Achievable Rates for the Fading Three-Hop Half-Duplex Relay Network using Buffer-Aided Relaying

Hebatallah Shoukry∗, Nikola Zlatanov†, Vahid Jamali∗, and Robert Schober∗
∗ Friedrich-Alexander University (FAU), Erlangen, Germany
† University of British Columbia (UBC), Vancouver, Canada

Abstract—The fading three-hop half-duplex relay network consists of a source, two half-duplex relays, and a destination connected in series where links are present only between adjacent nodes. We assume that the links are impaired by time-continuous fading and additive white Gaussian noise. For this network, we design new protocols based on buffer-aided relaying and derive their achievable average rates. We first develop a buffer-aided protocol which maximizes the average rate, but, as a side effect, introduces unbounded delay. Therefore, we also design a buffer-aided protocol which constrains the average delay, but at the expense of decrease of rate. Our numerical results show that the maximum average rate achieved with the developed buffer-aided protocol is larger than that of existing protocols for the considered network. Moreover, given a sufficiently large permissible average delay, the average rate achieved with the buffer-aided protocol with a delay constraint approaches the maximum average rate achieved without a delay constraint.

I. INTRODUCTION

Future generation communication networks are expected to include some form of cooperative, relay-based communication. Employing relays in a network increases coverage and/or throughput and/or reliability. This phenomenon was first observed in [1], [2] for the simple three-node relay network comprised of a source, a full-duplex relay, and a destination. As the distance between the source and the destination increases, more than one relay have to be employed between the source and the destination in order to combat the large path loss caused by the long distance. A suitable network architecture that can be employed for such a scenario is the multi-hop relay network. The multi-hop relay network is comprised of a source, multiple relays connected in series, and a destination, where links are present only between adjacent nodes. As a result of this practical application scenario, multi-hop relaying has gained considerable research interest, see [3]-[9].

The relays in the multi-hop relay network can be either full-duplex or half-duplex. The capacity of the full-duplex multi-hop relay network is known, and can be derived using the degraded relay channel model presented in [2]. However, due to self-interference limitations, it is difficult to realize an ideal full-duplex relay in practice. As a result, half-duplex relaying is preferred for practical cooperative networks. However, the capacity for the multi-hop half-duplex relay network with additive white Gaussian noise (AWGN) is still unknown. To the best of our knowledge, the largest reported achievable rate for the AWGN half-duplex multi-hop relay network was given in [9]. This rate is achieved when the relays receive in one time slot and forward the received information in the following time slot. However, in fading environments, using relays which switch successively between reception and transmission degrades the achievable rate, as shown in [10] and [11] for the one-way and two-way three node relay channels, respectively. In [10] and [11], the authors showed that the achievable average rates can be significantly improved if the relay’s switching between reception and transmission is conditioned on the instantaneous qualities of the receiving and transmitting links in the network. Such conditional switching was first proposed in [12] and can be achieved using buffer-aided relaying which enables temporary storage of information at the relay’s buffer until the quality of the transmit link becomes strong enough for the relay to transmit the stored information.

The aim of this paper is to develop new achievable average rates for the AWGN fading three-hop half-duplex relay network. In particular, we design a buffer-aided protocol for the considered network that achieves a larger average rate than existing protocols. Furthermore, the developed methodology for designing a buffer-aided protocol for the three-hop relay network can serve as a starting point for designing a buffer-aided protocol for the general multi-hop half-duplex relay network employing an arbitrary number of relays.

Relaying with buffers for the multi-hop half-duplex relay network was also considered in [7]. This protocol was designed for uncoded and/or fixed rate transmission when the fading gains of the links in the network are independent and identically distributed (i.i.d.). In order to use this protocol as a benchmark, we modify the protocol in [7] such that when a node transmits, it transmits with the maximum possible rate given by the capacity of the underlying transmit channel. However, the modified protocol used as benchmark is only applicable to the case when all links are impaired by i.i.d. fading and causes data loss, due to buffer overflow, when the fading is independent and non-identically distributed (non-i.i.d.). We note, however, that this is not due to our modifications since the phenomenon of buffer overflow also occurs for the protocol in [7] for uncoded and fixed rate transmission when the links of the network are non-i.i.d.

This paper is organized as follows. In Section II, we introduce the system model. In Section III, we present the proposed buffer-aided protocols for the considered network. In Section IV, we provide some numerical examples. Finally, Section V concludes the paper.

II. SYSTEM MODEL

We consider a three-hop half-duplex relay network consisting of a source $S$, a destination $D$, and two decode-and-forward half-duplex relays $R_k$, where $k = 1, 2$, as shown in Fig. 1. In this network, because of high attenuation, the source, relay 1, and relay 2 have direct links only to relay 1, relay 2, and the destination, respectively. Hence, in order for the source to transmit a message to the destination, it first has to transmit the message to relay 1, which then forwards it to relay 2, which finally forwards it to the destination. We assume that the relays are equipped with unlimited size buffers in which they store the received information and from which they extract the transmit information. Due to the half-duplex constraint, the relays cannot receive and transmit at the same time.

For this network, we assume that the nodes transmit their codewords with fixed power $P$, and that all links are impaired by
whether in the adjacent links cannot be simultaneously selected for transmission.

Because of the half-duplex constraint of the relays, two adjacent links cannot be simultaneously selected for transmission. Otherwise, it would mean that a relay can receive and transmit at the same time. Mathematically, the half-duplex constraint of the relays for the considered network can be expressed as

\[ q_k(i) + q_{k+1}(i) \in \{0, 1\}, \quad k \in \{1, 2\} \]  

Using \( q_k(i) \), for \( k = 1, 2, 3 \), we now define the vector \( q(i) \) as

\[ q(i) = [q_1(i) \ q_2(i) \ q_3(i)] \]  

Given the constraints in (5), the vector \( q(i) \) can only take the following values

\[ q(i) = \begin{cases} 
0 & \text{if all nodes are silent} \\
1 & \text{if S transmits, R}_1 \text{ receives, R}_2 \text{ silent} \\
0 & \text{if S transmits, R}_1 \text{ silent, R}_2 \text{ receives} \\
0 & \text{if S and R}_1 \text{ are silent, R}_2 \text{ transmits} \\
1 & \text{if S and R}_2 \text{ transmit, R}_1 \text{ receives.} 
\end{cases} \]  

The different values of vector \( q(i) \) in (7) are referred to as the modes of transmission. Hence, for the considered network, only the five modes of transmission given in (7) are possible.

For presenting the protocol, in the following, we discuss the transmission rates of the individual nodes in the network, and the dynamics of the queues in the buffers of the relays. To this end, let \( Q_1(i) \) and \( Q_2(i) \) denote the number of bits/symbol in the buffers of relay 1 and relay 2, at the end of time slot \( i \).

**Source transmits:** This happens when \( q_1(i) = 1 \) and \( q_2(i) = 0 \). In this case, the source transmits to relay 1 a Gaussian distributed codeword with rate \( R_1(i) = C_1(i) \). Relay 1 can decode this codeword and stores the information in its buffer. With this transmission, \( Q_1(i) \) is updated as

\[ Q_1(i) = Q_1(i-1) + R_1(i) \]  

**Relay 1 transmits:** This happens when \( q_1(i) = [0 \ 1 \ 0] \). In this case, relay 1 transmits to relay 2 a Gaussian distributed codeword with rate

\[ R_2(i) = \min\{Q_1(i-1), C_2(i)\} \]  

Note that \( R_2(i) \) is limited by \( Q_1(i-1) \) since relay 1 cannot transmit more information than what it has already stored inside its buffer. Relay 2 can decode this codeword and stores the information in its buffer. For this transmission, \( Q_1(i) \) and \( Q_2(i) \) are updated as

\[ Q_1(i) = Q_1(i-1) - R_2(i) \]  

\[ Q_2(i) = Q_2(i-1) + R_2(i) \]  

**Relay 2 transmits:** This happens when \( q_2(i) = 0 \) and \( q_3(i) = 1 \). In this case, relay 2 transmits to the destination a Gaussian distributed codeword with rate

\[ R_3(i) = \min\{Q_2(i-1), C_3(i)\} \]

Similarly to (9), \( R_3(i) \) is limited by \( Q_2(i-1) \) since relay 2 cannot transmit more information than what it has inside its buffer. The destination can successfully decode this codeword. With this transmission, \( Q_2(i) \) is updated as

\[ Q_2(i) = Q_2(i-1) - R_3(i) \]  

Using \( q_k(i) \) and \( R_k(i) \), \( k \), we can obtain the average rate transmitted by the source and received at relay 1, denoted by \( R_1 \), the average rate transmitted by relay 1 and received at relay 2, denoted by \( R_2 \), and the average rate transmitted by relay 2 and received at the destination, denoted by \( R_3 \). These average rates are obtained as

\[ \bar{R}_1 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_1(i) \ R_1(i) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_1(i) \ C_1(i) \]  

\[ \bar{R}_k = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_k(i) \ R_k(i) \]  

\[ = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_k(i) \ \min \left\{ Q_{k-1}(i-1), C_k(i) \right\}, \quad k \in \{2, 3\} \]
where \( q_1(i), q_2(i), \) and, \( q_3(i) \) satisfy the constraint in (5), \( \forall i \).
Any choice of the values of \( q_1(i), q_2(i), \) and, \( q_3(i), \) \( \forall i, \) which satisfies (5), will produce an average rate. However, in order for an average rate to be achievable, i.e., data loss does not occur, the choice of \( q_1(i), q_2(i), \) and, \( q_3(i), \) \( \forall i, \) should not only satisfy the constraints in (5), but also result in stable buffers at the two relays. Moreover, from all achievable average rates produced by the choices of the values for \( q_k(i), \forall k, i, \) one rate is the largest. In order to find this maximum achievable average rate, we provide the following useful lemma.

**Lemma 1:** For the maximum achievable average rate the following has to hold

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_1(i) R_1(i) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_2(i) R_2(i)
\]

Furthermore, when (16) holds, \( R_2 \) and \( R_3 \), given by (15), simplify to

\[
R_k = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_k(i) C_k(i), \quad k \in \{2, 3\}. \tag{17}
\]

**Proof:** Let \( A_k \) be defined as

\[
A_k = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_k(i) R_k(i), \quad k \in \{1, 2, 3\}.\]

Then, the average rate transmitted by relay 1 is given by \( R_2 = \min\{A_1, A_2\} \). \( R_2 = A_1 \) when \( A_1 < A_2 \) is a result of the conservation of flow in the buffer, and \( R_2 = A_2 \) when \( A_1 > A_2 \) is due to an absorbing buffer. Similarly, the average rate transmitted by relay 2 is given by \( R_3 = \min\{A_2, A_3\} \).

Now, assume that \( A_3 < A_2 \). Then, we can always increase \( A_1 \), and thereby increase \( R_2 \), by switching \( q_2(i) \) from one to zero, and for the same \( i \), switch \( q_1(i) \) from zero to one. Similarly, if \( A_1 > A_2 \), we can always increase \( A_2 \), and thereby increase \( R_2 \), by switching \( q_1(i) \) from one to zero, and for the same \( i \), switch \( q_2(i) \) from zero to one. Hence, \( A_1 = A_2 \) is the optimal point for \( R_2 \). Similarly, it can be shown that \( A_2 = A_3 \) is the optimal point for \( R_3 \).

In order for no data loss to occur, the average rates transmitted by relay 1 and relay 2 have to be equal, i.e., \( R_2 = R_3 \) has to hold. Hence, for \( R_2 = R_3, A_1 = A_2 = A_3 \) is the optimal point which maximizes both \( R_2 \) and \( R_3 \), when \( R_2 = R_3 \) holds. Furthermore, since the rate transmitted by the source is \( A_1 \), which satisfies \( A_1 = A_2 = A_3 = R_2 = R_3 \), the information transmitted by the source arrives at the destination without data loss, i.e., the buffers are stable. Finally, the proof that (17) holds when (16) is satisfied, closely follows [10, Appendix B].

Lemma 1 is useful in two ways. First, it shrinks the search space for the maximum achievable average rate by providing constraint (16). Secondly, it significantly simplifies the expression for \( R_2 \) and \( R_3 \) since it removes the dependence on \( Q_1(i) \) and \( Q_2(i) \), respectively. Using Lemma 1, we now state a maximization problem for maximization of the achievable average rate.

The maximum achievable average rate is found as the solution to the following optimization problem for \( N \to \infty \)

Maximize : \( \frac{1}{N} \sum_{i=1}^{N} q(i) C_3(i) \)
Subject to : \( C_1 : \frac{1}{N} \sum_{i=1}^{N} q(i) C_4(i) \) \( = \frac{1}{N} \sum_{i=1}^{N} q(i) C_5(i) \), \( C_2 : \frac{1}{N} \sum_{i=1}^{N} q(i) C_6(i) \) \( = \frac{1}{N} \sum_{i=1}^{N} q(i) C_7(i) \), \( C_3 : q_k(i) \in \{0, 1\}, \forall k, i \)
\( C_4 : q_1(i) + q_2(i) \in \{0, 1\}, \forall i \)
\( C_5 : q_2(i) + q_3(i) \in \{0, 1\}, \forall i. \)

In (18), constraint (16) is captured by \( C_1 \) and \( C_2 \), respectively, whereas constraint (5) is captured by \( C_4 \) and \( C_5 \), respectively.

The maximization problem in (18) can be solved analytically. However, before providing the solution of (18), we first define the following variables. Let \( \mu_1 \) and \( \mu_2 \) be constants satisfying \( 0 \leq \mu_1 \leq 1, 0 \leq \mu_2 \leq 1, \) and \( \mu_1 > \mu_2 \). These constants depend on the average statistics of the fading of all links and will be determined later. Using \( \mu_1 \) and \( \mu_2 \), we define \( A_1(i), A_2(i), \) and \( A_3(i) \) as

\[
A_1(i) = (1 - \mu_1) C_4(i), \tag{19a}
A_2(i) = (\mu_1 - \mu_2) C_5(i), \tag{19b}
A_3(i) = \mu_2 C_6(i).
\tag{19c}
\]

Furthermore, we define a random variable as the outcome of a coin flip. In particular, we define \( C(i) \in \{0, 1\} \) as the outcomes of the coin flip in the \( i \)-th time slot. The probabilities of the possible outcomes of the coin flip are given by \( \Pr \{C(i) = 1\} = p \) and \( \Pr \{C(i) = 0\} = 1 - p \). We are now ready to present the buffer-aided protocol which maximizes the achievable average rate.

**Theorem 1:** The buffer-aided protocol which maximizes the achievable average rate for the considered network, found as the solution to (18), is given as follows.

**Case 1:** If \( \mu_1 \neq 1 \) and \( \mu_2 \neq 0 \), vector \( q(i) \) is given by

\[
q(i) = \begin{cases} 
0 & \text{if } A_1(i) + A_3(i) > A_2(i) \\
1 & \text{if } A_1(i) + A_3(i) < A_2(i) \\
1 & \text{if } A_1(i) + A_3(i) = A_2(i).
\end{cases}
\tag{20}
\]

**Case 2:** If \( \mu_1 = 1 \), vector \( q(i) \) is given by

\[
q(i) = \begin{cases} 
0 & \text{if } A_3(i) > A_2(i) \text{ AND } C(i) = 1 \\
1 & \text{if } A_3(i) < A_2(i) \text{ AND } C(i) = 0 \\
1 & \text{if } A_3(i) = A_2(i).
\end{cases}
\tag{21}
\]

**Case 3:** If \( \mu_2 = 0 \), vector \( q(i) \) is given by

\[
q(i) = \begin{cases} 
0 & \text{if } A_1(i) > A_2(i) \text{ AND } C(i) = 1 \\
1 & \text{if } A_1(i) < A_2(i) \text{ AND } C(i) = 0 \\
1 & \text{if } A_1(i) = A_2(i).
\end{cases}
\tag{22}
\]

The values of constants \( \mu_1 \) and \( \mu_2 \), and probability \( p \) are found such that constraints C1 and C2 in (18) hold with equality.

The maximum achievable average rate is obtained when the corresponding elements of \( q(i), \forall i \), according to (20), (21), and (22) for \( \mu_1 \neq 1 / \mu_2 \neq 0, \mu_1 = 1, \) and \( \mu_2 = 0, \) respectively, are inserted into (14) or (17).

**Proof:** We first convert the problem to a linear optimization problem by relaxing the binary constraints \( q_k(i) \in \{0, 1\} \) in (18) to \( 0 \leq q_k(i) \leq 1, \forall k, i \). Then, we solve the relaxed problem via its dual where \( \mu_1 \) and \( \mu_2 \) are the Lagrange multipliers for constraints C1 and C2, respectively. The solution of the relaxed
problem can be shown to be at the boundaries, i.e., 0 or 1. A detailed proof is omitted here due to lack of space.

Remark 1: The intuition behind the existence of three different protocols in Theorem 1 is the following. When the S-R1 and R2-D links have approximately the same strength on average, then \( \mu_1 \neq 1 \) and \( \mu_2 \neq 0 \), and only the protocol in Case 1 is applicable. However, when the S-R1 link is much stronger than the R2-D link on average, then \( \mu_1 = 1 \) and the protocol in Case 2 is applicable. On the other hand, when the S-R1 link is much weaker than the R2-D link on average, then \( \mu_2 = 0 \) and the protocol in Case 3 is applicable.

Remark 2: To derive the maximum average rate of a general multi-hop network comprised of \( M \) links, a maximization problem similar to the one in (18) has to formulated. In particular, let \( q_k(i) \) and \( C_k(i) \) denote the link indicator variable and the capacity of the \( k \)-th link in the \( i \)-th time slot in the network, respectively. Then, one has to maximize \( \sum_{i=1}^{N} q_k(i) C_k(i) / N \), given that \( \sum_{i=1}^{N} q_k(i) C_k(i) / N = \sum_{i=1}^{N} q_k+1(i) C_k(i) / N \) holds for \( k = 1, \ldots, M-1 \), where \( N \to \infty \). Moreover, (5) has to hold for \( k = 1, \ldots, M-1 \) and C3 in (18) has to hold for all \( k = 1, \ldots, M \). Due to lack of space, the derivation of the maximum rate and the corresponding buffer-aided relayin protocol for general multi-hop networks having an arbitrary number of relay nodes is left for the journal version of this paper.

Using Theorem 1, we can find analytical expressions for \( \mu_1 \), \( \mu_2 \), \( p \), and the maximum achievable rate. This is discussed in the following lemma where index \( i \) is dropped since, due to the stationarity and ergodicity of the fading, the statistics are identical \( \forall i \). However, due to lack of space, we only present expressions for \( \mu_1 \), \( \mu_2 \), and the maximum achievable rate for Case 1 in Theorem 1.

Lemma 2: Let \( f_\gamma(x) \) denote the probability density function (PDF) of \( \gamma_k \) for \( k \in \{1, 2, 3\} \). Then, for Case 1 in Theorem 1, \( \mu_1 \) and \( \mu_2 \) can be found by solving the following system of equations\(^1\)

\[
\begin{align*}
\int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_1) f_{\gamma_1}(x_1) f_{\gamma_2}(x_2) f_{\gamma_3}(x_3) dx_1 dx_2 dx_3 & = \int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_2) f_{\gamma_2}(x_2) f_{\gamma_1}(x_1) f_{\gamma_3}(x_3) dx_1 dx_2 dx_3 \\
& = \int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_3) f_{\gamma_3}(x_3) f_{\gamma_1}(x_1) f_{\gamma_2}(x_2) dx_1 dx_2 dx_3 \\
& = \int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_3) f_{\gamma_3}(x_3) f_{\gamma_2}(x_2) f_{\gamma_1}(x_1) dx_1 dx_2 dx_3 \\
\end{align*}
\]

(23)

\[
\begin{align*}
&= \int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_3) f_{\gamma_3}(x_3) f_{\gamma_2}(x_2) f_{\gamma_1}(x_1) dx_1 dx_2 dx_3 \\
&= \int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_3) f_{\gamma_3}(x_3) f_{\gamma_1}(x_1) f_{\gamma_2}(x_2) dx_1 dx_2 dx_3 \\
&= \int_0^\infty \int_0^\infty \int_0^\infty \log_2(1+x_3) f_{\gamma_3}(x_3) f_{\gamma_1}(x_1) f_{\gamma_2}(x_2) dx_1 dx_2 dx_3
\end{align*}
\]

(24)

where

\[
\begin{align*}
g_1(x_2, x_3) & = \frac{(1+x_3)(1+x_2)^{1-\mu_2}}{(1+x_3)^{1-\mu_2}} - 1 \\
g_2(x_1, x_3) & = \frac{(1+x_3)(1+x_1)^{1-\mu_1}}{(1+x_3)^{1-\mu_2}} - 1 \\
g_3(x_1, x_2) & = \frac{(1+x_2)(1+x_1)^{1-\mu_2}}{(1+x_2)^{1-\mu_2}} - 1.
\end{align*}
\]

(25)

(26)

(27)

Let \( \mu_1^{\ast} \) and \( \mu_2^{\ast} \) denote the solutions of the system of equations in (23) and (24). Then the maximum achievable rate is given by inserting \( \mu_1^{\ast} \) and \( \mu_2^{\ast} \) in the left hand side of (23) or (24), or by inserting \( \mu_1^{\ast} \) and \( \mu_2^{\ast} \) in the right hand side of (23) or (24).

Proof: Due to lack of space the proof is omitted here.

\(^{1}\)A system of nonlinear equations can be solved e.g. by algorithms based on Newton’s method [13].

We note that the protocol proposed in Theorem 1 introduces unbounded delay. In order to bound the delay, in the following, we propose a buffer-aided protocol for delay limited communications.

B. Buffer-Aided Protocol for Delay Constrained Transmission

To bound the average delay, we limit the size of the buffers at relay 1 and relay 2 to \( Q_1^{max} \) and \( Q_2^{max} \) bits/symbol, respectively. Then, the buffer-aided protocol for a limited delay is the same as the protocol in Theorem 1, however, instead of the variables \( \tilde{\Lambda}_1(i), \tilde{\Lambda}_2(i), \) and \( \tilde{\Lambda}_3(i) \), in Theorem 1, we use the new set of variables \( \tilde{\Lambda}_1(i), \tilde{\Lambda}_2(i), \) and \( \tilde{\Lambda}_3(i) \), respectively, given by

\[
\begin{align*}
\tilde{\Lambda}_1(i) & = (1 - \mu_1) \min \{ C_1(i), Q_1^{max} - Q_1(i-1) \} \quad \text{(28a)} \\
\tilde{\Lambda}_2(i) & = (1 - \mu_2) \min \{ C_2(i), Q_2^{max} - Q_2(i-1) \} \\
\tilde{\Lambda}_3(i) & = \mu_2 \min \{ C_3(i), Q_2(i-1) \},
\end{align*}
\]

(28b)

(28c)

where \( \mu_1, \mu_2, \) and \( p \) are the same as in Theorem 1. The average delay for this protocol can be controlled via the buffer sizes \( Q_1^{max} \) and \( Q_2^{max} \). The average rate achieved by this protocol is given by

\[
R_3 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_3(i) \min \{ C_3(i), Q_2(i-1) \}.
\]

(29)

The average delay that this protocol introduces, denoted by \( E[T(i)] \), is the sum of the average delays introduced by the buffers at relay 1 and relay 2, denoted by \( E[T_1(i)] \) and \( E[T_2(i)] \), respectively. On the other hand, the average delay introduced by the buffer at the \( k \)-th relay, for \( k = 1, 2 \), can be found using Little’s formula [14] as \( E[T_k(i)] = E(Q_k(i))/E(A_k(i)) \), where \( E(A_k(i)) \) is the average arrival rate in the buffer of the \( k \)-th relay, found as

\[
\begin{align*}
E(A_1(i)) = & \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_1(i) \min \{ C_1(i), Q_1^{max} - Q_1(i-1) \}. \\
E(A_2(i)) = & \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} q_2(i) \min \{ C_2(i), Q_2^{max} - Q_2(i-1) \}.
\end{align*}
\]

In the following, we discuss the implementation of the proposed protocols.

C. Implementation of the Proposed Protocols

In this subsection, we discuss the implementation of the protocol given in Theorem 1. The delay limited protocol, given in Section III-B, can be implemented in a similar manner.

For implementation of the proposed protocols we require channel state information (CSI) for rate adaptation at the transmitters, for coherent detection at the receivers, and for mode selection at a central node. In order to perform rate adaptation and coherent detection, all nodes need local CSI information. In particular, at the beginning of each time slot, each node has to acquire the CSI of its own transmit and receive channels. This requires four pilot symbol transmissions at the beginning of each time slot. One pilot symbol transmission from the source to enable relay 1 to acquire the CSI of the S-R1 link. A second pilot symbol transmission is required from relay 1 to enable the source and relay 2 to acquire the CSI of the S-R1 and R1-R2 links, respectively. The third pilot symbol transmission is made by relay 2 to enable relay 1 and the destination to acquire
the CSI of the $R_1$-$R_2$ and $R_2$-$D$ links, respectively. Finally, the fourth pilot symbol transmission is made by the destination and is used by relay 2 to acquire the CSI of the $R_2$-$D$ link. Using the acquired CSI, the source can compute $C_1(i)$, relay 1 can compute $C_1(i)$ and $C_2(i)$, relay 2 can compute $C_2(i)$ and $C_3(i)$, and the destination can compute $C_3(i)$.

For mode selection, we choose relay 1 to be the central node in the network, although other nodes may assume this role as well. In order for relay 1 to perform mode selection, an additional feedback is required from relay 2 to relay 1 to enable relay 1 to acquire the CSI of the $R_2$-$D$ link, and thereby, be able to compute $C_3(i)$. Moreover, in order for the central node to perform mode selection, constants $\mu_1$ and $\mu_2$, and probability $p$ have to be acquired at the central node. These constants can be computed using Lemma 2, but this requires knowledge of the PDFs of the fading gains of all links before the start of transmission. Such a priori knowledge may not be available in practice. Therefore, the central node has to estimate these parameters in real-time using only the CSI knowledge until time slot $i$. Since $\mu_1$ and $\mu_2$ are actually Lagrange multipliers required for solving the linear optimization problem in (18), and since the probability $p$ is a parameter that ensures that the average rates $R_1$ and $R_3$ are equal, an accurate estimate of these parameters can be obtained using the gradient descent method [15]. In particular, using $C_k(i)$, for $k \in \{1, 2, 3\}$, relay 1 computes an estimate of $\mu_k$, for $k \in \{1, 2\}$, denoted by $\hat{\mu}_k(i)$, and an estimate of $p$, denoted by $\hat{p}(i)$, using Algorithm 1. In Algorithm 1, $\delta(i)$ is an adaptive step size which controls the speed of convergence of $\hat{\mu}_k(i)$ to $\mu_k$ and $\hat{p}(i)$ to $p$. The function $\delta(i)$ is properly chosen monotonically decreasing for $\delta(i)$ with $\delta(1) < 0$. Furthermore, since $\mu_1 > \mu_2$ has to hold, we initialize $\hat{\mu}_1(1) = 0.7$ and $\hat{\mu}_2(1) = 0.2$.

Once the transmission mode $q(i)$ has been computed, relay 1 broadcasts the transmission mode to source and relay 2. Only then begins the transmission of information for time slot $i$ according to the computed transmission mode.

### IV. Numerical Results

In this chapter, we present numerical results when all links in the network undergo Rayleigh fading, where $\Omega_1$, $\Omega_2$, and $\Omega_3$ denote the average powers of the fading of the $S$-$R_1$, $R_1$-$R_2$, and $R_2$-$D$ links, respectively.

As benchmark, we use the achievable rate given in [9, Eq. (41)]. This average rate is achieved when each relay switches between reception and transmission successively. In this case, two achievable rates are possible depending on the delay. The first achievable rate introduces unbounded delay and is given by

$$R_{\text{conv}, 1} = \min \left\{ \frac{C_1 C_2}{C_1 + C_2}, \frac{C_2 C_3}{C_2 + C_3} \right\},$$

(30)

where $C_k = E[C_k(i)]$, for $k \in \{1, 2, 3\}$ and $E[\cdot]$ denotes expectation. The second rate introduces a delay of two time slots and is given by

$$R_{\text{conv}, 2} = E \left\{ \min \left\{ \frac{C_2(i) C_3(i)}{C_1(i) + C_2(i)}, \frac{C_3(i) C_1(i)}{C_2(i) + C_3(i)} \right\} \right\}.$$  

(31)

We note that $R_{\text{conv}, 1} \geq R_{\text{conv}, 2}$. Furthermore, achieving both rates requires full CSI at all nodes.

Another benchmark scheme that we will use is our modification of the protocol proposed in [7]. This protocol was proposed for fixed rate transmission. Hence, in order to use it as a benchmark, we modify this protocol such that each node transmits with a rate equal to the capacity of its underlying transmit channel. In the protocol in [7], only the link with the highest SNR is activated in each time slot. Hence, the achieved average rate is

$$R_{\text{single-link}} = \frac{1}{3} E \left\{ \max\{C_1(i), C_2(i), C_3(i)\} \right\}.$$  

(32)

This protocol also introduces unbounded delay and requires full CSI at all nodes. Furthermore, this protocol is mainly applicable to the case when all links are affected by i.i.d. fading and will cause data loss due to buffer overflow, for non-i.i.d. This phenomenon occurs both for fixed and adaptive rate transmission.

In Figs. 2 and 3, we plot the average rates achieved with the buffer-aided protocols with unbounded delay, given in Theorem 1, and with bounded delay, given in Section III-B. Fig. 2 shows the average rates achieved for i.i.d. fading with $\Omega_1 = \Omega_2 = \Omega_3 = 1$, whereas Fig. 3 shows the average rates achieved for non-i.i.d. fading with $\Omega_1 = 0.5$, $\Omega_2 = 0.3$, and $\Omega_3 = 0.75$. In both figures, the average rates achieved with the proposed buffer-aided protocols are obtained via simulations, where $\hat{\mu}_1(i)$, $\hat{\mu}_2(i)$, and $\hat{p}(i)$ are found using the estimation method given in Algorithm 1 with $\delta(i) = 0.2/i^{0.6}$. These parameters are also plotted in Fig. 4 as function of time for $P/\sigma_n^2 = 0$ dB and

### Algorithm 1 Computation of $\mu_1(i)$, $\mu_2(i)$, $p'(i)$, and $q(i)$

1. Initialize $q(1) = [1 \ 0 \ 0]$, $\mu_1(1) = 0.7$, $\mu_2(1) = 0.2$, $p'(1) = 0.5$, and $R_k(0) = 0$ for $k \in \{1, 2, 3\}$.

2. For time slot $i \geq 2$ do:

   a. $R_k(i) = \frac{i-2}{i-1} R_k(i-2) + \frac{q(i-1)}{i-1} C_k(i-1)$, for $k \in \{1, 2, 3\}$

   b. $\mu_k(i) = \mu_k(i-1) + \delta(i)(R_k(i-1) - \bar{R}_k(i-1))$, for $k \in \{1, 2\}$

   c. $p'(i) = p'(i-1) + \delta(i)(R_k(i-1) - \bar{R}_k(i-1))$

   d. $p'(i) = \max\{0, \min\{1, p'(i)\}\}$

   e. $C(i) =$ outcome of the coin flip with $p'(i) = \Pr\{C(i) = 1\}$

   f. if $\mu_1(i) \leq 1$ and $\mu_2(i) \geq 0$ then

      i. $q(i) = [1 \ 0 \ 1]$

   g. else

   h. $q(i) = [0 \ 1 \ 0]$

   i. end if

j. end if

k. if $\mu_1(i) > 1$ then

   l. $\mu_1(i) = 1$

   m. if $\Lambda_3(i) > \Lambda_2(i)$ and $C(i) = 1$ then

   n. $q(i) = [1 \ 0 \ 1]$

   o. else if $\Lambda_3(i) > \Lambda_2(i)$ and $C(i) = 0$ then

   p. $q(i) = [0 \ 0 \ 1]$

   q. else

   r. $q(i) = [0 \ 1 \ 0]$

   s. end if

   t. end if

u. if $\mu_2(i) < 0$ then

v. $\mu_2(i) = 0$

w. if $\Lambda_1(i) > \Lambda_2(i)$ and $C(i) = 1$ then

x. $q(i) = [1 \ 0 \ 1]$

y. else if $\Lambda_1(i) > \Lambda_2(i)$ and $C(i) = 0$ then

z. $q(i) = [1 \ 0 \ 0]$

[...]

[...]

[...]

...
there is no curve for $p$ buffer-aided relaying. We presented two protocols, one which fading three-hop half-duplex relay network, achievable with all benchmark rates. With the buffer-aided relaying for $10$ average rate achieved with buffer-aided relaying protocol with $10$ rate. Moreover, the average rate achieved with the buffer-aided protocol with unbounded delay achieves the largest average fading this protocol produces data loss. Cannot use the protocol in [7] as benchmark since for non-i.i.d. channels.

As benchmark, in Figs. 2 and 3, we also plot the average rates obtained from Theorem 1, and Lemma 2. As can be seen, the simulated average rate coincides perfectly with the theoretical average rate.

As benchmark, in Figs. 2 and 3, we also plot the average rates achieved with conventional relaying, cf. (30) and (31). Moreover, in Fig. 2, we have also plotted the average rate achieved with the modified protocol from [7], cf. (32). We note that in Fig. 3 we cannot use the protocol in [7] as benchmark since for non-i.i.d. fading this protocol produces data loss.

Figs. 2 and 3 show that the proposed buffer-aided relaying protocol with unbounded delay achieves the largest average rate. Moreover, the average rate achieved with the buffer-aided relaying protocol with $10$ time slots delay closely approaches the average rate achieved with buffer-aided relaying protocol with unbounded delay. More importantly, the average rate achieved with the buffer-aided relaying for $10$ time slots delay outperforms all benchmark rates.

V. CONCLUSION

In this paper, we presented new average rates for the AWGN fading three-hop half-duplex relay network, achievable with buffer-aided relaying. We presented two protocols, one which maximizes the average rate but introduces unbounded delay, and a second protocol which bounds the delay. Our numerical results show that the maximum average rate achieved with the developed buffer-aided protocol is larger than the average rate achieved with existing protocols for the considered network. Furthermore, the developed methodology for designing the buffer-aided protocol for the three-hop relay network can serve as a starting point for designing a buffer-aided protocol for the general multi-hop half-duplex relay network employing arbitrary number of relays, as explained in Remark 2.

REFERENCES