Computer Communications 109 (2017) 24–37

Contents lists available at ScienceDirect

Computer Communications

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

Video frame size modeling for user-generated traffic in an enterprise-like environment

Marios Kastrinakis[a], Ghada Badawy[b], Mohammed N. Smadi[c], Polychronis Koutsakis[d,*]

[a] School of ECE, Technical University of Crete, Greece
[b] Computing Infrastructure Research Center, McMaster University, Canada
[c] ECE Department, McMaster University, Canada
[d] School of Engineering and Information Technology, Murdoch University, Australia

ARTICLE INFO

Article history:
Received 25 February 2017
Revised 12 May 2017
Accepted 15 May 2017
Available online 19 May 2017

ABSTRACT

Smart mobile devices have displaced personal computers in many daily applications such as internet browsing and email. However, for content creation, users still need to use a large display, keyboard and mouse. Many initiatives are currently working on enabling I/O functionality for content creation and peripheral access, and on preserving the grab-and-go experience where the mobile device is not tethered to the docking station but merely placed in proximity of it and the traffic is carried over Wi-Fi. Maintaining the Quality of Service (QoS) and Experience (QoE) of low-latency, high fidelity video (for example the desktop view of a smart device) when transmitted over a Wi-Fi link in heavily loaded environments has been proven problematic. In this work, we propose for the first time in the relevant literature to the best of our knowledge, a highly accurate video traffic model that is capable of predicting the volume of video traffic generated by an average user's computer during a day. Our modeling techniques are tested on real user-generated screen mirroring traffic from a large shared cube space similar to an enterprise environment, and can be easily used as source traffic generators in order to facilitate the study of H.264 transmission performance over wireless networks.

1. Introduction

Smart mobile devices such as smartphones and tablets are becoming more powerful every day with the advancement of mobile computing chips from the likes of Qualcomm, NVidia and Intel, while at the same time major software and operating system companies develop their products in a single code base suitable for many platforms [1]. In addition, a 2013 survey [2] placed smartphones' popularity at around 85%, surpassing all other kinds of computing devices. The above facts seem to point towards a future where smart mobile devices might replace computers completely, in personal and corporate environments.

On the other hand, users still want to use a large display, keyboard and mouse for content creation and when they are not on the move. This use case has been addressed using tether technology such as the Mobile High-Definition Link (MHL) [38], however to maintain the grab-and-go experience there is a trend of replacing the need for cables with a wireless link that connects the mobile device to a wireless docking station. In such a setup, the video on the mobile device display is mirrored unto a larger display with low-latency to ensure the interactivity of the end-user is maintained. The actual latency limits are application dependent. Due to its attachment rate in the mobile industry, Wi-Fi is a key candidate for carrying this docking traffic and Miracast [3], which was recently ratified as a standard to allow mobile device display mirroring over Wi-Fi, is also a candidate for enabling the video component of any docking station. At the heart of a Miracast source, an H.264 encoder streams over a Wi-Fi peer-to-peer (P2P) mode without needing access to an overlaid Wi-Fi network. However, in enterprise environments where an overlaid Wi-Fi network is not only present but heavily used, contention due to medium access between the Wi-Fi P2P docking link and the overlaid Wi-Fi network will have a direct impact on the QoE for the docking video. The impact and potential QoE degradation will depend, among many other factors, on the characteristics of the video traffic, which in turn is dependent on the video content being compressed, the encoder implementation and the target latency

* Corresponding author.
E-mail addresses: makastrinakis@electronics.tuc.gr (M. Kastrinakis), gbadawy@gmail.com (G. Badawy), smadim2@mcmaster.ca (M.N. Smadi), p.koutsakis@murdoch.edu.au (P. Koutsakis).

http://dx.doi.org/10.1016/j.comcom.2017.05.008

profile. For example, video chatting on mobile devices demands high bandwidth and if the offered bandwidth is decreased video quality may be compromised [41].

Streaming over best-effort networks that were not designed to provide stable QoS, especially wireless networks, makes it inefficient to use the same representation of a video for the duration of a streaming session. Instead, it must be adapted to dynamically varying networks conditions such as throughput, packet loss rate, and delay jitter [42]. These problems can be significantly mitigated if the volume of video traffic that will be generated in the network can be predicted.

Traffic characterization and modeling of multimedia services are required for an efficient network operation. The generated models can be used as traffic rate predictors, during the network operation phase (online traffic modeling), or as video generators for estimating the network resources, during the network design phase (offline traffic modeling). In the offline case, traffic models can be used as video generators, to select appropriate network parameters during the network design phase, such as utilization, and/or number of multiplexed sources that achieve an acceptable video quality. In this framework, the reliability of the network can be evaluated. For example, the probability of refusing a new video call or the probability of network overload can be estimated. On the other hand, online traffic models are very useful for traffic management algorithms and congestion control schemes, which prevent the network from possible overload [51]. In such schemes, the emphasis is to increase the network resources utilization while maintaining the desired level of QoS. Dynamic resource allocation (DRA) schemes are especially important for live streams, where the video stream characteristics are not known in advance. In order to provide an accurate estimation of the needed network resources for a certain flow, which indicate the cost of transmitting such a flow, the chosen DRA scheme has to be able to predict the required bandwidth for future video frames. To adjust the bandwidth assignment for a certain video stream, DRA renegotiates the assigned bandwidth for that flow. The main goals for a DRA scheme are: to predict the longest possible period with the least prediction error, and to provide the best possible resource utilization with the lowest achievable frame delay [50]. This will ensure that no degradation takes place in the QoE of accepted video sessions [43].

Hence, in this work we develop H.264 video traffic models for low-latency, candidate Miracast source implementation for content that resembles a typical desktop user in an enterprise-like environment. We discuss well-known models, which fail to accurately capture this type of video traffic and we propose a highly accurate model based on the combination of clustering with Markov chains and the use of the Jaccard index similarity coefficient. To the best of our knowledge, this is the first time that a model tested on real user-generated screen mirroring traffic is developed and presented. It is also the first time that the concept of the Jaccard index is used for video traffic modeling purposes. We make our dataset available to the community so that possible other modeling approaches can be tried by other researchers.

The paper is structured as follows. Section 2 discusses related work. Section 3 presents the video traffic encoding of the data that we worked with. Section 4 includes the data collection methodology, data statistics and the statistical tests that we have used during the development and testing of our models. In Section 5 we analyze our models and discuss the respective results. Section 6 includes our conclusions and ideas for future work.

2. Related work

Multiple efforts on video traffic modeling have been conducted in the literature. Video models which have been proposed include first order autoregressive (AR) models [5], discrete autoregressive (DAR) models [4,6,7,40], Markov renewal processes (MRP) [8], MRP transform expand sample (TES) [9,10], finite state Markov chain [11,12], a combination of wavelet and linear modeling [20] and gamma beta auto regression (GBAR) models [13,14]. In [15] the authors analyzed a number of mobile video streams and created a model that provides both video frame and RTP packet generators. The model was created and verified against “The Matrix” and “Lord of the Rings” movies. In [16], the authors create a video traffic model that takes interdependence between different frame types into consideration (I, P and B frames). The authors in [17] list a number of Variable Bit Rate (VBR) video traffic models and compare these models against three video traces “Star Wars IV”, “Tokyo Olympics” and “NBC 12 News”. They showed that some of the models work for some videos but not for others. They could not find a universal model that works with all types of videos. Other recent efforts include [45] in which the authors propose their models on top of the model in [20] and they exploit the hierarchical prediction structure inherent in H.264 for intra-GoP (Group of Pictures); [46], in which the authors propose a non-linear autoregressive model for long-range video traffic prediction (without separating traffic of different frame types) and they introduce adaptive algorithms to obtain the parameters of their model; [32], where the authors use linear regression to predict B-Frames’ sizes, with the goal of possibly dropping B-Frames (based on the prediction) in the case of network congestion. The works in [47,48] focus on modeling new types of video traffic, such as 4K and 3D video. The authors in [47] use a seasonal autoregressive model for modeling and prediction of 4K video traffic. They analyze over 17,000 video frames and show that their proposed methodology provides good accuracy in high definition video traffic modeling. The work in [48] proposes and evaluates a new model for multiview video (which is used to support 3D video applications) that is based on a Markov process.

The above-referenced papers model video traces from movies, which are significantly different, in terms of content, than those created by desktop applications used in an enterprise environment. Our goal in this study is to fill this gap by studying these applications and the video traffic they generate, in order to build a highly accurate model. There is very little work done on what applications are mostly used in enterprise environments. In [18] the author shows that employees use Microsoft Outlook the most from the Microsoft Office suite. The author does not mention which other applications are used. Also there is no characterization of the video generated by the Microsoft Office suite or similar tools.

3. Video encoding

We worked with two different datasets, encoded with the H.264 video coding standard. H.264 or MPEG-4 Part 10, AVC is a video coding standard developed by ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). It is the most widely accepted video coding standard (since MPEG-2) and it covers a wide area of video applications ranging from mobile services and videoconferencing to IPTV, DTV and HD video storage [19]. According to the H.264 standard, an encoded video trace features two distinct characteristics: (1) every video frame comes from one of three different types of frames, and (2) video frames are organized in groups with a specific structure.

There are three different types of frames, I-Frames (Intra-coded Frames), P-Frames (Predicted Frames) and B-Frames (Bi-directional predicted Frames). P-Frames are smaller than I-Frames and B-Frames are the smallest [22]. Video frames are grouped together in Group of Pictures (GOP) structures that specify the order in which intra- and inter-frames are arranged. A GOP pattern specifies the amount and order of P and B-Frames between two successive I-Frames. Every GOP contains a single I-Frame with which it starts.