Towards Content Delivery Optimization In Future Wireless Networks

Pavan Kamaraju
Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, USA
Email: pavan4@umbc.edu

I. Motivation

The last few years have seen an enormous growth in the usage of smart phones and this trend is expected to continue in the near future with increased processing power and larger form factors. This explosive growth has contributed to an unprecedented adoption of mobile data services and increase in data usage. The demand is also driving the development of next generation networks to support novel data hungry and user-experience focused applications such as virtual reality video. Global mobile data traffic is increasing at an unprecedented rate and video traffic alone currently constitutes about 50% of the total traffic and is predicted to grow up to 70% by 2021 as user preferences are shifting towards more video based applications relative to browsing. The explosion of traffic associated with video content poses significant challenges for mobile content provision. While, on the one hand, mobile video traffic surge is forecasted to require significant investments in bandwidth acquisition, infrastructure deployment and roll-out, on the other hand, users are not likely to be willing to pay significantly more than they are today. Operators would in turn like to charge the content providers for revenue. Furthermore, user expectations for high quality video content is constantly increasing.

In our work, we pose the following question: how to best deliver video content without additional investment from providers while still maintaining similar user Quality of Experience (QoE) as of today. We define QoE in terms of overall perception of service which include direct factors such as perceived quality of video content and indirect factors such as battery lifetime. We address the problem using three methods: (1) context-aware opportunistic pre-fetching of content, (2) video content optimization based on user perception and (3) video content delivery with personalized quality.

Our results show that opportunistic prepositioning of data objects which are likely to be consumed by users (pre-fetching) utilizing contextual parameters such as location, high data rates in an over-the-top fashion can reduce the energy consumption of transmission for video streaming by up to 40%. Next, by adapting the video based on user perception, where a video is broken into shorter segments which could be delivered in a satisfying resolution, we were able to show that file sizes of video can be reduced by up to 60% while achieving similar user perceived quality as the original video in fixed resolution. This technique allows us to create video content in varying perceptual qualities. Finally, we show that by grouping users based on similarities in viewing history for predicting and delivering a specific video quality (personalized) for individual users, we can reduce the overall video traffic in the network by up to 50%.

978-1-5090-2185-7/16/$31.00

The remainder of this extended abstract highlights the individual methods and their evaluation results in Section II followed by conclusion and future work of this doctoral thesis in Section III.

II. Methods

A. Context-aware opportunistic pre-fetching of content

Content consumption is currently agnostic to the context of user, network and terminal. However, this information could be used to optimize and deliver content more efficiently. For example, streaming video-on-demand content such as YouTube can be up to three times expensive when compared to pre-fetching over an HSDPA network [1], [2] and if this content could be pre-fetched on to the device, we can save energy costs and also improve the QoE for the user. We proposed and evaluated two novel schemes for pre-fetching: Over-the-Top (OTT) and Operator Co-operated Pre-fetching (OCP) using our implemented testbed *ActiveCast*. Fig. 1 highlights the working of the schemes on a broader level.

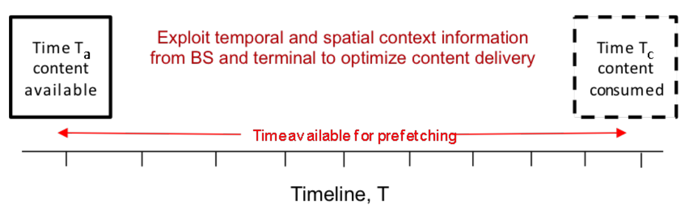

Fig. 1: Context aware opportunistic pre-fetching

Considering a data object which is available at time, T_a, and likely to be consumed at time, T_c, we opportunistically pre-fetch the object by optimizing mutually exclusive context information which includes (1) User: data cap for the user, price SLAs between content provider and operator, content type preferences etc., (2) Terminal: energy consumption, duration of pre-fetching, available data rates, signal quality information, network connectivity type, user location (and mobility pattern), battery charging status, remaining battery power, available memory, time of the day etc., and (3) Network: active users in the cell, current load etc. We assume that the content that is likely to be consumed can be identified utilizing various means such as viewing behavior from social media e.g., subscriptions on YouTube, saved lists on Netflix etc., explicitly gathering information from the user e.g., watch later on YouTube and also by opening the system to content providers for specifying content to be pre-fetched.

For the OTT scheme, we proposed downloading content based on an energy budget and sleep/wake-up cycle with download probes for channel data rate measurement. In particular, we assign an energy budget, data rate for a data object and notify the user terminals to pre-fetch. The terminals gather updated information on achievable data rates by monitoring the data rate performances within a given time window of a few seconds and use this to pause and retry content pre-fetching. We chose data rate as a pseudo indicator for energy as it is inversely proportional to energy in general. However, since the OTT case behaves more like a best effort system, the window for retrying pre-fetching of content can be optimized utilizing updated context information such as high signal strength, base station, location etc. We proposed a usage based model for estimating energy costs on-the-fly during pre-fetching which depends on the filesize of video, time to download and tails incurred. Utilizing this model, we track the energy consumed to pre-fetch and deliver the data object with a predictable energy cost. For the OCP scheme, we used the data rate traces from the real network to calculate potential gains in a simulated environment if the operator co-operates in actively notifying the terminals for pre-fetching.

We were able to show that current streaming solutions are extremely inefficient from an energetic perspective and opportunistic pre-fetching could potentially lead to decreased energy costs by a factor between 1.33 and about 5 times on 3G interface. We gathered relevant information to support the definition of novel and more flexible SLAs with content providers, providing strict guarantees on the upper bounds of the delivery delays while allowing to optimize the energy budget invested by both the users' devices and the wide area networks to complete the download of content (pre-fetch) in the user terminals.

B. Video content optimization based on user perception

There are several factors affecting the perceptual quality of a video. For example, a higher motion video tolerates lower frame quality at the expense of higher frame rate, to enable smooth motion of playback [3]. The *device display size* and *its pixel density* play a large role in determining the *minimum required frame resolution that can provide the best perceived video quality*. Increasing the resolution over this value *does not increase* the perceived video quality further, while undesirably increasing video size and bandwidth required to deliver this video to the end user.

Utilizing this notion of perceptual video quality, we proposed and evaluated a video optimization method where individual segments of the video could be reduced in resolution to generate video in different perceptual qualities. The method is shown graphically for a selected threshold in Fig. 2. [4]

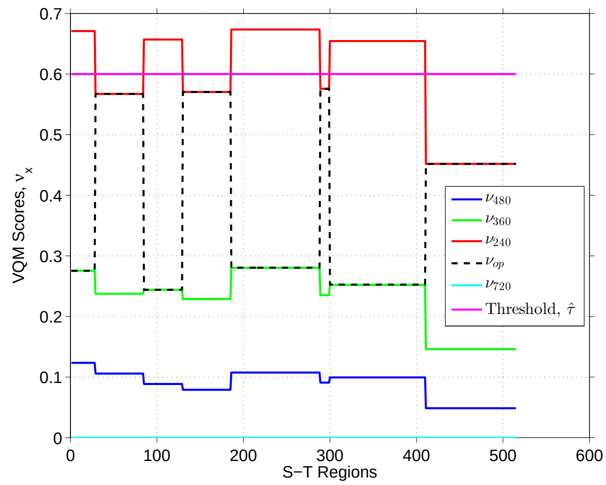

Fig. 2: Video optimization method. ν_x represents the VQM scores of video with resolution x (240,360 etc.,)

Video Quality Metric (VQM) software [6] generates a objective perceptual score between 0 and 1 for a source and reference clip, lower the score closer the perceptual gap to the source video. The software generates this score from individual perceptual parameters which are extracted from S-T (spatio-temporal) regions of the video. We chose this metric for our optimization method as it has high correlation with MOS grades (exceeds 0.9 on Pearson correlation coefficient). For the optimization method, we first generate the VQM scores [5] for individual segments of the video in different resolutions (e.g., depicted by ν_{720} for 720p in Fig. 2). Next, the optimization method identifies minimum resolution for individual segments that satisfies a chosen perceptual quality threshold, $\hat{\tau}$. Finally, the segments chosen by this threshold are used to compose the video (dotted line in Fig. 2)

Utilizing the optimization method, we were able to show that we could generate videos with varying perceptual qualities with up to 60% savings in terms of file size.

C. Video content delivery with personalized quality

Utilizing the optimization technique, movies could be composed with a perceptual quality as determined by the selected threshold ($\hat{\tau}$). Once composed, we need to be able to deliver a video with a quality threshold, $\hat{\tau}$, for a specific user. We propose clustering methods using collaborative filtering to predict the quality of video to be delivered to the user. The fundamental assumption behind collaborative filtering methods is that users opinion on quality can be filtered, selected and aggregated in such a way as to provide reasonable prediction of active users' preference. Fig. 3 shows one such approach by grouping users based on nearest neighbors with a radius threshold, ρ. By increasing the radius threshold, ρ, more users can be grouped as we can tolerate larger distances between user grades and as a result the error in prediction increases. However, by considering smaller thresholds, we group fewer users but with better prediction accuracy. This would mean some users might not be grouped. Once the users are grouped based on previous viewing history, a new video's quality threshold is predicted using the mean of the available grades for an active user(represented by the hollow 'x' in Fig. 3).

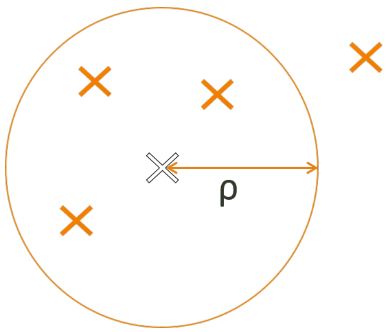

Fig. 3: Grouping based on similarities in viewing behavior to predict the QoE

We generated videos with different perceptual qualities and gathered subjective quality opinions on perceptual quality using the Amazon mturk platform. We validated our strategies on the gathered opinion scores and we were able to show that the median error in prediction was in the order of ±0.5 MOS points for delivering personalized video. We were also able to show that by delivering personalized quality, we consume significantly less bandwidth and attract relatively low dissatisfaction from users. [7]

III. Conclusion and Future Work

Our work in this dissertation presents an end-to-end scalable content delivery paradigm over mobile networks. We presented opportunistic context-aware pre-fetching schemes and compared the performance with current content delivery mechanisms in terms of energy cost. The schemes include OTT methods where data rates are measured by periodically probing the channel and an ideal OC case where data rate estimation is provided by the mobile operator. We gathered evidence that by adopting an opportunistic content pre-fetching scheme, which exploits information on content access probability to select appropriate data rate requirements, energy costs could be significantly reduced when compared with current video streaming mechanisms. We proposed and evaluated a method to optimize video content based on user perception. The proposed method reduces the size of video content by decreasing the spatial resolution based on perceptual quality determined by VQM. We performed subjective video quality testing for the optimized and fixed resolution videos to validate and compare performance of our method. We have shown that we can deliver perceptually similar high-quality video for a significantly lower cost in terms of file size. We also integrated our method to deliver content via streaming using DASH. Finally, we proposed a method to predict the QoE of content and deliver personalized quality to a user using clustering methods. We proposed and validated strategies to group users based on similarities in their viewing behavior and quality requirements. By delivering content with a personalized approach we show that we can deliver a superior QoE for lower cost with respect to total bandwidth.

Our work opened further avenues of research for future work. Specifically, opportunistic OTT pre-fetching has shown that significant energy savings can be achieved, but this approach has no knowledge of the underlying networks. *ActiveCast* testbed has the provision to integrate with an operator to test the efficiency of OCP methods with a real operator. Using *ActiveCast*, users' content preferences, viewing activity, time of access, connectivity information can be recorded. Machine learning techniques can be applied on this data to (1) increase the prediction accuracy for pre-fetching, (2) dynamically predict and modify pre-fetching policies and (3) dynamically adapt and adjust the tail time during data transfers. We currently utilize VQM software for extracting perceptual video quality information. Other quality indicators such as sharpness, blur and motion or a weighted combination of such factors could be used to increase the accuracy of the video optimization technique. We have promising results showing that by clustering users, we can greatly enhance the QoE and also reduce costs in terms of energy and bandwidth. However, large scale validation across many users with variety of content is necessary and we proposed a scalable crowd-sourced strategy to address this issue. Furthermore, we believe that clustering the content first into categories and applying the grouping strategies on the categories will further increase the prediction accuracy.

Acknowledgment

The author is grateful to Anupam Joshi, Zary Segall and Pietro Lungaro for their supervision and support.

References

[1] P. Kamaraju, P. Lungaro, and Z. Segall, "A novel paradigm for context-aware content pre-fetching in mobile networks," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, April 2013, pp. 4534–4539.

[2] P. Lungaro, C. Viedma, P. Kumar, and Z. Segall, "An experimental framework to investigate context-aware schemes for content delivery," in *Vehicular Technology Conference 2011*.

[3] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality," ITU-T SG 9, Contribution COM 9-80-E, Jun. 2000.

[4] A. Devlic, P. Kamaraju, P. Lungaro, Z. Segall, and K. Tollmar, "Qoe-aware optimization for video delivery and storage," in *2015 IEEE 16th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 2015, pp. 1–10.

[5] National Telecommunications & Information Administration (NTIA) Institute for Telecommunication Sciences (ITS). VQM software. www.its.bldrdoc.gov/resources/video-quality-research/software.aspx.

[6] M. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting*, vol. 50, pp. 312–322, Sep. 2004.

[7] P. Kamaraju, P. Lungaro, and Z. Segall, "Qoe aware video content adaptation and delivery," in *2016 IEEE 17th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 2016.