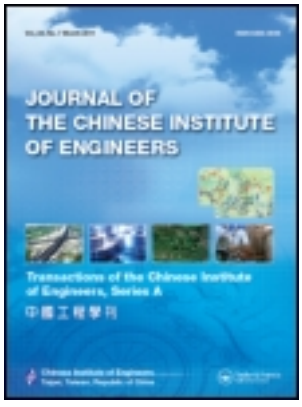

Journal of the Chinese Institute of Engineers
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcie20

Joint playout and FEC control for multi-stream voice over IP networks

Chun-Feng Wu [a] , Yuan-Chuan Chiang [b] & Wen-Whei Chang [a]
[a] Institute of Communications Engineering , National Chiao-Tung University , 1001 Ta Hsueh Road, Hsinchu , Taiwan
[b] Department of Special Education , National Hsinchu University of Education , 521 Nanda Road, Hsinchu , Taiwan
Published online: 06 Dec 2012.

To cite this article: Chun-Feng Wu , Yuan-Chuan Chiang & Wen-Whei Chang (2013) Joint playout and FEC control for multi-stream voice over IP networks, Journal of the Chinese Institute of Engineers, 36:2, 224-235, DOI: 10.1080/02533839.2012.726334

To link to this article: http://dx.doi.org/10.1080/02533839.2012.726334

Journal of the Chinese Institute of Engineers, 2013
Vol. 36, No. 2, 224–235, http://dx.doi.org/10.1080/02533839.2012.726334

Taylor & Francis
Taylor & Francis Group

Joint playout and FEC control for multi-stream voice over IP networks

Chun-Feng Wu[a], Yuan-Chuan Chiang[b] and Wen-Whei Chang[a]*

[a]*Institute of Communications Engineering, National Chiao-Tung University, 1001 Ta Hsueh Road, Hsinchu, Taiwan;*
[b]*Department of Special Education, National Hsinchu University of Education, 521 Nanda Road, Hsinchu, Taiwan*

(*Received 30 November 2010; final version received 4 September 2011*)

Packet loss and delay are the major network impairments for transporting real-time voice over internet protocol (IP) networks. In the proposed system, multiple descriptions of the speech are used to take advantage of packet path diversity. A new objective method is presented for predicting the perceived quality of multi-stream voice transmission. Also proposed is a joint playout buffer and forward error control (FEC) adjustment scheme that maximizes the perceived speech quality via delay-loss trading. Experimental results showed that the proposed multi-stream voice transmission scheme achieves significant reductions in delay- and packet-loss rates as well as improved speech quality.

Keywords: playout buffer; forward error control; multiple description coding; voice quality prediction model

1. Introduction

Quality of Service (QoS) has been one of the major concerns in the context of real-time voice communication over unreliable internet protocol (IP) networks. Iterative audio applications such as telephony and audio conferencing require high constraints on packet loss and end-to-end delay. In addition, the network delay experienced may vary for each packet depending on the level of congestion along the path. The variation in network delay, referred to as jitter, must be smoothed out since it obstructs the proper and timely reconstruction of the speech signal at the receiver end. The most common approach is to store recently arrived packets in a buffer before playing them out at scheduled intervals. By increasing the buffer size, the late loss rate is reduced, but the resulting improvement in voice transmission is off-set by the accompanying increase in the end-to-end delay. In balancing the impairment due to delay and packet loss, two current coding strategies, single description (SD) and multiple description (MD) transmissions, have used different playout buffer algorithms. In SD coding, a number of adaptive playout buffer algorithms have been proposed that react to changing network conditions by dynamically adjusting the playout delay. Most of them work by taking measurements on the network delays and either compressing or expanding silent periods between consecutive talkspurts. Although there are methods which focus on delay-loss performance (Moon *et al.* 1998), better algorithms have been proposed along with voice quality prediction models for perceptual optimization of playout buffer (Fujimoto *et al.* 2002, Sun and Ifeachor 2006). Taking a different approach, MD coding (Jiang and Ortega 2000, Liang *et al.* 2001, Balam and Gibson 2007) exploits packet path diversity such that each description can be individually decoded for a reduced quality reconstruction, but if all descriptions are available, they can be jointly decoded for a better quality reconstruction. For multi-stream voice transmission, Liang *et al.* (2001) proposed an algorithm which uses the Lagrangian cost function to trade delay versus loss by following a play-first strategy; that is, it plays out early arriving descriptions while discarding the later ones. Such a design was based on the assumption that human perceptual experience is more strongly impaired by high latency than packet loss. They neither consider the quality degradation due to frequent switching among playout scenarios nor try to optimize the perceived speech quality by way of a prediction model.

Packet loss and delay are the major network impairments for transporting real-time voice over IP networks. Packet loss in MD voice transmission is a result of not only network loss, but also late loss, which greatly impairs communication quality. Due to the stringent delay budget and the need to output speech continuously, packets experiencing sudden high delay

*Corresponding author. Email: wwchang@cc.nctu.edu.tw

have to be discarded at the receiver end if they arrive later than the scheduled playout deadline. There has been much interest in the use of packet-level forward error control (FEC) to mitigate the impact of packet losses (Lin and Costello 2004). Most current FEC mechanisms send additional information along with the media stream so that the lost data can be recovered in part from the redundant information. In many applications, however, the losses of successive packets are correlated and a packet loss may be followed by a burst packet loss, which significantly decreases the efficiency of FEC. Furthermore, the loss recovery of FEC is performed at the cost of increased end-to-end delay. This has motivated our investigation into trying to exploit the largely uncorrelated characteristics of packet loss and delay variation on multiple network paths using a joint control of MD and FEC. With an MD scheme coded with FEC we have now more freedom to tradeoff delay, late loss and speech reconstruction quality. Traditionally, the study of FEC for loss recovery and playout buffer adaptation for jitter compensation have proceeded independently. Most packet-level FEC mechanisms send some redundant information along with the media stream so that the lost data can be recovered in part from the redundant information embedded in the later arriving packets. In waiting for the arrival of a minimum required number of packets at the receiving end, loss recovery is performed at the cost of increased end-to-end delay. In view of this potential limitation and the coupling between FEC and playout buffer adaptation (Rosenberg *et al*. 2000, Boutremans and Boudec 2003), there is a need to develop a joint FEC and playout control scheme such that the additional delay due to FEC application is dealt with in the same optimization framework as for regular MD schemes. Previous efforts towards linking FEC with playout buffer for single-stream transmission can be found in Boutremans and Boudec (2003), but the assumption on which their algorithm was based may limit its applicability. Specifically, it was assumed that the single-stream network over which the voice packets are sent delivers packets in sequence, and thus if a given packet arrives after its playout time, then all the following packets will also arrive after the playout time of the given packet. This line of reasoning has been challenged by a number of related studies (Kuo *et al*. 2001) that addressed the possibility of packets delivered out of sequence because of network jitter. As such, the joint FEC and playout control scheme proposed below will ignore the constraints imposed by the no-reordering assumption made in Boutremans and Boudec (2003).

The concept of perceptual optimization is usually realized through the use of E-model (International Telecommunication Union 2000a) to predict the conversational speech quality. However, the E-model does not consider the dynamics of transmission impairments because it relies on static transmission parameters such as average packet loss and average end-to-end delay. Thus, the E-model may make invalid predictions in dealing with the overall quality issues that MD transmission is focused on. For example, the E-model may only suit single-path transmission with two conceivable playout scenarios; i.e., total loss versus no-loss of packets. A third scenario, partial loss, however, would rise with MD transmission. That is, with multiple streams sent along two paths, if packets from one path experience erasure or excessive delay, packets from the other path can often be used to conceal the absence of the lost packets. Although the partial loss is concealed, the resulting degraded playout quality may not be. In dealing with such a reconstruction scheme, the E-model is expected to show two limitations. First, it may fail to register impairments due to reconstruction based on information from a single path as opposed to from both paths, when no packets from either path are lost. Moreover, the resulting detrimental effects that accompany the change in the playout scenarios may thus be ignored and harm its prediction of the overall quality. Recognizing this, we propose a new objective method for predicting the perceived quality of multi-stream voice transmission. In addition to delay and packet loss, the model also takes into account the quality impairments due to frequent switch of playout scenarios. Based on the new model, we then propose the use of minimum overall impairment as a criterion for perceptual optimization of joint playout buffer and FEC adjustment.

2. System implementation

The implementation procedure consisted of description generation and description transmission over two independent network paths. Figure 1 shows a block diagram of the system with the first two components, MD speech coder and channel coder, responsible for description generation and the rest, for transmission and signal reconstruction. For description generation, the MD-G.729 based speech packetization scheme described in Balam and Gibson (2007) was used to generate two descriptions from the bitstream of the ITU-T G.729 codec (International Telecommunication Union 2000b). G.729 is a conjugate-structure algebraic code-excited linear prediction codec for encoding

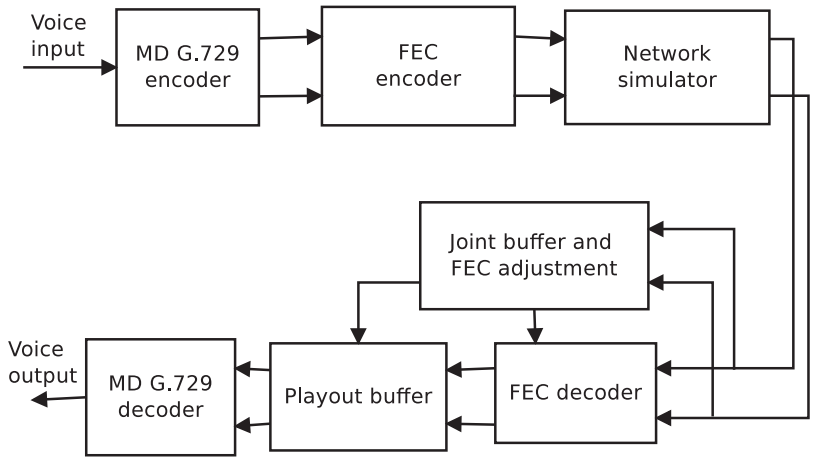

Figure 1. A multi-description voice transmission system.

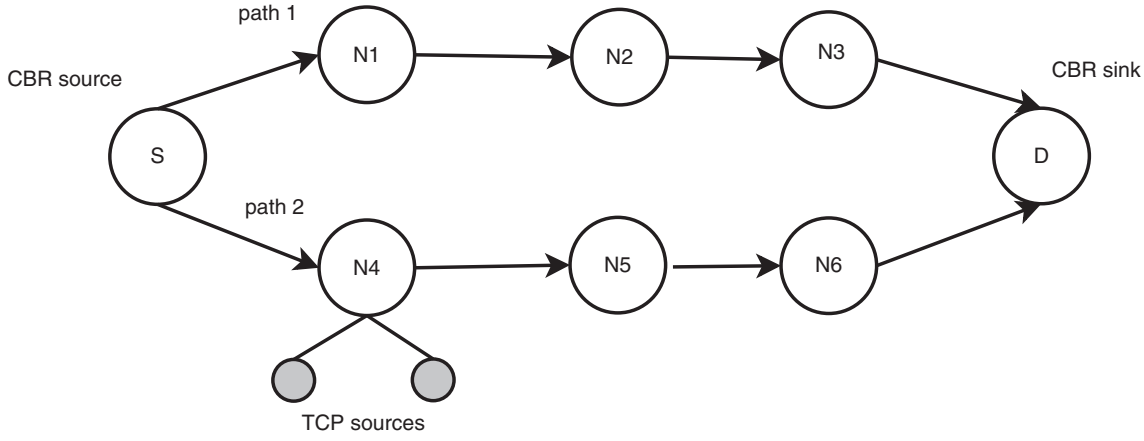

Figure 2. A multi-hop topology for network simulations.

narrowband speech at the rate of 8 kbps. It operates on 10-ms speech frames and each frame is divided into two subframes and all the parameters except the Linear Predictive Coding (LPC) coefficients are determined once per subframe. The MD-G.729 coder is designed to create two balanced descriptions; i.e., each description is of equal rate, 4.6 kbps and speech decoded from either description is of similar quality. After source coding, packet-level Reed–Solomon (N, K) codes (Lin and Costello 2004) are used for channel coding of individual descriptions. The channel encoder takes a codeword of K speech packets and generate $N-K$ additional FEC check packets for the transmission of N packets over the network. Such a code, denoted as an RS (N, K) code, is able to recover all losses in the block if and only if at least K out of N packets are received correctly.

During description transmission, the best-effort nature of IP networks results in packets experiencing varying amounts of loss and delay due to different levels of network congestion. To characterize this, we used the ns-2 network simulator (McCanne and Floyd 1997) to generate the traces of voice over inernet protocol (VoIP) traffic for different network topologies and varying network load. Meanwhile, traces were extended for varying link loss rates. A value ranging from 0% to 15% was used to simulate packet losses with different degrees of severity. Figure 2 shows the two path multi-hop network topology of our simulation, with transmission control protocol (TCP) data traffic on both paths contending simultaneously for network resources. The three nodes situated between source and destination on each path (N1 through N3 on the top path and N4 through N6 on the bottom) represent the data access points, each with a number of data sources attached, thus channelling in a large amount of incoming TCP traffic heading for different destinations. On each path a constant bit rate voice

stream is transmitted in 10-ms UDP packets at a rate of 4.6 kbps.

The receiver end features an adaptive playout buffer that smooths out the network jitter. The algorithm adjusts the playout buffer at the beginning of each talkspurt and subsequent packets of that talkspurt are played out with the generation rate at the sender. A joint design of FEC and playout buffer adaptation was further formulated as an optimization problem on the basis of a minimum overall impairment criterion. In addition to packet loss and delay that traditional systems sought to control, this design takes into account the dynamics of transmission impairments due to frequent switch of playout scenarios. As a prerequisite for obtaining impairments estimation on which the joint design could be based, a delay distribution model was established as it could provide a direct link to late loss rate in the presence of jitter. Previous work in Fujimoto *et al.* (2002) has found that the delay characteristics of VoIP traffic can be represented by statistical models which follow Pareto, normal and exponential distributions depending on applications. Finally, the MD-G.729 bitstream is decoded and degraded speech is generated. The decoder performs differently in dealing with the three description arrival situations: if both descriptions are lost, the error concealment algorithm (International Telecommunication Union 2000b) is used, while in other situations, speech packets are reconstructed depending on how many descriptions are received by the playout deadline. If both descriptions are received, the central decoder performs the standard G.729 decoding process after combining the two descriptions into one bitstream. If only one description is lost, the side decoder substitutes the missing information by using received parameters from the other description or information from the most recent correctly received frame (Balam and Gibson 2007).

3. Multi-stream voice quality prediction model

Conceptually, the proposed model followed the commonly used ITU E-model (International Telecommunication Union 2000a) in defining factors that affect the perceptual quality of the MD voice transmission. As an analytical model of conversational speech quality used for network planning purposes, the E-model combines individual impairments due to the signal's properties and the network characteristics into a single R-factor, ranging from 0 to 100. In VoIP applications (Cole and Rosenbluth 2001), the R-factor may be simplified as follows: $R = 94.2 - I_d - I_e$, where I_d represents the delay impairment. I_e is known as the equipment impairment and accounts for impairments due to speech coding and packet loss. The E-model, originally proposed for single-stream transmission, is only applicable to a limited number of speech codecs and network conditions, since deriving the I_e model requires time-consuming subjective tests. The delay impairment can be derived by a simplified fitting process in Cole and Rosenbluth (2001) with the following form

$$I_d(d) = 0.024d + 0.11(d - 177.3)H(d - 177.3), \quad (1)$$

where d is the end-to-end delay and $H(x)$ is the step function.

The task of defining the R-factor for multi-streams voice transmission lies in the fact that any subset can be used for signal reconstruction, and that the transmission quality improves with the size of the subsets. Thus, in addition to delay and packet loss, our prediction model aims to address the issue of impairments due to dynamic size allocations during the speech playout. For two-path transmission, each channel can either deliver or erase the transmitted description, so the two channels will always be in one of four possible states: no loss, loss in channel 1, loss in channel 2 and loss in both channels (packet erasure). Among them, only the speech resulting from the packet-erasure state is not affected by playout buffer operations. The receiver deals with the loss of both descriptions by using the error concealment algorithm of G.729 codec to conceal the erased packet. If, additionally, speech decoded from either MD-G.729 description is assumed to be of similar quality, we only need to consider two kinds of playout scenarios at the receiver end. Specifically, a packet is (1) fully restored with two descriptions and thus played with high quality and (2) partially restored with one description and thus played with degraded quality. For brevity, let $\mathbf{S}_k$ denote the scenario that k descriptions are received before the playout time. Conditioned on the event that the packet can be restored, we let r_k be the probability to play out the packet using k descriptions. Formally, it is given by $r_k = P(\mathbf{S}_k)/(P(\mathbf{S}_1) + P(\mathbf{S}_2))$. It is important to notice that quality degradation resulting from $\mathbf{S}_1$ and $\mathbf{S}_2$ are different perceptual experiences. Let $I_{e,k}$ denote the equipment impairment as a result of playing out k received descriptions. For scenario $\mathbf{S}_2$, the standard G.729 decoding process is carried out after combining the two descriptions into one bitstream. From the perceived QoS perspective, the MD-G.729 codec may be viewed as operating at two coding rates: 4.6 kbps for $\mathbf{S}_1$ and 8 kbps for $\mathbf{S}_2$. By taking frequent switches of coding rates into account, we define the

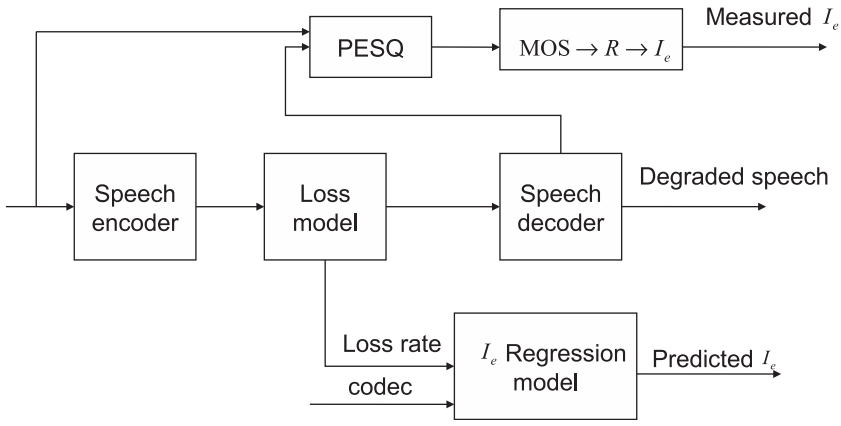

Figure 3. Schematic diagram for prediction of I_e model.

average equipment impairment due to MD-G.729 coding as follows:

$$I_e(e) = r_1 I_{e,1}(e) + r_2 I_{e,2}(e), \quad (2)$$

where e is the packet-erasure rate in percentage.

The next issue to be addressed is to derive an equipment impairment $I_{e,k}$ corresponding to each playout scenario $\mathbf{S}_k$. We followed the work of Sun and Ifeachor (2006), which describes an objective method for prediction of I_e regression model using the perceptual evaluation of speech quality (PESQ) algorithm (International Telecommunication Union 2001). As shown in Figure 3, each single measurement consists of three steps and is repeated several times with different transmission configurations. First, a speech sample is selected from an English speech database that contains 16 sentential utterances spoken by eight males and eight females. Each sample has a duration of 8 s and is sampled at 8 kHz. Second, the speech sample is encoded using MD-G.729 codec and then processed in accordance with the simulated loss model to generate the degraded speech. In our experiments, the decoder deals with packet erasure by using the error concealment algorithm of G.729 to conceal erased packets, while in other scenarios speech packets are reconstructed depending on how many descriptions are received by the playout deadline. Third, the reference speech and degraded speech are processed by the PESQ to obtain a mean opinion score (MOS). For each speech sample, a MOS value for one packet-erasure rate is obtained by averaging over 30 different erasure locations in order to remove the influence of erasure location. Further, these MOS values are averaged over all speech samples and then converted to a rating R to give an equipment impairment value $I_{e,k} = 94.2 - R$. The R-factor can be obtained from the average MOS with a conversion formula as follows:

$$R = 3.026MOS^3 - 25.314MOS^2 + 87.06MOS - 57.336. \quad (3)$$

Figure 4 shows that impact of transmission scenario $\mathbf{S}_k$ and packet-erasure rate e on the equipment impairment $I_{e,k}$ with a packetization of one frame per packet. The $I_{e,k}$ value for zero packet-erasure rate represents the codec impairment itself. It is obvious that the speech playout resulting from $\mathbf{S}_2$ has a lower codec impairment and has a high robustness to packet loss. From the curves, a nonlinear regression model can be derived for each $I_{e,k}$ by the least-squares data fitting method. The fitting curves are also shown in Figure 4. The derived $I_{e,k}$ model for scenario $\mathbf{S}_k$ has the following form: $I_{e,k}(e) = \gamma_{1,k} + \gamma_{2,k} \ln(1 + \gamma_{3,k} e)$. Our findings indicate that the regression model parameters $(\gamma_{1,k}, \gamma_{2,k}, \gamma_{3,k})$ for $\mathbf{S}_1$ are (52.61, 7.52, 10) and (21.96, 17.02, 16.09) for $\mathbf{S}_2$.

4. FEC in a Gilbert-model loss process

In Section 1, we stated the rationale for combining FEC into the playout buffer algorithm without following the no-reordering assumption underlying the work in Boutremans and Boudec (2003). Assume that MDs of the speech are transmitted over independent network paths and each path is characterized by a Gilbert-model loss process. The Gilbert model is a two-state Markov chain model in which state B represents a network loss and state G represents a packet reaching the destination. For each stream l, the parameters $p^{(l)}$ and $q^{(l)}$ denote, respectively, the probabilities of transitions from G to B states and from B to G states. A packet is said to be missing so long as the packet is either dropped in the network or discarded due to its

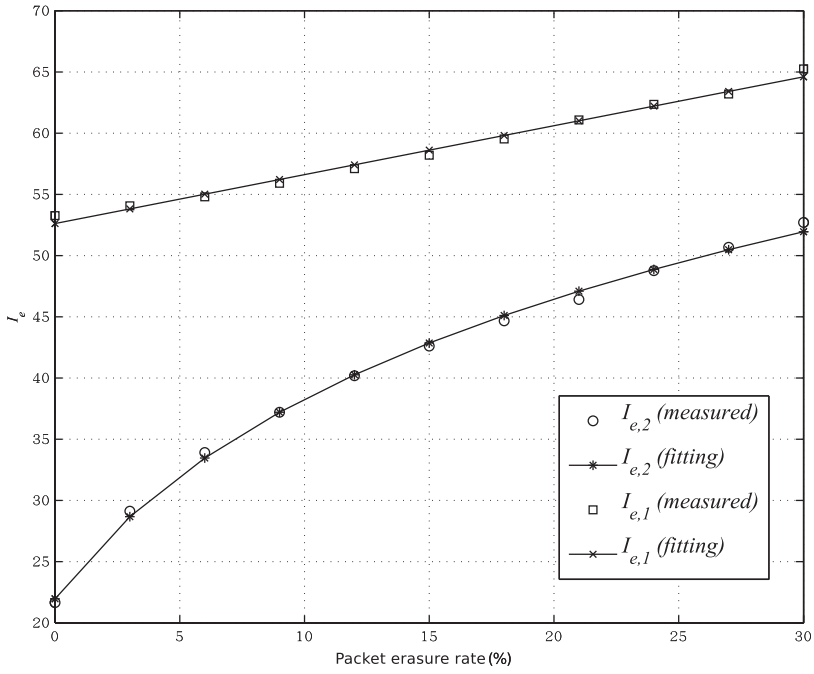

Figure 4. $I_{e,k}$ vs. packet-erasure rate e.

late arrival. For the sake of clarity, every packet i is assigned a variable $W_i \in \{0, 1, 2\}$, corresponding to the following three arrival scenarios: $W_i = 0$, arriving before its playout time, $W_i = 1$, a network loss and $W_i = 2$, arriving after its playout time. Following the development of Boutremans and Boudec (2003), let $R^{(l)}(m, n, D_{F,i})$ denote the probability that $m-1$ packets are missing (dropped or received late) in the next $n-1$ packets following the network loss of packet i, and let $S^{(l)}(m, n, D_{F,i})$ denote the probability that $m-1$ packets are missing in the next $n-1$ packets following the late loss of packet i. Similarly, let $\tilde{R}^{(l)}(m, n, D_{F,i})$ and $\tilde{S}^{(l)}(m, n, D_{F,i})$ denote the probability that $m-1$ missing packets occur in the last $n-1$ packets preceding packet i which is dropped and received late, respectively. As shown in the appendix, these probabilities can be computed by recurrence as follows:

$$
\begin{aligned}
&R^{(l)}(m,n,D_{F,i}) \\
&= \begin{cases}
q^{(l)}(1-p^{(l)})^{n-2} \cdot \prod_{h=1}^{n-1}(1-e_{b,i+h}^{(l)}), & m=1,\ n \geq 1 \\
(1-q^{(l)})R^{(l)}(m-1,n-1,D_{F,i+1}) & \\
\quad + \sum_{j=1}^{n-m}\Big\{ q^{(l)}(1-p^{(l)})^{j-1} \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}) & \\
\quad \cdot \big\{ p^{(l)} R^{(l)}(m-1,n-j-1,D_{F,i+j+1}) & \\
\quad\quad + (1-p^{(l)}) e_{b,i+j+1}^{(l)} S^{(l)}(m-1,n-j-1, & \\
\quad D_{F,i+j+1}) \big\} \Big\}, & 2 \leq m \leq n,
\end{cases}
\end{aligned}
\tag{4}
$$

and

$$
\begin{aligned}
&S^{(l)}(m,n,D_{F,i}) \\
&= \begin{cases}
e_{b,i}^{(l)}(1-p^{(l)})^{n-1} \cdot \prod_{h=1}^{n-1}(1-e_{b,i+h}^{(l)}), & m=1, n \geq 1 \\
\sum_{j=0}^{n-m}\Big\{ e_{b,i}^{(l)}(1-p^{(l)})^{j} \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}) & \\
\quad \cdot \big\{ p^{(l)} R^{(l)}(m-1,n-j-1,D_{F,i+j+1}) & \\
\quad + (1-p^{(l)}) e_{b,i+j+1}^{(l)} S^{(l)}(m-1,n-j-1,D_{F,i+j+1}) \big\} \Big\}, & \\
\quad 2 \leq m \leq n, &
\end{cases}
\end{aligned}
\tag{5}
$$

where $D_{F,i}$ is the FEC delay and $e_{b,i}^{(l)}$ is the estimated late loss probability of packet i in stream l. Table 1 summarizes the basic nomenclature used in the appendix.

With RS (N, K) code, each code takes a codeword of K voice packets and generates $N-K$ additional FEC packets for the transmission of N packets over the network. Such a code is able to recover any missing packet in the block if and only if at least K out of N packets in this block are received before their playout time. Viewed from this perspective, the probability to recover a dropped packet is given by

$$
P_{R1}^{(l)}(i) = \Pr(\text{packet } i \text{ can be recovered} \mid \text{packet } i \text{ is dropped in the network})
$$

Table 1. Basic nomenclature.

Nomenclature	Description
$D_{F,i}$	FEC delay of packet i
e_i	Packet-erasure probability of packet i
$e_{b,i}^{(l)}$	Late loss probability of packet i in stream l
$P_L^{(l)}(i)$	Resudual loss probability of packet i in stream l after FEC is used
$P_{R1}^{(l)}(i)$	Probability to recover a dropped packet i in stream l
$P_{R2}^{(l)}(i)$	Probability to recover a late lost packet i in stream l
$R^{(l)}(m, n, D_{F,i})$	Probability that $m-1$ packets are missing in the next $n-1$ packets following the network loss of packet i
$S^{(l)}(m, n, D_{F,i})$	Probability that $m-1$ packets are missing in the next $n-1$ packets following the late loss of packet i
$\tilde{R}^{(l)}(m, n, D_{F,i})$	Probability that $m-1$ packets are missing in the last $n-1$ packets proceeding the network loss of packet i
$\tilde{S}^{(l)}(m, n, D_{F,i})$	Probability that $m-1$ packets are missing in the last $n-1$ packets proceeding the late loss of packet i

$$
\begin{aligned}
&= \sum_{L=1}^{N-K} \Pr(L \text{ packets are missing in } W_1^N | W_i = 1) \\
&= \sum_{L=1}^{N-K} \sum_{m=0}^{\min(L-i,i-1)} \Pr(\ m \text{ packets are missing} \\
&\quad \text{in } W_1^{i-1} | W_i = 1) \cdot \Pr(\ L-m-1 \text{ packets} \\
&\quad \text{are missing in } W_{i+1}^N | W_i = 1) \\
&= \sum_{L=1}^{N-K} \sum_{m=0}^{\min(L-i,i-1)} \tilde{R}^{(l)}(m+1, i, D_{F,i}) \\
&\quad \cdot R^{(l)}(L-m, N-i+1, D_{F,i}),
\end{aligned} \tag{6}
$$

and the probability to recover a late lost packet is given by

$$
\begin{aligned}
P_{R2}^{(l)}(i) &= \Pr(\text{packet } i \text{ can be recovered} \mid \text{packet } i \\
&\qquad \text{is received late}) \\
&= \sum_{L=1}^{N-K} \Pr(L \text{ packets are missing in } W_1^N | W_i = 2) \\
&= \sum_{L=1}^{N-K} \sum_{m=0}^{\min(L-i,i-1)} \Pr(m \text{ packets are missing} \\
&\qquad \text{in } W_1^{i-1} | W_i = 2) \cdot \Pr(L-m-1 \text{ packets are} \\
&\qquad \text{missing in } W_{i+1}^N | W_i = 2) \\
&= \sum_{L=1}^{N-K} \sum_{m=0}^{\min(L-i,i-1)} \tilde{S}^{(l)}(m+1, i, D_{F,i}) \\
&\qquad \cdot S^{(l)}(L-m, N-i+1, D_{F,i}).
\end{aligned} \tag{7}
$$

Using these probabilities, we can compute the residual loss probability (after FEC is used) as follows:

$$
P_L^{(l)}(i) = e_n^{(l)}(1 - P_{R1}^{(l)}(i)) + (1 - e_n^{(l)}) e_{b,i}^{(l)}(1 - P_{R2}^{(l)}(i)), \tag{8}
$$

where $e_n^{(l)}$ represents the network loss probability measured in stream l. The packet-erasure probability e_i is defined as the probability that none of the descriptions of packet i arrives on time, and is given by

$$
e_i = \prod_{l=1}^{2} P_L^{(l)}(i). \tag{9}
$$

5. Joint FEC and playout control

The main attraction of multi-stream transmission arises from its flexibility in trading different sources of impairments against each other. Waiting for the arrival of both descriptions results in lower equipment impairment, but at the cost of higher delay impairment. On the other hand, playing out the voice description with lower delay avoids latency, but increases the equipment impairment. Since playout scheduling aims to improve the overall conversational speech quality, which hangs on the balance between delay and packet loss, full reconstruction of both descriptions may not always be the priority if the overall impairment does not justify the extra delay from waiting. Given that, the joint playout and FEC control must play around with switching between different playout scenarios in order to maximize the benefits of packet path diversity. To accomplish this goal, we formulated the system design as a perceptually motivated optimization problem and the adopted criterion relies on the use of the proposed multi-stream voice quality prediction model. Our efforts began by estimating the playout delay, which is defined as the time from the moment that packet is delivered to the network until it has to be played out. We applied an autoregressive algorithm (Moon *et al.* 1998) to

estimate the mean $\hat{d}$ and variance $\hat{v}$ of network delay, and use them to calculate the buffer delay $d_b = \hat{d} + \beta\hat{v}$. Waiting for the FEC check packets results in additional delay and, consequently, the playout delay is given by

$$d_{\text{play}} = \hat{d} + \beta\hat{v} + (N-1)T_p, \tag{10}$$

where T_p is the packet generation interval. The parameter β has a critical impact on the tradeoff between delay and late packet loss, which in turn influences the conversational speech quality. From Equation (10) it can be deduced that increasing β leads to lower late loss rate as more packets arrive in time, and yet the end-to-end delay also increases. Most playout buffer algorithms (Moon *et al*. 1998, Fujimoto *et al*. 2002, Sun and Ifeachor 2006) used a fixed value of β; e.g., $\beta = 4$, to set the buffer size, so that only a small fraction of the arriving packets should be lost due to late arrival. In this work, a β-adaptive algorithm is instead used to control the buffer size so that the reconstructed voice quality is maximized in terms of delay and loss.

Our general problem can be stated as follows: given estimates of the parameters characterizing the packet loss and delay distribution, find the optimal values of β and $\{N, K\}$ so as to minimize the overall impairment function subject to the rate constraint. Let d_i be the end-to-end delay experienced by the ith packet, which consists of encoding delay d_c and playout delay d_{play}. Now, we define an overall impairment function I_m with the following form:

$$I_m(d_i, e_1^K) = I_d(d_i) + \frac{1}{K}\sum_{j=1}^{K}\sum_{l=1,2} r_l, I_{e,l}(e_j), \tag{11}$$

where $e_1^K = (e_1, \ldots, e_K)$, $r_1 + r_2 = 1$ and the probability to receive both descriptions is given by

$$r_2 = \frac{1}{1-e_i}\prod_{l=1}^{2}(1 - P_L^{(l)}(i)). \tag{12}$$

Our optimization framework requires an analytic expression for the packet erasure probability e_i as a function of the parameter β. Notice that $e_{b,i}^{(l)}$ and the playout delay d_{play} are strongly correlated, and to find out their relationship, the network delays of stream l are assumed to follow a Pareto distribution which is defined as $F_D^{(l)}(d) = 1 - (g_l/d)^{\alpha_l}$. The parameters of Pareto distribution α_l and g_l can be estimated from past recorded delays using the maximum likelihood estimation method (Fujimoto *et al*. 2002). More specifically, given a set of past network delays $\{n_{i-1}^{(l)}, n_{i-2}^{(l)}, \ldots, n_{i-M}^{(l)}\}$, we compute $g_l = \min\{n_{i-1}^{(l)}, n_{i-2}^{(l)}, \ldots, n_{i-M}^{(l)}\}$ and $\alpha_l = M/\Sigma_{j=i-1}^{i-M}\log(\frac{n_j^{(l)}}{g_l})$. Then, the late loss probability of packet i in stream l can be computed as follows:

$$e_{b,i}^{(l)} = 1 - F_D^{(l)}(D_{F,i}) = (g_l/D_{F,i})^{\alpha_l}, \tag{13}$$

where $D_{F,i} = d_{\text{play}} - (i-1)T_p$. This reduces the expression of the packet-erasure probability e_i to be a function of the playout delay d_{play}, which in turn is a function of the parameter β.

Finally, we summarize the proposed multi-stream joint playout and FEC adjustment algorithm as below.

(1) Apply an autoregressive algorithm (Moon *et al*. 1998) to estimate the delay mean $\hat{d}_i^{(l)}$ and variance $\hat{v}_i^{(l)}$ for individual stream l ($l = 1, 2$) as follows:

$$\hat{d}_i^{(l)} = \mu\hat{d}_{i-1}^{(l)} + (1-\mu)n_i^{(l)}, \tag{14}$$

$$\hat{v}_i^{(l)} = \mu\hat{v}_{i-1}^{(l)} + (1-\mu)|n_i^{(l)} - \hat{d}_i^{(l)}|, \tag{15}$$

where $n_i^{(l)}$ is the network delay of packet i in stream l and $\mu = 0.998002$ is a weighting factor for convergence control.

(2) At the beginning of each talkspurt, update network delay records for the past $M = 200$ packets in every stream l ($l = 1, 2$), and use them to calculate the Pareto distribution parameters (α_l, g_l) by the maximum likelihood estimation method.

(3) Use the values of (α_l, g_l) to compute the late loss probability in Equation (13) and the packet-erasure probability e_i in Equation (9). Apply an exhaustive search method to determine the minimizer $(\hat{\beta}_i^{(l)}, \hat{N}^{(l)}, \hat{K}^{(l)})$ of the overall impairment function in Equation (11) subject to the code rate constraint $\frac{N}{K} \times \frac{9.2}{8} \le R_{\max}$. Here, the maximum overall code rate $R_{\max}$ is chosen to be 2.

(4) Set the playout delay and RS code parameters to

$$d_{\text{play}} = \hat{d}^{(l^*)} + \hat{\beta}_i^{(l^*)}\hat{v}^{(l^*)} + (\hat{N}^{(l^*)} - 1)T_p,$$
$$(N, K) = (\hat{N}^{(l^*)}, \hat{K}^{(l^*)}), \tag{16}$$

with $l^* = \arg\min\{I_m(\hat{\beta}^{(l)}, \hat{N}^{(l)}, \hat{K}^{(l)}), l = 1, 2\}$.

6. Experimental results

A set of experimental conditions was designed for the use of artificially degraded speech samples to verify the detrimental effects estimated by the proposed I_e

regression model in relation to the traditional E-model. The two models, despite their agreement in including packet loss as a main impairment factor, differ in how reconstruction in conditions with partial packet losses is treated. The proposed model differentiates partial reconstruction with one description from full reconstruction with two descriptions. The three states of frame reconstruction dictated by the model are (1) fully restored, when both descriptions are available and thus played with high quality, (2) partially restored, when only one description is available and thus played with less than optimal quality and (3) restored by the G.729 error concealment algorithm, when both descriptions are lost during transmission. In contrast, the traditional model treats the full and the partial reconstruction states uniformly as the no-loss state, leaving out any differentiation of the processes involved that lead to the no-loss at the receiver end. It is thus reasonable to hypothesize that the traditional model fails to register any quality impairment due to partial reconstruction. As such, if the I_e's estimated with the two models show significant differences in their closeness to the I_e's measured, then adding such a differentiation scheme into the modelling process should prove a valid approach. The speech samples considered here were one male and one female utterance. The G.729 speech codec and the proposed MD coding scheme were used sequentially, which turned each utterance into a bitstream of frames with two identical descriptions to be transmitted along separate dynamically changing paths. At the receiver end, each utterance was artificially degraded to render two tokens, each with its own composition of frames of the three reconstruction states. Since the proposed model diverges from the traditional model by treating the loss of one packet as a separate state from either total loss or no loss, the underlying variable being manipulated in the frame composition was the rate r_1 of partial loss. Thus, there was a total of four test conditions.

Table 2 lists for each condition the percentages of frames that are erased and restored with only one description, followed by the three corresponding I_e's as estimated by the traditional model, by the proposed model and as measured then converted with PESQ. The results showed that, unlike the traditional model that yielded poorer estimations for samples containing higher percentages of one description loss, the proposed model gave estimations that are quite robust regardless of the sample frame composition. For example, given the same percentage increases from 6.84% to 22% and from 14% to 31% in the female and the male utterance, respectively, the traditional model showed deviations from the measured I_e's that were increased from 1.38 to 6.43 and from 4.79 to 5.9, respectively, while the proposed model yielded across all conditions, more stable and smaller deviations that ranged from 0.6 to 1.6. Taken together, these comparison data suggest that independent evaluation of impairments due to loss of one versus both descriptions adds to the robustness of the proposed model.

Table 2. I_e comparison for different prediction models.

Speech	e (%)	r_1 (%)	Traditional I_e	Proposed I_e	Measured I_e
Female	9.88	6.48	43.16	44.35	44.54
	4.93	22	34.41	39.48	40.84
Male	4.84	14	31.78	34.97	36.57
	12	31	40.37	45.67	46.27

Computer simulations were carried out to evaluate the performances given by the four MD voice transmission schemes, MD1–4, which all used the MD-G.729 for source coding and RS(N, K) code for channel coding. The speech data fed into the simulations were two sentential utterances spoken by one male and one female, each sampled at 8 kHz and 8 s in duration. Both samples were encoded and then processed in accordance with the delay and loss characteristics of the trace data to degrade the speech. Among the four schemes, MD1 had its parameters $\{\beta, N, K\}$ dynamically adjusted according to the proposed voice quality prediction model, while MD2–4 shared a fixed $\beta = 4$ with (N, K) set at $(3, 2)$, $(5, 3)$ and $(10, 6)$, respectively. It should be pointed out that the last two (N, K) sets allowed MD3 and MD4 to perform at the same FEC coding ratio but with different lengths of delay, which gave us the opportunity to evaluate in our test environment the effect of packet loss versus delay. It was hypothesized that the performances of these schemes would be set apart mainly by the values of $\{\beta, N, K\}$ they each assumed, and that the best performance should come with the adaptive parameter adjustment scheme, or MD1 in the current case, whose calculation was based on link loss, packet-erasure loss and various transmission scenarios.

The performances of MD transmission schemes were also compared with an FEC-protected SD transmission scheme, which consists of an 8 kbps G.729 speech coder followed by an RS(N, K) channel coder. Following the work of Boutremans and Boudec (2003), the SD scheme applied a joint playout buffer and FEC adjustment scheme which jointly chooses both the playout delay d_{play} and the FEC scheme RS(N, K) so as to maximize the perceived voice quality. Figure 5 plots the perceived speech quality associated with the SD and four MD schemes for the case where the network

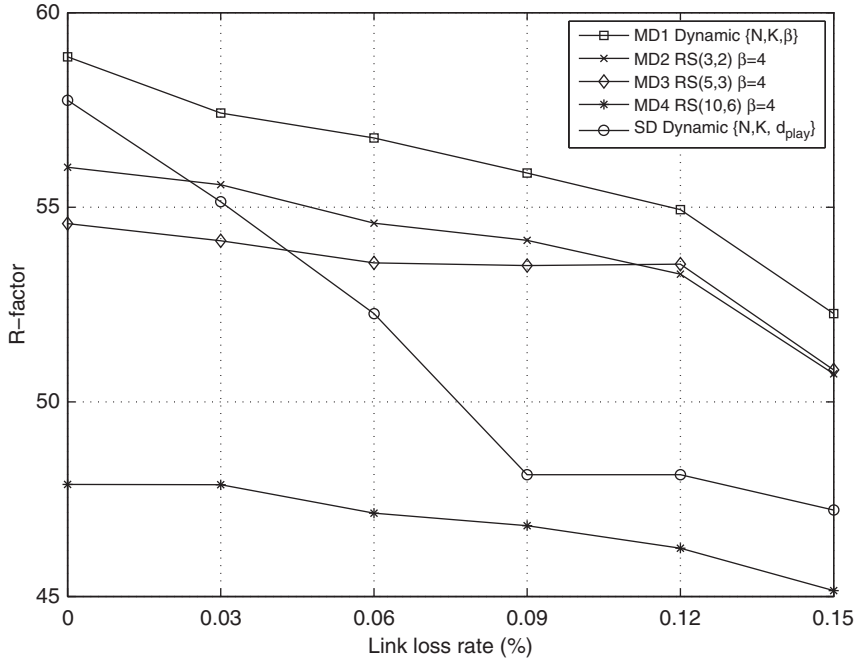

Figure 5. Performance comparison for different playout algorithms.

paths are subjected to Gilbert-model loss process with link loss rate ranging from 0% to 15%. As described in Section 2, the perceived quality was gauged by calculating the predicted average *R*-factor according to the E-model. It can be seen that the *R*-factor was decreased as the link loss rate was increased regardless of the scheme used. When applying a joint playout buffer and FEC control scheme, the results obtained using the MD1 clearly demonstrated an improvement over those obtained using the SD scheme, especially at high link loss rates. At link loss rates slightly beyond 6%, the SD scheme, despite its FEC feature, started showing incapability to recover the lost packets in the facing Gilbert-model link loss process. Among the four MD schemes, MD4, with the longest end-to-end delay, yielded the lowest *R*-factors, while MD3, with the same FEC coding ratio but shorter delays than those set for MD4, yielded higher *R*-factors than MD4, but lower *R*-factors than MD2. MD2 with the lower delay impairment allowed it to outperform MD3 and MD4, but its strength of packet recovery, as seen in Figure 5, receded faster as the link loss rate was increased, and at link loss rates greater than 12%, yielded lower *R*-factors than MD3. The best results in the plot, as hypothesized, were obtained with the currently proposed scheme MD1. Table 3 presents some of the varying parameters that shaped its performance and demonstrates the dynamic aspects of this scheme. At link loss rate = 12%, 10.25% of the descriptions were recovered with $(N, K) = (5, 3)$ while 89.7% (the rest) were recovered with $(N, K) = (3, 2)$; when the loss was increased to 15%, 25.6% of the descriptions were recovered with $(N, K) = (5, 3)$ and 74.35% (the rest) were recovered with $(N, K) = (3, 2)$. The average redundant bits thus obtained at the two link loss rates were $1.14 (= 10.25\% \cdot 2 + 89.7\% \cdot 1)$ and $1.255 (= 25.6\% \cdot 2 + 74.35\% \cdot 1)$, respectively. The plot showed that these settings allowed MD1 to outperform schemes with fixed settings in view of the transmission scenarios during testing. It follows that in multi-stream voice transmission scheme design, the pursuit of high performance of FEC does not guarantee high perceptual speech quality if delay fails to be jointly considered. The best performance seen in MD1 should therefore be taken as evidence attesting to the superiority of using an all encompassing algorithm proposed here that aims to lower the total impairment impacts by making adjustments adaptive to the on-going

Table 3. Average redundant bits comparison for different link loss rates.

Link loss rate (%)	RS(3,2)	RS(5,3)	Average redundant bits
12	89.7%	10.25%	1.14
15	74.35%	25.6%	1.255

interplay of delay, packet-erasure loss and various transmission scenarios.

7. Conclusion

In this article, we have proposed a perceptually motivated optimization criterion and a practically feasible new algorithm for multi-stream voice transmission. We start by considering the perceived voice quality as a function of playout scenario, the packet-erasure rate and the end-to-end delay. Adaptive joint playout buffer and FEC adjustment is then formulated as an optimization problem leading to the minimum overall impairment. Experimental results show that the proposed multi-stream voice transmission scheme can achieve a better delay-loss tradeoff and thereby improves the perceived speech quality.

Acknowledgment

This study was supported by the National Science Council, Republic of China, under contract NSC 96-2221-E-009-031-MY3.

References

Balam, J. and Gibson, J.D., 2007. Multiple descriptions and path diversity for voice communications over wireless mesh networks. *IEEE transactions on multimedia*, 9 (5), 1073–1088.

Boutremans, C. and Boudec, J., 2003. Adaptive joint playout buffer and FEC adjustment for internet telephony. *In*: *Proceedings of IEEE INFOCOM*, March. Vol. 1, San Francisco, CA, 658–662.

Cole, R. and Rosenbluth, J., 2001. Voice over IP performance monitoring. *ACM SIGCOMM computer communication reviews*, 31 (2), 9–24.

Fujimoto, K., Ata, S., and Murata, M., 2002. Adaptive playout buffer algorithm for enhancing perceived quality of streaming applications. *In*: *Proceedings of IEEE globecom*, November. Vol. 3, Taipei, Taiwan, 2451–4257.

International Telecommunication Union, 2000a. The E-model, a computational model for use in transmission planning, *ITU-T Recommendation G.107*.

International Telecommunication Union, 2000b. Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), *ITU-T Recommendation G.729*.

International Telecommunication Union, 2001. Perceptual evaluation of speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs *ITU-T Recommendation P.862*.

Jiang W. and Ortega A., 2000. Multiple description speech coding for robust communication over lossy packet networks. *In*: *International conference on multimedia and expo*, July, New York, USA, Vol. 1, 444–447.

Kuo, C.-C., Chen, M.-S., and Chen, J.-C., 2001. An adaptive transmission scheme for audio and video synchronization based on real-time transport protocol. *In*: *IEEE International conference on multimedia and expo*, August, Tokyo, Japan, 403–406.

Liang, Y.J., Steinbach, E.G., and Girod, B., 2001. Multi-stream voice over IP using packet path diversity. *In*: *Proceedings of IEEE fourth workshop on multimedia signal processing*, October. Cannes, France, 555–560.

Lin, S. and Costello, D.J., 2004. *Error control coding*. New Jersey: Pearson Prentice Hall.

McCanne, S. and Floyd, S., 1997. *Network Simulator ns-2*. Available from: http://www.isi.edu/nsnam/ns/ [Accessed 24 August 2009].

Moon, S.B., Kurose, J., and Towsley, D., 1998. Packet audio playout delay adjustment: performance bounds and algorithms. *Multimedia systems*, 6 (1), 17–28.

Rosenberg, J., Qiu, L., and Schulzrinne, H., 2000. Integrating packet FEC into adaptive voice playout buffer algorithms on the internet. *In*: *Processing IEEE INFOCOM 2000*, March, Tel Aviv. Vol. 3, Israel, 1705–1714.

Sun, L. and Ifeachor, E., 2006. Voice quality prediction models and their application in VoIP networks. *IEEE transactions on multimedia*, 8 (4), 809–820.

Appendix

This section gives the detailed computation of $R^{(l)}(m,n,D_{F,i})$ and $S^{(l)}(m,n,D_{F,i})$ when (1) a Reed–Solomon code (N,K) is used, (2) packets are sent over a Gilbert channel and (3) the FEC delay of packet i is $D_{F,i}$. For $m=1$, $n\geq 1$, $R^{(l)}(1,n,D_{F,i})$ is the probability that none of the packets are missing in the next $n-1$ packets following the network loss of packet i, and is given by

$$\begin{aligned} R^{(l)}(1,n,D_{F,i}) &= \Pr(W_{i+1}^{i+n-1}=0^{n-1}|W_i=1) \\ &= q^{(l)}(1-p^{(l)})^{n-2}\cdot\prod_{h=1}^{n-1}(1-e_{b,i+h}^{(l)}). \end{aligned} \tag{A1}$$

For $2\leq m\leq n$, we compute $R^{(l)}(m,n,D_{F,i})$ conditionally to the event $\{A_j,\ B_j,\ C_j,\ j=0,1,\ldots,n-m\}$ on the arriving states of packets:

$$\begin{aligned} A_j &= \{W_i^{i+j+1}=10^j1\}, \\ B_j &= \{W_i^{i+j+1}=10^j2\}, \\ C_j &= \left\{m-2 \text{ missing packets in } W_{i+j+2}^{i+n-1}\right\}, \end{aligned} \tag{A2}$$

where 0^j is a shorthand for j successive 0's. For a Gilbert loss model with parameters $p^{(l)}$ and $q^{(l)}$, we have

$$\Pr(A_j)=\begin{cases}(1-q^{(l)}), & j=0,\\ q^{(l)}(1-p^{(l)})^{j-1}p^{(l)}\prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}), & j\geq 1,\end{cases} \tag{A3}$$

$$\Pr(B_j) = q^{(l)}(1-p^{(l)})^j \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)})e_{b,i+j+1}^{(l)}, \quad j \geq 1, \quad \text{(A4)}$$

$$\begin{aligned}\Pr(C_j|A_j) &= \Pr(m-2 \text{ missing packets in } W_{i+j+2}^{i+n-1} | W_i^{i+j+1} = 10^j 1) \\ &= R^{(l)}(m-1, n-j-1, D_{F,i+j+1}), \end{aligned} \quad \text{(A5)}$$

$$\begin{aligned}\Pr(C_j|B_j) &= \Pr(m-2 \text{ missing packets in } W_{i+j+2}^{i+n-1} | W_i^{i+j+1} = 10^j 2) \\ &= S^{(l)}(m-1, n-j-1, D_{F,i+j+1}). \end{aligned} \quad \text{(A6)}$$

From the total probability theorem, $R^{(l)}(m,n,D_{F,i})$ can be computed as follows:

$$\begin{aligned} &R^{(l)}(m,n,D_{F,i}) \\ &= \sum_{j=0}^{n-m} \Pr(C_j|A_j)\Pr(A_j) + \Pr(C_j|B_j)\Pr(B_j) \\ &= (1-q^{(l)})R^{(l)}(m-1, n-1, D_{F,i+1}) \\ &\quad + \sum_{j=1}^{n-m} \Bigg\{ q^{(l)}(1-p^{(l)})^{j-1} \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}) \\ &\quad \cdot \left\{ p^{(l)} R^{(l)}(m-1, n-j-1, D_{F,i+j+1}) \right. \\ &\quad \left. + (1-p^{(l)}) e_{b,i+j+1}^{(l)} S^{(l)}(m-1, n-j-1, D_{F,i+j+1}) \right\} \Bigg\}. \end{aligned} \quad \text{(A7)}$$

Similarly, the probability $\tilde{R}^{(l)}(m,n,D_{F,i})$ can also be computed by recurrence as

$$\begin{aligned} &\tilde{R}^{(l)}(m,n,D_{F,i}) \\ &= \begin{cases} q^{(l)}(1-p^{(l)})^{n-2} \cdot \prod_{h=1}^{n-1}(1-e_{b,i-h}^{(l)}), & m=1, n \geq 1 \\ (1-q^{(l)})R^{(l)}(m-1, n-1, D_{F,i+1}) & \\ \quad + \sum_{j=1}^{n-m} \Bigg\{ q^{(l)}(1-p^{(l)})^{j-1} \prod_{h=1}^{j}(1-e_{b,i-h}^{(l)}) & \\ \quad \cdot \left\{ p^{(l)} \tilde{R}^{(l)}(m-1, n-j-1, D_{F,i}) \right. & \\ \quad \left. + (1-p^{(l)}) e_{b,i-j-1}^{(l)} \tilde{S}^{(l)}(m-1, n-j-1, D_{F,i}) \right\} \Bigg\}, & \\ \quad 2 \leq m \leq n. \end{cases} \end{aligned} \quad \text{(A8)}$$

Next, we give the detailed computation of $S^{(l)}(m,n,D_{F,i})$. For $m=1$, $n \geq 1$, $S^{(l)}(1,n,D_{F,i})$ is the probability that none of the packets are missing in the next $n-1$ packets following the late loss of packet i, and is given by

$$\begin{aligned} S^{(l)}(1,n,D_{FEC,i}) &= \Pr(W_{i+1}^{i+n-1} = 0^{n-1} | W_i = 2) \\ &= e_{b,i}^{(l)}(1-p^{(l)})^{n-1} \cdot \prod_{h=1}^{n-1}(1-e_{b,i+h}^{(l)}). \end{aligned} \quad \text{(A9)}$$

For $2 \leq m \leq n$, we compute $S^{(l)}(m,n,D_{F,i})$ conditionally to the event $\{C_j, D_j, E_j, j=0,1,\ldots,n-m\}$ on the arriving states of packets:

$$\begin{aligned} C_j &= \left\{ m-2 \text{ missing packets in } W_{i+j+2}^{i+n-1} \right\}, \\ D_j &= \left\{ W_i^{i+j+1} = 20^j 1 \right\}, \\ E_j &= \left\{ W_i^{i+j+1} = 20^j 2 \right\}. \end{aligned} \quad \text{(A10)}$$

For a Gilbert loss model with parameters $p^{(l)}$ and $q^{(l)}$, we have

$$\Pr(D_j) = e_{b,i}^{(l)}(1-p^{(l)})^j p^{(l)} \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}), \quad \text{(A11)}$$

$$\Pr(E_j) = e_{b,i}^{(l)}(1-p^{(l)})^{j+1} \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}) e_{b,i+j+1}^{(l)}, \quad \text{(A12)}$$

$$\begin{aligned} \Pr(C_j|D_j) &= P(m-2 \text{ missing in } W_{i+j+2}^{i+n-1} | W_i^{i+j+1} = 20^j 1) \\ &= R^{(l)}(m-1, n-j-1, D_{F,i+j+1}), \end{aligned} \quad \text{(A13)}$$

$$\begin{aligned} \Pr(C_j|E_j) &= P(m-2 \text{ missing packets in } W_{i+j+2}^{i+n-1} | W_i^{i+j+1} = 20^j 2) \\ &= S^{(l)}(m-1, n-j-1, D_{F,i+j+1}). \end{aligned} \quad \text{(A14)}$$

From the total probability theorem, $S^{(l)}(m,n,D_{F,i})$ can be computed as follows:

$$\begin{aligned} &S^{(l)}(m,n,D_{F,i}) \\ &= \sum_{j=0}^{n-m} \Pr(C_j|D_j)\Pr(D_j) + \Pr(C_j|E_j)\Pr(E_j) \\ &= \sum_{j=0}^{n-m} \Bigg\{ e_{b,i}^{(l)}(1-p^{(l)})^j \prod_{h=1}^{j}(1-e_{b,i+h}^{(l)}) \\ &\quad \cdot \left\{ p^{(l)} R^{(l)}(m-1, n-j-1, D_{F,i+j+1}) \right. \\ &\quad \left. + (1-p^{(l)}) e_{b,i+j+1}^{(l)} S^{(l)}(m-1, n-j-1, D_{F,i+j+1}) \right\} \Bigg\}. \end{aligned} \quad \text{(A15)}$$

Similarly, $\tilde{S}^{(l)}(m,n,D_{F,i})$ can be computed by recurrence as

$$\begin{aligned} &\tilde{S}^{(l)}(m,n,D_{F,i}) \\ &= \begin{cases} e_{b,i}^{(l)}(1-p^{(l)})^{n-1} \cdot \prod_{h=1}^{n-1}(1-e_{b,i-h}^{(l)}), & m=1,\ n \geq 1 \\ \sum_{j=0}^{n-m} \Bigg\{ e_{b,i}^{(l)}(1-p^{(l)})^j \prod_{h=1}^{j}(1-e_{b,i-h}^{(l)}) & \\ \quad \cdot \left\{ p^{(l)} \tilde{R}^{(l)}(m-1, n-j-1, D_{F,i-j-1}) \right. & \\ \quad \left. + (1-p^{(l)}) e_{b,i-j-1}^{(l)} \tilde{S}^{(l)}(m-1, n-j-1, D_{F,i-j-1}) \right\} \Bigg\}, & \\ \quad 2 \leq m \leq n. \end{cases} \end{aligned} \quad \text{(A16)}$$