Content-Aware Packet-Level Interleaving Method for Video Transmission over Wireless Networks

Jeong-Yong Choi and Jitae Shin

School of Information and Communication Engineering, Sungkyunkwan University, Suwon, 440-746, Korea
Email: {eldragon,jtshin}@ece.skku.ac.kr

Abstract. In the wireless network environment, the effect of transmission errors and losses on the video quality varies depending on the intensity of the burst and which parts of the video stream are lost. Among the existing transmission error control techniques, FEC and ARQ are good solutions for combating transmission errors, but they require redundant data. Although interleaving has no error correcting capability, it can improve subjective video quality without wasting additional bandwidth, because it allows the spreading of successive errors. In this paper, we propose a content-aware packet-level interleaving method, which uses a quantitative index to indicate the degree of content-importance of the video content, so that the effect of burst packet losses is distributed intelligently. The proposed scheme improves the overall video quality in comparison with content-blind interleaving methods.

Key words:interleaving, content-aware, wireless network, video transmission

1 Introduction

Two or one way streaming over unreliable and error-prone wireless channel is one of the major challenges for wireless video applications. There are many research efforts in wireless communication area to combat transmission error/loss over wireless channel such as channel coding, modulation, interleaving, etc. Besides these bit-level error control methods, there is also a need for application-aware techniques in order to provide multi-class service for diverse multimedia traffics, e.g., the four classes in Universal Mobile Telecommunications System (UMTS), viz. the conversational, streaming, interactive and background traffics. In addition, transmission error/loss control (TEC) needs to be both application content-aware and wireless channel-aware, therefore necessitating a cross-layer approach, which combines application quality of service (QoS) requests or content priorities and channel condition information.

Classic bit-level TEC in wireless communications can compensate for the time-varying channel effect, but still lacks the ability to accomplish the efficient transmission of diverse multimedia applications over packet-switched networks.

On the other hand, packet-level TEC would have the advantage of satisfying the need for application QoS requests and content-aware treatment. There are three main packet-level TEC methods: (1) *packet-level forward error correction (P-FEC)*; (2) *automatic repeat request (ARQ)*; and (3) *packet-level interleaving with packetization (P-interleaving)* [1].

There are trade-offs among these control techniques. FEC requires additional bits, but at the same time it can correct corrupted data without retransmission[2]-[5]. ARQ is inadequate for real-time video streaming because of delay-constraints, but it is more effective than FEC under relatively good channel conditions and loose delay requirements [4]-[7]. P-interleaving can spread burst errors and has no overhead, but causes additional delays associated with packet permutation. A low-delay interleaving method [7] was proposed using a video encoder buffer as part of interleaving memory, and Y. J. Liang et al.[8] determined the optimal interleaver minimizing the expected total distortion of the decoded video, subject to a delay constraint.

There have been few studies of interleaving which take video content into consideration, S. K. Chin et al.[9] proposed a content-aware interleaving method which considers the priority of the base layer and enhancement layer in the layered codec. They attempt to change packet-sending orders by randomizing and interleaving the base layer packets with several enhancement layer packets. However this technique provides only a quite coarse degree of content-aware interleaving, because it operates on only layer-level not packet-level.

In this paper, we propose a more general and finer content-aware P-interleaving method, which regulates burst error effects by spreading out the video packets in accordance with each packet's pre-calculated priority.

In section 2, we review general interleaving methods and describe the problem posed by the P-interleaver in video transmission. In section 3, we describe our own interleaving method, which is designed to solve the above mentioned problems. Lastly, we present the experimental results and further discussions in section 4.

2 General Packet-Level Interleaving

In the burst error-prone wireless network environment, consecutive packet losses happen more frequently, and this causes more serious degradation in the video quality than losses that are spread uniformly, for a similar average packet loss rate. P-interleaving is a method which is widely used to spread spatio-temporally contiguous packet losses. Y. J. Liang et al. [8] proposed an interleaver which minimizes the total distortion, given knowledge of the channel burst loss characteristics and the delay constraint. According to [8], given the channel loss characteristics and the delay constraint $\delta_{\max}$ determine the optimal interleaver $(n_{\mathrm{opt}}, d_{\mathrm{opt}})$, such that the total distortion of the decoded video sequence $D\left[I\left(n, d, K_{\mathrm{orig}}\right)\right]$ is minimized, i.e.,

$$(n_{\mathrm{opt}}, d_{\mathrm{opt}}) = \arg\ \min_{n,d:(n-1)\times(d-1)\leq\delta_{\max}} D\left[I\left(n, d, K_{\mathrm{orig}}\right)\right] \qquad (1)$$

where $I(\cdot)$ denotes the functional representation of the interleaver (n, d) indicating the interleaving data size, and K_{orig} denotes the indices of the lost packets before interleaving. Let us consider two methods of the (n, d) interleaver in video transmission, namely temporal interleaving and spatial interleaving.

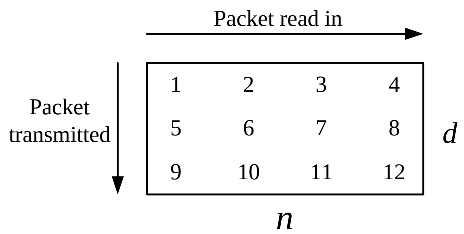


Fig. 1. An example of general (n, d) interleaver $(n = 4, d = 3)$

First, before going into depth, let us consider video sequence, $\boldsymbol{S}$, in Eq.(2)-(4).

$$\boldsymbol{S} = [\boldsymbol{s_0}; \boldsymbol{s_1}; \boldsymbol{s_2}; \ldots; \boldsymbol{s_m}; \boldsymbol{s_{m+1}}; \ldots] \tag{2}$$

where $\boldsymbol{S}$ is a matrix that represents the sequence of video packets, and is a vector that represents the sequence of video packets of the m-th picture, which is expressed as,

$$\boldsymbol{s_m} = [s_{m,0}, s_{m,1}, s_{m,2}, \ldots, s_{m,n}, s_{m,n+1}, \ldots] \tag{3}$$

where $s_{m,n}$ represents the n-th packet of the m-th picture of the video sequence.

Let us assume that burst errors impact on the video sequence $\boldsymbol{S}$ and the temporally interleaved video sequence, $\boldsymbol{S}_T$

$$\boldsymbol{S}_T = I_T\left(n_{\text{pic}}, d_{\text{pic}}, \boldsymbol{S}\right) \tag{4}$$

where $I_T\left(n_{\text{pic}}, d_{\text{pic}}, \boldsymbol{S}\right)$ is the functional representation of the $(n_{\text{pic}}, d_{\text{pic}})$ temporal interleaver on video sequence, $\boldsymbol{S}$. As shown in Fig. 2, the temporal inter-

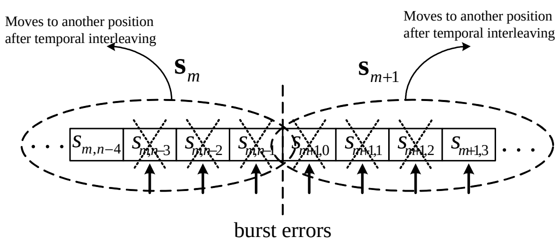


Fig. 2. Temporal interleaving

leaving process scatters the impact of burst errors, which would have affected two or more consecutive pictures, on non-consecutive pictures by changing the sequence of the pictures. Next, let us consider the corrupted video packets in the picture shown in Fig. 2. Although temporal interleaving scatters burst errors on non-consecutive pictures rather than consecutive pictures, some burst errors still remain in the pictures, as shown in Fig. 3, which result in serious quality degradation. For this reason, additional interleaving of the individual packets making up each picture is required to spread out the burst errors, and this is referred to as spatial interleaving. Although spatial interleaving spreads out burst errors, the number of packet errors within a picture, e.g., $\boldsymbol{s}_{m+1}$, remains the same.

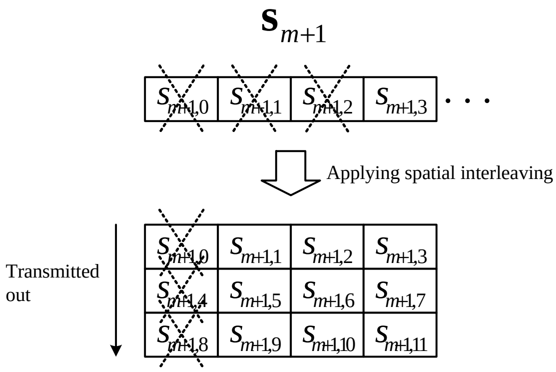


Fig. 3. Spatial interleaving within a picture after applying temporal interleaving

It is known that for the same amount of data lost from the video stream in a given communication channel, the effect on the end-to-end video quality varies considerably depending on the position of the data in the stream [1]. Thus, in the next section, we propose a content-aware packet-level interleaving technique that spreads the priority of the content, thereby resulting in improved protection against burst errors.

3 Proposed Content-Aware Packet-Level Interleaving

3.1 Quantitative Packet-Priority Metric with Video Content-Awareness

For a motion-compensated video coder such as ITU-T H.263 [10], the macroblock (MB)-based corruption can be modeled while taking into consideration the effects of error concealment, the temporal dependency controlled by coding modes and motion vectors, and prediction loop filtering. By assuming that the loss impact of each MB is independent, the impact of one MB loss can be expressed as the sum of the initial error and the propagation error. Also, assuming that the encoder is familiar with the error concealment method at the decoder, the initial error for the MB can be estimated by differentiating to a general error propagation behavior, where the initial error which propagates to subsequent frames is governed by the effects of the temporal dependency and prediction loop filtering. Under such a scenario and by making use of the results of our previous work [11], we can estimate the total impact of an MB loss in terms of its error energy by

$$\sigma_{MB}^2 = \sigma_u^2 + \sum_{m=1}^{M} \sum_{j=1}^{N} w_{n,j}^2 \left(m, j\right) \sigma_v^2 \left(m, j\right) \tag{5}$$

where σ_u^2 is the initial error and its value depends on the underlying error concealment scheme, $w_{n,j}\left(m, j\right)$ and $\sigma_v\left(m, j\right)$ stand for the temporal dependency weight and the propagating error impact on the j-th MB(among N MBs) of the m-th frame(among M subsequent frames), respectively. In addition,

$$\sigma_v^2 \left(m, j\right) = \frac{\sigma_u^2}{1 + \gamma_{m,j}} \tag{6}$$

where $\gamma_{m,j}$ is a parameter called the decaying factor that is governed by the strength of the prediction loop filtering and the frequency characteristic of the initial error. In order to transmit the video stream, it is efficient to packetize based on the synchronization code, where a start code can be inserted into the start of every GOB or slice data in H.263. Thus, we extend the above-mentioned MB-level corruption model, in order for it to be interleaved on GOB level.

Once σ_{MB}^2 is estimated, the GOB-level corruption effect, σ_{GOB}^2, as a packetization unit, can be estimated by averaging σ_{MB}^2 over the number of MBs in the GOB, N_{MB}. The estimated GOB-level relative priority index (RPI)-values (σ_{GOB}^2) are used as a parameter representing the effect on the end-to-end video quality. This RPI is adopted in our content-aware packet-level interleaver that is explained in Section 3.2.

3.2 The Algorithm of Content-Aware Packet-Level Interleaving

In this section, lets assume that only P-interleaving is available because of channel bandwidth constraints and that the maximum allowed transmission delay, $\delta_{\max}$, has been determined. In this case, temporal interleaving is necessary to limit the effect of burst errors on consecutive pictures. Besides temporal interleaving, we also need additional transmission procedures that minimize the degradation of the video quality. The spatial packet interleaving technique mentioned in Section 2, which interleaves the frames within each picture, can provide a solution for burst errors within a picture. However, content-blind interleaving can cause abrupt quality degradation when the most important portion of the packets contained in a burst pattern is lost. In order to solve this problem, we propose a content-aware interleaving technique designed to scatter burst errors intelligently with content-awareness.

Optimal Content-Aware Packet-Level Interleaving Basically, the packetized video sequence, $\boldsymbol{S}$, is transmitted in the order $\boldsymbol{s_0}, \boldsymbol{s_1}, \boldsymbol{s_2}, \ldots$. The proposed content-aware P-interleaving method consists of 3 steps, i.e., (1) *temporal interleaving*, (2) *spatial interleaving and packetization with RPI*, and (3) *packet transposition.*

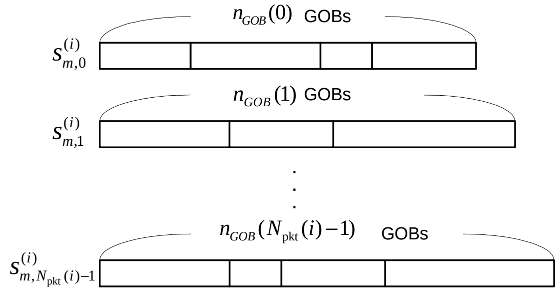


Fig. 4. Example of i-th selection set, $\boldsymbol{s_m^{(i)}}$, with $N_{\text{pkt}}^{(i)}$ packets of arbitrary m-th picture

First, we apply the temporal interleaving technique explained in Section 2, in order to spread the burst errors which would affect two or more consecutive pictures, so that a new picture sequence $\boldsymbol{S_T}$ is obtained. Next, all of the temporally interleaved pictures are spatially interleaved and packetized based on the RPI-value of the GOBs. Denoting N_S the number of possible packetization sets of $\boldsymbol{s_m} = \left\{\boldsymbol{s_m^{(0)}}, \boldsymbol{s_m^{(1)}}, \boldsymbol{s_m^{(2)}}, \cdots, \boldsymbol{s_m^{(N_S-1)}}\right\}$, the optimal content-aware interleaving and packetization is to find the selection set i_{opt}, so that $\boldsymbol{s_m^{(i_{\text{opt}})}}$ minimizes the quality-variation of packet sequence, $\left(\Delta_{\text{pkt}}^{(i)}\right)^2$, as described in Eqs.(7) and (8).

$$\boldsymbol{s_m^{(i_{\text{opt}})}} = \arg\min_{\boldsymbol{s_m^{(i)}}} \left(\Delta_{\text{pkt}}^{(i)}\right)^2 \tag{7}$$

$$\left(\Delta_{\text{pkt}}^{(i)}\right)^2 = \frac{1}{N_{\text{pkt}}^{(i)} - 1} \sum_{j=0}^{N_{\text{pkt}}^{(i)}-1} \sum_{k=0,k\neq j}^{N_{\text{pkt}}^{(i)}-1} \left[\sigma_{\text{pkt}}^2(j) - \sigma_{\text{pkt}}^2(k)\right]^2 \tag{8}$$

where $N_{\text{pkt}}^{(i)}$ is the number of packets in $\boldsymbol{s_m^{(i)}}$, $\sigma_{\text{pkt}}^2(j)$ is the averaged RPI-value of the j-th packet of $\boldsymbol{s_m^{(i)}}$. The values of $\left(\Delta_{\text{pkt}}^{(i)}\right)^2$ are evaluated for N_S number of $\boldsymbol{s_m^{(i)}}$'s in $\boldsymbol{s_m}$ and the optimal solution can be obtained from Eq.(7).

Lastly, once the spatio-temporally interleaved and packetized sequence of video packets, $\boldsymbol{S_{T,S}} = I_S(\boldsymbol{S_T})$, has been generated, where $I_S(\cdot)$ is the functional representation of the spatial interleaving, all of the packets in $\boldsymbol{S_{T,S}}$ should be transmitted in a manner that minimizes the expected degradation of the video quality. To accomlish this, we propose an additional transmission process, namely, packet transposition. To explain packet transposition, we need to go into detail about the packet transmission procedure. In the general packet transmission procedure, the spatio-temporally interleaved and packetized sequence, $\boldsymbol{S}_{T,S}$, is transmitted in the order $\boldsymbol{s'_0}, \boldsymbol{s'_1}, \boldsymbol{s'_2}, \ldots, \boldsymbol{s'_m}, \ldots$, where $\boldsymbol{s'_m}$ denotes the m-th picture after temporal interleaving in step 1. When transmitting in this order, the packets in $\boldsymbol{s'_{m+1}}$ should be transmitted only after $\boldsymbol{s'_m}$ has been completely transmitted, in order to limit the impact of burst errors on the picture sequence. To further reduce the impact of the error burst, we reschedule packet sequence $\boldsymbol{S}_{T,S}$ by applying a transpose operation to $\boldsymbol{S^{(i)}_{T,S}}$ in order to find the optimal solution, i.e., the selection set i, $\left(\boldsymbol{S^{(i)}_{T,S}}\right)^T$, which minimizes the effect of the burst error on the end-to-end video quality, by applying the rule contained in Eq.(8) to $\boldsymbol{S}_{T,S}$, thereby producing the packet sequence described in Eq.(9),

$$\left(\boldsymbol{S^{(i)}_{T,S}}\right)^T = \left[s^{(i)}_{0,0}{}', s^{(i)}_{1,0}{}', s^{(i)}_{2,0}{}', \ldots, s^{(i)}_{N_P,0}{}', s^{(i)}_{0,1}{}', s^{(i)}_{1,1}{}', \ldots, s^{(i)}_{N_P,1}{}', s^{(i)}_{2,0}{}', \ldots\right] \quad (9)$$

where $(\cdot)^T$ represents the transpose operation on the matrix, N_P is the number of picture in the interleaving interval and $s'_{m,n}$ is the n-th packet of the m-th picture. Since the above technique requires full scanning of the whole sequence of packets in an interleaving interval to obtain the optimal solution, which is too computationally expensive, we propose a more practical solution in the next subsection.

Practical Content-Aware Packet-Level Interleaving In this section, we propose a practical content-aware packet-level interleaving algorithm, designed to reduce the computational complexity of the optimal content-aware packet-level interleaving algorithm described in the previous subsection.

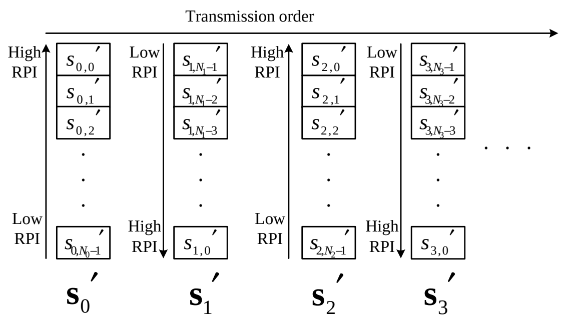


Fig. 5. *Packet transposition process for the spatio-temporally interleaved packet sequence.* ($\boldsymbol{S}_{T,S}$: N_0, N_1, and N_2 are the number of packets for the 0-th, 1st and 2nd picture, respectively.)

In order to simplify step 2 of the 3 steps described in the previous subsection, i.e., the spatial interleaving and packetization step, we consider the packetization unit as a GOB per packet. Of course, this is not the best method of minimizing the variation of the averaged RPI in a packet, however we compensate for this shortcoming by means of the packet transposition step. The next step is packet transposition. To implement this step, those packets in a picture with an even-numbered picture number are sorted in descending order of their RPI-value, σ^2_{pkt}, and those packets with an odd-numbered pictures are sorted in ascending order of their RPI-value, and this process is continued for all of the pictures in the interleaving interval. Then the packet sequence of $\boldsymbol{S}_{T,S}$ is transposed to $(\boldsymbol{S}_{T,S})^T$, which is the practical solution of the content-aware packet-level interleaving method.

4 Experimental Results and Discussion

In this section, we demonstrate the performance of the proposed content-aware interleaving scheme, by performing the simulations under the burst error environments (with 4, 6, 8 and 10 consecutive packet errors). A "Foreman" sequence (CIF, 352×288) was encoded using H.263 with the encoding parameters, shown in Table 1, and then transmitted through the burst error-prone channels using four different methods, i.e., (1) *no interleaving (neither temporal nor spatial)*, (2) *temporal interleaving only*, (3) *content-blind spatio-temporal interleaving*, and (4) *content-aware spatio-temporal interleaving*.

Table 1. Video encoder parameters used in the experiment

Parameter settings	
Video encoder	H.263
Sequence name	Foreman
Image format	CIF (352×288)
Number of encoded pictures	300
Encoding method	1 I-picture followed by 299 P-pictures with VBR
Channel setting	
Channel errors	4, 6, 8 and 10 burst packet errors every 15 pictures.

Figs. 6 and 7 show the experimental results for each of these four methods. The experimental result in Fig. 6 shows that the content-blind methods, i.e., (1), (2) and (3), are influenced by the burst errors, so that the PSNR distribution shows abrupt degradations. In contrast, the proposed content-aware method is less susceptible to the burst errors, with the result that the PSNR curve descends slowly and maintains a certain degree of video quality in spite of the severe burst errors.

The three content-blind methods do not consider the variable effect of the burst errors on the quality degradation, by making use of an index of content-awareness such as RPI. Therefore, as shown in Figs. 7(a) and 7(b), the variation

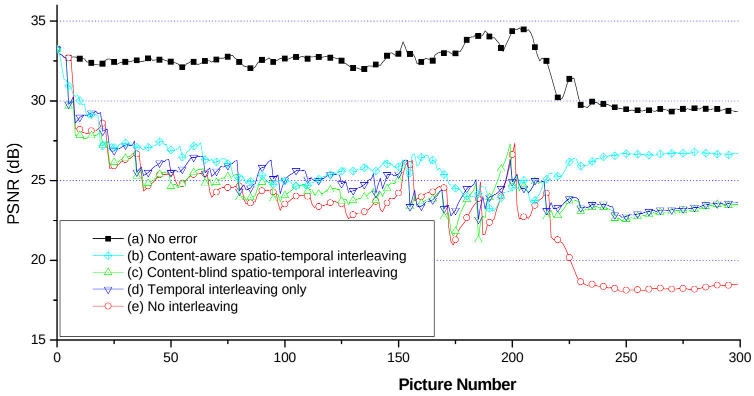
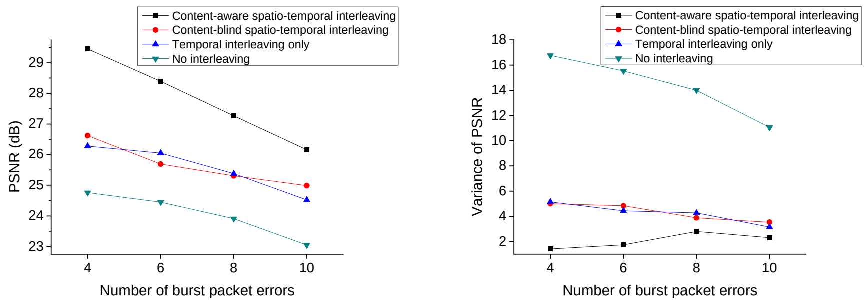


Fig. 6. *PSNR distribution of Y-component of transmitted sequence ((b)-(e): applied 10 burst packet errors with period of 15 pictures)*: (a) No error, (b) Content-aware spatio-temporal interleaving, (c) Content-blind spatio-temporal interleaving, (d) Temporal interleaving only, (e) No interleaving



(a) Average PSNR

(b) Variance of PSNR

Fig. 7. The perfermance comparison of the different interleaving methods

and average PSNR performances of these schemes are worse than those of the proposed one. In contrast, the content-aware scheme spreads the packet error bursts by re-ordering the packets spatially, according to the importance of the content of each packet with respect to the picture, and keeps the resultant quality degradation resulting from the burst errors as uniform as possible, through packet transposition. The experimental results show that the proposed scheme provides robust protection against burst errors.

From the experimental results, it can be inferred that the proposed content-aware P-interleaving method shows good performance from the viewpoint of the objective and subjective video quality by reducing the variance of the PSNR. These results provide some insight into why the simple interleaving method is insufficient to improve the overall video quality, in that although it spreads the quality degradation, the average PSNR remains at a similar level.

References

1. B. Girod and N. Färber, “Wireless Video,” in A. Reibman, M.-T. Sun (eds.), Compressed Video over Networks, Marcel Dekker, 2000
2. T. Nguyen, and A. Zakhor, “Distributed video streaming with forward error correction,” 12th International Packet Video Workshop (PV 2002), Apr. 2002
3. W. Tan and A. Zakhor, “Video multicast using layered FEC and scalable compression,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, Mar. 2001, pp. 373-386
4. D. Wu, Y. T. Hou, and Y. Q. Zhang, “Transporting real-time video over the Internet : Challenges and approaches,” Proceedings of the IEEE, Vol. 88, No.12, Dec. 2000, pp. 1855-1975
5. D. Wu, Y. T. Hou, W. Zhu, Y. Q. Zhang and J. M. Peha, “Streaming video over the internet: Approaches and directions,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No.3, Mar. 2001, pp. 282-300
6. N. Guo, and S. D. Morgera, “Frequency-Hopped ARQ for wireless network data services,” IEEE Journal on Selected Areas in Communications, Vol. 12, No. 8, Oct. 1994, pp. 1324-1337
7. S. Aramvith, C. W. Lin, S. Roy, and M. T. Sun, “Wireless video transport using conditional retransmission and low-delay interleaving,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No. 6, Jun. 2002, pp. 558-565
8. Y. J. Liang, J. G. Apostolopoulos and B. Girod, “Model-based delay-distortion optimization for video streaming using packet interleaving,” Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, Nov. 2002. Invited Paper
9. Suk Kim Chin and Braun, R, “Improving video quality using packet interleaving, randomisation and redundancy,” Local Computer Networks, 2001. Proceedings. LCN 2001. 26th Annual IEEE Conference on , 14-16 Nov. 2001, pp. 405-413
10. ITU-T, “Recommendation H.263, video coding for low bit rate communication,” Feb. 1998
11. J. Shin, J. G. Kim, J. W. Kim and C.-C.J. Kuo, “Dynamic QoS mapping control for streaming video in relative service differentiation networks,” European Transactions on Telecommunications, Vol. 12, No. 3, May-June 2001, pp. 217-230