Analytical Performance Derivation of Hybrid ARQ Schemes at IP layer

Christophe LE MARTRET [1#], *Aude LE DUC* [2],
Philippe CIBLAT [3], *Sébastien MARCILLE* [1,3]

Abstract

In this paper we derive performance metrics closed-form expressions of retransmission schemes such as Automatic Repeat reQuest (ARQ) and Hybrid ARQ (HARQ) in the case of memoryless block fading channels. As performance metrics, we consider the Packet Error Rate (PER), the efficiency, and the packet delay. The novelty of the paper is twofold: i) the metrics are considered at the Network level; ii) we introduce a new general framework which enables us to derive analytically the considered performance metrics for most retransmission schemes, including recent cross-layer strategies with the Network layer. The metrics at the Medium Access Control (MAC) level are obtained as a byproduct. Among the considered retransmission schemes, we especially consider the Incremental Redundancy HARQ scheme with different lengths redundancy packets.

Index Terms

Hybrid automatic repeat request, cross-layer, incremental redundancy, packet delay, packet error rate, efficiency.

[#] Contact author.

[1] THALES Land and Joint Systems, 160 Boulevard de Valmy, B.P. 82, 92704 Colombes Cedex, France.
Email: christophe.le_martret@fr.thalesgroup.com, sebastien.marcille@fr.thalesgroup.com.
Tel: +33 146132735, fax: +33 146132555.

[2] ESME Sudria
38, rue Molière, 94200 Ivry-sur-Seine, France. Email: leduc@esme.fr.

[3] Département Communications et Électronique, Telecom ParisTech,
46 rue Barrault, 75013 Paris, France. Tel: +33 145817728, fax: +33 145890020.
Email: philippe.ciblat@telecom-paristech.fr.

Part of this work has been published in IEEE conferences VTC'09, SPAWC'09, and SPAWC'10.

I. Introduction

Retransmission schemes like Automatic Repeat reQuest (ARQ) allow to establish reliable wireless links by retransmitting corrupted packets upon error detection. The Hybrid ARQ (HARQ) is an evolution of the ARQ that associates retransmission mechanisms with channel coding. Various types of HARQ have been proposed (see [1] and references herein for a global survey) and they are now included in modern standards such as IEEE802.16 [2] and 3GPP LTE [3]. These retransmisison schemes are implemented by the Medium Access Control (MAC) layer and their performance evaluation are usually done at the MAC layer.

Until a few years ago, communication systems were designed by considering independently the different layers following the spirit of the OSI model. Recently, the concept of the cross-layer design consisting in optimizing jointly the different layers has been introduced in order to improve the whole network performance. In [4], the authors have proposed an interesting performance improvement of the ARQ schemes by taking into account that successive MAC packets on which the ARQ scheme is applied independently belong to the same Network packet. As seen later in our paper, the work in [4] can be easily extended to any HARQ scheme. As a consequence, the performance metrics for the cross-layer strategy [4] are computed at the Network layer, and it is thus necessary to study the performance of the other retransmission schemes at the Network layer as well in order to compare them.

In order to have a complete view on the system performance, it is necessary to consider the Packet Error Rate (PER), the efficiency and the delay as performance metrics. Let us firstly consider the state-of-the-art concerning the analytical performance derivations at the MAC layer. In the context of infinite number of transmissions per packet, the efficiency and the packet delay have been computed analytically for HARQ schemes and can be found in [1], [5], [6], [7]. When the maximum number of transmissions per packet is finite, which is a more realistic assumption, the derivations are rather difficult: i) the Packet Error Rate (which does not vanish anymore for finite SNR) can be found in [1], [8] for any type of HARQ; ii) the efficiency in [9] for ARQ, in [10] for Chase-Combining HARQ (CC-HARQ) in [8], [11] for Incremental Redundancy HARQ (IR-HARQ), and in [12] for almost any type of HARQ; iii) the delays in [8], [12] for different types of HARQ. At the Network layer, to our best knowledge, only [4] focused on the analytical performance derivations. More precisely, in [4] one can find the expressions of the PER and the delay for ARQ scheme with their optimized cross-layer strategy and the standard one. One has to notice that in all above mentioned papers, the length of the transmitted packets is assumed to be identical which is not necessary true in practice. Unequal lengths for the transmitted packets lead to a non-incremental modification of the closed-form expressions of the efficiency (even at the MAC layer) and will be treated in the paper. Therefore *the goal of this paper is to derive in closed-form the PER, the efficiency, and the delay for a wide range of HARQ schemes, encompassing both conventional and cross-layer strategies, in the case of memoryless block fading channels at the Network layer*. As a byproduct, some new closed-form expressions will be provided at the MAC layer.

The document is organized as follows. In Section II we first depict the considered HARQ schemes, the layer model, the cross-layer strategy, as well as the performance metrics definitions. In Section III, we derive the closed-form expressions for the different metrics. We especially prove that the cross-layer strategy is always better in terms of PER, whatever the considered retransmission scheme. In Section V, we show numerical results of performance metrics for two different HARQ schemes.

II. Communication Scheme Description

A. The Layer Model

As depicted in Fig. 1, the layer model considered in this paper encompasses physical (PHY), MAC and Network layers. We assume that the Network layer is an Internet Protocol (IP) but our work can be applied to any Network layer protocol. In the sequel, the Network layer will be thus called IP layer/level. At the transmitter side, the MAC layer gets IP packets (IPP) of length L_{IPP} from IP layer. The kth IP packet, denoted by IPP(k), is fragmented by the Segmentation And Reassembly (SAR) block into N fragments (shorten to FR in the sequel) $\{\text{FR}_k(n)\}_{n=1}^{N}$ of length $L_{\text{FR}} = L_{\text{IPP}}/N$ (we neglect here the overhead added by the SAR). These packets are fed into the ReTransmission Manager (RTM) block which implements the retransmission scheme and generates for each fragment $\text{FR}_k(n)$ a sequence of packets usually called MAC packets (MP) $\text{MP}_{k,n}(m)$. The number of MP for a given (k, n) depends on the retransmission scheme considered and will be specified later on. Depending on the retransmission scheme, the MP may have different lengths and we note δ_m the length of $\text{MP}_{k,n}(m)$. The MP is then provided to the Physical layer in order to be modulated and sent through the propagation channel into a frame structure. After propagating through the channel, the signal is demodulated and the reconstituted MP is sent to the MAC layer into the RTM. The RTM processes the incoming MP and decides whether the transmission is successful or not and sends an ACKnowledgment (ACK) or a Negative ACKnowledgment (NACK) back to the transmitter accordingly. The RTM then delivers fragments that are sent to the SAR that reconstitutes the IPP. This scheme corresponds to a Stop-and-Wait one that covers most of the actual existing usages (in a multiple transmission mode). In the rest of the document, we will drop the subscripts for FR and MP for simplicity.

B. Retransmission Protocol Description

In order to give an accurate description of the different retransmission schemes, we propose to distinguish between the manipulations operated at the transmitter side, called the Retransmission Mechanism (RM), and the manipulations performed at the receiver side, called the Receiver Processing (RP). Note that usually the standards depict the RM while the RP is left to the manufacturer. Although our description framework fits to any retransmission protocol, due to the lack of space we will only focus on a few selected schemes below. In the rest of the document, we say that a packet is "received" (or equivalently "OK") when it is received without error.

1) Retransmission Mechanisms: We describe three RMs considered in this paper. The RM describes the way each FR is transformed into MPs and how the transmitter reacts upon continuously receiving NACKs (as soon as an ACK is received, each RM starts again the process with the next FR). For each RM, we assume that the FR includes the payload, the header and the Cyclic Redundancy Check (CRC) parity bits to check the packet integrity at the receiver side. In this paper we consider *truncated* HARQ schemes with a finite number of transmissions per MP as introduced in [13] in order to bound the maximum transmission delay. At the MAC layer, the *maximum transmission credit per FR* is denoted by $P_{\max}$ and will play an important role in the HARQ performance.

RM1: The MP is the FR. The transmitter retransmits the same MP until $P_{\max}$ is reached.

RM2: The MP is obtained by encoding the FR with a FEC of rate R. Then it works as RM1.

RM3: The FR is firstly encoded by a FEC of rate R_0 (known as the *mother code*). The redundancy bits are then split into n_{IR} MPs by following the *rate compatible coding* principle introduced in [14]. The length of the MPs are in general not constant and depend upon the selected set of puncturing patterns. The transmitter transmits the MP sequence $\{\mathrm{MP}(i)\}_{i=1}^{n_{\mathrm{IR}}}$ in a cyclic way until the maximum number of transmissions per FR is reached.

2) Receiver Processing: We describe four RPs considered in this paper. The RP depicts the way the received MPs are processed in order to decide if the decoded FR is corrupted or not.

RP1: The receiver decodes the MPs one by one.

RP2: The receiver combines the received MPs linearly according to the Maximum Ratio Combining (MRC) principle, and then applies RP1.

RP3: It consists in processing sequentially the incoming MPs: decode MP(1), then concatenate MP(2) with MP(1) and decode the resulting packet, then concatenate MP(3) with MP(2) and MP(1), and so on up to the concatenation with MP($n_{\rm IR}$). If the FR is still in error after receiving MP($n_{\rm IR}$), the memory is flushed and the process starts again.

RP4: After processing the $n_{\rm IR}$ first received MPs as in RP3, the incoming MP$_{k,n}(m)$, $m > n_{\rm IR}$, is linearly combined with the packets MP$_{k,n}(m')$ according to the MRC principle, where $m' = m \mod n_{\rm IR}$.

Notice that the standard ARQ corresponds to the scheme RM1+RP1. The type-I HARQ is the scheme RM2+RP1. The Chase-Combining (CC) HARQ developed in [15] is obtained as (RM1 or RM2)+RP2 and will be noted CC-HARQ. The Incremental Redundancy (IR) HARQ with flush memory is described by RM3+RP3 and will be noted IR-HARQ. The incremental redundancy HARQ associated with Chase Combining is described by RM3+RP4 and will be noted IC-HARQ.

C. Cross-layer Strategies

The conventional retransmission schemes are usually applied at the MAC layer, which according to the layer model depicted in Fig. 1 correspond to the fragment level after the SAR. For truncated schemes, once the $P_{\max}$th transmission fails, the fragment is dropped and the retransmission process is started again with the next fragment. Recently, based on the fact that if one fragment is missing at the receiver side the corresponding IP packet at the Network layer is dropped, the authors in [4] proposed to enhance the ARQ scheme by granting a global transmission credit, noted C, to the set of fragments belonging to the same IP packet before the SAR. Thus, rather than allowing each of the N fragments to be transmitted $P_{\max}$ times, the new scheme allocates C transmissions to the set of the N fragments. Results in [4] show that this cross-layer strategy outperforms the conventional one. In the later we will refer the conventional strategy to as Fragment-Based Strategy (FBS) and the cross-layer one to as IP-Based Strategy (IBS). In [4] the strategy was applied to the ARQ scheme and in this paper, we apply it to the HARQ (RP2, RP3, RP4).

D. Performance Metric Definitions

The metrics considered in this paper are: packet error rate, efficiency, and packet delay. For the FBS, all the metrics can be considered at the FR level and at the IP level, whereas for the IBS, only the IP level makes sense. For the notation, we will put the subscript 'F' for the FR level and 'I' for the IP level. We also put the superscript 'F' for the standard strategy FBS and 'I' for the cross-layer strategy IBS. Thus, to one metric "X", we will consider three different values: $\mathrm{X}_\mathrm{F}^\mathrm{F}$ (eventually denoted by X_F since $\mathrm{X}_\mathrm{F}^\mathrm{I}$ does not make sense) , $\mathrm{X}_\mathrm{I}^\mathrm{F}$, and $\mathrm{X}_\mathrm{I}^\mathrm{I}$.

The **packet error rate**, noted Π, is defined as the probability that a packet fails to be transmitted. The **efficiency**, noted η, is defined as the average number of correctly received bits per transmitted bit. The packet **delay**, noted n, is defined as the average number of MP needed to transmit one packet correctly from the transmitter to the receiver.

In previous definitions, if the FR level is considered, the mentioned packet is the FR whereas if the IP level is considered, the mentioned packet is the IPP.

III. Metric Analytical Expression Derivations

We first derive the efficiency that is the most challenging issue concerning the performance metrics derivations. We then focus on the derivations of the delay and the PER. Our main novelty consists in obtaining general closed-form expressions valid for the FBS and IBS strategies at IP level whatever the retransmission mechanism schemes. Afterward, we inspect some particular interesting cases such as Packet with Equal Length, or ARQ schemes. Finally, we treat the way to evaluate numerically these new closed-form expressions.

Before going further, let us introduce some notations. The term $p_M^{\mathrm{x}}(k)$ denotes the probability that M successive fragments are successfully received in exactly k MP transmissions for the xBS strategy. When $k < M$, $p_M^{\mathrm{x}}(k) = 0$. When $M = 1$, $p_M^{\mathrm{x}}(k)$ is independent of the strategy, and we will just write it $p_1(k)$.

A. Efficiency

1) Efficiency General Expression: A very general mathematical expression for the efficiency can be found by using the renewal theory [16]. The renewal theory was proposed to derive throughput expressions in [17], [18] for the ARQ scheme and [8] for the IR-HARQ scheme relying on capacity achieving codes. This approach remains valid for deriving the efficiency for any HARQ scheme at any level with any cross-layer strategy. Since the efficiency as no dimension (compared to throughput that is expressed in bits per second or bits per second per Hertz), the time index t in [16] as to be taken as the number of transmitted bits, and is noted k here in order to avoid confusion. Let R denote the random variable modeling the reward process and R_i its realization corresponding to the reward in bits for the ith successful packet reception. Let $N(k)$ be the number of decoded packets up to time k. Then, the average number of correct received bits per transmitted bit, *i.e.* the efficiency, is given by $\eta = \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{N(k)} R_i$. Let B denote the random variable modeling the number of transmitted bits between two rewards and B_i its realization corresponding to the number of bits transmitted between the events $(i-1)$ and i. According to the renewal reward theorem [16], we have

$$\eta = \frac{\mathbb{E}[R]}{\mathbb{E}[B]}. \tag{1}$$

In our case, the reward amount R_i is constant and is equal to the received packet size L. Notice that, in the rest of this paragraph, the "packet" refers to as i) the fragment (of length $L = L_{\mathrm{FR}}$) if FR level is studied or ii) the IP packet (of length $L = L_{\mathrm{IPP}}$) if IP level is studied. We thus have $\eta = L/\tau$, where $\tau := \mathbb{E}[B]$ is equal to the average number of transmitted bits between two successive received packets ($:=$ meaning *by definition*). If $\hat{n}$ denotes the average number of transmitted bits when the transmission is successful, and $\check{n}$ the average number of transmitted bits when the transmission fails, then when the packet is transmitted after k attempts (with failure), the average number of transmitted bits is equal to $\hat{n} + k\check{n}$ with probability $\Pi^k(1-\Pi)$ and we can write $\tau = \sum_{k=0}^{\infty}(\hat{n} + k\check{n})\Pi^k(1-\Pi)$, where Π is the PER of the considered packet. After some basic computation we get $\tau = \hat{n} + \check{n}\Pi/(1-\Pi)$, which finally leads to

$$\eta = \frac{L(1-\Pi)}{\hat{n}(1-\Pi) + \check{n}\Pi}. \tag{2}$$

For the sake of clarity, we have omitted the superscript and subscript for the variables in (2) since it is a very general expression that is valid at any layer and for any cross-layer strategy. The efficiency is thus characterized by three quantities $\check{n}$, $\hat{n}$, and Π that should be computed according to the considered layer and cross-layer strategy.

We now derive the expressions for $\hat{n}$ and $\check{n}$ at the FR level, then at the IP level for the FBS and the IBS.

2) Derivations for $\hat{n}_{\mathrm{F}}$: we introduce $w_k := \sum_{i=1}^{k} \delta_{(i \mod n_{\mathrm{IR}})}$ which gives the total number of bits transmitted at the kth transmission. The average number of transmitted bits per successful transmitted FR is then given by $\hat{n}_{\mathrm{F}} = \sum_{k=1}^{P_{\max}} w_k \Pr\{\text{FR received in } k \text{ transmissions} \mid \text{FR OK}\}$. Thanks to the Bayes's rule, we have

$$\hat{n}_{\mathrm{F}} = \frac{1}{1-\Pi_{\mathrm{F}}} \sum_{k=1}^{P_{\max}} w_k p_1(k). \tag{3}$$

3) Derivations for $\check{n}_F$: the average number of transmitted bits when the FR fails to be received is obviously equal to

$$\check{n}_{\mathrm{F}} = \sum_{i=1}^{P_{\max}} \delta_{(i \mod n_{\mathrm{IR}})} = w_{P_{\max}}. \tag{4}$$

Let us now move at the IP level which is the main purpose of this paper.

4) Derivations for $\hat{n}_I^F$: for the FBS, as the fragments are independent and identically distributed, we can easily show that the average number of transmitted bits is equal to N times the average number of transmitted bits per FR,

$$\hat{n}_{\mathrm{I}}^{\mathrm{F}} = N\hat{n}_{\mathrm{F}}. \tag{5}$$

5) Derivations for $\check{n}_I^F$: after very tedious derivations reported in Appendix A, we obtain the following simple expression

$$\check{n}_{\mathrm{I}}^{\mathrm{F}} = N\frac{\Pi_{\mathrm{F}}}{\Pi_{\mathrm{I}}^{\mathrm{F}}}(\check{n}_{\mathrm{F}} - \hat{n}_{\mathrm{F}}) + N\hat{n}_{\mathrm{F}}. \tag{6}$$

6) Derivations for $\hat{n}_I^I$: for the IBS, we have to list all the events corresponding to a successful IPP packet transmission. Each of these events is composed of N successful FR transmissions as long as the total number of MP retransmissions is less than C and can be described by the set $\hat{E}_{\mathrm{I}}^{\mathrm{I}}(\underline{i}) := \{\ F_1(i_1) \text{ and } F_2(i_2) \text{ and } \cdots\ F_N(i_N) \mid \underline{i} \in \mathcal{R}\}$ with $F_j(\ell) = \{\text{FR}(j) \text{ is received in } \ell \text{ MPs}\}$ and $\mathcal{R} = \{\underline{i} \in \mathbb{N}_*^N | \sum_{k=1}^{N} i_k \leq C\}$. Then the average number of transmitted bits in this case is given by $\hat{n}_{\mathrm{I}}^{\mathrm{I}} = \sum_{\underline{i}\in\mathcal{R}} e(\underline{i}) \Pr\{\hat{E}_{\mathrm{I}}^{\mathrm{I}}(\underline{i}) | \text{ IP packet OK}\}$ where $e(\underline{i})$ is the number of bits associated with the event $\hat{E}_{\mathrm{I}}^{\mathrm{I}}(\underline{i})$. One can easily deduce that $e(\underline{i})$ is equal to the total number of bits transmitted for the N fragments, and is equal to $e(\underline{i}) = \sum_{k=1}^{N} w_{i_k}$. Since events $F_j(k)$ are independent, we have $\Pr\{\hat{E}(\underline{i}) | \text{ IP packet KO}\} = (1-\Pi_{\mathrm{I}}^{\mathrm{I}})^{-1} \prod_{k=1}^{N} \Pr\{F_k(i_k)\}$. Since $\Pr\{F_i(k)\} = p_1(k)$, $\hat{n}_{\mathrm{I}}^{\mathrm{I}}$ takes the following form

$$\hat{n}_{\mathrm{I}}^{\mathrm{I}} = \frac{1}{1-\Pi_{\mathrm{I}}^{\mathrm{I}}} \sum_{\underline{i}\in\mathcal{R}} \Big[\prod_{k=1}^{N} p_1(i_k) \sum_{k=1}^{N} w_{i_k}\Big]. \tag{7}$$

7) Derivations for $\check{n}_I^I$: some non-trivial algebraic manipulations given in Appendix B lead to

$$\check{n}_{\mathrm{I}}^{\mathrm{I}} = \frac{1}{\Pi_{\mathrm{I}}^{\mathrm{I}}}\left[q(C-1)w_C + \sum_{j=2}^{N-1}\sum_{\underline{i}\in\mathcal{T}_j}\prod_{k=1}^{j-1} p_1(i_k)q(C-m_{\underline{i}}(j-1)-1)(r_{\underline{i}}(j-1)+w_{C-m_{\underline{i}}(j-1)}) \right.$$
$$\left. + \sum_{\underline{i}\in\mathcal{T}_N}\prod_{k=1}^{N-1} p_1(i_k)q(C-m_{\underline{i}}(N-1))(r_{\underline{i}}(N-1)+w_{C-m_{\underline{i}}(N-1)})\right], \tag{8}$$

where $m_{\underline{i}}(j) := \sum_{k=1}^{j} i_k$ is the number of MPs, $r_{\underline{i}}(j) := \sum_{k=1}^{j} w_{i_k}$ is the number of transmitted bits for transmitting the j first FRs, $\mathcal{T}_j := \{\underline{i} \in \mathbb{N}_*^{j-1} | \sum_{k=1}^{j-1} i_k < C\}$, and

$$q(i) := \begin{cases} 1 & \text{for } i = 0, \\ \prod_{k=0}^{i-1} \pi_k & \text{for } i > 0. \end{cases} \tag{9}$$

which correspond to the probability that one FR is still not correctly received (KO) after i transmissions.

Unlike $\check{n}_{\mathrm{I}}^{\mathrm{F}}$ given by (6), we have not found way to get a more compact expression for $\check{n}_{\mathrm{I}}^{\mathrm{I}}$.

8) Final expressions for η: We give here a summary of the new closed-form expressions for the efficiency proved in this paper. We only report η_{F} and $\eta_{\mathrm{I}}^{\mathrm{F}}$ since $\eta_{\mathrm{I}}^{\mathrm{I}}$ does not take a reasonable compact form even if $\eta_{\mathrm{I}}^{\mathrm{I}}$ will be numerically evaluated quite easily further (see Section IV).

When putting (3) and (4) into (2) and using (15), we get

$$\eta_{\mathrm{F}} = \frac{L_{\mathrm{FR}}\sum_{k=1}^{P_{\max}} p_1(k)}{(1-\sum_{k=1}^{P_{\max}} p_1(k))w_{P_{\max}} + \sum_{k=1}^{P_{\max}} w_k p_1(k)}. \tag{10}$$

When putting (5) and (6) into (2) and using (10) and (15), we obtain

$$\eta_{\mathrm{I}}^{\mathrm{F}} = \eta_{\mathrm{F}}(1-\Pi_{\mathrm{F}})^{N-1} \tag{11}$$

which shows that $\eta_{\mathrm{I}}^{\mathrm{F}}$ can be interestingly simply deduced from η_{F} by a multiplicative coefficient which depends on the PER at the FR level. We can also prove that since $(1-\Pi_{\mathrm{F}}) \leq 1$, we have $\eta_{\mathrm{I}}^{\mathrm{F}} \leq \eta_{\mathrm{F}}$ which represents the price to pay to transmit N FRs instead of a single one.

B. Delay

At the FR level, the average number of transmitted MPs per successful FR transmission for the FBS is given by $n_{\mathrm{F}} = \sum_{k=1}^{P_{\max}} k \Pr\{\text{FR received in } k \text{ MPs} \mid \text{FR received}\}$ which leads to

$$n_{\mathrm{F}} = \frac{1}{1-\Pi_{\mathrm{F}}}\sum_{k=1}^{P_{\max}} k p_1(k). \tag{12}$$

This previous equation has actually already been found in [12] for any HARQ scheme by using the framework of Markov chain instead of a combinational approach as done here which leads to a strongly different but equivalent shape for (12).

At the IP level for the FBS, the instantaneous delay for an IP packet is the sum of the N FR transmission delays that are independent and identically distributed. Basic computation leads to

$$n_{\mathrm{I}}^{\mathrm{F}} = N n_{\mathrm{F}}. \tag{13}$$

In [4], the delay for the FBS at the IP level has been introduced with a much more complicated form and only for ARQ scheme. Equation (13) is thus new for any HARQ scheme and is simple since $p_1(k)$ only is involved. For the IBS, the reasoning is the same as for n_{F} but at the IP layer. We then get

$$n_{\mathrm{I}}^{\mathrm{I}} = \frac{1}{1-\Pi_{\mathrm{I}}^{\mathrm{I}}} \sum_{k=N}^{C} k p_N^{\mathrm{I}}(k). \tag{14}$$

Equation (14) boils down to expression given in [4] for the ARQ scheme but is clearly now justified for any HARQ scheme.

C. Packet Error Rate

The PER is defined as $\Pi_{\ell}^{\mathrm{x}} = 1 - \Pr\{\text{packet received}\}$ where the packet is either the MP (if $\ell = \mathrm{F}$) or the IPP (if $\ell = \mathrm{I}$) . According to the definition of $p_M^{\mathrm{x}}(k)$, we are able to give this general expression of the PER

$$\Pi_{\ell}^{\mathrm{x}} = 1 - \sum_{k=K_{\min}}^{K_{\max}} p_M^{\mathrm{x}}(k), \tag{15}$$

where we have

- $K_{\min} = 1, K_{\max} = P_{\max}, M = 1$ for Π_{F}. For Π_{F}, this expression is already well-known and different approaches for the proof can be found in [8], [12] for several HARQ schemes.
- $K_{\min} = N, K_{\max} = NP_{\max}, M = N$ for $\Pi_{\mathrm{I}}^{\mathrm{F}}$. Concerning ARQ scheme, it is proven in [4] that $\Pi_{\mathrm{I}}^{\mathrm{F}} = 1 - (1-\pi_0)^N$ where π_0 denotes the FR error rate. Once again only for ARQ scheme, (15) has been given in [19] but not explicitly since we have to transform a Gilbert-Elliot channel into a single state based channel to obtain it. Inasmuch as the FR transmissions are independent, $\Pi_{\mathrm{I}}^{\mathrm{F}}$ can be straightforwardly simplified as follows: $\Pi_{\mathrm{I}}^{\mathrm{F}} = 1 - (1-\Pi_{\mathrm{F}})^N$.
- $K_{\min} = N, K_{\max} = C, M = N$ for $\Pi_{\mathrm{I}}^{\mathrm{I}}$. In [4], $\Pi_{\mathrm{I}}^{\mathrm{I}}$ has been provided with this form but only proven for the ARQ scheme. Therefore, our extension to any HARQ scheme is new.

The proposed framework enables us to obtain the following theorem proven in Appendix C.

Theorem 1 *For any retransmission scheme and for $C \geq NP_{\max}$, $\Pi_{\mathrm{I}}^{\mathrm{I}} < \Pi_{\mathrm{I}}^{\mathrm{F}}$.*

So, when $C \geq NP_{\max}$, it is always beneficial from a packet error rate point of view to use the cross-layer scheme IBS rather than the FBS one at the IP level.

D. Some particular cases

1) MPs with equal length: The efficiency expressions obtained in Subsection III-A can be greatly simplified when the MAC packets have the same length (equal to L_{MAC}). In such a case, we have $\delta_k = L_{\mathrm{MAC}}$ for each k, which implies $w_k = kL_{\mathrm{MAC}}$. As a consequence, by putting $\rho = L_{\mathrm{FR}}/L_{\mathrm{MAC}}$, we obtain

$$\eta_{\mathrm{F}} = \frac{\rho \sum_{k=1}^{P_{\max}} p_1(k)}{(1 - \sum_{k=1}^{P_{\max}} p_1(k))P_{\max} + \sum_{k=1}^{P_{\max}} kp_1(k)}, \tag{16}$$

and

$$\eta_{\mathrm{I}}^{\mathrm{F}} = \frac{\rho \left(\sum_{k=1}^{P_{\max}} p_1(k)\right)^N}{(1 - \sum_{k=1}^{P_{\max}} p_1(k))P_{\max} + \sum_{k=1}^{P_{\max}} kp_1(k)}. \tag{17}$$

Whereas (16) is already known and can be found in [8], [11], (17) is new.

In Appendix A, the term $r_{\underline{i}}(n) = \sum_{k=1}^{n} w_{i_k}$ can be replaced with $L_{\mathrm{MAC}} \sum_{k=1}^{n} i_k$, which leads after some tedious calculations to the following more simple expression for $\eta_{\mathrm{I}}^{\mathrm{I}}$

$$\eta_{\mathrm{I}}^{\mathrm{I}} = \frac{N\rho \sum_{k=N}^{C} p_N^{\mathrm{I}}(k)}{C(1 - \sum_{k=N}^{C} p_N^{\mathrm{I}}(k)) + \sum_{k=N}^{C} kp_N^{\mathrm{I}}(k)}. \tag{18}$$

This previous expression is new. It seems to be a trivial extension of (16) by replacing $p_1(k)$ with $p_N^{\mathrm{I}}(k)$ and $P_{\max}$ by C. Nevertheless, our derivations show that the extension was not straightforward and needed hard works.

2) ARQ scheme: We are obviously in the context of MP with equal length since the same MP is sent per FR. Then, the term $p_1(k)$ can be easily written as $(1 - \pi_0){\pi_0}^{k-1}$, and $p_N^{\mathrm{I}}(k) = \binom{k-1}{N-1}(1 - \pi_0)^N \pi_0^{k-N}$ as proved in [4]. We are able to prove the following results[1]:

$$\Pi_{\mathrm{I}}^{\mathrm{F}} = 1 - (1 - \pi_0)^N, \; \Pi_{\mathrm{I}}^{\mathrm{I}} = I_{\pi_0}(C - N + 1, N) \tag{19}$$

[1]Note that some of the results presented here have been obtained after tedious calculations, and proofs are omitted due the page limitation.

$$\eta_{\mathrm{I}}^{\mathrm{F}} = \rho(1-\pi_0)(1-\pi_0^{P_{\max}})^{N-1} \tag{20}$$

$$\eta_{\mathrm{I}}^{\mathrm{I}} = \frac{N\rho(1-\pi_0)\alpha(\pi_0)}{C(1-\pi_0)B_{\pi_0}(C-N+1,N)+N\alpha(\pi_0)-(1-\pi_0)^N\pi_0^{C-N+1}} \tag{21}$$

$$n_{\mathrm{I}}^{\mathrm{F}} = N\left(P_{\max}+\frac{1}{1-\pi_0}+\frac{P_{\max}}{\pi_0^{P_{\max}}-1}\right),\ n_{\mathrm{I}}^{\mathrm{I}} = \frac{N}{1-\pi_0}-\frac{(1-\pi_0)^{N-1}\pi_0^{C-N+1}}{\alpha(\pi_0)} \tag{22}$$

where $B_x(a,b)$ is the incomplete Beta function as defined in [20, Eq. (8.391)], $I_x(a,b)$ is the normalized incomplete Beta function as defined in [20, Eq. (8.392)], and $\alpha(\pi_0) = B(C-N+1,N)B_{1-\pi_0}(N,C-N+1)/B(N,C-N+1)$ with $B(a,b) = B_1(a,b)$ the complete Beta function. One can note that our previous expressions at the IP level for FBS and IBS are new, except for $\Pi_{\mathrm{I}}^{\mathrm{F}}$. To obtain the expressions for the MAC level, we just have to put $N = 1$. At the MAC level, $\Pi_{\mathrm{F}} = \pi_0$ and $\eta_{\mathrm{F}} = \rho(1-\pi_0)$ which are well-known results. In contrast, the expression of the delay is also new. Thanks to these expressions at the IP level, one can prove that, when $\pi_0 \to 1$ (corresponding to low SNR case), the limit for the delay is counter-intuitive since it is equal to $N(P_{\max}+1)/2$ for FBS and $N(C+1)/(N+1)$ for IBS and thus are different from the maximum delay.

IV. Performance Metrics computation

We can show that the $p_N^{\mathrm{I}}(k)$ can be expressed as a function of the $p_1(k)$, and as a consequence, all the performance metrics expressions established in the previous section can be computed from the $p_1(k)$. In the following we first derive the link between $p_N^{\mathrm{I}}(k)$ and the $p_1(k)$, and then provide insights about the evaluation of the $p_1(k)$.

A. *Computation of* $p_N^{\mathrm{I}}(k)$

The computation of the $p_N^{\mathrm{I}}(k)$ can be done by using (33) given in Appendix C. However, such a way requires a huge amount of calculations since we have to explicit all the elements of $\mathcal{Q}_{N,k}^{\mathrm{I}}$. As soon as k or N become large, the cardinality of $\mathcal{Q}_{N,k}^{\mathrm{I}}$ is large and exhibiting all the elements of $\mathcal{Q}_{N,k}^{\mathrm{I}}$ becomes an untractable problem. Here, we show that the $p_N^{\mathrm{I}}(k)$ can be computed recursively from the $p_1(k)$. By noticing that the probability that N MPs are received in k transmissions is equal to the probability that $(N-1)$ MPs are received in $(N-1)$ transmissions along with one MP received in $(k-N+1)$ transmissions, (*i.e.* $p_1(k-N+1)p_{N-1}^{\mathrm{I}}(N-1)$) or, that $(N-1)$ MPs are received in N transmissions along with one MP received in $(k-N)$

transmissions, (*i.e.* $p_1(k-N)p^{\mathrm{I}}_{N-1}(N)$), or so on until that $(N-1)$ MPs are received in $(k-1)$ transmissions along with one MP received in one transmission, (*i.e.* $p_1(1)p^{\mathrm{I}}_{N-1}(k-1)$), we can express $p^{\mathrm{I}}_N(k)$ as follows

$$p^{\mathrm{I}}_N(k) = \sum_{i=1}^{k-N+1} p_1(i)p^{\mathrm{I}}_{N-1}(k-i). \tag{23}$$

This algorithm has a polynomial complexity in *multiply and add* operations since the cardinality of these operations is equal to $(n-1)(k-n+1)(k-n+2)/2$.

B. Approximate Computation of $p_1(k)$

Except for very seldom cases, closed-form expressions for $p_1(k)$ can not be derived and we need to evaluate it numerically through tight approximations. To treat this problem, we split it into two categories: i) Type-I HARQ (*i.e.* RP1) and ii) Type-II HARQ (*i.e.* RP2, RP3, RP4).

1) Type-I HARQ: when the channel fades or noises each MP in an uncorrelated manner, we simply have $p_1(k) = (1-\pi_0){\pi_0}^{k-1}$ where we remind that π_0 is the FR/MP error rate. For non-coded schemes (ARQ, *i.e.* RM1-RP1), we have $\pi_0 = 1-(1-P_e)^{L_{\mathrm{FR}}}$ where P_e is the bit error rate. P_e can then be computed exactly, and its closed-form expression will depend on the nature of the channel (*e.g.* Gaussian/Rayleigh/Erasure), of the constellations (*e.g.* PAM/PSK/QAM) and can be found in [21]. For coded schemes (*i.e.* RM2-RP1), one can only found approximations for π_0 in the literature. Although an exhaustive review of this type of literature is out of the scope of our paper, we hereafter list some references dedicated to some FEC schemes. For convolutional codes one can use the union bound [14]. For turbo-codes, methods have been proposed to estimate the floor error probability [22] and the waterfall error probability [23] using the extrinsic information transfer (EXIT) chart technique. For LDPC codes, one can use the recent results in [24] applied to binary erasure channel.

2) Type-II HARQ: the term $p_1(k)$ cannot be longer expressed as a function of π_0 since the FR error rate depends now on the number of associated MPs. To overcome this problem, we rather write $p_1(k) = q(k-1)-q(k)$ as done in [8] and [18], where $q(k)$ is defined in (9). We now focus on the computation of $q(k)$ for which we resort to approximations. As done in [25], we consider that $q(k) \approx \nu_k$ where $\nu_i = \Pr\{K_i\}$ with the event $K_i = \{\text{FR is KO after } i \text{ transmissions}\}$. As $q(k)$ can equivalently be written as $q(k) = \Pr\left\{\cap_{i=1}^{k} K_i\right\}$, the approximation works as if $\nu_i \approx 1$ for $i < k$. Even if this approximation seems to be rough, we will see that it is actually tight enough. From now, our problem boils down to compute ν_k. The computation of ν_k can be done very similarly to the computation of π_0 in the type-I HARQ case. For instance, let us focus on the Gaussian channel for the sake of simplicity. When CC-HARQ (RM1/RM2-RP2) is employed, we have $\nu_k = Q\left(\sqrt{2k\gamma_1\gamma_2\mathrm{SNR}}\right)$ where γ_1 is a constant that depends on the constellation, γ_2 is the "coding gain" associated with the FEC scheme, and $Q(.)$ is the Gaussian tail function [21]. Other closed-form expressions for ν_k can be found in [21] for Rayleigh and multipath fading channels. For IR-HARQ (RM3-RP3), the fact that $q(k) \approx \nu_k$ is valid only when $P_{\max} \leq n_{\mathrm{IR}}$. Then ν_k is the codeword error rate associated with the kth puncturing coding scheme. When $P_{\max} > n_{\mathrm{IR}}$, $q(k)$ may be far away from ν_k as soon as $k > n_{\mathrm{IR}}$ since $q(k)$ should take into account that several MP have already failed whereas ν_k is not able to take into account this phenomenon if memory has been flushed. Nevertheless, it is preferable to combine the multiple packets replicas using IC-HARQ than flushing the memory. For IC-HARQ (RM3-RP4), $q(k) \approx \nu_k$ is *a priori* valid but further investigations need to be done to compute the value of ν_k. Note that in [8], similar approach has been used for evaluating $p_1(k)$, *i.e.* they use the fact that $p_1(k) = q(k-1) - q(k)$. But, in [8], $p_1(k)$ is defined as the complementary of the outage probability for IR-HARQ (and so independent of practical coding scheme), and K_i is the outage event of the system after i transmissions, and the approximation becomes an equality which means that $q(k) = \nu_k$.

When the approximation $q(k) \approx \nu_k$ is not accurate enough or/and when ν_k can not be approximated tightly enough, we have to use a semi-analytical approach consisting in estimating $p_1(k)$ by extensive simulation leading to run the HARQ scheme at the MAC level. All the performance metrics can be then computed at the IP layer for the various cross-layer strategies by replacing the estimated values of $p_1(k)$ in the expressions described in our paper. Notice that this approach guaranties that the evaluated performance are equal to the true ones since the analytical expressions do not contain any approximation (within about the $p_1(k)$ error estimation). Compared to a completely empirical evaluation approach, this semi-analytical approach allows nevertheless to save a lot of simulation time since the simulation at the IP level which is not necessary in this semi-analytical approach is much more demanding than at the MAC level. Moreover, one set of estimated $p_1(k)$ allows to compute performance for both FBS and IBS strategies and for various values of parameters, C, $P_{\max}$, N without running the system again.

V. Numerical Illustrations

We inspect the numerical accuracy of the proposed closed-form expressions of the considered performance metrics. We therefore plot three kinds of performance evaluation: i) the *analytical* ones; all the metrics are computed according to their respective closed-form expressions depending on $p_1(k)$. Then $p_1(k)$ is computed via the theoretical approximations suggested in Subsection IV-B. ii) the *semi-analytical* ones; once again, all the metrics are computed according to their respective closed-form expressions depending on $p_1(k)$. But the terms $p_1(k)$ are now estimated empirically by simulating the channel coding. This approach enables us to omit simulating the HARQ mechanism. Moreover the values of $p_1(k)$ can be put in pre-computed tables. iii) the *empirical* ones; the whole HARQ mechanism and the channel coding are simulating including the IP layer and the cross-layer strategy.

We have chosen two HARQ schemes:

- a CC scheme (RM2-RP2) implemented using the Universal Mobile Telecommunications System turbo code of rate $1/2$ [26] with uncoded block length of $K = 1152$,
- an IR scheme (RM3-RP3) implemented with the Rate Compatible Punctured Convolutional (RCPC) codes introduced in [14], with a mother code of rate $R_0 = 1/4$. We have chosen the rates as $\{8/9, 4/5, 4/7, 1/2\}$ and the uncoded block length as $K = 128$ for which the incremental redundancy packet lengths are equal to: $\delta_0 = 144$ $\delta_1 = 16$, $\delta_2 = 64$, $\delta_3 = 32$.

The number of fragments per IP packet is fixed to $N = 3$. The size of the FR is $L_{\text{FR}} = K$. The maximum number of transmissions per FR in the FBS is fixed to $P_{\max} = 4$ for the IR and $P_{\max} = 3$ for the CC. The global credit in the IBS is fixed to $C = NP_{\max}$ and thus equal to $C = 12$ for the IR and $C = 9$ for the CC. The modulation is a Binary Phase Shift Keying (BPSK) for the CC scheme and a Quaternary Phase Shift Keying (QPSK) for the IR one. We have considered both Gaussian and Rayleigh channels with an additive, white, zero-mean Gaussian noise.

In Fig. 2, we compare the estimated values of $q(k)$ and the approximate ones obtained by following the suggestion done in Subsection IV-B for the above-mentioned CC and IR schemes in Gaussian channel. The approximation is thus tight enough. From Fig. 3 to Fig. 8, we respectively plot the efficiency, the delay and the PER for the analytical, semi-analytical, and empirical approaches. As expected, there is a perfect agreement between the empirical and semi-analytical approaches which validates the proposed closed-form expressions. However, we observe slight difference between the analytical and semi-analytical approaches only due to the channel coding evaluation.

VI. Conclusion

In this paper we have derived analytical expressions of performance metrics (efficiency, delay, PER) for a wide range of truncated HARQ retransmission schemes encompassing new cross-layer strategies. We have especially proposed a new general expression for the efficiency even valid when the incremental redundancy packets do not have the same length. These expressions allowed us to speed up notably the metrics computation.

Appendix

A. Details for the calculation of $\check{n}_I^F$

Let us then note the event $G_{\underline{i}}(j) := \{\bar{F}_1$ and $\bar{F}_2$ and $\ldots$ and $\bar{F}_j$ and $F_{j+1}(i_1)$ and $\ldots$ and $F_N(i_{N-j})\ \}$, where $\bar{F}_k$ is the event FR(k) is not received. The event $G_{\underline{i}}(j)$, $1 \leq j \leq N$, corresponds to the case where the first j fragments are not received and the $(N-j)$ remaining ones are successfully transmitted. For a given j, the indices $\{i_1, i_2, \ldots, i_{N-j}\}$ belong to the set $\mathcal{S}_j := \{\underline{i} = [i_1, i_2, \cdots, i_{N-j}] \in \mathbb{N}_*^{N-j} | \forall k, i_k \leq P_{\max}\}$. Noticing that the number of possible events that have j FRs KO is $\binom{N}{j}$. As $\check{n}_{\mathrm{I}}^{\mathrm{F}}$ corresponds to the fact that at least one FR is not received given that the IPP is not received, we can write $\check{n}_{\mathrm{I}}^{\mathrm{F}} = \sum_{j=1}^{N} \binom{N}{j} \sum_{\underline{i} \in \mathcal{S}_j} g_{\underline{i}}(j) \Pr\{G_{\underline{i}}(j) |$ IP packet KO$\}$ where $g_{\underline{i}}(j)$ is the number of transmitted bits associated with the event $G_{\underline{i}}(j)$. This leads to

$$\check{n}_{\mathrm{I}}^{\mathrm{F}} = (\Pi_{\mathrm{I}}^{\mathrm{F}})^{-1} \sum_{j=1}^{N} \binom{N}{j} \sum_{\underline{i} \in \mathcal{S}_j} g_{\underline{i}}(j) \Pr\{G_{\underline{i}}(j)\}. \tag{24}$$

When a FR is not received, the corresponding number of transmitted bits is equal to $w_{P_{\max}}$ bits. Thus we can write $g_{\underline{i}}(j) = j w_{P_{\max}} + r_{\underline{i}}(N-j)$. For $j = N$, all the FRs are in error and thus we have $\Pr\{G_{\underline{i}}(N)\} = (\Pi_{\mathrm{F}})^N$ with $g_{\underline{i}}(N) = N w_{P_{\max}}$. For $j < N$, the probability that j FRs are not received is equal to $(\Pi_{\mathrm{F}})^j$ and the probability that the $(N-j)$ remaining FRs are received is equal to $\prod_{k=1}^{N-j} p_1(i_k)$, which gives $\Pr\{G_{\underline{i}}(j)\} = (\Pi_{\mathrm{F}})^j \cdot \prod_{k=1}^{N-j} p_1(i_k)$ for $1 \leq j < N$, and $\Pr\{G_{\underline{i}}(N)\} = (\Pi_{\mathrm{F}})^N$. Putting the previous expressions into (24) leads to a sum of three terms $\check{n}_{\mathrm{I}}^{\mathrm{F}} = A_1 + A_2 + A_3$ with

$$A_1 = N w_{P_{\max}} \frac{(\Pi_{\mathrm{F}})^N}{\Pi_{\mathrm{I}}^{\mathrm{F}}}, \tag{25}$$

$$A_2 = \frac{w_{P_{\max}}}{\Pi_{\mathrm{I}}^{\mathrm{F}}} \sum_{j=1}^{N-1} \binom{N}{j} j (\Pi_{\mathrm{F}})^j \sum_{\underline{i} \in \mathcal{S}_j} \prod_{k=1}^{N-j} p_1(i_k), \tag{26}$$

$$A_3 = \frac{1}{\Pi_{\mathrm{I}}^{\mathrm{F}}} \sum_{j=1}^{N-1} \binom{N}{j} (\Pi_{\mathrm{F}})^j \sum_{\underline{i} \in \mathcal{S}_j} \left(\prod_{k=1}^{N-j} p_1(i_k) \sum_{k=1}^{N-j} w_{i_k} \right). \tag{27}$$

Due to the lack of space, the proofs of the two following lemmas are omitted.

Lemma 1 $\sum_{\underline{i} \in \mathcal{S}_j} \prod_{k=1}^{N-j} p_1(i_k) = (1 - \Pi_{\mathrm{F}})^{N-j}$.

Lemma 2 $\sum_{\underline{i} \in \mathcal{S}_j} \left(\prod_{k=1}^{N-j} p_1(i_k) \sum_{k=1}^{N-j} w_{i_k} \right) = (N-j)(1-\Pi_{\mathrm{F}})^{N-j-1} \sum_{k=1}^{P_{\max}} w_k p_1(k)$.

From Lemma 1, we can write $A_2 = \frac{w_{P_{\max}}}{\Pi_{\mathrm{I}}^{\mathrm{F}}} \sum_{j=1}^{N-1} \binom{N}{j} j (\Pi_{\mathrm{F}})^j (1-\Pi_{\mathrm{F}})^{N-j}$ which implies that

$$A_2 = N w_{P_{\max}} \frac{\Pi_{\mathrm{F}}}{\Pi_{\mathrm{I}}^{\mathrm{F}}} (1 - (\Pi_{\mathrm{F}})^{N-1}). \tag{28}$$

From Lemma 2, we have $A_3 = \frac{1}{\Pi_{\mathrm{I}}^{\mathrm{F}}} \sum_{j=1}^{N-1} \binom{N}{j} (\Pi_{\mathrm{F}})^j (N-j)(1-\Pi_{\mathrm{F}})^{N-j} \hat{n}_{\mathrm{F}}$, which, after some calculations, leads to

$$A_3 = N \left(1 - \frac{\Pi_{\mathrm{F}}}{\Pi_{\mathrm{I}}^{\mathrm{F}}} \right) \hat{n}_{\mathrm{F}}. \tag{29}$$

From (25) and (28) we have $A_1 + A_2 = N w_{P_{\max}} \Pi_{\mathrm{F}} / \Pi_{\mathrm{I}}^{\mathrm{F}}$. Thus from (4) and (29), we obtain the final expression provided in (6).

B. Details for the calculation of $\check{n}_I^I$

For the IBS, we have to take into account that the N FRs share the same global transmission credit and cannot be treated independently. Thus, the event associated with the failure of the IP packet transmission can be represented by the event $\check{E}_{\mathrm{I}}^{\mathrm{I}} := \bigcup_{j=1}^{N} H(j)$ with $H(1) = \{$FR(1) consumes all the credit$\}$, $H(j) = \{$FR(1) OK, and FR(2) OK, and ... and FR$(j-1)$ OK, and FR(j) consumes the remaining credit$\}$, for $j < 1 < N$, and $H(N) = \{$FR(1) OK, and FR(2) OK, and ... and FR$(N-1)$ OK and FR(N) KO with the remaining credit$\}$. As we will see after, we need to separate these three different cases for $H(j)$: i) $j = 1$, ii) $1 < j < N$, and iii) $j = N$. Before going further, one can remark that $H(1) = \{\bar{F}_1\}$, $H(j) = \bigcup_{\underline{i} \in \mathcal{T}_j} H_{\underline{i}}(j)$, for $1 < j < N$, where $H_{\underline{i}}(j)$, for $1 < j < N$, is the event defined as $H_{\underline{i}}(j) := \{F_1(i_1)$ and $F_2(i_2)$ and $\ldots$

and $F_{j-1}(i_{j-1})$ and $\tilde{F}_j(\underline{i})\}$ where $\tilde{F}_j(\underline{i})$ is the event that FR(j) consumes the remaining credit whenever it is successfully received or not, and $H(N) = \bigcup_{\underline{i}\in\mathcal{T}_N} H_{\underline{i}}(N)$, where $H_{\underline{i}}(N) = \{F_1(i_1)$ and $F_2(i_2)$ and $\ldots$ and $F_{N-1}(i_{N-1})$ and $\bar{F}_N\}$.

We then can write

$$\check{n}_{\mathrm{I}}^{\mathrm{I}} = h(1)\Pr\{H(1)|\text{ IP packet KO}\} + \sum_{j=2}^{N}\sum_{\underline{i}\in\mathcal{T}_j} h_{\underline{i}}(j)\Pr\{H_{\underline{i}}(j)|\text{ IP packet KO}\},$$

which is also equal to $\check{n}_{\mathrm{I}}^{\mathrm{I}} = (\Pi_{\mathrm{I}}^{\mathrm{I}})^{-1}\left(h(1)\Pr\{H(1)\} + \sum_{j=2}^{N}\sum_{\underline{i}\in\mathcal{T}_j} h_{\underline{i}}(j)\Pr\{H_{\underline{i}}(j)\}\right)$, where $h_{\underline{i}}(j)$ (resp. $h(1)$) is the number of transmitted bits associated with the event $H_{\underline{i}}(j)$ (resp. $H(1)$).

We now present the computation of $h(1)$, $\Pr\{H(1)\}$, $h_{\underline{i}}(j)$ and $\Pr\{H_{\underline{i}}(j)\}$ for $1 < j \leq N$.

Case: $j = 1$. The probability $\Pr\{H(1)\} = \Pr\{\bar{F}_1\}$ is actually the probability that the first FR has consumed C transmissions whenever it is successfully received or not which means that the first FR is still KO after $C-1$ transmissions. Therefore, we have

$$\Pr\{H(1)\} = q(C-1), \tag{30}$$

and the corresponding number of transmitted bits $h(1)$ is w_C.

Case: $1 < j < N$. In this case, the event $H_{\underline{i}}(j)$ corresponds to the successful transmission of the first $(j-1)$ FRs, followed by the transmission of FR(j) which consumes the remaining credits whenever it is successfully received or not. The remaining credit left after the $(j-1)$ FRs transmission is equal to $C - m_{\underline{i}}(j-1)$. Thus, following the same reasoning as for $j = 1$, we have $\Pr\{\tilde{F}_j(\underline{i})\} = q(C - m_{\underline{i}}(j-1) - 1)$. Since FR transmissions are independent and remembering that $\Pr\{F_i(k)\} = p_1(k)$, we thus easily deduce that

$$\Pr\{H_{\underline{i}}(j)\} = \sum_{\underline{i}\in\mathcal{T}_j}\prod_{k=1}^{j-1} p_1(i_k) q(C - m_{\underline{i}}(j-1) - 1), \tag{31}$$

and that the corresponding number of transmitted bits $h_{\underline{i}}(j)$ is $r_{\underline{i}}(j-1) + w_{C-m_{\underline{i}}(j-1)}$.

Case: $j = N$. In that case, FR(N) is necessarily KO at the Cth transmission, which gives $\Pr\{\bar{F}_N\} = q(C - m_{\underline{i}}(N-1))$. We then deduce that

$$\Pr\{H(N)\} = \sum_{\underline{i}\in\mathcal{T}_N}\prod_{k=1}^{N-1} p_1(i_k) q(C - m_{\underline{i}}(N-1)), \tag{32}$$

and that the corresponding number of transmitted bits $h_{\underline{i}}(N)$ is $r_{\underline{i}}(N-1) + w_{C-m_{\underline{i}}(N-1)}$. Thus, from (30), (31), (32), and the corresponding number of transmitted bits, $\check{n}_{\mathrm{I}}^{\mathrm{I}}$ takes the form provided in (8).

C. Proof of Theorem 1

Noticing that the different events related to $p_N^{\mathrm{x}}(k)$ are constituted by successive independent successful transmissions of N FRs, we can write

$$p_N^{\mathrm{x}}(k) = \sum_{\underline{i}\in\mathcal{Q}_{N,k}^{\mathrm{x}}} \prod_{m=1}^{N} p_1(i_m) \tag{33}$$

where the sets $\mathcal{Q}_{N,k}^{\mathrm{x}}$ are defined as: $\mathcal{Q}_{N,k}^{\mathrm{F}} = \{\underline{i} = [i_1, i_2, \cdots, i_N] \in \mathbb{N}_*^N \mid \sum_{m=1}^{N} i_m = k, i_k \leq P_{\max}\}$, and $\mathcal{Q}_{N,k}^{\mathrm{I}} = \{\underline{i} = [i_1, i_2, \cdots, i_N] \in \mathbb{N}_*^N \mid \sum_{m=1}^{N} i_m = k\}$. With such set definitions, we can easily establish the following properties:

- **P1:** $\forall k, N \leq k < P_{\max} + N \Rightarrow \mathcal{Q}_{N,k}^{\mathrm{F}} = \mathcal{Q}_{N,k}^{\mathrm{I}}$,
- **P2:** $\forall k, P_{\max} + N \leq k \leq NP_{\max} \Rightarrow \mathcal{Q}_{N,k}^{\mathrm{F}} \subset \mathcal{Q}_{N,k}^{\mathrm{I}}$,
- **P3:** For the FBS, $\forall k, k > NP_{\max}, \Rightarrow p_N^{\mathrm{F}}(k) = 0$.

Let us note $\Delta = \Pi_{\mathrm{I}}^{\mathrm{F}} - \Pi_{\mathrm{I}}^{\mathrm{I}}$. From (33) along with **P1** and **P3**, we can write for $C \geq NP_{\max}$: $\Delta = \sum_{k=P_{\max}+N}^{NP_{\max}} (p_N^{\mathrm{I}}(k) - p_N^{\mathrm{F}}(k)) + \sum_{k=NP_{\max}+1}^{C} p_N^{\mathrm{I}}(k)$. Noting $\bar{\mathcal{Q}}_{k,N} := \mathcal{Q}_{N,k}^{\mathrm{I}} \setminus \{\mathcal{Q}_{N,k}^{\mathrm{I}} \cap \mathcal{Q}_{N,k}^{\mathrm{F}}\}$, then according to **P2**, we get: $\Delta = \sum_{k=P_{\max}+N}^{NP_{\max}} \sum_{\underline{\mathbf{q}}\in\bar{\mathcal{Q}}_{k,N}} p_1(q_1)p_1(q_2)\cdots p_1(q_n) + \sum_{k=NP_{\max}+1}^{C} p_N^{\mathrm{I}}(k)$. Thus, since $\bar{\mathcal{Q}}_{k,N} \neq \emptyset$ according to **P2** and since $p_1(k)$ and $p_N^{\mathrm{I}}(k)$ are strictly positive quantities, we deduce that $\Delta > 0$ which concludes the proof.

References

[1] D. Costello and S. Lin, *Error Control Coding*. New Jersey, USA: Pearson Higher Education, 2nd ed., 2004.

[2] M. Moh, T.-S. Moh, and Y. Shih, "On enhancing WiMAX hybrid ARQ: a multiple-copy approach," in *IEEE Consumer Communications and Networking Conference (CCNC)*, Jan. 2008.

[3] K. C. Beh, A. Doufexi, and S. Armour, "Performance evaluation of hybrid ARQ schemes of 3GPP LTE OFDMA system," in *IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2007.

[4] Y. Choi, S. Choi, and S. Yoon, "MSDU-based ARQ scheme for IP-level performance maximization," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Oct. 2005.

[5] W. Yafeng, Z. Lei, and Y. Dacheng, "Performance analysis of type III HARQ with turbo codes," in *IEEE Vehicular Technology Conference (VTC)*, Apr. 2003.

[6] F. Chiti and R. Fantacci, "A soft combining hybrid ARQ technique applied to throughput maximization within 3G satellite IP networks," *IEEE Trans. on Vehicular Technology*, vol. 56, pp. 594–604, Mar. 2007.

[7] M. W. E. Bahri, H. Boujernaa, and M. Siala, "Performance comparison of type I, II and III hybrid ARQ schemes over AWGN channels," in *IEEE International Conference on Industrial Technology (ICIT)*, Dec. 2004.

[8] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on Low-Density Parity-Check codes," *IEEE Trans. on Communications*, vol. 52, pp. 1311–1321, Aug. 2004.

[9] M. Zorzi and R. Rao, "Performance of ARQ Go-Back-N protocol in Markov channels with unreliable feedback: delay analysis," in *IEEE International Conference on Universal Personal Communications (ICUPC)*, Nov. 1995.

[10] J.-W. Cho, Y. Chang, Y. Kim, and J. Chon, "Analytic Optimization of hybrid ARQ performance in wireless Packet Data Systems," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2006.

[11] C. F. Ieanderson and G. Caire, "The performance of incremental redundancy schemes based on convolutional codes in the block fading gaussian collision channel," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 843–854, May 2004.

[12] L. Badia, M. Levorato, and M. Zorzi, "Markov analysis of Selective Repeat type I HARQ using block codes," *IEEE Trans. on Communications*, vol. 56, pp. 1434–1441, Sept. 2008.

[13] F. Adachi, S. Ito, and K. Ohno, "Performance analysis of a time diversity ARQ in land mobile radio," *IEEE Trans. on Communications*, vol. 37, pp. 177–187, Feb. 1989.

[14] J. Hagenauer, "Rate-Compatible Punctured Convolutional codes (RCPC) and their applications," *IEEE Trans. on Communications*, vol. 36, pp. 389–400, Apr. 1988.

[15] D. Chase, "Code Combining - A maximum likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. on Communications*, vol. 33, pp. 385–393, May 1985.

[16] S. H. Ross, *Stochastic Processes*. New York, USA: Wiley, 2nd ed., 1983.

[17] M. Zorzi and R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," *IEEE Trans. on Communications*, vol. 44, pp. 1077–1081, Sept. 1996.

[18] G. Caire and D. Tuninetti, "The throughput of hybrid ARQ protocols for the gaussian collision channel," *IEEE Trans. on Informations Theory*, vol. 47, pp. 1971–1988, July 2001.

[19] J.-B. Seo, N.-H. Park, H.-W. Lee, and C.-H. Cho, "Impact of an ARQ scheme in the MAC/LLC layer on upper-layer packet transmissions over a markovian channel," in *IEEE Vehicular Technology Conference (VTC)*, May 2006.

[20] I. S. Gradshteyn and M. Ryzhik, *Tables of integrals, series, and products*. Florida, USA: Academic Press, 6th ed., 2000.

[21] M.-S. Alouini and A. J. Goldsmith, "A unified approach for calculating error rates of linearly modulated signals over generalized fading channels," *IEEE Trans. on Communications*, vol. 47, pp. 1324–1334, 1999.

[22] S. Benedetto and G. Montorsi, "Unveiling Turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. on Information Theory*, vol. 42, pp. 409–428, Mar. 1996.

[23] S. ten Brink, "Convergence behaviour of iteratively decoded parallel concatenated codes," *IEEE Trans. on Communications*, vol. 49, pp. 1727–1737, Oct. 2001.

[24] I. Andriyanova and R. Urbanke, "Waterfall region performance of punctured LDPC codes over the BEC," in *IEEE International Symposium on Information Theory (ISIT)*, June 2009.

[25] S. Kallel, "Analysis of a type II Hybrid ARQ scheme with code combining," *IEEE Trans. on Communications*, vol. 38, pp. 1133–1137, Aug. 1990.

[26] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley, 2009.

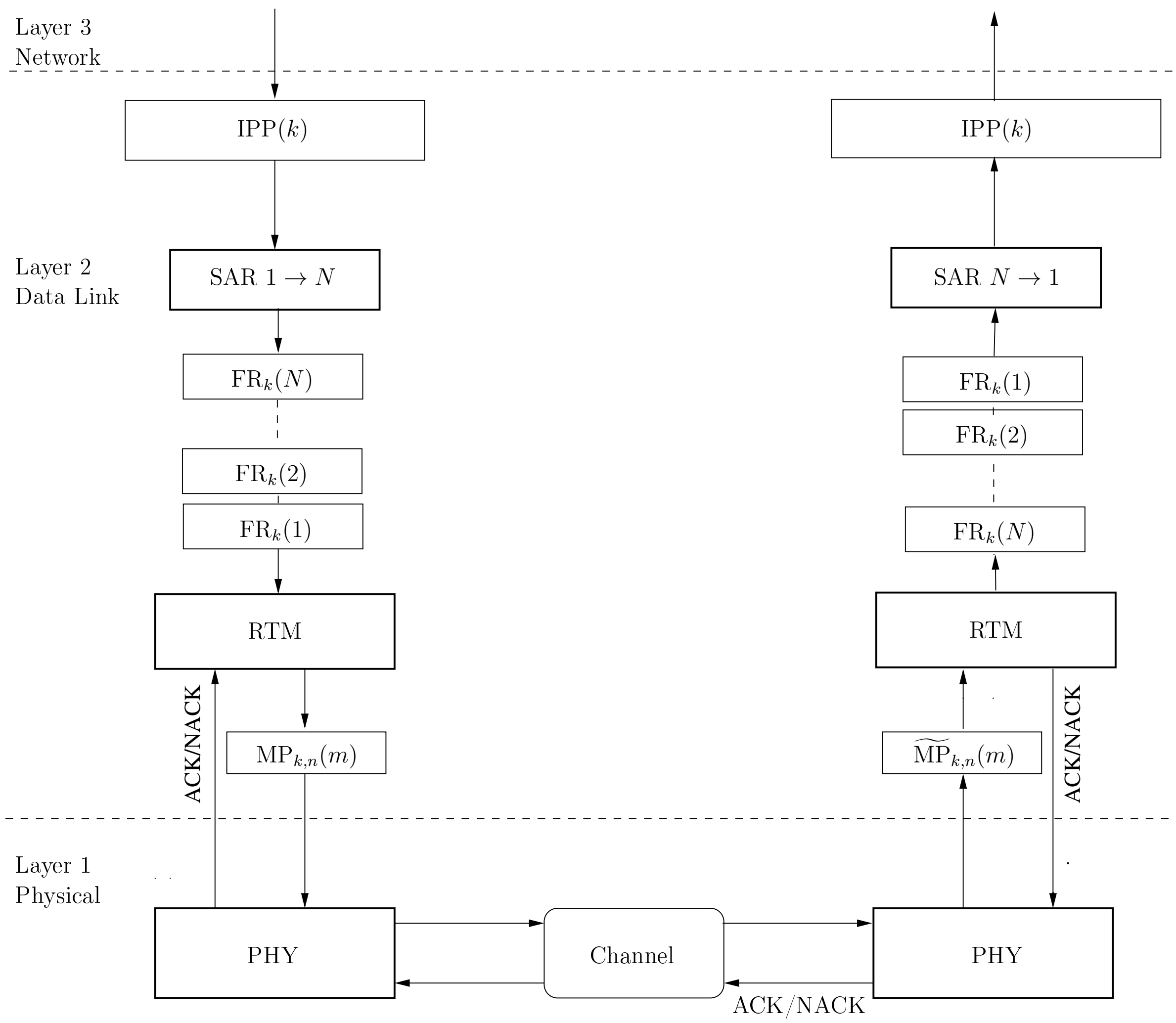

Fig. 1. Layer model.

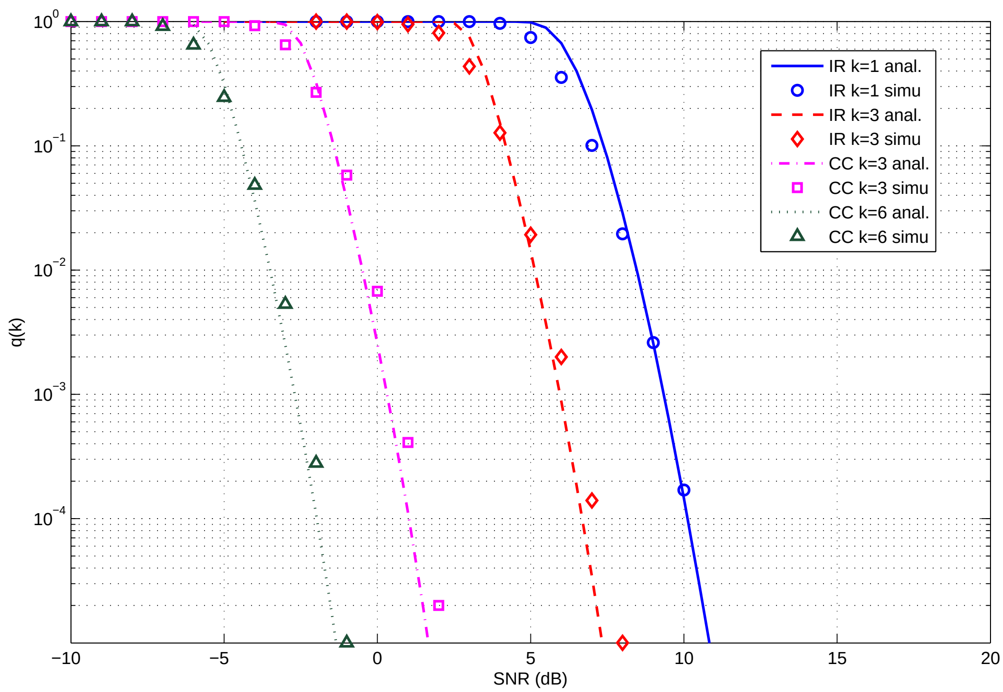

Fig. 2. Approximate and estimated $q(k)$ values for CC-HARQ and IR-HARQ.

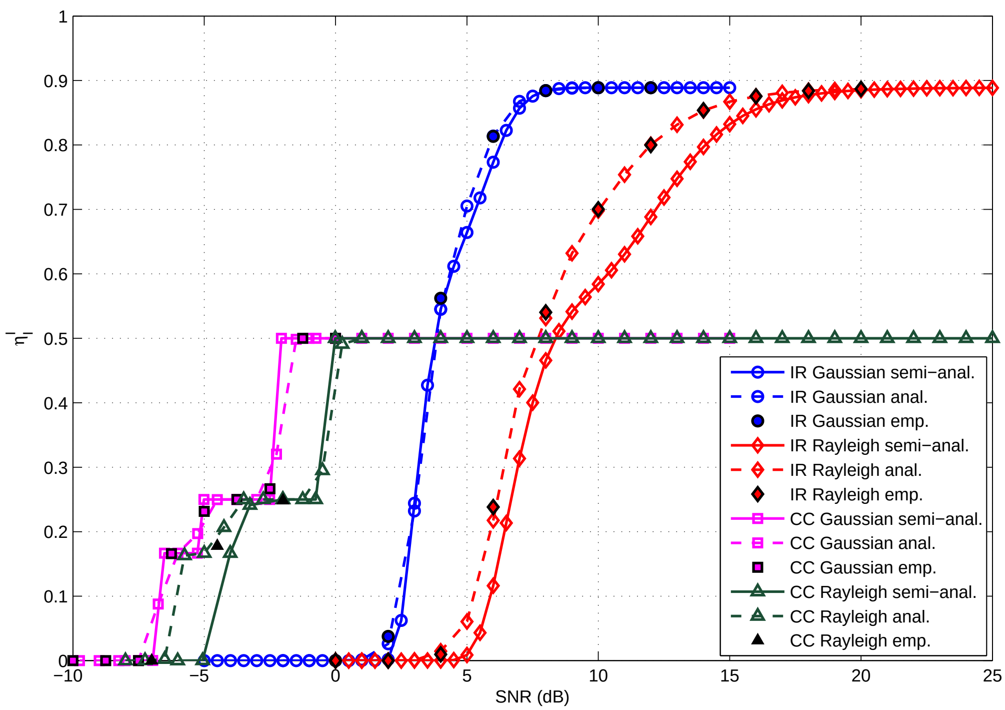

Fig. 3. Efficiency at IP layer for IBS with $N = 3$, for CC-HARQ ($C = 9$) and IR-HARQ ($C = 12$).

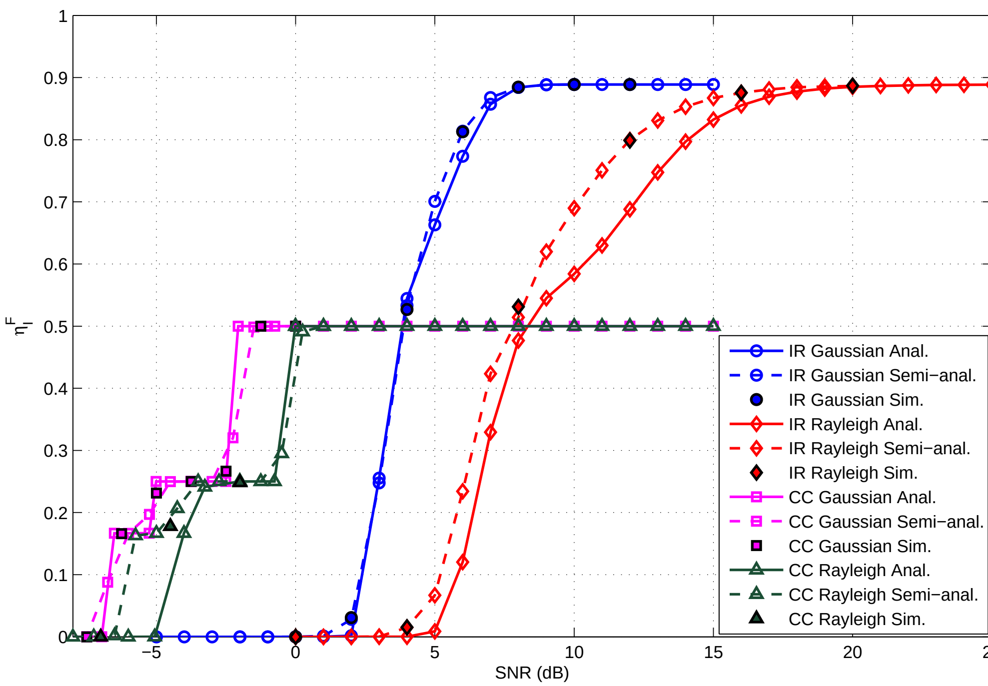

Fig. 4. Efficiency at IP layer for FBS with $N = 3$, for CC-HARQ ($P_{\max} = 3$) and IR-HARQ ($P_{\max} = 4$).

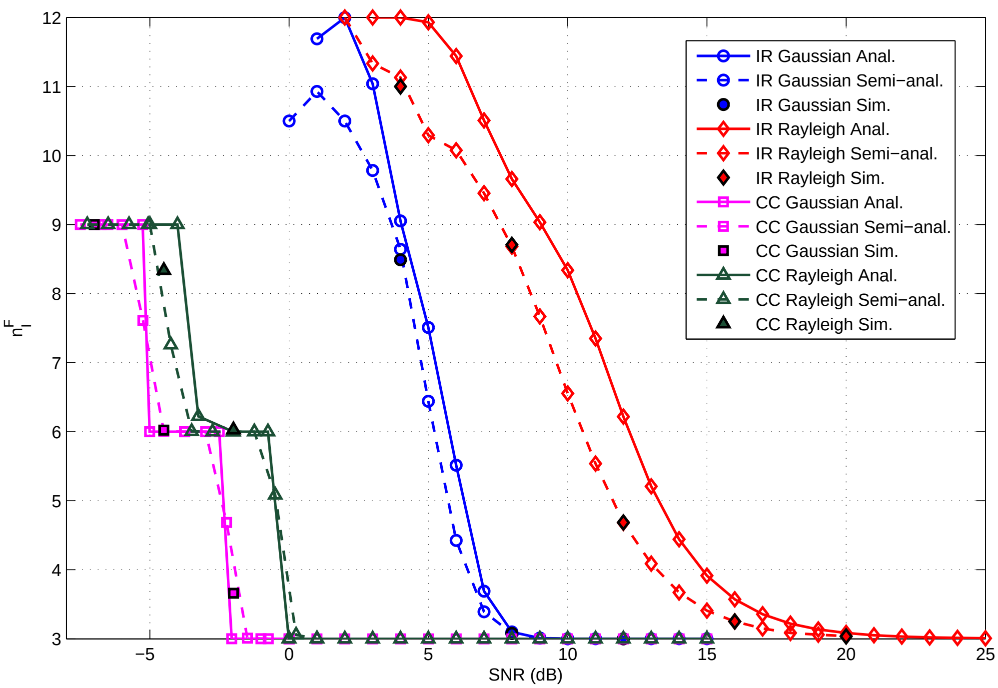

Fig. 5. Delay at IP layer for IBS with $N = 3$, for CC-HARQ ($C = 9$) and IR-HARQ ($C = 12$).

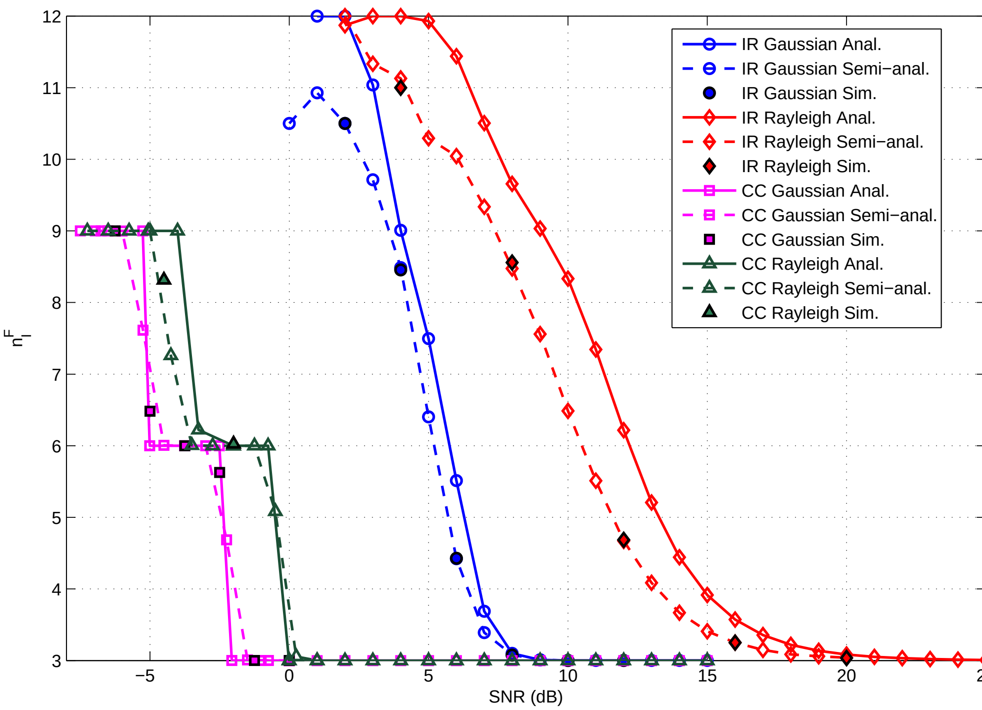

Fig. 6. Delay at IP layer for FBS with $N = 3$, for CC-HARQ ($P_{\max} = 3$) and IR-HARQ ($P_{\max} = 4$).

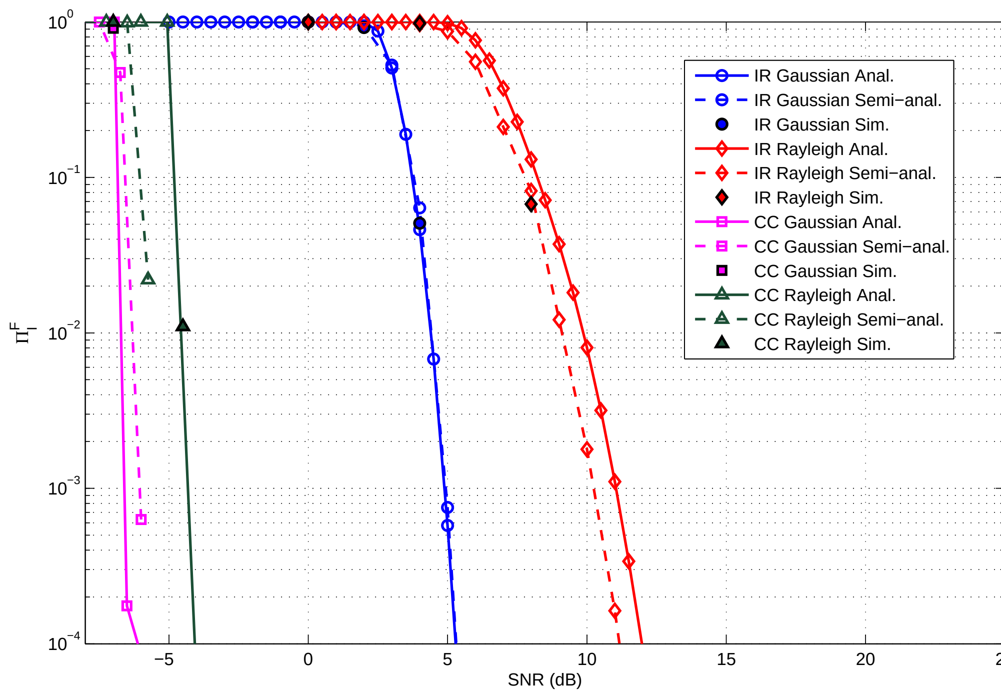

Fig. 7. PER at IP layer for IBS with $N = 3$, for CC-HARQ ($C = 9$) and IR-HARQ ($C = 12$).

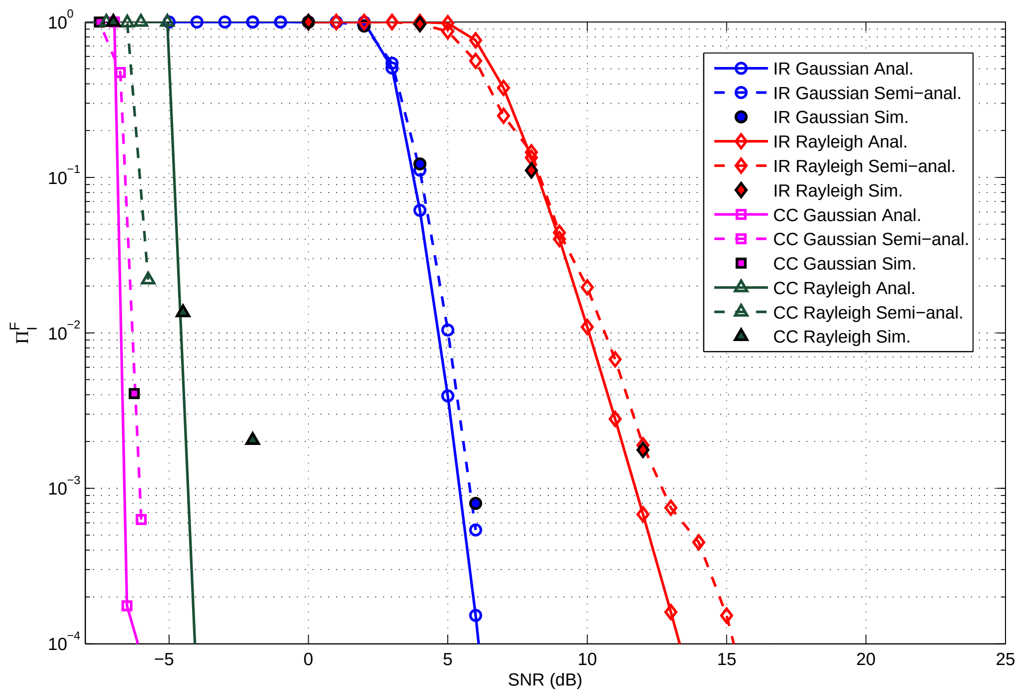

Fig. 8. PER at IP layer for FBS with $N = 3$, for CC-HARQ ($P_{\max} = 3$) and IR-HARQ ($P_{\max} = 4$).