On the Scalability of H-ARQ Systems in Wireless Multicast

Stefania Sesia
Eurecom Institute,
Sophia Antipolis, France.
stefania.sesia@eurecom.fr

Giuseppe Caire
Eurecom Institute,
Sophia Antipolis, France.
giuseppe.caire@eurecom.fr

Guillaume Vivier
Centre de Recherche Motorola,
Gif sur Yvette, France.
guillaume.vivier@crm.mot.com

We consider a wireless downlink scenario with one transmitter (base-station) that has to reliably deliver the same information to several receivers (user terminals). The channel is block-fading Gaussian. Transmission is slotted, every slot spans L complex dimensions and the channel fading coefficients for all users are i.i.d., constant on each slot. Since the users are completely symmetric and information is the same for all, the optimal delay-unconstrained coding transmission strategy is trivially given by coding at rate as close as desired to the ergodic capacity of the channel (the same for all users) $C(\gamma)$, where γ is the average received SNR. With a FEC based system, in order to achieve vanishing error probability, every codeword must span an arbitrarily large number of fading blocks, even if we back-off in rate and transmit at spectral efficiency $\eta = (1-\delta)C(\gamma)$, with δ being a fixed gap from capacity. In the single user case, zero error probability can be achieved for finite average delay for a fixed gap $\delta > 0$ by using a Hybrid ARQ protocol based on Incremental Redundancy (denoted by IR in the following) [1].

We consider the same HARQ protocol as in [1, 2] in the case of N users and we compute the throughput η from the base station point of view. As we expected, the average delay necessary to achieve any desired throughput tends to infinity as N increases, meaning that strictly speaking, the IR protocol is not scalable. In order to make the IR scheme scalable, i.e. the delay tends to a finite limit as $N \to \infty$, for any target throughput $\eta \in (0, C(\gamma))$, we have to allow for a fixed fraction $x \in (0,1)$ of users that do not decode successfully. We shall refer to these users as *unfulfilled*. In this case, the transmitter stops sending the current codeword and move to the next codeword if the number of unfulfilled users is not larger than xN. The general expression of the throughput η is a function of N, γ, x, of the ratio $R = b/L$, where b is the number of information bits per codeword and L is the number of dimensions per slot, and of $p(m)$ defined as the probability that a user u can not correctly decode the codeword up to slot m, as in [1]. It follows that

$$\eta(N, x, R, \gamma) = \frac{R}{1 + \sum_{m=1}^{\infty} \sum_{k=0}^{N-\lceil xN \rceil} \binom{N}{k} (1-p(m))^k \, p(m)^{N-k}}.$$

Notice that the probabilities $p(m)$ depend on both R and γ.

We show that for $x = 0$ and any finite R, $\lim_{N\to\infty} \eta = 0$, implying that the average delay goes to infinity. We study numerically the behavior of the average delay versus N for given target throughput $\eta = (1-\delta)C(\gamma)$, when we let R be a design parameter to be optimized. It is interesting to notice that the delay necessary to achieve throughput $\eta = (1-\delta)C(\gamma)$, increases quite slowly if δ is not too small. Hence, we argue that for typical values of N in a cellular system, and for target throughputs not too close to the ergodic capacity, the IR scheme is a viable solution for reliable multicast. In order to show that the IR is scalable in a strict sense, we study the limit $\eta_\infty(x, R, \gamma) = \lim_{N\to\infty} \eta$ for $x > 0$. This is given by

$$\eta_\infty(x, R, \gamma) = \frac{R}{1 + \sum_{m=1}^{\infty} 1\{x \le p(m)\} - \frac{1}{2}\delta(x - p(m))}$$

This limit coincides with the spectral efficiency of FEC coding over a number of slots equal to the delay of the IR system (that becomes a deterministic quantity for large number of users) and with error probability precisely equal to x. Hence, for $x > 0$ and $N \to \infty$ the IR scheme has the same performance (in terms of throughput, delay and error probability) of a FEC coding system. We notice also that in this limit, due to the *large-system hardening*, no explicit feedback channel is needed.

Figure 1 shows the comparison of the throughput maximized over R versus x, in the limit for $N \to \infty$, for $\gamma = 10$dB between IR and a protocol called Selective Repeat that consists on retransmitting the same codeword of length L instead of additional redundancy. We show that for sufficiently small x, the optimal throughput with respect to R, in the limit of large number of users, equals the ergodic capacity and it is obtained when letting $R \to \infty$ (as shown in the figure for $x \in (0, \sim 0.5)$). This extends the result of [1] valid for $x = 0$ and $N = 1$. Moreover, for sufficiently high error probability x the optimal throughput is obtained when the delay is equal to 1 (for $x \in (\sim 0.5, 1)$ in the figure). The proofs of the results can be found in [3].

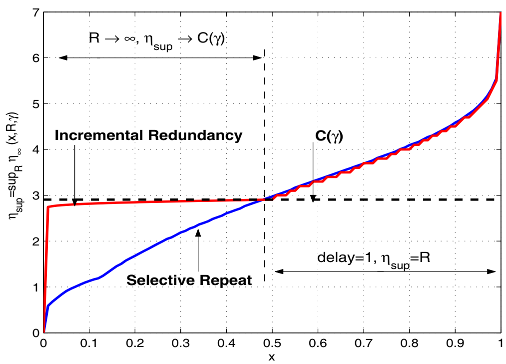

Figure 1: $\sup_{R\ge 0} \eta_\infty(x, R, \gamma)$, vs x for $\gamma = 10$dB.

References

[1] Caire G. and D. Tuninetti, "The Throughput of Hybrid-ARQ Protocols for the Gaussian Collision Channel," *IEEE Trans. on Information Theory*, Vol. 47, pp. 1971–1988, July 2001.

[2] S. Sesia, G. Caire, and G. Vivier, "Incremental Redundancy Hybrid ARQ Schemes Based on Low-Density Parity-Check Codes", to appear in *IEEE Trans. on Communication*, (Revised September 2003).

[3] Sesia S., Caire G. and Vivier G., "On the Scalability of Feedback Systems in Wireless Multicast," in preparation.