

Analytic Optimization of Hybrid ARQ Performance in Wireless Packet Data Systems

Jeong-woo Cho, Yong Chang, Yuchul Kim and JaeHong Chon

Telecommunication R&D Center, Samsung Electronics, Republic of Korea

E-mail: ggumdol@gmail.com

Abstract— We propose an analytical framework that optimizes the performance of various Hybrid Automatic Repeat reQuest (HARQ) schemes in wireless packet data systems. The optimization framework lends itself to maximizing the user throughput without assuming a specific HARQ scheme, a target packet error rate, and a mapping of signal-to-interference-noise ratios (SINR) into modulation and coding scheme (MCS) sets. Thus, a user selects an optimal amount of redundancy bits used for retransmissions, an optimal packet error rate, and an optimal SINR-MCS mapping such that the user throughput would be maximized. We show that there exists an optimal redundancy block size of incremental redundancy (IR) scheme when the number of retransmissions is limited, while as small as possible redundancy size is optimal when the number of retransmissions is unlimited. Our analytical results are verified through extensive simulations of the IEEE 802.16e system.

I. INTRODUCTION

Even with untiring efforts to satisfy service requirements of various applications, wireless packet data systems have not yet matched wired networks, *e.g.*, the Internet, in throughput. Since the ultimate demand of mobile users is to enjoy high data throughput comparable to that of wired networks, wireless packet data systems have been evolving into more advanced systems such as the evolved UTRA [1] and IEEE 802.16e [2] where Hybrid Automatic Repeat reQuest (HARQ) schemes as well as broader frequency bands and orthogonal frequency division multiple access (OFDMA) have been the essentials to their performance.

Unlike the simple ARQ scheme where a receiver scraps former erroneous packets and tries to decode each packet independently, an HARQ receiver buffers erroneous packets until it successfully decodes the original data by exploiting the signal information of the buffered packets. Since the soft information of symbols in the initial transmission are combined with that of the following retransmissions, each retransmission increases the probability of successful decoding at the receiver.

There are two fundamental characteristics of HARQ contributing to the performance improvement of systems adopting HARQ. That is to say, HARQ makes decoding attempts in 1) *combinative* and 2) *iterative* ways. Firstly, an HARQ receiver combines the signal information of packets received so far and that of the last packet to form an aggregate signal information. If an HARQ receiver does not exploit the combined information, HARQ might be reduced to a simple ARQ. Secondly, if the HARQ receiver does not try to decode

per packet reception, precious air resources are wasted since more than enough packets might be transmitted. In this case, HARQ is reduced to a simple forward error correction (FEC) scheme and cannot utilize the *early termination* probability.

To utilize these characteristics better, it stood to reason that various HARQ schemes should be introduced. In *Chase combining* (CC) scheme [3], a sender retransmits the original packet for each decoding failure of its corresponding receiver until the receiver successfully decodes the combined information. Thus the same packet is used for all retransmission requests in CC scheme. In *incremental redundancy* (IR) scheme [4], [5], a sender encodes the original (systematic) bits to produce the highest rate code of the corresponding code family. The sender transmits only the systematic bits at first and transmits one redundancy block for each retransmission failure of its corresponding receiver. In contrast to CC, a redundancy block is not necessarily the original packet but can be one of encoded blocks and the redundancy block size can be finely adjusted so that an appropriate amount of redundancy bits is delivered to the receiver. As the receiver combines more redundancy blocks with the original packet, the effective code rate becomes lowered to achieve a higher coding gain. In general, Turbo codes [4], Rate Compatible Punctured Convolutional (RCPC) [5] codes, and Low-Density Parity-Check (LDPC) [6] codes are used to generate redundancy blocks. Note that CC is a special case of IR and is thus implicitly supported in wireless packet data systems adopting IR scheme, *e.g.* [1] and [2].

II. MOTIVATION AND RELATED WORKS

To improve the actual performance of user applications, it is widely known that undisputed key performance indices are the user throughput, the packet error rate (PER), and the packet delay where the packet delay is closely related to the packet error rate. To better understand the importance of the performance indices, let us consider two representative applications one by one. A real-time application generally requires a lower packet delay which is achieved by controlling the packet error rate and the number of retransmissions. For a real-time application, packets received after a certain number of retransmissions is useless as packets should arrive timely within a specified time bound. Moreover, it is well known that transport layer protocols such as TCP NewReno and TCP Reno, that are mainly adopted by delay-tolerant applications,

are very vulnerable to even low packet error rates and large packet delays. The higher the packet error rate or the larger the packet delay, the lower the TCP throughput. Therefore, achieving sufficiently low packet error rates is crucial for both applications. Furthermore, it is certain that an advanced wireless packet data system should also provision abundant user throughputs to satisfy both applications. We can conclude that all of three performance indices are indispensable for satisfying the performance demand of user applications.

Whereas there have been a few approaches to improve the performance of HARQ from different angles, against our expectations, we have found that there has not been a unified approach that maximizes the throughput performance of HARQ while simultaneously maintaining the packet error rate and the packet delay at sufficiently low values. In almost work including [7] and [8], either only CC scheme was analyzed or IR scheme was analyzed without adjusting the redundancy block size or the mapping of signal-to-noise ratios (SINR) into modulation and coding scheme (MCS) sets was simply determined such that the packet error rate is less than or equal to a certain threshold. Unlike [7] and [8], Zheng and Viswanathan in their noticeable work [9], introduced a systematic and analytical approach that maximizes the user throughput of CC scheme in downlink transmissions of wireless packet data systems and proposed several scheduling algorithms. Though their framework is thought-provoking from diverse standpoints of HARQ optimization, they neither did consider IR schemes nor did impose a constraint upon the packet error rate.

In this paper, we provide a systematic and extended analytical framework pertinent to IR scheme with utilizing the inherent capability of IR scheme, *i.e.*, adjusting the redundancy block size. That is to say, we would like to answer the following questions in particular:

How much improvement can we make by optimally adjusting the redundancy block size of IR scheme?

To answer this question, we formulate an optimization problem which maximizes the user throughput in wireless packet data systems with considering diverse aspects of HARQ such as the redundancy block size, the PER and the mapping of SINR into MCS sets. Furthermore, the throughput optimization incorporates a practical constraint on the packet error rate so that packet delays and retransmission probability are controlled. We also derive an exact and simple expression of the user throughput and analyze the asymptotic behavior of HARQ schemes as the number of retransmissions goes to infinity, that were not addressed in previous works, either. Finally, we show through extensive simulations that our optimization approach with variable redundancy block size significantly improves the user throughput of an existing system, *i.e.*, IEEE 802.16e system, while maintaining a designated residual packet error rate.

III. PERFORMANCE ANALYSIS OF THE HARQ THROUGHPUT

There are various scheduler algorithms used for resource allocation such as the Proportional Fair Scheduler [10] and

Maximum Weighted-Rate Schedulers [11]. To better focus on the optimal operation of the underlying physical layer, we have the following major premise in this paper: each user is allocated a fixed amount of time-slots by a downlink scheduler in advance. If the available air resources comprise $\bar{\tau}$ time-slots in seconds and n_f subcarriers (*e.g.*, OFDMA systems like [1] and [2]), the two-dimensional resources can be regarded as $\tau_0 = \bar{\tau} \cdot n_f$ time-slots. Therefore, without loss of generality, it is assumed that τ_0 represents both the allocated time-slots and subcarriers. To facilitate the mathematical analysis, the following parameters are required:

- γ_0 : the signal-to-interference-noise ratio (SINR) of the initial transmission, which satisfies $\gamma_0 > 0$.
- γ : the effective SINR, which satisfies $\gamma > 0$.
- α : the ratio of redundancy block size to original packet size, which satisfies $\alpha > 0$.
- r : the channel rate achieved by a modulation and coding scheme (MCS), which satisfies $0 \leq r \leq \bar{r}$.
- \bar{k} : the maximum number of retransmissions.
- τ_0 : the amount of time-slots to be consumed at the initial transmission.
- $P_e(r, \gamma)$: the packet error rate (PER) for a given r and γ , which satisfies $0 < P_e(r, \gamma) < 1$.
- P_0 : the PER of the initial transmission, *i.e.*, $P_0 \triangleq P_e(r, \gamma_0)$.

For tractable analysis, it is assumed that the wireless channel condition is constant or slowly-varying during retransmissions as that of the initial transmission. That is, the initial transmission and retransmissions adopt the same MCS. In cases where this assumption does not hold, the decoder structure of the receiver as well as the performance analysis becomes very complex. In practical wireless systems, for a given r (which is decided by the MCS), the packet error rate $P_e(r, \gamma)$ decreases with γ . It is assumed that $\lim_{\gamma \rightarrow \infty} P_e(r, \gamma) = 0$ for all r .

The average throughput of the user of interest can be defined as the ratio of the expected number of successfully decoded systematic bits $E(B|\gamma_0, \alpha, r, \bar{k}, \tau_0)$ to the expected time-slots $E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0)$ occupied for the initial transmission and retransmissions:

$$T(\gamma_0, \alpha, r, \bar{k}, \tau_0) = \frac{E(B|\gamma_0, \alpha, r, \bar{k}, \tau_0)}{E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0)} \quad (1)$$

where the equality holds by a fundamental theorem of renewal reward processes [12].

A. Modeling The HARQ Combining Gain

Due to the intrinsic complexity of coding techniques used for IR scheme, the relation between α and SINR increment per retransmission, which is denoted by $\Delta\gamma$, has not been derived in a tractable form. However, Frederiksen et al. in their work [13], [14] suggested a manageable approximate relation between $\Delta\gamma$ and α and showed their modeling is quite exact through various simulations. Note that extensive simulation results in [15] also supports their model. For

mathematical tractability of HARQ performance analysis, we slightly modified the HARQ combining gain model in [13] and assume that the ratio of SINR increment per retransmission to original SINR, *i.e.*, $\Delta\gamma/\gamma_0$, is proportional to α as follows:

$$\Delta\gamma = g \cdot \alpha \cdot \gamma_0 \quad \text{where } g = \begin{cases} 1, & \text{if CC is adopted;} \\ g_{\text{IR}}, & \text{if IR is adopted,} \end{cases}$$

where the coding gain $g_{\text{IR}} \geq 1$ *does* depend on the modulation, coding scheme and packet size [13]–[16]. Note that the coding gain of IR scheme for $\alpha = 1$ is $g = g_{\text{IR}}$ because CC corresponds to a simplified version of IR. Since the energy of a redundancy block is α times that of the original packet, the HARQ combining model given in the above is fairly intuitive.

B. Maximizing The User Throughput

We investigate the user throughput performance of the following two cases since the qualitative characteristics of the one is clearly distinguishable from those of the other: when the number of retransmissions is 1) *limited* and 2) *unlimited*.

1) When the number of retransmissions is limited:

As each transmission attempt of an original packet and redundancy blocks occupies a portion of the receiver buffer, in general, receiver buffers required for HARQ schemes should be considerably large, and should support fast read/write operations into the bargain. Thus the buffers might cost too much if the number of redundancy blocks for an original packet is not upper-bounded. From an angle of performance, the packet delays incurred by repeated retransmissions should be kept within bounds. Thus we first consider the case when the number of retransmissions is limited.

We provide a systematic approach to maximize the user throughput while the average delay experienced by packets are controlled by confining the maximum number of retransmissions to \bar{k} . That is, an explicit expression of (1) is derived through Proposition 1 and Theorem 1. (All proofs are in Appendix.)

Proposition 1. *The expected time-slots spent by the transmission and retransmission attempts related to an original packet is given by*

$$E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0) = \tau_0 \left(1 + \alpha \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+jg\alpha)\gamma_0) \right).$$

Theorem 1. *The user throughput is given by*

$$T(\gamma_0, \alpha, r, \bar{k}) = \frac{r \cdot \left(1 - \prod_{j=0}^{\bar{k}} P_e(r, (1+jg\alpha)\gamma_0) \right)}{1 + \alpha \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+jg\alpha)\gamma_0)}. \quad (2)$$

Note that the case $\alpha = 1$ and $g = 1$ corresponds to CC scheme. In this case, the equation becomes

$$T(\gamma_0, 1, r, \bar{k}) = \frac{r \cdot \left(1 - \prod_{j=0}^{\bar{k}} P_e(r, (1+j)\gamma_0) \right)}{1 + \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+j)\gamma_0)}. \quad (3)$$

It should be remarked that (2) incorporates both IR and CC schemes and (3) is much simpler than the throughput

expression of CC scheme obtained in [9] because we simplified $E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0)$ in Proposition 1 while [9] applied a complicated expression of $E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0)$ (which is similar to (7) in Appendix) to (1) directly. If this theorem is applied to a practical wireless system, γ_0 and \bar{k} are predetermined so that only α and r can be varied. However, it is not clear what value of α and r would maximize (2). In fact, the maximization demands a complicated optimization with regard to α and r as (2) is a nonconvex function of the two variables.

From (2), it is easy to see that

$$\lim_{\alpha \rightarrow 0} T(\gamma_0, \alpha, r, \bar{k}) = r \cdot \left(1 - P_0^{\bar{k}+1} \right), \quad \lim_{\alpha \rightarrow \infty} T(\gamma_0, \alpha, r, \bar{k}) = 0. \quad (4)$$

If we define $T(\gamma_0, 0, r, \bar{k})$ as the above limit value and $T(\cdot, \alpha, \cdot, \cdot)$ is continuous, one can show that there exists an optimal $\alpha = \alpha^*$ by applying (4) to (2).

Observation 1. *If $T(\gamma_0, 0, r, \bar{k})$ is defined as $r \cdot (1 - P_0^{\bar{k}+1})$ and $T(\cdot, \alpha, \cdot, \cdot)$ is continuous on $[0, \infty)$, there exists an optimal $\alpha = \alpha^*$ at which the user throughput $T(\gamma_0, \alpha, r, \bar{k})$ is maximized.*

2) When the number of retransmissions is unlimited:

As mentioned above, the number of retransmissions should be kept within bounds in practical wireless systems. While the case when the number of retransmissions is limited ($\bar{k} < \infty$) requires an exhaustive optimization of (2), we can gain an insight into the relation between the user throughput and HARQ parameters by observing the asymptotic behavior of the case when the number of retransmissions is unlimited ($k = \infty$). Though we can derive $T(\gamma_0, \alpha, r, \infty)$ by an independent method, we begin from (2) for brevity:

$$\begin{aligned} T(\gamma_0, \alpha, r, \infty) &= \lim_{\bar{k} \rightarrow \infty} \frac{r \cdot \left(1 - \prod_{j=0}^{\bar{k}} P_e(r, (1+jg\alpha)\gamma_0) \right)}{1 + \alpha \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+jg\alpha)\gamma_0)} \\ &= \frac{r}{1 + \alpha \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} P_e(r, (1+jg\alpha)\gamma_0)} \end{aligned}$$

where the second equality holds by the assumption $\lim_{\gamma_0 \rightarrow \infty} P_e(r, \gamma_0) = 0$ for all r . Then, the supremum of the user throughput can be defined as

$$T_{\text{sup}}(\gamma_0, \infty) \triangleq \sup_{\substack{\alpha \in (0, \infty) \\ r \in [0, \bar{r}]}} \frac{r}{1 + \alpha \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} P_e(r, (1+jg\alpha)\gamma_0)}. \quad (5)$$

Theorem 2. *The user throughput of $\bar{r} - \delta$ is achievable for arbitrarily small $\delta > 0$. More precisely,*

$$T_{\text{sup}}(\gamma_0, \infty) = \sup_{\substack{\alpha \in (0, \infty) \\ r \in [0, \bar{r}]}} T(\gamma_0, \alpha, r, \infty) = \bar{r}, \quad \forall \gamma_0 > 0. \quad (6)$$

This theorem implies that the maximum data rate \bar{r} is achievable for any γ_0 by letting α be arbitrarily small. Though the packet error rate $P_0 = P_e(r, \gamma_0)$ increases as γ_0 decreases in practical wireless systems, IR scheme overcomes the high packet error rate by transmitting an infinitesimal redundancy block for each retransmission request. Note that IR scheme should face with the problem of uncontrolled packet delays, which is sacrificed for achieving \bar{r} . It is easy to see that the aggregate round-trip delay is proportional to

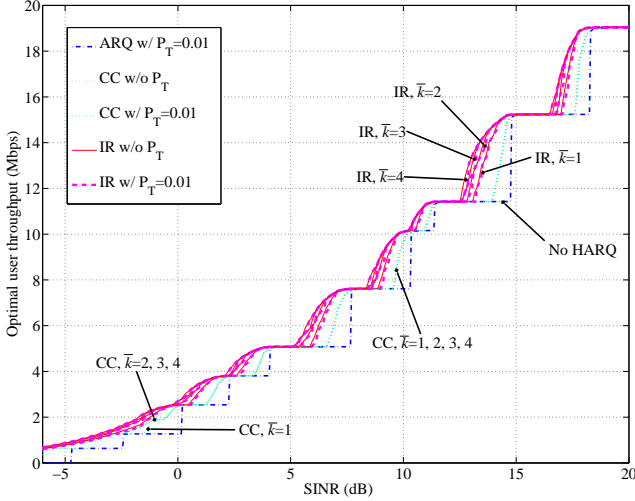


Fig. 1. Optimal user throughputs for various HARQ schemes.

$1 + \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} P_e(r, (1+jg\alpha)\gamma_0)$ where $P_e(r, (1+jg\alpha)\gamma_0)$ approaches P_0 as α approaches 0. In other words, the high user throughput in the above is achieved at the cost of uncontrollable delays.

Let us consider two special cases. A simple ARQ sender retransmits the original packet and does not combine the retransmitted packets with the original packet. In this case, the supremum of the user throughput becomes $T_{\text{sup}}^{\text{ARQ}}(\gamma_0, \infty) = \sup_{r \in [0, \bar{r}]} r \cdot (1 - P_0)$ where the factor $1 - P_0$ prevents the simple ARQ from achieving data rates arbitrarily close to \bar{r} . Readers are encouraged to derive the fact that $T_{\text{sup}}^{\text{ARQ}}(\gamma_0, \infty) = T_{\text{sup}}^{\text{ARQ}}(\gamma_0, \bar{k})$ for all $\bar{k} \geq 0$, which holds due to the memoryless operation of simple ARQ. In the case of CC scheme, we can lower-bound the user throughput $T_{\text{sup}}^{\text{CC}}(\gamma_0, \infty) \triangleq \sup_{r \in [0, \bar{r}]} T(\gamma_0, 1, r, \infty)$ as follows by replacing α with 1 in (9).

$$\underbrace{T_{\text{sup}}(\gamma_0, \infty)}_{\bar{r}} \geq T_{\text{sup}}^{\text{CC}}(\gamma_0, \infty) \geq \underbrace{T_{\text{sup}}^{\text{ARQ}}(\gamma_0, \infty)}_{\sup_{r \in [0, \bar{r}]} r \cdot (1 - P_0)}.$$

This relation implies that ARQ schemes can be enumerated as IR, CC and simple ARQ in descending order of throughput if \bar{k} is unlimited.

IV. SIMULATION RESULTS

Since a resource block size in real systems should be an integer multiple of a system-specific amount of bits, we assume that α is an integer multiple of α_{Δ} , i.e., $\alpha = k \cdot \alpha_{\Delta}$ for all integer $k \geq 1$. In this section, we use $\alpha_{\Delta} = 0.1$. As for the condition imposed on the number of retransmissions in Section II, we use only four \bar{k} values, i.e., 1~4 to achieve small packet delays and to mitigate receiver buffers and packet overheads. Moreover, upper protocol layers require that the packet error rate should be less than a certain threshold as we discussed in Section II. Thus we define the *aggregate packet error rate* as $P_{\Sigma} = \prod_{j=0}^{\bar{k}} P_e(r, (1+jg\alpha)\gamma_0)$, which is the probability

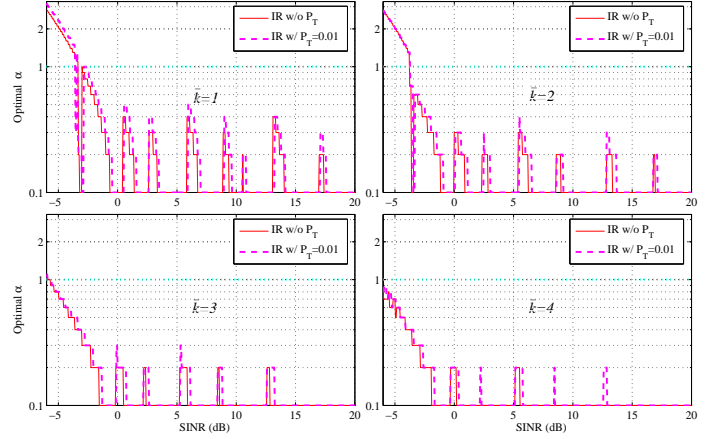


Fig. 2. Optimal α for various \bar{k} .

that the receiver fails in decoding after receiving the original packet and \bar{k} redundancy blocks. Then the aggregate packet error rate should not be greater than the *target aggregate packet error rate*, which is denoted by P_T . Therefore, for a given γ_0 and \bar{k} , we maximize (2) by varying α and r , such that $\prod_{j=0}^{\bar{k}} P_e(r, (1+jg\alpha)\gamma_0) \leq P_T$. We show through simulations that the deterioration of user throughputs caused by this constraint is negligible.

We adopt the TDD mode of WirelessMAN-OFDMA 8.75 MHz channel basic PHY Profile in the IEEE 802.16e mobile broadband wireless access system for our simulation as it supports both CC and IR schemes [2] and its performance is watched with keen interest. There are 10 MCS sets whose data rates range from 650Kbps to $\bar{r} = 19.0$ Mbps. The packet size used for MCS sets ranges from 60bytes to 480bytes. The packet error rates in additive white gaussian noise (AWGN) channel for MCS sets are obtained through extensive simulations based on the technical specification document [2].

One indispensable assumption is that $g_{\text{IR}} = 1$ all the time. It is apparent that, the performance gain of IR scheme is more emphasized when g_{IR} is greater than 1. However, it is not easy to optimize (2) since g_{IR} weakly depends on the modulation, coding scheme and packet size [13]–[16] in a complicated way. Since IR scheme with $g_{\text{IR}} = 1$ always performs worse than that with $g_{\text{IR}} > 1$, simulation results in this work present lower bounds for all cases with $g_{\text{IR}} \geq 1$. Moreover, we found through simulations that the main benefit of IR scheme lies in its capability to finely tune the redundancy block size rather than in its intrinsic coding gain.

In simple ARQ scheme, denoted by ‘ARQ w/ $P_T = 0.01$ ’, an original packet is retransmitted for each decoding failure at the receiver and the combined information is not exploited for decoding. Thresholds of MCS sets in simple ARQ are designated such that the packet error rates experienced with the specified MCS set for a given γ_0 is less than or equal to $P_T = 0.01$. In ‘CC w/o P_T ’ and ‘CC w/ $P_T = 0.01$ ’, two CC schemes are adopted with and without the condition on

the aggregate packet error rates, $P_\Sigma \leq P_T$, respectively. Two IR schemes, ‘IR w/o P_T ’ and ‘IR w/ $P_T = 0.01$ ’, are defined in a similar way. Note that all HARQ schemes are reduced to simple ARQ if $\bar{k}=0$ is used. In total, 17 schemes are simulated since each of 4 HARQ schemes is simulated for 4 different \bar{k} values, *i.e.*, $1 + 4 \cdot 4 = 17$.

A. User Throughput Under the Conditions on \bar{k} and P_Σ

User throughputs of 17 schemes for various γ_0 and \bar{k} are shown in Fig. 1. IR schemes significantly outperform CC schemes and simple ARQ scheme all the time. In particular, the superb performance of IR schemes is highlighted when γ_0 is near the MCS thresholds of simple ARQ. For example, if γ_0 is 14dB(=10^{1.4}), two IR schemes for $\bar{k} = 3$ have the same throughput of 14.8Mbps, while two CC schemes for $\bar{k} = 3$ have that of 11.7Mbps and the simple ARQ have that of 11.4Mbps. Though CC and IR schemes have the same packet error rate of 0.30, they have different values of α , that are $1(=10 \cdot \alpha_\Delta)$ and $0.1(=\alpha_\Delta)$, respectively. It can be calculated that $\alpha = 0.1$ results in the residual packet error rate of $P_e((1 + 3 \cdot 1 \cdot 0.1)\gamma_0) = 0.001$ and the aggregate packet error rate of $P_\Sigma = \prod_{j=0}^3 P_e((1 + j \cdot 1 \cdot 0.1)\gamma_0) = 3 \cdot 10^{-7}$, that are sufficiently small. Therefore, the superior performance of IR schemes at $\gamma_0=14$ dB arises from the fact that the redundancy block size is sufficiently small whereas CC schemes waste the air resource by retransmitting the entire original packet. Similar phenomena occur in Fig. 1 throughout. It can be observed that $\alpha=1$ is extravagant in many cases from a standpoint of user throughput.

User throughputs of CC schemes with and without the constraint $P_\Sigma \leq P_T$ are indistinguishable in Fig. 1 since the effective SINR, γ , is increased significantly per retransmission in wireless channels. Note also that, if we fix the value of \bar{k} , the throughput difference between two IR schemes with and without the constraint $P_\Sigma \leq P_T$ is also almost negligible. This observation means that the throughput-optimal α of IR schemes does not often result in high aggregate packet error rates. In other words, IR scheme effectively maximizes the user throughput even when the satisfactory packet error rates are achieved by the condition $P_\Sigma \leq P_T$.

Let us focus on the effect of \bar{k} . As shown in Fig. 1, the user throughputs of IR schemes are slightly different for various \bar{k} . However, it should be remarked that IR schemes even for $\bar{k}=1$ significantly outperforms CC schemes with $\bar{k}=4$, in user throughput. Since large \bar{k} is undesirable and performance indices for $\bar{k}=2\sim 4$ are nearly the same, $\bar{k}=2$ might be a reasonable choice for IR schemes, since both the high user throughputs and low packet delays are achievable with this choice.

The optimal α in Fig. 2 implies that the smaller the \bar{k} value or the smaller the SINR, the larger the required α value. It can be observed that IR schemes with the condition $P_\Sigma \leq P_T$ result in the reduction of α for some SINR values. Note that this result coincides with the Theorem 2, which implies that optimal α decreases with \bar{k} . Except for $\gamma_0 \leq 0$ dB, the optimal α is less than or equal to 0.5, 0.4, 0.3 and 0.2, respectively for

$\bar{k} = 1, 2, 3$ and 4. To simplify the selection of α , a real wireless packet data system might use these upper bounds (0.5, 0.4, 0.3 and 0.2) as its α for $\gamma_0 > 0$ dB.

Optimal MCS thresholds for all schemes are shown in Fig. 3. All MCS thresholds of HARQ schemes for a given SINR are always larger than those of simple ARQ. The point of Fig. 3 is that IR schemes adopt higher MCS sets than CC schemes does throughout the SINR. Note that HARQ schemes adopt very aggressive α values for low SINR in order to skip over low MCS thresholds such as 650Kbps and 1.3Mbps, as shown in Fig. 2. Since the ratios of channel rates determined by low MCS sets for low SINR is very high, *e.g.*, 1.3Mbps/650Kbps=2, HARQ schemes are inclined to skip over several MCS thresholds to adopt higher MCS sets. In other words, the gain obtained by adopting MCS sets is larger than the waste of time-slots incurred by high α . At low SINR, $\gamma_0 \leq 0$ dB, MCS selection is rather irregular since both MCS sets and α are not continuous but discrete.

V. CONCLUSION

Our contribution is three-fold: Firstly, we have presented a systematic and analytical framework that optimizes user throughputs in wireless packet data systems adopting general HARQ schemes by importing a moderately simplified model of HARQ combining gain. Moreover, the optimization framework provided in this paper is most suitable for identifying not only the optimal redundancy block size but also the optimal packet error rate and the the optimal mapping of signal-to-noise ratios into modulation and coding schemes.

Secondly, several mathematical characteristics of HARQ performance have been found. That is to say, the user throughput expressions derived in this paper are much simpler than those of previous work and pertain to be utilized in real wireless packet data systems. Moreover, we have proven that there exists an optimal redundancy block size when the number of retransmissions is limited and as small as possible redundancy block size is optimal when the number of retransmissions is unlimited.

Thirdly, to answer the question brought up in Section II, we have shown the excellent performance of IR schemes by simulating a real wireless packet data system. It is worth noticing that IR schemes significantly outperforms CC schemes in throughput while achieving both small packet error rates and low packet delays. Therefore, we can conclude that the strength of IR schemes mainly lies in its capability to adjust its redundancy block size adaptively rather than in its intrinsic coding gain.

APPENDIX: PROOFS

A. Proof of Proposition 1

Since we are assuming $\Delta\gamma = g \cdot \alpha \cdot \gamma_0$, $E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0)$ can be simplified as shown in (8).

B. Proof of Theorem 1

It is easy to see that the expected number of successfully decoded bits is given by $E(B|\gamma_0, \alpha, r, \bar{k}, \tau_0) =$

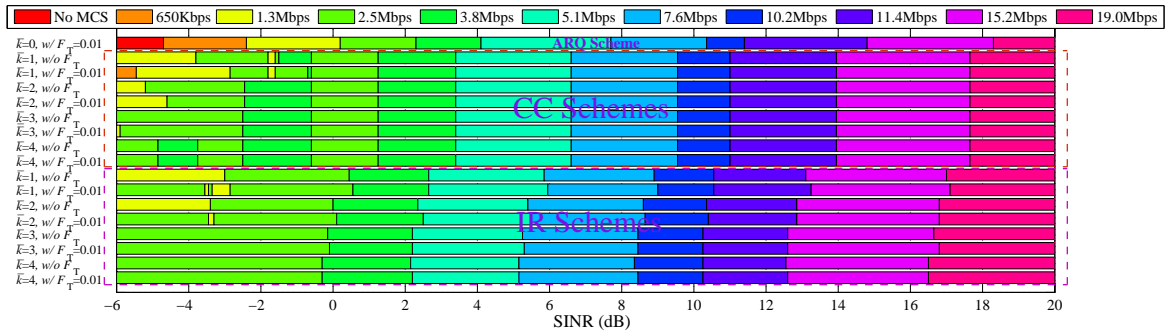


Fig. 3. Optimal MCS thresholds for various HARQ schemes.

$$\begin{aligned}
E(\tau|\gamma_0, \alpha, r, \bar{k}, \tau_0) &= \tau_0 \left(\sum_{k=0}^{\bar{k}} (1+k\alpha)(1-P_e(r, (1+k\alpha)\gamma_0)) \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) + (1+\bar{k}\alpha) \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right) \\
&= \tau_0 \left(\sum_{k=0}^{\bar{k}} ((1+k\alpha) \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) - (1+k\alpha) \prod_{j=0}^k P_e(r, (1+j\alpha)\gamma_0)) + (1+\bar{k}\alpha) \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right) \\
&= \tau_0 \left(\sum_{k=0}^{\bar{k}} ((1+(k-1)\alpha) \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) - (1+k\alpha) \prod_{j=0}^k P_e(r, (1+j\alpha)\gamma_0)) + \alpha \prod_{j=0}^{\bar{k}-1} P_e(r, (1+j\alpha)\gamma_0) + (1+\bar{k}\alpha) \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right) \\
&= \tau_0 \left(1 - \alpha - (1+\bar{k}\alpha) \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) + \alpha \sum_{k=0}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) + (1+\bar{k}\alpha) \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right) \\
&= \tau_0 \left(1 - \alpha + \alpha \sum_{k=0}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) \right) = \tau_0 \left(1 + \alpha \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) \right). \tag{8}
\end{aligned}$$

$B \cdot \left(1 - \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right)$. Thus, by Proposition 1, the user throughput becomes:

$$\begin{aligned}
T(\gamma_0, \alpha, r, \bar{k}) &= \frac{B \cdot \left(1 - \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right)}{\tau_0 \left(1 + \alpha \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0) \right)} \\
&= \frac{r \cdot \left(1 - \prod_{j=0}^{\bar{k}} P_e(r, (1+j\alpha)\gamma_0) \right)}{1 + \alpha \sum_{k=1}^{\bar{k}} \prod_{j=0}^{k-1} P_e(r, (1+j\alpha)\gamma_0)}.
\end{aligned}$$

C. Proof of Observation 1

Since $\lim_{\alpha \rightarrow \infty} T(\gamma_0, \alpha, r, \bar{k}) = 0$, by the definition of limit, there exists an $\bar{\alpha}$ such that $T(\gamma_0, \alpha, r, \bar{k}) < T(\gamma_0, 0, r, \bar{k})$ for all $\alpha \geq \bar{\alpha}$. Because the interval $[0, \bar{\alpha}]$ is a nonempty and compact set and $T(\cdot, \alpha, r, \cdot)$ is continuous on $[0, \bar{\alpha}]$, there should exist at least one optimal solution $\alpha^* \in [0, \bar{\alpha}]$ by Weierstrass' Theorem [17] and the inequality $T(\gamma_0, \alpha^*, r, \bar{k}) \geq T(\gamma_0, 0, r, \bar{k})$ is valid of course. By the definition of $\bar{\alpha}$, $T(\gamma_0, \alpha^*, r, \bar{k})$ is also larger than $T(\gamma_0, \alpha, r, \bar{k})$ for all $\alpha \geq \bar{\alpha}$ and becomes the maximum throughput in the interval $[0, \infty)$.

D. Proof of Theorem 2

Because the denominator of (5) is always greater than 1, we can easily see that

$$\bar{r} \geq T_{\text{sup}}(\gamma_0, \infty) \geq \frac{r}{1 + \alpha \sum_{k=1}^{\infty} (P_0)^k} = \frac{r}{1 + \alpha \frac{P_0}{1-P_0}} \tag{9}$$

for all $\alpha \in (0, \infty)$ and $r \in [0, \bar{r}]$. Since $P_0/(1-P_0)$ is finite, $T_{\text{sup}}(\gamma_0, \infty)$ is lower-bounded by $\lim_{\alpha \rightarrow 0} \frac{r}{1 + \alpha \frac{P_0}{1-P_0}} = r$. Then, the definition of supremum [18] implies (6).

REFERENCES

- [1] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Feasibility study for evolved UTRA and UTRAN (Release 7), 3GPP Std. Tech. Rep. 25.912, June 2006.
- [2] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, IEEE Std. 802.16e, 2005.
- [3] D. Chase, "Code combining—a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. 33, pp. 385–393, May 1985.
- [4] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo codes," *IEEE Trans. Commun.*, vol. 44, pp. 1261–1271, Oct. 1996.
- [5] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.
- [6] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, pp. 1311–1321, Aug. 2004.
- [7] Y. Kim, J. k. Han, H. Kwon, and D. Kim, "Performance of HARQ and the effect of imperfect power control in CDMA reverse link," in *Proc. IEEE Globecom*, Dallas, TX, Nov. 2004, pp. 3017–3021.
- [8] J. Gu, S. Nie, D. Wang, and X. Che, "Optimal HARQ procedures for reverse link," in *Proc. IEEE PIMRC*, Barcelona, Spain, Sept. 2004, pp. 875–879.
- [9] H. Zheng and H. Viswanathan, "Optimizing the ARQ performance in downlink packet data systems with scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 495–506, Mar. 2005.
- [10] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.
- [11] K. Kumaran and L. Qian, "Uplink scheduling in CDMA packet-data systems," in *Proc. IEEE Infocom*, San Francisco, CA, Mar. 2003, pp. 292–300.
- [12] M. Zorzi and R. R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," *IEEE Trans. Commun.*, vol. 44, pp. 1077–1081, Sept. 1996.
- [13] F. Frederiksen and T. E. Kolding, "Performance and modeling of WCDMA/HSDPA transmission/H-ARQ schemes," in *Proc. IEEE VTC Fall*, Vancouver, Canada, Sept. 2002, pp. 472–476.
- [14] T. E. Kolding, F. Frederiksen, and P. E. Mogensen, "Performance aspects of WCDMA systems with high speed downlink packet access (HSDPA)," in *Proc. IEEE VTC Fall*, Vancouver, Canada, Sept. 2002, pp. 477–481.
- [15] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *Proc. IEEE VTC Fall*, Atlantic City, NJ, Sept. 2001, pp. 1829–1833.
- [16] Results of L3NQS Simulation Study, 3GPP2 TSG-C WG5 Contribution C50-20 010 820-011, Oct. 2001.
- [17] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 2nd ed. John Wiley & Sons Inc., 1993.
- [18] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.