Diagnosing Cellular Video QoE Impairments in the Wild

M. Hammad Mazhar
Department of Computer Science
The University of Iowa
muhammadhammad-mazhar@uiowa.edu

Zubair Shafiq
Department of Computer Science
The University of Iowa
zubair-shafiq@uiowa.edu

Daniele Brancato
Nokia Labs
daniele.brancato@nokia.com

Samer Zeineddine
Nokia Labs
samer.zeineddine@nokia.com

Anand Shah
Nokia Labs
anand.shah@nokia.com

6 Ji Hui
Nokia Labs
hui.ji@nokia.com

***Abstract*—Streaming video is by far the largest constituent of Internet traffic on cellular networks today. To satisfy strict Quality of Experience (QoE) requirements of streaming video on resource constrained cellular networks, network operators currently diagnose the root causes of video QoE impairments using ad-hoc and manual analysis that is infeasible at scale. This diagnostic process is further limited because the manual analysis by domain experts to determine the underlying root causes of video QoE impairments does not consider interdependencies between different systems in the cellular network. We propose a machine learning approach that leverages fine-grained Radio Access Network (RAN) and Core Network (CN) performance characteristics in a systematic and automated fashion for root cause analysis of video QoE impairments, while taking into account interdependencies among these performance characteristics. Our in the wild deployment in a large operational cellular network shows that video QoE impairments are impacted by cellular network load and radio channel conditions, as well as resource management schemes that can be manipulated by network operators for mitigation.**

I. Introduction

Motivation: Streaming video already constitutes the majority of data traffic on cellular networks [1], [2] and is expected to account for more than three-quarters of mobile data traffic by 2021 [3]. This growth is driven by users increasingly turning to their mobile devices to stream video content from popular mobile apps such as YouTube, Netflix, ESPN, and Twitch [4]. Cellular network operators aim to satisfy strict Quality of Experience (QoE) requirements of video streaming by employing a variety of network management techniques.

Limitations of Prior Work: To assess user experience of video streaming, cellular network operators use video-specific QoE metrics which cover various aspects of video streaming such as rebuffering. While prior work has developed methods that can be used to monitor QoE metrics for non-encrypted [5], [6] and encrypted [7], [8], [9] video from the network's perspective, it lacks automated and systematic methods to understand the underlying network characteristics that impact these QoE metrics. Cellular network operators not only need to be able to monitor video QoE but also determine the potential root causes of video QoE impairments.

Problem Statement: Cellular network operators need to figure out which network characteristics are the potential root causes of video QoE impairments. Video QoE impairments can be caused by issues at the client's video player, the video server, or the cellular network over which the video stream is accessed. Cellular network operators need to be able to pinpoint the root causes of video QoE impairments by analyzing massive records of numerous Radio Access Network (RAN) and Core Network (CN) characteristics, and mitigate them by employing different network management techniques.

Technical Challenges: It is challenging to identify root causes of video QoE impairments in operational cellular networks due to the *high dimensionality* of RAN and CN characteristics, and the *interdependence* between these characteristics. The complexity of cellular networks is reflected in high dimensional RAN and CN data belonging to multiple systems within cellular networks, which are designed to achieve non-overlapping objectives. These systems interact with each other in complex ways, which leads to interdependencies between network characteristics that may not be easily discernable. As such, potential root causes can comprise of multiple combinations of network characteristics.

Proposed Approach: We propose a supervised machine learning approach to diagnose root causes video QoE impairments in operational cellular networks using RAN and CN characteristics. To this end, we first identify a comprehensive set of features that cover different interactions in RAN and CN during a video streaming session. Our RAN features cover characteristics such as cellular load, radio channel conditions, and radio resource management decisions. Our CN features cover characteristics such as latency and retransmissions. This leads to a high dimensional feature set where there may be significant interdependencies among different features. Machine learning targets these two challenges by (1) determining which features are important to the target variable while (2) taking into account how they may affect each other through interdependencies. We use supervised machine learning to train decision tree models that can be analyzed to identify

combinations of cellular network characteristics that cause video QoE impairments.

Key Findings: We deploy our machine learning approach in the wild on RAN and CN data collected from a tier-1 LTE cellular network in the United States covering a major metropolitan area. We summarize the key findings of our analysis of feature distributions and trained decision trees below. First, high *cellular load* induces more video QoE impairments. We observe that more active, prolonged connections led to more video QoE impairments in video streams. Second, degraded *radio channel conditions* are more likely to cause video QoE impairments. We observe that video sessions experiencing lower signal strength and higher interference are more likely to develop video QoE impairments. Finally, certain *radio resource management decisions* lead to video QoE impairments. We observe that resource allocation and transmission management procedures are prevalent in sessions with video QoE impairments. This leads us to believe that there are configuration issues in the cellular network that cause video QoE impairments.

II. Background

We provide a brief background of adaptive video streaming and LTE network architecture in this section.

A. Adaptive Video Streaming

Video Delivery Architecture: Over-The-Top (OTT) video streaming providers such as YouTube and Netflix use HTTP-based adaptive streaming protocols such as DASH (Dynamic Adaptive Streaming over HTTP). The video server stores videos as fixed-duration segments that are encoded at multiple quality levels. The video player then streams video by sending a series of HTTP GET requests, each of which correspond to a request for a video segment at a suitable quality level. The downloaded segments are stored in the player's buffer for smooth playback while waiting for future segments to download. The quality level for each segment request is determined by the quality adaptation logic, which is implemented in the player, based on factors such as available network bandwidth and remaining video buffer.

Video Quality of Experience (QoE) Metrics: Prior research has extensively conducted user studies to assess video QoE using subjective metrics such as Mean Opinion Score (MOS) scores [10]. Since MOS scores are difficult to obtain at scale, prior research has proposed objective video-specific metrics that correlate well with subjective MOS scores. In this paper, we use the following video QoE metrics.

• **Startup delay:** A video streaming process has two phases: startup and steady-state. In the startup phase, the player waits for the video buffer to fill up to a predefined level (typically 10 seconds [11]) before the playback starts. The time taken in this phase is known as *startup delay*. High startup delay is known to negatively impact QoE [12].

• **Video quality:** The video streaming service encodes video segments at different quality levels. The player requests video segments at varying quality levels based on the decisions made by the quality adaptation logic. These quality levels represent a nominal scale of video resolutions available to the video client. The resolution of video playback, represented by the modal quality level in a time period, is known as *video quality*. Fluctuations in video quality are known to negatively impact video QoE [13].

Note that these video QoE metrics can be directly measured at the video player, but are not visible to the network operator. As we discuss later, network operators use various approaches to estimate video QoE metrics.

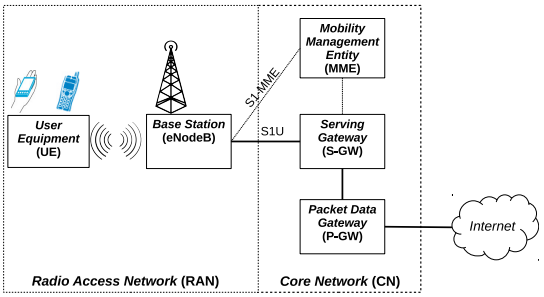


Figure 1. LTE network architecture

B. LTE Network Architecture

Overview: The LTE architecture is designed to allow User Equipment (UEs) like smartphones to access data services such as OTT video streaming. Figure 1 shows LTE network architecture which is composed of the Radio Access Network (RAN) and the Core Network (CN). The RAN is composed of UEs and cellular base stations (eNodeBs). The CN is composed of Mobility Management Entity (MME), Serving Gateway (S-GW), and the Packet Data Gateway (P-GW). LTE networks typically cover large geographic areas and consist of a pool of MMEs, S-GWs, and P-GWs for redundancy and load balancing. The RAN and CN connect via the S1U interface between eNodeBs and S-GWs and the S1-MME interface between eNodeBs and MMEs. eNodeBs typically control multiple cell sectors and are responsible for managing assignment of radio resources to UEs. The MME manages UE access authentication and eNodeB to S-GW attachment. The S-GW acts as a mobility anchor for UEs and is responsible for routing data packets as users move between different cell sectors and eNodeBs (i.e., handovers). The P-GW acts as the point of access to external networks including the Internet.

Resource Management: Cellular networks try to utilize the licensed radio spectrum as efficiently as possible because it is a scarce and expensive resource [14]. Cellular network operators divide the limited radio spectrum across different cell sectors and eNodeBs. Since cellular networks are oversubscribed, eNodeBs manage radio resources through *admission control* and *resource allocation*. Admission control determines whether a new connection request by a UE is accepted. In order to ensure acceptable service levels, the network operator can configure eNodeBs to accept a limited number of connections based on current cell load and the nature of the connection request. If the request is accepted, the eNodeB then determines suitable radio channel (e.g., shared, control) and transmission modes (e.g., modulation) based on measured channel conditions reported

by the UE and the current load on the eNodeB. Resource allocation determines how the available radio resources are assigned to different UEs already connected to the cellular network. Each UE requests resources based on its requirements, which can vary depending on the type of application (e.g., voice, data)[15]. In allocating resources to requesting UEs, the eNodeB takes into account a large number of characteristics such as radio channel conditions and mobility. The eNodeB also ensures that the resource allocation satisfies the objectives set by the network operator such as providing minimum service levels. The eNodeB allocates radio resources in terms of Physical Resource Blocks (PRBs) for transmissions to/from each UE.

Connection Establishment: To access data services, a connection has to be established between the UE and P-GW. First, the UE establishes a radio connection with an eNodeB, which then connects to a S-GW for packet data routing and mobility anchorage. Finally, the S-GW establishes a connection with the P-GW for the UE to access data services. These connections are maintained until either the UE has no more data to send/receive or remains inactive for a specified period of time. This set of connections from the UE to the P-GW is called an Evolved Packet System (EPS) bearer.

III. Video QoE Troubleshooting

Since RAN is oversubscribed, cellular network operators closely monitor RAN performance to detect and troubleshoot video QoE impairments. To this end, cellular network operators capture packet traces at the S1U interface between the RAN and CN. They then identify flows belonging to video streaming services by filtering on IP addresses (or hostnames) belonging to known video streaming services. If traffic is unencrypted, Deep Packet Inspection (DPI) based techniques [6], [16] can be used for estimating video QoE metrics. Otherwise, machine learning based techniques [9], [8], [7] can be used for estimating video QoE metrics. The estimated video QoE metrics are then analyzed to determine if there were any impairments. For instance, network operators may look at startup delay for video streaming sessions and label the sessions with high startup delay as problematic. If substantial impairments are detected, cellular network operators initiate a manual and ad-hoc process to identify the underlying root causes.

To this end, cellular network operators analyze fine-grained RAN event records which capture detailed information about connection establishment, resource management, and handovers, and packet traces are processed to extract features. For connection establishment, event records include information such as connection duration, the eNodeBs/cell sectors connected to, and the type of connection requested. For resource management, event records include information such as physical radio resource allocations, type of modulation, channel conditions, and transmission counters. For handovers, event records include information such the reason for handover, handover type, and channel conditions for the source/target cell sector. These event records are used to extract a variety of features that are then correlated with video QoE metrics. This statistical analysis is then reconciled with domain knowledge to determine root causes of video QoE impairments.

A. Features

Next, we discuss different RAN and CN features that are typically used in this troubleshooting process. *RAN features* capture radio network characteristics for video sessions, and *CN features* capture core network interactions for video sessions.

1) RAN features: RAN features are extracted from event records that capture detailed information about various aspects of RAN interactions such as channel and UE conditions. These RAN interactions are further qualified by the context in which they are captured such as connection establishment, resource management, and handovers. RAN features are extracted both for cell sectors and individual UEs. We first discuss RAN features based on what information they capture and then discuss the context in which they are captured. Table I lists different types of RAN features.

• *Channel condition features* provide information on the radio conditions such as signal strength and interference being experienced by the UE. The UE computes Channel Quality Indicator (CQI) values based on features such as Receive Signal Strength Indicator (RSSI) and Signal-to-Interference-plus-Noise Ratio (SINR) [17]. The eNodeB connected to the UE uses CQI values reported by the UE to determine transmission modes and PRB allocations, which in turn determine reliability and throughput of data transmission. If radio conditions are not conducive for data transmission, the eNodeB may decide to handover the UE to one of several neighboring cell sectors on the same or different eNodeB. To identify suitable target cell sector, the source eNodeB relies on radio condition measurements for other cell sectors by the UE such as Reference Signal Receive Power (RSRP) and Reference Signal Receive Quality (RSRQ).

• *Event count features* quantify occurrences of specific events during the life cycle of a LTE connection. These counts can provide information on how the eNodeB and UE behave during the connection. For instance, the number of handovers occurring for a specific UE can indicate mobility characteristics for the UE. As an another example, the number of active connections at a cell sector can indicate the load it is experiencing, which would influence how PRBs are allocated across existing and new UEs.

• *Event cause features* indicate the cause of certain events in the life cycle of a LTE connection. These features influence decision making at the eNodeB. For instance, the reason for ending a connection could be a Radio Link Failure (RLF), which would result in requesting a new radio connection. The reason for a failure may determine how the eNodeB may respond. Recurring failure causes can be indicative of configuration issues at the eNodeB.

• *UE condition features* contain information regarding UE characteristics. Feature Group Indicators (FGIs) encode LTE

Feature type	Example Features
Channel conditions	RSSI, SINR, RSRP, RSRQ, change in CQI, Frequency band
Event counts	Handovers, Failures, Connections, Periodics, Radio Link Failures
Event causes	Reason for handover, Reason for establishing S1U connection, Reason for Failure, Reason for establishing RRC connection
UE conditions	Distance from cell, Feature Group Indicator (FGI)
Transmission statistics	Throughput, Physical Resource Block (PRB) allocation, Buffer Status Reports (BSRs), Antenna usage, Power Head Room, Transmission mode (Single, Dual) statistics, Modulation and Control Scheme (MCS)

Table I
RAN FEATURE TYPES AND THEIR EXAMPLES

Context type	Example features
Connection	UE distance, device type, frequency band, reasons for S1U and RRC connections, starting and ending causes, UE identifiers, and radio channel conditions.
Periodic	Radio channel conditions for neighboring cells (RSRP, RSRQ), transmission statistics on the Medium Access Control (MAC) layer, uplink/downlink PRB allocations, delay, MCS, and change in CQI
Handover	Reason for handover, handover type, and radio channel conditions for source and target cell sector.
Failure	Reason for failure and radio channel conditions at time of failure

Table II
CONTEXT TYPE IN LTE CONNECTION LIFE CYCLE AND THEIR EXAMPLES

capabilities of the UE such as supported frequency bands and modulation schemes. FGIs determine how the eNodeB allocates radio resources to the UE which would lead to differences in throughput and delay. A UE's distance from the eNodeB determines its mobility characteristics such as near vs. far and stationary vs. mobile. A stationary UE can experience connectivity issues if it is far away from the eNodeB. A highly mobile UE can initiate a large number of handovers if it moves between cell boundaries.

• *Transmission statistic features* capture resource allocation information during transmissions carried out in a LTE connection. Since video streaming applications on LTE networks are served using Non Guaranteed BitRate (Non-GBR) bearers, the throughput achieved by video streaming applications can fluctuate. The eNodeB assigns PRBs to a UE before any transmissions take place. Depending on the frequency band, the eNodeB has a limited number of Physical Resource Blocks (PRBs) available to allocate per Transmission Time Interval (TTI). This allocation along with the Modulation and Control Scheme (MCS) used for transmissions determines the maximum available throughput. Data can be buffered at either the UE during uplink transmissions or the cell sector during downlink transmissions. The number of TTIs with buffered data indicates load at the eNode, which can impact video QoE.

The aforementioned features are reported in different contexts during the life cycle of a LTE connection. Table II lists these context types and their examples.

• *Connection* records capture the establishment and termination contexts of a LTE connection. Connection establishment details such as the reason for establishing various stages of the connection (connecting to the radio network, the core network), the frequency band used, and the UE FGI are recorded. Radio conditions such as RSRP and RSRQ and the UE's distance from the eNodeB are recorded during the life cycle of a LTE connection. The reason for connection termination is also recorded. Note that a UE can establish multiple connections over time, either to the same cell sector or to different cell sectors due to mobility or load issues.

• *Periodic* records are generated during the life cycle of a LTE connection. There are three types of periodic records. First, the UE records radio conditions for neighboring cell sectors in terms of RSRP and RSRQ when the eNodeB intends to initiate a handover. By monitoring these records, the eNodeB can determine a suitable target cell sector for handover. Second, Media Access Control (MAC) layer statistics such as the number of TTIs with uplink/downlink buffering and PRB allocations are recorded to capture MAC layer performance. Finally, transmission statistics such as Power Head Room (PHR), transmission modes, successful/failed transmission counters, Single/Dual Codeword mode transmission counts, and MCS are recorded to capture transmission decisions made by the cell sector and UE. Note that Single/Dual Codeword modes refer to Multiple-In-Multiple-Out (MIMO) capabilities. Transmissions made in Single Codeword mode use only one antenna, while those made in Dual Codeword mode use multiple antennas. Dual Codeword mode transmissions achieve higher throughput and are more resilient to errors. Acknowledgements (ACKs) and non-Acknowledgements (NACKs) are also recorded for transmissions made in these modes.

• *Handover* records capture the handover contexts during the life cycle of a LTE connection. Handover details such as the reason for handover, handover type, and radio channel conditions (RSRP and RSRQ) for the source and target cell sector are recorded.

• *Failure* records capture the context of failures in the life cycle of a LTE connection. Failure details such as the reason for failure and radio channel conditions (e.g., RSRP, RSRQ) at the time of failure are recorded. The reason for failure can impact subsequent decisions made by the UE or eNodeB to recover from the failure, which can impact video QoE.

2) CN features: These features are extracted from packet traces for video streaming sessions collected at the S-GW. These features capture TCP header information at the transport layer. These features are:

• *Packet counters* record the number of uplink (to the CDN server) and downlink (to the UE) packets. Downlink packet counts indicate how much video data is sent by the CDN server.

• *Retransmission ratios* record the ratio of retransmitted packets in uplink and downlink directions. Retransmissions are determined by analyzing TCP sequence numbers.

• *Full window ratios* record the ratio of packets that advertise full receive windows.

• *Uplink latency* records the time it takes for an upstream packet to be acknowledged by the CDN server at the S-GW. Higher uplink latencies are indicative of congestion and delay between the CDN server and the S-GW in the CN.

• *CDN host count* records the number of CDN servers with which a connection was established. A UE establishing connections to multiple CDN servers indicates load balancing by the CDN, possibly due to congestion.

Over these raw features, cellular networks may further compute summary statistics. For instance, mean, median, max, min, mode and standard deviation may be computed for numeric features, and mode for nominal features. This results in a total of 668 features in our data.

B. Correlation analysis

We now demonstrate the traditional troubleshooting process using the features outlined previously on data collected from an operational tier-1 LTE cellular network in the United States. This data was collected from 14 eNodeBs in the network for 1 hour for 6 days, totaling 6 hours of data. We collect LTE event records for all cell sectors on the selected eNodeBs, which capture RAN interactions for LTE connections to those eNodeBs including internal IP addresses assigned to the connections. Around 80K UEs connected to the selected eNodeBs on average per day. For the connected UEs, we collect event records for around 2.8M LTE connections, 38K radio failures, 800K handovers, and 40M periodic reports. We also collect packet capture traces at the S1U interfaces of the selected eNodeBs, which contain network and transport layer information for TCP connections, but not packet payloads due to the use of encrypted protocols such as HTTPS.

We use a proprietary tool [18] to identify 850 YouTube video streaming sessions and estimate video QoE metrics from the packet traces. The tool identifies these sessions by extracting host names from DNS records for destination IP addresses and comparing these host names to known video streaming services. LTE event records for these sessions are identified by comparing internal connection identifiers reported in the event records and packet traces and aligned based on timestamps. The event records and traces are then split into 10-second windows, which allows us to capture the fluctuating nature of both network features and video QoE metrics. After feature extraction and video QoE metric estimation, our dataset consists of more than 24K windows with RAN and CN features and video QoE metrics.

We perform correlation analysis using the Kendall's τ [19] coefficient whose value can range from -1 to +1, the magnitude indicates the strength and the sign indicates the direction (direct/inverse) of the correlation. Since we have hundreds of features, we limit our attention to features whose correlation magnitude exceeds a minimum threshold of 0.1.

Troubleshooting startup delay. Table III lists the feature correlation coefficients for startup delay. The maximum number of Radio Link Failures (RLFs) has the highest correlation (-0.1751) with startup delay. RLFs lead to transmission delays as the eNodeB and UE attempt to recover from them. It follows that such delays would translate in video streaming in the form of startup delays. We also see negative correlations with various radio condition indicators such as mean Physical Uplink Shared Channel (PUSCH) SINR (-0.1673), uplink SINR (-0.1305), RSRQ (-0.1233) and RSRP (-0.1035 for the serving cell). These indicate that degrading radio conditions can play a part in increasing startup delay, a point supported by domain knowledge. We also see negative correlations with CDN features, such as packets from/to the CDN, CDN retransmission ratios, full window ratios. More packets in either direction would indicate that the video streaming application is requesting and receiving more video chunks, which should reduce startup delay. However, the observation that having more retransmissions is correlated with a reduction in startup delay is counterintuitive and is likely an effect from some other underlying cause feature such as throughput.

Feature Name	**Kendall τ**
Maximum Radio Link Failures	0.1751
Median of Mean PUSCH SINR	-0.1673
Maximum Rank Indicator 1 Ratio	-0.1506
Median Dual Codeword Transmissions	-0.1313
Mean Rank Indicator Rank 1 Ratio	-0.1310
Mean Uplink SINR	-0.1305
Mean CDN Latency	-0.1301
Maximum Dual Codeword Transmissions	-0.1285
Minimum RSRQ of Serving Cell	-0.1233
Minimum UE Distance	-0.1222
Maximum Rank Indicator 3 Ratio	-0.1197
Std. Dev CDN Downlink Packets	-0.1193
Maximum PDSCH PRBs Allocated	-0.1161
Mean CDN Downlink Retransmission Ratio	-0.1157
Std. Dev. CDN Uplink Retransmission Ratio	-0.1156
Minimum CDN Downlink Full Window Ratio	-0.1132
Median Last Reported RSRP of Serving Cell	-0.1125
Mean CDN Uplink Full Window Ratio	-0.1081
Maximum RSRP of Target Cell	0.1051
Maximum of Last Reported RSRP of Serving Cell	-0.1035
Median of Last RSRQ of Serving Cell	-0.1029
Mean CDN Uplink Packets	-0.1026

Table III
CORRELATION COEFFICIENTS FOR STARTUP DELAY

Feature Name	**Kendall τ**
Maximum CDN Downlink Packets	0.2929
Std. Dev CDN Downlink Packets	0.2877
Mean CDN Downlink Packets	0.2366
Mean Downlink Received Transmissions	0.1627
Maximum Downlink Received Transmissions	0.1601
Median Downlink Received Transmissions	0.1575
QCI at time of Handover	-0.1500
Mean TTIs with Positive Uplink Buffer	0.1471
Std. Dev TTIs with Positive Uplink Buffer	0.1434
Maximum TTIs with Positive Uplink Buffer	0.1426
Minimum Received Transmissions	0.1393
Median Uplink/Downlink Non-GBR Throughput	0.1369
Maximum Uplink/Downlink Non-GBR Throughput	0.1366
Mean Uplink/Downlink Non-GBR Throughput	0.1365
Median TTIs with Positive Uplink Buffer	0.1346
Minimum Uplink/Downlink Non-GBR Throughput	0.1342
Median CDN Downlink Packets	0.1329
Modal QCI	-0.1261
Minimum PDSCH PRBs Allocated	0.1188
Maximum Single Codeword Transmissions	-0.1135
Maximum Single Codeword Starting Transmissions	-0.1132
Std. Dev Uplink/Downlink Non-GBR Throughput	0.1129
Starting Connections	-0.1007

Table IV
CORRELATION COEFFICIENTS FOR VIDEO QUALITY

Troubleshooting video quality. Table IV lists the feature correlation coefficients for video quality. We see positive correlations with RAN features such as received downlink transmissions, Non-GBR throughput and Physical Downlink

Shared Channel (PDSCH) PRBs allocated. The first two features can be considered to be a result of requesting higher video quality chunks, with higher data requirements requiring more throughput and transmissions to be completed. This can only be possible if the quality adaptation logic of the video player recognizes that network conditions (as seen by the player) support higher video quality. This effect also shows up with CDN features such as the number of downlink packets from the CDN, which has 3 features with the highest correlation seen thus far. We observe that 3 features are negatively correlated with video quality; QCI, Single Codeword transmissions and the number of starting connections. QCI values indicate the Quality of Service (QoS) requirements for LTE connections, and lower QCI values are used to indicate when QoS requirements are more stringent. It would follow that connections with lower QCI values would have priority for resources, so the correlation makes sense. Single Codeword transmissions are used when MIMO capabilities are not available resulting in lower throughput. Having more of such transmissions would result in the video player reducing video quality, hence the negative correlation.

C. Limitations

The main issue with this troubleshooting process is its manual nature—it puts the burden on domain experts to analyze hundreds of correlation coefficients to determine the potential root causes. Another key issue is that it assumes that different cellular network features are independent. The complex nature of LTE networks means that there are inherent interdependencies between various aspects. As such, the actual root cause of video QoE impairments is not clearly visible and can lead to misleading conclusions. First, our correlation analysis shows that RLFs and radio conditions are correlated with startup delay. Based on domain knowledge cellular network operators may conclude that improving radio conditions will reduce RLFs, thereby reducing startup delay. However, RLFs may also be happening due to configuration issues at eNodeBs. Second, our analysis shows that lower QCI improves video quality so cellular network operators may decide to assign lower QCI values to LTE connections more aggressively. This can have a knock-on effect of more impairments, as the stricter QoS requirements can make it harder for resource allocation algorithms to provide sufficient resources. Finally, our analysis shows that having a higher CDN downlink retransmission ratio leads to lower startup delay. Not only is this counterintuitive, but it is likely an effect of some other feature such as throughput. A cellular network operator may choose to ignore higher retransmission ratios due to this conclusion and in the process miss the actual root causes of high retransmission ratios.

In summary, traditional troubleshooting fails to produce meaningful root causes for video QoE impairments in cellular networks, due to manual analysis and limited view of how features may depend on each other.

IV. Proposed Approach

To alleviate the shortcomings of the current method of diagnosis of video QoE impairments used by cellular network operators, we propose a machine learning based approach to automatically identify the root causes of video QoE impairments. Machine learning techniques are designed to systematically evaluate multiple features and shortlist the most effective ones, while also evaluating interdependencies amongst those features. These techniques can then build models that describe how features are related to each other and to the target variable. In this case, we use machine learning techniques to evaluate RAN and CN features for correlations with video QoE metrics while taking into account any interdependencies that may exist between the features.

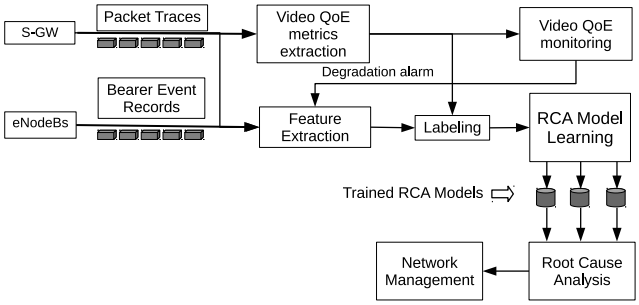


Figure 2. Overview of proposed approach

A. Overview

Figure 2 provides the overview of our proposed approach. The cellular network is configured to collect packet traces at the S-GW for ongoing video streaming sessions. Video QoE metrics for such sessions are estimated and monitored by an automated system which checks whether video QoE metrics are significantly degraded and raises alarm if applicable. If an alarm is raised, RAN event records for ongoing connections are retrieved. RAN event records and CN packet traces are then processed to extract an extensive set of features in a window based fashion to capture dynamics of RAN and CN characteristics. To build machine learning models, the extracted features are labeled with the estimated video QoE metrics and fed to supervised machine learning algorithms. The trained models are analyzed to gain insights into the root causes of video QoE impairments. Cellular network operators can leverage these insights to employ network management strategies to mitigate video QoE impairments.

B. Supervised Model Training

We use supervised machine learning techniques to train models that can use RAN and CN features to predict video QoE labels. Specifically, we use supervised decision trees built using the C4.5 [20] algorithm with 668 RAN and CN features as input and 2 QoE metrics as output. Note that we use the same features as in our correlation analysis. Decision trees partition data into smaller sets based on the discriminating power of features in terms of information gain. Compared to other supervised machine learning techniques, decision tree models are human interpretable which make them amenable for root cause analysis given the decision tree's output decision. Due to their accuracy and interpretability, decision trees have been widely deployed to detect and diagnose network performance issues [21], [22], [23], [24]. We analyze the

trained decision trees to pinpoint potential root causes of video QoE impairments.

C. Estabilishing Ground Truth

Our supervised machine learning based approach poses the problem of detecting video QoE impairments as a binary classification problem. Specifically, we design our approach to detect whether a QoE metric for a video streaming session is in a Good state or Bad state. To this end, we need to define what Good or Bad means for each video QoE metric. While this can be determined through subjective means i.e. by conducting user studies, we instead rely on data analysis techniques. We rely on the distributions of our features to tell us what is Good or Bad for each QoE metric. The key assumption here is that sessions undergoing Good or Bad QoE will result in feature distributions that are different. As such, QoE metric thresholds that result in the most significant differences will lead to better detection of causes of video QoE impairments.

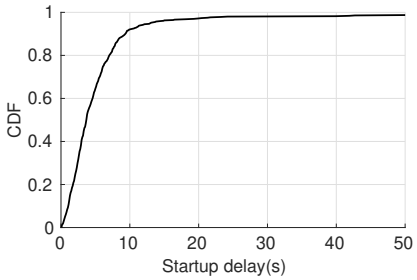
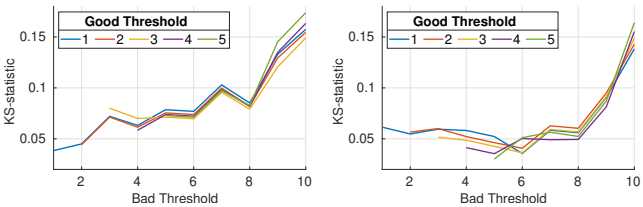


Figure 3. Distribution of startup delay across video streaming sessions

To perform this analysis, we rely on the two-sample Kolmogorov-Smirnov or KS test [19] which is designed to test if two empirical distributions differ. The test calculates a value called the *KS-statistic*, which is defined as a distance between two empirical distribution functions. To determine the best set of QoE metric thresholds, we partition our dataset into sets of Good and Bad sessions/windows using a "configuration" of thresholds for a specific video QoE metric. Then we calculate the KS-statistic between empirical distributions of the same feature from each set of sessions. We repeat the process for various configurations for the specific QoE metric, and select the configuration which produced the highest KS-statistic for that feature. We repeat this process for multiple features, which then provides us with a set of configurations for the QoE metric in question. From this set, we select the modal configuration i.e. the most common configuration that appeared for the QoE metric across our feature set.



(a) Maximum of Last Reported RSRP at Serving Cell

(b) Minumum of Last Reported RSRQ at Serving Cell

Figure 4. KS-statistic plots for QoE metric configurations of startup delay. X axis indicates the Bad theshold, each line indicates a Good threshold.

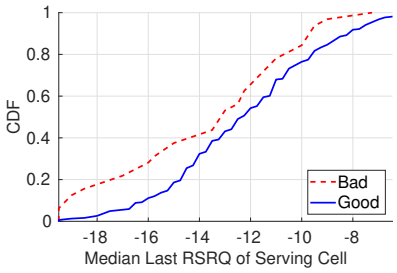
We demonstrate this process using startup delay. Figure 3 shows the distribution of startup delay across the video streaming sessions in our dataset, excluding any sessions with reported startup delay to be zero. We observe that more than 90% of sessions remaining experienced at most 10 seconds of startup delay. We limit our range of configurations between 1-10 seconds of startup delay. Each configuration consists of two thresholds; the Good threshold defines the maximum startup delay considered to be Good startup delay, and the Bad threshold defines the minimum startup delay to be Bad startup delay. We consider Good startup delay thresholds from 1-5 seconds and Bad thresholds from 1-10 seconds. Note that by definition, the Bad threshold cannot be less than the Good threshold. After creating configurations, we calculate KS-statistics for them for all features in our dataset.

Figure 4 shows plots for KS-statistics calculated for specific features. For both of these features, we see that Good $\leq$ 5 seconds, Bad $>$ 10 seconds of startup delay had the highest KS-statistic. This indicates that highest difference between Good and Bad sessions for these features was achieved at those thresholds for startup delay. Hence we select the configuration that appears the most across our features i.e. the modal configuration. For startup delay the modal configuration across all features was Good sessions with up to 5 seconds of startup delay and Bad sessions with at least 10 seconds of startup delay. Repeating this analysis for video quality, the modal configuration identified video resolutions of 480p and above as Good and below 480p as Bad video quality.

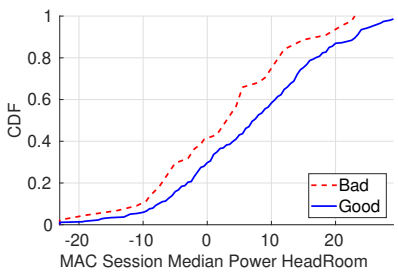
D. Feature Distributions

With our ground truth established, we now analyze cumulative distribution functions (CDFs) for RAN and CN features to understand how they differ based on our ground truth . Based on our method for establishing ground truth in Section IV-C, video streaming sessions with upto 5 seconds of startup delay are labeled Good while sessions with at least 10 seconds of startup delay are labeled Bad. Since startup delay is a one-time event in a video streaming session, we only consider features for the first 10 second window of the session for this analysis. For video quality, we label feature windows with resolutions of 480p and above as Good and below 480p as Bad. Note that video quality varies across the streaming session, so multiple windows are labeled for each session in our dataset. We then compare and discuss example feature distributions for both classes for each QoE metric. Figures 5 and 6 respectively show selected CDFs conditioned for startup delay and video quality. We now discuss these distributions to gain key insights into the characteristics in cellular networks that correlate with video QoE impairments.

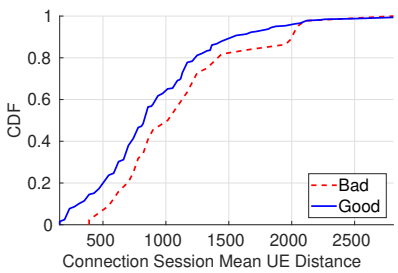
Figure 5(a) plots the median of last reported RSRQ for the serving cell with respect to startup delay. We first observe that sessions with Bad startup delay consistently had lower median reported RSRQ than sessions with Good startup delay, with larger differences at lower RSRQ values. RSRQ is used by the eNodeB for various purposes, such as deciding resource allocation and transmission modes. Next, Figure 5(b) plots



(a) Median Reported RSRQ of Serving Cell

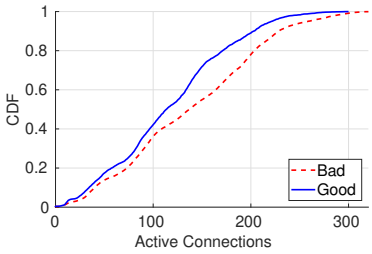


(b) Median Power Head Room

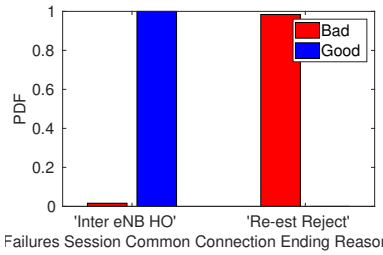


(c) Mean UE distance

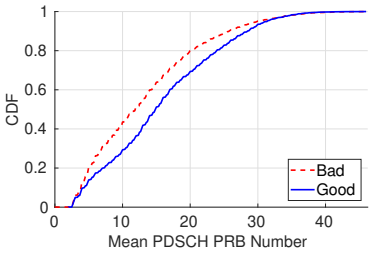
Figure 5. Radio Access Network (RAN) features for startup delay



(a) Active LTE connections



(b) Connection ending reason at failure



(c) Mean PDSCH PRBs Allocated

Figure 6. Radio Access Network (RAN) features for video quality

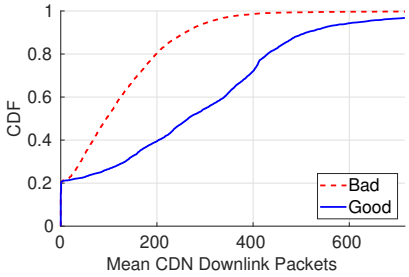


Figure 7. Mean CDN downlink packet count for video quality

median Power Head Room (PHR) with respect to startup delay, PHR is a measure of remaining transmission power available for the UE based on a limit. Positive values indicate that there is space for the UE to transmit at higher power, which increases available bandwidth. Negative values on the other hand indicate that the UE is already transmitting above allowed limits, mostly due to interference. Again, we see that sessions with Bad startup delay report lower PHR than sessions with Good startup delay. In short, we observe that video QoE impairments often correlate with degraded radio channel conditions.

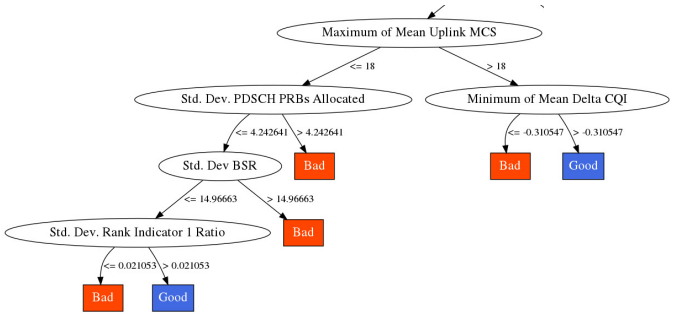
Figure 5(c) plots the distribution for average UE distance with respect to startup delay. We note that sessions with Bad startup delay were further away from the connected eNodeB, compared to sessions with Good startup delay. Larger distances lead to less reliable data transmissions due to increased path loss [25]. We find that video QoE impairments are more prevalent in UEs that connect under non-favorable conditions.

Figure 6(a) plots the distribution of active LTE connections in the serving cell for video session windows with respect to video quality. We observe that sessions with Bad video quality have higher cell load in terms of active connections Under high cell load, the eNodeB is constrained by the objective of providing minimum service levels to all connected UEs. This can lead to radio resource management decisions that lead to insufficient radio resources allocated to UEs running video streaming applications, hence we find that video QoE impairments tend to occur during periods of high cellular load.

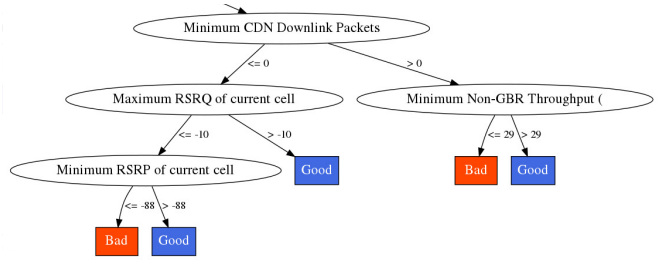
Figure 6(b) plots the distribution of the reasons for terminating a connection in the event of a failure with respect to video quality. All windows with Good video quality had connections terminated during failures by inter-eNodeB handovers, whereas the majority of Bad windows had the connection terminated by connection reestablishment rejection. By terminating a connection through a handover, the eNodeB ensures that the UE still receives service. A reestablishment rejection means that the UE will have to initiate a request for a new connection, which disrupts service. This leads us to believe that video QoE impairments are correlated with the occurrence of specific events in the life cycle of a LTE connection.

Figure 6(c) plots mean of Physical Downlink Shared Channel (PDSCH) PRB allocations with respect to video quality. These PRBs are allocated for downlink transmissions and are a major contributor to the final throughput observed by the UE. We observe that a significant portion of session windows with Bad video quality have lower PRBs allocated compared to windows Good video quality. We find that certain radio resource management decisions correlate with video QoE impairments, as resource management decisions determine factors such as the amount of PRBs allocated for data transmissions.

Figure 7 plots the distribution of mean number of packets seen downlink from the CDN server with respect to video quality. We observe that windows with Bad quality received fewer packets compared packets received by Good windows. Naturally, video segments of higher quality will require send-



(a) Subtree considering Uplink MCS, Delta CQI, PDSCH PRBs, BSR and Rank Indicator 1 ratio



(b) Subtree considering CDN Downlink Packets RSRP, RSRQ and non-GBR throughput

Figure 8. Subsections of decision tree trained for video quality. Each subsection shows how the decision tree considers various RAN and CN characteristics to estimate video quality. Red leaf nodes indicate that video quality will be Bad while blue leaf nodes indicate Good video quality

ing a larger number of packets. This reflects a feature that is not under control of the cellular network operator, and as such provides no actionable insight for the operator. As we discuss next, in our machine learning approach to root cause analysis, we target such CDN specific features by determining if they correlate to any RAN features, reconcile the correlation with domain knowledge and remove the CDN feature to avoid conclusions which provide no actionables to the cellular network operator.

E. Root Cause Analysis using Machine Learning

We now present and discuss the results of our machine learning based approach. To ensure that our root cause analysis focuses on the effects of RAN characteristics, we preprocess RAN-CN feature pairs to determine if there are any significant correlations between them. We extract Pearson's correlation coefficient for these feature pairs, investigate pairs with high correlation coefficients and utilize domain knowledge to determine whether the correlation implies causation. If the RAN-CN feature pair has a cause-effect relation, we remove the CN feature. After feature correlation analysis to remove CN features, we proceed to decision tree training. We label each window as Good or Bad based on the two video QoE metrics discussed in Section IV-C and IV-D. This provides us with a dataset of labeled instances that can used for training of binary classification trees, with each label representing a class. Subsequently, we train a separate tree for each video QoE metric and analyze the tree structure to gain insights into the root causes of video QoE impairments. Under 10-fold cross validation, our trained trees achieve 85%-80% accuracy in predicting video QoE. Figure 8 shows sections of the decision tree built to determine whether video quality during a 10-second window of an ongoing video streaming session is Good or Bad. We specifically focus on sections of the tree which are dominated by RAN features, since these are the factors that the cellular network operator can most influence.

Figure 8(a) shows one subtree that starts by considering the maximum of mean uplink MCS. MCS determines how data is encoded and transmitted and is selected by the eNodeB based on radio conditions, with transmissions at higher MCS achieving higher throughput at the cost of increased sensitivity to radio conditions. If mean MCS is above 18, the tree moves on to consider the minimum of mean delta CQI, which is the mean change in CQI observed between periodic reports. Negative changes in CQI indicate degradation in radio conditions, with negative changes beyond a point leading to transmission losses at higher MCS values. The video player likely picks up the transmission losses as impaired network conditions and reduces video quality. In the case when MCS is up to 18, the tree then considers standard deviations of PDSCH PRBs, BSR and Rank Indicator 1 Ratio. For the first two, having higher standard deviations leads to Bad video quality. As PRBs determine application-layer throughout, more variability would lead to fluctuating throughput which would induce the video player to be more conservative with video quality. BSR reports the amount of data buffered at the UE for transmission. Again, more variability here would indicate fluctuating network conditions to the video player. Finally, Rank Indicator 1 ratio measures the proportion of time the UE cannot engage in MIMO transmissions, which provide higher throughput. Less variability in Rank Indicator 1 indicates that there was less opportunity for the UE to send/receive MIMO transmissions, which lead to Bad Video Quality.

Figure 8(b) shows another subtree in the same decision tree which starts by considering the minimum number of downlink packets seen from a CDN server. If this is non-zero, the tree then looks at the minimum non-GBR throughput which if under 29 Mbps is considered to lead to Bad video quality. Note that other applications may be competing for bandwidth with the video player at the UE, hence having lower throughput overall can hurt video quality, as the video player will behave to limit it based on its observed throughput. In the case when the minimum downlink packets is zero, the tree then examines maximum RSRQ and minimum RSRP of the cell the UE is connected to. In both cases having higher values beyond a point (-10 dB for RSRQ and -88 for RSRP) indicates that video quality will be Good, otherwise video quality is Bad.

At surface, it may seem like our analysis of the decision trees uncovers patterns that are discussed in Section IV-D (e.g., cellular load and adverse radio conditions can cause video QoE impairments). However, there is a key advantage provided by machine learning: it takes into account *inter-dependencies* between different features. Every feature that appears in our decision trees is dependent upon features that appear above it.

These dependencies provide important context on how these features effect video QoE. In the examples above it is not just one aspect of the cellular network, but a combination of multiple aspects that leads to video QoE impairments. As such, our approach utilizing machine learning is key for cellular network providers to understand how interactions between different aspects of their network can lead to video QoE impairments. In the next section, we discuss how insights gained from our approach can be utilized by cellular network operators to mitigate QoE impairments for video streaming.

F. Actionable Insights

From the analysis of the decision trees for video QoE metrics, we observe that cellular load, radio channel conditions, and radio resource management decisions impact video QoE impairments. We now discuss how cellular network operators can act on these insights to mitigate video QoE impairments.

Cellular load: Cellular load can be of two types: *transient* and *systemic*. Transient cellular load occurs rarely and only persists for a short period of time. For example, a concert or a sports game event can cause transient cell load as more than usual UEs try to connect to the network [26]. To alleviate problems caused by such loads, cellular network operators can deploy more short term resources such as portable base stations called Cells On Wheels (COWs) or femto/pico base stations [27]. On the other hand, systemic cellular load can exist for longer periods of time. In this case, by monitoring long-term changes in cellular load over time, network operators can determine if they need to re-architecture the network to accommodate higher cellular load. Such network re-architecturing is typically part of long term planning of cellular network operators.

Radio channel conditions: Radio channel conditions are influenced by many factors such as the signal strength and interference in the wireless medium. Cellular network operators can use low signal strength to identify weak spots in their coverage. They may then either deploy additional radio resources or configure existing nearby eNodeBs to improve coverage in such weak spots. Interference in the wireless medium by components under the operator's control can be configured to mitigate their effect. Strategies for doing so include configuring non-overlapping frequency bands for nearby eNodeBs and cell sectors or through radio spectrum shaping [28]. These improvements will also impact MIMO transmissions by making it easier for UEs to detect multiple antennas. At the same time, understanding how other parts of the cellular network such as resource allocation and transmission scheduling algorithms react to changing radio conditions can also help.

Radio resource management: The decisions made by eNodeBs when managing radio resources are influenced by many factors such as the scheduling algorithm used for resource allocation and strategies for handling UE handovers. Cellular network operators can tweak their proprietary scheduling algorithm (not defined in the 3GPP standard) to improve QoE for UEs that are streaming video. The cellular network operator may also choose to employ resource management schemes that directly take into account video streaming performance and prevent degradation [29]. Specifically, if the cellular network is able to determine that a particular LTE connection is being used for video steaming, then it can choose to prioritize that connection over others for resource allocation.

V. Related Work

Monitoring Video QoE: Prior research has developed methods to help network operators monitor video QoE metrics. Earlier work used DPI to analyze HTTP headers for estimating video QoE of non-encrypted video traffic [5], [6]. As popular video streaming services have deployed encryption [30], [31], it is no longer possible for network operators to extract HTTP headers using DPI. More recently, researchers proposed machine learning based methods to monitor QoE for encrypted video traffic (using HTTPS and QUIC) using plain-text packet header information at the transport and network layers [7], [8], [9], [32]. This line of research focuses on estimating video QoE metrics from a network operator's perspective, but does not provide any insights to network operators about the potential root causes of QoE impairments. To address this gap, our work uses machine learning to diagnose the root causes of video QoE impairments in operational cellular networks.

Cellular Performance Diagnosis: Prior work has attempted to diagnose performance issues in cellular networks. Most notably, Iyer et al. [21] focused on identifying cellular RAN characteristics that lead to performance issues such as call drops. In contrast to our work, they does not consider the impact of cellular network characteristics on application-specific QoE metrics.

Cellular QoE Diagnosis: Balachandran et al. [23] analyzed 3G cellular RAN characteristics such as RSSI, handovers, and RRC failures to determine their impact on web QoE, which was measured in terms of partial HTTP downloads. The evaluation showed that higher network load and more handovers generally caused web QoE impairments. Shafiq et al. [22] studied the impact of 3G cellular RAN and CN characteristics on user engagement of non-adaptive video streaming over HTTP. They showed that RAN conditions such as higher load and more radio interference led to reduced user engagement. In contrast to this work, we focus on diagnosing QoE impairments for *encrypted* and *adaptive* video streaming in LTE cellular networks.

Video QoE Diagnosis at End-points: Prior work has also studied the root causes of QoE impairments at end-points (i.e, client video player and server). Huang et al. [33] showed that inaccurate bandwidth measurements used by bitrate adaption algorithms at end-points lead to video QoE impairments. Jiang et al. [34] used client-side measurements to shed light on video QoE impairments in the wild. Xu et al. [11] investigated how design choices at video client (e.g., buffer threshold in the startup phase) or server (e.g., available video bitrates) can cause cellular video QoE impairments. In contrast to these efforts, our work considers end-points as external factors that are beyond the control of cellular network operators. We only target RAN and CN characteristics that cellular network operators can influence to mitigate video QoE impairments.

VI. Conclusion

In this paper, we proposed a machine learning based approach to identify the root causes of video QoE impairments in cellular networks. We extracted a variety of features that capture different radio and core network characteristics. We showed how simpler analysis techniques such as feature correlation may provide misleading conclusions, representing effect rather than the underlying root cause or in some cases counterintuitive results. We argued that this is due to the manual nature of these analysis techniques, as well as their lack of accounting for interdependencies between various characteristics of the cellular network. To address these issues, we proposed a machine learning approach that leverages 668 radio and core network features to identify root causes of video QoE impairments. We deployed our approach in the wild on a tier-1 LTE cellular network in the United States. The key advantage of our machine learning approach lie in its systematic and automated application to pinpoint the underlying root causes of video QoE impairments while considering their interdependencies.

References

[1] "AT&T Named the Nationś Best Network for Streaming Video," http://about.att.com/newsroom/best_network_for_streaming_video.html, AT&T, Feb 2018.

[2] J. Erman, A. Gerber, K. K. Ramadrishnan, S. Sen, and O. Spatscheck, "Over the Top Video: The Gorilla in Cellular Networks," in *ACM Internet Measurement Conference (IMC)*, 2011.

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

[4] "App Annie. Video Streaming Apps: The Real Emmy Winners," https://www.appannie.com/en/insights/reports/video-streaming-apps-emmys/.

[5] R. Schatz, T. Hoßfeld, and P. Casas, "Passive YouTube QoE Monitoring for ISPs," in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, 2012.

[6] V. Aggarwal, E. Halepovic, J. Pang, S. Venkataraman, and H. Yan, "Prometheus: Toward Quality-of-experience Estimation for Mobile Apps from Passive Network Measurements," in *ACM HotMobile*, 2014.

[7] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki, "Measuring Video QoE from Encrypted Traffic," in *ACM Internet Measurement Conference (IMC)*, 2016.

[8] I. Orsolic, D. Pevec, M. Suznjevic, and L. Skorin-Kapov, "A machine learning approach to classifying YouTube QoE based on encrypted network traffic," *Multimedia Tools and Applications*, 2017.

[9] M. H. Mazhar and Z. Shafiq, "Real-time Video Quality of Experience Monitoring for HTTPS and QUIC," in *IEEE INFOCOM*, 2018.

[10] M. Katsarakis, R. C. Teixeira, M. Papadopouli, and V. Christophides, "Towards a Causal Analysis of Video QoE from Network and Application QoS," in *ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE)*, 2016.

[11] S. Xu, S. Sen, Z. M. Mao, and Y. Jia, "Dissecting VOD Services for Cellular: Performance, Root Causes and Best Practices," in *ACM Internet Measurement Conference (IMC)*, 2017.

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

[13] ITU-T, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport," Telecommunication standardization sector of ITU, 2016.

[14] ArsTechnica, "T-Mobile dominates spectrum auction, will boost LTE Network across US," https://arstechnica.com/information-technology/2017/04/t-mobile-dominates-spectrum-auction-will-boost-lte-network-across-us/, April 2017.

[15] I. Abdalla and S. Venkatesan, "A QoE preserving M2M-aware hybrid scheduler for LTE uplink," in *IEEE MoWNeT*, 2013.

[16] P. Casas, R. Schatz, and T. Hoßfeld, "Monitoring YouTube QoE: Is Your Mobile Network Delivering the Right Experience to your Customers?" in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.

[17] X. Chen, H. Yi, H. Luo, H. Yu, and H. Wang, "A novel CQI calculation scheme in LTE/LTE-A systems," in *IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, 2011.

[18] nTh Network Technologies, "Highsee," https://www.nth-people.com/highsee.html, 01 2019.

[19] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[20] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[21] A. P. Iyer, L. E. Li, and I. Stoica, "Automating Diagnosis of Cellular Radio Access Network Problems," in *ACM MobiCom*, 2017.

[22] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Understanding the Impact of Network Dynamics on Mobile Video User Engagement," in *ACM SIGMETRICS*, 2014.

[23] A. Balachandran, V. Aggarwal, E. Halepovic, J. Pang, S. Seshan, S. Venkataraman, and H. Yan, "Modeling Web Quality-of-experience on Cellular Networks," in *ACM MobiCom*, 2014.

[24] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "A Quest for an Internet Video Quality-of-experience Metric," in *ACM Workshop on Hot Topics in Networks (HotNets-XI)*, 2012.

[25] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communications*, 1999.

[26] Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A First Look at Cellular Network Performance during Crowded Events," in *ACM SIGMETRICS*, 2013.

[27] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi, "LTE-Advanced: Heterogeneous networks," in *European Wireless Conference (EW)*, 2010.

[28] J. Abdoli, M. Jia, and J. Ma, "Filtered OFDM: A new waveform for future wireless systems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015.

[29] A. E. Essaili, D. Schroeder, E. Steinbach, D. Staehle, and M. Shehada, "QoE-Based Traffic and Resource Management for Adaptive HTTP Video Delivery in LTE," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015.

[30] Google, "Adding YouTube and Calendar to the HTTPS Transparency Report," https://security.googleblog.com/2016/08/adding-youtube-and-calendar-to-https.html, Google, August 2016.

[31] Netflix, "Protecting Netflix Viewing Privacy at Scale," https://medium.com/netflix-techblog/protecting-netflix-viewing-privacy-at-scale-39c675d88f45, August 2016.

[32] S. Wassermann, N. Wehner, and P. Casas, "Machine learning models for youtube qoe and user engagement prediction in smartphones," *ACM SIGMETRICS Performance Evaluation Review*, 2019.

[33] T.-Y. Huang, N. Handigol, B. Heller, N. McKeown, and R. Johari, "Confused, Timid, and Unstable: Picking a Video Streaming Rate is Hard," in *ACM Internet Measurement Conference (IMC)*, 2012.

[34] J. Jiang, V. Sekar, I. Stoica, and H. Zhang, "Shedding Light on the Structure of Internet Video Quality Problems in the Wild," in *ACM CoNEXT*, 2013.