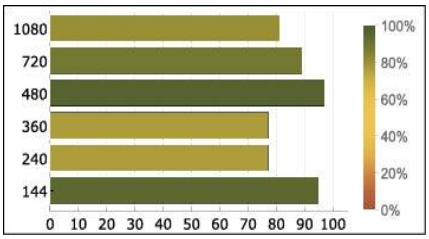
Decrypting Video Quality from Encrypted Streaming Traffic

Sarah Wassermann, Pedro Casas

With the advent of HTTP Adaptive Streaming (HAS) technology, the visual quality of the videos is no longer immaculate as recorded by the content creators. Instead, the adaptation of the visual-quality level of the streamed video can introduce a degradation to the Quality of Experience (QoE). Video quality has therefore become a paramount Key Performance Indicator (KPI) for Internet Service Providers (ISPs), which want to deliver a high video streaming QoE to satisfy their customers and avoid churn. We address the problem of real-time QoE monitoring of HAS, from the ISP perspective. Given the wide adoption of end-to-end encryption, we resort to machine-learning models to predict multiple QoE-relevant metrics directly from the analysis of the encrypted traffic, relying exclusively on lightweight network-level features. We introduce a machine-learning-based system for predicting QoE-relevant video-quality metrics in YouTube on the fly with very high accuracy. It is able to perform predictions in real time, during the course of an ongoing YouTube-streaming session, with the to date smallest time granularity of one second. Our framework focuses on the prediction of the video resolution, the average bitrate, and stalling occurrence. To do so, it analyzes ongoing streaming sessions using fine-grained time slots of 1 second. It computes multiple lightweight, snapshot-like statistical features from the video traffic in a stream-based fashion. Additionally, our technique uses two macro windows consisting of multiple time slots to capture trend and progression properties of the streaming session. More precisely, the system considers a first sliding window aggregating the last 3 time slots to compute trend features, and a second sliding window aggregating all past slots since the start of the session to compute session-progression features. At the end of each 1-second time slot, its features and those of the corresponding macro windows are fed into machine-learning models, which predict the video resolution (144p, 240p, 480p, 720p, or 1080p), the average bitrate, and stalling. *To the best of our knowledge, this is by now the finest time granularity for real-time prediction of video-quality metrics over encrypted traffic*.

For the evaluation of our framework, we use a very diverse dataset comprised of more than 15,000 different YouTube videos recorded under varying network conditions to obtain highly generalizable models. The videos were either streamed from a home or corporate WiFi network, or an LTE mobile network. For some sessions, a firewall was enabled, which blocked all QUIC traffic, such that the videos were streamed via TCP.

For each considered prediction target, we benchmark multiple machine-learning models and achieve very encouraging results with tree-based models. **For the video-resolution-prediction task**, our study suggests that a random forest with 10 trees is the most appropriate model: it is extremely lightweight and fast, and presents an excellent performance, with both recall and precision equal to at least 77% for each resolution class. **For the average-bitrate prediction**, we found an ensemble of 10 extremely randomized trees to be the best algorithm. It provides highly precise estimations, with a mean absolute error of only 93 kbps. Moreover, it runs much faster than the other tested algorithms. **For stalling detection**, bagging with 10 trees seems to be a good choice: even though it is slower than other models, the (recall, precision) tuple obtained with bagging is significantly higher than the one obtained by the other benchmarked algorithms. Nevertheless, this is the most challenging prediction task.



Recall for resolution prediction with random forest; precision scores are similar.

Confusion matrix for stalling detection with bagging.

Last but not least, we carried out a feature-importance analysis. We found that, for all targets, features summarizing the characteristics of the session since the beginning are the most relevant ones. For the video-resolution and the average-bitrate estimations, relying solely on these features improves the accuracy of the models. However, the results interestingly did not improve for the stalling detection, underlining that detecting rebufferings in videos is difficult.

Overall, we showed that our framework is very powerful for quality-metric inference in real-time scenarios.