Computer Networks 92 (2015) 134–147

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Real-time video streaming using prediction-based forward error correction

Yung-Tsung Weng[a], Chi-Huang Shih[b,*], Chun-I Kuo[a], Ce-Kuen Shieh[a]

[a] Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan, ROC

[b] Department of Computer Science and Information Engineering, Hung Kuang University, Taichung, Taiwan, ROC

A R T I C L E   I N F O

Article history:
Received 25 March 2015
Revised 17 September 2015
Accepted 19 September 2015
Available online 3 October 2015

Keywords:
Forward error correction
Video traffic prediction
Real-time video streaming

A B S T R A C T

Real-time video streaming applications typically use an on-line forward error correction (FEC) technique to recover transmission losses with a low delay overhead. However, transmitting prioritized video data over variable-rate transmission channels complicates the FEC rate allocation process. Specifically, on-line FEC schemes result in an inefficient utilization of the available FEC bandwidth in the absence of prior information regarding the statistics of video traffic. In most streaming networks, the optimal FEC configuration is computed off-line in accordance with an analytical model. However, the present study proposes an on-line FEC scheme in which real-time FEC allocation is performed by extending the analytical FEC model with the frame size prediction technique. In the proposed approach, the optimal FEC configuration is computed in advance on a frame-by-frame basis over a series of predicted video frames, thereby yielding a significant reduction in the data buffering delay. The performance effects of frame-size prediction errors are mitigated by continuously revising the FEC configuration each time a new frame arrives. Moreover, a transmission rate control mechanism is proposed to ensure that each video frames satisfies its presentation deadline. The simulation results show that the proposed prediction-based FEC scheme minimizes the FEC processing delay while achieving virtually the same perceived video quality as that obtained using the off-line optimal FEC model.

1. Introduction

Transmitting real-time video streams over the Internet is an extremely challenging task due to the need to meet stringent delay and packet loss requirements. The detrimental effects of packet losses on the perceived visual quality of the received video stream in end-to-end networks are generally mitigated using some form of forward error correction (FEC) scheme [1–11]. In application-layer FEC schemes, the source packets are grouped into blocks of a predetermined size at the sender end and redundant packets are then added to each block such that if losses occur during transmission, the source data can still be reconstructed at the receiving end. Compared with the automatic repeat request (ARQ) protocol, in which lost packets are retransmitted through an end-to-end acknowledgement mechanism, FEC is generally preferred for real-time video applications due to its lower end-to-end delay.

Recently, dynamic adaptive streaming over HTTP (DASH) is proposed as an adaptive bitrate streaming technique that splits video stream with different bit rates into small HTTP-based files for Internet transmissions [31]. Since HTTP is based on TCP, which provides the packet transport reliability by means of acknowledgement and retransmission, DASH is unsuitable for the real-time video applications using UDP.

* Corresponding author. Tel.: +886-4-26318652x6148; fax: +886-4-26324084.
E-mail address: chshih@sunrise.hk.edu.tw (C.-H. Shih).

http://dx.doi.org/10.1016/j.comnet.2015.09.020
1389-1286/

In this paper, we focus on UDP-based video streaming and utilize FEC to achieve transmission robustness.

The effectiveness of FEC schemes in transmitting priority-encoded video over bandwidth-constrained channels is reliant on an efficient control of the redundant packets. In the H.264/AVC coding formats, the video frames are assigned different levels of importance in accordance with their coding dependency. Compressed video sources are generally divided into two major categories, namely variable-bit-rate (VBR) stream and constant-bit-rate (CBR) stream. For VBR stream, the number of bytes in each encoded frame varies as a function of the scene complexity and the frame type used. For CBR stream, each encoded frame has a constant number of bytes by varying quantization matrix. When the VBR or CBR encoding method is considered in video service provisioning, the video stream is typically unaware of network changes and can experience the large losses and/or delays within a concentrated period [12]. As a result, a large number of redundant packets are required to ensure an acceptable video quality at the receiver end. However, the additional redundancy cost prompts a self-induced congestion effect due to buffer overflows and an increased queuing delay [13]. To mitigate this problem, the authors in [14,15] presented a TCP-friendly rate control (TFRC) mechanism designed to provide a congestion control mechanism for UDP-based video flows and a fair bandwidth sharing with TCP flows can be achieved accordingly. In practical end-to-end networks, the available FEC bandwidth is limited, and hence a differentiated redundancy allocation strategy (referred to as *unequal error protection*, or *UEP* for short) is used to smooth the quality degradation in the presence of high packet loss rates. More specifically, high priority video frames are allocated a greater number of redundant packets than low priority frames in order to ensure their successful reconstruction and decoding.

In practice, determining the optimal UEP configuration relies on a knowledge of the video frame size, packet loss rate and network transmission rate. Wu et al. [16] derived an analytical FEC model based on a temporal scaling concept and a TCP-friendly transmission rate constraint for optimizing the reconstruction quality of the group of pictures (GOP) in MPEG video streams. In the proposed model, redundant packets are injected at the frame level and the stream data rate is dynamically adjusted by discarding frames in accordance with their coding dependency [32,33]. To obtain a lower computational delay than that achieved using distortion-based methods [7,20,21], the video reconstruction quality in Wu's FEC model is estimated using a playable frame rate (PFR) metric, defined in terms of the expected number of playable frames at the receiver. A frame was considered to be playable if the entire frame was received correctly after FEC recovery and all of the frames upon which it depended were also playable. Yuan et al. [17] applied FEC at the GOP level rather than the frame level in order to obtain a larger PFR. However, the performance improvement is obtained at the expense of a greater computational complexity on average due to the larger volume of video data which must be processed.

The FEC models described above regard the transmission channel as a random binary symmetric channel (BSC) and assume that the packet losses are independent. By contrast, Kuo et al. [18,19] proposed a PFR-based FEC scheme based on a two-state Markov chain model of burst-loss transmission channels. Li et al. [20,21] presented an FEC transmission scheme based on the estimated mean-square-error (MSE) of the distortion caused by Markov model burst packet losses. In the proposed approach [21], the complexity of the MSE estimation process was reduced by means of a sliding window algorithm.

Given a knowledge of the video traffic pattern and the network transmission parameters, analytical FEC models provide an effective means of calculating the optimal UEP configuration. However, such models are generally implemented off-line since to perform the FEC allocation process in an on-line environment requires a buffer of a sufficiently large size to collect the entire GOP. The large buffer results in a coding buffering delay on the order of the GOP duration at the video receiver end before smooth video presentation can begin. As a result, the overall end-to-end delay of the video stream is significantly increased. In real-time video applications, however, the end-to-end delay should not exceed 150 ms if an adequate video quality is to be obtained [22].

Accordingly, the present study proposes an on-line FEC rate allocation scheme which minimizes the coding buffering delay through the use of an efficient frame size prediction technique. Based on previous video information, the proposed scheme constructs a virtual GOP in advance for the oncoming GOP and then uses an extended version of the PFR-based analytical model presented in [16] to calculate the optimal rate allocation results for the virtual GOP. When the oncoming GOP is actually processed, the FEC configuration is determined on a frame-by-frame basis based on the optimal results calculated for the virtual GOP. In addition, a transmission rate control mechanism is proposed to regulate the output data rate in such a way as to satisfy the presentation deadline of each frame subject to the specified transmission rate constraint. The main contributions given in this paper are: (1) A frame size predictor is utilized to predict the frame sizes within the oncoming GOP for the significant reduction in the data buffering delay. (2) A virtual GOP is built to calculate the optimal FEC allocation and temporal scaling decision. (3) A fast search algorithm is developed to reduce computation complexity in the optimization process.

Video traffic prediction is essential to satisfy the quality of service (QoS) requirements of real-time VBR video transmissions and to maximize the overall network utilization efficiency [23–28]. Broadly speaking, existing video traffic predictors can be classified as model-based predictors, least-mean-square (LMS)-based predictors, and neural network (NN)-based predictors. Model-based predictors use a stochastic video traffic model to capture the video content characteristics and then use this model to predict the future traffic trace from the current and past values for a given set of input parameters [23,24]. By contrast, LMS-based predictors use a linear combination of the current and previous traffic trace values to predict the traffic trace in the following frame and subsequently adjust the LMS parameters for the next prediction based on the difference between the predicted value and the actual value [25,26]. Finally, NN-based predictors utilize a multi-resolution learning paradigm to predict up to as many frames into the future as required [27,28]. Model-based predictors and NN-based predictors have the advantages of accurate prediction and long-term traffic