A probabilistic HARQ protocol for Demodulate-and-Forward (DMF) relaying network

Faton Maliqi, Francesca Bassi, Pierre Duhamel, Ilir Limani

A probabilistic HARQ protocol for Demodulate-and-Forward (DMF) relaying network

Faton Maliqi, Francesca Bassi, *Member, IEEE,* Pierre Duhamel, *Fellow, IEEE,* and Ilir Limani

***Abstract*—This paper considers wireless communication protocols implementing both cooperation and Hybrid Automatic Repeat reQuest (HARQ). Based on a simple example (one Source, one Relay and one Destination), we propose the use of probabilistic protocols as an alternative to classical, deterministic ones. Starting initially with the Finite State Markov Chain (FSMC) analysis of a deterministic protocol, the idea of probabilistic protocol comes as an association with the simplified FSMC of the deterministic protocol. This FSMC contains two parameters that can be optimized for finding the best performance. So far, probabilistic protocols have been proposed at higher layers of communication systems, while we consider here the physical and MAC layers. In our analysis, the Relay works in Demodulate-and-Forward (DMF) mode, and we demonstrate that (i) the FSMC analysis predicts accurately the performance of the actual system, and (ii) the performance of probabilistic protocol (when the two parameters are carefully tuned) outperforms the performance of a reference deterministic protocol. Analysis is checked upon Monte Carlo simulations.**

***Index Terms*—Hybrid Automatic Repeat reQuest (HARQ), Cooperative Communications, Wireless Networks, Finite-State Markov Chains.**

I. Introduction

Wireless channels suffer path-loss, fading, shadowing, interference, as a result of which the received signal is degraded, causing errors during the detection process. To combat these phenomena and to improve the communication the technique of Hybrid Automatic Repeat reQuest (HARQ) [1] can be used. HARQ consists in retransmitting information on the same packet from the same node. The HARQ technique can be combined with the cooperative diversity technique [2] by the introduction of relay nodes. One open question is the search of the best combination of these two techniques, and we are going to address it by designing tools to evaluate the performance of the combined protocols and by looking for the best way of maximizing it.

The combination of these two techniques has already been studied, for example in [3], where a communication protocol with multiple relays is proposed. The outage probability expressions of a cooperative wireless network with HARQ are derived in [4], while the energy efficiency of HARQ protocols on a cooperative network is discussed in [5]. Various tools have been proposed for a theoretical analysis of these systems, but we focus here on Finite State Markov Chain [6]. Similar frameworks are widely used in the literature, like in [7] for studying a bursty service model in a cooperative network with ARQ when the transmitting node is saturated with a current copy. In this situation, the process of retransmitting some packet is still ongoing when a new packet has to be transmitted, which results in some packet loss. In [8] the Markov framework is used to perform performance analysis for a cooperative two-path relay channel, where both the source and the relay are allowed to transmit simultaneously.

Most of these papers consider the case when relay works on the Decode-and-Forward (DCF) mode, while fewer consider the Amplify-and-Forward (AF) mode [9] where the source participates in all the retransmissions while relay retransmits only if its normalized accumulative mutual information is not lower than a certain threshold. One important aspect in these cases is that the action of the relay is complex (decode and re-encode) or conditioned on an evaluation of the efficiency of previous retransmissions. In contrast, we concentrate on the simplest mode for the relay, the Demodulate-and-Forward (DMF) mode [10], which would be of great interest in contexts where energy efficiency at the relay node is of great importance, such as in the Internet of Things. Note that in this case the action of the relay is not conditioned, and hence the relay always has something to retransmit.

So far, probabilistic protocols have mainly been proposed for higher layers of communication systems, like in [11] which studies a probabilistic routing protocol where the decision of packet forwarding from one node to another is based on a probabilistic metric which is updated each time one node encounters another one. Another example is found in [12] where the communication is based on an algorithm that pre-computes the probability that the communication is possible between a specified source and the destination. In contrast, expanding on ideas presented by the authors in [13], this paper proposes the use of a probabilistic protocol at the physical and MAC layers, and proposes a careful performance analysis and optimization. It has to be noted that the rather involved analysis is required only for performance optimization, and that the decoding algorithm remains unchanged with respect to classical deterministic protocols.

The paper is organized as follows: Section III describes an example of deterministic protocol on the relay network, and discusses its performance evaluation using FSMC for Type I and Type II decoders. Since the performance evaluation and optimization becomes computationally very demanding as the protocol gets more sophisticated, a simplified analysis using

F. Maliqi and P. Duhamel are with Laboratoire des Signaux et des Systèmes, L2S, Gif-sur-Yvette, France. Email: {faton.maliqi, pierre.duhamel}@l2s.centralesupelec.fr

F. Bassi is with Laboratoire des Signaux et des Systèmes, L2S, Gif-sur-Yvette, France and with ESME-Sudria, Ivry-sur-Seine, France. Email: francesca.bassi@l2s.centralesupelec.fr.

I. Limani is with Pristina University, Pristina, Kosovo. Email: ilir.limani@uni-pr.edu.

state aggregation on the FSMC is proposed. Then, by associating a probabilistic protocol to this simplified model, Section IV introduces the probabilistic protocol. The description by the simplified model, albeit able to represent the performance of the probabilistic protocol, is not well adapted to be used as a tool for parameter optimization, because of the mutual dependence it exhibits between the parameters. Then, for both Type I and Type II decoder, alternative FSMC descriptions are derived, which closely predict the performance of the probabilistic protocol and can be used for parameter optimization. In Section V, the theoretical results are validated using Monte Carlo simulations, and the performance of the probabilistic protocol is compared to the performance of the deterministic strategy.

II. System model

Consider a simple network [2] composed of one source, one relay, and one destination. The source generates information Packet Data Units (PDUs) of fixed length, to which a Cyclic Redundancy Check (CRC) sequence [14] is appended to enable the HARQ mechanism. Each PDU is channel encoded in a packet.

The transmission happens in time-slots. The transmission of a coded packet occupies a time-slot. The source and the relay take turns in occupying the channel to send the coded packet relative to the active PDU to the destination. At the end of each time-slot the destination decodes, performs error detection and broadcasts an HARQ control bit (ACK if the message has been correctly decoded, NACK otherwise), received by both the source and the relay without delay. In this work we consider both Type I HARQ, where the destination attempts decoding on only the most recently received coded packet; and Type II HARQ, where the destination exploits all coded packets relative to the same PDU received so far, performing Chase Combining [16]. Perfect CSI is assumed known at the destination.

Relays may work in Amplify-and-Forward (AF), Decode-and-Forward (DCF) or Demodulate-and-Forward (DMF) mode. In the AF mode the relay amplifies the received signal and forwards it if it meets a predefined quality criterion [9]. In the DCF mode the relay decodes the incoming sequence, checks for the integrity of the message using the CRC, and forwards it only if it is correct. In the DMF mode [10] the relay demodulates the received symbols (hard decision) and then forwards them, without checking on any conditions. In this work the relay is assumed to work in half-duplex, DMF mode. The DMF mode represents a good trade-off between performance and complexity, and is thus of great interest in contexts where energy efficiency at the relay node is critical.

The links source-destination (SD), source-relay (SR) and relay-destination (RD) are modeled as independent Rayleigh flat fading channels. The source and the relay use the same energy per symbol E_s, the same channel code and the same modulation scheme. Let d_{XY} be the distance between the nodes X and Y. The simplified path-loss factor between the nodes X and Y takes the form $d_{\text{XY}}^{-\alpha}$, with α path-loss exponent. The communication channels SD and SR for time-slot t are modeled as:

$$\mathbf{y}_{\text{SD},t} = \sqrt{E_s d_{\text{SD}}^{-\alpha}}\ h_{\text{SD},t}\ \mathbf{x}_t + \mathbf{w}_{\text{SD},t} \tag{1}$$

$$\mathbf{y}_{\text{SR},t} = \sqrt{E_s d_{\text{SR}}^{-\alpha}}\ h_{\text{SR},t}\ \mathbf{x}_t + \mathbf{w}_{\text{SR},t}, \tag{2}$$

where $\sqrt{E_s}\mathbf{x}_t$ is the vector of modulated symbols with energy E_s; $\mathbf{w}_{\text{XY},t}$ is the vector of white noise whose symbols have distribution $\mathcal{CN}(0, N_0)$; $h_{\text{XY},t}$ is the Rayleigh fading complex coefficient with distribution $\mathcal{CN}(0, 1)$. The communication channel RD for time-slot t is modeled as:

$$\mathbf{y}_{\text{RD},t} = \sqrt{E_s d_{\text{RD}}^{-\alpha}}\ \ h_{\text{RD},t}\ \widetilde{\mathbf{x}}_t + \mathbf{w}_{\text{RD},t}, \tag{3}$$

where $\widetilde{\mathbf{x}}_t$ is the vector of re-modulated symbols formed by the relay, based on hard-demodulation of $\mathbf{y}_{\text{SR},t}$. Obviously, these re-modulated symbols cannot be considered as being error free.

A. Performance metrics definition

Let ν be the total number of information PDUs generated by the source during the operation time, σ be the number of PDUs that have been ACK-ed by the destination, and ν_t be the number of PDUs that have been transmitted exactly t times. In order to express the performance of the system we consider the PDU Error Rate (PER), corresponding to the proportion of PDUs that were transmitted but never acknowledged by the destination; and the average number of transmissions per PDU, denoted by $\overline{T}$:

$$\text{PER} = 1 - \frac{\sigma}{\nu}, \qquad \overline{T} = \frac{1}{\nu}\sum_{t=1}^{N_{\max}} t \cdot \nu_t, \tag{4}$$

where $N_{\max}$ is the maximum number of times that a coded packet relative to the same PDU has been transmitted. In HARQ schemes an increase of $\overline{T}$ is beneficial in terms of PER, but represents a cost in terms of delay. A meaningful metric to account for the trade-off between PER and $\overline{T}$ is the *Goodput* G (also denoted as throughput), defined as the number of successfully delivered information PDUs per unit of time, [PDUs/tu]. Assuming that the transmission of one information PDU over the channel with code rate R_c takes 1 [tu], G takes the form

$$G = R_c \cdot \frac{1 - \text{PER}}{\overline{T}} \quad \left[\frac{\text{PDUs}}{tu}\right]. \tag{5}$$

Notice that the Goodput naturally expresses the trade-off between the performance of the system in terms of PER and the average cost per information PDU. In our system, it can be seen as a measure of the efficiency of the system in using the resources it spends.

III. The deterministic protocol

The communication protocol is the set of rules that allow each of the transmitting nodes to determine whether it is their turn to occupy the channel. The communication protocol is known also to the destination. In this Section we consider the class of communication protocols composed by deterministic rules, where the ACK/NACK message is the only random event involved in deciding who will transmit next. This means that the order of the nodes in transmitting is predetermined until ACK is observed, or the maximum allowed number $N_{\max}$ of transmissions are reached.

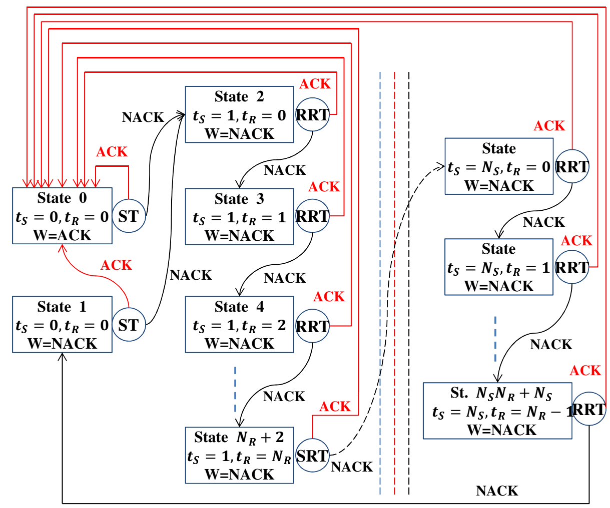
We have selected a specific instance of deterministic protocol on the network S-R-D that will be used as an example throughout the paper. It works as follows. The transmission happens in time-slots. The coded packet relative to the current information PDU is broadcast for the first time by the source and is received by both the destination and the relay. The destination decodes, uses the CRC code to detect errors, and sends the HARQ control message received by both source and relay. In case no error is found, the destination sends the ACK message and the source proceeds with the transmission of a new information PDU. In case an error is detected, the destination sends the NACK message and the protocol enters in the retransmission phase, where the relay and the source retransmit the coded packet of the active PDU in a deterministic order. The retransmission order is determined by the following rules:

1) the relay performs the first retransmission;
2) the relay retransmits N_R consecutive times after each (re-)transmission by the source, after which the control of the channel goes back to the source;
3) the source transmits the same PDU a maximum of N_S times (this includes the first transmission).

The deterministic protocol allows transmission of the same PDU for $N_{\max} = N_S + N_S N_R$ times at most. If after $N_{\max}$ transmissions the current PDU has not been acknowledged it is dropped, and the source starts the transmission of a new information PDU.

A. Description of the deterministic protocol using FSMs and performance analysis using FSMCs

Any deterministic protocol where each time-slot corresponds to an action by a transmitting node can be described using a Finite State Machine (FSM). The state of the FSM determines the action that takes place during the current time-slot, and the outcome of the action determines the transition to the next state. For the deterministic protocol considered as an example, the FSM description and the definition of the states are given in Figure 1. The states are defined according to the values taken by few parameters. At the beginning of the time-slot, t_S represents the number of times the source has already transmitted the current PDU; t_R represents the number of times the relay has already transmitted the current PDU after the last transmission from S; W represents the last control message issued by the destination. Different combinations of the values (t_S, t_R, W) define different states. Each state is associated with one of the possible actions: ST (Source Transmits a new PDU for the first time), SRT (Source Retransmits the current PDU) or RRT (Relay Retransmits the current PDU), and the scheme of all possible state transitions in the FSM is given in Figure 1a. All actions resulting in an ACK from the destination imply the transmission of a new PDU (hence a transition to state 0). Notice that for any deterministic protocol it is always possible



(a)

State				Action
0	$t_S = 0$	$t_R = 0$	W =ACK	ST
1	$t_S = 0$	$t_R = 0$	W =NACK	ST
2	$t_S = 1$	$t_R = 0$	W =NACK	RRT
3	$t_S = 1$	$t_R = 1$	W =NACK	RRT
…	…	…	…	…
N_R+2	$t_S = 1$	$t_R = N_R$	W =NACK	SRT
N_R+3	$t_S = 2$	$t_R = 0$	W =NACK	RRT
N_R+4	$t_S = 2$	$t_R = 1$	W =NACK	RRT
…	…	…	…	…
$N_S N_R$+N_S	$t_S = N_S$	$t_R = N_R - 1$	W =NACK	RRT

(b)

Fig. 1. Deterministic protocol, FSM description (a) and state definition (b).

to give a FSM description where an ACK from the destination triggers a transition to state 0, and NACK on the last allowed transmission triggers a transition to state 1.

Since, given that the machine is in a given state, the transition depends only on the last control message from the destination, the FSM associated to any deterministic protocol has the Markov Chain [15] property. Considering the case of Type II HARQ, the decoder combines all data relative to the same PDU received so far. The example deterministic protocol can then be described by a Finite State Markov Chain (FSMC), with probability transition matrix:

$$P_{\mathrm{II}} = \begin{pmatrix} 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 & \cdots & 0 & \cdots \\ 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 & \cdots & 0 & \cdots \\ 1-\pi_{[1,1]} & 0 & 0 & \pi_{[1,1]} & \cdots & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ 1-\pi_{[1,N_R]} & 0 & 0 & 0 & \cdots & \pi_{[1,N_R]} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ 1-\pi_{[N_S,N_S N_R]} & \pi_{[N_S,N_S N_R]} & 0 & 0 & \cdots & 0 & \cdots \end{pmatrix}, \quad (6)$$

where $\pi_{[A,B]}$ is the probability that the destination fails in decoding the current PDU, based on A copies received from the source and B copies received from the relay, knowing that previous decoding with fewer copies has been unsuccessful. Since in Type I decoding only the most recent received packet is used, the transition matrix P_{I} obtained for Type I decoder is the special case of matrix P_{II}, where $\pi_{[A,B]} = \pi_{[1,0]}$ on the rows corresponding to states with action ST, and $\pi_{[A,B]} = \pi_{[0,1]}$ on the rows corresponding to states with action RRT. For the time being we assume that each retransmission of the same PDU from the relay gives the same average probability of failure in decoding, $\pi_{[0,1]}$. This approximation is equivalent to consider the quality of packets received by consecutive retransmissions by the relay as statistically independent. The limits of this approximation will be discussed in Section IV-C. For numerical results, the probabilities $\pi_{[A,B]}$ are evaluated via Monte Carlo simulation.

As done with the example deterministic protocol, one can get the transition matrix P of the FSMC of any deterministic protocol. The steady state vector $\boldsymbol{p}$ is evaluated as the eigenvector associated to eigenvalue $\lambda = 1$ of the transition matrix P. The steady state probability p_k represent the fraction of time that the Markov Chain is in state k [17], and the reciprocal value of the steady state probability represent the expectation of recurrence time on state k. Observing that State 0 is visited only in case of correct acknowledgment of an information PDU, and that State 1 is visited only if a non-acknowledged PDU is dropped, the performance metrics (4) and (5) can be evaluated as:

$$\mathrm{PER} = \frac{p_1}{p_0 + p_1}, \qquad \overline{T} = \frac{1}{p_0 + p_1}, \qquad G = R_c \cdot p_0. \tag{7}$$

Notice however that this performance evaluation requires the computation of an eigenvector of the transition matrix, which becomes computationally demanding if the number of states is considerably higher than in the case considered as an example. This happens if $N_{\max}$ gets larger or if we consider other modes or other network topologies. For example, switching the relay in Decode-and-Forward mode implies that the definition of the state needs to encode whether the relay has the correct message, and this increases the number of possible states. Therefore, the next section addresses means of reducing the size of the FSMC, while keeping the performance evaluation unchanged.

B. Simplification of the FSMC description of the deterministic protocol via state aggregation

Consider a FSMC $(\mathcal{S}, P)$ defined by a state set $\mathcal{S} = \{0, \ldots, L\}$ and a transition matrix P. By aggregating multiple states of $(\mathcal{S}, P)$ it is possible to obtain a new FSMC $(\mathcal{S}', Z)$ with state set $\mathcal{S}' = \{0, 1, \ldots, M\}$, with $M < L$, and transition matrix Z. We are interested in imposing the following constraint on the steady state vectors $\boldsymbol{p}$ and $\boldsymbol{z}$ of the FSMCs $(\mathcal{S}, P)$ and $(\mathcal{S}', Z)$, respectively. Let I be the state in $(\mathcal{S}', Z)$ resulting from the aggregation of the set of states $\mathcal{I} \subseteq \mathcal{S}$ of $(\mathcal{S}, P)$. Then the steady state probability z_I of state I is the sum of the steady state probabilities $p_i,\ i \in \mathcal{I}$

$$z_I = \sum_{i \in \mathcal{I}} p_i. \tag{8}$$

The elements of the transition matrix Z then become [18]:

$$[Z]_{IJ} = \frac{\sum_{i \in \mathcal{I}} p_i \left(\sum_{j \in \mathcal{J}} [P]_{ij} \right)}{\sum_{i \in \mathcal{I}} p_i}. \tag{9}$$

We apply this procedure to the FSMC $(\mathcal{S}, P_{\mathrm{II}})$ defined by Figure 1b and (6), and we obtain the FSMC $(\mathcal{S}', Z_{\mathrm{II}})$. We aggregate the states in $\mathcal{S}$ associated with the same action, with the exception of States 0 and 1, which are left untouched, since their steady state probabilities contain all the information necessary to get the performance metrics. We obtain the state set $\mathcal{S}'$ described in Figure 2b. The states associated with the action RRT in $\mathcal{S}$ are grouped to State 2 in $\mathcal{S}'$; the states in $\mathcal{S}$ associated with the action SRT are grouped in State 3 in $\mathcal{S}'$. Notice that aggregating states associated with the same action can be applied to the FSMC description of any deterministic protocol, leading to a new FSMC composed by 4 grouped states.

The transition matrix Z_{II} can be evaluated from P_{II} and its steady state vector $\boldsymbol{p}$ using (9). It takes the following form

$$Z_{\mathrm{II}} = \begin{pmatrix} 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 \\ 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 \\ 1-\pi_{\mathrm{R}} & \gamma \cdot \beta \pi_{\mathrm{R}} & (1-\gamma)\,\pi_{\mathrm{R}} & \gamma\,(1-\beta)\,\pi_{\mathrm{R}} \\ 1-\pi_{\mathrm{S}} & 0 & \pi_{\mathrm{S}} & 0 \end{pmatrix}. \tag{10}$$

The term $(1-\pi_{\mathrm{R}})$ in (10) is defined via (9), and represents the average probability of successful decoding, knowing that the last transmission has been performed by the relay. Similarly, $(1 - \pi_{\mathrm{S}})$ is the average probability of successful decoding, knowing that the last *re*transmission has been performed by the source (notice that this excludes the first transmission). As a consequence, the parameter $(1 - \gamma)$ has the interpretation of the probability that the relay will perform next retransmission, knowing that it is transmitting in the current time-slot; and $(1 - \beta)$ represents the probability that the source will perform next transmission, knowing that in the current time-slot the relay is transmitting but will not be allowed to transmit in the next.

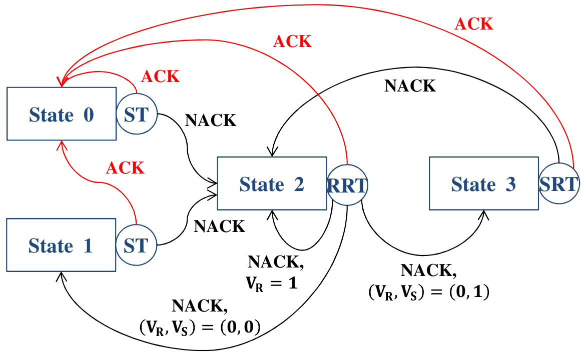
As before, we consider Z_{I} as a special case of Z_{II}. Since Type I decoding does not have memory of past transmissions, we have that $\pi_{\mathrm{R}} = \pi_{[0,1]}$ and $\pi_{\mathrm{S}} = \pi_{[1,0]}$. This is not true for Type II decoding, which exploits previously received copies, and where π_{R} and π_{S} depend on the parameters $(N_{\mathrm{S}}, N_{\mathrm{R}})$.

The parameters γ and β can be also found, for both Type I and Type II decoding, using a combinatorial approach. In the example deterministic protocol this yields, for Type I decoder,

$$\gamma = \frac{\pi_{[0,1]}^{N_{\mathrm{R}}}}{\pi_{[0,1]} + \pi_{[0,1]}^2 + \pi_{[0,1]}^3 + \cdots + \pi_{[0,1]}^{N_{\mathrm{R}}}}, \tag{11}$$

$$\beta = \frac{1}{1 + \sum_{t_{\mathrm{S}}=1}^{N_{\mathrm{S}}-1} \left(\dfrac{1}{\pi_{[1,0]}^{t_{\mathrm{S}}} \cdot \pi_{[0,1]}^{t_{\mathrm{S}} \cdot N_{\mathrm{R}}}} \right)}. \tag{12}$$

For the case of Type II decoder γ and β can be evaluated similarly. Since the values of γ and β depend on the elements of P_{I} for Type I decoding and on the elements of P_{II} for Type II decoding, they differ in the two cases.



(a)

State				Action
0			W = ACK	ST
1	V_{R}=0	V_{S}= 0	W = NACK	ST
2	V_{R}=1	V_{S}= 0	W = NACK	RRT
3	V_{R}=0	V_{S}= 1	W = NACK	SRT

(b)

Fig. 2. Probabilistic protocol, FSM description (a) and state definition (b).

IV. The probabilistic protocol

A. Association of a probabilistic protocol with the simplified FSMC

In Section III-B states associated with the same action are grouped together, with the aim of obtaining small FSMCs $(\mathcal{S}', Z_{\mathrm{I}})$ and $(\mathcal{S}', Z_{\mathrm{II}})$. Besides, it can easily be seen that these small FSMCs can be considered as being representative of a communication protocol, albeit different from the protocol described by $(\mathcal{S}, P_{\mathrm{I}})$ and $(\mathcal{S}, P_{\mathrm{II}})$.

The new transmission protocol is described by the FSM in Figure 2a. The transitions from States 0 and 1 are determined by the observed control message issued by the destination (ACK or NACK) only. Recall that γ is defined as the probability that the relay will not retransmit in the next time-slot knowing that it performs the current retransmission, and β as the probability that the source will not retransmit in the next time-slot, knowing that in the current time-slot the relay is transmitting and will not be allowed to transmit in the next time slot. The transitions from States 2 and 3 hence depend on the observed control message but also on the realization of two random variables, V_{S} and V_{R}. The variable V_{R} is Bernoulli distributed, with parameter γ. The variable $(V_{\mathrm{S}}|V_{\mathrm{R}} = 0)$ is Bernoulli distributed, with parameter β, and $(V_{\mathrm{S}}|V_{\mathrm{R}} = 1) = 0$ with probability 1. The definition of $\mathcal{S}'$ according to the values taken by V_{R} and V_{S} is in Figure 2b.

Clearly, since the actions depend on the realization of random variables, this protocol can be denoted as a probabilistic. This protocol is described by the FSM in Figure 2a and works as follows. The first transmission of a new information PDU is performed by the source. In case retransmissions are needed the following rules are applied:

1) the first retransmission after any (re)-transmission by the source is performed by the relay;
2) if the relay is retransmitting, the next action is determined by drawing $(V_{\mathrm{R}}, V_{\mathrm{S}})$. If $V_{\mathrm{R}} = 1$, the relay is allowed to retransmit in the next time slot (probability $(1-\gamma)$); if $(V_{\mathrm{R}}, V_{\mathrm{S}}) = (0, 1)$, the source is allowed to retransmit in the next time-slot (probability $\gamma(1-\beta)$); if $(V_{\mathrm{R}}, V_{\mathrm{S}}) = (0, 0)$ neither the relay or the source are allowed to retransmit in the next time slot (probability $\gamma \cdot \beta$);
3) if neither the relay or the source are allowed to retransmit, the current PDU is dropped, and the source transmits a new PDU for the first time.

As done for the deterministic protocol, the performance of the probabilistic protocol for Type I (Type II) decoder can be deduced by the steady-state vector elements z_0 and z_1 of the matrix Z_{I} (Z_{II}). Since the States 0 and 1 have been excluded from grouping, and because of the constraint (8) imposed in the aggregation procedure, one has that $z_0 = p_0$ and $z_1 = p_1$, for both Type I and Type II decoders. It follows that the probabilistic protocol associated to $(\mathcal{S}', Z_{\mathrm{I}})$ and $(\mathcal{S}', Z_{\mathrm{II}})$ matches the performance of the deterministic protocol associated to $(\mathcal{S}, P_{\mathrm{I}})$ or $(\mathcal{S}, P_{\mathrm{II}})$.

This example shows that given a deterministic protocol allowing $N_{\max}$ transmissions it is always possible, via the state aggregation procedure, to find the parameters (γ, β) such that the probabilistic protocol matches the performance of the deterministic protocol. We conclude that the probabilistic protocol is at least as good as any deterministic one, with a possibility of improvement. This is addressed below.

B. Parameter optimization for the probabilistic protocol

This section addresses the problem of finding the best values of parameters γ and β optimizing the performance of the probabilistic protocol. Since the goodput G represents the trade-off between the PER and the average number of transmissions per PDU $\overline{T}$, an appropriate criterion is:

$$\max_{\gamma,\beta} G = \max_{\gamma,\beta} R_c \frac{1-\mathrm{PER}}{\overline{T}}. \tag{13}$$

For any choice of (γ, β) the performance metrics PER, $\overline{T}$ and G of the probabilistic protocol can be predicted using the steady-state probabilities associated with Z_{I} and Z_{II}. This, however, requires knowledge of the probabilities π_{S} and π_{R} in Z_{I} and Z_{II}.

Consider the Type I decoder first. Since the decoder processes only the most recently received copy, the average probability of ACK after a retransmission from the source remains the same on each retransmission, hence $(1-\pi_{\mathrm{S}}) = (1-\pi_{[1,0]})$. We have so far assumed that also the average probability of ACK after a retransmission from the relay remains constant, which implies $(1-\pi_{\mathrm{R}}) = (1-\pi_{[0,1]})$. Though this assumption holds in many situations, it is not the case when the relay works on the DMF mode, because the demodulation errors at the relay have to be taken into account. Therefore, the probability of ACK cannot be assumed constant on successive retransmissions from the relay, especially when it is far from the source, and $(1-\pi_{\mathrm{R}})$ clearly depends on the probabilities that successive retransmissions are performed, and hence on γ. This makes the steady-state analysis of Z_{I} not very convenient for parameter optimization, since for each variation of γ the probability π_{R} needs to be re-evaluated, too. Therefore we propose in Section IV-C an FSMC model which accurately approximates the performance of the probabilistic protocol with Type I decoder and at the same time is convenient to use in the optimization process.

Consider now the Type II decoder. Since the decoder has memory, the average probabilities $(1-\pi_{\mathrm{R}})$ and $(1-\pi_{\mathrm{S}})$ involved in Z_{II} are statistically dependent, via the parameters γ and β. For example, for $\beta \longrightarrow 1$ the parameter $(1-\pi_{\mathrm{R}})$ depends only on γ, and can be evaluated as

$$(1-\pi_{\mathrm{R}}) = \sum_{k=1}^{\infty}(1-\pi_{[1,k]})(1-\gamma)^{k-1}\left(\prod_{h=0}^{k-1}\pi_{[1,h]}\right). \qquad (14)$$

Since the number of retransmissions is potentially infinite, it is clear from (14) that an accurate evaluation of the parameters $\pi_{[A,B]}$ for high values of A and B are needed to evaluate good approximations of $(1-\pi_{\mathrm{R}})$ and $(1-\pi_{\mathrm{S}})$. Moreover, the optimization process based on the steady-state analysis of Z_{II} is not convenient, since for each choice of (γ,β) the values of π_{R} and π_{S} need to be re-evaluated, too. In order to overcome the complexity issue associated with this problem, Section IV-D proposes an FSMC model that can predict very closely the performance of the probabilistic protocol with Type II decoder, and which at the same time is convenient to use in the optimization process.

C. FSMC model for performance approximation of the probabilistic protocol, Type I decoder

The problem we are facing is due to the assumption that the relay works on the DMF mode. Consider two successive transmissions by the relay. The relay may be forwarding a corrupted version of the coded packet, due to demodulation errors. Hence, the quality of the second transmission is strongly correlated to the quality of the first, and the two transmissions cannot be assumed as independent. This implies that one cannot assume that the probability of decoding failure after a retransmission by the relay, $\pi_{[0,1]}$, remains constant independently on how many times in a row the relay had been retransmitting. Define $\pi_{[0,1]}(1)$ as the probability that decoding fails on the first re-transmission by the relay given failure in all previous transmissions, $\pi_{[0,1]}(2)$ as the probability that it fails on the second retransmission given failure on the first, and so on. These probabilities are evaluated for various location of the relay using Monte Carlo simulations and plotted on Figure 3. As we may see, even when Type I decoding is considered and only the most recently received packet is used in decoding, the probability of packet error depends on how many times the relay has been retransmitting in a row, confirming that the hypothesis of independence of successive transmissions from the relay does not hold, in general. As the relay is located close to the source, the probability of failure after the k-th retransmission from the relay, $\pi_{[0,1]}(k)$, changes very slowly as k increases, while as the relay moves further toward the destination $\pi_{[0,1]}(k)$ changes more rapidly as k increases.

Notice that π_{R} in Z_{I} is, by definition, the average probability of NACK when R is transmitting, which depends, other than on the values $\pi_{[0,1]}(k)$, also on the retransmission probability, and hence on γ, and this makes Z_{I} not well-adapted to

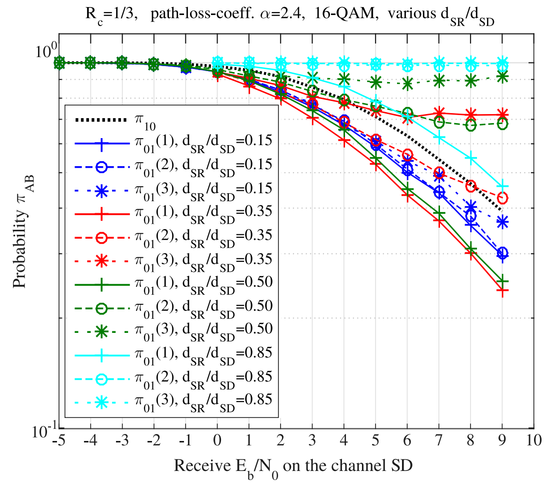


Fig. 3. Monte Carlo simulation of probabilities $\pi_{[1,0]}$ and $\pi_{[0,1]}(k)$, Type I decoder, various $d_{\mathrm{SR}}/d_{\mathrm{SD}}$.

be used in the optimization process. Therefore, we want to develop a FSMC model where the probabilities of ACK and NACK related to any given transmission do not depend on the parameters (γ,β) objective of the optimization.

Since $\pi_{[0,1]}(k)$ changes very slowly when the relay is located close to the source, and more rapidly (with the upper-bound probability close to 1) as the relay approaches the destination, we propose to keep track on probability variation up to the third successive retransmission from the relay and then assume that after the third retransmission the probability of failure for the k-th retransmission $(k>3)$ remains constant, that is:

$$\pi_{[0,1]}(k) \approx \pi_{[0,1]}(3), \quad k>3. \qquad (15)$$

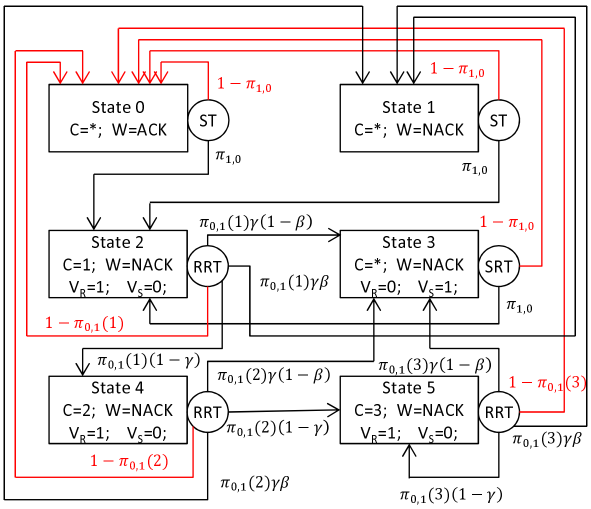
Under this approximation, the probabilistic protocol can be represented by a FSM with the state definitions shown in Figure 4b. The parameters V_{S}, V_{R} and W are the same as defined in Figure 2b, while there is one additional parameter, C, representing the number of successive retransmissions from R corresponding to the action RRT. The total number of states remains reasonable. The corresponding FSMC is shown in Figure 4a. Therefore, the corresponding state transition matrix for Type I decoder, Q_{I}, is now the 6×6 matrix in (16), which allows an easy performance evaluation for each value of γ and β.

$$Q_{\mathrm{I}} = \begin{pmatrix} 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 & 0 & 0 \\ 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 & 0 & 0 \\ 1-\pi_{[0,1]}(1) & \gamma\beta\pi_{[0,1]}(1) & 0 & \cdots & \cdots & \cdots \\ 1-\pi_{[1,0]} & 0 & \pi_{[1,0]} & 0 & 0 & 0 \\ 1-\pi_{[0,1]}(2) & \gamma\beta\pi_{[0,1]}(2) & 0 & \cdots & \cdots & \cdots \\ 1-\pi_{[0,1]}(3) & \gamma\beta\pi_{[0,1]}(3) & 0 & \cdots & \cdots & \cdots \end{pmatrix} \qquad (16)$$

The numerical results showing that this model provides a very accurate approximation of the performance of the probabilistic protocol with Type I decoder are provided in Section V.

D. FSMC model for performance approximation of the probabilistic protocol, Type II decoder

The optimization of the probabilistic protocol for Type II decoder using steady-state analysis of Z_{II} is very complex



(a)

State		Action
0	$C = *$, W=ACK	ST
1	$C = *$, W=NACK	ST
2	$C = 1$, $V_R = 1$, $V_S = 0$, W=NACK	RRT
3	$C = *$, $V_R = 0$, $V_S = 1$, W=NACK	SRT
4	$C = 2$, $V_R = 1$, $V_S = 0$, W=NACK	RRT
5	$C = 3$, $V_R = 1$, $V_S = 0$, W=NACK	RRT

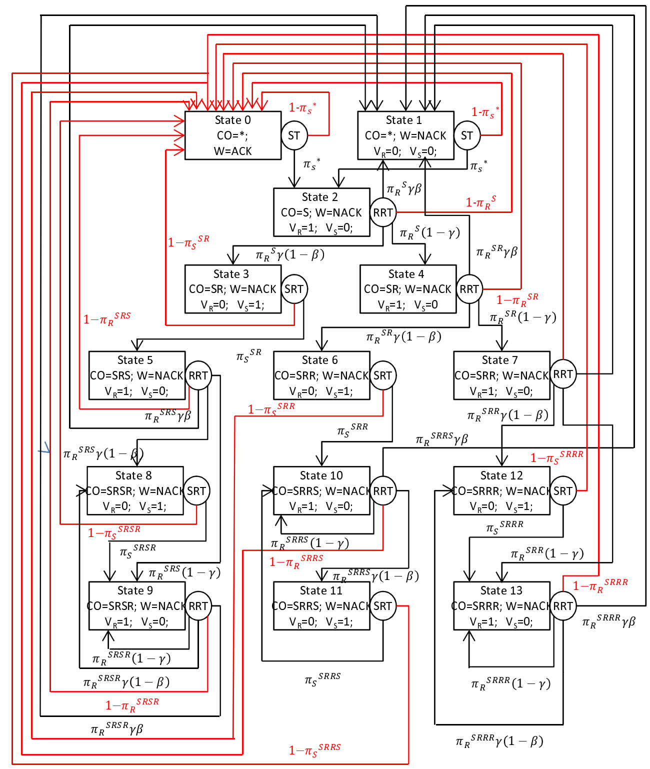
(b)

Fig. 4. FSMC description (a) and state definition (b) of the model for performance approximation of the probabilistic protocol, Type I decoder.

because $(1-\pi_S)$ and $(1-\pi_R)$ depend on the parameters (γ, β) target of the optimization, and because of the computational cost of getting the parameters $\pi_{[A,B]}$ as the values of A and B increase. As done for the case of Type I decoder, we want to develop a tractable FSMC model that allows reliable prediction of the performance of the protocol.

We consider the probabilistic protocol with a variation of the Type II decoder that simultaneously processes at most $C_{\max}$ received packets. We express this by saying that the decoder has a buffer limited to $C_{\max}$ entries. As the probabilistic protocol runs, the received copies at the decoder are stored in the buffer. If the buffer is filled to capacity $C_{\max}$ but a new copy is received, this overwrites an existing one, with the rule that a packet from the source can overwrite only packets from the source, and a packet from the relay can overwrite only packets from the relay. The packet to be replaced is chosen as the one with the worst quality using the criterion detailed in Section IV-D1. Notice that the limited buffer decoder for $C_{\max} \to \infty$ is the standard Type II decoder used by the probabilistic protocol. We recall that the intent in considering a finite $C_{\max}$ is to provide, as done for Type I decoding, a FSMC model that would be tractable for optimizing the parameters.

The states of the FSMC for decoder with limited buffer need to consider the possible filling levels of the buffer, in particular how many copies from the source and from the relay are in the buffer, but also the order with which the buffer has been filled for the first time. This is again related to the DMF mode of the relay, which implies that a copy from the relay carries different amounts of information depending on its position in the streak it belongs to. As a consequence the



(a)

State		Action
0	CO=*, W=ACK	ST
1	CO=*, $V_R = 0$, $V_S = 0$, W=NACK	ST
2	CO=S, $V_R = 1$, $V_S = 0$, W=NACK	RRT
3	CO=SR, $V_R = 0$, $V_S = 1$, W=NACK	SRT
4	CO=SR, $V_R = 1$, $V_S = 0$, W=NACK	RRT
5	CO=SRS, $V_R = 1$, $V_S = 0$, W=NACK	RRT
6	CO=SRR, $V_R = 0$, $V_S = 1$, W=NACK	SRT
7	CO=SRR, $V_R = 1$, $V_S = 0$, W=NACK	RRT
8	CO=SRSR, $V_R = 0$, $V_S = 1$, W=NACK	SRT
9	CO=SRSR, $V_R = 1$, $V_S = 0$, W=NACK	RRT
10	CO=SRRS, $V_R = 1$, $V_S = 0$, W=NACK	RRT
11	CO=SRRS, $V_R = 0$, $V_S = 1$, W=NACK	SRT
12	CO=SRRR, $V_R = 0$, $V_S = 1$, W=NACK	SRT
13	CO=SRRR, $V_R = 1$, $V_S = 0$, W=NACK	RRT

(b)

Fig. 5. FSMC description (a) and state definition (b) of the model for performance approximation of the probabilistic protocol, Type II decoder, $C_{\max} = 4$.

average probability of success in decoding at the destination depends not only on the number A of copies received from the source and the number B of copies received from the relay, but also on their order.

The states of the FSMC of the probabilistic protocol with limited buffer are defined in Figure 5b for the case of $C_{\max} = 4$. The parameters of state definitions V_S, V_R and W remain the same as defined in Figure 2b, and there is the additional parameter CO (Copy Order), representing the order with which the packets have filled the buffer, before the current time-slot. The number of states obviously increases with the size of the buffer $C_{\max}$. For size $C_{\max} = 4$ the number of states is 14, for $C_{\max} = 5$ the number of states is 23 and for $C_{\max} = 6$ the number of states is 37. The FSMC is represented in Figure 5a. Let Q_{II} be the transition matrix of the FSMC with state set defined by Figure 5b. It has the following form:

$$Q_{\mathrm{II}} = \begin{pmatrix} \left(1-\pi_{\mathrm{S}}^{*}\right) & 0 & \pi_{\mathrm{S}}^{*} & \cdots & \cdots \\ \left(1-\pi_{\mathrm{S}}^{*}\right) & 0 & \pi_{\mathrm{S}}^{*} & \cdots & \cdots \\ \left(1-\pi_{\mathrm{R}}^{\mathrm{S}}\right) & \gamma\cdot\beta\ \pi_{\mathrm{R}}^{\mathrm{S}} & \gamma\left(1-\beta\right)\pi_{\mathrm{R}}^{\mathrm{S}} & \cdots & \cdots \\ \left(1-\pi_{\mathrm{S}}^{\mathrm{SR}}\right) & \gamma\cdot\beta\ \pi_{\mathrm{S}}^{\mathrm{SR}} & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \left(1-\pi_{\mathrm{R}}^{\mathrm{SRRR}}\right) & \gamma\cdot\beta\ \pi_{\mathrm{R}}^{\mathrm{SRRR}} & 0 & \cdots & \cdots \end{pmatrix}. \tag{17}$$

The term $(1-\pi_{\mathrm{S}}^{\xi})$ represents the average probability of successful decoding at the destination, when the current transmission is performed by the source and the buffer has configuration $\mathrm{CO} = \xi$ just before the transmission happens. The term π_{R}^{ξ} is defined similarly. Notice that the average probabilities $(1-\pi_{\mathrm{R}}^{\xi})$ and $(1-\pi_{\mathrm{S}}^{\xi})$ relative to filling patterns ξ strictly shorter than $C_{\max}$ do not depend on parameters γ and β. This does not hold for ξ' of length $C_{\max}$. In this case the current retransmission results in an overwriting in the buffer, and the probability $(1-\pi_{\mathrm{R}}^{\xi'})$ denotes the average probability of success after the current transmission from the relay overwrites an element in the buffer. Let W_{R} be the number of substitutions that have happened in the buffer before the current time-slot, and let $q_{\mathrm{R}}(w)$ be the probability $\mathbb{P}(W_{\mathrm{R}} = w)$. Then the average probability $(1-\pi_{\mathrm{R}}^{\xi'})$ is defined as

$$(1-\pi_{\mathrm{R}}^{\xi'}) = \sum_{w}\left(\left(1-\pi_{\mathrm{R}}^{\xi'}(w)\right)|W_{\mathrm{R}} = w\right)\ q_{\mathrm{R}}(w). \tag{18}$$

Note that the probabilities $q_{\mathrm{R}}(w)$ are functions of the parameters β and γ. Since if $C_{\max}$ is not too small many overwritings are rare events, we choose to simplify the estimate, and we approximate (18) as the value taken for the first overwriting

$$(1-\pi_{\mathrm{R}}^{\xi'}) \approx \left(\left(1-\pi_{\mathrm{R}}^{\xi'}(1)\right)|W_{\mathrm{R}} = 1\right), \tag{19}$$

i.e. we make the approximation that the probability of success from the relay after performing $W_{\mathrm{R}} > 1$ overwritings remains unchanged compared to the probability of success after the first overwriting. With this approximation, then all probabilities in (17) do not depend on γ and β and can easily be evaluated using Monte Carlo simulations. The numerical results showing that this model provides a very accurate approximation of the performance of the probabilistic protocol with Type II decoder are provided in Section V.

Via steady-state analysis of Q_{I} and Q_{II} it is easy to predict the performance of the probabilistic protocol with Type I and Type II decoder. Moreover, it is possible to fine-sample the two parameters (γ,β) to find the values that optimize the performance with a reduced computational cost. Note that, by doing so, all possible choices of deterministic algorithms can be reached, and hence there is no need for choosing one of them as a basis for the optimization process. Via this optimization, we clearly observe two different regimes depending on the propagation quality of the physical channels:

1) **Bad medium**: in this regime it is rare event that a PDU can be successfully decoded with few transmissions, therefore decreasing the probability $\gamma\cdot\beta$ of dropping a non-acknowledged PDU causes $\overline{T}$ to increase and PER to decrease. In this region varying the values of γ and β affects the trade-off between PER and $\overline{T}$.
2) **Good medium**: in this regime it is a rare event that a PDU cannot be successfully decoded with few transmission. Therefore, decreasing the probability $\gamma\cdot\beta$ of dropping a non-acknowledged PDU causes PER to decrease without increasing $\overline{T}$. Then, since there is no trade-off to be achieved, choosing $\beta \to 0$, *i.e.* make the drop an almost impossible event, becomes the optimum strategy.

We illustrate the result of the optimization procedure in Section V.

1) Ranking of the received sequences at the Destination: For the overwriting process in the buffer, we need to rank the received sequences in order to overwrite the worst one. Note that sometimes the incoming copy could be worse than the worst copy inside the buffer: even in this case we choose to overwrite the worst copy. The ranking of the received sequences is based on the *a posteriori* probabilities of the coded bits, which can be evaluated as it follows:

$$\begin{aligned} p\left(c_{i,n}=0\,|\,y_n\right) &= \frac{p\left(y_n\,|\,c_{i,n}=0\right)}{p\left(y_n\,|\,c_{i,n}=0\right)+p\left(y_n\,|\,c_{i,n}=1\right)} \\ p\left(c_{i,n}=1\,|\,y_n\right) &= \frac{p\left(y_n\,|\,c_{i,n}=1\right)}{p\left(y_n\,|\,c_{i,n}=0\right)+p\left(y_n\,|\,c_{i,n}=1\right)} \end{aligned} \tag{20}$$

where i is the index of the coded bit of the n-th symbol, and

$$p\left(y_n\right)=p\left(y_n|c_{i,n}=0\right)p\left(c_{i,n}=0\right)+p\left(y_n|c_{i,n}=1\right)p\left(c_{i,n}=1\right) \tag{21}$$

and where, $p\left(c_{i,n}=0\right) = p\left(c_{i,n}=1\right) = 1/2$.

Then, we define B_s as the belief that hard demodulation of the received sequence with N symbols and k coded bits per symbol would produce the correct sequence of coded bits. This metric takes the form

$$B_s = \sum_{n=1}^{N}\left(\sum_{i=1}^{k}\overline{p}\left(c_{i,n}\,|\,y_n\right)\right), \tag{22}$$

where $\overline{p}\left(c_{i,n}\,|\,y_n\right) = \max\{p\left(c_{i,n}=0\,|\,y_n\right), p\left(c_{i,n}=1\,|\,y_n\right)\}$ for each coded bit of each symbol of the received sequence. A sequence with the higher value of B_s represents a better sequence. The worst sequence in the buffer is the one with the smallest B_s.

V. Numerical results

Here we present the numerical analysis of the probabilistic protocol using simulations and FSMC, for both types of decoders. The physical layer configuration is discussed in Section II. We use convolutional coding with code rate $R_c = 1/3$, and 16-QAM modulation. Each information PDU is of length 1000 bits. The average energy per modulated symbol is the same at the source and at the relay, $E_s = 1$, and as a result the transmit E_b/N_0 is the same on the channels SD and RD. The path-loss exponent is $\alpha = 2.4$. The probabilities $\pi_{[1,0]}$, $\pi_{[0,1]}(k)$, π_{S}^{ξ} and π_{R}^{ξ} needed for the FSMC representation can be evaluated using Monte Carlo simulations.

For both Type I and Type II decoder we initially perform the analysis when the relay is located at equal distance from the

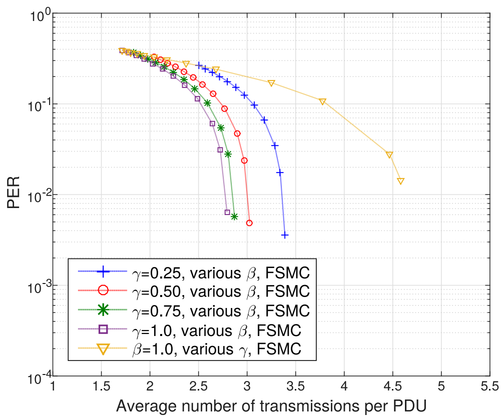


Fig. 6. Trade-off PER vs $\overline{T}$, several γ and β, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$, $E_b/N_0 = 5$ dB, Type I decoder.

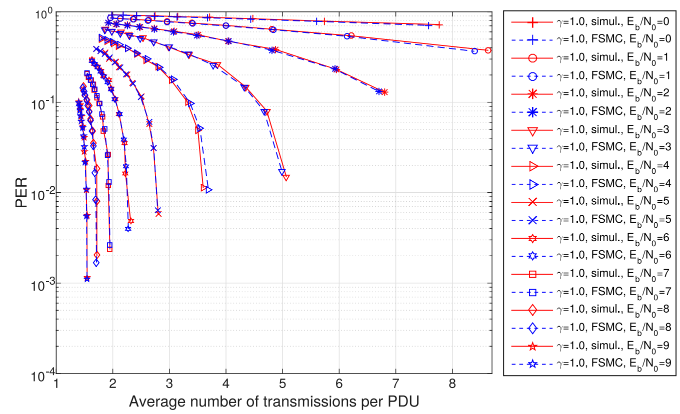


Fig. 7. Trade-off PER vs $\overline{T}$, $\gamma = 1$ and best β, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$, Type I decoder.

source and the destination, respectively, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$, and then we extend the analysis to other locations of the relay. By varying the values of γ and β, and using the FSMC analysis for evaluating the performance, one obtains the best possible combination of transmit parameters.

A. Numerical analysis of probabilistic protocol, Type I decoder

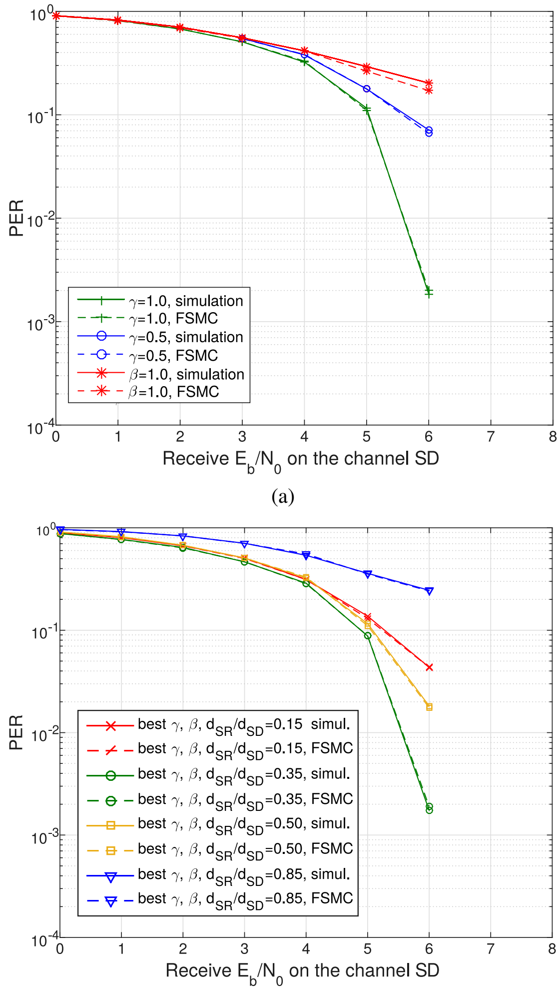
The parameter optimization procedure is repeated for each point of E_b/N_0. One example for $E_b/N_0 = 5$ is shown in Figure 6, which provides the trade-off between the PER and the average number of transmissions per PDU $\overline{T}$ corresponding to several choices of the values of γ and β. From Figure 6 we see the tendency that the best performance is achieved if we do not allow the relay to retransmit twice in a row (by setting $\gamma = 1.0$), as in this way we allow the relay to refresh the demodulated copy from the source before it can perform the next retransmission. This is due to the DMF mode, which makes the packet from the relay not as reliable as a packet from the source, even though the former has higher SNR at the destination due to the shorter distance from R to D. Following this tendency, we fix $\gamma = 1$ and we plot in Figure 7 the different trade-offs of PER vs $\overline{T}$ obtained by varying β, for various points of E_b/N_0. The prediction of the performance via the FSMC (dashed lines) is compared with the simulated protocol (solid lines), confirming that the FSMC analysis is accurate.

In Figure 7 are visible the bad and good medium regimes described in Section IV-B. For a given value of E_b/N_0, the PER decreases as the value of β decreases. This is due to the fact that β controls the probability that the system drops a packet. We observe that after some point, further decrease of β provokes a decrease in PER but $\overline{T}$ and Goodput remain unchanged. This is due to the fact that as the quality of the channel improves, the best strategy to decrease the PER is to allow for longer streaks of retransmissions, which however become rarer events and do not have major impact on $\overline{T}$. We refer to this as the *saturation point*. Therefore, in our analysis we are interested to make the optimization before the $\overline{T}$ reaches the saturation point.

Let us consider that we set up $\overline{T}_{\max} = 2.5$, and we want to optimize for $\overline{T} \leq \overline{T}_{\max} = 2.5$, by keeping the relay at the same distance as earlier, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.5$. We want to compare the following three cases: 1) we set the value of $\beta = 1.0$ (the source performs only the first transmission and the rest are performed by the relay), and we select γ that maximizes G under the constraint $\overline{T} \leq \overline{T}_{\max}$; 2) we set the value of $\gamma = 1.0$ (the relay is not allowed to retransmit successively two times), and we select β that maximizes G under $\overline{T} \leq \overline{T}_{\max}$; 3) we set the value of $\gamma = 0.5$ (we allow more successive retransmissions from the relay), and we select β that maximizes G under $\overline{T} \leq \overline{T}_{\max}$. We note that the optimization is performed independently for each point of E_b/N_0. The optimized PER from this comparison is shown in Figure 8a. As we may see, the best performance is achieved if we allow the relay to refresh its own demodulated copy from the source before the next retransmission (associated with $\gamma = 1.0$), which was also seen in Figure 6. Obviously this observation may change when the location of the relay changes: if the relay comes closer to the source, demodulation errors will be less likely, and refreshing demodulated copies will certainly not be a necessity.

Therefore, let us check what happens if the relay appears at other locations, by keeping the constraint $\overline{T}_{\max} = 2.5$. We perform the same analysis as before, and we plot the best PER curves at each location of the relay in Figure 8b as a function of the receive E_b/N_0 on the channel SD. The optimal values of γ and β for each location of the relay are in Table I.

As expected, it can be seen from Table I that, as the relay is located close to the source, for the lower points of E_b/N_0 the best performance it is achieved if we allow only the relay to perform successive retransmissions (associated with $\beta = 1.0$). As the channel improves then the best performance it is achieved with $\gamma = 0.5$ (successive retransmissions by R allowed) and at the same time by leaving some probability that relay refreshes its own demodulated copy from the source (associated with $\beta < 1$). As the relay moves further away from the source, we see that the best performance it is achieved if we do not allow the relay to retransmit successively more than once (associated with $\gamma = 1.0$), and we push the



(a) (b)

Fig. 8. PER obtained optimizing one parameter at the time (a), and optimizing γ and β jointly (b), various $d_{\mathrm{SR}}/d_{\mathrm{SD}}$, limited $\overline{T} = 2.5$, Type I decoder.

	$d_{\mathrm{SR}}/d_{\mathrm{SD}}$			
E_b/N_0	0.15	0.35	0.50	0.85
0	$\gamma = 0.6315$ $\beta = 1.0$	$\gamma = 0.6280$ $\beta = 1.0$	$\gamma = 1.0$ $\beta = 0.7740$	$\gamma = 1.0$ $\beta = 0.7845$
1	$\gamma = 0.5960$ $\beta = 1.0$	$\gamma = 0.5845$ $\beta = 1.0$	$\gamma = 1.0$ $\beta = 0.7440$	$\gamma = 1.0$ $\beta = 0.7665$
2	$\gamma = 0.5335$ $\beta = 1.0$	$\gamma = 0.5220$ $\beta = 1.0$	$\gamma = 1.0$ $\beta = 0.6895$	$\gamma = 1.0$ $\beta = 0.7325$
3	$\gamma = 0.4390$ $\beta = 1.0$	$\gamma = 0.50$ $\beta = 0.8840$	$\gamma = 1.0$ $\beta = 0.5950$	$\gamma = 1.0$ $\beta = 0.6710$
4	$\gamma = 0.3065$ $\beta = 1.0$	$\gamma = 1.0$ $\beta = 0.4195$	$\gamma = 1.0$ $\beta = 0.4555$	$\gamma = 1.0$ $\beta = 0.5850$
5	$\gamma = 0.50$ $\beta = 0.3495$	$\gamma = 1.0$ $\beta = 0.1610$	$\gamma = 1.0$ $\beta = 0.1940$	$\gamma = 1.0$ $\beta = 0.4350$
6	$\gamma = 0.50$ $\beta = 0.1670$	$\gamma = 1.0$ $\beta = 0.005$	$\gamma = 1.0$ $\beta = 0.0445$	$\gamma = 1.0$ $\beta = 0.3915$

TABLE I
OPTIMIZED VALUES OF γ AND β, VARIOUS $d_{\mathrm{SR}}/d_{\mathrm{SD}}$ AND E_b/N_0, TYPE I DECODER.

relay to refresh its own demodulated copy before the next retransmission (associated with $\beta < 1$).

A global conclusion from Figure 8b is that the highest performance it is achieved when the relay is located at distance $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.35$. This means that as we move the relay from

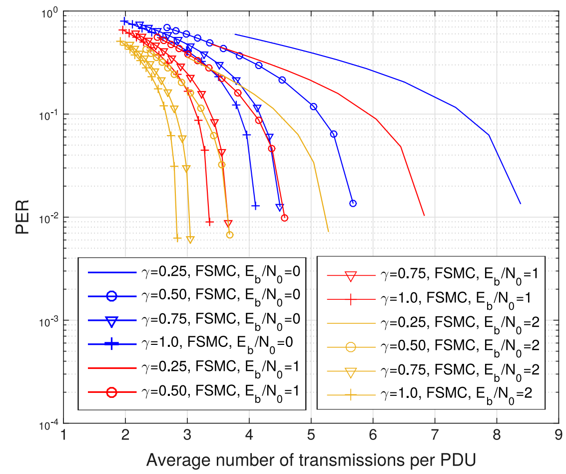


Fig. 9. Trade-off PER vs $\overline{T}$, several γ and β, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$, Type II decoder.

the source toward the destination, the performance improves up to distance $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.35$, where it has its maximum performance, and then the performance decreases as the relay moves closer to the destination, due to the high number of demodulation errors. In other words, the optimized location accounts for a trade-off between the demodulation errors on the channel SR and successful retransmissions for the remaining distance RD, which in our case appears to be in 1/3 of the distance SD.

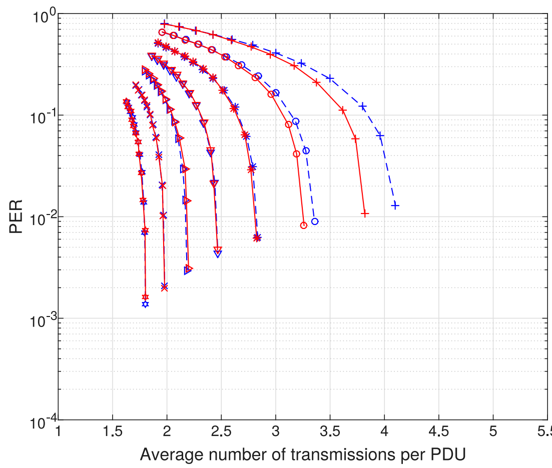
B. Numerical analysis of the probabilistic protocol, Type II decoder

The same procedure is applied for the case of Type II decoder. Initially, we stick to the location of the relay at distance $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$. The results of the optimization procedure for various values of γ and β for few points of E_b/N_0 are shown in Figure 9, where PER is expressed vs the average number of transmissions per PDU, $\overline{T}$.

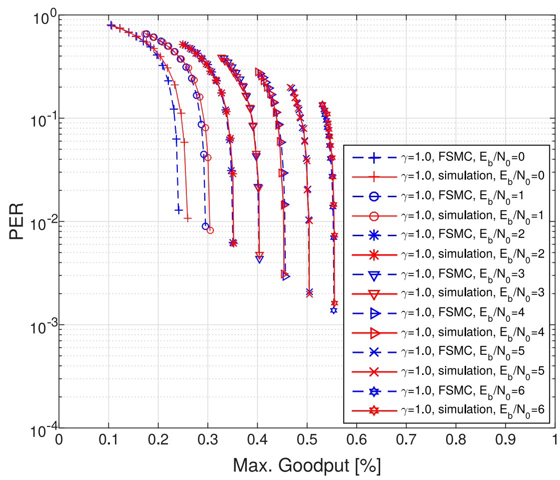
From Figure 9 we see that also in the case of Type II decoder there is tendency that the best performance is achieved if we set $\gamma = 1.0$. Following this tendency, then in Figures 10a and 10b there are shown the comparison of PER vs corresponding $\overline{T}$ and Goodput, respectively, for various points of E_b/N_0 when $\gamma = 1.0$ and continuosly decreasing β. As we may see from both Figures, performance prediction of the probabilistic protocol via the FSMC analysis is accurate, except for high levels of noise on the channels. Here we can also verify the presence of bad and good regimes as described in Section IV-B, and see the saturation point. Like on the case of Type I decoder, we are interested to make the optimization before the $\overline{T}$ reaches the saturation point.

Keeping the constraint that $\overline{T} \leq \overline{T}_{\max} = 2.5$, then the comparison for various values of γ and β is shown in Figure 11. Even here, it is confirmed that the best performance it is achieved if we allow R to refresh its own demodulated copy from S before the next retransmission (associated with $\gamma = 1.0$).

In the end, we perform the same optimization procedure for the other locations of the relay by keeping the constraint



(a)



(b)

Fig. 10. Trade-off PER vs $\overline{T}$ (a) and trade-off PER vs goodput G (b), $\gamma = 1$ and best β, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$, Type II decoder.

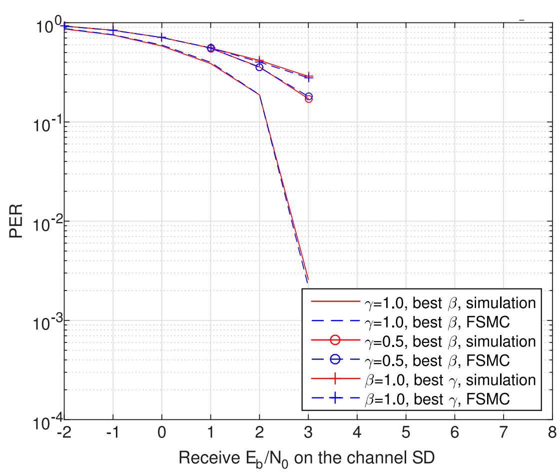
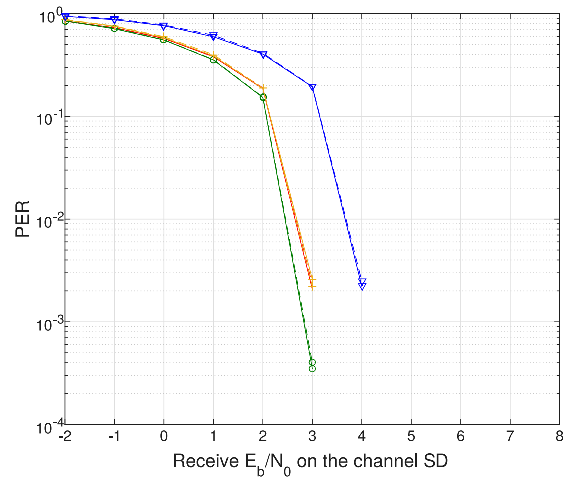
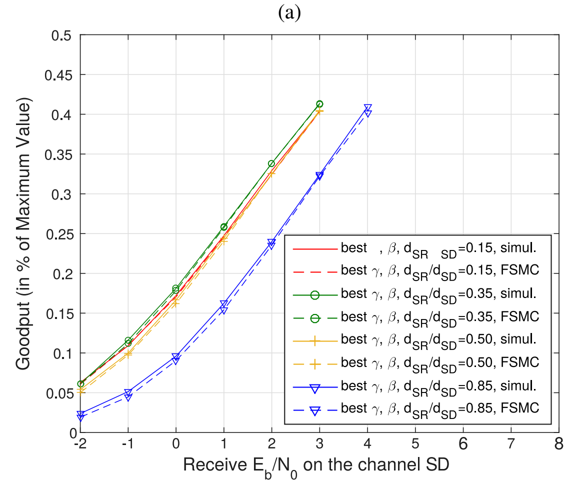


Fig. 11. PER obtained optimizing one parameter at a time, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.50$, limited $\overline{T} = 2.5$, Type II decoder.



(a)



(b)

Fig. 12. PER (a) and Goodput G (b) obtained optimizing both γ and β, various $d_{\mathrm{SR}}/d_{\mathrm{SD}}$, limited $\overline{T} = 2.5$, Type II decoder.

	$d_{\mathrm{SR}}/d_{\mathrm{SD}}$			
E_b/N_0	0.15	0.35	0.50	0.85
-2	$\gamma = 1.0$ $\beta = 0.76$	$\gamma = 1.0$ $\beta = 0.7595$	$\gamma = 1.0$ $\beta = 0.7670$	$\gamma = 1.0$ $\beta = 0.7850$
-1	$\gamma = 1.0$ $\beta = 0.7165$	$\gamma = 1.0$ $\beta = 0.7160$	$\gamma = 1.0$ $\beta = 0.7265$	$\gamma = 1.0$ $\beta = 0.7655$
0	$\gamma = 1.0$ $\beta = 0.6485$	$\gamma = 1.0$ $\beta = 0.6410$	$\gamma = 1.0$ $\beta = 0.6575$	$\gamma = 1.0$ $\beta = 0.7250$
1	$\gamma = 1.0$ $\beta = 0.5255$	$\gamma = 1.0$ $\beta = 0.5045$	$\gamma = 1.0$ $\beta = 0.5360$	$\gamma = 1.0$ $\beta = 0.65$
2	$\gamma = 1.0$ $\beta = 0.3230$	$\gamma = 1.0$ $\beta = 0.2795$	$\gamma = 1.0$ $\beta = 0.3230$	$\gamma = 1.0$ $\beta = 0.5175$
3	$\gamma = 1.0$ $\beta = 0.005$	$\gamma = 1.0$ $\beta = 0.001$	$\gamma = 1.0$ $\beta = 0.005$	$\gamma = 1.0$ $\beta = 0.3050$
4	———- ———-	———- ———-	———- ———-	$\gamma = 1.0$ $\beta = 0.005$

TABLE II
OPTIMIZED VALUES OF γ AND β, VARIOUS $d_{\mathrm{SR}}/d_{\mathrm{SD}}$ AND E_b/N_0, TYPE II DECODER.

$\overline{T}_{\max} = 2.5$, and the best curves at each location of the relay are shown in Figures 12a and 12b, comparing of PER and Goodput as a function of various points of receive E_b/N_0 on the channel SD, respectively. The results confirm that even in the case of Type II decoder, the highest diversity is achieved when the relay is located at distance $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.35$. The details of the optimized values of γ and β for each location of the relay are shown in Table II.

It can be concluded from the above simulations that the

E_b/N_0	
-2	$\gamma = 0.9975$, $\beta = 0.59$
-1	$\gamma = 0.995$, $\beta = 0.5325$
0	$\gamma = 0.9975$, $\beta = 0.45$
1	$\gamma = 1.0$, $\beta = 0.35$
2	$\gamma = 0.9975$, $\beta = 0.24$
3	$\gamma = 0.9925$, $\beta = 0.11$
4	$\gamma = 1.0$, $\beta = 0.03$
5	$\gamma = 1.0$, $\beta = 0.005$

TABLE III
VALUES OF γ AND β FOR EACH POINT OF E_b/N_0, IN THE COMPARISON OF PROBABILISTIC WITH DETERMINISTIC PROTOCOL.

constraints on the delay may play a large role in the attainable performance in terms of goodput. This is quite difficult to evaluate in general for the probabilistic protocol, since there is no real control of this parameter with reference to classical deterministic protocols. Obviously, deterministic protocols also have the same interaction, but the delay is mostlly determined once the protocol is chosen. Therefore, we have chosen to compare our proposed probabilistic protocols with a reference one, found in the literature.

C. Comparison of the achievable performance of deterministic and probabilistic protocols

The aim of this section is to compare the performance of the probabilistic protocol with the performance attainable by any deterministic strategy defined on the same network.

We consider the S-R-D network, with relay working in the DMF mode. We consider Type II decoding at the destination. We want to explore all deterministic protocols such that the maximum number of transmissions per PDU is $N_{\max} = 3$. There are four deterministic protocols corresponding to this constraint, which we distinguish by the order with which the source S and the relay R take turns in transmitting: det(S,S,S), det(S,S,R), det(S,R,S), det(S,R,R). We want to compare the performance of the probabilistic protocol to the performance of all possible deterministic strategies. The chosen metric is the Goodput, which expresses, for each strategy, the trade-off between the quality of the received data (the PER), and the average cost per information PDU.

We impose that the probabilistic protocol cannot spend more resources per information PDU of what is allowed to a deterministic protocol. We then define $\overline{T}_{\max}$ as the biggest average number of transmissions per information PDU generated by a deterministic protocol, and we optimize the parameters of the probabilistic protocol under the constraint $\overline{T} \leq \overline{T}_{\max}$. As it will be verified by simulations, the value $\overline{T}_{\max}$ corresponds to the protocol det(S,S,S).

In the simulations, we set $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.35$. We consider convolutional coding with code rate $R_c = 1/3$, and 16-QAM modulation. The comparison between the protocols is performed for each point of E_b/N_0 on the channel SD. The optimized values of the parameters γ and β of the probabilistic protocol are given in Table III.

Figure 13 presents the average number of transmissions per PDU $\overline{T}$ obtained using the deterministic and the probabilistic protocols. This allows to verify that $\overline{T}_{\max}$ corresponds to the det(S,S,S) protocol, and to verify that the probabilistic

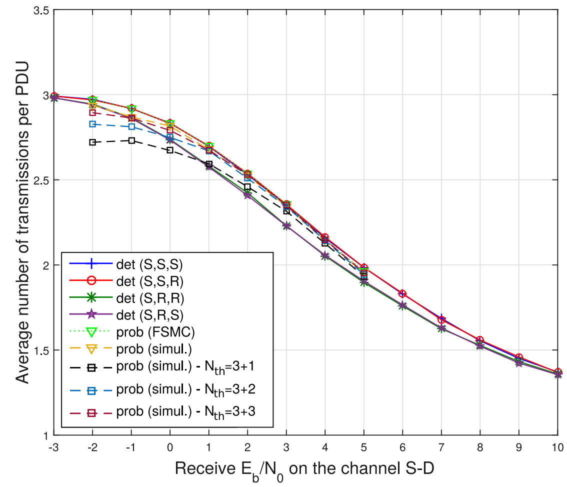


Fig. 13. Comparison of $\overline{T}$ achieved with the deterministic and the probabilistic protocols, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.35$, 16-QAM modulation scheme.

protocol respects the constraint $\overline{T} \leq \overline{T}_{\max}$. The comparison of PER as a function of receive E_b/N_0 on the SD channel is shown in Figure 14a. As we can see, the probabilistic protocol greatly outperforms any deterministic strategy in terms of PER, while keeping the constraint $\overline{T} \leq \overline{T}_{\max}$. This increased efficiency with respect to the deterministic strategy is visible when looking at the goodput, in Figure 14b. The best deterministic strategy consists of the protocol det(S,R,S), which minimizes the PER achievable with a deterministic protocol and is associated with the smallest $\overline{T}$.

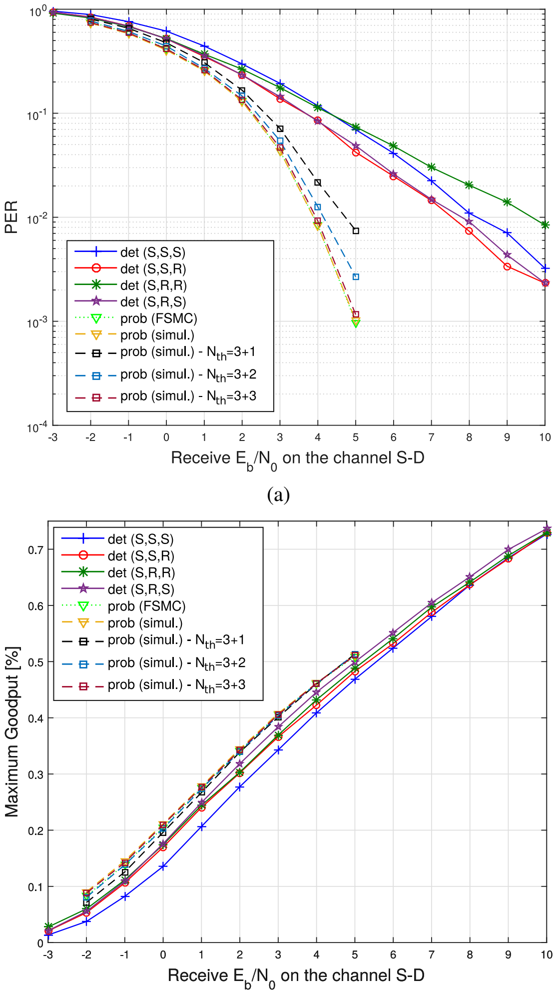
We repeat the same experiment with a variation of the modulation scheme, in order to verify that the probabilistic protocol consistently outperforms any deterministic strategy. Figures 15a and 15b show the results for BPSK and 4-QAM modulations schemes, respectively.

This experiment verifies that the probabilistic protocol is able not only to match the performance of any deterministic strategy, but even to improve it.

We underline the fact that the optimization of the deterministic protocol has required to explore all possible strategies combinatorially. This operation becomes very complex as the number of maximum allowed transmissions $N_{\max}$ increases. On the other hand, the complexity of the optimization of the probabilistic protocol does not increase as $\overline{T}_{\max}$ increases.

Since the number of retransmissions in the probabilistic protocol is potentially unbounded, this may give rise, occasionally, to very long delays, which may be incompatible with some applications where latency is a very important parameter. We hence explore a variation of the practical implementation of the probabilistic protocol, where the maximum number of transmissions is constrained to be less than a finite value $N_{\mathrm{th}} \geq N_{\max}$, while the values of the parameters γ and β are chosen according to Table III, corresponding to the case $N_{\mathrm{th}} = \infty$.

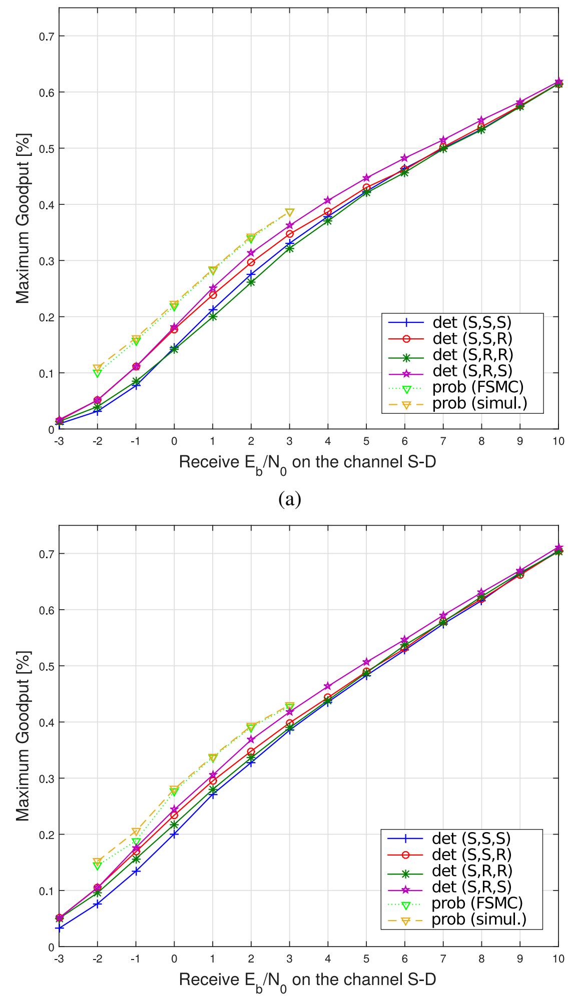
As visible in Figure 13, the value of $\overline{T}$ for any finite N_{th} never exceeds the value obtained for $N_{\mathrm{th}} = \infty$. Intuitively, smaller values of N_{th} yield lower $\overline{T}$. The effect of finite N_{th} is degradation of the PER, as visible in Figure 14a. However,



(a)

(b)

Fig. 14. Comparison of PER (a) and goodput G (b) achieved with the deterministic and the probabilistic protocols, $d_{\mathrm{SR}}/d_{\mathrm{SD}} = 0.35$, 16-QAM modulation scheme.



(a)

(b)

Fig. 15. Comparison of Goodput achieved with the deterministic and the probabilistic protocols for (a) BPSK modulation scheme and (b) 4-QAM modulation scheme.

the modest value $N_{\mathrm{th}} = N_{\mathrm{max}} + 3$ already gives a performance which is very close to the case $N_{\mathrm{th}} = \infty$, as it is observed in Figure 14a. For any considered finite N_{th}, the degradation of the PER is somewhat compensated by the gain in $\overline{T}$, so that the performance in terms of goodput G is very close to the case of $N_{\mathrm{th}} = \infty$, as visible in Figure 14b.

In conclusion, the performance for finite N_{th} converges to the performance of the probabilistic protocol with unbounded number of transmissions, as N_{th} increases; very modest values of N_{th} already approach the optimum performance.

VI. Conclusions

In this paper we have proposed a HARQ probabilistic protocol in a cooperative network. Starting with an example of deterministic protocol on a Source-Relay-Destination network, we have shown that its performance analysis using Finite State Markov Chain (FSMC) becomes more complex as the protocol gets more sophisticated. Therefore, we have shown how can we simplify its FSMC analysis, and moreover, we have shown that the simplified FSM with four-states can be associated with probabilistic protocol which contains two parameters that can be adjusted for performance optimization. Once this transmit protocol has been defined, we studied the optimization of the involved parameters, by a similar strategy: defining FSMCs with reduced number of states that closely approximate the performance of the decoder. We have shown that for Type I decoder we can use a six-states transition matrix for performance prediction and optimization, while for Type II decoder we need a larger matrix, but which still makes possible a much easier and more efficient optimization of the performance than for any deterministic protocol. Monte Carlo simulations and theoretical results show that the performance as evaluated via these FSMCs matches almost exactly the performance of the actual system, and that its performance can be adjusted easily such that it outperforms that of the deterministic protocols. Note also that we concentrated on the demodulate and forward strategy at the relay, because it is the simplest one to implement at the relay, a fact which could prove to be very useful in complexity constrained situations, such as those met in Internet of Things (low throughput, low energy device to device communication).

References

[1] Hoang Anh Ngo, Lajos Hanzo, "Hybrid Automatic-Repeat-reQuest Systems for Cooperative Wireless Communications," *IEEE Communications Surveys & Tutorials*, Year: 2014, Volume: 16, Issue: 1.

[2] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communications in wireless networks," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74-80, 2004.

[3] B. Zhao and M. Valenti, "Practical relay networks: A generalization of hybrid-ARQ," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 7-18, 2005.

[4] S. Tomasin, M. Levorato, and M. Zorzi, "Analysis of outage probability for cooperative networks with HARQ," in Proc. *IEEE International Symposium on Information Theory* (ISIT 2007), pp. 2716-2720, 2007.

[5] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, "Energy efficiency of non-collaborative and collaborative Hybrid-ARQ protocols," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 326-335, 2009.

[6] Frederick C. Hennie, "Finite-State Models for Logical Machines", *Wiley*, New York, 1968.

[7] D. Xu, Ch. Yang, Y. Yao, Ch. Zhang and B. Xia, "Performance Analysis of Bursty Service in Relay Systems with ARQ Protocol", *IEEE International Conference on Wireless Communications & Signal Processing (WCSP)*, Pages: 1-5, 2016.

[8] Q. Li, M. Yu, A. Pandharipande, T. Han, J.Zhang and X.Ge, "Cooperative Two-Path Relay Channels: Performance Analysis Using A Markov Framework", *IEEE International Conference on Communications (ICC)*, Pages: 3573-3578, 2015.

[9] M.S. Fazel Falavarjani, R. Hoshyar, and R. Tafazolli, "Performance Evaluation of A Flexible Amplify and Forward (AF) Combined with HARQ", *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Pages: 488-493, 2010.

[10] D. Chen and J. N. Laneman, “Modulation and demodulation for cooperative diversity in wireless systems”, *IEEE Transactions on Wireless Communications*, vol. 5, no. 7, pp. 1785-1794, July 2006.

[11] A. Lindgren, A. Doria, and O. Scheln, "Probabilistic routing in intermittently connected networks", *ACM SIGMOBILE mobile computing and communications review*, vol.7, no.3, pp.19-20, July 2003.

[12] J. Fukuyama, "A Probabilistic Protocol for Multi-Hop Routing in VANETs", *ICC Workshops*, 14-18 June 2009.

[13] F. Maliqi, F. Bassi, P. Duhamel, and I. Limani. “A Probabilistic Retransmission Protocol on a Relaying Network”, *IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC)*, Oct 2017, Montreal, Canada.

[14] T. V. Ramabadran, S. S. Gaitonde, " A tutorial on CRC computations", *IEEE Journals & Magazines*, Volume: 8, Issue: 4, Pages: 62-75, Year: 1988.

[15] J.R.Norris, "Markov Chains", *Cambridge University Press*, 1997.

[16] D. Chase, "Code combining: A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Transactions on Communications*, Vol. Com-33, No.5, May 1985.

[17] Robert G. Gallager, "Stochastic Processes, Theory for Applications", *Cambridge University Press*, book chapter 4, February 2014.

[18] H. De Sterck, Thomas A. Manteuffel, Stephen F. Mccormick, Quoc Nguyen, and John Ruge, "Multilevel adaptive aggregation for Markov Chains, with application to web ranking," *SIAM Journal on Scientific Computing*, Volume:30, Issue:5, Pages: 2235-2262, Year: 2008.

Faton Maliqi was born in Kosovo in 1983. He received the BSc degree in 2005 and the MSc degree in 2012, both in Electrical Engineering from the University of Prishtina, and the PhD degree in 2017 in Networks, Information and Communications under the supervision of Dr. Pierre Duhamel from the Laboratory of Signals and Systems (L2S), University of Paris-Sud, affiliated also with CentraleSupelec and CNRS, in Paris, France. From 2005 to 2014 he held several positions as Electrical Engineer at several companies in Kosovo and USA (remote work for AT&T via Tellabs). From 2011 he has been continuously teaching at the Faculty of Electrical and Computer Engineering, University of Prishtina. His research is focused mainly on signal processing tools for communications, in which field he has published several papers.

Francesca Bassi was born in Cremona, Italy. She received the BSc and the MSc degrees in Telecommunications Engineering from the University of Bologna, Italy, in 2003 and 2006 respectively. She received the Ph.D. degree from University Paris-Sud, Orsay, France, and from the University of Bologna, in 2010. In 2007 and 2009, she has been a visiting Ph.D. student at the University of Bologna. From 2006 to 2010, she was with Laboratoire des Signaux et des Systèmes, Gif-sur-Yvette, France. Between 2011 and 2012, she was a Postdoctoral Researcher at the University of Michigan, Ann Arbor, MI, USA. Since 2012, she has been an Assistant Professor at ESME-Sudria, Ivry-sur-Seine, France, and an Associate Researcher with the Laboratoire des Signaux et des Systèmes.

Pierre Duhamel (Fellow, IEEE, 1998, Fellow, Eurasip, 2008, Grand Prix France Télécom of the French Science Academy, 2000) received the Eng. Degree in Electrical Engineering from the National Institute for Applied Sciences (INSA) Rennes, France in 1975, the Dr. Eng. Degree in 1978, and the Doctorat ès sciences degree in 1986, both from Orsay University, Orsay, France.

From 1975 to 1980, he was with Thomson-CSF (now Thales), Paris, France, where his research interests were in circuit theory and signal processing. In 1980, he joined the National Research Center in Telecommunications (CNET, now Orange Labs), Issy les Moulineaux, France, where his research activities were mainly concerned with the design of recursive CCD filters. Later, he worked on fast algorithms for computing various signal processing functions (FFT’s, convolutions, adaptive filtering, wavelets). From 1993 to Sept. 2000, he has been professor at ENST, Paris (National School of Engineering in Telecommunications) with research activities focused on Signal processing for Communications. He was head of the Signal and Image processing Department from 1997 to 2000. He is now with CNRS/L2S (Laboratoire de Signaux et Systèmes, Gif sur Yvette, France), where he developed studies in Signal processing for communications and signal/image processing for multimedia applications, including joint source- protocol-channel coding/decoding. He is currently investigating the connections between communication theory and networking as well as Information theory and AI.

He was Distinguished lecturer, IEEE, for 1999, and was co-technical chair of ICASSP 06, Toulouse, France, and of WCNC 2012, Paris, France. The paper on subspace-based methods for blind equalization, which he coauthored, received the “Best paper award” from the IEEE Transactions on SP in 1998. He was awarded the “grand prix France Telecom” by the French Science Academy in 2000.

Dr Duhamel published more than 100 papers in international journals, more than 300 papers in international conferences, and holds 29 patents. He is a co-author of the book “Joint Source and Channel Decoding: A cross layer perspective with applications in video broadcasting” which appeared in 2009, Academic Press. He successfully advised or co-advised 60 PhD students, two of them being now fellow, IEEE.

Ilir Limani received Dipl.Ing. degree in Electrical Engineering from University of Prishtina, in 1978 and Mr.sc. degree in Electrical Engineering from University of Prishtina, in 1988. He received Ph.D. degree in Electrical Engineering from University of Prishtina in 1997. He is currently Professor in the department of Automatic Control, and Communications, Faculty of Electrical and Computer Engineering, University of Prishtina. His research area of interests include statistical signal processing, system identification and recursive estimation.