tradeoff between outage and delay that can be achieved by adjusting the threshold levels. The simulation results supported the theoretical analysis and demonstrated the high performance gains that can be reaped from the proposed relaying scheme with relatively small buffer sizes not exceeding three.

Appendix A

We first prove that the set $\mathcal{S}$ given in (22) is asymptotically closed with $t_{(l_1,l_2),(l_1',l_2')} \to 0$ for all $(l_1,l_2) \in \mathcal{S}$ and $(l_1',l_2') \notin \mathcal{S}$. As such, after a certain number of transitions, the MC will move to a state inside $\mathcal{S}$ and remains confined to this subset since the probability of leaving this subset tends to zero. As such, in the asymptotic regime, instead of considering all $(L+1)^K$ states of the MC, the analysis can be simplified by focusing only on the four states inside $\mathcal{S}$ with:

$$\sum_{(l_1,l_2)\in\mathcal{S}} \pi_{l_1,l_2} \to 1 \quad \text{for} \quad \bar{\gamma} \gg 1. \tag{49}$$

From (1), the outage probabilities $\mathfrak{p}_k$ and $\mathfrak{q}_k$ scale asymptotically as $\bar{\gamma}^{-1}$. Therefore, the product of n terms among $\{\mathfrak{p}_1,\mathfrak{p}_2,\mathfrak{q}_1,\mathfrak{q}_2\}$ scales asymptotically as $\bar{\gamma}^{-n}$ implying that such products can be neglected compared to $\{\mathfrak{p}_k,\mathfrak{q}_k,1-\mathfrak{p}_k,1-\mathfrak{q}_k\}_{k=1}^2$ for $n \geq 2$ at high SNR.

We consider the four states in $\mathcal{S}$ separately. (i): Consider the state $(l_{\text{th},1},l_{\text{th},2})$ where the transitions from this state are given in (13)-(17). Since $l_{\text{th},1} \neq L$ and $l_{\text{th},2} \neq L$ from (7), then $(p_k(l_k),q_k(l_k)) = (\mathfrak{p}_k,\mathfrak{q}_k)$ for $k=1,2$. Replacing these probabilities along with $\Delta_1 = \Delta_2 = 0$ in (13), (14), (15), (16) and (17) results in $t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1},l_{\text{th},2})} = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_1\mathfrak{q}_2 \to 0$, $t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1}+1,l_{\text{th},2})} = (1-\mathfrak{p}_1)\mathfrak{p}_2 \approx \mathfrak{p}_2$, $t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1},l_{\text{th},2}+1)} = 1-\mathfrak{p}_2$, $t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1}-1,l_{\text{th},2})} = \mathfrak{p}_1\mathfrak{p}_2(1-\mathfrak{q}_1) \to 0$ and $t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1},l_{\text{th},2}-1)} = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{q}_1(1-\mathfrak{q}_2) \to 0$, respectively. As such, the possible transitions from the state $((l_{\text{th},1},l_{\text{th},2})$ are limited asymptotically to:

$$t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1}+1,l_{\text{th},2})} = \mathfrak{p}_2 \ ; \ t_{(l_{\text{th},1},l_{\text{th},2}),(l_{\text{th},1},l_{\text{th},2}+1)} = 1-\mathfrak{p}_2. \tag{50}$$

(ii): Consider the state $(l_{\text{th},1},l_{\text{th},2}+1)$. The corresponding transition probabilities are given in (18) where the non-zero asymptotic probabilities are given by:

$$t_{(l_{\text{th},1},l_{\text{th},2}+1),(l_{\text{th},1}+1,l_{\text{th},2}+1)} = \mathfrak{q}_2 \quad ; \quad t_{(l_{\text{th},1},l_{\text{th},2}+1),(l_{\text{th},1},l_{\text{th},2})} = 1-\mathfrak{q}_2. \tag{51}$$

(iii): For the state $(l_{\text{th},1}+1,l_{\text{th},2})$, the non-zero transition probabilities in (19) are given by:

$$t_{(l_{\text{th},1}+1,l_{\text{th},2}),(l_{\text{th},1}+1,l_{\text{th},2}+1)} = \mathfrak{q}_1 \quad ; \quad t_{(l_{\text{th},1}+1,l_{\text{th},2}),(l_{\text{th},1},l_{\text{th},2})} = 1-\mathfrak{q}_1. \tag{52}$$

(iv): Finally, for the state $(l_{\text{th},1}+1,l_{\text{th},2}+1)$, the transition probabilities are given in (20)-(21). Replacing $\Delta_1=\Delta_2=1$ in these equations results in:

$$t_{(l_{\text{th},1}+1,l_{\text{th},2}+1),(l_{\text{th},1}+1,l_{\text{th},2})} = \mathfrak{q}_1 \quad ; \quad t_{(l_{\text{th},1}+1,l_{\text{th},2}+1),(l_{\text{th},1},l_{\text{th},2}+1)} = 1-\mathfrak{q}_1. \tag{53}$$

Equations (50)-(53) imply that the transitions from any state of $\mathcal{S}$ are always confined within the same set $\mathcal{S}$. Therefore, MC cannot exit the set $\mathcal{S}$ implying that this set is closed asymptotically.

The balance equations are given by $\pi_{l_{\text{th},1},l_{\text{th},2}} = (1-\mathfrak{q}_1)\pi_{l_{\text{th},1}+1,l_{\text{th},2}} + (1-\mathfrak{q}_2)\pi_{l_{\text{th},1},l_{\text{th},2}+1}$, $\pi_{l_{\text{th},1},l_{\text{th},2}+1} = (1-\mathfrak{p}_2)\pi_{l_{\text{th},1},l_{\text{th},2}} + (1-\mathfrak{q}_1)\pi_{l_{\text{th},1}+1,l_{\text{th},2}+1}$, $\pi_{l_{\text{th},1}+1,l_{\text{th},2}} = \mathfrak{p}_2\pi_{l_{\text{th},1},l_{\text{th},2}} + \mathfrak{q}_1\pi_{l_{\text{th},1}+1,l_{\text{th},2}+1}$ and $\pi_{l_{\text{th},1}+1,l_{\text{th},2}+1} = \mathfrak{q}_1\pi_{l_{\text{th},1}+1,l_{\text{th},2}} + \mathfrak{q}_2\pi_{l_{\text{th},1},l_{\text{th},2}+1}$. Solving any three of these equations along with (49) results in the steady-state probabilities provided in (23).

Appendix B

As in A, the asymptotic analysis revolves around neglecting the product of two or more terms in $\{\mathfrak{p}_k,\mathfrak{q}_k\}_{k=1}^K$. From (7), $(p_k(l_k),q_k(l_k)) = (\mathfrak{p}_k,\mathfrak{q}_k)$ for $k=1,\dots,K$. For simplicity, the four states of the set $\mathcal{S}$ in (43) will be denoted by: $\mathbf{l}_1 \triangleq (l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1},l_{\text{th},K})$, $\mathbf{l}_2 \triangleq (l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1},l_{\text{th},K}+1)$, $\mathbf{l}_3 \triangleq (l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1}+1,l_{\text{th},K})$ and $\mathbf{l}_4 \triangleq (l_{\text{th},1},l_{\text{th},2},\dots,l_{\text{th},K-1}+1,l_{\text{th},K}+1)$.

Consider the state $\mathbf{l}_1$. This state belongs to case 1 in Section V-A1 where all relays are in the Rx mode with $\Delta_k = 0$ for $k=1,\dots,K$. Following from (33), and since all Δ_k are equal, then the farthest relay from S will be chosen to receive a packet. Therefore, whenever the S-R$_K$ link is available (with probability $1-\mathfrak{p}_K$), R$_K$ is chosen to receive a packet. If the link S-R$_K$ is not available, then the previous relay R$_{K-1}$ will be chosen to receive a packet if the link S-R$_{K-1}$ is available with a probability of $\mathfrak{p}_K(1-\mathfrak{p}_{K-1}) \approx \mathfrak{p}_K$. The selection of any of the remaining relays to receive will require the unavailability of both links S-R$_K$ and S-R$_{K-1}$ with probability $\mathfrak{p}_K\mathfrak{p}_{K-1}$ that tends to zero asymptotically. On the other hand, the selection of a relay to transmit a packet will require the unavailability of all S-R links with probability of $\prod_{k=1}^K \mathfrak{p}_k \to 0$ as shown in (36). Therefore, for large SNR, the possible transitions from $\mathbf{l}_1$ are given by:

$$t_{\mathbf{l}_1,\mathbf{l}_2} = 1-\mathfrak{p}_K \quad ; \quad t_{\mathbf{l}_1,\mathbf{l}_3} = \mathfrak{p}_K, \tag{54}$$

while the probabilities of the remaining transitions will tend to zero.

Consider the state $\mathbf{l}_2$. This state belongs to case 3 in Section V-A3 where only R$_K$ is in Tx mode ($\mathcal{T}_b = \{K\}$) while all remaining relays are in Rx mode ($\mathcal{R}_b = \{1,\dots,K-1\}$). In this case, the first priority is given for R$_K$ to transmit when the link R$_K$-D is available with probability $1-\mathfrak{q}_K$. Since $\Delta_1 = \dots = \Delta_{K-1} = 0$, the second priority is given for R$_{K-1}$ to receive (relay in $\mathcal{R}_b$ that is the farthest from S) with probability $\mathfrak{q}_K(1-\mathfrak{p}_{K-1}) \approx \mathfrak{q}_K$. Therefore:

$$t_{\mathbf{l}_2,\mathbf{l}_1} = 1-\mathfrak{q}_K \quad ; \quad t_{\mathbf{l}_2,\mathbf{l}_4} = \mathfrak{q}_K, \tag{55}$$

while the probability of selecting any other transition will include the term $\mathfrak{q}_K\mathfrak{p}_{K-1} \to 0$ following from (41) and (42).

Consider the state $\mathbf{l}_3$ where $\mathcal{T}_b = \{K-1\}$ and $\mathcal{R}_b = \{1,\dots,K-2,K\}$. Interchanging the roles of relays R$_K$ and R$_{K-1}$ in the analysis of state $\mathbf{l}_2$ results in the following possible transitions:

$$t_{\mathbf{l}_3,\mathbf{l}_1} = 1-\mathfrak{q}_{K-1} \quad ; \quad t_{\mathbf{l}_3,\mathbf{l}_4} = \mathfrak{q}_{K-1}. \tag{56}$$

Consider the state $\mathfrak{l}_4$. This state belongs to case 3 in Section V-A3 with $\mathcal{T}_b = \{K-1, K\}$ and $\mathcal{R}_b = \{1, \dots, K-2\}$. Following from (39), the highest probability is given for transmission from R_{K-1} followed by the transmission from R_K:

$$t_{\mathfrak{l}_4,\mathfrak{l}_2} = 1 - \mathfrak{q}_{K-1} \quad ; \quad t_{\mathfrak{l}_4,\mathfrak{l}_3} = \mathfrak{q}_{K-1}(1-\mathfrak{q}_K) \approx \mathfrak{q}_{K-1}. \quad (57)$$

On the other hand, the probabilities in (40), (41) and (42) will tend to zero since the products $\prod_{j\in\mathcal{T}_b} \mathfrak{q}_j$ and $\prod_{j\in\mathcal{R}_b} \mathfrak{p}_j$ involve two or more terms and, hence, can be neglected.

Equations (54)-(57) show that the states $\mathfrak{l}_1$, $\mathfrak{l}_2$, $\mathfrak{l}_3$ and $\mathfrak{l}_4$ form a closed subset. Therefore, the MC analysis can be limited asymptotically to this subset that encompasses the self-contained transitions with the highest probabilities. Solving the obtained balance equations along with the relation $\sum_{i=1}^{4} \pi_{\mathfrak{l}_i} = 1$ results in the solution provided in (44).

REFERENCES

[1] M. Agiwal, R. Abhishek, and S. Navrati, “Next generation 5G wireless networks: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 3rd Quarter 2016.

[2] N. Nomikos, T. Charalambous, I. Krikidis, D. N. Skoutas, D. Vouyioukas, M. Johansson, and C. Skianis, “A survey on buffer-aided relay selection,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1073–1097, 2nd Quarter 2015.

[3] R. Kumar and A. Hossain, “Survey on half-and full-duplex relay based cooperative communications and its potential challenges and open issues using markov chains,” *IET Communications*, vol. 13, no. 11, pp. 1537–1550, May 2019.

[4] Z. Tian, G. Chen, Y. Gong, Z. Chen, and J. A. Chambers, “Buffer-aided max-link relay selection in amplify-and-forward cooperative networks,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 2, pp. 553–565, Feb. 2015.

[5] N. Zlatanov, R. Schober, and P. Popovski, “Buffer-aided relaying with adaptive link selection,” *IEEE J. Select. Areas Commun.*, vol. 31, no. 8, pp. 1530–1542, Aug. 2013.

[6] N. Zlatanov and R. Schober, “Buffer-aided relaying with adaptive link selectionfixed and mixed rate transmission,” *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2816–2840, May 2013.

[7] C. Abou-Rjeily, “Toward a better comprehension of decode-and-forward buffer-aided relaying: Case study of a single relay,” *IEEE Communications Letters*, vol. 24, no. 5, pp. 1005–1009, May 2020.

[8] I. Krikidis, T. Charalambous, and J. S. Thompson, “Buffer-aided relay selection for cooperative diversity systems without delay constraints,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1957–1967, May 2012.

[9] Z. Tian, Y. Gong, G. Chen, and J. A. Chambers, “Buffer-aided relay selection with reduced packet delay in cooperative networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2567–2575, Mar. 2016.

[10] S. L. Lin and K. H. Liu, “Relay selection for cooperative relaying networks with small buffers,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6562–6572, Aug. 2016.

[11] M. Oiwa, C. Tosa and S. Sugiura, “Theoretical analysis of hybrid buffer-aided cooperative protocol based on max-max and max-link relay selections,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9236–9246, Nov. 2016.

[12] A. A. M. Siddig and M. F. M. Salleh, “Balancing buffer-aided relay selection for cooperative relaying systems,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8276–8290, Sep. 2017.

[13] B. Manoj, R. K. Mallik, and M. R. Bhatnagar, “Performance analysis of buffer-aided priority-based max-link relay selection in DF cooperative networks,” *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 2826–2839, July 2018.

[14] S. Luo and K. C. Teh, “Buffer state based relay selection for buffer-aided cooperative relaying systems,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5430–5439, Oct. 2015.

[15] P. Xu, Z. Ding, I. Krikidis, and X. Dai, “Achieving optimal diversity gain in buffer-aided relay networks with small buffer size,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8788–8794, Oct. 2015.

[16] W. Raza, N. Javaid, H. Nasir, N. Alrajeh and N. Guizani, “Buffer-aided relay selection with equal-weight links in cooperative wireless networks,” *IEEE Communications Letters*, vol. 22, no. 1, pp. 133–136, Jan. 2018.

[17] N. Nomikos, D. Poulimeneas, T. Charalambous, I. Krikidis, D. Vouyioukas and M. Johansson, “Delay-and diversity-aware buffer-aided relay selection policies in cooperative networks,” *IEEE Access*, vol. 6, pp. 73531–73547, Nov. 2018.

[18] P. Xu, Z. Yang, Z. Ding, I. Krikidis, and Q. Chen, “A novel probabilistic buffer-aided relay selection scheme in cooperative networks,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4548–4552, Apr. 2020.

[19] P. Xu, G. Chen, Z. Yang, and H. Lei, “Buffer-state-based probabilistic relay selection for cooperative networks with delay constraints,” *IEEE Wireless Commun. Letters*, 2020, Early Access.

[20] C. Li, P. Hu, Y. Yao, B. Xia, and Z. Chen, “Optimal multi-user scheduling for the unbalanced full-duplex buffer-aided relay systems,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3208–3221, Apr. 2019.

[21] N. Nomikos, T. Charalambous, D. Vouyioukas, R. Wichman, and G. K. Karagiannidis, “Integrating broadcasting and NOMA in full-duplex buffer-aided opportunistic relay networks,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9157–9162, June 2020.

[22] J. Ren, X. Lei, P. D. Diamantoulakis, Q. Chen, and G. K. Karagiannidis, “Buffer-aided secure relay networks with SWIPT,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6485–6499, June 2020.

[23] C. Abou-Rjeily and W. Fawaz, “Buffer-aided relaying protocols for cooperative FSO communications,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8205–8219, Dec. 2017.

[24] J. D. Little and S. C. Graves, “Little’s law,” in *Building intuition*. Springer, 2008, pp. 81–100.