Computer Communications 35 (2012) 748–758

Contents lists available at SciVerse ScienceDirect

Computer Communications

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

Feed Forward Bandwidth Indication (FFBI): Cooperation for an accurate bandwidth forecast

Rami Haddad [a], Michael P. McGarry [b,*]

[a] *Dept. of Mechanical and Electrical Engineering, Georgia Southern University, 1332 Southern Drive, Statesboro, GA 30458, United States*
[b] *Dept. of Electrical and Computer Engineering, University of Texas at El Paso, 500 West University Ave., El Paso, TX 79968, United States*

A R T I C L E I N F O

Article history:
Received 10 August 2011
Received in revised form 27 November 2011
Accepted 6 January 2012
Available online 18 January 2012

Keywords:
Bandwidth forecasting
Bandwidth prediction
Multimedia
Video

A B S T R A C T

Video bandwidth forecasts can empower video transport mechanisms with a new intelligence that can increase the efficiency of Dynamic Bandwidth Allocation. We exploit the fact that for pre-recorded video, the size of every video frame is known prior to the video being delivered. We propose Feed-Forward Bandwidth Indication (FFBI) which feeds video frame sizes forward in a sequence of video frames. We extend FFBI to live video by introducing a delay at the source equivalent to the forecast window. We compare FFBI to the most accurate forecast methods found in the literature. With network transport of video projected to supplant other transport mechanisms over the next few years, we conduct a performance analysis of FFBI within Ethernet Passive Optical Networks (EPONs). We find that the use of FFBI can provide a 50% reduction in queueing delay compared to the use of no forecasting and a 35% reduction in queueing delay compared to other forecasting methods. In addition, we find that FFBI can provide a very significant reduction in queueing delay variation compared to the use of no forecasting or other forecasting methods.

1. Introduction

According to a report published by the Discovery Institute [1] the third phase of the Internet is being driven by increased bandwidth to subscribers and growing popularity of communicating video information. Optical access networks such as Ethernet Passive Optical Networks (EPONs) are providing increased bandwidth to individual subscribers. Cisco Systems has projected that video traffic, in all of its forms, will constitute 90% of consumer IP traffic in the very near future [2]. Bandwidth usage forecasts are clearly indicating that video will be the dominant information exchange in communication networks of the future. Therefore, the future success of packet switched networks hinges upon the discovery, development, and deployment of methods that can successfully deliver video content over packet switched networks.

Video content exists in three forms. There is *pre-recorded* on-demand video content (e.g., YouTube, Hulu, broadcast networks, IPTV), *live broadcast* video content (e.g., live content from the broadcast networks), and *live interactive* video content (e.g., video telephony). These different forms of video content have similarities (e.g., encoding) as well as dissimilarities (e.g., delay constraints). Most video is compressed to reduce bandwidth requirements.

A characteristic of compressed digital video, that is common among all of its forms, is its time varying frame sizes. Each frame size is dependent on the amount of spatial and temporal information that a compression algorithm can remove at any instant in time while maintaining a certain fidelity to the original. A compression algorithm can remove redundant information as well as information determined to be unimportant for a desired fidelity to the original information. A compression algorithm that only removes redundant information is referred to as loss-less, otherwise it is referred to as lossy.

The most popular standardized video compression algorithms (e.g., MPEG-2, H.263, H.264) are all lossy compression algorithms. The amount of redundant and unimportant information at some instant in time depends on the content of the video at that time. In general, the precise probability distribution of frame sizes is unknown although it is generally considered to have an asymmetrical bell shape that has a right tail with more weight than the left tail [3,4]. Frame size distributions have a large difference between the mean and maximum values [5,6]. As an example, a particular 1080p resolution video compressed using the H.264 standard has a mean frame size of 26973.52 bytes and a maximum frame size of 379005 bytes; the maximum is 14.05 times greater than the mean.

The bandwidth requirement, over a frame period, for transmitting a video is determined by each frame size divided by the frame period. The frame period is constant, therefore the time varying

---

[*] Corresponding author.
E-mail address: mpmcgarry@utep.edu (M.P. McGarry).

0140-3664/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2012.01.004

frame sizes are what contributes to time varying bandwidth requirements (i.e., bit rate) from frame to frame. As a result of the Variable Bit Rate (VBR), it is difficult to allocate communication resources to accommodate video without underutilizing those resources. The simplest approach is to allocate bandwidth for the maximum rate of the video. Unfortunately, with a large difference between the mean and maximum rate of the video, network utilization will be quite low [7,8].

Dynamic Bandwidth Allocation (DBA) provides the most efficient means of allocating network resources to video traffic sources. However, without predictive abilities it is difficult for a DBA algorithm to anticipate the bandwidth requirements of a video traffic source. With an accurate forecast of the bandwidth requirements, a DBA algorithm can successfully accommodate the short term bandwidth needs of video.

The existing body of research on bandwidth forecasting investigates different forecast model structures and model coefficient computation methods. We propose a forecast method that differs significantly from the existing approaches. Specifically, we exploit the fact that for pre-recorded video all of the video frame sizes are known at the point of storage and feed this information forward into the network. So, the video source cooperates with the network elements to provide an accurate bandwidth forecast. More details of how our work differs from what can be found in the existing research literature are provided in the next section.

This article is organized as follows. In Section 2 we review the existing research literature on video bandwidth forecasting and differentiate our approach from this existing body of work. In Section 3 we describe our cooperative approach to bandwidth forecasting that we call Feed Forward Bandwidth Indication (FFBI). In Section 4 we discuss several applications of bandwidth forecasts that can improve network performance measures. In Section 5 we discuss the details of one of these applications, exhaustive queue service in Ethernet Passive Optical Networks. In Section 6 we analyze the performance of FFBI and several other bandwidth forecasting methods. Finally, in Section 7 we conclude this article with a summary of our findings and an outline for future investigation.

2. Background and related work

Many video bandwidth forecasting methods have been proposed and evaluated in the research literature. These methods rely on auto-correlation among video frame sizes. More specifically, several auto-regressive forecast models and neuro-computational methods have been proposed and evaluated. Video traffic can be characterized by two time series: one representing video frame arrivals, the other representing the size of the video frame (i.e., workload associated with the arrival) [9]. Let A_n represent the time between the nth arrival and the $(n-1)$th arrival, and W_n represent the size of video frame n. A_n is both deterministic and constant (typically, $\frac{1}{30}$ s). W_n depends on the compression and is therefore a stochastic process. A succinct review of the literature related to forecasting W_n using the proposed forecast method as a taxonomy follows.

2.1. Forecasting models

An Auto-Regressive (AR) model has been used to forecast: (1) the size of each frame type (i.e., I, P, and B) separately [10–15], (2) the size of a GoP rather than individual frames [11,16,17], and (3) the size of each frame of an aggregate of videos [15,18]. An AR model has also been used to forecast the differences between: (1) individual frame sizes in adjacent GoPs [19] and (2) the sizes of I frames in adjacent GoPs [20].

An AR model has also been used to forecast the sizes of B frames from previous sizes of P and B frames [21–23].

Some researchers have used transformations of frame sizes to potentially improve forecast accuracy. Specifically, an AR model has been used with inputs transformed using both the discrete wavelet transform [24–26] and the Fourier transform [27,28].

A Moving Average (MA) model with error feedback has been used to forecast frame sizes [13,14]. An Auto-Regressive Moving Average (ARMA) model has been used to forecast frame sizes [29]. An Auto-Regressive Integrated Moving Average (ARIMA) model has been used to forecast frame sizes [30,31]. A model that requires the bit content of the frame to be forecast, to compute the spatial variance, is used to predict the compressed frame size [32].

2.2. Neuro-computational methods

Several neuro-computational structures have been used for bandwidth forecasting. Some examples are:

- Multilayer perceptron [33–45].
- Radial basis function network [43,46].
- Time-delay neural network [27,28].
- Time-lagged feed-forward neural network [37,43,44,47].
- Recurrent neural network [39,43,45,48–50].
- Neuro-fuzzy network [38,41,51,52].

2.3. Exploiting frame size information

Our method of bandwidth forecasting differs significantly from the other methods in that we do not propose a new forecast model that relies on auto-correlation in the video frame size stochastic process. Instead, we exploit the knowledge of all the video frame sizes of pre-recorded video at the point of storage [53] and feed this information into the network as bandwidth indications. The video source *cooperates* with packet switching devices to share this valuable information to improve network performance. This method can be extended to live video by introducing a delay at the source. This method is referred to as Feed Forward Bandwidth Indication (FFBI) and is described in detail in Section 3.

Other researchers have identified that the video frame sizes for pre-recorded video are known in advance and can be used to determine an effective bandwidth for a video source [5,8,40,53–55]. However, we are the first to propose a method to feed this information to packet switching devices and to investigate methods to utilize this information to improve network performance measures.

In addition, other researchers have proposed introducing delay to live video for the purpose of improving online smoothing in conjunction with prefetch buffers at the destination [54,55]. However, we are the first to explore introducing delay for live video to obtain an accurate bandwidth forecast.

3. Feed-Forward Bandwidth Indication (FFBI)

We propose a method called Feed-Forward Bandwidth Indication (FFBI) that feeds future bandwidth requirements of a video into its video frame headers and then onto packet headers to provide this information to packet switching devices at all protocol layers.

3.1. FFBI for pre-recorded video

The time varying frame sizes of a pre-recorded video are known for the entire length of the video at the point of storage in advance of its transmission through a packet switched network. However, future frame sizes are unknown at individual network elements as the pre-recorded video is transmitted through a packet switched network. At the point of storage, we feed video frame sizes forward