Cooperative ARQ with Relay Selection: An Analytical Framework Using Semi-Markov Processes

Nikolaj Marchenko, *Student Member, IEEE,* and Christian Bettstetter, *Senior Member, IEEE*

***Abstract*—An important building block in cooperative diversity is relay selection, which has to ensure that a well-suited node is employed as relay. The required coordination among nodes causes signaling overhead, which, in turn, can significantly devalue performance benefits gained by cooperative diversity. A relay update policy defines when a new relay is selected; it can balance the tradeoff between performance and overhead.**

This tradeoff is studied using mathematical methods. We consider three relay selection schemes: *permanent*, *reactive*, and *adaptive*, which have different relay update rules. We develop an analytical framework using semi-Markov processes to evaluate the throughput and energy efficiency of cooperative Automatic Repeat reQuest (ARQ) protocols in time-correlated multipath fading channels. Results reveal potential performance gains of different selection schemes under various conditions.

The reactive and adaptive schemes make use of better-suited relays due to frequent selections. If their selection overhead, however, is significant, a permanent relay can achieve higher throughput due to negligible overhead. The impact of temporal correlation of fading channels on throughput and energy efficiency is also shown. These insights can be applied for development of cooperative communication protocols.

***Index Terms*—Cooperative communication, relay selection, ARQ, semi-Markov process, fading.**

I. Introduction

WIRELESS communications in multipath propagation environments suffers from fading. During a deep fade, correct data reception becomes impossible due to very low signal power at the destination. In particular, in slow fading channels, a communication link can remain in outage for long time periods in the order of seconds [1].

Cooperative diversity [2] is a retransmission technique aiming to decrease link outage rates by utilizing a relay node that overhears source-destination transmissions and forwards received messages to the destination. In this way, the destination gains additional signal diversity since it has the possibility to obtain the same message via different paths. It is basically a form of spatial and temporal diversity.

Manuscript submitted August 01, 2012; revised January 09, 2013 and March 04, 2013; accepted April 14, 2013. Date of current version June 07, 2013.

This work was performed as part of the project RELAY of Lakeside Labs. It was funded by the ERDF, KWF, and BABEG under grant 20214/15935/23108.

Authors are with the Mobile Systems Group, Institute of Networked and Embedded Systems, University of Klagenfurt, 9020 Klagenfurt, Austria (e-mail: firstname.lastname@aau.at). Christian Bettstetter is also with Lakeside Labs GmbH, 9020 Klagenfurt, Austria.

The relay selection procedure plays a critical role in resulting performance benefits of cooperative retransmissions. One relay [3] or multiple relays [4] should be selected from a set of potential candidates and assigned to overhear and retransmit messages to the destination. It was shown in [5] that already cooperative diversity protocols with a single preassigned relay are superior in outage probability compared to non-cooperative source-destination links.

A. Aspects of Relay Selection

Let us briefly discuss some design choices for relaying.

1) Which metrics determine the optimal relay candidate? Channel quality estimations between source, destination, and potential relays are the most important and most used metrics as they determine successful message delivery to the destination [3]. Further aspects, such as residual energy of the nodes [6], [7] or spatial efficiency [8] can also be included as selection metrics to optimize network performance.

2) How are nodes coordinated? Most relay selection proposals are contention-based, i.e., surrounding nodes nominate themselves in a distributed manner as relays either using timers [9] or transmitting in a slotted contention window [10], [11]. The particular message exchange depends on application goals and the wireless technology.

3) When is relay selection performed? Relay selection can be triggered by special events, e.g. failed packets, expired timers. Various update rules can be used to select an optimal relay with low selection overhead.

The article at hand puts emphasis on the analysis of the third aspect — the impact of relay update policy and its timing on throughput and energy efficiency. Moreover, our analysis focuses on incremental relaying, i.e., relaying is performed when the destination is unable to decode the data sent by the source directly. Such incremental relaying is basically a cooperative automatic repeat request (ARQ) protocol [12] if no signal combining [13] is employed. It is assumed here that only one node at a time is selected to act as relay.

Three common selection schemes are compared:

1) *Permanent* selection: a relay is selected for a long period of time (at least several magnitudes longer than the duration of a data message).
2) *Reactive* selection: a relay is selected anew each time the destination fails to receive a data message from the source directly [3].
3) *Adaptive* selection: a relay is selected anew each time the destination fails to receive a data message, i.e., neither

the source nor the currently active relay could deliver the message.

B. Contributions

We propose an analytical framework using semi-Markov processes [14] to evaluate performance of cooperative ARQ protocols in time-correlated fading channels. This framework provides expected throughput and energy efficiency of relaying protocols taking into account relay selection overhead and energy required for transmitting and receiving data messages. Results are derived for a one-dimensional grid network with Rayleigh fading. They illustrate the tradeoff between throughput and selection overhead with reactive and adaptive selection. The throughput gain achieved through selection diversity can be diminished if selection delay is nonnegligible and relay updates are triggered frequently. We also study the impact of time correlation of fading on throughput and energy efficiency, and derive close-form throughput expressions for two channel correlation bounds (quasi-static and i.i.d. channels).

We treat this topic in a systematic manner using well-defined analytical methods. Although the analysis is limited to three selection schemes, the proposed framework is flexible enough to be extended to suit other cooperative retransmission schemes, e.g., proactive selection [3]. The presented comparison yields novel insight into the relay selection process and can be used in the development of cooperative protocols.

C. Organization

Section II gives an overview of related work. Section III describes cooperative ARQ with permanent, reactive, and adaptive relay selection. Section IV introduces modeling assumptions used by the framework and performance analysis. Section V explains the framework based on semi-Markov processes. Section VI discusses the throughput and energy efficiency results of the three cooperative ARQ schemes.

II. Related Work

Zhao et al. [15] evaluate the performance of a cooperative H-ARQ protocol in line networks and compare it with traditional multi-hop point-to-point transmissions. Authors compare three selection schemes where a node is selected as a relay based on the instantaneous Signal-to-Noise Ratio (SNR), average SNR, or randomly. They show potential benefits in throughput, energy, and latency efficiency. However, timing of relay selection, required selection overhead, and channel time correlation are not taken into account.

Dianti et al. [16] investigate a cooperative ARQ scheme where several permanently selected relay nodes can simultaneously retransmit data using distributed space-time codes (DSTC) if the source fails to deliver the message to the destination directly. The authors consider time-correlated Rayleigh fading channels using Markov chains to model their cooperative ARQ scheme and obtain results for throughput and delay performance. Mahitan et al. [17] also use Markov models to model a cooperative ARQ protocol where a preassigned relay r always retransmits source messages to the destination as long as r is able to decode it. The authors consider error-correlated Nakagami-m fading, and do not take into account any relay selection aspects. The authors of [18] also assume correlated Nakagami-m fading, and derive guidelines for relay selection and optimal power allocation. The authors of [19] consider time-orthogonal transmissions for cooperative ARQ, but assume only permanent relay selection.

Yu et al. [12] study cooperative ARQ protocols with reactive relay selection based on feedback from destination. The resulting packet error rate after single retransmission is presented without consideration of the relay selection overhead.

Madan et al. [20] analyze energy efficiency of cooperative relaying with various relay selection rules. They take into account the energy required for signaling and derive an optimal selection rule to maximize overall energy per message. Shah et al. [21] analyze the tradeoff between selection duration and resulting throughput and energy benefits from cooperative transmission. In contrast to incremental relaying, they assume that a relay is selected after the source transmission and always retransmits data to the destination. It is shown that selection overhead can significantly decrease benefits of cooperation.

The particular aspect of relay update rules in cooperative transmissions has been addressed in [22] and [23]. In both articles the active relay is changed when the resulting SNR at the destination down-crosses certain SNR threshold. The resulting switching rates versus the SNR threshold and number of potential relays are obtained and presented.

In this article, we take a different approach: we investigate various relay update schemes within a unified analytical framework based on the use of semi-Markov processes. Furthermore, we investigate the resulting throughput and energy efficiency of cooperative protocols, and discuss the tradeoff between throughput and relay selection overhead. The current article extends our previous work on cooperative relay selection published in preliminary form in [24]–[26].

III. Cooperative ARQ with Relay Selection

This section explains cooperative ARQ with permanent, reactive, and adaptive relay selection. Each protocol may use a contention procedure to choose a single node from a node set; solutions for this are mentioned at the end of this section.

A. Permanent Relay Selection

In permanent relay selection, a relay is selected once and serves as a single relay for a period of time at least several magnitudes longer than the duration of a data message.

After the relay selection, the source s can send DATA messages to the destination d and the selected relay r. If d receives the message correctly, it sends a positive acknowledgment (ACK), and a new DATA transmission can begin. The relay r retransmits DATA only if it has received it correctly and d has not. The retransmission can be triggered explicitly by a negative acknowledgment (NACK) from d, or implicitly if an ACK is missing. For simplicity of analysis, we assume that r retransmits DATA again and again until d receives it. If neither r nor d receive DATA from s, s retransmits it.

The selected relay is intended to assist on many s-d transmissions. Therefore, certain long-term characteristics should be employed to select the best-suited relay. For the purpose of our study, the expected SNR values of the s-r and r-d channels are reasonable and sufficient. The selected relay should be statistically most capable of receiving messages from s and delivering them successfully to d.

The signaling overhead can be neglected in comparison to number of DATA messages sent over the cooperative link.

B. Reactive Relay Selection

In reactive relay selection, s broadcasts a DATA message to d and all nodes surrounding s. Relay selection takes place after each failed s-d transmission. A node is selected to become the relay for a failed DATA message if it has a correct copy of the message and a sufficiently good channel to deliver DATA to d. The channel state information is obtained through a NACK from d, which in turn triggers a contention procedure. If more than one node receives DATA from s correctly, the node with the best channel to d (if good enough) should be preferred. The chosen node then delivers the DATA message to d. If no candidates are available for relaying, s retransmits.

Since all nodes overhear direct transmissions, the advantage of reactive selection is in the usage of selection diversity at each failed packet.

C. Adaptive Relay Selection

Compared to reactive selection, a new relay is selected if both possible paths (i.e., s-d and s-r-d) fail. If there is currently no assigned relay, s first broadcasts a short relay request (RREQ) indicating that relay selection is starting. Then it transmits a new DATA message to d and surrounding nodes. The destination sends a short acknowledgment (ACK or NACK) to the relay candidates that allows them to evaluate the channel. A node that has received DATA from s and has a sufficiently good channel to d can be selected as relay. If multiple nodes fulfill this requirement, the node with the best long-term channel characteristics such as for the permanent selection should be preferred.

After a relay r is selected through the contention procedure, it remains an assigned relay as long as the cooperative link remains good, i.e., as long as d receives DATA messages either from s or r. If both d and r are unable to decode DATA, or if d fails to receive the forwarded DATA message from r, s broadcasts an RREQ and retransmits the failed DATA. Neighboring nodes receiving this DATA message and the corresponding ACK from d participate in a new relay contention.

D. Node Contention Procedure

The objective of node contention is to identify, in a distributed manner, a single node out of a set of candidate nodes and to assign it as relay r to a given s-d pair. The selection criteria in general depend on the network application. In this work, we use the expected and instantaneous SNR values between r, s, and d.

Two classes of node contention are commonly used: timer-based and slot-based contention. Using timer-based contention [9], each node that fulfills the requirements to become a relay sets a timer. Upon expiration of a timer, a node listens to the channel and, if it is idle, sends a reply message (RREP) (for permanent selection) or starts retransmitting the DATA message itself (for reactive selection). A given timer function maps local channel information (or other selection metrics) in such a way that nodes with better metrics transmit first [9].

Using slot-based contention [10], [11], the contention window is divided into time slots of fixed duration. Based on local information, a relay candidate randomly selects a time slot and transmits its RREP message in it. The receiver can collect nominations from multiple candidates and choose a node (or several nodes) with the best characteristics. In both methods, the selection is typically triggered by a request message RREQ sent either from s or d, depending on the protocol implementation. The decoding of this message by other nodes is the starting point for timers or the contention window, respectively. In this way a local synchronization among competing potential relays is guaranteed.

Nevertheless, collisions of messages can occur due to hidden terminals. There are various ways to maximize the contention success: using channel listening and random congestion backoff times, increasing contention window size, or repeating contentions. We assume that contention is always successful if there is at least one node during contention that fulfills the selection requirements. The intention here is to leave out implementation-specific details and keep the analytical framework generic and mathematical analysis more comprehensible. In spite of that the presented analytical framework can be extended to consider imperfect contention, which highly depends on the particular implementation and topology.

Numerous variations of these two contention methods can be developed to improve contention efficiency for particular network setups and applications [8], [11], [27].

IV. MODELING ASSUMPTIONS

A. Protocol Assumptions

The following assumptions are made on the operation of cooperative ARQ protocols with all three relay selection schemes described earlier in Section III:

- All transmissions are orthogonal in time.
- All nodes use the same transmission rate and power.
- All DATA messages have same duration T.
- Signaling messages for relay selection and acknowledgments are error-free.
- Relays operate in decode-and-forward mode [15].
- Receivers perform selection combining on the message level [13]. Energy accumulation from different transmissions is not possible.
- A relay contention results in the selection of an optimal available relay candidate according to the selection requirements of a particular selection scheme.
- The selection overhead is the time interval T_w needed for a relay selection procedure. Typically, it consists of a contention window and a number of implementation-specific coordination messages from source and destination. We assume that this time remains constant for all

three schemes. If a relay is not selected after time T_w, the source transmits data without an assisting cooperative relay.

The ratio of the relay selection time to the data transmission time is $w = T_w/T$. The duration of other signaling messages is either ignored or included in the DATA message duration.

- Energy for a DATA message transmission is E_{tx}. At the receiver, energy is used only when a data message is received correctly. The corresponding energy per message is E_{rx}. If the channel is bad, the receiver can detect it at the beginning of the message and stop receiving to save energy.
- Energy consumption during relay contention is not considered since it heavily depends on particular implementation and network setup. However, our analytical framework can be extended to include this energy.

B. Radio Channel

We consider symmetrical wireless links with time-correlated block fading. Time is divided into slots indexed by $k \in \mathbb{N}$ of duration T during which the signal level is constant. We assume that T is also the transmission time of a DATA message.

The SNR between nodes i and j over time is represented as a series of SNR samples $\{\gamma_{ij}(k)\}$. If the current SNR is higher than the decoding threshold, $\gamma_{ij}(k) > \gamma_{\min}$, the channel is in the *good* state, and can receive a DATA message without errors. Otherwise, it is in the *bad* state, i.e., an outage event occurs, thus the DATA cannot be decoded by the receiver.

A binary random process $\{c_{ij}(k)\}$ describes the channel states between nodes i and j over time:

$$c_{ij}(k) = \begin{cases} \text{“Good” (G)}, & \gamma_{ij}(k) \geq \gamma_{\min}, \\ \text{“Bad” (B)}, & \gamma_{ij}(k) < \gamma_{\min}. \end{cases} \tag{1}$$

Generally, the process can be time-correlated, and we can model it as a two-state Markov chain [28], [29]. The corresponding transition probability matrix of the channel states,

$$\mathbf{C}_{ij} = \begin{bmatrix} \Pr(\text{G}|\text{G})_{ij} & \Pr(\text{B}|\text{G})_{ij} \\ \Pr(\text{G}|\text{B})_{ij} & \Pr(\text{B}|\text{B})_{ij} \end{bmatrix}, \tag{2}$$

defines the channel behavior. Here, $\Pr(b|a)_{ij}$, $a, b \in \{\text{G}, \text{B}\}$, is the probability that the next channel state is $c_{ij}(k+1) = b$ given that the current channel state is $c_{ij}(k) = a$.

The approach of [28], [30] is applied to obtain $\mathbf{C}_{ij}$ for Nakagami-m fading channels with given fading margin ψ_{ij}, Doppler spread f_D, and message duration T. Fading is considered as slow if $f_D T < 0.1$ and fast if $f_D T > 0.2$ [28].

The fading margin ψ_{ij} characterizes the received signal power in relation to the receiver SNR threshold,

$$\psi_{ij} = \frac{\overline{\gamma}_{ij}}{\gamma_{\min}}. \tag{3}$$

The term $\overline{\gamma}_{ij}$ denotes the expected SNR at the receiver and is calculated according to a simple pathloss model by

$$\overline{\gamma}_{ij} = \frac{p_i}{p_n} \left(\frac{\Delta_{ij}}{\Delta_0} \right)^{-\alpha}, \tag{4}$$

where p_i is the transmission power of node i, p_n is the noise power, Δ_{ij} is the distance between nodes, Δ_0 is a reference distance, and α is the pathloss exponent. Note that these values are linear and not in dB.

Since the results in this article are calculated for Rayleigh fading, we provide here only the outage probability for this special case of Nakagami-m fading ($m = 1$):

$$\varepsilon_{ij} = \Pr[\gamma_{ij} < \overline{\gamma}_{\min}] = 1 - \exp\left(-\frac{1}{\psi_{ij}} \right). \tag{5}$$

For detailed information on Nakagami-m fading see [31].

If a conventional Stop-and-Wait (SW) ARQ protocol is employed on such a channel, i.e., s keeps retransmitting a data message until it is received by d, with negligible and error-free feedback, the resulting normalized throughput at the receiver is $\eta = 1 - \varepsilon_{sd}$, which does not depend on channel time correlation [32].

V. Cooperative ARQ as a Semi-Markov Process

A network consists of a source s, destination d, and N surrounding nodes indexed by $n \in \{1, 2, \ldots, N\}$. The following notation is used to describe a cooperative ARQ protocol:

1) $c_{ij}(k)$ is the radio channel state between two nodes $i, j \in \{s, 1, 2, \ldots, N, d\}$. The channel behavior is defined by the channel state transition probability matrix $\mathbf{C}_{ij}$ as discussed in Section IV-B.
2) $Y = \{y_1, y_2, \ldots, y_L\}$ is a set of L operational states of a particular cooperative ARQ protocol. E.g., a protocol state can be a transmission of a new message by s, retransmission by relay $r \in \{1, \ldots, N\}$, or relay selection procedure. The detailed description of protocol states for three considered cooperative ARQ schemes is provided later in this section.
3) $y(k) \in Y$ is the protocol state at time slot k. Similar to a radio channel, the protocol states over time can be presented as random process $\{y(k)\}$.
4) Tuple $\mathbf{z}(k)$ includes the protocol state and channel states at given time,
$$\begin{aligned} \mathbf{z}(k) = \Big(& y(k), c_{sd}(k), c_{s1}(k), c_{1d}(k), \\ & c_{s2}(k), c_{2d}(k), \ldots, c_{sN}(k), c_{Nd}(k) \Big). \end{aligned} \tag{6}$$
Here, radio channels between nodes $n \in \{1, \ldots, N\}$ are not included since communication between potential relays is not considered in the proposed cooperative ARQ protocols.
5) Z is the set of all permitted unique tuples $\mathbf{z}(k)$ for a given protocol. The size of the set is
$$|Z| = L \cdot 2^{2N+1}. \tag{7}$$
In cases when the tuple set size becomes too large to handle, boundary cases have to be used as described later in this section.
6) Function $f : Z \to Y$ defines the protocol state transition from $y(k)$ to $y(k+1)$, which depends on the current channel states in the network and the protocol state.

Each tuple $\mathbf{z} \in Z$ can be seen as a state of a Markov chain incorporating protocol and channel transitions. The transition from tuple $\mathbf{z}_a$ to tuple $\mathbf{z}_b$ (both $\in Z$; $a, b \in \{1, \ldots, |Z|\}$) in one time step is possible only when $y^{(b)} = f(\mathbf{z}_a)$, and $y^{(b)} \in \mathbf{z}_b$, i.e., the protocol state of the next tuple is the same as defined by the function f for the current tuple $\mathbf{z}_a$. The transition probability is defined by the channel state transitions in $\mathbf{z}_a$ to ones in $\mathbf{z}_b$. The transition probability matrix $\mathbf{P}$ contains the probabilities of transitions between the tuples. Its elements are calculated by

$$P_{ab} = \begin{cases} \Pr(c_{sd}^{(b)}|c_{sd}^{(a)}) \prod_{n=1}^{N} \Pr(c_{sn}^{(b)}|c_{sn}^{(a)}) \Pr(c_{nd}^{(b)}|c_{nd}^{(a)}) \\ \quad \text{for } y^{(b)} = f(\mathbf{z}_a), \\ 0 \quad \text{otherwise}, \end{cases} \tag{8}$$

where $c_{ij}^{(a)}$ is the corresponding channel state between nodes i and j in tuple $\mathbf{z}_a$. Channel state transition probabilities are obtained from $\mathbf{C}_{sd}$, $\mathbf{C}_{sn}$, and $\mathbf{C}_{nd}$.

Vector $\boldsymbol{\pi} = \begin{bmatrix} \pi_1 \ \pi_2 \ \cdots \ \pi_{|Z|} \end{bmatrix}$ contains the limiting-state probabilities of the defined Markov process, i.e., element π_a is the probability that in its steady state after numerous transitions the Markov process will be in state $\mathbf{z}_a$.

If the Markov chain is irreducible and aperiodic, $\boldsymbol{\pi}$ can be obtained by solving the following set of linear equations:

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi} \quad \text{with} \sum_{a=1}^{|Z|} \pi_a = 1. \tag{9}$$

In general, before making the transition from state $\mathbf{z}_a$ to $\mathbf{z}_b$ the protocol waits for a holding time H_{ab}. If this time is equal for all state transitions, the process is considered Markov. If H_{ab} varies for some pairs $(\mathbf{z}_a, \mathbf{z}_b)$, or it has some random distribution, the system is semi-Markov and is defined by two matrices: transition probability matrix $\mathbf{P}$ of the embedded Markov chain and holding time matrix $\mathbf{H}$.

To consider the relay selection overhead we set different holding times to different transitions. We define the semi-Markov processes later for each relay selection scheme.

Next, we assign a delivery reward $X_{ab} = 1$ to any transition from tuple $\mathbf{z}_a$ to tuple $\mathbf{z}_b$ that results in a successful packet delivery to the destination. Otherwise the reward is set to 0. The cumulative reward of the process at time τ is called reward function $X(\tau)$. In the long term, $X(\tau)/\tau$ corresponds to the *normalized throughput* of the protocol and is calculated according to the fundamental renewal-reward theorem [33] by

$$\eta = \lim_{\tau \to \infty} \frac{X(\tau)}{\tau} = \frac{\sum_{a=1}^{|Z|} \pi_a \sum_{b=1}^{|Z|} P_{ab} X_{ab}}{\sum_{a=1}^{|Z|} \pi_a \sum_{b=1}^{|Z|} P_{ab} H_{ab}}. \tag{10}$$

Here, in the numerator, the inner sum $\sum_{b=1}^{|Z|} P_{ab} X_{ab}$ is the expected reward (delivered packets) gained by transitions starting in state $\mathbf{z}_a$. In the denominator, the inner sum $\sum_{b=1}^{|Z|} P_{ab} H_{ab}$ is the corresponding expected waiting time in the state $\mathbf{z}_a$ before a transition. The outer sums provide the expected reward and waiting time of the whole semi-Markov process in the steady state. More detailed explanations can be found in [33].

In a similar way, we can assign energy rewards E_{ab}, i.e., energy consumed for data transmission and reception during the state transition from $\mathbf{z}_a$ to $\mathbf{z}_b$. The *expected energy per delivered packet* in the long run can be calculated similar to (10) with additional division by throughput η,

$$\xi = \frac{1}{\eta} \lim_{\tau \to \infty} \frac{E(\tau)}{\tau} = \frac{\sum_{a=1}^{|Z|} \pi_a \sum_{b=1}^{|Z|} P_{ab} E_{ab}}{\sum_{a=1}^{|Z|} \pi_a \sum_{b=1}^{|Z|} P_{ab} X_{ab}}, \tag{11}$$

and is independent from holding times $\mathbf{H}$ and overhead w.

The computational complexity of using this analytical framework basically corresponds to the complexity of solving the system of linear equations (9). It varies from $O(n^3)$ floating operations for a dense matrix to $O(n)$ for a sparse matrix [34], where n in our case equals $|Z|$.

A. Limiting Cases of Time-Correlated Channels

Two channel time correlation boundaries can be used to simplify the analysis of the protocol performance: a) independent and identically distributed (*i.i.d.*) channels, and b) *quasi-static* channels.

In an i.i.d. channel, the next state of the channel between nodes i and j does not depend on the current state and is defined solely by the error rate ε_{ij}. The corresponding channel transition probability matrix is simply

$$\mathbf{C}_{ij} = \begin{bmatrix} 1 - \varepsilon_{ij} & \varepsilon_{ij} \\ 1 - \varepsilon_{ij} & \varepsilon_{ij} \end{bmatrix}. \tag{12}$$

If each channel is considered to be i.i.d., the system Markov chain can be drastically reduced to the number of protocol states so that $|Z| = |Y| = L$. The transition probabilities from protocol state $y^{(a)}$ to state $y^{(b)}$ can still be calculated by (8). Thus, taking into account (12), the resulting probabilities are independent of the current channel states. The corresponding throughput and energy efficiency are calculated by (10) and (11).

A quasi-static channel is the limiting case when $f_D T \to 0$, and, as a result, the corresponding channel transition probability matrix approaches its limit

$$\lim_{f_D T \to 0} \mathbf{C}_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{13}$$

To calculate the throughput at this limit, we need to identify all state transitions within our semi-Markov model that can take place when channel states do not change. This means that transitions between tuples in Z become deterministic. Therefore, the transitions between tuples that lead to changes of the channel states can be ignored. Taking this into account, we calculate the expected reward $\widetilde{X}$ on possible tuple transitions and overall mean time between transitions $\widetilde{H}$ when the semi-Markov process is in steady state. The resulting throughput boundary is then

$$\lim_{f_D T \to 0} \eta = \frac{\widetilde{X}}{\widetilde{H}}. \tag{14}$$

As we show later, the throughput in such channels can be derived as closed-form expression.

The throughput of time-correlated channels with $0 < f_D T < 0.35$ always lies between the throughput of these two bounds. Therefore, the bounds can be used to assess protocol throughput without extended calculations of full semi-Markov models.

B. Permanent Relay Selection

In cooperative ARQ with a permanent relay, a relay node is selected to assist the transmission for a significantly long time. After a relay $r \in \{1, \ldots, N\}$ has been selected as described in Section III-A, the cooperative ARQ protocol can be in one of the following states:

Tx: s transmits a message to d and r. Depending on whether the previous message was delivered successfully, it can be a new message transmission or a retransmission of the failed packet.

R: r relays the source message to d.

Fig. 1 shows protocol states and transitions between them.

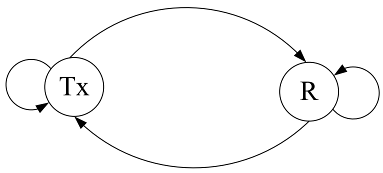


Fig. 1. Markov chain for cooperative relaying with permanent relay selection

Only s-d, s-r, and r-d radio channels are needed to model the cooperative ARQ protocol operation. The set Z contains all valid combinations for the quadruple $\mathbf{z}_a = (y^{(a)}, c_{sd}^{(a)}, c_{sr}^{(a)}, c_{rd}^{(a)})$. In total, there are $|Z|= 16$ unique tuples that cover all possible state transitions in the systems. The function $y^{(b)} = f(\mathbf{z}_a)$ describing protocol state transitions of cooperative ARQ with a permanent relay can be written as:

$$y^{(b)} = \begin{cases} \text{Tx} & \text{for } y^{(a)} = \text{R},\ c_{rd}^{(a)} = \text{G}, \\ & \text{or } y^{(a)} = \text{Tx},\ c_{sd}^{(a)} = \text{G}, \\ & \text{or } y^{(a)} = \text{Tx},\ c_{sd}^{(a)} = \text{B}, c_{sr}^{(a)} = \text{B}, \\ \text{R} & \text{for } y^{(a)} = \text{R},\ c_{rd}^{(a)} = \text{B}, \\ & \text{or } y^{(a)} = \text{Tx},\ c_{sd}^{(a)} = \text{B},\ c_{sr}^{(a)} = \text{G}. \end{cases} \tag{15}$$

State transition probabilities from tuple $\mathbf{z}_a$ to tuple $\mathbf{z}_b$ are obtained according to (8).

Whenever a packet is successfully delivered to d, the protocol returns to state Tx. We assign a reward X_{ab} $(a, b \in \{1, 2, \ldots, 16\})$ in the following way

$$X_{ab} = \begin{cases} 1 & \text{for } y^{(a)} = \text{Tx}, c_{sd}^{(a)} = \text{G}, \\ & \text{or } y^{(a)} = \text{R}, c_{rd}^{(a)} = \text{G}, \\ 0 & \text{otherwise}. \end{cases} \tag{16}$$

The holding time is the same for each state transition and corresponds to the duration of a single message transmission, which we normalize to one. Since the duration of relay selection can be neglected in the long run, the resulting throughput is calculated by

$$\eta = \sum_{a=1}^{16} \pi_a \sum_{b=1}^{16} P_{ab} X_{ab}. \tag{17}$$

If all channels are i.i.d., the Markov process describing tuple transition can be reduced to the chain in Fig. 1 with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 1 - \varepsilon_{sd} + \varepsilon_{sd}\varepsilon_{sr} & \varepsilon_{sd}(1 - \varepsilon_{sr}) \\ 1 - \varepsilon_{rd} & \varepsilon_{rd} \end{bmatrix}. \tag{18}$$

The resulting throughput η is obtained by solving (9) and (10) and can be written as a closed-form expression

$$\eta = \Pr(\text{Tx}) = \pi_1 = \frac{1 + \varepsilon_{sd}\varepsilon_{sr}\varepsilon_{rd} - \varepsilon_{sd}\varepsilon_{sr} - \varepsilon_{rd}}{1 + \varepsilon_{sd} - \varepsilon_{sd}\varepsilon_{sr} - \varepsilon_{rd}}. \tag{19}$$

If all channels are approaching static states, throughput reward $X_{ab} = 1$ is earned only when a) the s-d channel is good; or b) the s-d channel is bad AND both the s-r and r-d channels are good. In the second case, reward $X_{ab} = 1$ is assigned only when a protocol transition R→Tx takes place, which makes up half of all transitions. Since all holding times are the same, we have $\widetilde{H} = 1$. Therefore, the resulting limit for the throughput is

$$\lim_{f_D T \to 0} \eta = \widetilde{X} = 1 - \varepsilon_{sd} + 0.5\varepsilon_{sd}(1 - \varepsilon_{sr})(1 - \varepsilon_{rd}). \tag{20}$$

To shorten our next expressions, we use the indicator function for channel state c_{ij}:

$$\mathbb{1}_{\text{G}}(c_{ij}) = \begin{cases} 1, & c_{ij} = \text{G}, \\ 0, & c_{ij} = \text{B}. \end{cases} \tag{21}$$

Using this indicator function, the energy consumed at state transition $\mathbf{z}_a \to \mathbf{z}_b$ is

$$E_{ab} = \begin{cases} E_{\text{tx}} + E_{\text{rx}}\Big(\mathbb{1}_{\text{G}}(c_{sd}^{(a)}) + \mathbb{1}_{\text{G}}(c_{sr}^{(a)})\Big) \\ \qquad\qquad\qquad\qquad \text{for } y^{(a)} = \text{Tx}, \\ E_{\text{tx}} + \mathbb{1}_{\text{G}}(c_{rd}^{(a)})E_{\text{rx}} \quad \text{for } y^{(a)} = \text{R}. \end{cases} \tag{22}$$

The corresponding energy efficiency per delivered packet is calculated according to (11).

C. Reactive Relay Selection

In reactive relay selection, all nodes are listening to data transmissions originated from s. A new relay selection takes place after a direct s-d transmission fails.

A node n is an available candidate during selection procedure when: a) it receives the message from s (i.e., the current s-n channel state is good), AND b) currently its channel to d is also good. If a node fulfills the selection requirements it can always deliver the message to the destination. It is thus not important as to which node out of the set of available candidates is chosen. To simplify the calculations, however, we assume that the nodes are sorted in order of preference, and a node with the lowest index in the candidate set is selected. This manipulation does not have any impact on the resulting throughput and energy efficiency of the protocol.

Cooperative ARQ with reactive relay selection is described by the same underlying Markov chain as cooperative ARQ with a permanent relay in Fig. 1. However, the protocol states have a different meaning:

Tx: s transmits a message. If the previous message was not delivered and no relay was selected, s retransmits the same message again. If the message was successfully delivered, a new message is transmitted.

R: A relay has been selected and delivers the message to d.

Since channels to multiple potential relays are considered now, the size of set Z with valid tuples according to (7) becomes $|Z| = 2^{2N+2}$. Tuple $\mathbf{z}_a$ in (6) is modified in a way that $c_{rd}^{(a)}$, $r \in \{1, 2, \ldots, N\}$, corresponds to the channel state in the subsequent step $k+1$. It is a valid manipulation since the s-r and r-d channels are independent from each other. The state of the r-d channel is obtained during the relay selection after each failed direct transmission and assumed to not change for the message transmission after the selection. The transitions between protocol states are defined as follows:

$$y^{(b)} = \begin{cases} \text{R} & \text{for } y^{(a)} = \text{Tx}, c_{sd}^{(a)} = \text{B}, \\ & \quad \sum_{r=1}^{N} \mathbb{1}_{\text{G}}(c_{sr}^{(a)}) \mathbb{1}_{\text{G}}(c_{rd}^{(a)}) \geq 1, \\ \text{Tx} & \text{for } y^{(a)} = \text{R}, \\ & \text{or } y^{(a)} = \text{Tx}, c_{sd}^{(a)} = \text{G}, \\ & \text{or } y^{(a)} = \text{Tx}, c_{sd}^{(a)} = \text{B}, \\ & \quad \sum_{r=1}^{N} \mathbb{1}_{\text{G}}(c_{sr}^{(a)}) \mathbb{1}_{\text{G}}(c_{rd}^{(a)}) = 0. \end{cases} \tag{23}$$

The corresponding system state transition probabilities are calculated according to (8).

When a direct s-d transmission fails, the holding time of the process consists of the DATA message duration and the time of relay selection overhead. If a direct transmission succeeds, the holding time equals only the data message duration. The elements of the holding time matrix $\mathbf{H}$ are

$$H_{ab} = \begin{cases} 1 + w & \text{for } y^{(a)} = \text{Tx}, c_{sd}^{(a)} = \text{B}, \\ 1 & \text{otherwise.} \end{cases} \tag{24}$$

The resulting throughput is calculated according to (10).

For $N \to \infty$, the throughput approaches

$$\lim_{N \to \infty} \eta = 1 - \frac{1+w}{2+w} \varepsilon_{sd}, \tag{25}$$

as the selection of a relay is always possible. The consumed energy per delivered message, however, goes to infinity for $E_{\text{rx}} > 0$, since infinitely many nodes overhear the message.

Next, we obtain the throughput when the time correlation of the channels approaches the quasi-static bound. First, we define a variable ε_R which is the probability that no node satisfies the relay selection criteria,

$$\varepsilon_R = \prod_{n=1}^{N} \Big(1 - (1 - \varepsilon_{sn})(1 - \varepsilon_{nd})\Big). \tag{26}$$

Instantaneous channel knowledge becomes irrelevant in a quasi-static environment. The corresponding expected rewards are assigned in a similar manner as for permanent relay selection (20), but instead of a single relay state there are multiple relaying states that can be combined:

$$\widetilde{X} = 1 - \varepsilon_{sd} + 0.5\varepsilon_{sd}(1 - \varepsilon_R) = 1 - 0.5\varepsilon_{sd}(1 + \varepsilon_R). \tag{27}$$

To calculate the expected holding time between tuple state transitions, the probability of a state is multiplied with the time spent in this state before the transition assuming quasi-static channel states:

$$\begin{aligned} \widetilde{H} &= 1 - \varepsilon_{sd} + 0.5\varepsilon_{sd}(1 - \varepsilon_R) \\ &\quad + 0.5\varepsilon_{sd}(1 - \varepsilon_R)(1 + w) + \varepsilon_{sd}\varepsilon_R(1 + w) \\ &= 1 + 0.5\varepsilon_{sd}(1 + \varepsilon_R)w. \end{aligned} \tag{28}$$

The resulting throughput in quasi-static channels approaches

$$\lim_{f_D T \to 0} \eta = \frac{\widetilde{X}}{\widetilde{H}} = \frac{1 - 0.5\varepsilon_{sd}(1 + \varepsilon_R)}{1 + 0.5\varepsilon_{sd}(1 + \varepsilon_R)w}. \tag{29}$$

Similar to permanent relay selection, for each transition from $\mathbf{z}_a$ to $\mathbf{z}_b$ we assign energy rewards:

$$E_{ab} = \begin{cases} E_{\text{tx}} + E_{\text{rx}} \Big(\mathbb{1}_{\text{G}}(c_{sd}^{(a)}) + \sum_{r=1}^{N} \mathbb{1}_{\text{G}}(c_{sr}^{(a)}) \Big) \\ \qquad \text{for } y^{(a)} = \text{Tx}, \\ E_{\text{tx}} + E_{\text{rx}} \quad \text{for } y^{(a)} = \text{R}, y^{(b)} = \text{Tx}, \\ 0 \qquad \text{otherwise.} \end{cases} \tag{30}$$

The resulting average energy consumed per delivered message is calculated according to (11).

D. Adaptive Relay Selection

As explained in Section III-C, adaptive relay selection is triggered when not only the direct transmission (as in reactive selection) but also the relay retransmission fails. The selected node remains active relay until the cooperative link fails again. Similar to reactive selection, node n is an available relay candidate, if during the selection it has the message from s and currently a good channel to d. However, in addition, during relay contention, it reports its expected SNR values on the s-n and d-n channels, as in permanent selection. A candidate node that provides the most reliable relaying path is preferred. Based on the received expected SNR values, s can estimate the most suitable relay node.

Without loss of generality, but for simplicity of calculation, we assume here that an index is assigned to each node to reflect the reliability of a two-hop path through this node. As in reactive relay selection, if multiple nodes fulfill selection requirements, the one with the lowest index is selected. This index is just used for analysis but is not required in the real protocol implementation.

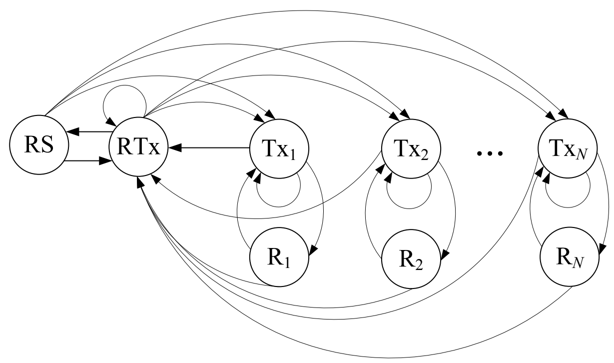


Fig. 2. Protocol states and transitions for cooperative ARQ with adaptive relay selection.

Fig. 2 shows the corresponding $L = 2+2N$ protocol states and transitions between them.

Tx_r: s transmits a new message. Node $r \in \{1, 2, \ldots, N\}$ is assigned as a cooperative relay.

R_r: The current relay r forwards the message to d if it receives the message from s but d does not.

RTx: The state can play two roles depending on the state transition resulting in RTx.

1) First role of RTx state: s retransmits the *same* message itself and requests a new relay selection. This situation occurs in the following cases:
   a) s can deliver the message neither to d nor to the currently assigned relay r ($\mathrm{Tx}_r \rightarrow \mathrm{RTx}$)
   b) The current relay r receives the message from s but can not deliver it to d after a failed s-d transmission ($\mathrm{R}_r \rightarrow \mathrm{RTx}$)
   c) If the s-d pair does not have an assigned relay yet, s transmits a data message and requests a new relay selection. If it fails to deliver the message to d, and the relay selection does not provide any relay, the source needs to transmit the same message again and requests a new selection. This takes place after $\mathrm{RS} \rightarrow \mathrm{RTx}$ transition.
2) Second role of RTx state: s transmits a *new* message and indicates the need for a relay. Assuming the s-d pair has no assigned relay, it can happen sometimes that the source delivers the message to d directly without selecting a relay. This corresponds to the transition $\mathrm{RTx} \rightarrow \mathrm{RTx}$.

If the source delivers the message to d and successively a relay node is assigned, the protocol state changes to the corresponding Tx_r state ($\mathrm{RTx} \rightarrow \mathrm{Tx}_r$).

RS: A relay selection procedure is performed when there is currently no assigned relay and s was unable to deliver the message directly to d ($\mathrm{RTx} \rightarrow \mathrm{RS}$). If a relay is selected successfully, it delivers the message to d, and the protocol moves from state RS to the corresponding state Tx_r. A new message transmission can start. If the relay cannot be selected, the protocol returns to the state RTx, s retransmits the message and starts relay selection anew.

For the purpose of better readability, formal definitions of the protocol state transitions together with corresponding holding time matrix $\mathbf{H}$ and energy reward matrix $\mathbf{E}$ are omitted here and collected in the Appendix. The corresponding transition probability matrix $\mathbf{P}$ is calculated according to (10).

The throughput reward of one is assigned to transitions resulting in a successful message delivery to d:

$$X_{ab} = \begin{cases} 1 & \text{for } y^{(b)} = \mathrm{Tx}_r, \\ & \text{or } y^{(a)} = y^{(b)} = \mathrm{RT}, \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

The resulting throughput and energy per delivered data message are calculated according to (10), and (11), respectively.

Similar to reactive relay selection instantaneous channel knowledge becomes irrelevant in a quasi-static environment. The mean reward per transition is calculated in the same way as in (27). The expected holding time between transitions is

$$\begin{aligned} \widetilde{H} &= (1-\varepsilon_{sd})(1-\varepsilon_R) + \varepsilon_{sd}\,(1-\varepsilon_R) \\ &\quad + (1-\varepsilon_R)(1-\varepsilon_{sd})\varepsilon_R \\ &\quad + (1-\varepsilon_{sd})\varepsilon_R^2(1+w) + \varepsilon_{sd}\varepsilon_R(1+w) \\ &= 1 + \varepsilon_R(\varepsilon_{sd} + \varepsilon_R\,(1-\varepsilon_{sd})\,)w, \end{aligned} \tag{32}$$

and the throughput when all channels approach quasi-static states is

$$\lim_{f_D T \rightarrow 0} \eta = \frac{\widetilde{X}}{\widetilde{H}} = \frac{1 - 0.5\varepsilon_{sd}(1+\varepsilon_R)}{1 + \varepsilon_R(\varepsilon_{sd} + \varepsilon_R\,(1-\varepsilon_{sd})\,)w}. \tag{33}$$

VI. Performance Analysis

A. Network Scenario

Our framework can be used for performance analysis of arbitrary network topologies. In this article, we evaluate performance in linear network topologies. Networks in many transportation or production systems can be modeled as one-dimensional networks [35]. Similar modeling is also performed in [15] for studying cooperative Hybrid-ARQ in practical relay networks. Despite the topological simplicity, a linear network still enables us to apprehend distinctively the qualitative differences among the relay selection schemes in all considered aspects. Performance analysis with a two-dimensional or three-dimensional node placement would not necessarily give additional insight in the protocol behavior.

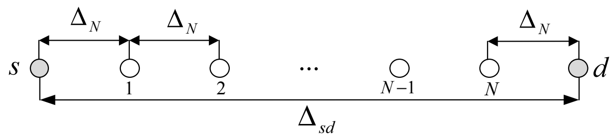


Fig. 3. Network topology.

Fig. 3 shows the used topology. There are N nodes located between source and destination at equal distances $\Delta_N = \Delta_{sd}/(N+1)$. These nodes can overhear s-d communication if necessary and act as relays.

The pathloss exponent α is 3, and, for the sake of simplicity, we assume that all communication channels experience the same time correlation. All radio channels experience Rayleigh block fading. The corresponding channel state transition matrices are obtained according to [28].

The relay position that maximizes end-to-end delivery ratio clearly depends on the channel characteristics. For simplicity we assume here that the optimal position is the middle point between s and d. A relay at this position provides near-optimal throughput performance in our scenario [24] and is straightforward for network setup.

Since permanent relay selection always provides the node that is closest to the midpoint between s and d, the performance of cooperative ARQ is determined by the availability of such a node. To allow better comparison among schemes, all plotted results of cooperative relaying with permanent relay are calculated for a relay in the midpoint.

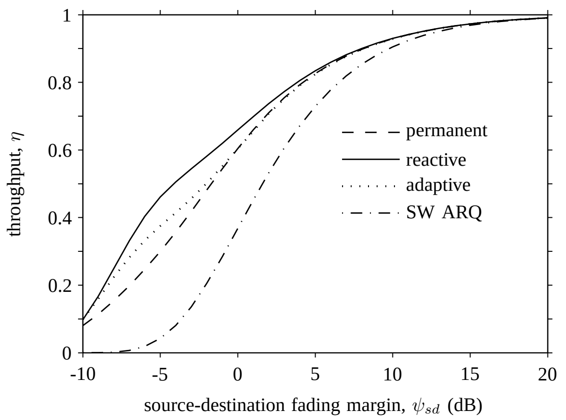


Fig. 4. Throughput η as a function of source-destination fading margin ψ_{sd}. Number of potential relays $N = 5$, channel time correlation $f_DT = 0.1$, selection overhead $w = 0$.

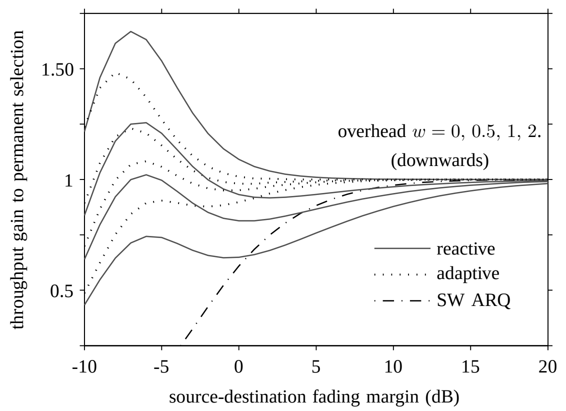


Fig. 5. Throughput gain to permanent relay selection as a function of source-destination fading margin ψ_{sd} and selection overhead w. Number of potential relays $N = 5$, channel time correlation $f_DT = 0.1$.

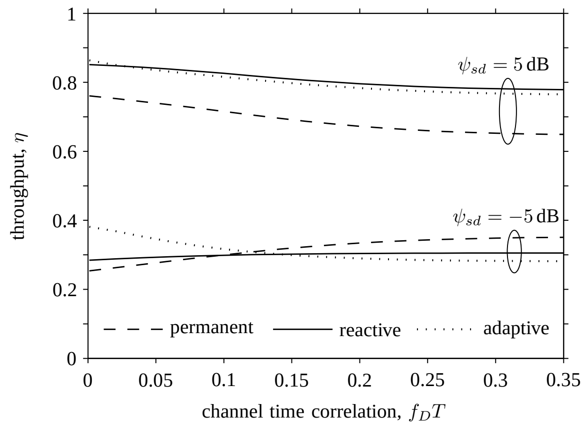


Fig. 6. Throughput η versus channel time correlation: $f_DT \to 0-$ quasi-static channels, $f_DT \approx 0.35-$ i.i.d. channels. Number of potential relays $N = 5$, selection overhead $w = 1$.

B. Throughput

Fig. 4 shows the throughput η versus the s-d fading margin ψ_{sd} for cooperative ARQ and SW ARQ when selection overhead is neglected, $w = 0$, and the number of intermediate nodes is $N = 5$. All cooperative schemes perform better than SW ARQ. Reactive relay selection provides the highest throughput, since all N nodes overhear source transmissions, and in case of packet decoding failure at d, there is a higher probability of a successful relay retransmission. Adaptive selection outperforms permanent selection for $\psi_{sd} < 0$ dB for the same reasons. However, when a relay is selected, all nodes except the selected relay ignore s-d transmissions, and in case the cooperative link fails, a retransmission by s and a new relay selection is triggered. Therefore, the throughput for adaptive selection becomes lower than that of reactive relaying. For $\psi_{sd} > 5$ dB all schemes provide nearly the same throughput, since the relay selection and relay transmission are almost always successful at such channel conditions.

However, when selection overhead becomes larger, throughput performance changes significantly. We take the throughput of cooperative relaying with permanent relay selection as a baseline which is independent of w. Fig. 5 shows the ratio to this baseline of two other relay selection schemes with different selection overhead w.

The throughput of reactive relaying (solid lines) significantly suffers from the selection overhead. At some conditions, e.g., $w = 2$, $\psi_{sd} > 0$ dB, it is even lower than the throughput of non-cooperative SW ARQ. Throughput of cooperative ARQ with adaptive relay selection (dotted lines) is decreasing with increase of w as well. However, the impact of the overhead is smaller than that of reactive relaying. We can see that adaptive relaying always outperforms the reactive one for $w \geq 1$. Finally, we observe that cooperative ARQ with permanent relay selection (the gain equals one), which uses only one preassigned relay, can perform better than other selection schemes that employ selection diversity among multiple relays but require additional selection overhead.

Next, we study the impact of channel time correlation on throughput η. Fig. 6 shows the throughput for $f_DT \in (0, 0.35]$. As explained in Section IV-B, $f_DT \to 0$ corresponds to a quasi-static environment, where channel states do not change. $f_DT \approx 0.35$ corresponds to time-uncorrelated channels when the next channel state does not depend on the current state. Channel correlation can result in a difference of throughput performance from 10% to 35%. At $\psi_{sd} = 5$ dB, all selection schemes perform better in slower fading channels.

At s-d margin $\psi_{sd} = -5$ dB most transmissions require a retransmission by the relay. The s-r and r-d channels are now more prone to errors. As a result, in fast fading channels and given $N = 5$ relays, reactive and adaptive schemes often cannot select any relay since they require both s-r and r-d channels to be good. The probability can be improved by higher N, with the limiting case of $N \to \infty$, when a suitable relay node can always be found. Permanent relay selection, however, allows the selected relay to retransmit data multiple times until the message is delivered to d. Together with zero selection overhead, this results in higher throughput. At slow fading channels, the channels to potential relays remain rather constant, and adaptive relay selection provides best throughput, since it makes use of various available relay nodes, but keeps selection overhead at minimum.

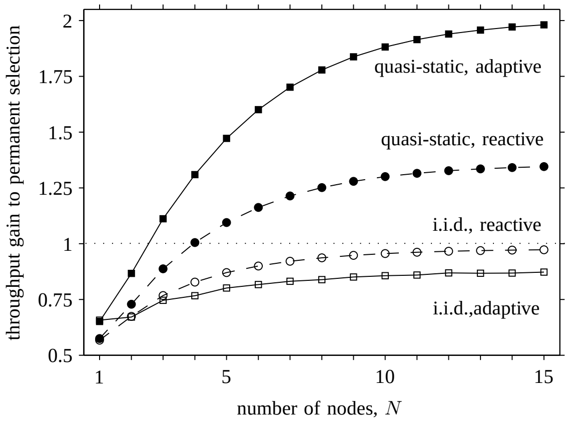
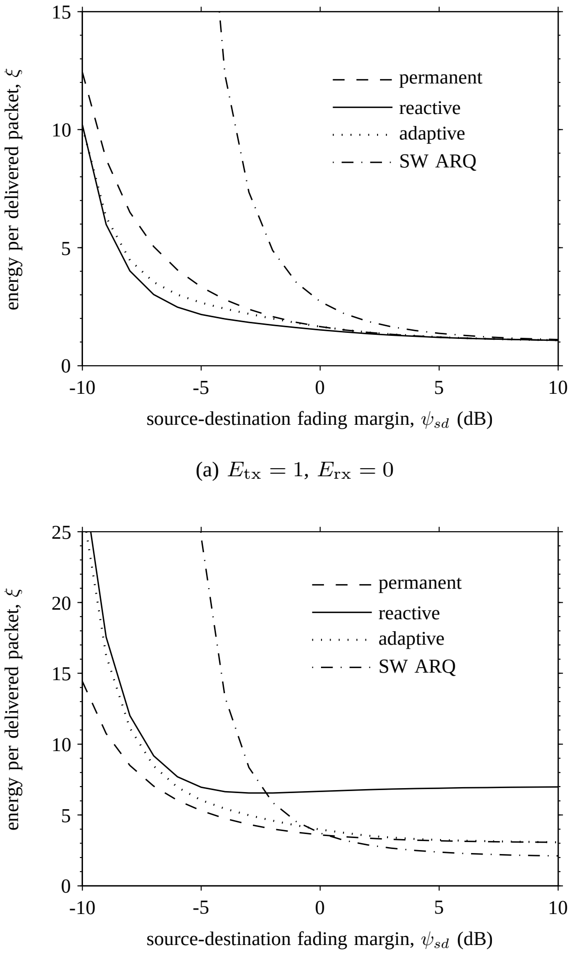


Fig. 7. Throughput gain at correlation boundaries as a function of number of nodes N. The throughput gain here is the ratio of throughput for reactive/adaptive selection to the corresponding throughput of cooperative ARQ with the permanent relay selection. Selection overhead $w = 1$, s-d fading margin $\psi_{sd} = -5\,\text{dB}$.



(a) $E_{\text{tx}} = 1$, $E_{\text{rx}} = 0$

(b) $E_{\text{rx}} = E_{\text{tx}} = 1$

Fig. 8. Expected energy per delivered packet ξ over source-destination fading margin ψ_{sd}. Number of nodes $N = 5$, channel time correlation $f_D T = 0.1$.

Finally, we investigate the impact of the number of nodes N on throughput. Cooperative ARQ with a permanent relay at the midpoint between s and d is used as a baseline (independent of N) for comparison. For fading margins $\psi_{sd} > 5\,\text{dB}$ the throughput of the two other schemes does not depend that much on N, since already with one or two available nodes a good relay can be selected in most cases, and the throughput limit is achieved. Fig. 7 shows the throughput ratio of reactive and adaptive relay selection schemes to that of permanent relay selection at i.i.d. and quasi-static channel bounds and $\psi_{sd} = -5\,\text{dB}$. Throughput ratios for other time-correlated channels lie within given bounds. Results show that permanent relay selection performs better in i.i.d. channels, even when other schemes can make use of other available potential relays N. The channel is too dynamic, which means selection of a good relay is less probable, and the selection overhead takes a lot of resources. For quasi-static channels, both adaptive and reactive selection schemes show significant benefits, since they can make use of more nodes and their stable channels. Particularly, adaptive relaying is highly beneficial in slow fading channels and high N, since new relay selection is performed less frequently when using reactive relay selection.

C. Energy Efficiency

We use the total energy consumption per delivered data message to evaluate the energy efficiency of the protocols. For comparison, we also present the corresponding energy of SW ARQ which is calculated by

$$\xi = \frac{1}{\eta} E_{\text{tx}} + E_{\text{rx}}. \tag{34}$$

In this section we set $E_{\text{tx}} = 1$.

Fig. 8a shows the expected energy per delivered DATA packet when energy consumption on the receiver side is neglected, i.e., $E_{\text{rx}} = 0$. This corresponds to the inverse of the throughput η in Fig. 4. As a result, reactive relay selection requires the least energy, since it provides the highest throughput. SW ARQ performs worst at low fading margins, since a packet delivery becomes nearly impossible. The energy consumption for $\psi_{sd} > 10\,\text{dB}$ changes only insignificantly for all schemes and approaches one energy unit.

However, it is more practical to also consider the energy required for packet reception. In this article we make a simplified assumption that the energy required to correctly receive a data packet is equal to the energy used for its transmission ($E_{\text{rx}} = E_{\text{tx}} = 1$) [36]. Fig. 8b shows that, as a result, the energy efficiency changes significantly. Reactive relaying performs worst among all cooperative ARQ protocols. At $\psi_{sd} > 0\,\text{dB}$ its energy per delivered packet is proportional to $N + 2$, since almost all overhearing nodes receive DATA messages with high probability. Permanent relay selection requires the lowest amount of energy, and, as shown in Fig. 5, provides best throughput. Adaptive relay selection can adapt to the channel quality, and it requires the same amount of energy at higher ψ_{sd} as permanent relay selection. At lower fading margins, however, relay selection is triggered more and more frequently. This means that the source broadcasts its data to all surrounding nodes, and the energy efficiency of adaptive selection approaches that of reactive selection.

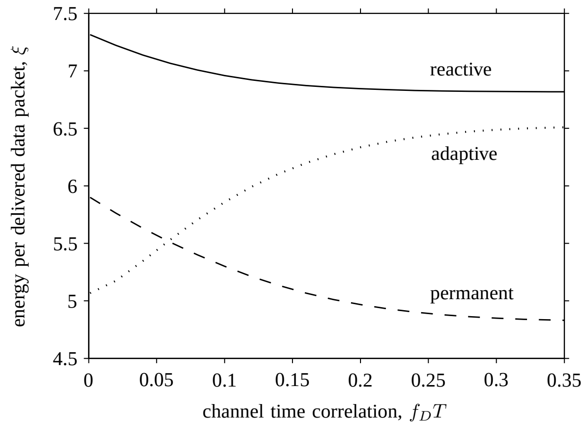


Fig. 9. Expected energy per delivered packet versus channel time correlation: $f_D T \to 0$ – quasi-static channels, $f_D T \approx 0.35$ – i.i.d. channels. Number of potential relays $N = 5$, $E_{\rm rx} = E_{\rm tx} = 1$, channel time correlation $f_D T = 0.1$, s-d fading margin $\psi_{sd} = -5$ dB.

Fig. 9 shows the impact of channel correlation on energy efficiency. Channels with higher correlation (lower $f_D T$) require more energy for relaying with permanent and reactive selection. This is due to the decreasing throughput (see Fig. 6), i.e., additional message retransmissions decrease the energy efficiency. Adaptive relay selection, in contrast, performs significantly better in slow fading channels ($f_D T < 0.1$), as new relay selections occur less frequently, and mostly only one relay needs to overhear s-d transmissions.

VII. Conclusions

This article provided a framework based on semi-Markov processes enabling us to model cooperative ARQ protocols with relay selection. Within this framework, we obtain the protocol performance in terms of throughput and energy efficiency taking into account relay selection overhead and temporal correlation of fading channels. Three relay selection schemes applying different relay update policies were studied.

The results obtained in a one-dimensional network with Rayleigh fading show that there is a significant tradeoff between relay selection overhead and throughput for reactive and adaptive relay selection, which can devalue throughput gains achieved through selection diversity. In contrast, the selection overhead for cooperative ARQ with a permanent relay can be neglected, and its actual throughput can be higher compared to reactive and adaptive schemes.

We also showed that time correlation of a radio channel has significant impact on the performance of cooperative ARQ protocols, particularly at low fading margins. The framework also introduces two channel correlation bounds: quasi-static channel and i.i.d. channel, which can be used to obtain expected throughput boundaries in a simple way.

Finally, we compared the expected energy consumption per delivered message. If the energy needed for packet reception is taken into account, reactive selection performs worst, since it requires all neighboring nodes to listen to source transmissions. In contrast, a permanent relay requires only a single listening relay. Adaptive selection adapts its behavior according to dynamics of radio channels, and is more energy efficient in slow time-correlated channels, where relay selections are less frequent.

These results show that relay selection is a critical part of cooperative relaying protocols, and that relay update rules have significant impact on the throughput and energy performance benefits. Adaptive relay selection methods should be taken into account in the design of new cooperative networking protocols.

Appendix
Cooperative ARQ with Adaptive Relay Selection

Protocols state transitions depend on current state and channel states. Here are the rules for protocol state transitions of cooperative ARQ with adaptive relay selection as described in Section III-C and shown in Fig. 2. The transition probability matrix $\mathbf{P}$ is calculated with (10).

$$y^{(b)} = \begin{cases} \mathrm{Tx}_r & \text{for } y^{(a)} = \mathrm{Tx}_r, c_{sd}^{(a)} = \mathrm{G}, \\ & \text{or } y^{(a)} = \mathrm{R}_r, c_{rd}^{(a)} = \mathrm{G}, \\ & \text{or } y^{(a)} = \mathrm{RTx}, c_{sd}^{(a)}, c_{sr}^{(a)}, c_{rd}^{(a)} = \mathrm{G}, \\ & \quad \sum_{n=1}^{N} \mathbb{1}_{\mathrm{G}}(c_{sn}^{(a)}) \mathbb{1}_{\mathrm{G}}(c_{nd}^{(a)}) = 0, \\ & \text{or } y^{(a)} = \mathrm{RS}, c_{sr}^{(a)} c_{rd}^{(a)} = \mathrm{G}, \\ & \quad \sum_{n=1}^{r-1} \mathbb{1}_{\mathrm{G}}(c_{sn}^{(a)}) \mathbb{1}_{\mathrm{G}}(c_{nd}^{(a)}) = 0, \\ \mathrm{R}_r & \text{for } y^{(a)} = \mathrm{Tx}_r, c_{sd}^{(a)} = \mathrm{B}, c_{sr}^{(a)} = \mathrm{G}, \\ \mathrm{RTx} & \text{for } y^{(a)} = \mathrm{Tx}_r, c_{sr}^{(a)}, c_{sd}^{(a)} = \mathrm{B}, \\ & \text{or } y^{(a)} = \mathrm{R}_r, c_{rd}^{(a)} = \mathrm{B}, \\ & \text{or } y^{(a)} = \mathrm{RTx}, c_{sd}^{(a)} = G, \\ & \quad \sum_{n=1}^{N} \mathbb{1}_{\mathrm{G}}(c_{sn}^{(a)}) \mathbb{1}_{\mathrm{G}}(c_{nd}^{(a)}) = 0, \\ & \text{or } y^{(a)} = \mathrm{RS}, \\ & \quad \sum_{n=1}^{N} \mathbb{1}_{\mathrm{G}}(c_{sn}^{(a)}) \mathbb{1}_{\mathrm{G}}(c_{nd}^{(a)}) = 0, \\ \mathrm{RS} & \text{for } y^{(a)} = \mathrm{RTx}, c_{sd}^{(a)} = \mathrm{B}. \end{cases} \tag{35}$$

The holding times are assigned with consideration of selection overhead as following:

$$H_{ab} = \begin{cases} 1 + w & \text{for } y^{(a)} = y^{(b)} = \mathrm{RTx}, \\ & \text{or } y^{(a)} = \mathrm{RTx}, y^{(b)} \in \{\mathrm{Tx}_r, \mathrm{RS}\}, \\ 0 & \text{for } y^{(a)} = \mathrm{RS}, y^{(b)} = \mathrm{RTx}, \\ 1 & \text{otherwise}. \end{cases} \tag{36}$$

Energy rewards for cooperative ARQ protocol with adaptive relay selection:

$$E_{ab} = \begin{cases} E_{\rm tx} + E_{\rm rx}\Big(\mathbb{1}_{\mathrm{G}}(c_{sd}^{(a)}) + \mathbb{1}_{\mathrm{G}}(c_{sr}^{(a)})\Big) & \\ \qquad\qquad \text{for } y^{(a)} = \mathrm{Tx}_r, & \\ E_{\rm tx} + \mathbb{1}_{\mathrm{G}}(c_{rd}^{(a)}) E_{\rm rx} \quad \text{for } y^{(a)} = \mathrm{R}_r, & \\ \qquad\qquad \text{or } y^{(a)} = \mathrm{RS}, y^{(b)} = \mathrm{Tx}_r, & \\ E_{\rm tx} + E_{\rm rx}\Big(\mathbb{1}_{\mathrm{G}}(c_{sd}^{(a)}) + \sum_{n=1}^{N} \mathbb{1}_{\mathrm{G}}(c_{sr}^{(a)})\Big) & \\ \qquad\qquad \text{for } y^{(a)} = \mathrm{RTx}, & \\ 0 \qquad \text{for } y^{(a)} = \mathrm{RS}, y^{(b)} = \mathrm{RTx}. & \end{cases} \tag{37}$$

Acknowledgments

The authors would like to thank S. Toumpis, H. Adam, E. Yanmaz, U. Schilcher, and K. Lienbacher.

REFERENCES

[1] A. Abdi, K. Wills, H. Barger, M.-S. Alouini, and M. Kaveh, "Comparison of the level crossing rate and average fade duration of Rayleigh, Rice and Nakagami fading models with mobile channel data," in *Proc. Veh. Technol. Conf. (VTC-Fall)*, Boston, US-MA, Sep. 2000.

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

[3] A. Bletsas, H. Shin, and M. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3450–3460, Sep. 2007.

[4] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1414–1423, Mar. 2009.

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

[6] D. Chen, H. Ji, and X. Li, "An energy-efficient distributed relay selection and power allocation optimization scheme over wireless cooperative networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011.

[7] Y. Wei, F. Yu, and M. Song, "Distributed optimal relay selection in wireless cooperative networks with finite-state Markov channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2149–2158, Jun. 2010.

[8] N. Marchenko, E. Yanmaz, H. Adam, and C. Bettstetter, "Selecting a spatially efficient cooperative relay," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Honolulu, US-HI, Dec. 2009.

[9] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[10] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Proc. IEEE INFOCOM*, Hong Kong, Mar. 2004.

[11] V. Shah, N. Mehta, and R. Yim, "Splitting algorithms for fast relay selection: Generalizations, analysis, and a unified view," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1525–1535, Apr. 2010.

[12] G. Yu, Z. Zhang, and P. Qiu, "Cooperative ARQ in wireless networks: Protocols description and performance analysis," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Istanbul, Turkey, Jun. 2006.

[13] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill Science/Engineering/Math, Aug. 2000.

[14] R. A. Howard, *Dynamic Probabilistic Systems Volume II. Semi-Markov and Decision Processes.* Dover Publications, 2007.

[15] B. Zhao and M. Valenti, "Practical relay networks: A generalization of Hybrid-ARQ," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 7–18, Jan. 2005.

[16] M. Dianati, X. Ling, K. Naik, and X. Shen, "A node-cooperative ARQ scheme for wireless ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 1032–1044, Jun. 2006.

[17] V. Mahinthan, H. Rutagemwa, J. W. Mark, and X. S. Shen, "Cross-layer performance study of cooperative diversity system with ARQ," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 705–719, Feb. 2009.

[18] Y. Lee and M.-H. Tsai, "Performance of decode-and-forward cooperative communications over Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1218–1228, Mar. 2009.

[19] D. Leong, P.-Y. Kong, and W.-C. Wong, "Performance analysis of a cooperative retransmission scheme using Markov models," in *Proc. Int. Conf. Inform., Commun. and Signal Process.*, Singapore, Dec. 2007.

[20] R. Madan, N. Mehta, A. Molisch, and J. Zhang, "Energy-efficient cooperative relaying over fading channels with simple relay selection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3013–3025, Aug. 2008.

[21] V. Shah, N. Mehta, and R. Yim, "The relay selection and transmission trade-off in cooperative communication systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2505–2515, Aug. 2010.

[22] D. Michalopoulos, A. Lioumpas, G. Karagiannidis, and R. Schober, "Selective cooperative relaying over time-varying channels," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2402–2412, Aug. 2010.

[23] C. Xiao and N. Beaulieu, "Node switching rates of opportunistic relaying and switch-and-examine relaying in Rician and Nakagami-m fading," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 488–498, Feb. 2012.

[24] N. Marchenko, C. Bettstetter, and W. Elmenreich, "Incremental cooperative relaying in time-correlated Rayleigh fading channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Miami, US-FL, Dec. 2010.

[25] N. Marchenko and C. Bettstetter, "Throughput and energy efficiency of cooperative diversity with relay selection," in *Proc. European Wireless (EW)*, Vienna, Austria, Apr. 2011.

[26] ——, "Impact of relay selection overhead in cooperative diversity protocols," in *Proc. Veh. Technol. Conf. (VTC-Fall)*, San Francisco, CA, Sep. 2011.

[27] R. Yim, N. Mehta, and A. Molisch, "Fast multiple access selection through variable power transmissions," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1962–1973, Apr. 2009.

[28] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 445–455, May 1997.

[29] ——, "Error statistics in data transmission over fading channels," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1468–1477, Nov. 1998.

[30] A. Ramesh, A. Chockalingam, and L. Milstein, "A first-order Markov model for correlated Nakagami-m fading channels," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, New York, US-NY, May 2002.

[31] M. K. Simon and M. Alouini, *Digital Communications over Fading Channels*, 2nd ed. John Wiley & Sons, Inc., 2005.

[32] D. L. Lu and J. F. Chang, "Performance of ARQ protocols in nonindependent channel errors," *IEEE Trans. Commun.*, vol. 41, no. 5, pp. 721–730, May 1993.

[33] M. Zorzi and R. R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1077–1081, Sep. 1996.

[34] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. John Hopkins Studies in Mathematical Sciences, Oct. 1996.

[35] S. C. Ng, G. Mao, and B. D. O. Anderson, "On the properties of one-dimensional infrastructure-based wireless multi-hop networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2606 –2615, Jul. 2012.

[36] L. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *Proc. IEEE INFOCOM*, Anchorage, US-AK, Apr. 2001.

Nikolaj Marchenko graduated from RWTH Aachen University, Germany, in 2007 with Diploma degree in Computer Engineering. He wrote his thesis while at Siemens Corporate Research in Princeton, NJ. Since January 2008 Nikolaj has been research and teaching staff member at Networked and Embedded Systems institute, Klagenfurt University, Austria, where he is working towards his PhD on cooperative communication protocols. In 2011 Nikolaj spent four months as a visiting scholar at Georgia Institute of Technology, Atlanta, GA.

Christian Bettstetter studied electrical engineering at TU München, receiving the Dipl.-Ing. in 1998 and the Dr.-Ing. (summa cum laude) in 2004. He was then a senior researcher at DOCOMO. Since 2005, Christian has been a professor and head of Networked and Embedded Systems at the University of Klagenfurt, leading a team of 15 researchers. He is scientific director and founder of Lakeside Labs, a cluster on self-organizing networked systems. He received paper awards from the IEEE Vehicular Technology Society and the German ITG.