Is Burst Loss Worse Than Random Loss For Video Transmission?

Jimin Xiao*†, Tammam Tillo†, Chunyu lin‡
* Dept. of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK
Email: jimin.xiao@liverpool.ac.uk
† Dept. of Electrical & Electronic Engineering Xi'an Jiaotong-Liverpool University, Suzhou, China
Email: tammam.tillo@xjtlu.edu.cn
‡ Institute of Information Science, Beijing Jiaotong University, Beijing, China
Email: yuailian@gmail.com

***Abstract*—Many papers have shown that burst loss causes more severe damage to the reconstructed video quality than random loss for the same packet loss rate. With traditional interleaving methods, the consecutive lost packet positions will be interleaved to random positions, and after interleaving, the PSNR of burst loss can be close to the PSNR of random loss for the same packet loss rate. In this paper, a novel slice reorder and packetization (SRP) approach is proposed which fully utilizes the temporal correlation between slices in successive frames. Theoretical analysis is presented that explains why with the proposed approach, burst loss can be better than random loss for the same packet loss rate. The effect of the proposed approach is validated with H.264/AVC coded video. The simulation results show that with the proposed approach, the PSNR of the burst loss can be up to more than 2dB higher than random loss without adding any redundant information.**

***Index Terms*—Rate Distortion, H.264/AVC, error resilience, slice, burst loss, random loss.**

I. INTRODUCTION

The latest video coding standard, H.264/AVC [1] provides higher coding efficiency and stronger network adaptation capability in comparison to all the previously developed video coding standards. Most video coding standards are based on a hybrid coding method, which use transform coding with Motion-Compensated Prediction (MCP). As a result, when transmitting the hybrid-coded video in packet-loss environment, it suffers from error propagations and this leads to the well-known drifting phenomenon [2][3].

Many papers have addressed the actual network loss behavior, and most of them agree that Internet packet loss often exhibits finite temporal dependency, which means if current packet is lost, then the next packet is also likely to be lost. This leads to burst packets loss [7]. The distortion caused by burst loss is modeled in [4][5]. However, in both [4] and [5], only P-frame is used when modeling the effect of burst loss, this is not suitable for the case when I-frame is periodically inserted, like video broadcasting. Another concern in both [4] and [5] is that, it is assumed that one frame is encoded into one slice, and therefore, each frame is carried in one packet. However, this packetization scheme is not practical, especially for high resolution and/or high quality compressed video communications, as the bits of each frame is much higher than the Maximum Transmission Unit (MTU) of the underlying network. One practical method of packetization is to use fixed slice length to meet the MTU limitation, this approach has been used to create slices in the proposed approach. In the case of transmitting one frame with more than one slice (packet), burst error generally leads to lose a large area in the current frame, then the decoder can not effectively use the neighboring correlation information to recover the lost pixels, eventually, burst loss leads to more severe damage to the reconstructed video quality than random loss for the same Packet Loss Rate (PLR). In this paper, a novel Slice Reorder and Packetization (SRP) scheme is proposed, which fully utilizes the temporal correlation between slices of successive frames to tackle the burst cases. Unlike traditional error-resilience method as MDC and ULP, the advantage of the proposed approach is that the encoder does not need to insert any redundant information to the bitstream, which can save the precious bandwidth resource.

The rest of the paper is organized as follows. In Section II the basic principles of the proposed approach are presented, and then the implementation algorithm is described in detail. Section. III presents the theoretical analysis of the proposed approach. In Section. IV simulation results are presented. Finally, some conclusions are drawn in Section V.

II. SLICE REORDER AND PACKETIZATION

In burst loss scenario, if the current packet is lost, the next packet is also likely to be lost. Based on this, the basic idea of the proposed approach is to arrange the I-slice and its corespondent P-slices into adjacent positions for transmission. Then, in burst loss scenario, if an I-slice or P-slice is lost, the following P-slices will also be lost with high probability. The proposed SRP approach can be divided into three steps. First step is slicing, which assures the maximum length of each slice is not larger than the MTU of the network, and meanwhile, the I-slice and its corespondent P-slices cover the same spatial region. The second step is slice reorder, in which I-slice and its corespondent P-slices are rearranged to adjunct transmission positions. The final step is packetization stage, in this stage slices will be incapsulated into one packet until all the MTU bytes are used.

Before describing the algorithm, some preliminaries are presented. The MTU of the underlying network is M in byte, the GOP length of the video sequence is L. Only the first frame in one GOP is I-frame, while the following frames are P-frames. Both I-frame and P-frame contain N slices. The i-th frame of the video sequence is denoted by f_i, and the j-th slice in frame f_i is denoted by $S_{i,j}$. The length of slice $S_{i,j}$ in byte is denoted by $b_{i,j}$, while the number of macroblocks included in slice $S_{i,j}$ is denoted by $m_{i,j}$.

A. Slicing step

The proposed slicing step has two requirements. First, for any j, slices $S_{1,j}$, $S_{2,j}$, ..., $S_{L,j}$ should cover the same area in the frame they belong to, that is to say, all the slices in the same spatial position should include the same number of macroblocks, which is denoted as m_j.

$$m_j = m_{1,j} = m_{2,j} = ... = m_{L,j}, \forall j \tag{1}$$

In general, this condition leads to slices at the same spatial position, however in different frames, having different length. Meanwhile, because we need to meet the MTU limitation, we put as many macroblocks into one slice as possible, under the condition that the slice length fits into one MTU. In order to ensure the previous requirements, let us assume $b_j = max\{b_{1,j}, b_{2,j}, ..., b_{L,j}\}$, in this case, the value of m_j should meet following requirement.

$$\begin{cases} b_j \le M & \text{if } \forall i,\ S_{i,j} \text{ contains } m_j \text{ macroblocks} \\ b_j > M & \text{if } \forall i,\ S_{i,j} \text{ contains } m_j + 1 \text{ macroblocks} \end{cases} \tag{2}$$

For the majority of the video sequences, the I-frame generates more bits than following P-frames. Based on this, the requirement (2) can be simplified as (3).

$$\begin{cases} b_{1,j} \le M & \text{if } S_{1,j} \text{ contains } m_j \text{ macroblocks} \\ b_{1,j} > M & \text{if } S_{1,j} \text{ contains } m_j + 1 \text{ macroblocks} \end{cases} \tag{3}$$

where $b_{1,j}$ is the byte length of the I-slice $S_{1,j}$.

B. Reorder step

In this step, I-slice and its corespondent P-slices would be rearranged, so as to be transmitted in sequence of adjacent positions. Without reorder, the slice transmission order is sending all slices in the first frame, and all slices in the second frame, and so on. For example in this case, the slice sending order is $S_{1,1}$, $S_{1,2}$, $S_{1,3}$, ..., $S_{1,N}$, $S_{2,1}$, $S_{2,2}$, $S_{2,3}$, ..., $S_{2,N}$, ..., $S_{L,1}$, $S_{L,2}$, $S_{L,3}$, ..., $S_{L,N}$.

After slice reorder, slices with the same spatial position in successive frames are sent together. For example, with slicing scheme in previous section, the sending order is $S_{1,1}$, $S_{2,1}$, $S_{3,1}$, ..., $S_{L,1}$, $S_{1,2}$, $S_{2,2}$, $S_{3,2}$, ..., $S_{L,2}$, ..., $S_{1,N}$, $S_{2,N}$, $S_{3,N}$, ..., $S_{L,N}$.

C. Packetization Step

It is worth noticing that b_j is the maximum length in byte of slices $S_{1,j}$, $S_{2,j}$, $S_{3,j}$, ..., $S_{L,j}$, and slice $S_{1,j}$ is encoded with Intra mode, while other slices are encoded with Inter mode. In most cases, Intra coding is not as efficient as Inter coding, which means that $S_{1,j}$ is much larger than other slices. Consequently, it is expected that most of the slices $S_{1,j}$, $S_{2,j}$, $S_{3,j}$, ..., $S_{L,j}$ are smaller than the MTU size M, so usually one slice per packet would not be efficient, because of the IP/UDP/RTP [6] header cost. In the proposed SRP approach, more than one slice are transmitted in one packet. The pseudo-code of the Packetization is presented in Algorithm 1, and the aim of it is to incapsulate as many slices into one packet as possible.

Algorithm 1 Packetization algorithm

```
packet_byte ⇐ 0
for j = 1 to N do
  for i = 1 to L do
    if packet_byte + b_{i,j} ≤ M then
      Append S_{i,j} to current packet
      packet_byte = packet_byte + b_{i,j}
    else
      Send current packet
      encapsulate S_{i,j} into new packet
      packet_byte = b_{i,j}
    end if
  end for
end for
```

III. Why burst loss can be better?

When there are slices loss during transmission, the lost slices would be concealed with concealment algorithm. For the P-slice loss, concealment algorithm estimates the lost motion vector, then use temporal replacement (TR) approach to recover the lost slice. However, the estimated motion vector is not accurate, especially when the movement is fast and not translational or the lost slice covers a large area. As a result, the inaccuracy of the estimated motion vector is a major reason that leads to concealment distortion. In order to demonstrate the effect of inaccurate estimated motion vectors, we compare the video quality in three cases: I) the compared video with, no slice is lost. II) 10% P-slices are lost, but the lost motion vectors can be recovered correctly, which means the estimated motion vectors of the lost P-slices are identical to the original motion vectors. III) 10% P-slices are lost, and concealed using the concealment algorithm in H.264 JM14.0, and consequently in this case the estimated motion vectors are inaccurate. In all the three cases, the quantization parameter is set to 28 for intermediate compression video quality, the video sequence GOP is set to 30. The PSNR for case I, II and III for some standard CIF video sequences is shown in Table.I. Clearly, the distortion caused by estimated motion vector error is larger than that caused by residual coefficient loss. Therefore, in the following analysis, we take the error of estimated motion vectors as the key factor that leads to concealment distortion.

We compare two different scenarios: I) the random loss case, this case is depicted in Fig.1.a, where macroblocks A and G are lost; II) the burst loss case where macroblocks A and B are lost together as shown in Fig.1.b. Let us suppose that

TABLE I
VIDEO QUALITY FOR CASE I, II AND III

Video Sequence	I	$II\ (I-II)$	$III\ (I-III)$
Bridge Close	35.65	34.32(1.33)	33.50(2.15)
Bridge Far	37.35	37.30(0.05)	37.20(0.15)
Bus	35.81	30.32(5.49)	22.95(12.86)
Coastguard	35.39	32.74(2.65)	28.73(6.66)
Container	36.34	36.03(0.31)	34.04(2.30)
Foreman	37.05	36.13(0.92)	29.29(7.76)
Flower	36.35	32.18(4.17)	25.12(11.23)
Highway	38.85	38.29(0.56)	31.62(7.23)
Mobile	35.29	31.08(4.21)	24.07(11.22)
Mother Daughter	39.36	39.29(0.07)	37.95(1.41)
News	38.45	37.80(0.65)	32.81(5.64)
Silent	36.16	35.85(0.31)	31.86(4.30)
Stefan	36.57	33.38(3.19)	23.69(12.88)
Tempete	35.43	32.63(2.80)	27.71(7.72)
Waterfall	35.37	35.11(0.26)	31.56(3.81)
Average	36.63	34.83(1.80)	30.14(6.49)

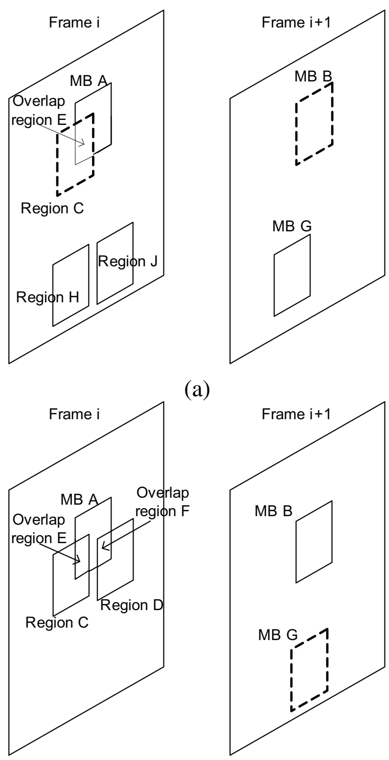


Fig. 1. Model for random loss (a) and burst loss (b). For random loss, macroblocks A and G are lost together; for burst loss, macroblocks A and B are lost together.

macroblocks A and B are in the same spatial position in Frame i and Frame $i+1$, and the positions of macroblocks A and G are far away from each other, so the distortion caused by losing macroblock A will not affect macroblock G. Let us suppose that Macroblock B is originally predicted from region C in Frame i, if this macroblock is lost, the estimated motion vector would point to region D. Similarly, for Macroblock G, it is originally predicted from region H in Frame i, however, if this macroblock is lost, the estimated motion vector would point to region J. We compare the distortions caused by losing macroblocks A and G with that of losing macroblocks A and B. To do so, we can start from the observation that in the two cases, the distorion in Frame i is exact same, and the distortion in the regions outside macroblocks B and G in Frame $i+1$ is almost similar, in fact, for simplicity we ignore the effect of INTRA prediction. Therefore, we just need to compare the total distortion in the macroblocks B and G for the two cases.

For the random case, macroblocks A and G are lost, the distortion in macroblock B is denoted by $D_r(B)$, which is caused by distortion propagation from region C. $E[D_r(B)] = E[n_1{\delta_1}^2]$, where n_1 is the overlap pixel number in the region E, ${\delta_1}^2$ is the average distortion per pixel, which is caused by the inaccuracy of the estimated motion vector in the concealment process. The distortion in macroblock G is denoted by $D_r(G)$, $E[D_r(G)] = E[n{\delta_1}^2]$, here n is the pixel number in one macroblock. Adding the distortion from macroblocks B and G together, we can get

$$E[D_r(B)] + E[D_r(G)] = E[(n_1 + n)]{\delta_1}^2 \quad (4)$$

For the burst case, macroblocks A and B are lost, the distortion in macroblock B is denoted by $D_b(B)$, while the distortion in macroblock G is 0 in this case. Here for macroblock B, the distortion is caused by two parts. One part is because of the inaccuracy of estimated motion vector, as macroblock B is originally predicted from region C, but now the estimated motion vector points to region D. Another part is error propagation from macroblock A, as macroblock A and region D have an overlap region F. We assume there are n_2 pixels in the region F, $E[D_b(B)] = E[(n-n_2){\delta_1}^2 + n_2{\delta_2}^2]$, here ${\delta_2}^2$ is the average per pixel distortion caused by error prediction to an already error region. As the average value of n_1 and n_2 are same, $E[n_1] = E[n_2]$, we get

$$E[D_b(B)] + E[D_b(G)] = E[(n-n_1)]{\delta_1}^2 + E[n_1]{\delta_2}^2 \quad (5)$$

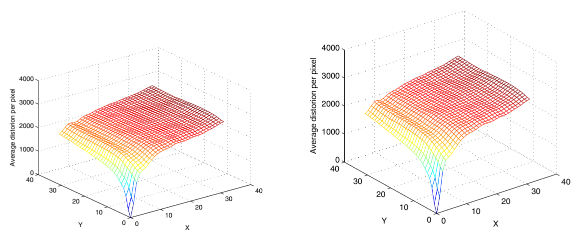
From equations (4) and (5), we can see if ${\delta_2}^2 = 2{\delta_1}^2$, burst loss and random loss cause same amount of distortion, if ${\delta_2}^2 < 2{\delta_1}^2$, burst loss leads to less distortion than random loss. In general, ${\delta_1}^2$ is caused by predicting from wrong region, and the distortion is enormous, then wrong prediction from an already error region would not double the distortion. Therefore, it is reasonable to assume ${\delta_2}^2 < 2{\delta_1}^2$. Fig.2 shows the average pixel distortion versus the motion vector error for some video sequences. Obviously, the distortion will not increase linear with the motion vector error, when the distortion reaches one certain amount, the distortion increases very slowly. The relation between average pixel distortion and the motion vector error also demonstrates our assumption of ${\delta_2}^2 < 2{\delta_1}^2$. Therefore, losing two co-located P-slices from two consecutive frames causes less total distortion than losing two far away uncorrelated P-slices.

IV. Experimental results

In the simulation, we compare the proposed SRP approach with standard JM software. The proposed SRP approach is implemented based on H.264 JM14.0 software. To get a better

TABLE II
VIDEO QUALITY FOR RANDOM LOSS AND SRP BURST LOSS ENVIRONMENT, PLR IS SET TO 10% AND 5%, QP IS SET TO 22 AND 28

Video Sequence	PLR = 10%				PLR = 5%			
	QP = 22		QP = 28		QP = 22		QP = 28	
	Random	SRP	Random	SRP	Random	SRP	Random	SRP
Bridge Close	29.20	29.95	28.77	29.34	32.90	33.20	31.68	31.97
Bridge Far	33.77	34.36	31.68	31.81	36.96	37.19	34.79	35.13
Bus	21.74	23.31	21.61	23.39	25.01	26.88	24.68	27.15
Coastguard	27.01	28.25	26.86	27.75	29.95	31.23	29.42	30.46
Container	28.09	28.33	27.13	27.47	31.78	32.73	31.14	31.46
Foreman	27.03	27.63	26.59	28.14	30.27	31.48	29.52	31.22
Flower	23.09	24.37	22.96	24.29	26.15	27.66	25.70	27.36
Highway	29.93	31.08	29.28	31.56	32.74	34.19	31.82	34.73
Mobile	21.53	22.53	21.35	22.91	24.60	25.97	24.24	25.76
Mother Daughter	35.41	35.61	33.15	33.62	37.93	38.97	35.52	36.52
News	28.01	29.14	27.90	28.92	31.71	33.07	31.10	32.31
Silent	28.33	29.46	27.73	29.02	32.31	33.30	31.04	32.31
Stefan	22.10	23.72	21.80	23.69	25.01	27.08	24.72	27.12
Tempete	24.83	25.69	24.33	25.55	27.92	29.02	27.31	28.80
Waterfall	29.86	30.73	29.17	30.11	32.75	33.76	31.41	32.14
Average	27.33	28.28	26.69	27.84	30.53	31.72	29.61	30.96



(a) (b)

Fig. 2. Average pixel distortion versus motion vector error (a) Flower sequence (b) Stefan sequence

understand of the proposed SRP approach, we compare the results of SRP with JM random loss. In all the simulations, GOP structure is IPPP...; GOP length L is 30 frames, among which only one previous frame is used as reference frame; the MTU size of network M is set to 500 bytes; burst length is 5; and PLR is set to 5% and 10%. In Table.II, the average PSNR is shown for SRP and random scheme for 15 standard video sequences. With the SRP scheme, burst loss is always better than random loss. It is also interesting to notice that the faster the movement in the video sequence, the bigger the PSNR gap between SRP and SRP random scheme would be. This is because for the fast movement sequence, the distortion caused by inaccuracy of estimated motion vector is huge. For some fast movement video sequence, like sequence bus and stefan, SRP can be more than 2dB better than random case.

V. CONCLUSIONS

In this paper we have proposed a novel video slice re-order and packetization scheme to combat network burst loss. From simulation results, it is shown that, with the proposed approach, the PSNR of burst loss can be up to more than 2dB higher than random loss for the same packet loss rate and source bit rate. Unlike other error resilience technology as multiple description coding or unequal loss protection, the proposed approach does not need to insert any redundant information to the video stream. The only cost of the proposed approach is one GOP of delay, but for non-conversation application like video broadcasting, one GOP of delay is usually acceptable. To the best of our knowledge, this is the first approach that shows burst loss can lead to better performance than random loss for the same packet loss rate.

ACKNOWLEDGMENT

This work was supported in part by National Natural Science Foundation of China (No.60972085, No.60903066), and Sino-Singapore JRP (No.2010DFA11010).

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[2] S. Wenger, "H.264/AVC over IP," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 645 - 656, July 2003.

[3] T. Stockhammer, M. M. Hannuksela, and T. Wiegand "H.264/AVC in Wireless Environments," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 657 - 673, July 2003.

[4] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, " Analysis of Packet Loss for Compressed Video: Effect of Burst Losses and Correlation Between Error Frames," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 861-874, July 2008.

[5] Z. C. Li, J. Chakareski, X. D. Niu, Y. J. Zhang, and W. Y. Gu " Modeling and Analysis of Distortion Caused by Markov-Model Burst Packet Losses in Video Transmission," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 917-931, July 2009.

[6] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," *Internet Engineering Task Force - RFC 1889*, Jan. 1996.

[7] D. Loguinov and H. Radha, "End-to-end internet video traffic dynamics: Statistical study and analysis," in *Proc. of Proc. IEEE INFOCOM '02*, Jun. 2002, pp. 723 - 732.