

Procédé de retransmission à redondance incrémentale adapté aux paquets IP fragmentés

A. LE DUC^{1,2}, C. LE MARTRET¹, P. CIBLAT², H. LABIOD²

¹Thales Communications EDS/SPM/WFD, 160 Bd de Valmy, BP 82, F-92704 Colombes Cedex, France - Tel. : +33 (0)146133248 et +33 (0)146132735 - (aude.leduc,christophe.le_martret)[@fr.thalesgroup.com](mailto:fr.thalesgroup.com)

²ENST, 46 rue Barrault, 75013 Paris, France - Tel. : +33 (0)145817728 - (philippe.ciblat,houda.labiod)@enst.fr

Résumé – Le problème traité concerne l’optimisation des performances au niveau de la couche réseau (couche 3 de l’OSI) dans le cadre des transmissions sans fil lorsque l’adaptation à la couche accès radio (couches 1 et 2 de l’OSI) nécessite une fragmentation du paquet de la couche 3. L’originalité du travail réside dans l’intégration et l’adaptation du concept de schéma d’ARQ hybride basé sur la redondance incrémentale à la méthode décrite dans [1] qui prend en compte le fait que les paquets de la couche 3 doivent être fragmentés. Il constitue une approche d’optimisation inter-couches. Le premier résultat important est que les performances obtenues par notre solution sont toujours supérieures aux méthodes conventionnelles de l’état de l’art. De plus, nous déduisons des règles permettant de paramétrer notre solution en fonction du type de performances à améliorer (délai, gigue, ...).

Abstract – The problem treated in this article relates to the optimization of the performances at the network layer level (layer 3 of the OSI model) within the framework of wireless transmissions when the adaptation to the radio access layer (layers 1 and 2 of the OSI model) requires a fragmentation of the layer 3 packet. The originality of the work lies in the integration and the adaptation of the concept of Hybrid ARQ scheme based on the incremental redundancy to the method described in [1] which takes into account the fact that the network layer packets must be fragmented. It constitutes an approach of cross-layer optimization. The first important result is that the performances obtained by our solution are always better than the conventional methods of the state of the art. Moreover, we deduce rules allowing parameterization of our solution according to the type of performances (e.g. delay, jitter) to be improved.

1. Introduction

La qualité des liens radio dans les réseaux sans fil est sujette à des variations importantes dans le temps compte tenu du canal de propagation et des interférences locales. La dégradation du lien physique qui s’en suit a des répercussions importantes au niveau des couches supérieures (par exemple, arrêt de transfert ou déconnexion au niveau TCP dans un réseau IP), il est donc nécessaire de disposer d’accès radio adaptatifs garantissant les meilleures performances offertes par le canal de transmission. Nous nous plaçons dans cette étude dans le cadre d’une modélisation de type TCP/IP comprenant successivement (des couches basses vers les couches hautes) les couches : accès radio, réseau, TCP (transport) et application. La couche accès radio comprend la partie couche physique (PHY) et couche d’accès au medium (MAC).

Cette étude s’inscrit dans le paradigme de l’optimisation inter-couches (cross-layer optimization), dans lequel on optimise un procédé en prenant en compte la connaissance a priori globale ou partielle des différentes couches du système. Nous nous focalisons ici sur les couches accès radio et réseau. Un exemple de cette optimisation est fourni dans [1] lorsque les paquets de la couche réseau doivent être fragmentés pour s’adapter à la taille des paquets autorisée par la couche accès radio (paquets MAC). Les auteurs proposent d’utiliser au niveau de la couche accès radio le fait que les paquets MAC appartiennent à un même paquet provenant de la couche réseau. Se basant sur une couche accès radio de type ARQ (Automatic Repeat

reQuest) conventionnelle et sur une couche réseau de type IP, leur idée est de transférer la persistance (nombre maximal de retransmissions autorisées par paquet MAC) de l’ARQ pour chacun des paquets MAC en un crédit global de transmissions pour l’ensemble des paquets MAC constituant le paquet IP. En effet, dès lors qu’un fragment d’un paquet IP n’a pu être transmis sans erreur, il est inutile de continuer à transmettre les autres fragments du même paquet IP sachant qu’une retransmission sera demandée au niveau TCP. Les auteurs ont montré qu’on améliore fortement les performances en termes de probabilité d’erreur paquet résultante au niveau IP.

Il existe également des techniques d’optimisation inter-couches au niveau accès radio, parmi lesquelles on trouve les ARQ hybrides (HARQ) à redondance incrémentale. Les mécanismes HARQ associent de manière automatique un code correcteur d’erreur à un protocole ARQ, réalisant ainsi le meilleur compromis entre les deux procédés. Les HARQ étant plus performants en termes de throughput (rapport entre le nombre de paquets MAC reçus sans erreurs et le nombre de paquets MAC transmis) que les ARQ, l’objet de notre travail a été de les intégrer et de les adapter à la méthode décrite en [1]. Nous présentons d’abord l’algorithme proposé, puis nous nous intéressons au calcul de ses performances analytiques lorsqu’on l’adapte à l’HARQ. Enfin, nous étudions les performances obtenues par notre solution et discutons de l’apport de l’HARQ par rapport à l’ARQ.

2. Présentation de l'algorithme

Lorsque l'on adopte au niveau de la couche MAC la stratégie de retransmission ARQ et que les paquets de la couche réseau doivent être fragmentés pour s'adapter à la taille des paquets MAC, une même quantité de ressources est allouée pour tous les paquets MAC obtenus après fragmentation des paquets IP. Cette quantité de ressources correspondant au nombre maximal de transmissions autorisées par paquet MAC est égale à $M+1$, où M est la persistance au niveau MAC. Cependant, selon la qualité du canal, un paquet MAC pourrait avoir besoin d'être transmis plus de $M+1$ ou, au contraire, moins de $M+1$ fois pour être reçu sans erreur. Dans un tel contexte, il est préférable d'attribuer la valeur de la persistance pour un paquet MAC en fonction de la qualité du canal. Ainsi, lorsque la fragmentation des paquets IP est nécessaire, il est judicieux de tenir compte du fait que les paquets MAC appartiennent à un même paquet IP et de transférer la persistance pour chaque paquet MAC en un crédit global de transmissions accordé à l'ensemble des paquets MAC constituant le paquet IP [1]. Ainsi, pour un même effort (nombre moyen de retransmissions) au niveau de la couche accès radio, on améliore les performances au niveau TCP/IP. En admettant qu'un paquet IP est fragmenté en N paquets MAC et que la taille des paquets IP est fixe (c'est donc la taille des paquets MAC qui varie en fonction de N), nous accordons un crédit global de transmissions, noté C , à l'ensemble des N paquets MAC appartenant à un même paquet IP. Ce compteur est décrémenté d'une unité à chaque transmission d'un paquet MAC. Nous considérons la technique HARQ à redondance incrémentale (HARQ-IR) de rendement $1/r$. Le paquet MAC à transmettre est tout d'abord encodé par un code Cyclic Redundancy Check (CRC) puis par un code de rendement $1/r$. Le mot de code est alors constitué de r paquets de tailles égales : un paquet contenant l'information encodée (paquet MAC avant codage + CRC) et $r-1$ paquets contenant la redondance. L'HARQ-IR, de par sa redondance, introduit donc un degré de liberté supplémentaire par rapport à l'ARQ. Par la suite, nous désignons les paquets contenant l'information par l'acronyme PMI (Paquet MAC d'Information) et les paquets contenant la redondance par l'acronyme PMR (Paquet MAC de Redondance).

Le compteur C est initialisé à la valeur $r(M+1)N$ puis décrémenté à chaque tentative de transmission d'un paquet MAC (PMI ou PMR). Pour chaque tentative, le processus HARQ-IR s'opère classiquement en transmettant la redondance de manière incrémentale. En cas de succès de la transmission d'un paquet MAC, on passe alors au paquet suivant. Dans le cas contraire on retransmet le même paquet. Le processus s'arrête lorsque le compteur est inférieur au nombre de paquet MAC restant à transmettre.

Par la suite, on distinguera donc deux modèles : le modèle conventionnel (noté MPDU) dans lequel une persistance M est accordée à chaque paquet MAC et le modèle proposé (noté MSDU), dans lequel un crédit global de transmissions, égal à $r(M+1)N$, est accordé à l'ensemble des N paquets MAC appartenant au même paquet IP.

3. Calcul des performances analytiques

Dans [1] les auteurs démontrent analytiquement l'avantage à considérer une approche MSDU associée à de l'ARQ sur la Probabilité de Bonne Réception (PBR) des paquets IP et sur leur délai de transmission. Par contre, le calcul de la PBR des paquets MAC ou IP n'a a priori jamais été publié dans le cas de l'HARQ à redondance incrémentale, quel que soit le modèle considéré (MPDU ou MSDU). Dans cette section, nous calculons les performances analytiques des modèles MPDU et MSDU pour l'HARQ-IR en fonction de i , où i est défini comme étant le nombre maximum de transmissions autorisées pour transmettre l'ensemble des N paquets MAC.

On introduit les notations suivantes :

- π_j , avec $0 \leq j \leq r-1$, le Packet Error Rate (PER) associé à la transmission d'un paquet codé avec un code de rendement $1/(j+1)$. Le paquet correspondant au rendement $1/(j+1)$ est obtenu après concaténation des $j+1$ premiers paquets.
- $P_{N,i}$, la PBR des paquets IP lorsque ceux-ci sont fragmentés en N paquets MAC et qu'on dispose d'un crédit de i transmissions pour transmettre chaque paquet IP.

La transmission de N paquets MAC nécessite au minimum N transmissions, de sorte que l'on considère l'étude de $P_{N,i}$ pour $i \geq N$ ($P_{N,i} = 0$, pour $i < N$). On verra de plus que l'on peut borner l'étude de $P_{N,i}$, celle-ci devenant constante à partir d'une certaine valeur de i que l'on déterminera.

Dans le cas de l'HARQ-IR, $P_{N,i}$ est définie par :

$$P_{N,i} = \Pr\{\text{recevoir } N \text{ paquets sans erreur} \mid i, r, M\} \quad (1)$$

que l'on peut exprimer sous la forme :

$$P_{N,i} = \Pr\left\{ \bigcap_{k=N}^i E_k \mid r, M \right\} \quad (2)$$

où E_k est l'événement correspondant à la réception sans erreur des N paquets MAC en k transmissions. Autrement dit, E_k est réalisé lorsque le dernier paquet est reçu à la dernière (k -ième) transmission. La probabilité d'un tel événement est notée : $p_N(k) = \Pr\{E_k \mid r, M\}$.

On peut noter que lorsque k est strictement supérieur au crédit total de transmissions C accordé à la transmission de l'ensemble des N paquets MAC, on a :

$$p_N(k) = 0, \text{ pour } k > C \quad (3)$$

Le canal étant sans mémoire, les événements sont indépendants de sorte qu'on peut mettre (2) sous la forme :

$$P_{N,i} = \sum_{k=N}^i p_N(k). \quad (4)$$

Il s'agit maintenant de déterminer l'expression de $p_N(k)$. D'une manière générale, la transmission correcte d'un paquet est précédée d'une suite de j échecs successifs dus aux erreurs, j variant de 0 à J . Dans le cas MPDU, on a $J = C_p - 1$ où $C_p = r(M+1)$ désigne le crédit de transmissions autorisées pour chaque paquet MAC. Dans le cas MSDU, on a $J = C_s - N$ où C_s est le crédit de transmissions autorisées pour l'ensemble des N paquets

MAC, avec C_s quelconque. Soit Q_j , la probabilité qu'un paquet soit reçu de cette manière. On trouve simplement :

$$Q_j = (1 - \pi_j) \prod_{l=0}^{j-1} \pi_l \quad (5)$$

Ainsi, la probabilité $p_N(k)$ de recevoir N paquets MAC en k transmissions s'exprime de manière générale en fonction des probabilités Q_j , $0 \leq j \leq J$ sous la forme :

$$p_N(k) = \sum_{\underline{n} \in \Theta(k)} Q_0^{n_0(k)} Q_1^{n_1(k)} Q_2^{n_2(k)} \dots Q_J^{n_J(k)}, \quad (6)$$

où $n_l(k)$, $0 \leq l \leq J$, représente le nombre de paquets MAC reçus correctement après l échecs, $\underline{n} = [n_0(k), \dots, n_J(k)]$ et $\Theta(k)$ l'ensemble des vecteurs \underline{n} décrivant la somme (6). Ces indices sont donc contraints à vérifier $\sum_{l=0}^J n_l(k) = N$. De plus, le nombre de transmissions consommées pour chaque $n_l(k)$ non nul étant égal à $n_l(k)(l+1)$ et le nombre maximum de transmission étant égal à k , les indices doivent aussi vérifier $\sum_{l=0}^J n_l(k)(l+1) = k$. Ces deux contraintes caractérisent complètement l'ensemble $\Theta(k)$.

D'après (6), l'expression (5) se met alors sous la forme :

$$P_{N,i} = \sum_{k=N}^{\min(i,C)} \sum_{\underline{n} \in \Theta(k)} Q_0^{n_0(k)} Q_1^{n_1(k)} Q_2^{n_2(k)} \dots Q_J^{n_J(k)}, \quad (7)$$

expression générale valable pour les cas MPDU et MSDU.

On va maintenant donner des expressions particulières de la forme (7) en fonction de i et du modèle considéré. Quelle que soit la méthode adoptée, lorsque $i = N$, la solution est simple puisqu'elle correspond à N transmissions successives sans erreur, d'où :

$$P_{N,N} = (1 - p_0)^N = Q_{0N}. \quad (8)$$

Lorsque $i \geq C$, la condition (3) associée à (4) nous indique que $P_{N,i}$ devient constante. Dans un tel cas, $P_{N,i}$ est notée P_N^P pour le modèle MPDU et P_N^S pour le modèle MSDU. Dans le cas MPDU, le crédit total est égal au nombre de paquets à transmettre multiplié par le crédit par paquet C_p , soit $C = r(M+1)N$. Dans le cas MSDU on a directement $C = C_s$.

Nous allons maintenant établir l'expression de P_N^P . Tout d'abord, il convient de remarquer que comme $i \geq C$, on a $P_N^P = (P_1^P)^N$ (avec $P_1^P = P_N^P|_{N=1}$), de sorte qu'il suffit de calculer l'expression de P_1^P .

D'après (3) et (4), on déduit que $P_1^P = \sum_{k=i}^C p_1(k)$. De plus, d'après la définition de Q_j dans le cas MPDU, on identifie simplement $p_1(k) = Q_{k-1}$, ce qui amène :

$$P_1^P = \sum_{j=0}^{C_p-1} Q_j. \quad (9)$$

Cette dernière expression peut se simplifier en remarquant qu'il est possible d'écrire les probabilités Q_j pour $j \geq r$, en fonction des probabilités Q_0, Q_1, \dots, Q_{r-1} . En effet, lorsque $j \geq r$, si l'on décompose j sous la forme $j = kr + m$, on peut alors interpréter l'événement correspondant à Q_j comme la transmission successive de k groupes composés de r transmissions erronées (de probabilité $\prod_{l=0}^{r-1} \pi_l$), suivie de la transmission correcte du paquet après m erreurs (de probabilité Q_m). Ainsi, on peut écrire $Q_{kr+m} = Q_m \left(\prod_{l=0}^{r-1} \pi_l \right)^k$, de sorte qu'en utilisant la

décomposition $j = kr + m$, on réexprime (9) sous la forme $P_1^P = \sum_{k=0}^M \sum_{m=0}^{r-1} Q_{kr+m}$, qui devient alors :

$$P_1^P = \sum_{k=0}^M \left(\prod_{l=0}^{r-1} \pi_l \right)^k \times \sum_{m=0}^{r-1} Q_m. \quad (10)$$

En appliquant la définition (5), on montre facilement que $\sum_{m=0}^{r-1} Q_m = 1 - \prod_{j=0}^{r-1} \pi_j$ ce qui reporté dans (10) permet d'obtenir l'expression recherchée :

$$P_1^P = \left(1 - \prod_{l=0}^{r-1} \pi_l \right) \times \sum_{k=0}^M \left(\prod_{l=0}^{r-1} \pi_l \right)^k. \quad (11)$$

Dans le cas MSDU, nous n'avons pas trouvé d'expression analytique pour P_N^S faisant intervenir uniquement les probabilités π_l . Toutefois, on peut montrer que les performances de la méthode MSDU sont supérieures à celle de la méthode MPDU dans le cas $i \geq C$ ce qui justifie son intérêt.

Afin de se mettre dans une situation comparable, on se place dans le cas où C est identique pour les deux méthodes. Soit $\Theta^S(k)$ et $\Theta^P(k)$ les ensembles des vecteurs d'indices de la formule (7) respectivement pour le cas MSDU et MPDU. On peut tout d'abord remarquer que les éléments de ces ensembles étant à valeurs dans \mathbb{N}^{J+1} , et que J vaut respectivement $C-N$ pour la méthode MSDU et $C/N-1$ pour la méthode MPDU, la dimension des vecteurs d'indice est plus grand dans le cas MSDU. De plus, le fait d'autoriser un crédit de transmissions global, fait que l'ensemble des événements du cas MSDU inclut celui du cas MPDU. On a alors $\Theta^S(k) = \Theta^P(k) \cup \bar{\Theta}(k)$, avec $\bar{\Theta}(k) \neq \{\emptyset\}$. Ainsi, dans le cas MSDU, (8) devient $P_N^S = \sum_{k=N}^C \sum_{\underline{n} \in \Theta^S(k)} Q_0^{n_0(k)} \dots Q_{C-N}^{n_{C-N}(k)}$ qui peut se décomposer comme :

$$P_N^S = \sum_{k=N}^C \left\{ \sum_{\underline{n} \in \Theta^P(k)} Q_0^{n_0(k)} \dots Q_{C/N-1}^{n_{C/N-1}(k)} + \sum_{\underline{n} \in \bar{\Theta}(k)} Q_0^{n_0(k)} \dots Q_{C-N}^{n_{C-N}(k)} \right\},$$

expression dans laquelle on identifie la probabilité P_N^P et qui s'exprime alors :

$$P_N^S = P_N^P + \sum_{k=N}^C \sum_{\underline{n} \in \bar{\Theta}(k)} Q_0^{n_0(k)} \dots Q_{C-N}^{n_{C-N}(k)}. \quad (14)$$

Comme $\bar{\Theta}(k) \neq \{\emptyset\}$, le terme de droite de l'expression (14) est supérieur à 0, d'où l'on en déduit que $P_N^S > P_N^P$. Ainsi, lorsque $i \geq C$, la méthode MSDU est la plus performante.

4. Étude des performances

Nous présentons dans cette section les performances comparées en termes de PBR et de délai de la méthode MSDU adaptée aux HARQ-IR et ARQ proposée dans [1]. Le délai est proportionnel au nombre de paquets MAC transmis. Aussi, à taille de paquet IP constante, le délai sera estimé par le nombre de paquets MAC transmis, divisé par N . L'HARQ-IR est implémenté en utilisant les codes convolutifs à rendements compatibles [2] avec un rendement du code mère égal à $1/4$ ($r=4$). Le canal de transmission est modélisé par un canal binaire symétrique de probabilité d'erreur p_e .

Nous évaluons les performances dans le cas $i \geq C$, en fonction des paramètres M et N et cherchons à déterminer les meilleurs compromis pour le choix de ces valeurs en

fonction du critère choisi (PBR ou délai). Sur la figure 2.a (resp. 2.b), nous avons tracé la PBR (resp. le délai) pour notre solution et celle proposée en [1] en fonction de p_e pour différentes valeurs de M , le paramètre N étant fixé.

Tout d'abord nous remarquons que l'amélioration des performances est très significative en faveur de l'HARQ. Dans notre exemple, le seuil à partir duquel les performances commencent à se dégrader passe de $p_e = 3 \cdot 10^{-3}$ les performances de notre solution par rapport à celles de la méthode décrite en [1] sont toujours meilleures, en termes de PBR et de délai

Nous remarquons que plus M est grand, plus la PBR est grande, ce qui était prévisible. De plus, pour $p_e < 10^{-1}$, le délai reste constant quel que soit M , car, pour de telles valeurs de p_e , le PMI est reçu correctement après un faible nombre de retransmissions (toutes les ressources de retransmissions ne sont pas utilisées). Quand la PEB est supérieure à 10-1, le délai diminue quand M augmente : en effet, le délai au niveau MAC est plus long quand M est élevé car les PMI sont transmis un grand nombre de fois, mais de ce fait, le délai au niveau IP est plus court car le paquet IP est reçu correctement après un plus faible nombre de retransmissions. Ce phénomène est bien visible sur l'ARQ. Pour l'HARQ, il existe néanmoins mais à des valeurs très élevées de PEB. Enfin,

Sur les figures 3.a et 3.b, nous faisons varier le paramètre N , le paramètre M étant fixé, et nous comparons les résultats obtenus au niveau IP avec l'algorithme proposé à ceux obtenus avec la méthode décrite en [1] : nous constatons, d'une part, que plus N est important, plus la PBR est grande et, d'autre part, que plus N est grand, plus le délai diminue pour l'ARQ. En revanche l'HARQ semble insensible au choix de N dans la gamme de p_e considérée. Dans le cadre de l'ARQ, cela signifie qu'il est préférable de transmettre de petits paquets MAC. Cette remarque n'est pas si simple qu'elle n'y paraît puisque, certes, un paquet MAC de petite taille a plus de chances d'être reçu correctement qu'un paquet de grande taille mais plus de petits paquets MAC (qui forment l'unique paquet IP) doivent être reçus correctement afin que le paquet IP associé le soit également. Enfin, nous constatons que les performances obtenues avec l'HARQ sont grandement supérieures, en termes de PBR et de délai, à celles obtenues avec l'ARQ.

En conclusion, les résultats obtenus suite aux différentes simulations montrent qu'il est possible d'optimiser les performances au niveau de la couche réseau en tenant compte du fait que les paquets MAC à transmettre appartiennent à un même paquet IP et en utilisant l'HARQ-IR avec codes convolutifs à rendements compatibles au niveau accès radio. Cette technique peut toutefois être étendue aux Turbo-codes et aux Low Density Parity Check Codes (LDPC), la solution proposée ici étant indépendante du type de code. Nous avons également établi des règles permettant de paramétrer notre solution en fonction des performances que l'on souhaite améliorer au niveau IP. Dans le contexte de l'ARQ, les valeurs de la persistance MAC (M) et du nombre de paquets MAC (N) constituant un paquet IP doivent être élevées si on souhaite à la fois

augmenter la PBR et diminuer le délai. Le système basé sur l'HARQ est beaucoup moins sensible au choix des paramètres M et N et est évidemment bien meilleur qu'un système basé sur l'ARQ.

Références

- [1] Youngkyu Choi, Sunghyun Choi and Seokhyun Yoon, "MSDU-ARQ Scheme for IP-Level Performance Maximization", IEEE Globecom 2005 proceedings, St. Louis, Missouri, 2005.
- [2] Joachim Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications", IEEE Transactions on Communications, Vol. 36, n°4, p.389 – 400, April 1988.
- [3] Leonardo Badia, Marco Levorato and Michele Zorzi, "Analytical Investigation with Markov Models of Selective Repeat Type II Hybrid ARQ", IEEE Globecom 2006 proceedings, San Francisco, California, 2006.
- [4] Shu Lin and Philip S. Yu, "A Hybrid ARQ Scheme with Parity Retransmission for Error Control of Satellite Channels", IEEE Transactions on Communications, Vol. 30, n°7, July 1982.

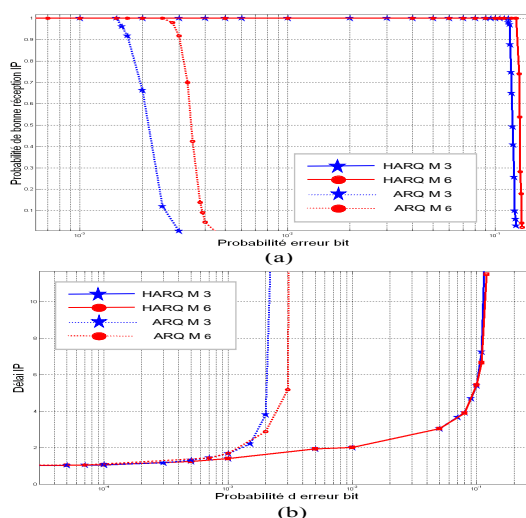


Fig 2 : PBR IP (a) et délai IP (b) avec HARQ et ARQ en fonction de M ($N = 30$)

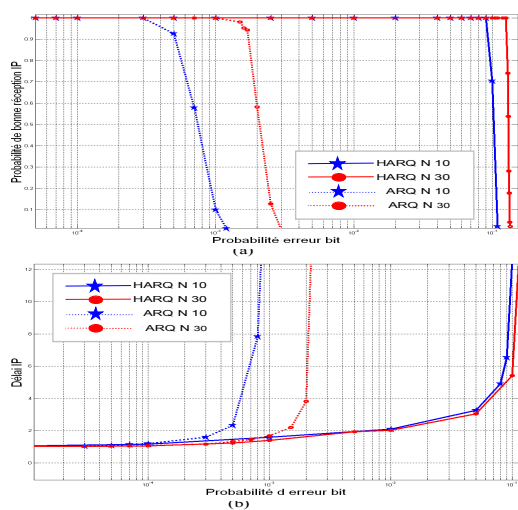


Fig 3 : PBR IP (a) et délai IP (b) avec HARQ et ARQ en fonction de N ($M = 3$)