Computer Communications 116 (2018) 118–131

Contents lists available at ScienceDirect

Computer Communications

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

A user-oriented performance comparison of video hosting services

Alessio Botta*[a,b], Aniello Avallone[a], Mauro Garofalo[a], Giorgio Ventre[a,b]

[a] University of Napoli Federico II, Napoli, Italy
[b] NM2 srl, Napoli, Italy

ARTICLE INFO

Keywords:
Network neutrality
Internet video streaming
Network performance

ABSTRACT

Internet streaming is responsible for a significant fraction of Internet traffic. It has been reported to account up to 70% of peak traffic in North American fixed access networks, and this figure is expected to reach 80% by 2020 [1]. Regarding such a killer service of the Internet, much has been discussed regarding if and how video hosting providers violate or may violate neutrality principles, in order to give users a "better" service compared to their competitors or other services. In this paper, we provide a contribution to this discussion studying three video hosting providers (i.e. YouTube, Vimeo, and Dailymotion). Specifically, we analyze their delivery infrastructures, including where the servers that provide videos are located, and the performance from a user viewpoint. To assess the performance, we measure throughput and RTT as experienced by users watching real videos of different popularity from several locations around the world and at different day hours.

We uncover the performance differences of these providers as a function of the different variables under control and move a step forward to understand what causes such differences. We also study the changes in the infrastructures and related performance over time, performing different measurement campaigns over different years.

Our results allow to understand what are the real performance users currently get from these providers and if the performance differences observed can be due or considered as a violation of network neutrality principles.

1. Introduction

There is a long ongoing debate on the so-called "network neutrality". Several definitions exist for this term, and they all share the common idea that data on the Internet should be treated in the same way despite several its characteristics such as technology, device, application, service, user, provider, and the country they come from or go to. An early debate about network neutrality regarding Internet traffic management policies appeared in 2003 [2], but concerns about possible threats to the end-to-end nature of the Internet raised already in the late 1990s [3]. Nowadays the debate has gained momentum also because of recent events such as the one involving the provider Comcast, which was slowing down uploads from peer-to-peer file sharing applications [4]. The discussion on whether the Internet should be fully neutral, or rather providers should be allowed to use techniques to differentiate traffic does not concern only economic aspects, but also and increasingly both legal and regulatory ones. For example, a research work regarding legal aspects was presented in 2011 [5]. In our paper, we do not want to advocate a position pro or against network neutrality. We rather aim at providing a contribution to understanding the current situation from a user viewpoint.

Our work focuses on three Video Hosting Services, YouTube, Vimeo, and Dailymotion, for which we measured the performance achievable by end users depending on video popularity and user location (i.e. the country) and inferred the characteristics of the infrastructure used for video delivery. The choice of these providers was driven by two properties: the global diffusion of the service, in order to identify every possible country-specific violations of network neutrality (e.g. due to censorship issues), and the service model of the providers, not asking subscription fees to the users. We analyze traffic related to video streaming because this service accounts a very high share of Internet traffic (about 70% according to CISCO [6]). The highlights of our work could be summarized as follow:

- We introduce a provider-independent methodology that allows to capture, analyze, and compare the performance statistics of the video hosting services.
- We measure, analyze, and compare these statistics for YouTube, Vimeo, and Dailymotion, from several locations around the world.
- We compare the performance obtained by the different providers.
- We provide insights on the topology of the infrastructures and routing policies used by the video hosting services to deliver their

* Corresponding author.
E-mail addresses: a.botta@unina.it (A. Botta), anie.avallone@studenti.unina.it (A. Avallone), mauro.garofalo@unina.it (M. Garofalo), giorgio@unina.it (G. Ventre).

https://doi.org/10.1016/j.comcom.2017.11.016
Received 20 February 2017; Received in revised form 6 October 2017; Accepted 26 November 2017
Available online 02 December 2017
0140-3664/

content.

- We compare the results obtained over different years to investigate recent changes to the performance and the infrastructure.

The remainder of the paper is structured as follows. Related work and highlights of the novel aspects of this work are presented in the next section. Available information regarding the infrastructures of video hosting providers is given in Section 3. In Section 4, we present the way we collected the dataset and our analysis methodology. Section 5 describes in detail the results of two measurement campaigns and with a longitudinal comparison. This section is followed by a discussion about geographical location (Section 6). Finally, we conclude the paper in Section 7.

2. Related work

There are several interesting works related to OTT streaming video services. A comparison of the content delivery frameworks of YouTube, Dailymotion, and Metacafe was provided by Saxena et al. [7]. The performance measures were collected using PlanetLab [8] nodes deployed in 9 different countries to have a global perspective. This work covered three aspects: the measurement of QoS, the investigation of the service delay variation, and the analysis of QoE. These evaluations considered the different geographical locations of end-users and how video content storage and distribution were impacted by the meta information associated with videos, such as popularity (i.e. the number of views) and video ages. Finally, the authors inferred the content delivery frameworks of the providers, showing that all providers relied on one or more CDNs (Content Delivery Network) to deliver their contents. With respect to this work, our measurements comprise Dailymotion, for which several sources of information revealed a “centralized” infrastructure with all its servers deployed in France.

Often in literature, the analysis of video hosting services is focused only on YouTube, since it generates a significant share of Internet traffic. An extensive data-driven analysis of YouTube concerning users behavior, video popularity, and their evolution was presented in 2007 [9]. Firstly, the authors compared YouTube and Daum, two video providers of User Generated Content (UGC), with non-UGC video providers (such as Netflix and Yahoo! Movies). Secondly, the authors made an extensive analysis of meta-information of videos, to investigate user behavior and video popularity distribution patterns. The data collection was related to several years and involved video information both fixed (such as category and length), and time-varying (such as the number of views and ratings). A tool to measure QoS and QoE of YouTube was designed in 2012 by Plissonneau et al. [10]. Metrics collected by a hundred of volunteers were analyzed by the authors to infer the video delivery policies of YouTube and understand how these metrics were impacted by Internet Service Providers (ISPs). Finally, a comparison of YouTube policies in the US and Europe was presented. One of the earliest analysis of HTTP video streaming with a comparison between YouTube and Dailymotion was presented in 2012 [11]. The authors performed passive measurements from a residential ISP network (i.e. Orange) to infer video characteristics (such as duration and encoding rate) and TCP-level performance (such as RTT and packet loss rate). A recent work [12] studied Netflix and Hulu, two leading providers of subscription-based video streaming services in the US. Both providers use the same three CDNs (i.e. Akamai, Limelight, and Level3) for video content delivery. The authors performed passive and active measurements using both residential users and PlanetLab nodes as vantage points. The aim was to uncover the provider architectures and their different CDN selection strategies. Results show that neither Netflix nor Hulu used the network conditions to choose the CDN. Consequently, the authors proposed an alternative CDN selection strategy to improve QoE to end-users. Unlike our work, all the vantage points were located in the US and both services needed a subscription fee to be accessed. Furthermore, rather than studying the QoE perceived by end-users and how to improve it, we aim to understand the reasons behind the performance differences experienced by end-users when accessing videos of different providers from different geographical locations around the world.

To the best of our knowledge, only our previous work [13] evaluated the performance of Vimeo. At the time of that measurement campaign (from 2014 to early 2015), Vimeo used only Akamai CDN as the content delivery platform. In the first quarter of 2016, during our second measurement campaign, describe in this paper, Vimeo has switched to Fastly CDN. For this reason, we also relate our work with ones regarding the evaluation of these two CDNs. For a comprehensive description on CDN in general and more specifically on the Akamai infrastructure the reader is referred to [14] and [15]. The former work gave the first insights into the overall infrastructure of Akamai. The authors presented an overview of the Akamai network, describing the mechanisms used for redirect user requests and the approaches to optimize the content delivery to end-users. Moreover, this paper provided an “agenda” of technical issues encountered in the development of the Akamai CDN. Several aspects of Akamai infrastructure with an overview of all components composing its platform and their capabilities were reported by the latter. Moreover, a comprehensive description of how Akamai redirects client requests to the “nearest” available server (load balancing system) and how its servers deliver content including video streaming (delivery policy) was provided. An extensive measurement of the Akamai network was performed in 2009 [16]. The aim of this work was to infer information about network condition, measuring network paths and refresh frequency of Akamai DNS server. The probing phase relied on 140 PlanetLab vantage points. Measurements were performed sending DNS requests to Akamai customers and then gathering the IP address of the Akamai edge servers. The analysis showed that redirection depends on the latency between clients and edge servers. Using DipZoom, a peer-to-peer Internet measurement platform, authors of [17] exposed the distribution of Akamai edge servers and performed active measurements to estimate the performance of Akamai infrastructure. Furthermore, a performance comparison between the Akamai CDN configuration and a possible consolidated configuration was presented, where a high amount of servers were clustered.

More focused on the geographic location of the infrastructures, Padmanabhan and Subramanian [18] built a service that pairs the IP address of an Internet host with its geographical location. The authors proposed three techniques to infer the target host position. *GeoTrack*, based on information provided by DNS server about the target and its “neighbors”. *GeoPing*, exploiting the correlation between RTT and geographic distance between target and vantage points with a well-known location. Finally, *GeoCluster*, grouping IP addresses into clusters assuming that all hosts in a cluster are geographically near, and combining partial host-to-location mapping information and BGP prefix to infer the host location. A recent work gathered all servers of Google infrastructure in serving sites then localize them using a technique called Client-Centric Geolocation (CCG). The CCG is based on the hypothesis that clients that are directed to the server are likely to be topologically, and probably geographically, close to the server [19].

Summarizing, studies more relevant to our work investigated either the CDN infrastructure and performance measures or the geographic location of such infrastructure. Our work moves a step forward with respect to existing literature. To the best of our knowledge, our work is the first study that presents a comparative analysis of the performance of YouTube, Vimeo, and Dailymotion. We also present the results of a second measurement campaign made one year after the first one, to evaluate how the delivery infrastructure of each provider evolved over time.

Unlike works related to residential ISPs networks, which involve a large number of volunteers or needs measurements from the network of the ISPs, we perform active measurements using a globally distributed research infrastructure (i.e.PlanetLab). Performance indicators