Christian Koch and David Hausheer, Optimizing Mobile Prefetching by Leveraging Usage Patterns and Social Information
In Proc. of the IEEE International Conference on Network Protocols (ICNP), ISBN: 978-1-4799-6204-4

2014 IEEE 22nd International Conference on Network Protocols

# Optimizing Mobile Prefetching by Leveraging Usage Patterns and Social Information

Christian Koch
Peer-to-Peer Systems Engineering Lab, TU Darmstadt
Email: christian.koch@ps.tu-darmstadt.de

David Hausheer
Peer-to-Peer Systems Engineering Lab, TU Darmstadt
Email: hausheer@ps.tu-darmstadt.de

***Abstract*—Real-time entertainment constitutes the majority of traffic in today's mobile networks. The data volume is expected to increase in the near future, whereas the mobile bandwidth capacity is likely to increase significantly slower. Especially peak-hour traffic often leads to overloaded mobile networks and poor user experience. This increases costs for the mobile operator, which has to adapt to the peak demand by capacity over-provisioning. The new approach proposed in this paper aims to leverage the user's context and video meta-information to unleash the potential of video prefetching. Based on observed user interactions with social networks, the videos a user consumes from social neighbours can be predicted. Moreover, the user's daily routine even enables a prediction of the time when videos are consumed as well as the network capabilities available at that point. First results show that partial prefetching based on content categories provides a potential for efficiently offloading mobile networks. Additionally, the user experience can be improved as freezing playbacks of videos can be decreased. Initial results show a high potential for category-based prefeching.**

## I. Introduction

Real-time entertainment content represents the largest single source of mobile data traffic in North America and Europe [1]. The growing number of videos watched mobile leads to a heavy load on the networks, especially during peak-hours. In North America, real-time entertainment traffic accounts for one third of the mobile traffic during peak-hours, followed by traffic caused by Online Social Networks (OSN) and content portals. The most popular platform for user generated content (UGC) is YouTube. Six billion hours of videos are watched each month on YouTube. Currently, 40% of the contents are watched mobile. Besides YouTube, which is likely to remain the single largest source of UGC traffic, further applications such as Vine are expected to contribute more to mobile traffic in the near future. It is expected that the per-month data volume will increase tenfold whereas the mobile bandwidth capacity will increase only twofold by the year 2018. As a result, mobile operators face new challenges since the frequency spectrum is limited, while the data volume growth is not. Despite the increasing adoption of LTE, the gap between supply and demand on mobile network capacities keeps increasing [2]. Mobile carriers react to this by adding caps to LTE data plans.

To this end, a new social- and context-aware approach to relieve overloaded mobile networks at peak-hours is proposed in this paper. The approach leverages prefetching mechanisms to offload videos from the mobile network, mainly over WiFi. Less peak-hour traffic leads to reduced costs for mobile operators, since capacity over-provisioning can be decreased. Furthermore, the mobile network frequency spectrum is used less intensively, which is important considering an increasing bandwidth capacity gap between supply and demand.

The proposed approach differs from existing work in this area by considering both, the network operator and the user needs, as well as using specific user-centric social information. This information is derived from the user's OSN profiles (e.g., YouTube, Facebook, or Vine), from the user's watching behaviour (e.g., partial/repeated watching, source, and topic), and the used smartphone sensors (e.g., connectivity, location, time, and movement). There is a huge number of potentially interesting videos for a user on content portals, which is hard to predict. Using OSN information has proven to be predictive for a users watching behaviour [3]. The main contributions of the proposed work include (i) a mechanism which predicts the probability of videos being watched by a user. The prediction includes the portion of a video being watched and the number of times a user will watch it. (ii) The user's location and time are used to determine the optimal prefetching sequence. (iii) The best time to prefetch is determined based on the time left until a user is likely to watch a video and based on a location-connectivity model. Initial results on the analysis of users' preference patterns and partial video views show a high potential for category-based prefeching.

The remainder of this paper is structured as follows: Section II discusses existing work. Section III describes the approach on mobile video offloading. Section IV shows and discusses first results. Section V summarizes this paper and gives an outlook on possible further optimization.

## II. Related Work

Content selection is the most important choice for prefetching systems. Zhao et al. [3] use machine learning to predict relevant videos from the user's Facebook feed. They use post interactions, the number of private messages exchanged, the number of viewed videos from friends or pages, and the post popularity to determine promising videos. To derive the user's engagement, their own Facebook app needs to be used, which introduces a bias, since the post ordering and the look-and-feel of Facebook cannot be imitated. The approach proposed works with the native Facebook view, while using an own video player and its focus is on the network operators.

A key aspect in prefetching is how much shall be prefetched. An approach using a fixed chunk length of 1200 bytes is presented in [4]. In [5] the authors compare caching, prefetching, and a mixed approach implemented at a proxy or a client device. They are using prefix-prefetching. Based on the available bandwidth, the video duration, and the video bitrate, the prefix size is determined. This approach increases the user experience, but does not respect the network operator needs.

The authors of NetTube [6] present a prefix-based prefetching mechanism. They aim to increase the user experience, which has shown to be sensitive to stalling events at the video playback start [7]. The approach in this paper determines the segment's length with respect to the probability that it is watched. Therefore, the user perceives less stalling events.

Energy saving is important for the user. The measurements conducted by [8] show that transmissions over WiFi are about ten times more energy efficient compared to 3G. Thus, it is beneficial to download content when WiFi is available.

In [9], the behaviour of mobile streaming apps with iOS and Android in combination with the streaming players of YouTube, Netflix, and Hulu is compared. An interesting observation of the authors is that up to 15% of traffic in video streaming sessions is caused by redundant traffic. This is caused by video quality adaptation due to varying network conditions. The proposed approach circumvents this by offloading on stable network conditions in a fixed quality.

Overall, the approach presented here differs, as it makes predictions based on videos' meta-data and the user's engagement towards these aspects. Therefore, watching probabilities based on social information are leveraged in the approach.

## III. Approach

Offloading videos from the mobile network requires accurate watching prediction, the estimation of how much is watched, and a download scheduler. This problem is divided into multiple modules depicted in Fig. 1. The implementation is currently ongoing work. As inputs, the *Network Offloader* uses information from the OSNs, the observed user interactions, and the network conditions. The proposed architecture consists of three monitoring modules, targeting the OSNs, the user interactions, or the environment. These modules provide input for the *Prefetching Predictor*. This module determines videos which are most relevant for a user. The module's submodules formulate, update, and leverage a model of the user involvement, the probability for a video being watched to a certain percentage, and the connectivity. The predicted videos are passed to the *Prefetching Scheduler*, which schedules the download based on time, location, and consumption patterns.

**OSN Monitor.** Content published on the user's OSNs has turned out to be much more predictive than focusing on popular content. This module provides a view on all contents that are presented to the user by OSNs. In OSNs, the publishing behaviour is assumed to be non-deterministic. Due to this, the module has to request for video updates regularly. Video-related data is stored in a local database.

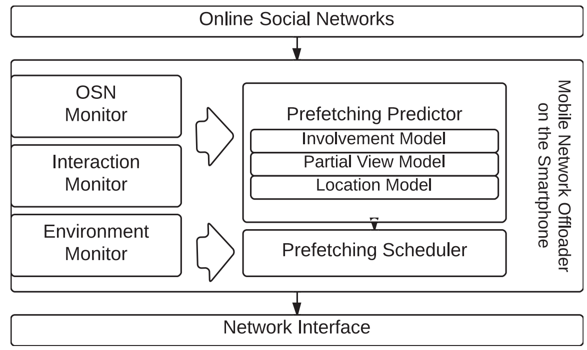


Fig. 1. Architecture overview of the Mobile Network Offloader

**Interaction Monitor** The *Interaction Monitor* senses interactions between the content and the user. Subscriptions, assignments, and friendships of a user are used as a baseline. Information like when, where, and for how long a user watched a video is stored in a local database. Compared to other works, this approach does not force a replacement of the user's Facebook app. Instead, a video player which is able to collect detailed information about the video playbacks is used.

**Environment Monitor.** This module aims for the identification of user habits. Therefore environmental data, e.g., location, time, and bandwidth are captured. The entries of the *OSN Monitor's* DB are enriched with this information.

**Prefetching Predictor.** Based on the monitoring module's information, this module models the user's key behavioural aspects. Not all aspects are equally predictive for different users. This module weights them in a user-specific manner. The *Prefetching Predictor's* sub-modules build the core of this work and are described in the following.

**Involvement Model.** This module determines which video features are most predictive. Users are assumed to be interested in certain topics or in specific sources. A source is defined as the combination of OSN and publisher. Topics are identified by common words in the video name or description. Topics are especially interesting for predicting videos from unknown sources. This part of watched videos accounts for up to 44% of watched videos [10]. Sources and topics cover interest-driven OSNs like Quora and source-driven OSNs like YouTube.

**Partial View Model.** Not all videos are watched completely [11]. This module uses the portion watched as an engagement measure. Depending on the user's context and the content's meta-data, a prediction model is developed. This model is further used to decide how much shall be prefetched. As shown by [12], videos are watched repeatedly more often than others, depending on their categories. The proposed approach uses a cache management which leverages the user's re-watch behaviour by keeping them longer or shorter.

**Location Model.** Most people follow a daily location pattern [13]. This is assumed to apply for video consumption in many cases. The *Location Model* strives to model the user's access patterns based on his location. At work, the user might be interested in significantly different topics, than in his leisure time. This information is important since the time between content recognition by the *OSN Monitor* and the user watching it can be quite short. Therefore, an optimal prefetching sequence should consider this information, especially when only a few videos can be prefetched timely.

**Prefetching Scheduler.** The time between the content publishing and its consumption allows to plan the download for a certain time window. An energy-efficient scheduling uses WiFi whenever possible. Sometimes this is not possible in the given time window. In this case, the module uses a model which includes the daily movement pattern and the observed connectivity for an optimal scheduling. The network operator and the user have to agree on this scheduling plan. Therefore, a request, including video ID, time when the prefetching must be completed, and the videos size, is sent to the operator automatically. The reply message includes the ID and when to start downloading.

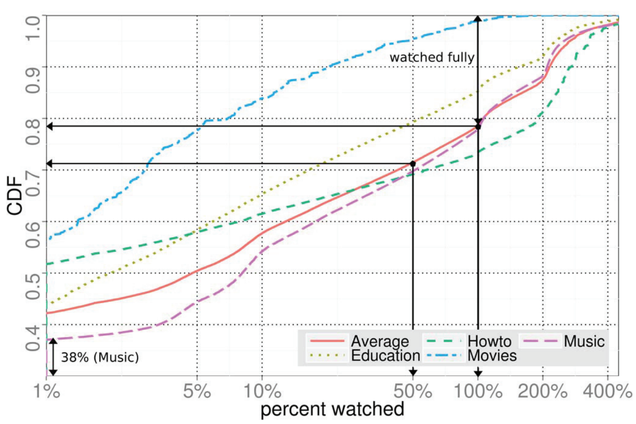


Fig. 2. CDF of percentage requested, depending on the YouTube category

## IV. Initial Results and Discussion

A study is currently conducted on a large dataset of mobile YouTube video views (10M requests to 1.6M videos by 700k users over two weeks). The focus is on the portion of a video watched based on the video's category. Fig. 2 shows that only a minority of videos are watched fully. The overall average is shown as a solid line. The other four lines correspond to specific video categories. This information was retrieved from the YouTube Data API. The scale of the x-axis in Fig. 2 goes beyond 100% because a user may watch videos multiple times, jumps back in the video, or due to packet retransmissions. For about 70% of the videos only 50% or less of it were watched. This heavily varies depending on the video category. The number of videos watched less than 1% indicates which categories are browsed more random. E.g., *Music* is watched less than 1% in 38% of the observed views, while *Howtos* account for over 50%. In only about 20% of all samples, *Music* is watched fully or repeatedly. After the 100% mark, a sharp incline is observed for some categories. This indicates that if more than 100% of the video is watched, it includes only a small part in most cases. A similar effect is visible at 200%.

This initial result clearly shows the potential for prefetching by considering videos as a group of segments with different consumption probabilities. The video's category is suited for building probabilistic models. These models can be used to optimize segment-based prefetching mechanisms. Categories which are randomly browsed are likely to be watched only a few seconds. To this end, it is reasonable to prefetch more prefix-segments of related videos than for other categories. E.g., *Music* seems to be more specifically targeted by the users, so that greater numbers of prefixes or even the entire video should be prefetched. It is planned to evaluate the proposed approach with a segment-based hit rate and precision metric.

## V. Summary and Outlook

This paper proposed an architecture for prefetching videos on mobile devices. The related work has been investigated and a new approach has been proposed which leverages the user's context and social information. The approach enables to relieve the mobile network from peak-hour traffic. Initial results on the analysis of users' preference patterns and partial video views show a high potential for category-based prefeching.

In future work, the sketched architecture will be implemented and verified using real user traces. Possible extensions may consider internal and external smartphone sensors or new mobile network technologies.

## Acknowledgments

This work has been supported in parts by the EU (FP7/#317846, SmartenIT and FP7/#318398, eCOUSIN). The authors acknowledge valuable comments by their colleagues and project partners, especially Ali Gouta and Julius Rückert.

## References


[1] Sandvine Incorporated ULC, "Global Internet Phenomena 1H 2014 Report," 2014.

[2] A. Solheim, "Microwave Backhaul Radios Meet The Evolving Traffic Challenge," Mobile Dev & Design, February 2013.

[3] Y. Zhao *et al.*, "O2SM: Enabling Efficient Offline Access to Online Social Media and Social Networks," in *Middleware, LNCS*. Springer, 2013, Vol. 8275, pp. 445–465.

[4] Z. Wang *et al.*, "Prefetching Strategy in Peer-assisted Social Video Streaming," in *ACM MM*, 2011.

[5] S. Khemmarat *et al.*, "Watching User Generated Videos with Prefetching," *Signal Processing: Image Communication*, Vol. 27, No. 4, pp. 343 – 359, 2012.

[6] X. Cheng and J. Liu, "NetTube: Exploring Social Networks for Peer-to-Peer Short Video Sharing," in *IEEE, INFOCOM*, April 2009.

[7] S. S. Krishnan and R. K. Sitaraman, "Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-experimental Designs," in *ACM IMC*, 2012.

[8] C. Gross *et al.*, "EnerSim: An Energy Consumption Model for Large-scale Overlay Simulators," in *IEEE, LCN*, 2013.

[9] A. Mansy *et al.*, "Characterizing Client Behavior of Commercial Mobile Video Streaming Services," in *ACM, MoViD*, 2014.

[10] GfK Media & Entertainment, "Original Digital Video Consumer Study," April 2014.

[11] P. Gill *et al.*, "Characterizing User Sessions on YouTube," in *ACM/SPIE MMCN*, 2008.

[12] X. Cheng *et al.*, "Insight Data of YouTube from a Partner's View," in *ACM NOSSDAV*, 2013.

[13] T. M. T. Do and D. Gatica-Perez, "Where and What: Using Smartphones to Predict Next Locations and Applications in Daily Life," *Pervasive and Mobile Computing*, Vol. 12, pp. 79 – 91, June 2014.