QoS-Based Adaptive Error Control for Wireless Networks

Neda Nikaein, Christian Bonnet
Institut Eurécom
2229, Route des Crêtes
B.P. 193
06904 Sophia Antipolis, France
{neda.nikaein, christian.bonnet}@eurecom.fr

Abstract

Wireless channels are highly affected by unpredictable factors such as cochannel interference, adjacent channel interference, propagation path loss, shadowing and multipath fading. The unreliability of media degrades the transmission quality seriously. Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC) schemes are frequently used in wireless environments to reduce the high bit error rate of the channel. In this paper, we propose an adaptive error control scheme for wireless networks based on dynamic variation of error control strategy as a function of the channel bit error rate, desired QoS and number of receivers. Reed-Solomon erasure codes are used throughout this study because of their appropriate characteristics in terms of powerful coding and implementation simplicity. Simulation results show that our adaptive error control protocol decreases the waste of bandwidth due to retransmissions or extra coding overheads while satisfying the QoS requirements of the receivers.

Keywords: QoS, adaptive error control, ARQ, FEC, Markov model, wireless networks.

Recently, the emergence of new multimedia applications has created a strong need for the support of *Quality of Service* (QoS). In response, the Internet is moving from a best effort model to a system, capable of supporting a range of traffic characteristics and service requirements. The main obstacle in order to enable users to have access to Internet and multimedia applications in wireless environments is the high error rate of the wireless channels. In fact, wireless channels are highly affected by unpredictable factors such as cochannel interference, adjacent channel interference, propagation path loss, shadowing and multipath fading. As a result, most of the wireless systems are equipped with a complementary error control protocol at the link layer.

Basically, there are two main error recovery mechanisms: *Automatic Repeat Request* (ARQ) and *Forward Error Correction* (FEC). ARQ tries to retransmit the lost packets while

FEC transmits some redundant data with the original ones. FEC is frequently used in wireless environments but it can not assure full reliability unless coupled with ARQ. The scheme combining ARQ with FEC is normally called *Hybrid ARQ*. In this paper, we propose an adaptive protocol capable of choosing either an ARQ or a hybrid ARQ error control protocol as a function of the QoS requirements of receivers as well as the wireless channel conditions.

We take a multicast communication mode. If we consider multicast communication as a general communication mode where the traffic is sent to a set of receivers, unicast and broadcast communications can be viewed as special cases where the traffic is only sent to one receiver or to all receivers respectively. Having a framework for QoS provisioning in the case of multicast communications means that the same general rule can be applied to any communication mode. Note that we suppose a single QoS per multicast session. It means that all members of a given group are supposed to have the same QoS requirements. Other approaches like layered multicast [1] may be used in the case of receivers with different QoS needs but this is not the subject of this study.

[2] showed that the use of hybrid ARQ protocol improves the performance of error control schemes for wireless links in most cases. However, the choice of the coding scheme depends on several parameters. A high degradation of the channel bit error rate may cause a high retransmission rate. On the other hand, even in good channel conditions, the retransmission rate increases enormously if there is a high number of receivers in a session. Hence, choosing a fixed coding scheme may cause the waste of bandwidth during the normal behavior of the channel since the redundant information is not required due to the low bit error rate of the channel. On the other hand, during the temporary degradation of the network, the amount of redundancy may not be sufficient for receivers to recover from transmission errors. Even with good channel conditions, if there is a high number of receivers, the redundancy level of a code may not be sufficient. Therefore, the use of adaptive coding schemes for wireless channels is an issue that requires further study.

An adaptive algorithm needs to estimate the channel conditions of all receivers listening to the same session in order to adjust its parameters dynamically based on an optimization criteria. Adaptive schemes have already been proposed in different contexts. It has been proposed for real-time applications in order to cope with retransmission delays in Internet [3] [4] [5] as well as in wireless networks [6] [7] [8] [9]. It has also been proposed for multicast communications [10] [11]. We observe that all the adaptive schemes designed for multicast communications are based on a fixed environment. The other works have considered a wireless network but their adaptation scheme is designed for a point-to-point communication mode. Our proposed approach is different from other adaptive algorithms since it is capable to adapt itself not only to the channel conditions but also to the number of receivers. It is based on a predictive mechanism in the sense that it forwards a certain number of redundant packets in the network before their necessity. It attempts to decrease the used bandwidth while maintaining the desired QoS parameters.

We take a finite state Markov chain in order to model the radio channel. The advantage of such a model lies on its facility to capture the burstiness of the error process as well as to predict the future states of the channel based on its present state. Prediction is useful due to the memory that exists in the physical channel. Our proposed scheme tries to take advantage of the channel memory in order to obtain better performance. We use *Reed-Solomon Erasure* (RSE) codes because of their appropriate characteristics in terms of powerful coding and

implementation simplicity.

The paper is organized as follows. We start by some background information about coding and Reed-Solomon Erasure codes. After describing the protocol models, we explain the QoS metrics that have been taken in order to analyze the effect of our adaptive scheme. Then, we present the finite state Markov model used in this paper. Our proposed prediction method as well as our adaptation policy are presented afterwards. Finally, we illustrate some simulation results comparing the performance of our adaptive error control protocol with other protocols.

1 Coding Aspects

Coding consists of adding redundant information to data in order to allow the receiver to recover the original data even in the presence of transmission errors. Basically, a code transforms a *data block* of k symbols $d = (d_{k-1}, d_{k-2}, ..., d_0)$ into a *coded block* of n symbols $C = (c_{n-1}, c_{n-2}, ..., c_0)$. In a system that uses FEC for error control, the sender and the receiver use a mutually agreed code to protect the data. If a coded block can be divided into the data part and the redundancy part, then the code is said to be a *systematic code*. A systematic code generates a coded block consisting of an unaltered copy of the data block followed by the $h = n - k$ redundant symbols. The advantage of a systematic code is that in case a receiver receives the data block correctly, no decoding is necessary.

A Reed-Solomon erasure code is a Reed-Solomon code with symbols defined over the *Galois Field* $GF(2^m)$, designed to recover from erasures. It is represented as $RSE(n, k)$ and it has a symbol size of m bits. A Reed-Solomon erasure code has the capacity to recover from h erasures with only h redundant symbols. This characteristic makes this kind of code particularly powerful to cope with transmission packet losses. The parameters of such a code are:

$$\begin{array}{ll} \text{Number of symbols in a coded block:} & n = 2^m - 1, \\ \text{Number of redundant symbols:} & h = n - k, \end{array}$$

In the sender side, the RSE encoder takes k data packets and generates h redundant packets to form a coded block of $n = k + h$ packets. If the receiver gets at least k packets out of the $k + h$ transmitted packets correctly, it can reconstruct the original data. Note that the loss unit is a packet and a packet payload is considered as a symbol.

1.1 Implementation Issues

McAuley proposed a hardware architecture for RSE codes in [12] using a symbol size $m = 8$ and $m = 32$. Rizzo proposed a software implementation of RSE codes in [13] with a symbol size in the range of $m = 4, ..., 16$. RSE coders with large symbol size are difficult to implement. Normally, the packet size is in the order of hundreds or thousands of bits. In this case, we need to consider a packet as l symbols of m bits and the coding can be implemented using l parallel RSE coders, each operating on a symbol size of m bits.

Since the number of elements of the $GF(2^m)$ with a symbol size of m is limited to 2^m, it is important to choose an RSE code with $n < 2^m$. If we take $m = 8$, we will have a maximum block length $n = 255$ which is sufficient in our case.

In the following, we use the software RSE coder developed by Rizzo in the systematic form with a symbol size $m = 8$. The encoding and decoding speeds of this software coder have been tested in various platforms from high speed workstations to small portable systems [13] [14] and have been shown to be in the order of Mega Bytes per second. One important observation is that for all the tested platforms the encoding and decoding speeds, c_e and c_d respectively, remain approximately constant over a wide range of k and h with c_d slightly smaller than c_e due to additional overheads in decoding. As a result, we consider them as constants during this study.

In order to have variable error correcting capabilities, we are interested to modify the coding parameters k and h of an RSE code. This is feasible by using *shortening* and *puncturing* techniques [15]. Shortening consists of adding a certain number of information symbols equal to zero to the original information in the encoding phase. Let's consider a Reed Solomon erasure code of $RSE(n, k)$. We can generate a set of shortened code $RSE(n-b, k-b)$ with $1 \leq b \leq k-1$ and an error correcting capability, h', equal to h. These shortened codes have their b high order information symbols equal to zero. Code puncturing involves not transmitting (deleting) certain redundant symbols. Puncturing allows a coder to change its number of redundant packets h while shortening allows it to change its number of data packets k. The shortened and punctured codes can use the same encoder/decoder pair as their original code.

We consider an *original code* $RSE(N_{max}, K_{max})$ with $H_{max} = N_{max} - K_{max}$. Using the shortening technique, we can derive a *basic code* $RSE(N, K)$ with the same number of redundant packets $H = H_{max} = N - K$. From this basic code, we can create a large set of RSE codes $RSE(n, k)$ with $k \leq K$ and $h \leq H$ using the shortening and puncturing techniques. The software coder proposed by Rizzo can be easily extended to support multiple block sizes and multiple redundant packets as in [14]. The only implication of such a coder is that it needs to support the maximum data block size K which is normally bigger than the actual data block size k. However, taking the maximum data block size allows us to use a single generator matrix that can support up to K data packets which is important if we need to vary our coding parameters.

It is important to note that having a $K \geq k$ leads to a higher space complexity of the software coder. However, the time complexity is unaffected by the value K. The time complexity is only affected by the actual data block size k used in encoding/decoding. Therefore, choosing a high K does not affect the encoding and decoding speeds.

2 Protocol Models

As stated before, our adaptive protocol can select two different error control protocols. The first protocol, P1 uses an ARQ mechanism. The second protocol, P2, uses a hybrid ARQ scheme with RSE codes. We consider that both protocols send k data packets before waiting for a feedback. Protocol P1 proceeds as follows:

- At the sender side

    1. Send k data packets.
    2. Send a POLL message for feedback to indicate the end of the transmission and start a timer.
    3. If no NAK is received after the time out, then no packet is lost. Proceed to the transmission of the next k new packets.
    4. If there are R received NAKs or if there is a timeout but there are less than R NAKs, include the lost packets as well as new packets in the next transmission.
    5. Go to the first step.

- At the receiver side

    1. Buffer the received packets.
    2. If k packets are received, send them to the higher layer.
    3. If a POLL is received but there is less than k packets in the buffer, send a NAK to the sender indicating the sequence numbers of the lost packets. Send the correctly received packets to the higher layer.
    4. Go to the first step.

Protocol P2 uses a hybrid ARQ scheme. We assume that the receiver can distinguish a data packet from a redundant packet thanks to the header information of each packet. If the receiver receives all the k data packets correctly, no decoding is necessary. Protocol P2 performs as follows:

- At the sender side

    1. Send k data packets followed by h redundant ones.
    2. Send a POLL for feedback to indicate the end of the transmission and start a timer.
    3. If no NAK is received after the time out, then no packet is lost. Proceed to the transmission of the next coded block composing of k new packets and h redundant packets.
    4. If there are R received NAKs or if there is a timeout with less than R NAKs, then include the lost packets as well as new packets followed by h redundant ones in the next transmission.
    5. Go to the first step.

- At the receiver side

    1. Buffer the data as well as redundant packets.
    2. If all the k data packets are received correctly, send them to the higher layer.

3. If there are lost data packets but with enough redundant packets (at least k packets out of the $k + h$ transmitted packets are correctly received), recover the lost packets by decoding and send the k data packets to the higher layer.
4. If there are lost data packets but there are not enough redundant packets to recover (less than k packets are correctly received), generate a NAK including the sequence numbers of the lost packets. Send the correctly received data packets to the higher layer.
5. Go to the first step.

Note that both protocols are NAK based and that we send only one NAK for a block of k packets. The receivers send a NAK message if they have not received a packet correctly. If a receiver has received all the packets correctly, it does not send any feedback. It is clear that in this case the sender needs a timer mechanism in order to proceed the transmission if it receives no NAK message from receivers.

We have considered a non-continuous mode protocol where the sender sends a POLL message in order to inform the receivers about the end of the transmission. The sender then starts a timer and stops sending packets until it receives a NAK from each receiver or until it makes a timeout. This is because in case of multicast communication, the reception of one NAK can not trigger the transmission of the next block since it may be other receivers that have lost a packet in the transmitted block but they have not yet sent their feedbacks. Hence, the sender can not start the transmission of the next block. However if it has received either no NAK or less than R NAKs after the timeout, it means that other receivers have correctly received their packets and thus the sender can proceed to the transmission of the next block. It is clear that the timer value must be large enough in order to allow all receivers to send their NAKs.

The protocols presented above do not have any feedback suppression mechanisms. In case of multicast communication, we may have several NAKs coming from different receivers to the sender for the same block causing a feedback implosion at the sender. A feedback suppression mechanism, like the one proposed in [20], is useful to reduce the processing load of the received feedbacks in the sender. However, it adds an extra delay in case of retransmission due to its mechanism for feedback suppression.

3 QoS Metrics

Several aspects must be taken into account when using an adaptive scheme as a QoS control mechanism. The first aspect to consider is its effect on bandwidth. In other words, we have to evaluate how much overhead our proposed scheme adds compared to other schemes. We take *efficiency* as a measure of the used bandwidth and we define it as the inverse of the average number of transmissions required by all receivers to receive a packet correctly.

The second issue is to evaluate the loss probability before and after our adaptive mechanism. We define *packet loss rate* as the probability that at least one receiver can not receive a packet correctly after the first transmission. This metric allows us to observe the decrease of

loss rate due to the utilization of our adaptive protocol. It gives us a precise measure of the effectiveness of our protocol in reducing the loss rate.

The last issue is the effect of our adaptive protocol on delay. The transmission delay of a packet is composed of several components. At the sender side, the delay is affected by the processing, queuing and encoding time. At the network side, we have to account for the transmission and propagation delays. At the receiver side, there are also processing, queuing and decoding delays that have to be considered. We ignore the effect of the queuing delay since this delay can be influenced by other flows of data and it depends on the congestion state of the network and also on the scheduling algorithm used. Here our objective is to see the effect of our protocol on the average delay of a packet. In this sense, we are interested in comparing the average delay of a packet in our protocol with other protocols.

In the protocol models used in our adaptive scheme, a receiver delivers the packets to the higher layer if it receives all the k packets in a block correctly or if it receives a POLL message. In other words, a packet can not be delivered to the higher layer before the end of the reception of its block. Therefore, the delay of each packet is nearly equivalent to the delay of its block since it must wait for the reception of all the packets in its block before going to the higher layer. Moreover in case of packet losses, the receiver has to wait for the transmission of the next block. Therefore, we define the *average delay* as the average time spanning from the beginning of the transmission of a block until it has been successfully received by all receivers.

Throughout this paper, we suppose that the loss events at different receivers are independent. We assume that all bit errors in a received packet are detected thanks to its CRC field and no control messages are lost. In case of multicast, the traffic is transmitted to all receivers using the broadcast mechanism of the radio rather than sending a separate copy for each receiver.

4 Finite State Markov Model

Markov chains have been extensively used in the literature to capture the bursty nature of the error sequences generated by a wireless channel. Previous studies [16], [17] show that a first order Markov chain provides a good approximation of the error process in fading channels. Furthermore, the parameters of the model can be easily mapped to real physical quantities in case of a Rayleigh fading channel. We take a finite state Markov model as in [18]. This model is depicted in Figure 1. As it can be seen, the channel states associated with consecutive symbols are assumed to be neighboring states. This assumption is true for a slow fading channel where the SNR varies slowly compared to the symbol interval T.

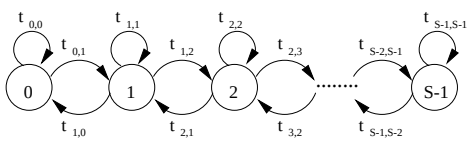


Figure 1: Finite state Markov model

Let $0 = \lambda_0 < \lambda_1 < \lambda_2 < ... < \lambda_S = \infty$ be the thresholds of the received SNR. The channel is said to be in state s where $s \in \{0, 1, 2, ..., S-1\}$ if the received SNR is in the interval $[\lambda_s, \lambda_{s+1})$. Associated with each state, there is a Binary Symmetric Channel (BSC) with the error probability e_s. Recall that Rayleigh fading results in an exponentially distributed distortion of the signal [19]. The probability density function of the SNR, $f(\lambda)$, follows an exponential distribution:

$$f(\lambda) = \frac{1}{\bar{\lambda}} exp(-\frac{\lambda}{\bar{\lambda}}) \qquad \lambda > 0 \tag{1}$$

where $\bar{\lambda}$ is the average SNR.

Assuming that the channel fades slowly with respect to the symbol interval, T, the Markov transition probabilities can be approximated using the level crossing rate and the SNR density function as follows:

$$t_{s,s+1} \approx \frac{1}{\pi_s} exp(-\frac{\lambda_{s+1}}{\bar{\lambda}}) f_d T \sqrt{\frac{2\pi\lambda_{s+1}}{\bar{\lambda}}} \tag{2}$$

$$t_{s,s-1} \approx \frac{1}{\pi_s} exp(-\frac{\lambda_s}{\bar{\lambda}}) f_d T \sqrt{\frac{2\pi\lambda_s}{\bar{\lambda}}} \tag{3}$$

$$t_{s,s} = 1 - t_{s,s-1} - t_{s,s+1} \tag{4}$$

$$t_{0,0} = 1 - t_{0,1} \tag{5}$$

$$t_{S-1,S-1} = 1 - t_{S-1,S-2} \tag{6}$$

where f_d is the maximum doppler frequency given by $f_d = \frac{v f_c}{c}$ with v the vehicle speed, f_c the carrier frequency and c the speed of light ($3 \times 10^8 m/s$). The steady state probabilities, π_s, are:

$$\pi_s = \int_{\lambda_s}^{\lambda_{s+1}} f(\lambda) d\lambda = exp(\frac{-\lambda_s}{\bar{\lambda}}) - exp(\frac{-\lambda_{s+1}}{\bar{\lambda}}) \tag{7}$$

The error probabilities of each state e_s can be related to the received SNR according to the modulation scheme used in the system.

$$e_s = \frac{\int_{\lambda_s}^{\lambda s+1} f(\lambda) e_m(\lambda) d\lambda}{\int_{\lambda_s}^{\lambda s+1} f(\lambda) d\lambda} = \frac{1}{\pi_s} \int_{\lambda_s}^{\lambda s+1} f(\lambda) e_m(\lambda) d\lambda \tag{8}$$

where $e_m(\lambda)$ is the modulation function relating bit error probability to received SNR. For a *Binary Phase Shift Keying* (BPSK) scheme, we have:

$$e_m(\lambda) = 1 - F(\sqrt{2\lambda})$$

where

$$F(\lambda) = \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})dx$$

Simplified expression for e_s is provided in [18] for a BPSK scheme. The average error rate of the model can be found as $e = \sum_{s=0}^{S-1} \pi_s e_s$.

5 Prediction Method

We consider a finite state Markov model as described in the previous section. In order to investigate the effect of packet level FEC, we are interested to model the process of successful or unsuccessful packet transmission. [21] showed that a Markov approximation for a packet loss process is a good model for a broad range of parameters. In fact for typical data rates (e.g. more than 64 Kb/s) and for environments commonly considered (e.g. carrier frequency of about 1-2 GHZ and typical pedestrian and vehicular speeds), we can assume that the channel is constant during a packet interval T. With this assumption, the packet loss probability of each state, p_s, can be calculated as in a BSC model with the error probability e_s. For a packet of length L bits, we have:

$$p_s = 1 - (1 - e_s)^L \tag{9}$$

5.1 Efficiency

In order for our adaptive algorithm to change its coding parameters dynamically, it must be able to predict the performance of each of the available error control schemes for the next block before actually transmitting it. Let us first consider the protocol P1 where a sender multicasts data to R receivers using an ARQ scheme. The original packet is retransmitted if there is at least one receiver that has not received the packet correctly. The sender sends k packets at a time before waiting for a feedback. Generally, it is not aware of a packet loss unless it receives a negative feedback from one of the receivers. In this case, it can only retransmit the lost packet after the retransmission delay. We assume that the channel remains at the same state during the time spanning the end of the transmission of a block and the beginning of the transmission of the next block. It is clear that if this time interval is longer than the correlation time of the channel, the assumption that the channel stays at the same state is not correct. We define L_r as the number of times that a packet gets lost by a receiver.

We assume that the state transitions occur at the beginning of a time slot of unit length and then a packet is transmitted. The probability that a receiver loses a packet exactly l times is:

$$P(L_r = l) = \sum_{s=0}^{S-1} P_s(L_r = l), \tag{10}$$

$$P_s(L_r = l) = \begin{cases} \sum_{i=0}^{k-1} \Big[P_s(i, k-1) t_{s,s}(1-p_s) \\ + P_{s-1}(i, k-1) t_{s-1,s}(1-p_s) \\ + P_{s+1}(i, k-1) t_{s+1,s}(1-p_s) \Big] & l = 0 \\ \sum_{i=0}^{k-1} \Big[P_s(i, k-1) t_{s,s} p_s \\ + P_{s-1}(i, k-1) t_{s-1,s} p_s \\ + P_{s+1}(i, k-1) t_{s+1,s} p_s \Big] & l = 1, 2, ... \end{cases}$$

$P_s(L_r = l)$ is the probability that a receiver loses a packet l times with the channel ending in state s. $P_s(i, k-1)$ represents the probability to have i packet losses in $k-1$ packet transmissions with the channel ending in state s. [22] calculated the probability to have i errors in j transmissions in a Gilbert-Elliot model using recursion. Using the same approach, we can calculate the probability to have i packet losses among j transmitted packets, $P(i, j)$, in a finite state Markov chain. Let $P_s(i, j)$ be the probability to have i packet losses among j transmitted packets with the channel ending in state s. As before, we assume that state transitions occur at the beginning of a time slot of unit length and then a packet is transmitted. Extending the equation from a Gilbert-Elliot model to a finite state Markov model, the probability to have i packet losses in j packet transmissions is:

$$P(i, j) = \sum_{s=0}^{S-1} P_s(i, j), \qquad \text{for } i = 0, 1, 2, ..., s \text{ and } j = 1, 2, 3... \tag{11}$$

$$\begin{aligned} P_s(i, j) = & P_s(i, j-1) t_{s,s}(1-p_s) \\ & + P_{s-1}(i, j-1) t_{s-1,s}(1-p_s) \\ & + P_{s+1}(i, j-1) t_{s+1,s}(1-p_s) \\ & + P_s(i-1, j-1) t_{s,s} p_s \\ & + P_{s-1}(i-1, j-1) t_{s-1,s} p_s \\ & + P_{s+1}(i-1, j-1) t_{s+1,s} \end{aligned}$$

We assume that the adaptive algorithm is informed about the channel state of all the receivers at the beginning of the transmission of each block. Once the channel state of all receivers at instant t is known, the algorithm can predict the evolution of channel conditions

of the receivers for the next block taking advantage of the fact that the future states of the Markov chain depends only on its present state. Assuming that a receiver is in state s' at the beginning of the transmission, the initial conditions for $P_s(i,j)$ in equation (10) are:

$$P_s(0,0) = \begin{cases} 1 \quad \text{if} \quad s = s' & l = 0,1 \\ 0 \quad \text{otherwise} & l = 0,1 \\ P_s(L_r = l-1) & l = 2,... \end{cases}$$

Using the above initial conditions, we get S different values for $P(i,j)$ and $P(L_r = l)$ depending on the state where the receiver was at the beginning of the transmission. We represent these probabilities by $P(L_r = l|s')$ and $P(i,j|s')$ where s' is the state of a receiver at the beginning of the transmission. We represent the number of receivers in each of the states of the Markov chain by $\{r_0, r_1, .., r_{S-1}\}$ and the total number of receivers as R. It is clear that we have $\sum_{s=0}^{S-1} r_s = R$. The algorithm estimates the efficiency of P1 for R receivers as follows:

$$Eff = \frac{1}{E[M]} = \frac{1}{\sum_{m=1}^{\infty} \left(1 - \prod_{s'=0}^{S-1} (1 - P(L_r = m-1|s'))^{r_{s'}}\right)} \tag{12}$$

Now, we consider protocol P2 where the sender uses an RSE code with a coded block size of n packets containing k original packets and h redundant packets. In this case, the sender sends k original packets followed by h redundant ones. Each receiver can recover from loss if it receives correctly k packets out of the $n = k + h$ transmitted packets. If the receiver can not recover from loss, it asks for a retransmission.

We define $Q(L_r = l)$ as the probability that a receiver loses a packet exactly l times in the case of hybrid ARQ. $Q(L_r = l)$ is again the sum of $Q_s(L_r = l)$, the probability of a receiver to lose a packet exactly l times with the channel ending in state s. In the presence of FEC, a packet is retransmitted if it is lost by the FEC receiver and if more than $h-1$ out of the other $n-1$ packets of the coded block are lost. In the same way, a packet is considered to be correctly received if it has not been lost or if it has been lost but there are at least $h-1$ packets out of the other $n-1$ packets of the coded block that have been correctly received. Once again, we assume that the channel does not change its state during the interval $T + t$ where t corresponds to the time between the end of the transmission of the last packet of a coded block and the beginning of the transmission of the first packet of the next coded block.

$$Q(L_r = l) = \sum_{s=0}^{S-1} Q_s(L_r = l), \tag{13}$$

$$Q_s(L_r = l) = \begin{cases} \sum_{i=0}^{h-1} \Big[P_s(i, n-1) t_{s,s} p_s \\ + P_{s-1}(i, n-1) t_{s-1,s} p_s \\ + P_{s+1}(i, n-1) t_{s+1,s} p_s \Big] + \\ \sum_{i=0}^{n-1} \Big[P_s(i, n-1) t_{s,s} (1-p_s) \\ + P_{s-1}(i, n-1) t_{s-1,s} (1-p_s) \\ + P_{s+1}(i, n-1) t_{s+1,s} (1-p_s) \Big] & l = 0 \\ \sum_{i=h}^{n-1} \Big[P_s(i, n-1) t_{s,s} p_s \\ + P_{s-1}(i, n-1) t_{s-1,s} p_s \\ + P_{s+1}(i, n-1) t_{s+1,s} p_s \Big] & l = 1, 2, ... \end{cases}$$

In order to estimate the efficiency of protocol P2, the adaptive algorithm needs to estimate $Q(L_r = l)$ first. Assuming that a receiver is in state s' at the beginning of a transmission, the initial conditions for $P_s(i, j)$ in equation (13) are:

$$P_s(0,0) = \begin{cases} 1 \quad \text{if} \quad s = s' & l = 0, 1 \\ 0 \quad \text{otherwise} & l = 0, 1 \\ Q_s(L_r = l - 1) & l = 2, ... \end{cases}$$

Note that once again, we have different $Q(L_r = l)$ probabilities depending on the channel state at the beginning of the transmission. We represent these probabilities by $Q(L_r = l | s')$ where s' is the channel state of a receiver at the beginning of the transmission. The algorithm predicts the efficiency of protocol P2 as follows:

$$Eff = \frac{1}{E[M]} = \frac{k}{n} \frac{1}{\sum_{m=1}^{\infty} \left(1 - \prod_{s'=0}^{S-1} (1 - Q(L_r = m-1 | s'))^{r_{s'}} \right)} \tag{14}$$

Note that in all the above formulas we have $t_{s-1,s} = 0$ for $s = 0$ and $t_{s,s+1} = 0$ for $s = S - 1$.

5.2 Packet Loss Rate

The next QoS metric that the adaptive algorithm must estimate is the packet loss rate which is the probability to have at least one receiver that has not received a packet correctly after the first transmission. Considering protocol P1, the probability to receive a packet correctly after the first transmission is $P(L_r = 0)$. Once again, we define r_s as the number of receivers in state s. $P(L_r = 0|s')$ is the probability of a receiver to receive a packet correctly after the first transmission with the receiver being in state s' at the beginning of the transmission. The packet loss rate of protocol P1 is estimated as follows:

$$PLR = 1 - \prod_{s'=0}^{S-1} \Big[P(L_r = 0|s') \Big]^{r_{s'}} \tag{15}$$

In case of protocol P2, the probability that a receiver gets a packet correctly after the first transmission is $Q(L_r = 0)$. We represent this probability with the receiver being in state s' at the beginning of the transmission by $Q(L_r = l|s')$. The adaptive algorithm estimates the packet loss rate of protocol P2 as below:

$$PLR = 1 - \prod_{s'=0}^{S-1} \Big[Q(L_r = 0|s') \Big]^{r_{s'}} \tag{16}$$

5.3 Average Delay

Variable	Definition
D	Packet delay
D_d	Transmission time of a data packet
D_c	Transmission time of a control packet
X_c	Control packet (NAK, POLL) processing time at sender
X_e	Encoding delay per packet at sender
T_s	The timer value at sender
Y_d	Data packet processing time at receiver
Y_c	Control packet (NAK, POLL) processing time at receiver
Y_e	Decoding delay per packet at receiver
N_r	Number of transmission rounds for a packet
A	Number of received NAKs at the sender
R	Number of receivers
l	Number of lost packets
L_B	Packet size in Bytes

Table 1: The definition of the variables used in the delay analysis

The last metric that the adaptive algorithm has to consider is the average delay. The list of all variables involved in the delay analysis is presented in Table 1. In this analysis, we ignore the effect of propagation delay.

Let us begin with protocol P1 which uses only an ARQ mechanism. Considering that each packet is transmitted $E[M]$ times in average, the total delay due to the transmission of a block is $kE[M]D_d$. Since the protocol does not work in a continuous way, the sender has to either receive R NAKs or make a timeout in order to continue the transmission of the next block. Therefore, a packet can not be retransmitted immediately. We define *Round Trip Delay* (RTD) as the minimum time that it takes between the end of the transmission of a block and the beginning of the transmission of the next block. Note that this minimum time corresponds to the situation where the sender receives R NAKs before the expiration of its timer. In case the sender receives less than R NAKs, it has to wait for the timeout. The timer value must be high enough to allow all receivers to send their NAKs. The sender starts the timer right after the transmission of the POLL message. Therefore, the retransmission only takes place after the POLL processing delay at the sender, the POLL transmission time and the RTD in the best case.

In order to calculate the RTD, we have to take into account the average processing time of the POLL message at the receiver, $E[Y_c]$, the average processing time of a NAK message at the receiver, $E[Y_c]$, the NAK transmission time, D_c, and finally the average NAK processing time at the sender, $RE[X_c]$. Therefore, we have:

$$RTD = RE[X_c] + 2E[Y_c] + D_c$$

In order to estimate the average delay, we need to estimate the average number of transmissions for a packet first. This parameter has already been estimated in equation (12). Moreover, we have to predict the probabilities of having R NAKs and no NAK for a block of k packets. Recalling that every receiver sends only one NAK for k packets, the probability to have a NAKs, $\Phi(A = a)$, is the probability to have only and only a receivers that have lost at least one packet among the k transmitted packets. The probabilities of having R NAKs and no NAK for a block of k packets can be estimated as follows:

$$\Phi(A = R) = \prod_{s'=0}^{S-1} \Big[1 - P(0,k|s')\Big]^{r_{s'}} \tag{17}$$

$$\Phi(A = 0) = \prod_{s'=0}^{S-1} \Big[P(0,k|s')\Big]^{r_{s'}} \tag{18}$$

$P(0,k|s')$ is the probability to have zero losses in k transmissions with the receiver in state s' at the beginning of the transmission. This probability can be calculated from equation (11) using the following initial values.

$$P_s(0,0) = \begin{cases} 1 & \text{if } \ s = s' \\ 0 & \text{otherwise} \end{cases}$$

The average delay in protocol P1 is then determined as follows. Note that the RTD and T_s are the minimum and the maximum times that it takes before the sender starts transmitting

the next block. The $E[X_c]+D_c$ in the following formula corresponds to the POLL processing and transmission delays.

$$\begin{aligned} E[D] =& kE[M]D_d + kE[Y_d] \\ &+ (E[M]-1)\Big(E[X_c] + D_c + \Phi(A=R)RT\,D \\ &+ (1-\Phi(A=R)-\Phi(A=0))T_s \Big) \end{aligned} \tag{19}$$

The delay of protocol P2 is essentially the same but with the addition of coding delays. Recall that if a packet is correctly received no decoding is necessary since we use systematic codes. Decoding is only used when a packet is lost but there are enough redundant packets in order to recover. For this protocol, the adaptive algorithm needs to estimate the probability of decoding as well as the probabilities to have R NAKs and no NAK before estimating the average delay. We take P_d as the probability of decoding which is the probability to have at least one receiver which has lost a packet but it has enough redundant packets to recover from the loss (at least k packets out of the other $n-1$ packets are correctly received). The protocol estimates the probability of decoding P_d as:

$$P_d = \prod_{s'=0}^{S-1} \Big[Q(L_r=0|s') \Big]^{r_{s'}} - \prod_{s'=0}^{S-1} \Big[P(L_r=0|s') \Big]^{r_{s'}} \tag{20}$$

We define $E[N_r]$ as the average number of transmission rounds necessary for a packet to be correctly received by all receivers. In fact, $E[N_r]$ is the average number of transmissions required for a packet to be correctly received by all receivers without taking into account the coding overhead:

$$E[M] = \frac{n}{k} E[N_r] \tag{21}$$

where $E[M]$ in the above equation is found from equation (14). Rizzo has calculated the time required to produce h parity packets in [13] as follows:

$$X_e = \frac{kL_B}{c_e}(n-k) \tag{22}$$

where L_B is the packet size in Bytes and c_e is the encoding constant in Byte/sec. In the same way, the decoding time is $Y_e = \frac{kL_B}{c_d} l$ where l is the number of lost packets and c_d is the decoding constant in Byte/sec. Assuming that the number of lost packets in a coded block is equal to the number of redundant packets in a block, the decoding delay that each packet undergoes is:

$$Y_e = \frac{kL_B}{c_d}(n-k) \tag{23}$$

The last thing that we have to estimate is the probabilities to have R NAKs and no NAK in case of P2:

$$\Phi(A=R) = \prod_{s'=0}^{S-1}\Big[1-\sum_{i=0}^{h} P(i,n|s')\Big]^{r_{s'}} \tag{24}$$

$$\Phi(A=0) = \prod_{s'=0}^{S-1}\Big[\sum_{i=0}^{h} P(i,n|s')\Big]^{r_{s'}} \tag{25}$$

where again $P(i,n|s')$ represents the probability to have i losses in n transmissions with th receiver in state s' at the beginning of the transmission. The initial values are the same as equations (17) and (18).

The overall average delay of protocol P2 can be determined as below. The RTD is exactly the same as in protocol P1. We have simplified the decoding delay by assuming that the number of lost packets in a block is equivalent to the number of redundant packets of the block. The implications of this simplification are that we calculate the decoding delay in the worst case and that only k packets among the n transmitted ones are correctly received at the last transmission round.

$$\begin{aligned} E[D] =& nE[N_r]D_d + E[N_r]X_e + kE[Y_d] + P_dY_e \\ &+ (E[N_r]-1)\Big(E[X_c] + D_c + \Phi(A=R)RTD \\ &+ (1-\Phi(A=R)-\Phi(A=0))T_s\Big) \end{aligned} \tag{26}$$

6 Adaptation Policy

Let $C = \{c_0, c_1, ..., c_k\}$ be the set of RSE codes available at the sender. The sender can either choose a hybrid ARQ error control protocol (protocol P2) with an RSE code in C or a pure ARQ protocol (protocol P1). According to the variations of SNR, the receiver channel may be in one of the states of the Markov model at each instant t. We assume that the sender knows the state of the Markov chain at the transmission time for all receivers. Let's define the *transmission status* at time t as the set of all tuples (s, r_s) where $s \in \{0, 1, ..., S-1\}$ is the channel state in the Markov model and r_s is the number of wireless receivers in state s at time t.

Before transmitting, the adaptive algorithm in the sender must estimate the efficiency, packet loss rate and delay of the hybrid ARQ protocol using all the available coding schemes as well as the ARQ protocol as a function of the transmission status. It then tries to find

the protocol satisfying the desired packet loss rate and delay. If there are several protocols satisfying these criteria, the algorithm must choose the one with the highest efficiency in order to minimize the use of bandwidth. Note that our adaptive approach is predictive rather than reactive since the sender tries to predict the channel conditions as well as the evolution of QoS metrics for all receivers before actually sending a block. The sender then chooses a protocol according to its predictions.

The time is divided into transmission rounds. Each transmission round corresponds to the transmission of n packets in case of FEC and k packets in case of ARQ. A transmission round ends when the sender is informed about the reception states of all receivers. The adaptive algorithm is repeated at the end of each transmission round. Basically, the algorithm goes through the following steps:

1. At the beginning of the algorithm, the sender determines the desired packet loss rate and delay of the session. It also determines the transmission status.
2. The sender estimates the packet loss rate and the delay of the ARQ as well as the hybrid ARQ using all the available coding schemes, based on the transmission status. If it finds several protocols satisfying the QoS metrics of the session, it chooses the one with the highest efficiency. It then adjusts its parameters and starts the transmission of the block.
3. At the end of a transmission round, the sender again determines the transmission status. It then repeats the step 2.

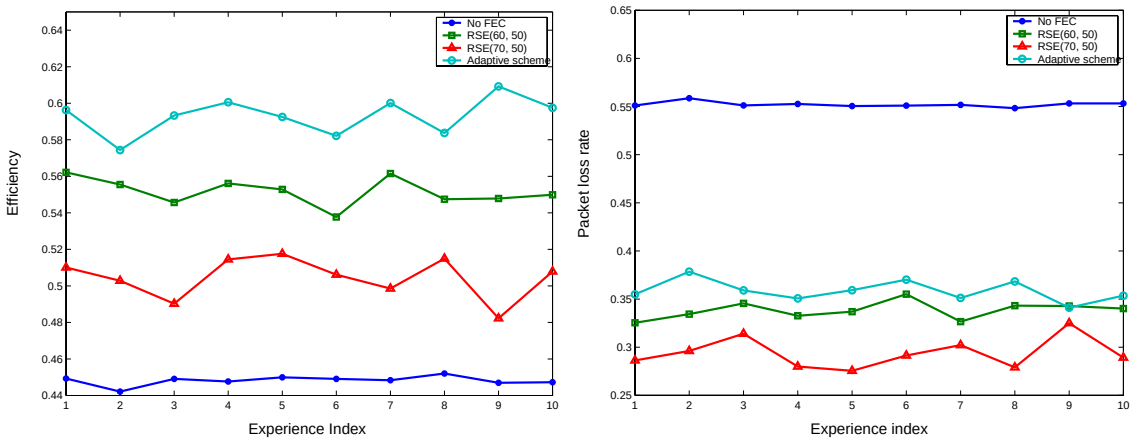
In the above algorithm, the sender finds the best error control mechanism in real-time right before transmitting a block. One possible optimization is for the sender to make a table of optimal mechanisms called FEC_TABLE indexed on transmission status. Each time, the sender estimates the QoS metrics for a transmission status, it adds an entry to the FEC_TABLE. In this way, if a transmission status occurs again, the sender can find the best mechanism by a simple table lookup. If the service differentiation is limited to some classes with predefined QoS metrics, then a table can be generated for each QoS class off-line.

One may argue that the sender risks to consume a lot of memory if there is a high number of receivers or if the Markov model used to model the wireless channel has many states. It is clear that the choice between using a precomputed table or an online estimation is a tradeoff between the consumed memory and the complexity of the algorithm. However, the simulation results show that the choice of the best error control protocol does not vary significantly for a wide range of transmission status.

7 Simulation Results

We have carried out several simulations in OPNET which is an event-driven simulation tool. We take a date rate of 20 Mb/s for our wireless network. The carrier frequency is 5.2 GHz. The data and control packets have 54 and 9 bytes respectively. The timer value of the protocols is fixed at $T_s = 2RTD$. We use a BPSK modulation scheme. The average SNR is 34

dB corresponding to an average bit error probability of 10^{-4}. All the receivers are located within a distance of 23 meters from their base station. The wireless channel is modeled by a 3 state Markov model in the OPNET environment. The state s_0 of the Markov model corresponds to a Bad state with $e_0 \approx 1$, the state s_1 corresponds to an intermediate state with a non-zero error probability $e_1 \approx 2 \times 10^{-5}$ and the state s_2 corresponds to a Good state with a zero error probability $e_2 \approx 0$. In order to have an error probability of $e_0 \approx 1$, λ_1 must be equal to 2dB in BPSK. For a zero error probability $e_2 \approx 0$ in state s_2, we also need to fix λ_2 at 34dB in BPSK. Knowing the threshold values of the Markov model, all the other parameters can be easily found as in Section 4. For our adaptive scheme, we take an original code $RSE(255, 235)$ and a basic code $RSE(70, 50)$. Using this basic code, we can vary the coding parameters such that for any used code $RSE(n, k)$, we have $k \leq 50$ and $h \leq 20$.



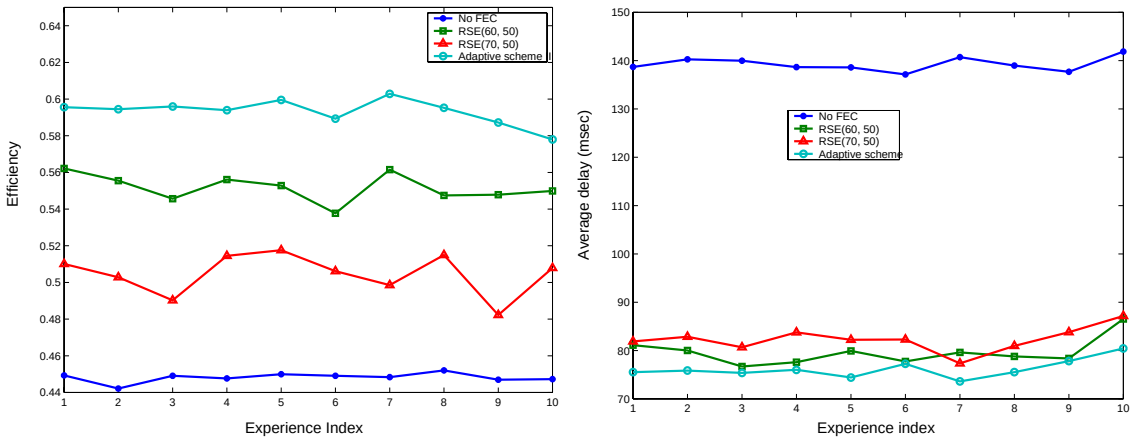
(a) Efficiency

(b) Packet loss rate

Figure 2: Simulation results for $PLR = 50\%$

For each scenario, we have carried out ten different simulations, each with a different seed. Figure 2 compares the efficiency and the packet loss rate of our proposed adaptive scheme with a pure ARQ protocol, a hybrid ARQ protocol using $RSE(60, 50)$ and another hybrid protocol using $RSE(70, 50)$. The number of receivers is fixed at 1000. This figure corresponds to a scenario where non-real-time traffic such as data is transmitted over the network. The protocols try to retransmit a packet until it is correctly received by all receivers. We have chosen a $PLR = 50\%$ in order to reduce the retransmission rate by a half in case of adaptive scheme. From this figure, we can observe that the adaptive scheme provides the best efficiency. It also has a PLR less than 50% as it was expected. Although other fixed hybrid protocols provide better packet loss rates, they have a lower efficiency compared to our adaptive scheme.

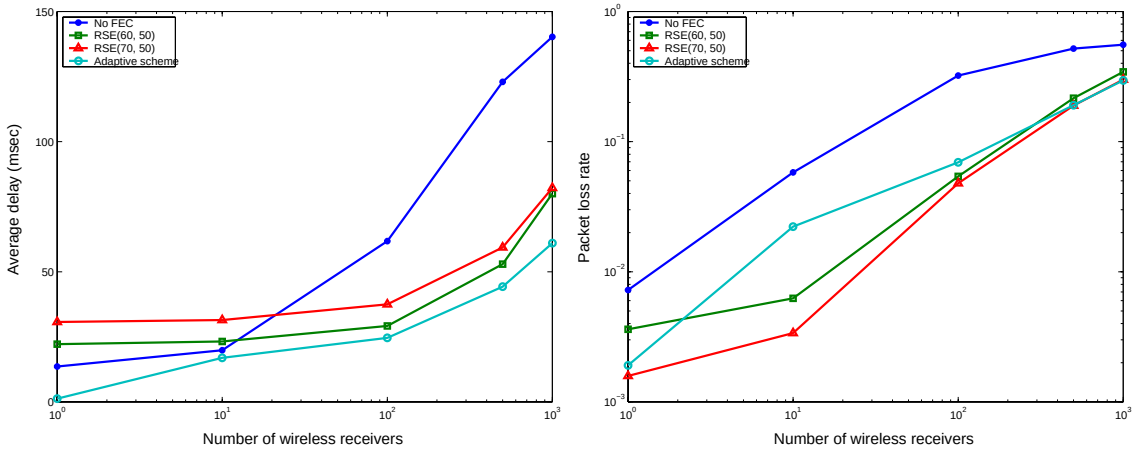
Figure 3 compares the efficiency and the average delay of our proposed adaptive scheme with the same error control mechanisms as in the previous figure. The number of receivers is again 1000. This scenario corresponds to a traffic with delay constraints. The maximum lifetime of each packet is fixed at 100 msec. We observe that our proposed adaptive scheme provides the lowest average delay while maximizing the efficiency. The other hybrid ARQ protocols also provide an average delay lower than 100 msec but they have a lower efficiency



(a) Efficiency (b) Average delay

Figure 3: Simulation results for a packet life time of 100 msec

than our protocol.



(a) Average delay (b) Packet loss rate

Figure 4: Simulation results for a packet life time of 50 msec and a PLR of 10%

Figure 4 depicts the average delay and the packet loss rate of our adaptive algorithm and the same error control protocols as before as a function of the number of receivers. The packet life time is fixed at 50 msec and the packet loss rate is 10%. Our adaptive mechanism can guarantee an average delay of 50 msec up to 600 receivers while the ARQ mechanism, for example, can guarantee this delay up to 30 receivers as it can be seen in the first plot. The packet loss rate is also less than 10% up to 200 receivers in our adaptive scheme and less than 20 receivers in the case of ARQ protocol.

8 Conclusion

In this paper, we proposed an adaptive algorithm capable to switch between an ARQ and a set of hybrid ARQ error control protocols. The coding scheme used in the hybrid ARQ protocols is based on RSE codes. The adaptive algorithm chooses the best error control mechanism as a function of the channel bit error rate, the channel state of the receivers and the desired QoS metric of the receivers while maximizing efficiency. We used a finite state Markov chain as our wireless channel model. This model allowed us to predict the future states of the channel for each receiver based on its current channel state. Simulation results showed that the use of adaptive mechanism is useful in order to save bandwidth while maintaining the QoS metrics below their thresholds.

We considered efficiency, packet loss rate and average delay for our analysis. The effect of our adaptive protocol on other QoS metrics such as jitter, dropping rate and power consumption of mobile terminals is an interesting direction for our future work.

References

[1] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in *Proceedings of SIGCOMM 96*, Stanford, CA, August 1996, pp. 117–130.

[2] N. Nikaein and C. Bonnet, "On the performance of FEC for multicast communication on a fading channel," in *Proceedings of International Conference on Telecommunications IEEE ICT'00*, Acapulco, Mexico, May 2000.

[3] Tsunyi Tuan and Kihong Park, "Multiple time scale redundancy control for QoS-sensitive transport of real-time traffic," in *Proceedings of INFOCOM'00*, Tel Aviv, Israel, Mar. 2000.

[4] Jean Bolot and Thierry Turletti, "Adaptive error control for packet video in the internet," in *proccedings of ICIP'96*, Lausanne, Switzerland, Sept. 1996.

[5] Jean-Chrysostome Bolot, Sacha Fosse-Parisis, and Don Towsley, "Adaptive FEC-Based error control for interactive audio in the internet," in *proceedings of INFOCOM'99*, New York, March 1999.

[6] Chi-Yuan Hsu, Antonio Ortega, and Masoud Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE JSAC*, vol. 17, no. 5, pp. 756–773, May 1999.

[7] Daji Qiao and Kang G. Shin, "A two-step adaptive error recovery scheme for video transmission over wireless networks," in *proceedings of INFOCOM'00*, Tel Aviv, Israel, Mar. 2000.

[8] M. Khansari, A. Jalali, E. Dubois, and P. Mermelstein, "Low bit-rate video transmission over fading channels for wireless microcellular systems,," *IEEE Transactions on Circuits and Systems for Video Technolog*, vol. 6, no. 1, pp. 1–11, 1996.

[9] H. Liu and M. El Zarki, "Delay bounded type-II hybrid ARQ for video transmission over wireless networks," in *Proceedings of Conference on Information Sciences and Systems*, Princeton, NJ, March 1996.

[10] P. Godlewski, M. Braneci, and A. Serhrouchni, "An error control scheme for multicast communications over an ATM network," in *Proceedings of Singapore International Conference on Communication Systems (ICCS'94)*, Singapore, November 1994.

[11] Dan Rubenstein, Jim Kurose, and Don Towsley, "Real-time reliable multicast using proactive forward error correction," in *NOSSDAV'98*, Cambridge, England, July 1998, pp. 279–293.

[12] Anthony J. McAuley, "Reliable broadband communication using a burst erasure correcting code," in *Proceedings of SIGCOMM'90*, Philadelphia, Pennsylvania, September 1990.

[13] Luigi Rizzo, "On the feasibility of software FEC," *Computer Communication Review*, April 1997,
Source code available at http://www.iet.unipi.it/ luigi/fec.html.

[14] Dan Rubenstein, "Increasing the functionality and availability of reed-solomon fec codes: a performance study," Tech. Rep. UMass CMPSCI Technical Report 98-31, University of Massachusetts, August 1998.

[15] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[16] M. Zorzi, R.R. Rao, and L.B. Milstein, "On the accuracy of a first-order Markov model for data block transmission on fading channels," in *Proceedings IEEE ICUPC'95*, November 1995, pp. 211–215.

[17] H.S. Wang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," in *Proceedings IEEE ICUPC'94*, 1994, pp. 160–164.

[18] Hong Shen Wang and Nader Moayeri, "Finite-state Markov channel - a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, February 1995.

[19] J. G. Proakis, *Digital Communications*, New York: MacGrawhill, 2nd edition, 1989.

[20] S. Floyd, V. Jacobson, C. Liu, S. McCanne, and L. Zhang, "A reliable multicast framework for ligth-weight sessions and application level framing," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 784–803, December 1997.

[21] M. Zorzi, R.R. Rao, and L.B. Milstein, "Error statistics in data transmission over fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 1468–77, November 1998.

[22] James R. Yee and Edward J. Weldon, "Evaluation of the performance of error-correcting codes on a Gilbert channel," *IEEE Transactions on Communications*, vol. 43, no. 8, pp. 2316–2323, August 1995.