Relay Selection for Cooperative Relaying Networks with Small Buffers
Syuan-Li Lin and Kuang-Hao Liu, Member, IEEE.

Abstract—Buffer-aided relaying, which combines relay selection and packet buffering, has recently been studied to mitigate throughput and multiplexing gain loss of two-hop relaying protocols. However, existing buffer-aided relaying schemes require large buffers and a long initialization process for minimizing the outage probability. In this paper, a new buffer-aided relaying scheme is proposed that considers both instantaneous buffer status and channel state information (CSI) for making the relay selection decision. Theoretical analysis is performed to get some insights of the proposed scheme. We present numerical results to validate the analysis accuracy and demonstrate the improvement of the proposed scheme in terms of outage probability, buffer size requirement, and average delay, compared with existing buffer-aided relaying schemes.

Keywords—Buffer, channel state information (CSI), cooperative diversity, delay, fading channel.

I. INTRODUCTION

Achieving diversity gain through cooperative relaying has received recent research interests due to its merits in improved reliability and coverage for wireless communications with simple hardware. The idea of cooperative relaying is to let one or several intermediate nodes between the source and the destination nodes forward the source message by creating a two-hop relaying link. Via such a two-hop transmission, the source node can reach a distant destination node without using very large transmission power and thus it is more energy efficient than the conventional point-to-point transmission [1]. In addition, cooperative relaying offers diversity gain that enhances the quality-of-service (QoS) for cell-edge users [2], [3] as currently considered in 3GPP Long-Term Evolution Advanced (LTE-A) standard [4]. Some recent work also reveals that cooperative relaying is appealing to throughput improvement in various network configurations [5]–[7]. In practice, the source message may traverse across the network via several consecutive relaying links, leading to the multi-hop transmission. Since the two-hop relaying serves as the building block of the more involved multi-hop scenario, this work focuses on the two-hop setting.

The diversity gain achieved via cooperative relaying normally comes at the expense of multiplexing gain loss and a higher resource consumption. For example, the implementation of fixed relaying (FR) protocol in a network with \( K \) relays would need \( K \) orthogonal channels (in the form of time or frequency) to achieve full diversity gain (equal to the number of diversity paths \( K \)) but incurs the spectral efficiency penalty of \( 1/K \) [8]. Relay selection (RS) has been regarded as a promising countermeasure to improve the spectral efficiency of cooperative relaying. By choosing the best relay to assist the source node, the achieved diversity gain can scale with the number of relay nodes without consuming extra resources. The most well-known RS scheme is opportunistic relaying (OR), where the best relay is selected as the one with the strongest bottleneck of the two-hop relaying link [9]. Another example is the selection cooperation (SC) [10], which selects the relay with the best relay-destination link from the set of relays that can decode successfully, referred to as the decoding set. Both OR and SC rely on channel state information (CSI) to choose the best relay. RS has also been combined with other mechanisms for reducing the throughput loss, such as incremental opportunistic relaying (IOR) [11]–[13] that utilizes destination feedback to decide whether additional resources should be allocated to the selected relay.

A. Related Work

In conventional RS schemes, such as OR [9], SC [10], and their variants [11]–[15], the selected relay forwards the received signal immediately to the destination. In addition, the same relay is used to receive and forward the source signal. However, the source-relay and relay-destination links associated with the selected relay usually experience independent fading. Therefore, conventional RS schemes may fail to fully utilize all available diversity paths. Some recent efforts have explored relay buffers as another design dimension for improving the performance of RS aiming at two design objectives.

- Throughput gain: Due to the capacity disparity between the source-relay and relay-destination links, conventional RS schemes may suffer the bottleneck effect that severely degrades the throughput of relay networks [8]. In [16] and [17], a store-and-forward strategy with relay scheduling is proposed, independently, to minimize the bottleneck effect and in turn increase the end-to-end throughput by taking advantage of the time-varying nature of wireless channels.
- Multiplexing gain: To reduce the hardware complexity and self interference, half-duplex radios are widely
adopted in cooperative networks but they suffer multiplexing gain loss because only one packet can be delivered in two time slots. RS can remedy this shortcoming by selecting different relays for transmission and reception as in [18] so that two packets can be processed in two slots. The similar idea is also considered in [19]–[22], where the relay with the best source-relay link is selected for receiving the source signal and this relay is scheduled to transmit when its relay-destination link is superior than the others. These schemes, referred to as link adaptive relaying (LAR), differ in that the source and the selected relay follow a fixed schedule to transmit in [18] and [20], whereas their transmission sequence is non-deterministic in [19], [21], [22].

Readers may also refer to [23] for an elegant overview of buffer-aided cooperative communications.

B. Paper Contribution

In this work, we aim to enable buffer-aided relaying with small buffers. While this consideration is more practical than the assumption of infinite buffer size in order to achieve full diversity gain [18], [20], [21], it greatly challenges the protocol operation. With finite relay buffers, the probability of buffer overflow is non-zero. Also, there is a non-zero probability that the relay buffer becomes empty if it is selected for transmission more often than the other relays. Both cases will lead to significant performance degradation. To remedy the above-mentioned problems, a new RS scheme is proposed to reduce the likelihood that the selected relay for transmission (respectively for reception) sees an empty (respectively a full) buffer. The main contributions of this work are highlighted below.

- We propose a novel RS scheme, which utilizes both CSI and buffer status information to equalizes buffer lengths of all relays. Consequently, large relay buffers are not required as assumed in the existing buffer-aided relaying schemes [16]–[22] that only consider CSI. Even with very small relay buffers, we show that the proposed scheme can achieve full diversity gain as in previous works [18]–[22] which necessitates the relay buffers to be sufficiently large such that the probability of full relay buffers is negligible.
- We analyze the outage performance of the proposed scheme by using the Markov-chain model that captures the evolution of all relay buffers. Although the Markovian modeling approach has been considered in [20], [21], our analysis provides more insights such as the equilibrium behavior and the diversity gain of the proposed scheme.
- Extensive numerical results along with comparisons with existing buffer-aided RS schemes are presented to validate the accuracy of the analytical results and to demonstrate the performance of the proposed scheme including the outage probability and average delay under various system parameters. In addition to some ideal settings, practical conditions including outdated CSI and non-identical channel distributions are also investigated via simulations. It is shown that the proposed scheme can greatly reduce the delay and the required buffer size for minimizing the outage probability.

The remainder of this paper is organized as follows. Sec. II describes the system model and the relevant benchmark schemes. A new buffer-aided relaying scheme is proposed in Sec. III. In Sec. IV, we analyze the outage probability of the proposed scheme in the general setting followed by the closed-form derivation for a simplified scenario in Sec. V. Next, we present the outage and delay performance of the proposed scheme in comparison with other related schemes in Sec. VI. Finally, concluding remarks are drawn in Sec. VII.

II. System Model and Review of Relevant Schemes

A. System Model

As shown in Fig. 1, we consider a two-hop relaying network consisting of one source node $S$, one destination node $D$, and $K$ relays $R_1, \ldots, R_K$, all performing decode-and-forward (DF). Each relay is half duplex radio and equipped with a buffer. The buffer size denoted by $B$ is identical to all relays, and $L_i$ represents the buffer length of relay $R_i$. Node $S$ communicates to node $D$ via relays according to a two-phase protocol. In phase I, node $S$ broadcasts a packet, and one of the relays is selected to receive and store the source packet in its buffer. In phase II, a relay is selected to transmit a packet from its buffer to node $D$. The relay selection rule will be elaborated in Sec. III. Assuming a time slotted system, each packet transmission time corresponds to one slot duration. The channel coefficients of $S - R_i$ and $R_i - D$ links, $i = 1, \ldots, K$ are denoted as $g_i$ and $h_i$, respectively. The following assumptions on made throughout this paper.

- All links are assumed to be independent and identically distributed (i.i.d.) Rayleigh fading channels, where $g_i$ and $h_i$ are mutually independent complex Gaussian distributed random variables with zero mean and unit
variance. The channel coefficients remain constant during one slot, but change independently from one slot to another. The case with independent and non-identically distributed (i.n.d.) fading will be discussed in Sec. VI through numerical results.

- Denote \( \frac{P}{N_0} \) as the unfaded signal-to-noise ratio (SNR), where \( P \) is the constant transmit power and \( N_0 \) is the additive white Gaussian noise (AWGN) power. Both are assumed to be identical to all nodes. Thus the instantaneous SNR between \( S \) and \( R_i \) can be represented as \( \gamma_{g_i} = |g_i|^2 \frac{P}{N_0} \) with mean \( \gamma_{g_i} \). Likewise, the instantaneous SNR of the \( R_i - D \) link is symbolized as \( \gamma_{h_i} = |h_i|^2 \frac{P}{N_0} \) with mean \( \gamma_{h_i} \).

- All the nodes transmit in a fixed rate \( R \) b/s/Hz. In information-theoretic sense, the source transmits to relay \( R_i \) in outage if the mutual information of the \( S-R_i \) channel does not exceed the required transmission rate, i.e., \( 1/2 \log_2(1 + \gamma_{g_i}) < R \), which is equivalent to \( \gamma_{g_i} > T \) with \( T = 2^R - 1 \). Likewise, relay \( R_i \) transmits to \( D \) in outage if \( \gamma_{h_i} < T \).

The above assumptions are commonly adopted in the related work [20] and are used to simplify the theoretical analysis in this paper.

### B. Benchmark Schemes

We review two existing schemes relevant to our study and will be used as the performance benchmark in Sec. VI.

1) **Nonbuffer-Aided Scheme:** OR scheme proposed in [9] is a simple yet effective RS scheme that achieves full diversity gain based on the max-min selection criterion. In OR, the best relay denoted as \( R^* \) is defined as the one with the strongest bottleneck of the dual-hop relaying link, as expressed by

\[
R^* = \arg \max_{R_i} \left( \min \left\{ \gamma_{S,R_i}, \gamma_{R_i,D} \right\} \right). \tag{1}
\]

Without buffering, the selected relay \( R^* \) will receive the source packet in phase I and immediately forward the received packet in phase II. As a result, OR suffers the bottleneck effect because the end-to-end link capacity associated with the selected relay is limited by the weakest hop.

2) **Buffer-Aided Scheme:** To mitigate the bottleneck effect, the max-max relay selection (MMRS) proposed in [20] chooses the relay with the best source-relay link to receive in phase I and that with the best relay-destination link to transmit in phase II, respectively, as expressed by

\[
(R^*_r, R^*_t) = \arg(\max_{R_i} \{\gamma_{S,R_i}, \max_{R_i} \{\gamma_{R_i,D}\}\}). \tag{2}
\]

For relays with infinite buffer size, MMRS achieves the performance upper bound of any RS schemes using two phases.

### III. PROPOSED SCHEME

In this section, we propose a new RS scheme to improve the performance of buffer-aided relaying. As discussed in Sec. I-A, buffer-aided relaying can fully utilize both the first- and the second-hop links as compared to traditional nonbuffer-aided counterparts. However, relays are not able to perform relaying whenever their buffers are full or empty. Increasing the buffer size may not be a promising strategy because the resultant buffering delay grows with the buffer size. Given a limited buffer size, we are interested in the method to reduce the delay cost of buffer-aided relaying while still achieving a full diversity gain. In the following, we first explain the key ideas of the proposed scheme, followed by detailing the protocol.

The proposed scheme aims to reduce the outage probability due to buffer overflow and underflow in buffer-aided relaying schemes. This is achieved by two steps. The first step is to define two sets of relays that are eligible to perform cooperation. Generally, whether a relay can receive the source message successfully depends on the individual link quality of the first hop. In this work, we define the set of relays that can receive the source message successfully as the reception set \( \mathcal{R} = \{R_i|\gamma_{g_i} \geq T\} \), where \( T \) is the SNR threshold defined in Sec. II-A. Similarly, the set of relays that can transmit successfully is defined as transmission set \( \mathcal{T} = \{R_i|\gamma_{h_i} \geq T\} \).

Given the two sets of relays defined above, RS is based on the following rule: the relay in \( \mathcal{R} \) with the shortest buffer length is chosen for reception in phase I, and the relay in \( \mathcal{T} \) with the longest buffer length is chosen for transmission in phase II, respectively. Since the source packet is filled in the shortest relay buffer in \( \mathcal{R} \) and taken out from the longest relay buffer in \( \mathcal{T} \), the proposed scheme is referred to as shortest-in-longest-out (SILO), which chooses the best relay pair for reception and transmission according to

\[
(R^*_r, R^*_t) = \arg(\min_{R_i \in \mathcal{R}} \{L_i\}, \max_{R_i \in \mathcal{T}} \{L_i\}), \tag{3}
\]

where \( L_i \) represents the buffer length of relay \( R_i \) at the relay selection epoch.

The proposed RS rule can effectively alleviate the requirements of large relay buffers and a long initialization process, both resulting in undesired queueing delay. Intuitively, increasing the buffer size helps to reduce the chance of buffer overflow at the cost of longer queueing delay. SILO addresses this problem by choosing the relay for reception according to the “shorted queue first” discipline and that for transmission based on the “longest queue first” principle. As a consequence, different relay buffers tend to maintain the same length, approximately equal to the number of packets filled into the buffer initially. Different from existing buffer-aided relaying schemes that require the initial buffer length to be at least half full to minimize the outage probability [18], [20], the proposed scheme can achieve the same goal with a very small number of packets filled in the initialization phase. This will become more clarified in Sec. VI when we discuss the numerical results.

With finite buffer size, buffer overflow or underflow is inevitable. We remedy this problem by combining SILO and OR, as in [20]. Specifically, SILO is adopted under the following conditions: when both \( \mathcal{R} \) and \( \mathcal{T} \) are not empty as well as when the buffer of the receiving relay is not full and the buffer of the transmitting relay is not empty. Otherwise, OR will be used. Hence, the best relay pair for reception and transmission can
be expressed as

\[
(\hat{R}_t^e, \hat{R}_t^o) = \begin{cases} (R_t^e, R_t^o), & \text{if } (R \neq \emptyset) \cap (T \neq \emptyset) \\
\cap (L_{R_t} < B) \cap (L_{R_t} > 0) \quad (4) \\
(R^e, R^o), & \text{otherwise},
\end{cases}
\]

where \( R^e, R^o \) and \( R^o_t \) are defined in (1) and (3), respectively. In the rest of this paper, we call the combined scheme as combined relay selection (CRS). A complete description of CRS is as follows.

- Initialization: Each relay buffer is fed with a certain number of packets. The length of the initialization phase presents an important tradeoff between the outage probability and the resultant delay. This will be discussed further in Sec. VI using numerical results.
- Phase I: Each relay reports their buffer length and channel condition to the source node, which performs relay selection according to (4). Then the source starts to transmit a packet with the relay selection result piggybacked. Only the relay \( R^o_t \) selected for reception will listen to the source message and store the received packet in its buffer. The rest of relays are idle during phase I.
- Phase II: The relay \( R^e_t \) selected for transmission sends its HOL packet to the destination node and remove the transmitted packet from the buffer.

The protocol proceeds by repeating phase I and phase II. Next, we discuss some implementation issues of the proposed scheme.

**Remark:** The relays considered in this work are assumed to be half-duplex. SILO also works for full-duplex radios with a slight modification by always selecting different relays for transmission and reception as in [18]. Secondly, SILO relies on the information of buffer status fed back from relays to the source node to perform RS. Alternatively, a timer-based approach similar to that in [9] can be used to eliminate the feedback overhead. Finally, a common drawback of buffer-aided relaying is that packets may not arrive at the destination node in sequence. The destination node needs to maintain a reordering buffer as used in the Automatic Repeat-reQuest (ARQ) protocol.

IV. OUTAGE PROBABILITY ANALYSIS

In analyzing the performance of CRS, recall that CRS switches between SILO or OR according to the instantaneous channel quality and the buffer status. When SILO is used, the relay selected from the reception set \( R \) ensures an outage-free reception in phase I while the relay selected from the transmission set \( T \) guarantees an error-free transmission in phase II. Consequently, the outage probability with SILO being used is zero. On the other hand, outage may occur if OR is used that occurs when (i) \( R \) or \( T \) is empty with probability denoted as \( p_{\text{empty}} \), or (ii) when the buffer of \( R^o_t \) is full or that of \( R^o_t \) is empty with probability denoted as \( p_{\text{null}} \), according to (4). By the law of total probability, the outage probability of CRS can be expressed as

\[
P_{\text{out}}^{\text{CRS}} = \sum_{a \in A} p_a \cdot P_{\text{out},a}
\]

where \( A = \{\text{empty}, \text{null}\} \), and \( P_{\text{out},a} \) is the conditional outage probability on \( a \in A \).

To find \( p_a \) and \( P_{\text{out},a} \), we model the evolution of relay buffers at each RS epoch as a Markov chain (MC) denoted by \( \Psi \), where the \( i \)-th state is defined as \( s_i = (X_1, X_2, \cdots, X_K) \) with \( X_j, j = 1, \cdots, K \) being the buffer length of relay \( R_j \). Each state in the constructed MC needs to meet the following constraints

\[
0 \leq X_j \leq B - 1, \quad j = 1, \cdots, K
\]

(6)

\[
\sum_{j=1}^{K} X_j = \sum_{j=1}^{K} L_{0,j} \neq L,
\]

(7)

where \( L \) accounts for the total number of packets in the network and \( L_{0,j} \) denotes the initial buffer length of relay \( R_j \). Specifically, (6) follows the definition of “full” buffer, and (7) is the consequence of the constant transmission rate such that at the end of any time slot, the total number of packets in the \( K \) relay buffers must be identical to the total number of packets fed during the initialization phase.

A. Properties of the Markov Chain

\( \Psi \) is characterized by the state transition probability matrix (TPM) \( \mathbf{P} = [p_{ij}] \) and the steady-state probability vector \( \pi = [\pi_1, \cdots, \pi_J] \), where \( J \) denotes the total number of states. In \( \mathbf{P} \), the entry \( p_{ij} = \mathbb{P}[s_i \rightarrow s_j] \) represents the transition probability from state \( s_i \) to state \( s_j \) in two consecutive slots.

**Lemma 1.** The MC \( \Psi \) has the following general properties:

1) It is regular, i.e., the MC is irreducible, positive recurrent, and aperiodic.
2) The transition probability matrix \( \mathbf{P} \) is of dimension \( J \times J \) where

\[
J = K + L - 1 \cdot C_L
\]

(8)

with \( C_L \) denoting the binomial coefficient and \( L \) is specified in (7).
3) \( \mathbf{P} \) is asymmetric, i.e., \( p_{i,j} \neq p_{j,i} \).
4) \( \mathbf{P} \) is row stochastic but not column stochastic.
5) The steady-state probability of the initial state \( s_i = (L_0, L_0, \cdots, L_0) \) is higher than that of any other states when SNR is sufficiently high.
6) The steady-state probability vector \( \pi \) exists, and it is unique and strictly positive.

**Proof:** The proofs for the above properties follow the standard definitions of Markov chains and the features of the proposed RS scheme. Due to space limit, we delegate the details to [24] for interested readers.

The steady-steady probability vector \( \pi \) can be obtained by solving a system equation \( \pi \mathbf{P} = \pi \) along with the normalisation equation \( \sum_{i=1}^{J} \pi_i = 1 \). Given \( \pi \), then we can
find the probability \( p_s \) defined in (5) by
\[
p_s = P_s \pi^T
\]
where \( P_s = [P_{s_1}, \ldots, P_{s_5}] \) with the \( j \)th entry specifying the probability that case \( a \in A \) takes place at state \( s_j \) and \( \pi^T \) is the transpose of \( \pi \). Since there does not exist a closed-form expression for \( \pi \) in general, an illustrate example instead is employed in the next section to gain insights of the proposed scheme.

V. ILLUSTRATIVE EXAMPLE

In this section, the outage probability of CRS is derived in closed form for a simplified setting with two relays \((K = 2)\), the buffer size \( B = 6 \) and the initial buffer length \( L_0 = 2 \) of each relay. According to (8), the constructed MC has five states, including (4,0), (3,1), (2,2), (1,3), and (0,4). As shown, this MC has a symmetric structure. To ease the presentation, define the following notations: \( \alpha_c = e^{-T/c} \) and \( \alpha_c = 1 - \alpha_c \) where \( c \in \{g, h\} \). Recall that \( \gamma_g \) and \( \gamma_h \) defined in Sec. II represent the average SNR of the S-R link and R-D link, respectively, where the relay-dependent subscript has been omitted due to the i.i.d. assumption.

We start from deriving the steady-state probabilities of the MC \( \Psi \) according to whether the same relay is selected for reception and transmission.

A. Steady-State Probabilities

Case 1: \( \bar{R}^*_i = \bar{R}^*_i \), which occurs when \((R \neq \emptyset) \cap (T \neq \emptyset)\) such that SILO is used. Given the current state \( s_i = (X_1, X_2) \) where \( i = 1, \ldots, 5 \), the state transition of the MC can be further classified into two sub-cases depending on the buffer status.

Case 1.1: \( X_1 = X_2 \), corresponding to state \( s_3 \). The state transition probability associated with \( s_3 \) can be obtained as
\[
\begin{align*}
P[s_3 \rightarrow s_j] &= P[R^*_i \neq R^*_i | X_1 = X_2] \\
&= (\alpha_g - \frac{1}{2} \alpha_g^2) (\alpha_h - \frac{1}{2} \alpha_h^2) \triangleq C_1, \quad j = (2, 4).
\end{align*}
\]
(10)

Case 1.2: \( X_1 \neq X_2 \), corresponding to states \( s_1, s_2, s_4, \) and \( s_5 \). The associated transition probability is
\[
\begin{align*}
P[s_i \rightarrow s_j] &= P[R^*_i \neq R^*_i | X_1 \neq X_2] \\
&= \alpha_g \alpha_h \triangleq C_2, \quad (i, j) = (1, 2), (2, 3), (4, 3), (5, 4).
\end{align*}
\]
(11)

Notice that \((i, j) = (2, 1)\) and \((4, 5)\) are special cases that occur only when \(|R| = |T| = 1\). Take \( s_2 = (3, 1) \) for example, it transits to state \( s_1 = (4, 0) \) only if \( R = \{R_1\} \) and \( T = \{R_2\} \). Hence the corresponding transition probability is
\[
P[s_i \rightarrow s_j] = P[R^*_i \neq R^*_i | X_1 \neq X_2] \\
= \alpha_g \alpha_h \triangleq C_3, \quad (i, j) = (2, 1), (4, 5).
\]
(12)

Case 2: \( \bar{R}^*_i = \bar{R}^*_i \), which implies the transition from state \( s_i \) to itself for all \( i = 1, \ldots, 5 \). Depending on the buffer status, we further consider three sub-cases.

Case 2.1: \( X_1 = X_2 \), corresponding to state \( s_3 \). If both the sets \( R \) and \( T \) are nonempty, SILO will be used and the transition probability \( P[s_3 \rightarrow s_3] \) is identical to (10). If either \( R \) or \( T \) is empty, OR will be used and the resulting transition probability is
\[
P[s_3 \rightarrow s_3] = P[\bar{R}^*_i = \bar{R}^*_i | (R = \emptyset) \cup (T = \emptyset)] \\
= \alpha_g^2 + \alpha_h^2 - \alpha_g^2 \alpha_h^2 \triangleq C_4.
\]
(13)

Case 2.2: \( X_1 \neq X_2 \), corresponding to states \( s_1, s_2, s_4, \) and \( s_5 \). If \( (R \neq \emptyset) \) and \( (T \neq \emptyset) \), SILO will be used, whereby the same relay \( \bar{R}^* \) is selected for transmission and reception only if \((R = \{R^*_i\}) \cap (R^*_i \in T)\) or \((R^*_i \in R) \cap (T = \{R^*_i\})\). Therefore, the associated transition probability is
\[
P[s_i \rightarrow s_i | (R \neq \emptyset) \cap (T \neq \emptyset)] \\
= P[\bar{R}^*_i = \bar{R}^*_i | (R = \{R^*_i\}) \cap (R^*_i \in T)] \\
+ P[\bar{R}^*_i = \bar{R}^*_i | (R^*_i \in R) \cap (T = \{R^*_i\})] \\
= \alpha_g \alpha_g \alpha_h + \alpha_h \alpha_h \alpha_g, \quad i = (2, 4).
\]
(14)

On the other hand, OR will be used if either \( R \) or \( T \) is empty, in which case the transition probability is identical to (13). Therefore, the overall transition probability \( s_1 \rightarrow s_i \) for \( \bar{R}^*_i \) is equal to the sum of (13) and (14), and \( C_3 + C_4 \triangleq C_5 \).

Case 2.3: \( X_1 \neq X_2 \), \( X_1 = 0 \) or \( X_2 = 0 \) but not both, corresponding to states \( s_1 \) and \( s_5 \). This case is similar to Case 2.2, except that one of the relay buffers is empty. If the relay selected for transmission has an empty buffer, OR will be used instead, leading to the same relay being selected for transmission and reception. Take \( s_1 = (4, 0) \) for example. \( R_2 \) must be selected when \( T = \{R_2\} \) and \( R \neq \emptyset \) using OR. The transition probability of \( s_1 \rightarrow s_1 \) can be derived as
\[
P[s_1 \rightarrow s_1] = P[\bar{R}^*_i = \bar{R}^*_i | (R = \{R_1\}) \cap (R_1 \in T)] \\
+ P[\bar{R}^*_i = \bar{R}^*_i | (R = \emptyset) \cup (T = \emptyset)] \\
= \alpha_g \alpha_g \alpha_h + \alpha_h \alpha_h \alpha_g, \quad i = (2, 4).
\]
(15)

where the first probability accounts for the case when SILO is used, while the latter two correspond to the case when OR is used. Due to the symmetry of the MC, the transition probabilities \( s_1 \rightarrow s_5 \) is identical to (15).

Based on the given results, the transition probability matrix \( P \) can be shown as
\[
P = \begin{bmatrix}
C_6 & C_2 & 0 & 0 & 0 \\
C_3 & C_5 & C_2 & 0 & 0 \\
0 & C_1 & C_4 & C_1 & 0 \\
0 & 0 & C_2 & C_5 & C_3 \\
0 & 0 & 0 & C_2 & C_6
\end{bmatrix}.
\]
(16)

It can be verified that \( P \) in (16) is regular (i.e., sum of all values in each row is one and each element is non-negative). According to Lemma 1.6, there exists a unique steady-state probability vector \( \pi \) that satisfies \( P \pi = \pi \) and \( \sum_{i=1}^{N} \pi_i = 1 \). In general, solving \( \pi \) involve solving a system of linear equations whose number grows with the dimension.
Lemma 2. Regardless of the initial state, the MC \( \Psi \) tends to stay on state \( s_3 = (2, 2) \), and the tendency is more pronounced as the average SNR increases.

Proof: See Appendix B.

Lemma 2 implies that for any initial setting, SILO can be utilized to reduce the number of required linear equations by half. In Appendix A, we obtain the steady-state probability of each state as given by

\[
\pi_1 = \pi_5 = \frac{C_1C_3}{C_2^2 + 2C_1(C_2 + C_3)}
\]

\[
\pi_2 = \pi_4 = \frac{C_1C_2}{C_2^2 + 2C_1(C_2 + C_3)}
\]

\[
\pi_3 = \frac{C_2^2}{C_2^2 + 2C_1(C_2 + C_3)}.
\]

The symmetry of the MC \( \Psi \) also suggests the following equilibrium behavior of SILO.

Lemma 2.

Proof: See Appendix B.

We proceed to derive \( p_{\text{out}} \), namely the average probability of event \( a \in A = \{ \text{empty}, \text{null} \} \) in (9).

1) \( a = \text{empty} \): This corresponds to the case whenever \( R \) or \( T \) is empty. Hence

\[
\begin{align*}
\tilde{p}_{\text{empty}} = & \ P(\{R = \emptyset \} \cup \{T = \emptyset \}) = \tilde{\alpha}_g^2 + \tilde{\alpha}_h^2 - \tilde{\alpha}_g^2\tilde{\alpha}_h^2. \\
\end{align*}
\]

2) \( a = \text{null} \): This is the case when the buffer of the selected relay for transmission (or reception) is empty (or full). In the above toy example, only states \( s_1 = (4, 0) \) and \( s_5 = (0, 4) \) satisfy this condition, where \( R_2 \) and \( R_1 \) are selected for transmission and reception, respectively. For state \( s_1, R_2 \) is selected for transmission if \( R \neq \emptyset \) and \( T = \{R_2\} \). For state \( s_5, R_1 \) is selected for transmission if \( R \neq \emptyset \) and \( T = \{R_1\} \). Consequently, the average outage probability of the null condition associated with state \( s_i \) can be found as

\[
\begin{align*}
P_{\text{null},i} = & \ (1 - \tilde{\alpha}_g^2)\tilde{\alpha}_h\tilde{\alpha}_h, \quad i = 1, 5, 0, \quad i = 2, 3, 4. \\
\end{align*}
\]

Substituting (21) into (9), we obtain

\[
\tilde{p}_{\text{null}} = 2\pi_1(1 - \tilde{\alpha}_g^2)\tilde{\alpha}_h\tilde{\alpha}_h,
\]

where we use the symmetry \( \pi_1 = \pi_5 \).

C. Conditional Outage Probability and Diversity Order

To compute the outage probability in (5), we first find the conditional outage probability on \( a \in A = \{ \text{empty}, \text{null} \} \). In either case, OR will be used that results in the same relay \( \tilde{R} \) being selected for reception and transmission. Furthermore, the case \( a = \text{empty} \) takes place when either \( R \) or \( T \) is empty, or equivalently, \( \tilde{\gamma}_{gR} < T \) or \( \tilde{\gamma}_{bR} < T \). For OR with DF relays, the outage probability is given by [9]

\[
P_{\text{out, empty}} = \mathbb{P}(\min(\tilde{\gamma}_{gR}, \tilde{\gamma}_{bR}) < T) = 1.
\]

For \( a = \text{null} \), the conditional outage probability with \( K \) relays is equal to

\[
P_{\text{out, null}} = \mathbb{P}(\min(\tilde{\gamma}_{gR}^K, \tilde{\gamma}_{bR}^K) < T) = \tilde{\alpha}_g^K,
\]

where \( \tilde{\alpha}_c = 1 - e^{\frac{-\pi}{\gamma_c}} \). When \( \gamma_c \) is sufficiently large, the approximation \( \tilde{\alpha}_c \approx \frac{T}{\gamma_c} \) holds according to the Taylor series approximation. Asymptotically, \( \tilde{p}_{\text{out, null}} \) can be approximated as

\[
\begin{align*}
\tilde{p}_{\text{out, null}} &= \mathbb{P}(\min(\tilde{\gamma}_{gR}^K, \tilde{\gamma}_{bR}^K) < T) \\
&= \tilde{\alpha}_g^K + \tilde{\alpha}_h^K - \tilde{\alpha}_g^K\tilde{\alpha}_h^K \\
&\approx \left( \frac{T}{\gamma_c} \right)^K + \left( \frac{T}{\gamma_c} \right)^K - \left( \frac{T}{\gamma_c} \right)^K. \\
\end{align*}
\]

The above three schemes employ a deterministic two-phase schedule to arrange the source and relay transmissions. Instead, the Max-Link scheme proposed in [21] uses a non-deterministic schedule by choosing the best link (either the source-relay or the relay-destination link) at each transmission opportunity.
Fig. 2. Outage probability vs. SNR for $B = 6$ and $L_0 = 2$.

A. Outage Performance

We first show the outage probability versus SNR with fixed buffer size $B = 6$, initial buffer length $L_0 = 2$, and varying number of relays $K$ in Fig. 2. As can be seen, CRS performs the same as the ideal MMRS and achieves diversity order of $K$, which validates Lemma 3. When $K$ increases, CRS provides a larger coding gain over HRS and OR at high SNR. The improvement can be explained by Lemma 1.5, according to which all the relay buffers are maintained at their initial status, i.e., non-full and non-empty, using CRS. On the contrary, in HRS, relay buffers can be of any length with an equal probability [20, Lemma 1]. In other words, CRS reduces the chance of a selected relay seeing an empty (or full) relay buffer that causes outage, as compared with HRS. Among all the considered schemes, Max-Link performs the best in terms of outage probability because each slot is allocated to the best link out of total of $2 \cdot K$ links. This provides the maximum diversity order equal to $2 \cdot K$ as compared to $K$ achieved by CRS. We note that each relay buffer in Max-Link needs to be sufficiently large to achieve the maximum diversity order but this introduces adverse effect to the delay performance. Also, Max-Link needs to dynamically determine the transmission sequence between the source and the selected relay nodes. To enable dynamic scheduling, additional signaling should be added.

Fig. 3 studies the impact of buffer size $B$ to the outage probability of different schemes for SNR $= 15$ dB and $K = 2$ and 3. Here we focus on the RS schemes using deterministic scheduling and thus Max-Link is not included. Following [20], each relay buffer is set to be half full initially, i.e., $L_0 = B/2$, except ideal MMRS that assumes $B \to \infty$. Consequently, the outage probability of ideal MMRS is independent from buffer size $B$ and serves as the lower bound of buffer-aided relaying schemes. As the buffer size increases, the outage probabilities of HRS and the proposed CRS approach to that of ideal MMRS. For $K = 2$, CRS achieves the minimum outage probability when $B = 6$, while HRS needs $B \geq 30$ (i.e., five times the buffer size) to minimize the outage probability. For $K = 3$, the minimum outage probabilities of HRS and CRS occur at $B = 100$ and $B = 6$, respectively. The results indicate that CRS significantly reduces the buffer size requirement for minimizing the outage probability.

B. Impact of Initial Buffer Length

The effect of initial buffer length to the delay performance is investigated in Fig. 4. For buffer-aided RS schemes such as HRS and the proposed CRS, maintaining non-empty relay
buffers initially is essential to the normal protocol operation. Otherwise, the relay selected for transmission in the first few slots might have nothing to send. Even each relay buffer may be filled with packets in the subsequent slots, the chance that a relay buffer becomes empty may still be high if the number of packets filled initially is not sufficient. Suppose the packet arrives at the relay buffer in a first-come first-served sense, the average delay is defined as the average duration measured in slot when a packet arrives a relay buffer until it leaves the relay buffer, which happens if the relay is selected for transmission. Here, we fix SNR = 15 dB and $B = 100$, and vary the number of relays $K = 3$ and 4. As depicted, the delay curves of both CRS and HRS are both linear with the slope approximately equal to $K \times L_0$. This result coincides with that in [20, Fig. 10]. One can see that as the number of relays $K$ increases, the average relay becomes higher because each relay tends to wait longer to be selected for transmission. It is worth noting that increasing $K$ is beneficial to the diversity and coding gains as shown in Fig. 2. Also, the delay performance in Max-Link is nearly invariant to the initial buffer length $L_0$ because Max-Link can switch between the source-relay and the relay-destination link on a per-slot basis, whichever is more appropriate than the other. Therefore, the occurrence of an empty relay buffer does not lead to outage unless all relay buffers are empty. It is interesting to see that by choosing $L_0 \leq B/2$, CRS maintains a lower delay than Max-Link.

Under the same setting, Fig. 5 investigates the impact of initial buffer length $L_0$ to the outage probability for various numbers of relays. We first note that the flat curves of OR and MMRS are due to the fact they do not consider the buffer status in relay selection. As shown, the outage probability curves for both HRS and CRS reveal a valley shape, because a small (a large respectively) $L_0$ increases the risk of an empty (a full respectively) buffer. In either case, both schemes will switch to use OR that is less efficient in exploiting the diversity gain as HRS and CRS. Furthermore, CRS only needs $L_0 = 2$ to minimize the outage probability for both $K = 3$ and 4. However, the minimum outage probability of HRS occurs when $L_0 = 50$, that is, half of the buffer is full initially. Hence, we can conclude that although filling more packets to the relay buffer initially can slightly reduce the outage probability, the associated price in terms of average delay is notable as seen from the linear curve in Fig. 4.

### C. Outage Probability vs. Average Delay

Thus far, we have shown that the average delay of HRS and CRS is a function of the initial buffer length $L_0$, which can be chosen to minimize the outage probability. To further study the interaction between the outage probability and the average delay, we compare the delay cost for minimizing the outage probability of different schemes. In Fig. 6, we consider $K = 3$ relays with SNR = 5 dB and 15 dB, respectively. The $y$-axis can be interpreted as the delay cost for achieving a certain outage probability. The results of OR and MMRS are also included as the benchmark.

As can be seen, the decrease of the outage probability results in the exponential increase of the average delay. Comparing to HRS, the delay cost of CRS is much lower. For example, HRS incurs the average delay of 60 slots to achieve the same minimum outage probability as MMRS at SNR = 5 dB while the average delay of CRS is only 6 slots, corresponding to 90% delay reduction under the same condition. At SNR = 15 dB, the delay reduction of CRS is about 94% over HRS. From the discussion, we confirm the superiority of CRS over existing buffer-aided RS schemes in terms of both reliability and average delay.


D. Impact of I.N.D. Fading

The above results are obtained for i.i.d. fading channels. We further extend our discussion to the case with i.n.d. fading based on simulations. Following [20], we consider $K = 3$ and set $\sigma^2_{g1} = 1\text{ dB}, \sigma^2_{g2} = 0\text{ dB}, \sigma^2_{g3} = -0.5\text{ dB}, \sigma^2_{h1} = 0\text{ dB}, \sigma^2_{h2} = 0.5\text{ dB,} \sigma^2_{h3} = -1\text{ dB, and SNR} = 15\text{ dB.}$

Our result coincides with that in [20, Fig. 11], which shows that there exists a larger performance gap between HRS and the ideal MMRS for i.n.d. fading (c.f. Fig. 5). On the contrary, CRS can still minimize the outage probability with small relay buffers under the i.n.d. fading. We omit the results for other SNR values because they reveal the same trend.

E. Impact of Outdated CSI

All the above results are obtained by assuming perfect CSI for RS. In general, the fading coefficient may vary with time and thus the CSI at the measurement instant may differ from that at the transmission instant. The impact of outdated CSI on the outage performance of CRS is investigated and compared with the benchmark scheme OR in the following using simulations. Let $\hat{h}$ denote the outdated channel coefficient. Based on the Jakes’ scattering model, the discrepancy between $|h|^2$ and $|\hat{h}|^2$ can be measured by the power correlation coefficient given by $\rho = J_0^2(2\pi f_d \tau)$, where $J_0(\cdot)$ denotes the zero-order Bessel function of the first kind, $f_d$ represents the maximum Doppler frequency shift [25], and $\tau$ is the feedback delay. A smaller $\rho$ implies that the outdated CSI is more deviated from the actual version, and $\rho = 1$ corresponds to the ideal case of perfect CSI. The method proposed in [26] is used to generate two correlated fading envelopes. For simplicity, both S-R and R-D links have the same correlation coefficient.

Fig. 8 illustrates the outage probability of CRS, HRS, and Max-Link versus $\rho$ with varied correlation coefficients $\rho$ for $K = 3$ relays, $B = 100, L_0 = B/2,$ and SNR $= 10\text{ dB and} 15\text{ dB, respectively. Under the present setting and perfect CSI ($\rho = 1$), CRS performs identically to HRS, as shown in Fig. 3. With outdated CSI ($\rho < 1$), all the considered RS scheme suffer a significant performance loss, which is more pronounced at higher SNR. When $\rho$ is smaller, corresponding to the more deviated CSI scenario, the increasing trend of outage probability in CRS is minor compared to HRS, primarily because choosing relays based on buffer status adopted in CRS attenuates the impact of deviated CSI.

VII. Conclusion

For cooperative networks with multiple relays, this paper addressed the problems of existing buffer-aided relaying schemes that require large relay buffers and a long initialization process. The proposed scheme, referred to CRS, takes into account both CSI and buffer status when making the relay selection decision. Performance analysis is conducted to assess the outage probability and diversity gain of CRS assuming i.i.d. fading and accurate CSI. Through extensive numerical results, we show that CRS achieves full diversity and requires only small relay buffers to perform as good as the CSI-based counterpart that assumes infinite buffer. Additionally, CRS significantly reduces the delay cost for minimizing the outage probability. We also investigated the performance of CRS under i.n.d. fading and outdated CSI, and the results show that CRS still outperforms other benchmark schemes in the i.n.d. case but it is vulnerable to inaccurate CSI. Future work include the modification of the proposed scheme to mimic full-duplex relays, the use of quantized feedback to reduce feedback overhead, and the extension to heterogeneous setting where relay buffers are not identical in size.
APPENDIX A

STEADY-STATE PROBABILITIES

Generally the steady-state probabilities of a MC can be obtained by solving a set of linear equations \( P \pi^T = \pi^T \) and normalizing \( \pi \) to the nullspace. Due to the symmetry of the considered MC and the assumption of i.i.d. fading, we have \( \pi_1 = \pi_5 \) and \( \pi_2 = \pi_4 \). Therefore, the system of equations \( P \pi^T = \pi^T \) can be simplified as

\[
\begin{bmatrix}
C_0 & C_2 & 0 \\
C_1 & C_3 & C_0 \\
0 & 2C_1 & C_1
\end{bmatrix}
\begin{bmatrix}
\pi_1 \\
\pi_2 \\
\pi_3
\end{bmatrix}
= \begin{bmatrix}
\pi_1 \\
\pi_2 \\
\pi_3
\end{bmatrix}. 
\tag{A.1}
\]

Since \( A \) is singular, we need another independent equation to solve \( x \). By replacing the last row in \( A \) with the normalization equation \( \sum_{i=1}^{5} \pi_i = 1 \), (A.1) can be reduced to

\[
\begin{bmatrix}
C_0 - 1 & C_2 & 0 \\
C_1 & C_3 - 1 & C_0 \\
2 & 2 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}. 
\tag{A.2}
\]

Now \( x \) can be solved as \( x = \tilde{A}^{-1}b \). After some straightforward algebra, we obtain the steady-state probabilities as shown in (17)–(19).

APPENDIX B

PROOF OF LEMMA 2

First, subtracting \( \pi_1 \) in (17) from \( \pi_2 \) in (18), we obtain

\[
\pi_1 - \pi_2 = \frac{C_1}{C_2 + 2C_1(C_2 + C_3)}(C_3 - C_2) \triangleq f_1(\gamma),
\tag{A.3}
\]

where \( C_3 - C_2 = \alpha_g^2\hat{\alpha}_h^2(\alpha_h^2 + \alpha_h - 1) < 0 \) because \( 0 < \alpha_g, \alpha_h < 1 \). Since \( C_1 \)'s are all positive and \( C_3 < C_2 \), we have \( f_1(\gamma) < 0 \) and thus \( \pi_1 < \pi_2 \). Likewise,

\[
\pi_2 - \pi_3 = \frac{C_2}{C_2 + 2C_1(C_2 + C_3)}(C_1 - C_2) \triangleq f_2(\gamma),
\tag{A.4}
\]

where \( C_1 - C_2 = \frac{1}{4}\alpha_g^2\hat{\alpha}_h^2(\alpha_g^2 + \alpha_h + \alpha_g \alpha_h - 4) < 0 \). Combining the above results, we have \( (\pi_1 = \pi_5 < (\pi_2 = \pi_4) < \pi_3 \), i.e, the MC \( \Psi \) tends to stay on state \( s_3 \). To understand how such a tendency interacts with SNR, we can verify that \( f_1(\gamma) \) is a monotonic decreasing function of \( \gamma \) by checking its first derivative with respect to \( \gamma \). Asymptotically, \( f_1(\gamma) \rightarrow -1/6 \) as \( \gamma \rightarrow \infty \). Similarly, \( f_2(\gamma) \) is strictly negative, monotonically decreasing, and approaches to \( -1/2 \) as \( \gamma \rightarrow \infty \). Since the MC \( \Psi \) is irreducible according to Lemma 1, its stationary distribution is unique and thus the above results hold regardless of the initial state.

REFERENCES


Syuan-Li Lin received the B.S. degree from Yuan Ze University, Taiwan, in 2011, and the M.S. degree from National Cheng Kung University, Taiwan, in 2013, both in electrical engineering. Since 2013, he has been an engineer with EDIMAX Technology Co., Ltd., Taiwan. His research focuses on cooperative communications and performance analysis for wireless communications systems.

Kuang-Hao (Stanley) Liu (S’06-M’08) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2008. From 2000 to 2002, he was an engineer with Siemens Telecom. System Ltd., Taiwan. From 2004-2008, he was a Research Assistant with the Broadband Communications Research (BBCR) Group at the University of Waterloo, Canada. He is currently an Associated Professor with the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan. His recent research focuses cover cooperative communications, energy-efficient communications, and small-cell networks.

Dr. Liu has been participated in organizing several international conferences, including Chinacom’09 and ’10, Wicon’10, IEEE PIMRC’12, and IEEE SmartGridComm’12. He has served as Technical Program Committee Member for many IEEE conferences, such as the IEEE International Conference on Communications and the IEEE Global Telecommunications Conference. He was the Guest Editor for the IET Communications (Special Issue on Secure Physical Layer Communications). He is now the Editor for the IEEE Wireless Communications Magazine. He is the recipient of the Best Paper Award from the IEEE Wireless Communications and Networking Conference (WCNC) 2010.