Inferring Streaming Video Quality from Encrypted Traffic: Practical Models and Deployment Experience

Francesco Bronzino[1,2], Paul Schmitt[3], Sara Ayoubi[2],
Guilherme Martins[4], Renata Teixeira[2], and Nick Feamster[4]
[1]Nokia Bell Labs, France, [2]Inria Paris, [3]Princeton University, [4]University of Chicago
[1]francesco.bronzino@nokia-bell-labs.com, [2]{firstname.lastname}@inria.fr
[3]pschmitt@cs.princeton.edu, [4]{gmartins, feamster}@uchicago.edu

Abstract

Inferring the quality of streaming video applications is important for Internet service providers, but the fact that most video streams are encrypted makes it difficult to do so. We develop models that infer quality metrics (*i.e.*, startup delay and resolution) for encrypted streaming video services. Our paper builds on previous work, but extends it in several ways. First, the models work in deployment settings where the video sessions and segments must be identified from a mix of traffic and the time precision of the collected traffic statistics is more coarse (*e.g.*, due to aggregation). Second, we develop a single composite model that works for a range of different services (*i.e.*, Netflix, YouTube, Amazon, and Twitch), as opposed to just a single service. Third, unlike many previous models, our models perform predictions at finer granularity (*e.g.*, the precise startup delay instead of just detecting short versus long delays) allowing to draw better conclusions on the ongoing streaming quality. Fourth, we demonstrate the models are practical through a 16-month deployment in 66 homes and provide new insights about the relationships between Internet "speed" and the quality of the corresponding video streams, for a variety of services; we find that higher speeds provide only minimal improvements to startup delay and resolution.

CCS Concepts

• **Information systems** → **Multimedia streaming**; • **Networks** → **Network measurement**; *Network management*; • **Computing methodologies** → *Classification and regression trees.*

Keywords

DASH; Quality Inference; Encrypted Traffic; Network Measurements

ACM Reference Format:
Francesco Bronzino[1,2], Paul Schmitt[3], Sara Ayoubi[2], and Guilherme Martins[4], Renata Teixeira[2], and Nick Feamster[4] . 2020. Inferring Streaming Video Quality from Encrypted Traffic: Practical Models and Deployment Experience. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '20 Abstracts), June 8–12, 2020, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3393691.3394194

ACM ISBN 978-1-4503-7985-4/20/06.
https://doi.org/10.1145/3393691.3394194

1 Introduction

Video streaming traffic is by far the dominant application traffic on today's Internet, with some projections forecasting that video streaming will comprise 82% of all Internet traffic in just two years [3]. Optimizing video delivery depends on the ability to determine the quality of the video stream that a user receives. In contrast to video content providers, who have direct access to video quality from client software, Internet Service Providers (ISPs) must infer video quality from traffic as it passes through the network. ISPs need to measure video streaming quality because it represents a more direct metric of customer experience than performance metrics typically extracted from network flows such as data rates. ISPs monitor video quality metrics to detect network issues that affect customer experience and mitigate problems before they generate user complaints. On longer timescales, trends in video quality can facilitate capacity planning. For example, when all clients in a neighborhood experience poor video quality, the ISP can plan to upgrade capacity for those users. Yet, monitoring video quality is not straightforward for ISPs: With end-to-end encryption becoming more common, as a result of increased video streaming content over HTTPS and QUIC [10, 11], ISPs cannot directly observe video quality metrics such as startup delay and video resolution from the video streaming protocol [2, 5]. The end-to-end encryption of the video streams thus presents ISPs with the challenge of inferring video quality metrics solely from properties of the network traffic that are directly observable.

Previous approaches infer the quality of a specific video service, typically using modeling and prediction that is based on an offline trace in a controlled laboratory setting [4, 7, 8]. Unfortunately, these models are often not directly applicable in practice because real-world networks (1) have other traffic besides the video streams themselves, creating the need to identify video services and sessions; (2) have multiple video services, as opposed to just a single one. Transferring the existing models to practice turns out to introduce new challenges due to these factors.

First, inference models must take into account the fact that real network traffic traces contain a mix of traffic. Often gathered at coarse temporal granularities due to monitoring constraints in production networks, video session traffic is mixed with non-video cross-traffic. In a real deployment, the models must then identify the video sessions accurately, especially given that errors in identifying applications can propagate to the quality of the prediction models. Second, the prediction models should apply to a range of services, which existing models tend not to do. A model that can predict quality across multiple services is hard to devise because both video streaming algorithms and content characteristics can vary significantly across video services (*e.g.*, buffer-based [6] versus throughput-based [12] rate adaption algorithm, fixed-size [6] versus variable-size video segments [9]).

2 Contributions

This work takes a step towards making video inference models practical, tackling the challenges that arise when the models must operate on real network traffic traces and across a broad range of services.

Robust quality inference models. We build on previous work that infers video quality for specific services or in controlled settings, extending the state of the art in several ways. First, we design models that infer startup delay and resolution delay more precisely, attempting to infer the specific values of these metrics instead of coarse-grained indicators. Our results show that prediction models that use a combination of network- and application-layer features outperforms models that rely on network- and transport-layer features, for both startup delay and resolution. Models of startup delay achieve less than one second error for most video sessions; the average precision of resolution models is above 0.93.

Second, we develop composite models that can predict quality for a range of services: YouTube, Netflix, Amazon, and Twitch. We find that models that are trained across all four services (*composite* models) perform almost as well as the service-specific models that were designed in previous work, provided that the training data contains traffic from each of the services for which quality is being predicted. On the other hand, we fall short of developing a truly general model that can predict video quality for services that are not in the training set. An important challenge for future work will be to devise such a model. Towards the goal of developing such a model, we release our training data to the community [1], which contains over 13,000 video sessions labeled with ground truth video quality, as a benchmark for video quality inference, so that others can compare against our work and build on it.

Third, we design models that are robust in a deployment setting where challenges arise in accurately detecting the start and end of a video session in the presence of unrelated cross-traffic and due to the granularity of training data versus what is practical to collect in an operational system. We apply techniques such as domain adaptation to make the trained models more robust to the noise that appears in the deployment. Our results show improvements in startup and resolution inference suggesting that domain adaptation is a promising approach for bridging the gap between lab-trained models and real network deployments.

Year-long deployment. As a proof of concept that a general model that applies across a range of services in a real deployment can be designed and implemented, we studied the four major streaming services across a 16-month period, in 66 home networks in the United States and France, comprising a total of 216,173 video sessions. To our knowledge, this deployment study is the largest public study of its kind for video quality inference. Beyond the practical models themselves, the deployment study that we performed as part of this work has important broader implications for both ISPs and consumers at large. We conducted this study in collaboration with *The Wall Street Journal* to understand the effect of home internet speeds on video streaming quality [13].

We find that the speeds that consumers purchase from their ISPs have considerably diminishing returns with respect to video quality. Specifically, Internet speeds higher than about 100 Mbps of downstream throughput offer negligible improvements to video quality metrics such as startup delay and resolution. This new result raises important questions for operators and consumers. Operators may focus on other aspects of their networks to optimize video delivery; at the same time, consumers can be more informed about what downstream throughput they actually need from their ISPs to achieve acceptable application quality.

Acknowledgments

We thank our shepherd Zubair Shafiq and the anonymous reviewers for their helpful comments. We also thank Israel Marquez Salinas, Srikanth Sundaresan, and Sarah Wassermann for their help with early stages of this project and Mark Crovella for the valuable discussions on domain adaptation. This work was supported by the ANR Project N^o ANR-15-CE25-0013-01 (BottleNet), NSF Award CNS-1704077, the Google Faculty Research Award program, the Comcast Tech Research Grant program, and Inria through the IPL BetterNet and the associate team HOMENET. Special thanks to our collaborators at *The Wall Street Journal*, especially Shalini Ramachandran, Tom Gryta, Kara Dapena, and Patrick Thomas, for their efforts in managing the deployment and recruiting participants.

References

[1] 2019. Labeled video sessions dataset. https://nm-public-data.s3.us-east-2.amazonaws.com/dataset/all_traffic_time_10.pkl.

[2] GSM Association. 2015. Network Management of Encrypted Traffic: Version 1.0. https://www.gsma.com/newsroom/wp-content/uploads/WWG-04-v1-0.pdf.

[3] Cisco 2017. Cisco Visual Networking Index: Forecast and Methodology, 2016-2021. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html.

[4] Giorgos Dimopoulos, Ilias Leontiadis, Pere Barlet-Ros, and Konstantina Papagiannaki. 2016. Measuring video QoE from encrypted traffic. In *Proceedings of the 2016 Internet Measurement Conference*. ACM, 513–526.

[5] Keith Dyer. 2015. How encryption threatens mobile operators, and what they can do about it. http://the-mobile-network.com/2015/01/how-encryption-threatens-mobile-operators-and-what-they-can-do-about-it/.

[6] T. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson. 2014. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In *ACM SIGCOMM*. Chicago, IL.

[7] Vengatanathan Krishnamoorthi, Niklas Carlsson, Emir Halepovic, and Eric Petajan. 2017. BUFFEST: Predicting Buffer Conditions and Real-time Requirements of HTTP(S) Adaptive Streaming Clients. In *MMSys'17*. Taipei, Taiwan.

[8] M. Hammad Mazhar and Zubair Shafiq. 2018. Real-time Video Quality of Experience Monitoring for HTTPS and QUIC. In *INFOCOM*. Honolulu, HI.

[9] Abhijit Mondal, Satadal Sengupta, Bachu Rikith Reddy, MJV Koundinya, Chander Govindarajan, Pradipta De, Niloy Ganguly, and Sandip Chakraborty. 2017. Candid with YouTube: Adaptive Streaming Behavior and Implications on Data Consumption. In *NOSSDAV'17*. Taipei, Taiwan.

[10] Openwave Mobility. 2018. Mobile Video Index. https://landing.owmobility.com/mobile-video-index/.

[11] Sandvine. 2015. Global Internet Phenomena Spotlight: Encrypted Internet Traffic. https://www.sandvine.com/hubfs/downloads/archive/global-internet-phenomena-spotlight-encrypted-internet-traffic.pdf.

[12] T. Stockhammer. 2011. Dynamic adaptive streaming over HTTP: standards and design principles. In *ACM Conference on Multimedia Systems (MMSys '11)*. San Jose, CA.

[13] The Wall Street Journal 2019. The Truth About Faster Internet: It's Not Worth It. https://www.wsj.com/graphics/faster-internet-not-worth-it/.