Characterizing Web-based Video Sharing Workloads

Siddharth Mitra¶ Mayank Agrawal¶ Amit Yadav¶
Niklas Carlsson‡ Derek Eager§ Anirban Mahanti†
¶Indian Institute of Technology Delhi, India
‡University of Calgary, Canada
§University of Saskatchewan, Canada
†NICTA, Australia
{sidmitra.del, mayankiitd, amitkyiitd, anirban.mahanti} @gmail.com
ncarlsso@cpsc.ucalgary.ca, eager@cs.usask.ca

Categories and Subject Descriptors: C.2.0 [Computer-Communications Networks]: General
General Terms: Measurement, Human Factors
Keywords: Workload Characterization, Video Sharing, UGC

1. INTRODUCTION

A video sharing service allows "user generated" video clips to be uploaded, and users of the service can view, rate, and comment on uploaded videos. Prior work has focused mostly on the YouTube video sharing service [1,2]. While YouTube is arguably the most popular video sharing service, studying the workload characteristics of other video sharing services, and identifying invariant properties as well as significant differences, is an important step towards building a broader understanding of this type of service.

With the aforementioned objective, we collected traces from four video sharing services: Dailymotion, Yahoo! video, Veoh, and Metacafe. Dailymotion is France's leading video sharing service and caters mostly to French-speaking demographics, while Yahoo! video, Veoh, and Metacafe are US-based services. While all four host user generated video clips, Veoh, in addition, also serves content from major studios and independent production houses, and utilizes peer-to-peer technology to distribute longer videos. Metacafe is distinctive among these services in its use of a revenue sharing model in which content creators are paid for videos that exceed a certain threshold of views. These services cover a spectrum of possibilities in the realm of video sharing.

2. SUMMARY OF CONTRIBUTIONS

Our key contributions are summarized below:

- We present and analyze workload data from *four* video sharing services. In aggregate, our traces contain metadata on 1.8 million videos which together acquired more than 6 billion views.
- We identify seven key invariants of these workloads, concerning aspects such as the video popularity distribution, use of social and interactive features, and the uploading of new content.
- We also find some significant differences across these services. For example, while the number of video uploads by users follows the Pareto principle, the fraction of multi-time uploaders is almost two times larger with Veoh (65%) than with Yahoo (33%).
- We show that video popularity can be measured in different ways and argue that one commonly used metric, specifically the number of views a video has received since it was uploaded, may not be appropriate when studying issues such as potential for caching. We define alternative metrics for quantifying video popularity that may be appropriate for such purposes.

For a complete description of our measurement methodology and a discussion of the above results we refer to our full paper [3]. Here we briefly describe invariants pertaining to video popularity.

3. VIDEO POPULARITY

We distinguish between two different measures of popularity that have differing applications: the total number of views to videos since they were uploaded, referred to here as the *total views popularity*, and the rate with which videos accumulate new views, referred to here as the *viewing rate popularity*.

The total views popularity distribution is useful for understanding service features such as "all time" most popular listings, but does not provide an accurate picture of the distribution of the rates at which videos are viewed. The latter is very important when attempting to model the video reference process, and in understanding the potential of different content distribution and caching architectures. For example, with the total views popularity metric, an older video with many views in the past may appear to be more popular than a recently uploaded video (and, erroneously, a better caching candidate) simply because the newer video has not been available for enough time to acquire more views.

We note that the viewing rate is highly non-stationary. To measure the *viewing rate popularity*, i.e., the rate with which videos accumulate views, we measure the *average* rate over some particular time period. One approach to obtaining such a measure for a site is to crawl the site multiple times. With two crawls, the (average) viewing rate popularity of a video can be obtained as the *increase* in the number of total views between the two crawls, divided by the time between the measurements. In the absence of at least two crawls, another measure of (average) viewing rate popularity can be obtained using the *average viewing rate since upload*, which we define as the number of views received since a video was

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.