Adaptable Packet Significance Determination Mechanism for H.264 Videos over IP Dual Stack Networks

Chu-Chuan Lee
Chunghwa Telecommunication Laboratories,
Chinese Taipei

Email: cclee@vaplab.ee.ncu.edu.tw

Ya-Ju Yu and Pao-Chi Chang
Department of Communication Engineering,
National Central University,
Chinese Taipei
Email: {yjyu, pcchang}@vaplab.ee.ncu.edu.tw

***Abstract*—The video streaming applications are full of potentials in the IP dual stack network that supports IPv4 and IPv6 protocols simultaneously. However, the significances of video packets belonged to various video sequences are different. An equal error protection to all video packets in the IP network will degrade the video quality significantly. This paper proposes an Adaptive Significance Determination Mechanism in Temporal and Spatial domains (ASDM-TS) for H.264 videos over IP dual stack network with DiffServ model. ASDM-TS determines the video packet significance simultaneously in temporal and spatial domains. From the temporal domain, ASDM-TS evaluates the packet significance based on the estimated error propagation if a packet is lost. From the spatial domain, ASDM-TS computes the packet significance based on the content complexity belonging to a packet. Moreover, ASDM-TS is adaptive to various video sequences with a self-learning method. Compared with traditional schemes, simulation results show that the proposed scheme significantly improves the accuracy of signification determination up to 15% and effectively improves the received video quality up to 0.7dB in PSNR.**

I. Introduction

There is a strong consensus today that IP will be the foundation of next-generation networking [10]. The continuous growth of the Internet world requires that its overall architecture can evolve to accommodate new technologies for satisfying the growing numbers of users, applications, and devices. IPv6 is designed to satisfy these requirements and allow the return to a global end-to-end environment where the addressing rules of network are transparent to applications again [1]. IPv6 quadruples the number of network address bits from 32bits of IPv4 to 128bits, which provides enough globally unique IP addresses for every network device on the planet. In addition to unlimited IP addresses, IPv6 also enhances the QoS capability with the Traffic Class field and the Flow Label field [3], [8]. On the other hand, most video coding methods exploit both temporal and spatial redundancies to reduce required transmission rate and achieve high compression efficiency. In the spatial domain, there exists a high correlation within a picture. In the temporal domain, there usually exists a high similarity between successive pictures. However, the received video quality is highly sensitive to packet loss. When a video packet that belongs to I-frame is lost due to network congestion, all frames belonging to the same GOP (Group of Picture) are hurt due to error propagation in the decoding process. This phenomenon causes significant degradation of received picture quality. Moreover, all succeeding frames belonging to the same GOP are also hurt if a video packet that belongs to P-frame is lost. Therefore, a robust network mechanism that can provide sufficient protections to video data is essential for received picture quality [11].

Unfortunately, the default QoS (Quality of Service) strategy of Internet is the best-effort transmission, which is lack of QoS guarantee to encoded video data. The Internet Engineering Task Force (IETF) had proposed the Differentiated Services (DiffServ) [9] to solve the problem and to manage the allocation of limited network bandwidth [7], [12]. In DiffServ, packets are assigned and classified to one of several classes. However, the fact that every video packet has various significance and different picture quality influence in the video decoding process generally complicates DiffServ operations The congestion loss of video packets is possible in the DiffServ network if too many packets with the same class simultaneously arrive at an output port of router/switch. Moreover, the degree of picture quality degradation due to packet loss is different among various video sequences. It is unsuitable if a fixed packet significance classification scheme is utilized to all video sequences regardless of the difference of picture complexity among videos. To prevent the unexpected packet loss of significant video frames such as I-frames in DiffServ network, an unequal priority assignment scheme is required for video packets at the video sender side. The priority of video packet also implies the distortion effect induced by packet loss.

Many research results were developed in past years. In [6], the intra-refreshed MacroBlock (MB) technique is used to alleviate the error propagation. In the spatial domain, the content of each packet is evaluated to determine the packet significance, according to the ratio of the number of intra-refreshed MBs to the total number of MBs in a packet. However, the error propagation effect of each packet in the

$$QD_{k,i} = D_{k,i} + \sum_{j=1}^{P} \frac{r^{NP_{k+1,j}} - 1}{r-1} \cdot r \cdot D_{k,i} \cdot \Re_{k+1,j} \tag{1}$$

$$\Re_{k+1,j} = \frac{\text{Number of pixels in } i\text{-th packet refered by } j\text{-th packet of } k+1\text{-th frame}}{\text{Total number of pixels in } i\text{-th packet of } k\text{-th frame}} \tag{2}$$

$$NQD_{k,i} = \frac{QDR_{k,i}}{\max_{k,i}(QDR_{k,i})} = \frac{r-1}{r^N - 1} + \sum_{j=1}^{P} \frac{r^{NP_{k+1,j}} - 1}{r^N - 1} \cdot r \cdot \Re_{k+1,j}, \text{ where } QDR_{k,i} = \frac{QD_{k,i}}{D_{k,i}} \tag{3}$$

temporal domain is not examined in [6]. In [4], the authors analyzed the distortion effect of each lost MB and utilized a fixed model to approximate the statistic results. Nevertheless, a fixed model cannot satisfy the varied properties of video sequences. In [2], the authors determined the priority of a video packet according to the evaluation from temporal and spatial domains simultaneously. A weighting factor α $(0 \leq \alpha \leq 1)$ was used to decide the proportion of spatial domain and temporal domain considerations. However, the value of α is not easy to decide since it depends on the complexity of video sequence. In [5], the error propagation influence of each video frame is estimated at the sender side according to its temporal position in a GOP. However, the method in [5] assumed the different video sequences have the same error propagation influences.

To solve the above problems, this paper proposes an Adaptive Significance Determination Mechanism in Temporal and Spatial domains (ASDM-TS) for H.264 videos over IP dual stack network with DiffServ model. ASDM-TS determines the video packet significance simultaneously in temporal and spatial domains. From the temporal domain, ASDM-TS evaluates the packet significance based on the estimated error propagation if a packet is lost. From the spatial domain, ASDM-TS computes the packet significance based on the content complexity belonging to a packet. Moreover, ASDM-TS is adaptive to various video sequences with a self-learning mechanism. Compared with traditional schemes, simulation results show that the proposed mechanism can significantly improve the accuracy of signification determination up to 15% and effectively improve the received video quality up to 0.7dB in PSNR. The detailed process of the proposed scheme is presented in Section 2. Simulation environment and results are discussed in Section 3. Finally, Section 4 concludes this paper.

II. Packet Significance Classification among video flows

The scheme in [5], where we refer to it as Frame Based Classification (FBC) in this paper, assumed that the error propagation effect exhibits a nearly linear relationship to the following frames. As shown in following results in this section, this study observes that using linear and fixed mathematics model cannot accurately describe the behaviors of multiple video sequences.

To classify the significance difference among packets in the same frame and enhance the approximation accuracy of error propagation influence, this study utilizes a geometric series to describe the estimated total Quality Degradation (QD) due to error propagation when the i-th packet of k-th frame is lost, as shown in (1).

Where $D_{k,i}$ is the actually local distortion due to the loss of the i-th video packet of k-th video frame, $\sum_{j=1}^{P} \frac{r^{NP_{k+1,j}}-1}{r-1} \cdot r \cdot D_{k,i}$ is the quality degradation due to behaviors of error propagation, NP is the number of propagation frames $(0 \leq NP \leq N-k)$, N is the GOP size in a video sequence, and P is the total packet number of a frame. Regarding the H.264 error resilient operations, the Cyclic Intra Refresh (CIR) and the Random Intra Refresh (RIR) mechanisms are provided. This study uses the CIR method and computes the Number of Propagated frames (NP) between the frame having a packet loss and the frame enabling the intra refresh coding operation in the same vertical position of the lost packet. The proposed method always buffers the current (k-th) and next ($k+1$-th) frames and computes the reference ratio ($\Re$) of the j-th packet of $k+1$-th frame to the i-th packet of current k-th frame. $\Re_{k+1,j} = 1$ means that all pixels of i-th packet of current k-th frame are referred by j-th packet of next $k+1$-th frame.

Moreover, this study defines the Quality Degradation Ratio (QDR) to describe the degree of error propagation and then defines the Normalized Quality Degradation (NQD) to normalize the degree of error propagation ratio, as shown in (3). We can adjust r, the common ratio of a geometric series, to obtain an estimated NQD distribution, which can approximate the actual NQD distribution of a video sequence. In our proposed method, different video sequences use various values of r, where r is renamed as error propagation ratio in this paper.

As shown in Fig. 1, this study can utilize (3) and adjust the value of r to obtain three approximate curves for Stefan, News, and Akiyo video sequences, respectively. The values of r are set to 0.85, 1.06, and 1.28 for Stefan, News, and Akiyo, respectively. Obviously, no actual NQD distributions of three video sequences exhibit a linear relationship.

Given (4), the approximated QD value for the i-th packet of k-th frame, which is called Significance Index (SI) in this paper, can be easily obtained by

$$SI_{k,i} = NQD_{k,i} \cdot D_{k,i} \tag{4}$$

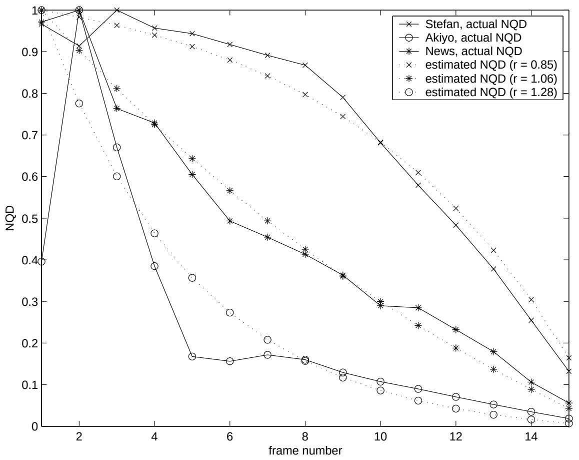


Fig. 1. Actual and estimated significance curves of various video sequences

Using (4), we can obtain the estimated quality degradation in the unit of packet for video sequences easily and effectively.

Although (4) can provide the estimated QD value to each packet of a video sequence effectively, the current proposed method still requires selecting the suitable value of r for a given video sequence by complicated manual process. Therefore, this study includes a self-learning algorithm to ASDM-TS. The improved operation of ASDM-TS is independent of the video sequence type and the value of r is automatically adjusted while each GOP begins in a video sequence.

The flowchart of the self-learning algorithm is shown in Fig. 2. Each GOP has two values of r in ASDM-TS. One is the estimated r and the other is the actual r, which are expressed as r_{next} and r_{new}, respectively. The value of r_{next} is given when a GOP begins and is used to calculate the SI value of each packet in the GOP. On the other hand, the value of r_{new} is automatically computed whenever the encoding process for the GOP is finished. In general, the computed r_{new} of i-th GOP will be the r_{next} of $(i+1)$-th GOP directly. However, to reduce the undesirable oscillation phenomenon of r_{next}, a smoothing process is utilized in ASDM-TS. If the difference between the r_{new} computed from current i-th GOP and the ”Mean” value calculated from previous r_{new} is less than a threshold, the r_{new} computed from i-th GOP will be the r_{next} of $(i+1)$-th GOP directly. In contrast, if the above condition violates, ASDM-TS will average the computed r_{new} of i-th GOP to the ”Mean” value and the result is as the r_{next} of next GOP. In ASDM-TS, the initial r_{next} and ”Mean” value are set to 0.99 for the first GOP of a video sequence and the current used r is denoted as $r_{current}$ in Fig. 2. Note that the initial value of 0.99 is determined by averaging the values of r of video sequences that belong to the Class B video type. In addition, to decrease the computational complexity for obtaining r_{new}, ASDM-TS uses four frames, including the 2nd, 6th, 9th, and 13th frames of a GOP, to determine the value of r_{new}, where the positions of four frames are located to front, center, and rear of GOP.

The robustness of ASDM-TS is verified in Fig. 3, where three methods are compared with each other. Regarding the ASDM-TS-upbound method, i-th GOP is buffered at first, and the r_{new} of i-th GOP is calculated and then assigned to the $r_{current}$ of i-th GOP. Although the current i-th GOP can use the most suitable value of r to estimate the QD values of packets, extra delay and additional computation complexity are generated in ASDM-TS-upbound method. However, the results of ASDM-TS-upbound can act as the upper bound in this simulation scenario. The second method is the ASDM-TS-directly mechanism that the smoothing operations for avoiding oscillation of r_{next} are inhibited, where this study intends to examine the influence of oscillation of r_{next}. In Fig. 3, we observe that ASDM-TS-directly method generates obvious oscillation of r in the range of 6th - 9th GOPs in Stefan sequence and the difference between ASDM-TS-upbound and ASDM-TS-directly methods is explicit. Similar situation results also happen in the case of Akiyo. In contrast to ASDM-TS-directly method, the proposed ASDM-TS algorithm decreases the oscillation effectively and the generated results of r are close to that of ASDM-TS-upbound method. Even though some scene changes happen, the proposed ASDM-TS algorithm still works well.

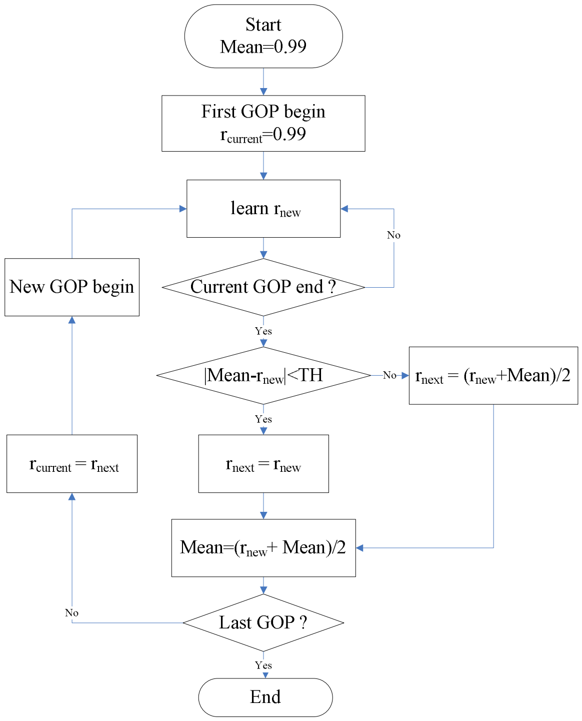
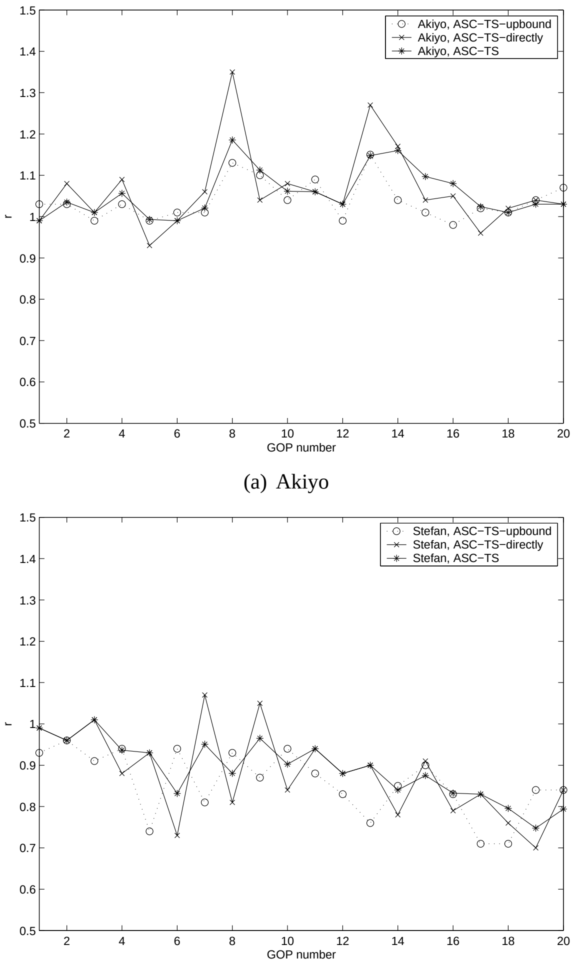


Fig. 2. The flowchart of ASDM-TS

III. Simulation Results

A. Definition of packet classification accuracy

During the process, the grade of estimated SI values for video packets may appear in a wide range. Therefore, the mapping between the SI values and the limited Differential Service (DiffServ) levels are required, which is so-called QoS mapping. For comparison, we define the accuracy of packet



(a) Akiyo

(b) Stefan

Fig. 3. Comparison of ASDM-TS performance

QoS mapping as:

$$Accuracy = \frac{A_{num}}{M} \cdot 100\% \tag{5}$$

where

$$A_{num} = \sum_{i=1}^{M} a(i),\ a(i) = \begin{cases} 1, & \text{if } q(i) = \hat{q}(i) \\ 0, & \text{otherwise} \end{cases}$$

Assuming DiffServ network can provide Q aggregated levels (DiffServ levels), $q(i)$ is the DiffServ level to which the i-th packet is mapped, where $1 \leq q(i) \leq Q$, $1 \leq i \leq M$, and the total packet number of a video sequence is M. In our simulation scenarios, the significance classification of packets and the corresponding QoS mapping are executed based on the actual QD values and the estimated QD results obtained from ASC-TS, respectively. While actual QD values are used, the QoS mapping results are denoted by $\vec{q} = \{q(1), q(2), \ldots, q(M)\}$. When the results of ASC-TS are used, the QoS mapping results are denoted by $\hat{\vec{q}} = \{\hat{q}(1), \hat{q}(2), \ldots, \hat{q}(M)\}$.

TABLE I
ACCURACY OF PACKET CLASSIFICATION IN CASE OF LAB(%)

	FBC	ASC-TS-directly	ASC-TS	ASC-TS-upbound
Akiyo	85.68	93.57	93.91	94.89
Container	85.07	93.78	94.17	95.33
Foreman	76.61	88.54	89.11	90.98
News	88.13	93.41	93.7	94.67
Stefan	80.7	91.81	92.78	94.67
Bus	78.38	91.67	92.13	93.98

TABLE II
ACCURACY OF PACKET CLASSIFICATION IN CASE OF MAB (%)

	FBC	ASC-TS-directly	ASC-TS	ASC-TS-upbound
Akiyo	84.39	93.65	94.13	95.56
Container	84.67	93.33	93.56	94.87
Foreman	75.44	89.76	90.13	91.54
News	90.2	94.39	95.17	96.15
Stefan	80.07	91.31	91.78	94.41
Bus	75.79	91.85	91.94	93.29

TABLE III
ACCURACY OF PACKET CLASSIFICATION IN CASE OF HAB (%)

	FBC	ASC-TS-directly	ASC-TS	ASC-TS-upbound
Akiyo	84.57	93.96	94.37	95.41
Container	85.46	92.33	93.07	94.52
Foreman	76.94	90.28	90.93	93.02
News	91.76	93.72	94.65	96.06
Stefan	82	92.56	93.19	95.22
Bus	75.37	92.18	92.27	93.7

B. Simulation environment and results

In this experiment, we use Network Simulator version 2 (NS-2) to simulate the managed IP network, and use H.264 JM10.2 codec to compress videos at a target rate of 1M bps. The length of video sequence is 300 frames. In addition, the IPPP video format with GOP size 15 frames is adopted in encoding, and P frames only refer to previous one frame. Three network conditions are considered in the following simulation scenarios, including low, medium, and high available bandwidth cases (LAB, MAB, and HAB). In addition, three DiffServ levels are chosen in this paper.

Using the accuracy defined in (5), Table 1, Table 2, and Table 3 show the accuracy comparison of packet classification between the proposed ASDM-TS mechanism and traditional FBC method in cases of LAB, MAB, and HAB, respectively. In these tables, two additional ASDM-TS-directly and ASDM-TS-upbound methods defined in session 2 are used for comparison. In these simulation results, we observe that the accuracy of packet classification using the proposed ASDM-TS is better than that of FBC up to 15%. Moreover, as mentioned in Section 2, the computational complexity of the proposed ASDM-TS is less than that of ASDM-TS-directly method. However, the accuracy of packet classification using ASDM-TS is better than that of ASDM-TS-directly method up to 1%.

IV. Conclusion

The data of various video sequences always exhibits different significances and different error propagation characteristics. Using a fixed model to classify the priorities of video data for various sequences is ineffective and thus degrades the received video quality due to undesirable loss of important video packets. The proposed ASDM-TS mechanism adaptively and effectively solves above problems by evaluating the significance of video packets in temporal and spatial domains simultaneously with a self-learning process. Compared with traditional FBC scheme, the proposed mechanism can significantly improve the accuracy of significance classification up to 15%. Moreover, delivering video data with ASDM-TS on IP dual stack DiffServ network outperforms FBC priority strategy up to 0.7dB in PSNR.

References

[1] A. Durand. “Deploying IPv6”. *IEEE Internet Computing*, 5(1):79–81, Jan.-Feb. 2001.

[2] C. C. Lee, P. C. Chang, and S. J. Chuang. “Unequal Priority Arrangement for Delivering Streaming Videos over Differentiated Service Networks”. *Lecture Notes in Computer Science (LNCS)*, 4319(1):812–821, Dec. 2006.

[3] E. B. Fgee, J. D. Kenney, W. J. Phillips, W. Robertson, and S. Sivakumar. “Implementing an IPv6 QoS management scheme using flow label & class of service fields”. *Canadian Conference on Electrical and Computer Engineering*, May 2004.

[4] F. D. Vito, D. Quaglia, and J. C. D. Martin. “Model-based distortion estimation for perceptual classification of video packets”. *IEEE 6th Workshop on Multimedia Signal Processing*, 2004.

[5] F. Zhang, M. R. Pickering, M. R. Frater, and J. F. Arnold. “Streaming MPEG-4 Video over Differentiated Services Network”. *Workshop on Internet, Telecommunication and Signal Processing, Wollongong,*, Dec. 2002.

[6] G. Cote and F. Kossentini. “Optimal Intra Coding of Blocks for Robust Video Communication over The Internet”. *Signal Processing on Image Communication*, Sep. 2009.

[7] I H. Huang, C. S. Lin, C. S. Chen, and C. Z. Yang. “Design of a QoS Gateway with Real-time Packet Compression”. *Annual IEEE Region 10 Conference (TENCON)*, Nov. 2007.

[8] M. Tatipamula, P. Grossetete, and H. Esaki. “IPv6 integration and coexistence strategies for next-generation networks”. *IEEE Commun. Mag.*, 42(1):88–96, Jan. 2004.

[9] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. “An Architecture for Differentiated Services”. *RFC2475*, Dec. 1998.

[10] S. I. Maniatis, E. G. Nikolouzou, and I. S. Venieris. “End-to-end QoS specification issues in the converged all-IP wired and wireless environment”. *IEEE Commun. Mag.*, 42(6):80–86, June 2004.

[11] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra. “Overview of the H.264/AVC video coding standard”. *IEEE Trans. on Circuits System for Video Technology*, 13(7):560–576, July 2003.

[12] V. Fineberg,. “A Practical Architecture for Implementing End-to-End QoS in an IP Network”. *IEEE Commun. Mag.,*, 40(1):122–130, Jan. 2002.