

Reducing Retransmissions in Point-to-point Satellite Transmissions

Jérôme Lacan and Tanguy Pérennou

Univ. Toulouse, ISAE, 10 av. Edouard Belin, BP 54032, 31055 Toulouse Cedex 4, France

LAAS-CNRS, Univ. Toulouse, 7 av. du Colonel Roche, 31077 Toulouse Cedex 4, France

This paper presents a mechanism that reduces the number of retransmissions needed to ensure reliable unicast transmissions over a shared broadcast channel. This improves the bandwidth occupancy by information packets and potentially reduces the end-to-end jitter and delay. After each block transmission, the mechanism determines the minimal set of information packets to retransmit, builds a specific Reed-Solomon code from this set and then sends repair packets produced with this code instead of retransmitting the information packets. The mechanism was analysed theoretically and evaluated numerically under many different conditions (variable packet loss rate, number of receivers, etc.). Both the erasure channel and the error channel were explored. In most cases, the system can be set up such that it uses less retransmissions than a classical ARQ mechanism: most of the time, 40% up to 95% less. Such a mechanism is clearly useful in satellite contexts.

I. Introduction

This paper deals with reliable unicast transmissions using a satellite communication system. Satellite communication systems mainly use the radio channel to send information packets to terminals or to exchange other control or management packets. As many other wireless systems (2.4GHz and 5GHz Wireless LANs, WiMax, cellular phone systems, etc.), this broadcast channel can have a high bit error rate (BER) that leads to a lot of packet losses as compared to wired communication systems. In addition, they exhibit a comparatively low data rate. Improving the goodput of such systems is therefore a useful task.

A first solution consists in reducing the BER of the channel using improved antennas and signal processing. Additionally, the reliability of satellite transmissions is classically ensured by pro-active mechanisms such as error correcting codes at the physical layer. Indeed, contrary to Wireless LANs, reactive mechanisms such as the Automatic Repeat reQuest (ARQ) are generally not implemented at the link layer because the large round-trip time on the satellite link can lead to unacceptable delays (especially with a geostationary satellite) when several consecutive retransmissions are needed.

If ARQ does not seem appropriate to improve the satellite network performance, error correcting codes or erasure codes (also called Forward Error Correction (FEC)) provide a well-known alternative to retransmissions by generating a set of repair packets from a set of information packets. Both information and repair packets are sent on the channel, and if some information packets are lost or received with bit errors, they can be recovered using the received repair packets. In some cases, the joint use of retransmission techniques and error correcting techniques, called Hybrid-ARQ (H-ARQ), allows to combine the advantages of these two techniques (see *e.g.* Ref. 1).

In this paper we propose to use a generalization of H-ARQ techniques at the data link level, in such a way that each packet retransmitted on the channel can recover other lost or erroneous packets in several concurrent point-to-point transmissions. We first consider a *packet erasure channel* in which all packets containing errors are erased and do not reach the receiver: such packets lost in concurrent transmissions are considered as information packets in an erasure code. Each retransmitted packet is then a repair packet generated from the lost information packets. We then generalize the proposed mechanism to an *error channel* in which packets containing errors do reach the receiver and are considered as information packets in an error correcting code. Similarly, each retransmitted packet is a repair packet generated from the erroneous information packets. The paper will demonstrate theoretically and numerically that the amount



of retransmissions dramatically decreases as compared to the classical ARQ approach.

The paper first reviews related work (Section II). The principle of our proposition is then detailed (Section III). Potential gains are then evaluated in a theoretical manner (Section IV) and numerical results are presented and discussed (Section V). Then Section VI shows how the proposed mechanisms can be applied to several satellite communication contexts. Finally, Section VII concludes and presents future work in this direction.

II. Related Work

In the general context of this paper, several users share a broadcast channel for point-to-point transmissions. We assume that the channel can only broadcast a packet at a time and that transmissions are temporally multiplexed at the packet level. This type of channel is used in wireless local or metropolitan area networks (Bluetooth, 802.11, WiMAX), mobile networks (GSM, GPRS, UMTS) or satellite networks.

The above types of network rely on the radio resource, which is particularly characterized by a low transmission rate and a high bit error rate (BER) as compared to wired networks. Error control mechanisms are used at the physical level and allow the correction of a part of bit errors.

Other mechanisms allow the correction of the remaining errors so as to ensure end-to-end reliability. In the case of the Transmission Control Protocol (TCP)² for instance, information packets that are lost are not acknowledged by the receiver and several successive retransmissions are attempted upon expiration of a timer. As TCP was designed to handle congestions for wired networks where router congestions are the principal cause of packet losses, it is not optimal when packet losses are due to bit errors. TCP's additive increase multiplicative decrease (AIMD) congestion control reacts to a packet loss by both a decrease of the sending rate and a doubling of the retransmission time-out. If the loss is due to a bit error, those actions are useless and simply decrease bandwidth occupancy although there is no congestion in the network.³⁻⁵ The same problem exists for more other transport protocols using congestion control like TFRC,⁶ DCCP⁷ or SCTP.⁸ Several propositions have been made to implement end-to-end loss differentiation algorithms and improve end-to-end performance.⁹

In general wireless networks, retransmission mechanisms are used at the data link layer to recover as much as possible packet losses due to bit errors. In the 802.11 family MAC layer¹⁰ for example, each data frame may be retransmitted several times until it is acknowledged by the receiver; if several retransmissions in a row fail, the auto-rate fallback mechanism is activated and switches to a modulation that is more robust to error bits (at the expense of a lower transmission rate). Retransmissions are also used in the Radio Link Control Protocol (RLC) layer of UMTS in its acknowledgement mode¹¹ as well as in recent propositions for satellite link layers.¹²

The use of retransmissions at the lower layers however causes a number of problems. When the number of retransmissions gets too high, the useful bandwidth occupancy decreases while end-to-end delays increase. When those delays increase too much, higher level error recovery mechanisms such as those cited above for TCP are automatically unleashed, and the overall network performance unavoidably decreases.¹³

On the erasure channel, a well-known alternate solution is the use of an erasure code which allows the correction of a certain amount of losses in a set of information packets by using repair packets. For instance, in the Maximum Distance Separation (MDS) family of erasure codes (including Reed-Solomon codes), the MDS(k, n) allows the computation of $n - k$ repair packets for a block of k initial information packets. As soon as the receiver receives any k packets among the $k + (n - k)$ packets, it will be able to decode the original k information packets.¹⁴ Other families of erasure codes were defined. The most famous are Low-Density Parity-Check (LDPC) codes and Raptor codes which have very fast coding and decoding algorithms but which are not optimal in term of correction capability.

Most of these erasure codes (except Raptor codes) are derived from classical error correcting codes. The most famous are turbo codes, LDPC codes or algebraic codes like Reed-Solomon or BCH codes.

Erasure and error correcting codes can be used in a proactive way (Forward Error Correction or FEC) or in a reactive way (Hybrid ARQ). In the reactive way, the sender generates and sends redundancy following the feedbacks of the receiver.

Error correcting codes are generally used in proactive way, however, some of the most recent wireless standards such 3GPP HSDPA¹⁵ or "Mobile Wimax" 802.16e¹⁶ define a framework to implement hybrid ARQ schemes. Erasure codes are mainly used for real-time transmissions and multicast/broadcast transmissions (see *e.g.* Ref. 17). In both applications, when a return channel is available, the level of redundancy can be

adapted to the channel conditions so as to improve the global performance.

Recently, the concept of opportunistic coding was introduced to improve practical network coding performance.¹⁸ The idea is to use some information on the data needed by the neighbors (*i.e.* the receivers) to determine how to combine the information packets in order to generate "optimal" redundant packets. In Ref. 19, this concept was applied to packet losses recovery in the context of unicast and multicast transmissions on a broadcast channel. The main result is the construction of an "optimal" repair packet with XOR operations.

Our proposition can be considered as a mix between opportunistic and hybrid ARQ techniques. To recover losses or errors of various unicast transmission on a broadcast channel, we first opportunistically determine the minimal set of needed information packets. Then, we generate the repair packets with classical but efficient erasure/error correcting codes. This proposal allows a dramatical reduction of the number of retransmissions compared to classical ARQ mechanism.

III. Applying Forward Error Correction Codes to Retransmissions

A. Principle

The mechanism proposed in this paper essentially aims at reducing the number of retransmissions at the data link level. This will allow both an increase of the proportion of *information* packets broadcasted on the channel, thus improving bandwidth occupancy, and as a side effect a decrease of the end-to-end delay and jitter.

We make the basic assumption that transmissions are all issued from a unique sender. This assumption corresponds for instance to satellite transmissions or transmissions relayed by an access point in a cellular network. We also make the assumption that the access to the broadcast channel is shared. The kind of multiplexing is generally dynamic, as in the Carrier Sense Multiple Access (CSMA) found in Ethernet and 802.11, or in the Demand-Assigned Multiple Access (DAMA) used in DVB-RCS return link of satellite communications.

We adapt the Hybrid ARQ type II mechanism¹ to point-to-point transmissions by coupling the filtering function and the retransmission mechanism at the data link level.

The proposed mechanism is based on the fact that on a broadcast channel each receiver receives at the physical level all the broadcasted packets (promiscuous reception). In our proposition, instead of filtering out packets according to their destination address, each receiver temporarily archives all information packets received at the physical level. Then each receiver acknowledges only the packets addressed to him. Then the sender computes the list of non acknowledged packets and builds repair packets based on that list. The list of unacknowledged packet IDs is broadcasted to all receivers, then the sender broadcasts redundancy packets until all receivers acknowledge all the previously lost or rejected information packets. A receiver can acknowledge such a packet as soon as he has received enough redundancy: in the case of a packet erasure channel, he must have received at least as many repair *packets* as lost packets in the list, and for an error channel, he must have received enough repair *symbols* to correct erroneous information symbols in the list.

B. Illustration

Figure 1 illustrates the proposition for an erasure channel. First, sender S , typically a satellite, relays information packets to receivers R_1 and R_2 using a simple interleaving scheme. Each receiver should receive four information packets for himself. We note $\text{info}(i, j)$ the i^{th} information packet sent to the j^{th} receiver.

R_1 does not receive his first information packet while R_2 does not receive his second and third information packets. Both receivers acknowledge all correctly received packets that were addressed to them to S . S can then establish that the list of unacknowledged information packets IDs is $L = \{(1, 1); (2, 2); (3, 2)\}$. S broadcasts the L list to R_1 and R_2 , then builds a Reed-Solomon code based on the information packets of this list (Reed-Solomon codes are not the only option, especially when this mechanism is adapted to an error correcting channel).

S then sends repair packets made with the previously built code until all receivers have acknowledged their four packets. A first repair packet – $\text{repair}(1)$ – is broadcasted to both receivers, which receive it without error. Now R_1 has buffered packets $\text{info}(2,2)$, $\text{info}(3,2)$ and $\text{repair}(1)$, so he can decode the L list and recover missing packet $\text{info}(1,1)$. R_1 then acknowledges its block to S . R_2 , however, misses information packets $\text{info}(2,2)$ and $\text{info}(3,2)$ from L and even with $\text{repair}(1)$ he still does not have enough redundancy to

decode L . S has not received the block acknowledgement from R_2 and therefore sends repair(2), which is lost by R_2 . S still does not receive any block acknowledgement from R_2 and sends repair(3) which is correctly received by R_2 . With info(1,1), repair(1) and repair(3), R_2 decodes the L list and recovers missing packets info(2,2) and info(3,2), and acknowledges its block to S .

In this example, only 3 retransmissions were needed for both users to get their information packets. With a classical ARQ mechanism, 4 retransmissions would have been needed (the second retransmitted packet is lost and needs additional retransmission).

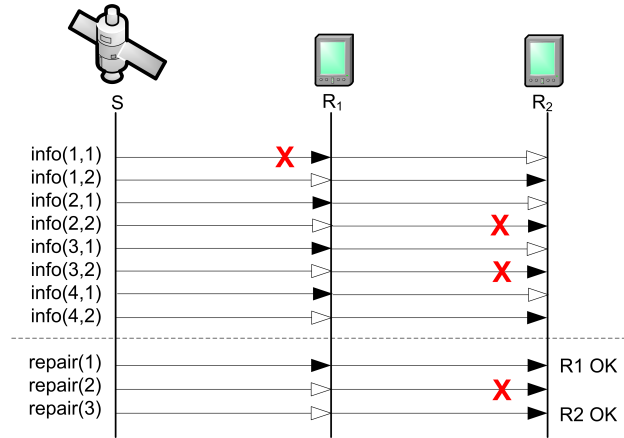


Figure 1. Our modified Hybrid ARQ type II mechanism.

C. Sender and Receiver Algorithms

We consider an erasure channel and we assume a static temporal multiplexing: if R receivers simultaneously receive their own N packets at the data link level, and if we note info(i, j) the transmission of the i^{th} information packet addressed to the j^{th} receiver on the broadcast channel, then the order of the interleaved information packets on the channel will be info(1,1), ..., info(1, R), info(2,1), ..., info(N, R).

The sender's algorithm is basically as follows:

1. send a block made of $N \times R$ interleaved information packets (B_n);
2. determine the complete list of packets lost by their addressee in this block (L_n);
3. consider the packets in L_n as the k information packets of a Reed-Solomon (RS) code and build code RS_n ;
4. broadcast the list of the L_n packet IDs;
5. broadcast repair packets produced by the RS_n code until all receivers have sent an acknowledgement for the full block B_n .

Concerning receivers, each one promiscuously listens to the broadcast channel and uses the following algorithm:

1. temporarily archive all packets received in block B_n including the packets not explicitly addressed to him;
2. acknowledge each received packet that is addressed to him;
3. receive the list of packet IDs forming the new RS_n code;
4. receive and archive the successive repair packets of the RS_n code;
5. as soon as there are enough repair packets to decode L_n , recover lost packets addressed to him from the decoded L_n and acknowledge block B_n .

For the error channel, the sender's algorithm is the same. On the receiver side, each receiver acknowledges only the error-free packets addressed to him. He stores all received packets, erroneous or not, so as to exploit as much valid symbols as possible. Upon reception of each repair packet, the receiver tries to decode the erroneous packets of the list L_n with the set of received repair packets. When the decoding is successful, it sends a block acknowledgment to the sender.

In both erasure and error channels, the repair packets are built from a Reed-Solomon (see Figure 2). For the erasure channel, the packets are built following the classical construction¹⁴ by considering the packets as columns of a matrix and by encoding the rows. For the error channel, each packet is virtually split into G groups of b symbols. The m repair packets are also made of G groups of repair symbols, such that each "row" made of $(N.R + m) \times b$ symbols is a code-word with $N.R.b$ information symbols and mb repair symbols. Each row can be then decoded independently from the others.

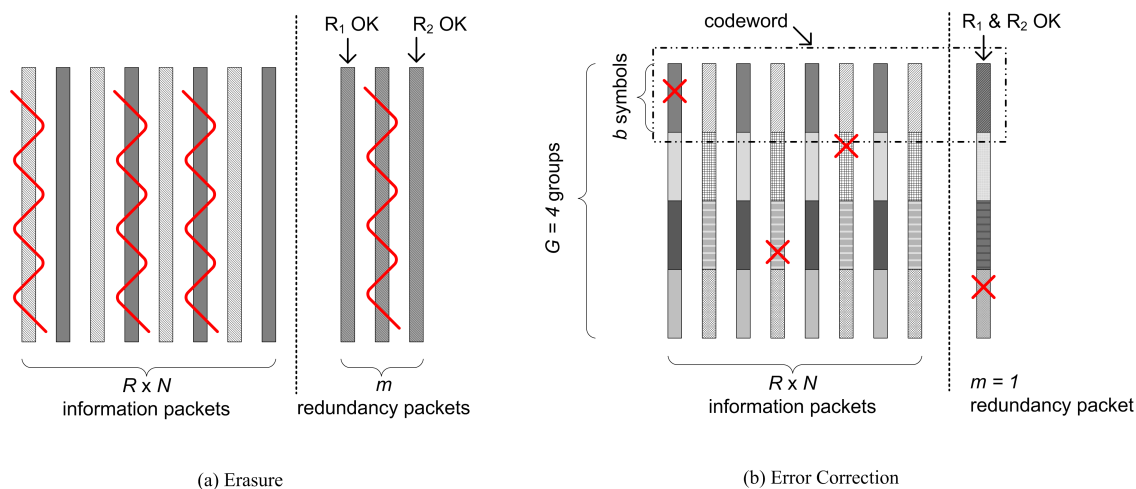


Figure 2. Retransmissions with erasure and error correcting codes

Note that Reed-Solomon codes are used in this algorithm because: (1) the repair symbols generated by a Reed-Solomon code can be used for both erasure and error channels and (2) the properties of these codes simplify the analytical evaluation of the algorithm provided in Section IV. However, the principle of the algorithm remains valid if other erasure codes such *e.g.* LDPC, Raptor, XOR-based codes or other error correcting codes such *e. g.* turbo or LDPC codes are used.

IV. Analytical Evaluation

In this Section we provide an analytical evaluation of the mechanism described in Section III in the case of both the erasure channel and the error channel. We use the following assumptions:

- Homogeneous receivers: all R receivers have the same properties.
- Same number of information packets per receiver per block: each receiver is addressed N packets in each block.
- Fixed-sized packets: all packets have the same size. They are made of G groups of b symbols. These groups may be used to perform independent error correction attempts on different groups.
- Independent erroneous packets: the distribution of erroneous packet does not depend on time nor on receivers (no temporal nor spatial correlation). The probability that a packet contains errors is p and therefore the probability that it does not contain any error is $1 - p$. On an erasure channel, erroneous packets are considered lost. On an error channel, erroneous packets are considered received, but with a subset of erroneous symbols.
- Independent symbol errors: when a packet is erroneous, its errors are uniformly distributed with a probability q (symbol error rate) over the $G.b$ symbols of the packet.

The remainder of this Section evaluates the *overhead* of the proposed mechanism for the erasure channel and the error channel, measured as the amount of retransmissions needed for each receiver to recover its N information packets.

A. ARQ Overhead

In a classical ARQ mechanism¹ each erroneous packet is retransmitted until reception without error. The overhead is simply computed as the mean number of retransmissions for one packet multiplied by the number of packets. This overhead is denoted O_{ARQ} .

Let A be the random variable (r.v.) representing the number of trials needed to successfully transmit a single packet. Under the above uniform independent assumptions, we have $P(A = a) = p^{(a-1)}(1 - p)$ and the amount of trials needed is $E[A] = 1/(1 - p)$. The amount of retransmissions needed for one packet is $1 - E[A]$, and therefore:

$$O_{\text{ARQ}} = \frac{N.R.p}{1 - p} \quad (1)$$

This value will be used as a basis of comparison for the numerical evaluation of our mechanism in Section V.

B. Hybrid ARQ Proposal Overhead

Let M be the r.v. representing the *minimal* amount of repair packets such that all R receivers can decode the N information packets addressed to them. The overhead of the proposed hybrid ARQ mechanism is therefore the expectation of M : $O_{\text{HARQ}} = E[M]$.

Let L be the r.v. representing the size of the list of erroneous packets discussed in Section III. To compute $P(M = m)$, we successively switch to the CDF of M , introduce the possible values of L and apply Bayes' theorem:

$$P(M \leq m) = \sum_{\ell=0}^{R.N} P(M \leq m | L = \ell).P(L = \ell) \quad (2)$$

L follows a binomial law $B(N.R, p)$:

$$P(L = \ell) = \binom{N.R}{\ell} p^{\ell} (1 - p)^{N.R - \ell} \quad (3)$$

Let S be the r.v. representing the amount of repair packets needed by a *single* receiver to decode all its information packets knowing that $L = \ell$. The law of M is the maximum of the $S_j, j = 1, \dots, R$ of each receiver, which translates to the product of CDFs:

$$P(M \leq m) = \sum_{\ell=0}^{R.N} P(S \leq m | L = \ell)^R . P(L = \ell) \quad (4)$$

To determine $P(S \leq m | L = \ell)$ for a given receiver, we introduce the following random variables: let X be the r.v. describing the amount of erroneous information packets that were addressed to him, let Y be the r.v. describing the amount of packets received with errors by this receiver in the L list (including the X erroneous packets), and let Z be r.v. describing the amount of repair packets received with errors by this receiver. Using Bayes again and exploiting independence properties, we can show that:

$$\begin{aligned} P(S \leq m | L = \ell) &= \sum_{x,y,z} P(S \leq m | Y = y \wedge Z = z).P(X = x | L = \ell).P(Y = y | X = x \wedge L = \ell).P(Z = z) \end{aligned} \quad (5)$$

In the above equation, r.v. Z follows a binomial law $B(m, p)$; r.v. Y with $X = x$ and $L = \ell$ fixed can be determined knowing that $Y - x$ follows a binomial law $B(\ell - x, p)$; and r.v. X with $L = \ell$ fixed follows a hypergeometric law $H(\ell, N, N.R - N)$. The three last terms of Eq. 5 are therefore:

$$P(X = x | L = \ell) = \binom{N}{x} \cdot \binom{N.R - N}{\ell - x} / \binom{N.R}{\ell} \quad \forall x, 0 \leq x \leq \min(\ell, N) \quad (6)$$

$$P(Y = y | X = x \wedge L = \ell) = \binom{\ell - x}{y - x} p^{y-x} (1-p)^{\ell-y} \quad \forall y, x \leq y \leq \ell \quad (7)$$

$$P(Z = z) = \binom{m}{z} p^z (1-p)^{m-z} \quad \forall z, 0 \leq z \leq m \quad (8)$$

All other values of x , y and z lead to null terms.

1. Erasure Channel

With a Reed-Solomon code on the erasure channel, the receiver can recover erroneous packets erased by the channel if and only if he has received at least as many repair packets as erroneous packets (information or repair), i.e. $m \geq y + z$. So when y and z are fixed:

$$\begin{aligned} P(S \leq m | Y = y \wedge Z = z) &= 1, \text{ if } m \geq y + z; \\ &= 0, \text{ otherwise.} \end{aligned} \quad (9)$$

Hence for the erasure channel:

$$\begin{aligned} P(S \leq m | L = \ell) &= \sum_{x,y,z} P(X = x | L = \ell) \cdot P(Y = y | X = x \wedge L = \ell) \cdot P(Z = z) \\ &\text{with:} \\ &0 \leq x \leq \min(\ell, N), \\ &x \leq y \leq \ell, \\ &0 \leq z \leq \max(0, m - y). \end{aligned} \quad (10)$$

2. Error Channel

On the error channel, we must consider the symbols in every packet, including erroneous information and repair packets. In Eq. 5, $P(S \leq m | Y = y \wedge Z = z)$ must be determined in the context of a Reed-Solomon error correcting code. The decoding is applied independently on the G code words constituted by the concatenation of all groups of b symbols located at the same offset in all stored information and repair packets. So here the information consists of ℓ groups of b symbols and the redundancy consists of m groups of b symbols. With a Reed-Solomon code, a code word can be decoded if and only if the amount of erroneous symbols is lesser or equal to half the amount of redundancy symbols, so at most $m.b/2$ erroneous symbols can be corrected.

Let D be the r.v. such that $P(D \leq m)$ is the probability for a receiver to decode a code word with m or less repair packets. Under the independence assumptions above, the G code words can be decoded independently from each other, so:

$$P(S \leq m | Y = y \wedge Z = z) = P(D \leq m | Y = y \wedge Z = z)^G \quad (11)$$

Let E be the r.v. describing the total number of erroneous symbols in a single code word. As stated above, the code word can be decoded if and only if $E \leq m.b/2$:

$$P(D \leq m | Y = y \wedge Z = z) = P(E \leq m.b/2 | Y = y \wedge Z = z) \quad (12)$$

$$= \sum_{e=0}^{m.b/2} P(E = e | Y = y \wedge Z = z) \quad (13)$$

Each code word contains $(y + z)b$ symbols being part of erroneous packets. As symbol errors are independent in each erroneous packet, E follows a binomial law $B((y + z)b, q)$:

$$P(E = e | Y = y \wedge Z = z) = \binom{(y+z)b}{e} q^e (1-q)^{(y+z)b-e} \quad (14)$$

Hence for the error channel:

$$P(S \leq m | L = \ell) = \sum_{x,y,z} \left[\sum_{e=0}^{m.b/2} P(E = e | Y = y \wedge Z = z) \right]^G \times P(X = x | L = \ell).P(Y = y | X = x \wedge L = \ell).P(Z = z)$$

with:

$$\begin{aligned} 0 &\leq x \leq \min(\ell, N), \\ x &\leq y \leq \ell, \\ 0 &\leq z \leq m. \end{aligned} \quad (15)$$

V. Numerical Evaluation

In this Section, we compare the overhead of the mechanism described in Section III with a classical ARQ retransmission mechanism. For both the erasure channel and the error channel, the evaluation is carried out using the equations given in Section IV. We define the *gain* of our mechanism over ARQ as the percentage of retransmissions needed by ARQ and not needed by HARQ:

$$\text{gain} = 100 \times \frac{O_{\text{ARQ}} - O_{\text{HARQ}}}{O_{\text{ARQ}}}$$

A positive gain means that the proposed mechanism is better than classical retransmissions. An increasing gain denotes an increasing performance of HARQ, *i.e.* proportionally less and less retransmissions. The theoretical maximum gain of 100% is obtained when $O_{\text{HARQ}} = 0$ and $O_{\text{ARQ}} > 0$. A negative gain means that classical retransmissions are less numerous and behave better than our mechanism.

An extensive campaign of evaluation was made for both channels, varying all parameters: the packet error rate p , the size of symbol groups in a packet G such that $G.b = 256$ symbols, the number of receivers R and the amount of packets addressed to each receiver N . The symbol error rate $q = 0.10$ was fixed.

The best results are systematically obtained for $G = 1$ and $b = 256$. In that configuration, the error channel always allows a better gain than the erasure channel. Figure 3 shows the results for $p = 0.01$ (top figures) and $p = 0.15$ (bottom figures), N varying from 1 to 8 packets per block and R varying from 1 to 64 receivers. Both gains increase with N as well as with R . The error channel reaches a gain of 81% for $p = 0.01$ and of 95% for $p = 0.15$, while the erasure channel reaches a gain of 73% for $p = 0.01$ and of 72% for $p = 0.15$. These maximum values are all reached for the maximum $N.R$ value. The Figure also shows that in most cases better gains are obtained for $p = 0.15$. In the case of the error channel, this is intuitively explained by the fact that there is a lot of erroneous packets but not a lot of erroneous symbols: a single packet containing 256 repair symbols can repair all the erroneous symbols needed by all the receivers.

The error channel does not always outperforms the erasure channel. The gain obtained with the error channel decreases as the number of groups per packet G increases from 1 to 256, and for many (N, R) couples there is a minimum value of G such that $\text{gain}_{\text{errors}} < \text{gain}_{\text{erasure}}$. This is particularly true for $p = 0.01$ where such a minimum G exists for all investigated (N, R) couples, as illustrated in Figure 4. The minimum G value then varies between 32 and 128 (we did not perform an exhaustive campaign of measure and stucked to powers of 2 for N, R and G , so the actual threshold values may be lower than that). Going beyond that G value leads to a dramatic decrease of the error channel performance. In the worst case, *i.e.* $N = 1$ and $R = 2$, setting $G = 256$ on the error channel leads our mechanism to use nearly 5 times more retransmissions than on the erasure channel.

To sum up, both proposed hybrid ARQ type II mechanisms provide potentially high gains on the number of retransmissions. The largest gains obtained correspond to higher values of the packet error rate (*i.e.* 15% rather than 1%) and large values of N and R . When a small amount of group symbols per packet can be chosen, the performance of the error channel is clearly better than the performance of the erasure channel.

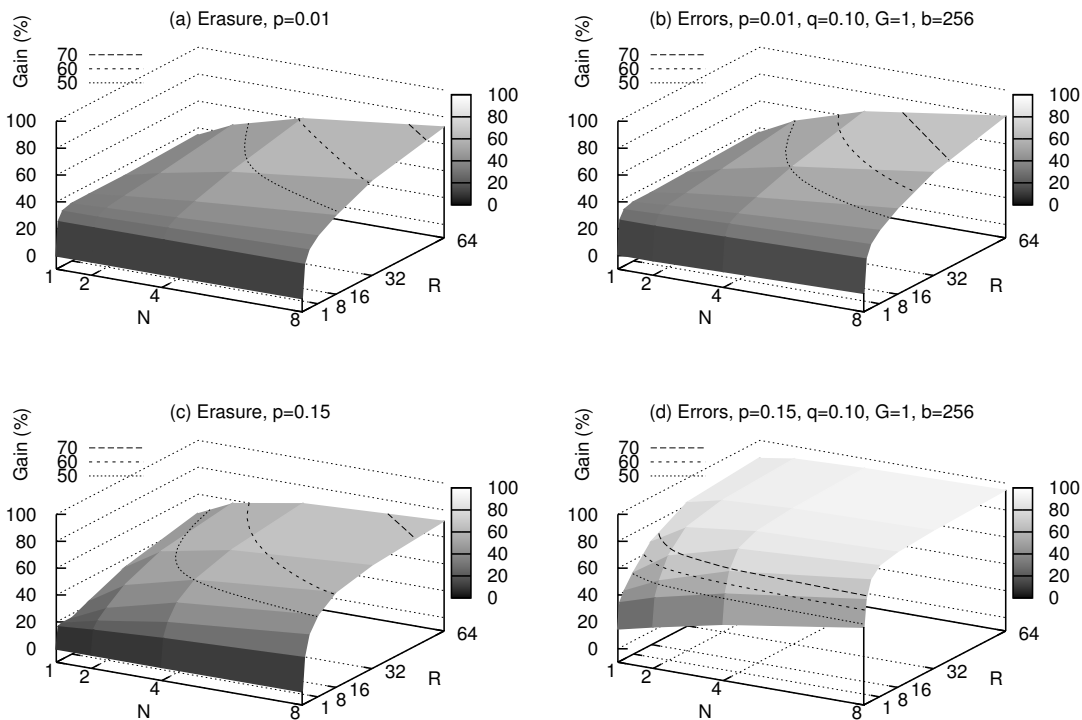


Figure 3. Impact of N and R on the gain

Discussion

A number of constraints related to the mechanism proposed in this paper should be carefully examined. One of these constraints is the storage capacity needed to temporarily store the packets addressed to other receivers. The needed storage capacity is a function of parameters R (number of receivers) and N (number of information packets per receiver and per block), as well as a function of the transmission rate. Yet, the higher $N \times R$, *i.e.* the number of packets per block, the higher the efficiency of the proposed mechanism will be.

Among these three parameters, only N and R may be adjusted by the sender so as to find a good compromise between the obtained gains (see Section V) and the amount of storage capacity. The number of receivers R may be controlled through the constitution of subgroups of jointly managed receivers. For instance on a broadcast channel shared among 10 receivers, 2 subgroups of 5 receivers can be formed, each receiver of a group ignoring the packets sent in the other subgroup. More sophisticated solutions may be elaborated, *e.g.* grouping receivers according to their storage capacity.

The number of information packets per receiver and per block N can be adjusted too. However augmenting N leads to an additional delay: in case of packet loss, the N information packets addressed to a given receiver will be delivered to the upper communication layers only after decoding the block. This can generate delay and jitter on the first information packets and make the use of *e.g.* the TCP protocol at the transport layer rather erratic, because the resulting delays may trigger unwanted retransmissions at the TCP level as well as a decrease of the transmission window¹³ although the lost information packet will eventually be decoded by the data link layer. Yet, considering all N packets addressed to a receiver within a block, the last packet will be delivered faster than without the HARQ mechanism presented here. In practice the N parameter should be adjusted according to the maximum delay accepted by upper layers (especially when using TCP) and the storage capacities of the receivers.

Another potential implementation issue is the size of packets. A number of data link level protocols use a variable packet size, like the 802.11 protocol. The implementation should provide an operation that combines

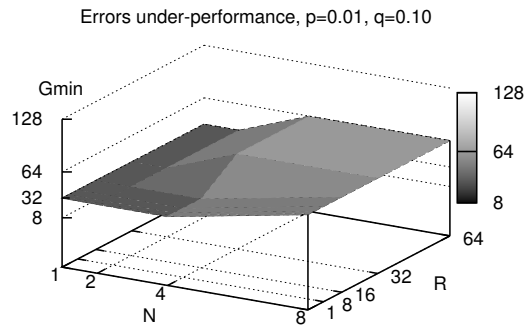


Figure 4. Error Channel under-performance wrt Erasure Channel

two packets of different sizes. In some very specific cases, for instance if a large packet is combined with several small packets, several large repair packets may be retransmitted and thus increase channel occupancy overhead. However this phenomenon can be partially avoided by the constitution of blocks with packets of the same size.

The evaluation of the previous section assumes that packets losses are uniformly distributed over time, and that they are independent for different receivers. This assumption does not cover all kinds of broadcast channels. In particular in 802.11 wireless networks, packet losses are correlated in time and space, and it has been shown that these correlations have a strong impact on the performance of reliable multicast transmissions.²⁰ A classical solution to cope with bursts of losses or errors is to use interleaving techniques. The mechanism proposed in this paper may therefore be evaluated under stronger assumption of spatial and temporal correlation in future work.

For the error channel, the choice of the parameter G is important. As shown on Figure 4, if the code has a limited length (for example, classical octet-based Reed-Solomon implementations do not work for a number of encoding symbol greater than 256), the parameter G should be set as small as possible. If, for the minimal value G , the decoding performances are not acceptable, other error correcting codes supporting a large number of encoding symbols (such *e.g.* LDPC codes) should be chosen.

VI. Applications in Satellite contexts

Several architectures can be considered for satellite communications. In this Section, we consider 3 classes of architecture: bi-directional point-to-point satellite communications, satellite broadcast with terrestrial repeaters, and satellite broadcast without repeaters, but with an ad-hoc network of receivers. For each architecture, we describe the general principle allowing to apply our proposal.

A. Satellite-only

In the satellite-only case, we consider that several bidirectional point-to-point communications share the same satellite link. A classical example is an IP communication using DVB-S or DVB-S2 on the forward link and DVB-RCS or a terrestrial link on the return link. The proposed mechanism can be implemented at the link layer between the gateway and the satellite terminals. It optimizes the link utilization by reducing significantly the number of retransmitted packets.

Last sections have presented the proposed mechanism and have evaluated its performance in terms of the number of retransmissions. This result has a direct consequence on the *bandwidth occupancy*.

The bandwidth management and the transmission delays are probably two of the main problems of satellite communications. Indeed, the bandwidth-over-number of potential users ratio is largely lower than for wired Internet or even for wireless LAN. So, the bandwidth must be carefully managed and all possible mechanisms should be used to cope with this problem. A perfect illustration of this is that the next generation of protocols for satellite communications such as *e.g.* DVB-S2²¹ integrate adaptive mechanisms to switch

between optimal classes of error correcting codes/modulation.

Another consequence of the reduction of the number of retransmissions is the reduction of the average *transmission delay*. Clearly, avoiding one retransmission decreases the delay by at least one Round Trip Time (RTT). This RTT being roughly 500 ms for geostationary satellites, any reduction of the number of retransmissions could be very useful for satellite communications. Note that even if large RTTs make it difficult to implement retransmission strategies on satellite communications, several works have shown that ARQ or hybrid ARQ mechanisms can be successfully used in this context^{22,23}. Moreover, the intrinsic adaptability of HARQ mechanisms has lead WLAN physical layers designers to integrate them in most of recent standards.^{15,16}

In our context, we consider an architecture with several users connected to the Internet through a geostationary satellite. Several bidirectional point-to-point communications simultaneously share the same broadcast channel. We consider that the traffic is asymmetric (most of the traffic is sent from the Internet toward the users) with a high rate forward link between the gateway and the user terminals and a low rate return link between the user terminal and the gateway. Note that the return link can also be terrestrial.

In this context, the proposed mechanism can be implemented as described in Section III between the gateway and the user terminals, the gateway acting as the sender.

This proposition applies to many broadcast contexts and some adaptations must be done in order to integrate the satellite-only context. The first point is to determine the layer in which the mechanism can be implemented. At the physical layer, it seems better to consider an error channel where the redundant data are generated with an error correcting code. It can be noted that in the classical mode, a terminal decodes all the sent data (including the data for the other terminals) and then filters them in order to only keep its own data. The modification needed for our mechanism is simply to store temporarily the decoded data to use them for decoding. Thus, it does not imply additional decoding operations upon reception. The implementation of an hybrid-ARQ mechanism was studied²³ in order to use HSDPA in satellite communications. From this framework, very few modifications are needed to implement our proposal in this context.

To study the implementation of our proposal in DVB-S2,²¹ we can observe that the FECFRAMEs, *i.e.* the codewords, are generated from information data belonging to different flows. Thus, the implementation of our mechanism can be done simply by implementing classical HARQ type II in DVB-S2. This can be simply done by adding a unique identifier to each FECFRAME header and by allowing some FECFRAME to only contain redundancy which is built from information data sent in a previous FECFRAME.

When implemented at the link layer, the mechanism will be preferably used in erasure mode. The redundant data will then be generated with an erasure code. In a practical context where the link layer is MPE, ULE or GSE, an unique identifier must be added to the data unit headers. Note that this is also a requirement for the implementation of a secure version of these protocols.²⁴ Under this assumption, the generation of repair data unit should be easily implemented since it was already standardized with MPE-FEC²⁵ or at least proposed.²⁶

B. Satellite and terrestrial repeaters

Present or future architectures for TV or radio broadcast from satellite to mobile receivers include terrestrial repeaters (or "gap fillers") that ensure a good reception of the signal in urban areas. Examples of such architectures are MBSAT in South Korea and Japan, Sirius and XM systems in North America or DVB-SH²⁷ in Europe.

In most of these systems, both satellite and terrestrial links are unidirectional, *i.e.* there is no return link from the mobile receivers to the satellite or the repeater.

However, with the convergence of technologies (*e.g.* with 3G technologies), bi-directional links between mobiles and repeaters will probably be available in next-generation architectures.

Under this assumption, we propose the following application of our mechanism: the sender role is split between the satellite and the repeaters.

- the satellite broadcasts the data (toward the mobiles and the repeaters)
- the mobiles send their list of lost data to the repeater
- the repeater aggregates the lists, then generates and sends redundant data
- the mobiles recover their lost data and acknowledge the final recovery

The main interest of this scheme is to improve the role of the repeaters in the system. Indeed, in a classical scheme, the repeaters simply broadcasts the data without taken into account the properties of the each mobile. In the propose mode, they can concentrate its efforts to help specific set of mobiles. In addition, the delay between the repeater and a terminal if much smaller than with the satellite, so the overall delay is greatly decreased, thus reducing the impact of retransmissions on upper layers such as TCP.

If we consider the recent standard DVB-SH,²⁷ it seems easier to implement our mechanism at the link layer instead of the physical layer. Indeed, at the link layer, the standard describes MPE-IFEC, a framework allowing to implement erasure codes. Thus, with the addition of an unique identifier to the MPE data unit, our mechanism should be implemented without any strong modification.

C. Satellite and retransmissions done by peers

In more future architectures, packets lost on the satellite link may be recovered by an ad-hoc network composed of the satellite receivers. In this context, each receiver broadcasts the list of its lost packets and the receivers with good signal reception generate and transmit redundant packets built from these lists of lost packets. This ad-hoc recovery system is quite similar to the network coding COPE system introduced in.²⁸ It has the advantage of not requiring any repeater and not having the satellite or gateway manage retransmissions. However such a system should be carefully designed to take into account isolated terminals as well as ad-hoc routing delays.

VII. Conclusion

In this paper we have presented a hybrid ARQ type II mechanism that opportunisticly builds a specific Reed-Solomon code based on a minimal set of information packets for each R -receiver transmission block, and produces repair packets using that code. An analytical evaluation of the mean number of retransmissions has been carried out as well as numerical evaluations, that showed that the mechanisms performs very well compared to classical ARQ retransmissions. The mechanism may allow up to a 95% decrease of the amount of retransmitted packets in the case of an error channel. This in turn increases the overall goodput, while the end-to-end delay and jitter are decreased.

This mechanism may be used on any erasure or error broadcast channel, and we have illustrated how to use it in a variety of satellite communications contexts. The most promising one is a mixed satellite/repeater context such as DVB-SH, where the management of the Reed-Solomon code and the transmission or repair packets is done by the repeater, thus even more diminishing the problem of additional end-to-end delays due to retransmissions.

Future work in that direction may include the use of weaker assumptions in our analysis, in particular the introduction of loss correlations in time (*e.g.* a Gilbert-Eliott error model) or among groups of receivers.

References

- ¹Lin, S. and Costello, D., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
- ²Postel, J., "Transmission Control Protocol," *Request For Comments 793*, IETF, Sept. 1981.
- ³Caceres, R. and Iftode, L., "Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments," *IEEE Journal of Selected Areas in Communications*, Vol. 13, No. 5, 1995, pp. 850–857.
- ⁴Balakrishnan, H., Padmanabhan, V. N., Seshan, S., and Katz, R. H., "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, 1997, pp. 756–769.
- ⁵Holland, G. and Vaidya, N. H., "Analysis of TCP Performance Over Mobile Ad Hoc Networks," *Proceedings of IEEE/ACM MOBICOM '99*, August 1999.
- ⁶Floyd, S., Handley, M., Padhye, J., and Widmer, J., "Equation-based congestion control for unicast applications," *SIGCOMM 2000*, Stockholm, Sweden, August 2000, pp. 43–56.
- ⁷Kohler, E., Handley, M., and Floyd, S., "Designing DCCP: congestion control without reliability," *SIGCOMM Comput. Commun. Rev.*, Vol. 36, No. 4, 2006, pp. 27–38.
- ⁸Stewart, R. R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H. J., Taylor, T., Rytina, I., Kalla, M., Zhang, L., and Paxson, V., "Stream Control Transmission Protocol," *Request For Comments 2960*, IETF, Oct. 2000.
- ⁹Cen, S., Cosman, P. C., and Voelker, G. M., "End-to-end differentiation of congestion and wireless losses," *IEEE/ACM Trans. Netw.*, Vol. 11, No. 5, 2003, pp. 703–717.
- ¹⁰ANSI/IEEE Std 802.11, "Wireless LAN MAC and PHY Specifications," 1999.
- ¹¹3GPP, "Radio Link Control (RLC) Protocol Specification , 3G TS RAN 25.322," 1999.

¹²Kojo, M., Astuti, D., Daniel, L., Nyrhinen, A., and Raatikainen, K., "Improving TCP Performance over Wireless WANs using TCP/IP-Friendly Link Layer," *1st International Conference on E-Business and Telecommunication Networks (ICETE)*, August 2004.

¹³Fairhurst, G. and Hood, L., "Advice to Link Designers on Link Automatic Repeat reQuest (ARQ)," *Request For Comments 3366*, IETF, August 2002.

¹⁴Rizzo, L., "Effective Erasure Codes For Reliable Computer Communication Protocols," *ACM Computer Communication Review*, Vol. 27, No. 2, April 1997, pp. 24–36.

¹⁵3GPP, "Technical Specification Group Radio Access Network; High Speed Downlink Packet Access (HSDPA), 3GPP TS 25.308-V8.0.0," 2007.

¹⁶"IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1," *IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004)*, 2006, pp. 1–822.

¹⁷Luby, M., Vicisano, L., Gemmell, J., Rizzo, L., Handley, M., and Crowcroft, J., "The Use of Forward Error Correction (FEC) in Reliable Multicast," *Reliable Multicast Transport (rmt) Internet group, RFC 3453*, December 2002.

¹⁸Katti, S., Rahul, H., Hu, W., Katabi, D., Médard, M., and Crowcroft, J., "XORs in the air: practical wireless network coding," *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, 2006, pp. 243–254.

¹⁹Rozner, E., Iyer, A. P., Mehta, Y., Qiu, L., and Jafry, M., "ER: Efficient Retransmission Scheme for Wireless LANs," *International Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, December 2007.

²⁰Lacan, J. and Pérennou, T., "Evaluation of Error Control Mechanisms for 802.11b Multicast Transmissions," *Second International Workshop On Wireless Network Measurement (WinMee 2006)*, April 2006.

²¹ETSI, "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2), EN 302 307," 2004.

²²Lin, S. and Yu, P., "A Hybrid ARQ Scheme with Parity Retransmission for Error Control of Satellite Channels," *Communications, IEEE Transactions on [legacy, pre - 1988]*, Vol. 30, No. 7, Jul 1982, pp. 1701–1719.

²³Azizan, A., Quddus, A. U., and Evans, B. G., "Link Level Performance Analysis of Satellite High Speed Downlink Packet Access (S-HSDPA)," *24th AIAA International Communications Satellite Systems Conference (ICSSC), San Diego*, 2006.

²⁴Michael Noisternig and Bernhard Collini-Nocker, "Lightweight security extension for the Unidirectional Lightweight Encapsulation (ULE) protocol," IETF draft, draft-noisternig-ipdvb-ulesec-00, work in progress, 2008.

²⁵ETSI, "Specifications for Data Broadcasting, EN 301 192," .

²⁶Lei, J., Granados, G. S., and Castro, M. V., "MPE/ULE-FEC vs GSE-FEC Efficiency Comparison of IP Datagram Transmission over DVB-S2," *25th AIAA Int. Comm. Satellite Systems Conf. (ICSSC 2007)*, Seoul, Korea, 2007.

²⁷ETSI, "Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz (draft EN 302 583 V1.1.1)," .

²⁸Katti, S., Rahul, H., Hu, W., Katabi, D., Médard, M., and Crowcroft, J., "XORs in the air: practical wireless network coding," *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, 2006, pp. 243–254.