Retransmission Scheme for MPEG Streams in Mission Critical Multimedia Applications

Sugh-Hoon Lee, Sungyoung Lee
Department of Computer Engineering,
Kyung-Hee University, Korea
{shlee, sylee@oslab.kyunghee.ac.kr}

Abstract

Since multimedia data are characterized by continuity and massive volume, compression techniques such as MPEG are used for efficient transmission. However, a transmission error occurring at a frame tends to be propagated to other frames because MPEG frames cross-reference one another upon presentation. Due to time constraints, it is usual to discard without retransmission the data which have transmission errors or which cause serious delay. But in cases of multimedia applications where retransmission is required (for example, mission critical applications), the QoS may be severely deteriorated by the continuous loss caused by propagation if damaged frames are simply discarded. In this paper, we propose a selective retransmission method where the number of retransmissions can be reduced without losing QoS and data continuity. Since the proposed method keeps track of error propagation ranges using virtual buffers which have information about the sizes and types of frames before transmission, the number of retransmissions from the transmitter to the receiver can be minimized by selecting and retransmitting the recoverable data only. Simulation results show the proposed method reduces retransmission cost considerably while maintaining QoS similar to the case of unconditionally retransmitting all the damaged data.

1. Introduction

Under high-speed network environments, multimedia data with transmission errors or serious delay are usually discarded without retransmission due to time constraints. However, in the cases of mission-critical multimedia applications where the QoS (Quality of Service) is very important, it is necessary to retransmit such damaged data to provide high quality results. For example, when a doctor remote-examines the position or transition of cancerous cells through the network, it is possible for him/her to make a wrong diagnosis due to the data loss occurring during transmission. Even in normal applications, it is desirable to retransmit if the data is recoverable without violating the time constraint [2].

MPEG is widely used as a standard compression technique for moving pictures because it gives a higher data compression rate by removing not only spatial but also temporal redundancy [1]. It uses three types of frames: independently decodable I (Intra-mode) frames, P (Predicted-mode) frames which reference leading I or P frames, and B (Bidirectional) frames which bidirectionally reference leading and trailing I or P frames. These frames are arranged in a uniform sequence. Once an error occurs at an I or P frame, it is propagated to the other frames and causes consecutive frame losses [3]. Generally, the Cell Loss Concealment technique is used to solve the frame loss problem. The technique however, is not an adequate solution because it simply enhances the picture quality by repairing the damaged portions in order for users not to recognize them.

Therefore, in this paper, we propose a selective retransmission method, in retransmission-required multimedia applications, where the retransmission cost can be reduced and the QoS and data continuity can be guaranteed. To do so, we introduce the concept of a virtual buffer which grants predictability to the data by maintaining information about the data such as sizes and types of frames. By using virtual buffer, it is possible to selectively retransmit the damaged data only when they are recoverable, since we can tell at the decoder which frame the error occurred at and know whether it would affect other frames or not at the decoder. The frame information for the virtual buffer is maintained in a file in advance when the transmitter encodes or pre-processes the MPEG data. Retransmission requests can be made through the control channel generally used to send control commands such as play, stop or rewind, while an independent retransmission narrowband channel is established to send a series of frame information for the virtual buffer and the data to be retransmitted.

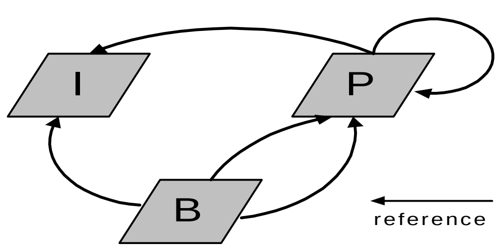
The remainder of this paper is organized as follows: Section 2 highlights the features of the MPEG. Section 3 describes the proposed model and its assumptions. Section 4 depicts the error detection and selective retransmission method using the virtual buffer. Section 5 shows simulation results. Section 6 summarizes our work.

2. Characteristics of MPEG

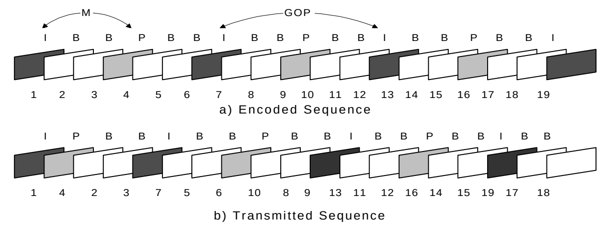
MPEG-2 (Motion Picture Expert Group-2) uses the DCT (Discrete Cosine Transform) to remove space-axis redundancy and the motion compensation to remove time-axis redundancy. It defines three types of frames: I, P and B frames. I frames are intra-coded using the DCT and are independently decodable without referencing other frames. P frames are decoded by referencing leading I or P frames located before them using motion compensation. B frames are decoded by bidirectionally referencing I or P frames located before and after them using motion compensation (Figure 1).

Figure 2-a shows an encoded sequence of MPEG frames. MPEG defines a group of frames from an I frame to the B frame right before the next I frame as GOP (Group of Pictures), and the period between an I frame and a P frame (or a P frame and another P frame) as M. Actual MPEG data are composed of a sequence of GOPs. Since P or B frames decoded through motion compensation must have the frames to reference in advance before they are decoded, the actual transmitted sequence of frames must be rearranged as shown in Figure 2-b. In Figure 2-a, in order for the number 2 B frame to be decoded, the P frame following the B frame as well as the leading I frame must arrive in advance even if the P frame is late in the decoding sequence.

Each frame can be prioritized according to its importance. Figure 3 shows an example of MPEG data where the GOP is 12 and M is 3. If frame I1 is damaged, the error is propagated to 14 other frames, B3 through B12. If frame P2 is damaged, B5 through B12 frames referencing it are affected. According to the importance of each frame within the GOP when an error occurs, the frame removal priority can be determined as in Figure 3 (The lower priority frame should be removed first). B frames must be removed first because they are affected by other frames and do not affect others. Although the loss of one frame is negligible in video, consecutive losses are not desirable. That is why the priorities of the adjacent B frames differ. Since P frames reference leading I or P



Ficture 1. Reference Relations among I,P,B Frame



Ficture 2. Encoded and Transmitted Frame Sequences (M=3, GOP=6)

Frame Type	P_1	B_1	B_2	I_1	B_3	B_4	P_2	B_5	B_6	P_3	B_7	B_8	P_4	B_9	B_{10}	I_2	B_{11}	B_{12}	P_5
Frame Priority		6	5	1	6	5	2	6	5	3	6	5	4	6	5	1	6	5	2

Ficture 3. Frame Priority with GOP

frames when decoded, the latter P frame (P4) has lower priority than the former (P2) within a GOP.

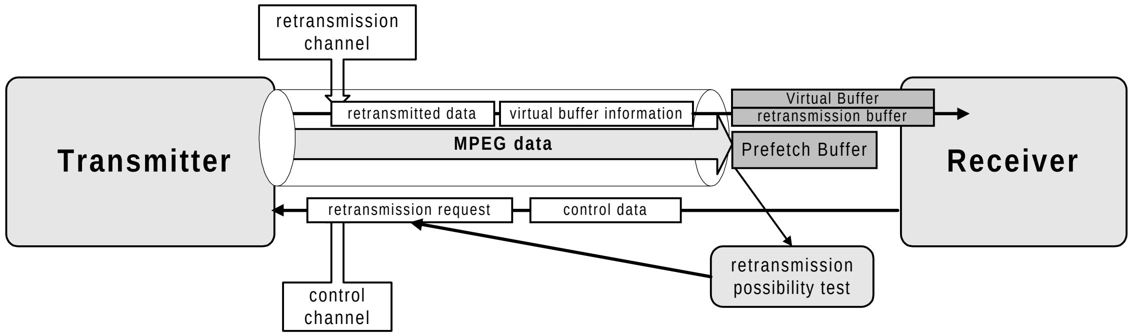
The sizes of I, B and P frames in MPEG differ according to the compression method. That is, the size of I frames is relatively larger than that of P or B frames because I frames are encoded using the DCT without motion compensation, while P or B frames use motion compensation to store only the changed parts. The sizes of the same types of frames may differ due to different compression rates according to the kinds of pictures to be compressed. Generally, the size of a P frame is one-third of an I frame, and the ratio between P and B frames ranges from 2:1 to 5:1.

Meanwhile, features of MPEG are very dependent on the contents of actual pictures. The time interval from one scene change to another is not regular, while scenes between one change and another are very similar. Furthermore, an intra-coded frame may appear even when a P or B frame must be present, because the cross-referencing is not meaningful any more once a scene change occurs. These characteristics of MPEG make it difficult to predict the sizes of frames. For instance, TV commercials or kungfu movies with many scene changes must have larger frame sizes than soap operas or artistic movies.

According to Krunz's experiments [5] in which he tried to model MPEG traffic mathematically based on real MPEG data, the sizes of I, B, P frames are analogous to lognormal distribution and the time interval from one scene change to another is analogous to a geometric distribution.

3. The Model and Assumptions

The proposed system model appears in Figure 4. As shown in Figure 4, the transmitter sends MPEG-encoded data to the receiver. It is assumed that the data stay in a buffer at the receiver and transmitter until decoded for, at most, 500 ms, which is the approximate time used to



Ficture 4 . System Architecture

Virtual Buffer

			28	27	26	25	24	23	22	21	20	19	18	17	16	15	← frame sequence
	....		B	I	B	B	P	B	B	P	B	B	P	B	B	I	← frame type
			22	131	31	44	92	13	12	73	32	14	60	25	20	120	← frame size

Ficture 5. Structure of Virtual Buffer

decode a 15-frame GOP. In the system, the virtual buffer and the retransmission channel are added to an existing system while the control channel for retransmission request uses an existing one.

As discussed in Section 2, it is not easy to predict the size of each MPEG frame. Consequently, it is hard to tell which frame contains the damaged cell, and to what extent the error will be propagated. Therefore, we introduce a mechanism that utilizes a virtual buffer at the receiver through which a selective retransmission can be made when an error occurs by ensuring predictability upon transmission of the MPEG data. The virtual buffer has information about the MPEG data stream such as the frame sequence, type and size. The structure of the virtual buffer is shown on Figure 5.

In case an error occurs during transmission, the possibility of retransmission is checked before the damaged data get into the buffer. In ATM (Asynchronous Transfer Mode), the error detection is made by inspecting the cell sequence number at the ATM Adaption Layer. The process of error detection and retransmission will be covered in Section 4. Once it is determined, through inspection, that retransmission is desirable, a retransmission request is sent to the transmitter through the control channel generally used to send commands such as play, stop, fast forward or rewind. Then, a fresh version of the loss cell is sent back to the receiver through the retransmission channel. When there is no retransmission request, brief information about the MPEG data to be transmitted, such as frame number, type and size are sent to the receiver through the retransmission channel in order to maintain the virtual buffer. It is also assumed that a band of several K-bytes/sec is enough to cover the retransmitted data and that no error occurs during retransmission.

4. Selective Retransmission Method Using a Virtual Buffer

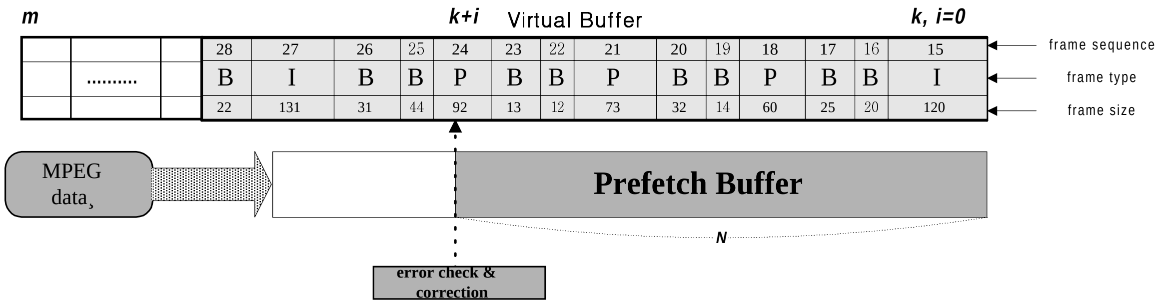
In this section we describe the selective retransmission scheme using a virtual buffer. At the point of time the error occurs, the frame type and the error propagation range can be caught by comparing the data quantity in the (prefetch) buffer with the virtual buffer. The minimum quantity of data in the virtual buffer to be maintained by the receiver will be the number of frames the real buffer can hold at maximum added to the number of frames in a GOP.

A damaged frame is found as follows: starting from the frame in the virtual buffer corresponding to the frame to be presented next, the sizes of frames to be presented are added one by one to the quantity of data in the real buffer at the point of time an error is detected. The frame corresponding to the very point of time at which the sum of frame sizes becomes greater than the size of data in the real buffer is the damaged frame. This is formally expressed as

$$\sum_{i=0}^{m} VB[k+1].frame\_size \rangle N \text{ ....}(1)$$

In the expression, VB is the virtual buffer, k is the sequence of the frame to be presented next, m is the number of frames currently maintained in the virtual buffer and N is the size of the data stream in the prefetch buffer at the point of time the error occurs. The real frame corresponding to the (k+i)th frame in the virtual frame that first satisfies the above expression (1) as i increases is the damaged frame.

The way to find frames affected by an error is to locate the end of the GOP following the referencing chain from the damaged frame in the virtual buffer. For example, in



Ficture 6. Error Detection Using Virtual Buffer

Figure 6, as the error occurred at the 24th P-frame, the error is propagated down to the 26th frame. Since it takes a large buffer and plenty of processing time for the transmitter to send the frame information in real time, we assume the information for the virtual buffer is maintained in a file in advance when the transmitter encodes or pre-processes the MPEG data. Though additional memory and bandwidth is needed, the size is relatively small compared with the massive amount of multimedia data.

After the damaged frame is found, the possibility of recovery through retransmission is examined. Generally, multimedia applications on a network have a fixed amount of prefetching buffer in order to absorb delay and jitter and secure data continuity [4]. Accordingly, there must be some prefetching delay from the moment of the data arrival until presentation. Therefore, an error is recoverable if, from the retransmission request, the time the data returns is less than the prefetching delay. This is expressed as

$$t_p(n) - t_a(n) \ge Max\ Round\ Trip\ Delay \ \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where ta(n) is the time the nth frame arrives at the buffer, and tp(n) is the expected time for the nth frame to be presented. The tp(n) is calculated by counting the number of frames in the buffer using the virtual buffer, and the Max Round Trip Delay is the time it takes for the damaged data to return after the retransmission request gets to the server through the control channel and the server retransmits the data through the retransmission channel. Since the error occurred at an I or P frame is propagated down to the end of the GOP, the errors caused by the damaged I or P frame are recoverable if the damaged data come back within the time all the frames in the GOP are presented. The expression determining retransmission is as follows:

$$t_p(n_{\text{lost picture of GOP}}) - t_a(n) \ge Max\ Round\ Trip\ Delay \ \ldots\ldots\ldots\ldots (3)$$

where tp(nlast picture of GOP) is the time the last frame affected by the error caused by the nth frame is presented. Therefore, if a damaged frame satisfies expression (3), the error is recoverable through retransmission.

Since an ATM network does not provide the data retransmission function, it should be done at the user's level. Further, one way to reduce the number of retransmissions is to avoid retransmitting when B frames are damaged, because errors are not propagated. Though the number of B frames is larger than that of I or P frames, there is relatively less probability of error occurrences at B frames since the size of B-frames is smaller.

Figure 7 shows the selective retransmission algorithm proposed in this paper. N is the number of cells in the buffer at the moment cell loss occurs. The sequence number and type of the MPEG frame containing the damaged cell can be found as we compare N with the sum of each frame size by looking into the virtual buffer (line 2-6). If the damaged frame is of type B, it is ignored; otherwise the possibility of retransmission is tested (line 8). After finding the last frame influenced by the cell loss

1 ***when cell loss occurred***
2   *tot = 0; i=0;*
    { find frame type and frame number influenced by cell loss }
3   ***while*** *(tot < N)* ***do***
4   *tot = tot + VB[k+i].frame_size;*
        { k is next presented frame }
5   *i = i + 1;*
6   ***end while***
7   *n := k + I;*
    { cell loss occurred at k+i th frame }
8   ***if*** *(VB[k+i].frame_type is I or P)* ***do***
        { selective retransmission }
9   *find last frame influenced by* n^{th} *frame using virtual buffer: l ;*
10   *calculate presentation time of* $t_p(l)$;
11   *calculate arrival time of cell loss frame n:* $t_a(n)$;
12   ***if*** *(* $t_p(l) - t_a(n) >$ *maximum round trip delay)* ***do***
13   *request retransmission;*
14   ***end if***
15   ***end if***

Ficture 7. Algorithm for Detecting and Retransmitting Damaged Frames

1 *set Scene_max := desired number of scenes*
2 *set values for compression pattern: L, Q*
3 *set scene length parameter: q*
4 set the values for $\tilde{\mu}I, \tilde{\sigma}I, \mu P, \sigma P, \mu B, \sigma B, a_1, a_2, \sigma_\varepsilon^2$

5 *X = frame size(in cells)*
6
7 ***for** j=1 to Scene_max **do***
8 *generate the length of scene j:* $N_j \sim Geometric(p)$
9 *generate* $\tilde{X}_I(j) \sim \log normal(\tilde{\mu}_I, \tilde{\sigma}_I)$
10 ***for** I=1 to Nj **do***
11 ***for** k=1 to L **do***
12 ***if** k=1, **then***
13 $X = \tilde{X}_I(j) + \Delta_I(i)$, where
14 $\Delta_I(i) = \hat{a}_1\Delta_I(i-1) + \hat{a}_2\Delta_I(i-2) + \varepsilon(i)$
15 ***else***
16 ***if** remainder(k/Q) = 1, **then***
17 $X \sim \log normal(\hat{\mu}_P, \hat{\sigma}_P)$
18 ***else***
19 $X \sim \log normal(\hat{\mu}_B, \hat{\sigma}_B)$
20 ***else if***
21 ***else if***
22 ***else for***
23 ***else for***
24 ***else for***

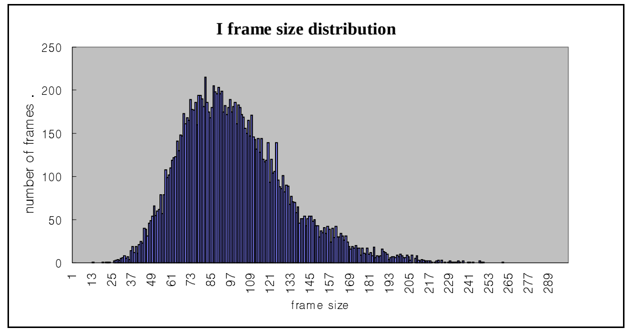
Ficture 8. Pseudocode of MPEG Traffic Generator

using the virtual buffer (line 9), the presentation time of the frame is calculated (line 10). The expected presentation time can be calculated by multiplying the frame period by the number of frames presented. After subtracting the arrival time of the cell loss frame from the expected presentation time of the last frame influenced by the cell loss, if the result is greater than the maximum round trip delay, a retransmission is requested (line 12-13).

In summary, the process of proposed selective retransmission method can be highlighted as follows:

Step 1: confirmation of error occurrence.
Step 2: detection of damaged frame type.
Step 3: finding error propagation range.
Step 4: testing possibility of retransmission.
Step 5: retransmission request through control channel and retransmitting through retransmission channel.
Step 6: recovery and presentation of damaged frame.

5. Simulation



Ficture 9. Distribution of I Frames by Size

Statistic (Kbits)	All	I	P	B
Sample Mean	41.7	197.1	58.0	19.6
Sample STDV	51.7	63	37.3	5.7
Maximum	343.1	343.1	284.6	60.0
Minimum	0.56	36.2	0.56	0.57
GOP	15			
M	3			

Figure 10. Simulation Parameters (1.2 Mbps (MPEG1))

We carried out simulations on a Pentium processor, with 32 megabytes of main memory, running under Windows 95. Simulations were implemented using the 32-bit console application program of MS-Visual C++ 4.0.

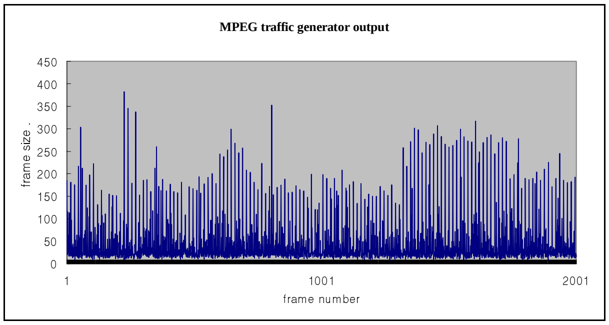
5.1 Traffic Generator and Simulation Model

As seen in the previous sections, MPEG traffic is characterized by different sizes of I, B, P frames because each type of frame uses different data compression method. Even frames of the same type have different sizes because compression rates differ depending on the contents of pictures to be compressed. On the other hand, adjacent MPEG frames are related to each other because similar scenes continue for a certain amount of time and intra-coded frames may appear because of scene changes. So, it is not possible to predict the traffic distribution.

Therefore, for the simulation, we created an MPEG traffic generator using the pseudo-code in Figure 8 following Krunz's experimental results.

According to his experiments, the sizes of I, B, P frames are analogous to lognormal distribution and the time interval from one scene change to another is analogous to geometric distribution. Figure 9 shows the distribution of I frames by size produced by the traffic generator, which is very similar to lognormal distribution.

Initial values for the traffic generator are shown in Figure 10. Those are the mean and standard deviation of all frames and respective I, P, B frames. It is also assumed that the GOP is 15, M is 3 and the interval between frames is 34 ms. Pseudo MPEG traffic generated from the input parameter amounts to about 1.2 Mbps on average.



Ficture 11. Output of Traffic Generator

a	do not retransmit
b	selectively retransmit
c	retransmit only when, if retransmitted, the current frame is recoverable
d	unconditionally retransmit if an error occurs

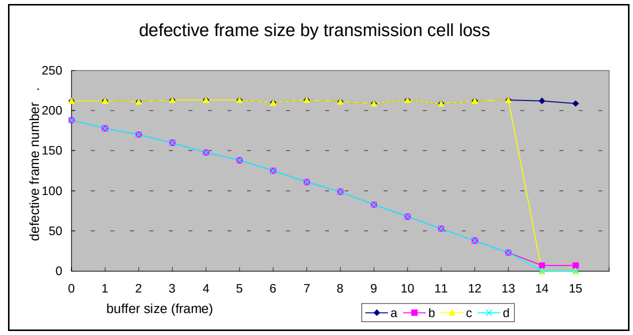
Ficture 12. Simulation Cases

Figure 11 shows an output of 2,000 frames produced from the generator based on the input parameters in Figure 10.

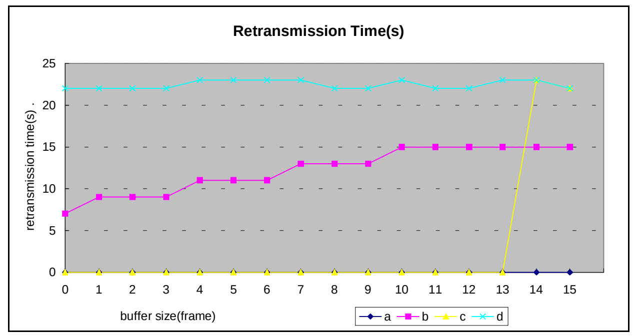
We tested a case of presenting 2-hours, that is, 216,000 frames of MPEG data having a traffic pattern as in Figure 11. We also assumed that the frequency in error occurrence on network is 10-6, the Round Trip Delay is 450 ms, and errors do not occur in the retransmitted data because the amount is very small. Four cases as in Figure 12 were simulated, comparing, as the buffer size increased, the number of frames affected by cell loss, the number of retransmissions, and the number of damaged frames after an error was propagated to more than 3 frames.

5.2 Simulation Results

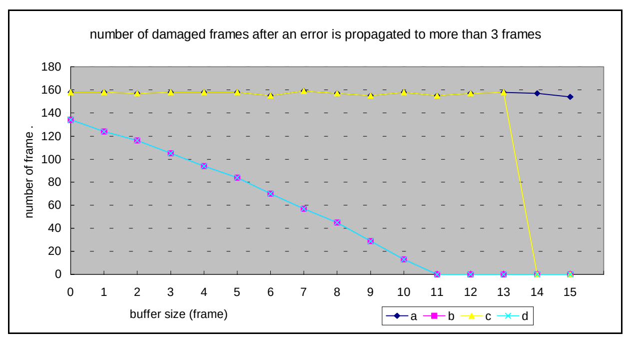
Figure 13 shows the number of frames affected by cell loss as the buffer size increases. In case (a), there is no change in the number of damaged frames regardless of buffer size because no retransmission is made. In case (b), the method proposed in this paper, where frames are retransmitted comparing the buffer size with the Round-Trip Delay and GOP, the number of retransmittable frames increases as the buffer size increases. Accordingly, the number of damaged frames decreases. However, the number of lost frames keeps steady after the frame size goes beyond 15 (approximately 500 ms), since damaged B frames, which are negligible, are not retransmitted. In (c), the number of damaged frames becomes 0 at the frame size 14 where the retransmission gets started since the Round-Trip Delay becomes larger than the buffer size. In (d), the number decreases as in case (b) until frame size 15, after which point, it becomes 0. When the buffer size is small, there is no big difference among cases (a), (b) and (d) because errors can



Ficture 13. Number of Frames Affected by Cell Loss



Ficture 14 . Number of Retransmission



Ficture 15 . Number of Damaged Frames After An Error Is Propagated to More Than 3 Frames

not be recovered through retransmission.

Figure 14 shows the change in the number of retransmissions as the buffer size increases. In case (a), the number is always 0 because there is no retransmission. In case (b), the proposed method, the number of retransmissions increases in inverse proportion to the number of frames affected by cell loss. The reason why a stepped pattern appears from 0 to 10 is that errors in B frames are ignored upon the retransmission possibility test in GOP.

In (c), the retransmission starts at 14 because the Round-Trip delay gets larger than the buffer size. In (d), the number of retransmissions is largest all the time since unconditional retransmissions are made when errors occur. When the buffer size is small, the number of retransmissions decreases because there is little chance of recovery through retransmission.

Figure 15 shows the number of damaged frames after an error is propagated to more than 3 frames. Generally, one frame loss is unrecognizable, so only the cases of more than 3 consecutive losses were considered. Case (a) shows no change in the number of losses. Case (b) is similar to case (d) which is the unconditional retransmission.

According to the experiments, the selective retransmission method, where only the data that can contribute to the error recovery are chosen to retransmit, showed results similar to the unconditional retransmission method in spite of the smaller number of retransmissions. However, the proposed method is hard to apply to cases having serious network delay because the possibility of error recovery through retransmission is reduced. The number of damaged frames or retransmissions was relatively small because approximately 1.2 Mbps of MPEG data were used in the simulation. But, in a high-speed network environment where several or scores of Mbps high-quality video can be transmitted, it is expected that the number of retransmissions will increase as the frequency of error occurrences increases.

6. Conclusion

In this paper, we proposed a selective retransmission method where the number of retransmissions can be reduced without losing QoS and data continuity in multimedia applications using MPEG-2. Error propagation due to the MPEG frames' cross-referencing mechanism was taken into account in the method. Retransmissions are executed only when errors are recoverable through them. This can be done by catching hold of the error propagation range in advance. To achieve the goal, we introduced a virtual buffer maintaining information about the data to be presented such as sizes and types of frames. The proposed method works well when the network delay is less than the transmission time of the GOP and is suitable for multimedia applications where minor damage in pictures may seriously affect the QoS.

References

[1] V. Bhaskaran, K. Konstantinides, Image and Video Compression Standards Algorithms and Architecture, Kluwer Academic Publishers, 1995

[2] Miguel Correia, Paulo Pinto, "Low-Level Synchronization Algorithms on Broadband Networks", IEEE Proceeding of MULTIMEDIA'95, pp. 423-434, 1995.

[3] Ching-Chin Han, Kang G. Shin, "Scheduling MPEG-Compressed Video Streams with Firm Deadline Constraints", IEEE Proceeding of Broadband Networks", IEEE Proceeding of MULTIMEDIA'95, pp.423-434, 1995.

[4] Joseph Y. Hui, Ezhan Karasan, Jun Li, Junbiao Zhang, "Client-Server Synchronization and Buffering for Variable Rate Multimedia Retrievals", IEEE Journal on Selected Areas in Communications, Vol.14, No.1, pp.226-237 Jan. 1996.

[5] Marwan Krunz, Ron Sass, Herman Hughes, "Statistical Characteristics and Multiplexing of MPEG Streams", Proceeding of INFOCOMM'95, pp.455-462, 1995.