VILNIUS TECH
Vilniaus Gedimino technikos universitetas

Darius CHMIELIAUSKAS

QUALITY OF EXPERIENCE EVALUATION AND ENHANCEMENT OF VIDEO STREAMING SERVICES IN CELLULAR NETWORKS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES
ELECTRICAL AND ELECTRONIC ENGINEERING (T 001)

Vilnius, 2026

2026-024-M

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Darius CHMIELIAUSKAS

QUALITY OF EXPERIENCE EVALUATION AND ENHANCEMENT OF VIDEO STREAMING SERVICES IN CELLULAR NETWORKS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
ELECTRICAL AND ELECTRONIC ENGINEERING (T 001)

Vilnius, 2026

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2020–2026.

Supervisor

Prof. Dr Šarūnas PAULIKAS (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001).

The Dissertation Defense Council of the Scientific Field of Electrical and Electronic Engineering of Vilnius Gediminas Technical University:

Chairman

Prof. Dr Darius PLONIS (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001).

Members:

Dr Rimvydas ALEKSIEJŪNAS (Vilnius University, Physics – N 002),
Prof. Dr Vaidotas BARZDĖNAS (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001),
Prof. Dr Algirdas BAŠKYS (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001),
Dr Habil. Jakub NIKONOWICZ (Poznań University of Technology, Poland, Electrical and Electronic Engineering – T 001).

The dissertation will be defended at the public meeting of the Dissertation Defence Council of Electrical and Electronic Engineering in the *Aula Doctoralis* Meeting Hall of Vilnius Gediminas Technical University at **10:00 a. m. on 2 June 2026**.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel. +370 5 274 4956; fax +370 5 270 0112; e-mail: doktor@vilniustech.lt

A notification on the intended defense of the dissertation was sent on 30 April 2026.
A copy of the doctoral dissertation is available for review at the Vilnius Gediminas Technical University repository https://etalpykla.vilniustech.lt, the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania), and the Wroblewski Library of the Lithuanian Academy of Sciences (Žygimantų g. 1, LT-01102 Vilnius, Lithuania).

Vilnius Gediminas Technical University book No. 2026-024-M
https://doi.org/10.20334/2026-024-M

darius.chmieliauskas@vilniustech.lt

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Darius CHMIELIAUSKAS

VAIZDO PERDAVIMO PASLAUGŲ PATIRTIES KOKYBĖS VERTINIMAS IR GERINIMAS JUDRIOJO RYŠIO TINKLUOSE

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
ELEKTROS IR ELEKTRONIKOS INŽINERIJA (T 001)

Vilnius, 2026

Disertacija rengta 2020–2026 metais Vilniaus Gedimino technikos universitete.

Vadovas

prof. dr. Šarūnas PAULIKAS (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001).

Vilniaus Gedimino technikos universiteto Elektros ir elektronikos inžinerijos mokslo krypties disertacijos gynimo taryba:

Pirmininkas

prof. dr. Darius PLONIS (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001).

Nariai:

dr. Rimvydas ALEKSIEJŪNAS (Vilniaus universitetas, Fizika – N 002),
prof. dr. Vaidotas BARZDĖNAS (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001),
prof. dr. Algirdas BAŠKYS (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001),
habil. dr. Jakub NIKONOWICZ (Poznanės technologijos universitetas, Lenkija, Elektros ir elektronikos inžinerija – T 001).

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje **2026 m. birželio 2 d. 10 val.** Vilniaus Gedimino technikos universiteto *Aula Doctoralis* posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel. +370 5 274 4956; faksas +370 5 270 0112; el. paštas *doktor@vilniustech.lt*

Pranešimai apie numatomą ginti disertaciją išsiųsti 2026 m. balandžio 30 d.

Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto talpykloje http://etalpykla.vilniustech.lt/, Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva) bei Lietuvos mokslų akademijos Vrublevskių bibliotekoje (Žygimantų g. 1, LT-01102 Vilnius, Lietuva).

Abstract

This doctoral dissertation investigates the evaluation and enhancement of Quality of Experience (QoE) in 4G and 5G cellular networks. This research belongs to the field of Electrical and Electronics Engineering, specifically communications engineering, signal processing, and network systems management. It addresses the key challenges faced by mobile network operators (MNOs) in managing the increasing demand for multimedia streaming services.

The dissertation consists of an introduction, four chapters, and the general conclusions. The First Chapter introduces the definition of QoE in cellular networks and reviews data sources for QoE evaluation. It surveys subjective, objective, and parametric QoE evaluation methods, together with transport-layer indicators and Radio Access Network (RAN) metrics. Building on these insights, the chapter identifies the lack of an operator-side framework that fuses multi-layer Key Performance Indicators (KPIs) to model, predict, and optimize QoE, with particular emphasis on practical evaluation for uplink real-time services. The Second Chapter focuses on encrypted traffic classification. The chapter proposes a novel approach that analyzes bitstream shapes using Convolutional Neural Networks (CNN), thereby avoiding the complexities of traditional deep packet inspection. The Third Chapter proposes a comprehensive QoE evaluation methodology that integrates multiple MNO data sources, including network-level KPIs, user flows, and application metrics. It applies this pipeline to real-world video streaming scenarios, analyzing key factors such as bandwidth, Radio Frequency (RF) conditions, handovers, and congestion. The performance of streaming protocols such as Real-Time Messaging Protocol (RTMP) and Web Real-Time Communication (WebRTC) is evaluated under diverse conditions, revealing correlations between RF measurements, network KPIs, and objective QoE. The Fourth Chapter introduces methods to enhance QoE in cellular networks through three main approaches: traffic forecasting, antenna parameter selection, and configuration tuning.

This research presents a practical, data-driven framework for MNOs to proactively manage and enhance QoE, a significant contribution to both academic and industrial fields of cellular network management.

The main results of the dissertation were published in eight scientific publications. Three were published in peer-reviewed journals, including two indexed in Clarivate Analytics Web of Science and with an impact factor. Additionally, the research results were presented at five scientific conferences with indexed proceedings.

Reziumė

Disertacijoje tiriama vartotojo patirties kokybė ir jos gerinimas 4G LTE ir 5G NR judriojo ryšio tinkluose. Šie tyrimai priklauso elektros ir elektronikos inžinerijos mokslo krypčiai ir apima ryšių inžineriją, signalų apdorojimą bei tinklo sistemų valdymą. Disertacijoje sprendžiamos pagrindinės problemos, su kuriomis susiduria mobiliojo ryšio tinklų operatoriai, valdydami augančią daugialypės terpės srautinių paslaugų paklausą.

Disertaciją sudaro įvadas, keturi skyriai ir bendrosios išvados. Pirmame skyriuje apibrėžiama patirties kokybės sąvoka judriojo ryšio tinkluose ir apžvelgiami duomenų šaltiniai, taikomi patirties kokybei vertinti. Taip pat apžvelgiami subjektyvūs, objektyvūs ir parametriniai patirties kokybės vertinimo metodai, transporto lygmens rodikliai ir radijo prieigos tinklo rodikliai. Apibendrinant šį skyrių nustatoma problema, kad tinklo operatoriams stinga integruotos sistemos (angl. *framework*), kuri apimtų kelių tinklo lygmenų rodiklius ir galėtų modeliuoti, prognozuoti bei gerinti klientų patirtą kokybę realaus laiko (angl. *real-time*) vaizdo perdavimo metu. Antrasis skyrius skirtas šifruoto duomenų srauto klasifikavimui. Jame siūlomas metodas, kuris analizuoja bitų srauto formas naudojant konvoliucinius neuroninius tinklus, užuot taikius tradicinį paketų tikrinimo metodą. Trečiame skyriuje siūloma išsami patirties kokybės vertinimo sistema, integruojanti kelis operatoriui pasiekiamus duomenų šaltinius, įskaitant tinklo kokybės rodiklius, vartotojų duomenų srautus ir taikomojo lygmens metrikas. Ši sistema taikoma srautinio vaizdo transliavimo paslaugoms. Analizuojami tokie pagrindiniai veiksniai kaip tinklo pralaidumas, radijo signalo kokybė, perjungimai (angl. *handover*) ir tinklo celių apkrova. Taip pat vertinamas srautinio transliavimo protokolų, tokių kaip realaus laiko pranešimų perdavimo protokolas (RTMP) ir realaus laiko komunikavimas žiniatinklyje (WebRTC), veikimas įvairiomis ryšio sąlygomis, ieškant koreliacijos tarp radijo ryšio signalo, tinklo elementų kokybės rodiklių ir objektyviai vertinamos patirties kokybės. Ketvirtame skyriuje pristatomi patirties kokybės gerinimo metodai korinio ryšio tinkluose, naudojant tris būdus: srauto prognozavimą, antenų išdėstymo parinkimą ir parametrų derinimą.

Pagrindiniai disertacijos rezultatai paskelbti aštuoniose mokslinėse publikacijose. Trys iš jų – recenzuojamuose žurnaluose, dvi iš jų indeksuotos *Clarivate Analytics Web of Science* duomenų bazėje ir turi cituojamumo rodiklį. Be to, tyrimo rezultatai pristatyti penkiose mokslinėse konferencijose, kurių pranešimų rinkiniai įtraukti į indeksuojamas duomenų bazes.

Notations

Symbols

$f^{(j)}$ – Scaling factor for carrier j (liet. *nešlio j mastelio koeficientas*);

J – Number of aggregated component carriers (liet. *agreguotų komponentinių nešlių skaičius*);

$N_{\mathrm{PRB}}^{(j,\mu)}$ – Number of PRBs for carrier j and numerology μ (liet. *fizinių resursų blokų skaičius pagal nešlį j ir numerologiją* μ);

$\mathrm{OH}^{(j)}$ – Protocol overhead for carrier j (liet. *protokolo tarnybinė informacija nešliui j*);

$Q_m^{(j)}$ – Maximum modulation order for carrier j (liet. *didžiausias moduliacijos laipsnis nešliui j*);

R – Data rate, in Mbit/s (liet. *duomenų sparta, Mbit/s*);

R_{max} – Maximum coding rate (liet. *didžiausias kodavimo santykis*);

$T_{\mathrm{s}}^{(\mu)}$ – OFDM symbol duration for numerology μ (liet. *OFDM simbolio trukmė numerologijai* μ);

$v_{\mathrm{layers}}^{(j)}$ – Maximum number of MIMO layers for carrier j (liet. *didžiausias MIMO sluoksnių skaičius nešliui j*).

Abbreviations

3GPP – 3rd Generation Partnership Project (liet. *Trečiosios kartos partnerystės projektas*);

4G – 4th Generation Mobile System (liet. *ketvirtosios kartos mobilioji ryšio sistema*);

5G – 5th Generation Mobile System (liet. *penktosios kartos mobilioji ryšio sistema*);

5QI – 5G QoS Identifier (liet. *5G paslaugos kokybės identifikatorius*);

AVC – Advanced Video Coding (liet. *pažangusis vaizdo kodavimas*);

CCO – Coverage and Capacity Optimization (liet. *aprėpties ir talpos optimizavimas*);

CDN – Content Delivery Network (liet. *turinio pristatymo tinklas*);

CNN – Convolutional Neural Network (liet. *konvoliucinis neuroninis tinklas*);

CQI – Channel Quality Indicator (liet. *kanalo kokybės rodiklis*);

DASH – Dynamic Adaptive Streaming over HTTP (liet. *dinaminis adaptyvusis srautinis perdavimas per HTTP*);

DPI – Deep Packet Inspection (liet. *gilioji paketų analiz*);

E2E – End-to-End (liet. *perdavimas nuo galo iki galo*);

eMBB – Enhanced Mobile Broadband (liet. *patobulintas mobilusis plačiajuostis ryšys*);

FDD – Frequency Division Duplex (liet. *dažninio padalijimo dupleksas*);

GBR – Guaranteed Bit Rate (liet. *garantuotoji bitų sparta*);

GNSS – Global Navigation Satellite System (liet. *Pasaulinė palydovinės navigacijos sistema*);

HARQ – Hybrid Automatic Repeat Request (liet. *hibridinė automatinė pakartotinio siuntimo užklausa*);

HAS – HTTP Adaptive Streaming (liet. *adaptyvusis srautinis perdavimas per HTTP*);

HEVC – High Efficiency Video Coding (liet. *didelio efektyvumo vaizdo kodavimas*);

HLS – HTTP Live Streaming (liet. *tiesioginis srautinis perdavimas per HTTP*);

KPI – Key Performance Indicator (liet. *pagrindinis veiklos rodiklis*);

KQI – Key Quality Indicator (liet. *pagrindinis kokybės rodiklis*);

LTE – Long Term Evolution (liet. *ilgalaikė evoliucija*);

MAC – Medium Access Control (liet. *prieigos prie perdavimo terpės valdymo lygmuo*);

MCS – Modulation and Coding Scheme (liet. *moduliacijos ir kodavimo schema*);

MDT – Minimization of Drive Tests (liet. *matavimų važiuojant mažinimas*);

MIMO – Multiple-Input Multiple-Output (liet. *kelių įėjimų ir išėjimų sistema*);

ML – Machine Learning (liet. *mašininis mokymasis*);

mMTC – Massive Machine Type Communications (liet. *masinis mašininio tipo ryšys*);

MNO – Mobile Network Operator (liet. *mobiliojo / judriojo ryšio operatorius*);

MOS – Mean Opinion Score (liet. *vidutinis subjektyvusis įvertis*);

MSE – Mean Squared Error (liet. *vidutinė kvadratinė paklaida*);

NR – New Radio (liet. *naujasis radijas*);

NWDAF – Network Data Analytics Function (liet. *tinklo duomenų analitikos funkcija*);

O-RAN – Open Radio Access Network (liet. *atvirasis radijo prieigos tinklas*);

OCR – Optical Character Recognition (liet. *optinis simbolių atpažinimas*);

OTT – Over-the-Top (liet. *internetu teikiamos paslaugos*);

PDB – Packet Delay Budget (liet. *paketų delsos biudžetas*);

PDCP – Packet Data Convergence Protocol (liet. *paketinių duomenų konvergencijos protokolas*);

PER – Packet Error Rate (liet. *paketų klaidų dažnis*);

PHY – Physical Layer (liet. *fizinis lygmuo*);

PRB – Physical Resource Block (liet. *fizinis išteklių blokas*);

PSNR – Peak Signal-to-Noise Ratio (liet. *pikinis signalo ir triukšmo santykis*);

QoE – Quality of Experience (liet. *patirties kokybė*);

QoS – Quality of Service (liet. *paslaugų kokybė*);

QUIC – Quick UDP Internet Connections (liet. *spartusis UDP pagrindu veikiantis perdavimo protokolas*);

RAN – Radio Access Network (liet. *radijo prieigos tinklas*);

RIC – RAN Intelligent Controller (liet. *išmanusis radijo prieigos tinklo valdiklis*);

RL – Reinforcement Learning (liet. *skatinamasis mokymasis*);

RLC – Radio Link Control (liet. *radijo sąsajos valdymas*);

RRC – Radio Resource Control (liet. *radijo išteklių valdymas*);

RSRP – Reference Signal Received Power (liet. *etaloninio signalo priėmimo galia*);

RSRQ – Reference Signal Received Quality (liet. *etaloninio signalo priėmimo kokybė*);

RTMP – Real-Time Messaging Protocol (liet. *realaus laiko pranešimų perdavimo protokolas*);

RTP – Real-Time Transport Protocol (liet. *realaus laiko transporto protokolas*);

RTT – Round-Trip Time (liet. *dvikryptis vėlinimas*);

SINR – Signal-to-Interference-plus-Noise Ratio (liet. *signalo ir trukdžių bei triukšmo santykis*);

SSIM – Structural Similarity Index (liet. *struktūrinio panašumo indeksas*);

SVR – Support Vector Regressor (liet. *atraminių vektorių regresorius*);

TBS – Transport Block Size (liet. *transporto bloko dydis*);

TCP – Transmission Control Protocol (liet. *perdavimo valdymo protokolas*);

TDD – Time Division Duplex (liet. *laiko padalijimo dvipusis ryšys*);

UDP – User Datagram Protocol (liet. *naudotojo datagramų protokolas*);

UE – User Equipment (liet. *naudotojo įrenginys*);

URLLC – Ultra-Reliable Low-Latency Communications (liet. *itin patikimas mažos delsos ryšys*);

VMAF – Video Multimethod Assessment Fusion (liet. *daugiametodžio vaizdo kokybės vertinimo metrika*);

VVC – Versatile Video Coding (liet. *universalusis vaizdo kodavimas*);

WebRTC – Web Real-Time Communication (liet. *realaus laiko komunikavimas žiniatinklyje*).

Contents

Introduction

Problem Formulation

Mobile data traffic is growing steadily at an annual rate greater than 20%, with video traffic accounting for 76% of total mobile data (Ericsson, 2025). Despite increased mobile network bandwidth with the evolution of 5G, ensuring high Quality of Experience (QoE) remains a significant challenge for Mobile Network Operators (MNOs). It is especially significant for video streaming services. High-quality video streaming is both bandwidth-demanding and delay-sensitive.

This dissertation addresses the problem of QoE evaluation and management by proposing a framework that considers modern network complexities. Achieving this goal requires addressing several interconnected issues. The first issue is limited service visibility due to traffic encryption. Once a service is identified, the next challenge is to assess the QoE of video streams. The wireless channel is the last mile for delivering data, video, and voice services. In the wireless channel, the user's achievable throughput and delay fluctuate due to: (a) varying Radio Frequency (RF) signal quality, (b) contention for the shared medium with other users, and (c) the unknown delay and bandwidth demand of the current session. The stochastic RF environment and the random access behavior of users result in unpredictable Quality of Service (QoS), which leads to an even more obscure QoE evaluation. Although downlink streaming has been extensively investigated, the growing relevance of uplink-intensive applications, such as live content creation and remote control, requires dedicated evaluation models. Finally, a high QoE

depends on network coverage and performance. Therefore, achieving this goal requires shifting from equipment performance tuning to user-centric, QoE-oriented network management. The dissertation is situated within the field of Electrical and Electronics Engineering, specifically at the intersection of communications engineering, signal processing, and network systems management. The research involves reliably recognizing services within encrypted traffic, inferring and forecasting QoE for uplink real-time streaming, and enhancing RAN configuration and coverage to maximize QoE under stochastic radio conditions.

Relevance of the Dissertation

The strategic importance of mobile network services is growing rapidly, driven by ubiquitous multimedia streaming for entertainment and by industrial applications such as live journalism, influencer content creation, and remote-control operations in smart-city infrastructure. Accurate QoE evaluation is essential to prevent service churn, financial losses, and critical safety risks. Academia and the telecommunications industry are actively developing 5G and beyond networks with architectural innovations. These include network slicing, the Open RAN (O-RAN) architecture with its RAN Intelligent Controller (RIC) for QoE-oriented RAN optimization, and the Network Data Analytics Function (NWDAF) for QoE management.

Research Object

The object of this research is operator-side evaluation and management of objective QoE for video streaming services in 4G LTE and 5G NR cellular networks.

Aim of the Dissertation

The dissertation aims to develop and validate a network-side framework that enables MNOs to recognize services, evaluate and forecast QoE, and support proactive QoE management for video streaming services.

Tasks of the Dissertation

To achieve the aim of the dissertation, the following research tasks are defined:

1. Analyze the end-to-end data delivery path in 4G LTE and 5G NR networks, identifying which radio, transport, and application layer metrics are observable to the MNO.

2. Collect and classify MNO-accessible data sources (radio KPIs, transport protocol metrics, and media-layer data) and establish their relationship to QoE definitions for video streaming services.

3. Develop a lightweight and encryption-resilient algorithm for recognizing video streaming services based solely on network-side data.

4. Design and validate a model that maps RAN and IP metrics to objective QoE estimates for uplink real-time video streaming.

5. Develop and evaluate machine learning and reinforcement learning models to predict RAN capacity and support QoE-driven RAN parameter adaptation.

Research Methodology

Research methodology combines a systematic literature review of 4G LTE and 5G NR network architectures and QoE evaluation methods for video streaming services. To validate the hypotheses, comprehensive data acquisition systems were designed and implemented, integrating commercial cellular networks, UE-side and network-side measurements, and full-reference media-quality assessment. Large-scale field measurements were conducted in live LTE and 5G environments for service recognition, objective QoE regression, traffic forecasting, and QoE-oriented RAN enhancement. Beyond the data collection, comprehensive data engineering was performed. It includes Android-based data collection, Wireshark/pcap analysis, OBS Studio and RTC Studio for stream generation, FFmpeg/ffprobe and OCR-assisted frame alignment, Python-based machine-learning workflows using Keras, scikit-learn, and Prophet, as well as MATLAB-based QuaDRiGa simulations and a Gym-based RL environment.

Scientific Novelty of the Dissertation

This dissertation presents the following scientific contributions:

1. A unified operator-side framework is developed for LTE and 5G NR networks to perform objective QoE evaluation based on RAN and IP transport-layer metrics.

2. A lightweight traffic classification method using a CNN is proposed, which analyzes the bitstream shape as a time series to recognize video services, achieving accuracy on par with complex DPI-based methods.

3. Regression models are developed to predict objective QoE for real-time uplink video streaming directly from observable RAN air interface metrics.

4. Network management methods are introduced, such as a Prophet-based traffic and congestion prediction algorithm and a Q-learning-based dynamic antenna azimuth selection algorithm.

Practical Value of the Research Findings

A formalized relationship between RAN metrics and objective QoE for real-time video services is established, enabling automated QoE-oriented network management in operational LTE, 5G NR, and future mobile networks. In addition, a traffic-shape-based method is developed that operates solely on 1 s byte sequences, enabling MNO-side service recognition with low computational complexity and privacy preservation. The proposed models are trained and validated on large-scale measurements collected from live LTE and 5G networks rather than simulations, and the datasets and source code are openly available for further research.

Defended Statements

1. Objective QoE for uplink real-time video streaming, expressed by VMAF, can be predicted from operator-observable path loss, MCS, and allocated PRBs, with R^2 values exceeding 0.79.

2. Video-streaming traffic can be recognized from aggregated 1 s byte-count traces using a convolutional neural network without packet inspection, achieving classification accuracy above 0.9.

3. Busy-hour cell traffic and congestion can be forecast up to 28 days ahead using a modified two-stage Prophet-based approach, with forecasting performance reaching R^2 values up to 0.89.

4. Encrypted video-streaming service recognition and objective QoE forecasting can be achieved using an operator-side QoE-centric framework based solely on MNO-observable RAN measurements and IP transport metrics.

Approval of the Research Findings

The main results of this dissertation have been published in eight scientific papers: two papers in foreign journals indexed in the *Web of Science* database with a citation index (Chmieliauskas & Paulikas, 2025; Chmieliauskas & Paulikas, 2023), one paper in a local peer-reviewed journal (Chmieliauskas, 2021), and five papers in international conference proceedings. Additionally, the research results were presented at the following conferences:

- Open Conference of Electrical, Electronic, and Information Sciences (eStream), 2019, Vilnius, Lithuania.
- IEEE Open Conference of Electrical, Electronic, and Information Sciences (eStream), 2021, Vilnius, Lithuania.
- IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), 2022, Vilnius, Lithuania.
- Workshop on Microwave Theory and Techniques in Wireless Communications (MTTW), 2023, Riga, Latvia.
- IEEE 11th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), 2024, Valmiera, Latvia.

Structure of the Dissertation

This dissertation consists of four main chapters, general conclusions, references, a list of the author's publications, and a summary in Lithuanian. The dissertation is 152 pages, including 54 figures, 26 equations, and 24 tables. Additionally, it includes 142 references.

Acknowledgments

For the one and only, thank you for your patience.

1

General Introduction to Quality of Experience Evaluation in Cellular Networks

QoE in cellular networks is a critical measure of service performance and user satisfaction. Although the general definition of QoE remains ambiguous and open to interpretation, evaluating it from the MNO side is even more challenging because operators have limited direct visibility into user-perceived quality. Therefore, any QoE evaluation in cellular networks should begin with a clear and consistent definition of QoE. This chapter provides a comprehensive overview of QoE evaluation in mobile networks, focusing on cellular network architecture, protocol stacks, and data collection strategies, and explaining their roles in accurate QoE modeling and prediction. This chapter builds on concepts published in (Chmieliauskas & Paulikas, 2025) and (Chmieliauskas & Paulikas, 2023).

1.1. Background on Data Sources in Cellular Networks

According to Schatz et al. (2018), one of the most comprehensive and widely used definitions of QoE comes from the EU Qualinet community (COST Action IC1003: European Network on QoE in Multimedia Systems and Services): "Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in light of the user's personality and current state. In the context of communication services, QoE is influenced by service, content, device, application, and context of use" (Brunnström et al., 2013). To align with the definition of QoE and evaluate it within MNO infrastructure, it is essential to understand the network architecture and service delivery mechanisms. The following key components form the foundation for accurate QoE modeling and prediction in cellular networks:

- Cellular Network Architecture and Protocol Stack: understanding the data path and the behaviors of layers and interfaces.
- QoE Evaluation Methods: assessing the applicability of QoE models to the target service type.
- Service-Type Recognition: identifying encrypted traffic classes for selecting appropriate QoE models.
- Network Coverage and Performance: coverage conditions and capacity factors that drive QoS-to-QoE mapping and assurance.

The first step in evaluating QoE from the MNO perspective is to understand what, how, and where data can be collected. Objective end-to-end QoE for a cellular service or application can be evaluated using data from the user equipment (UE) application and corresponding server-side information. Only these endpoints have access to the original data and can track how it changes during transmission. However, MNOs typically do not own application servers on the Internet, as customers use Over-the-Top (OTT) applications provided by third-party Internet companies. In such scenarios, the data stream passes through the MNO's network elements and interfaces, which are controlled by the signaling and control planes. Accurate information about network and data flows enables effective modeling, evaluation, and prediction of QoE for cellular network users. Therefore, a detailed examination of the cellular network architecture and its protocol stacks is essential for identifying the data-collection points that enable accurate QoE modeling from the MNO perspective.

1.2. Cellular Network Architecture and Protocol Stack

Most commercial cellular networks use the 5G Non-Standalone (NSA) architecture, in which the new radio (NR) air interface is used exclusively for data transfer. Control signaling between the UE and the RAN is managed by LTE (Liu et al., 2020). This approach enables faster 5G rollout without requiring the deployment of a completely new 5G Core Network (5GC). Figure 1.1 illustrates the architecture of a 5G NSA network, from the UE application on one end to the Content Delivery Network (CDN) and cloud server on the other.

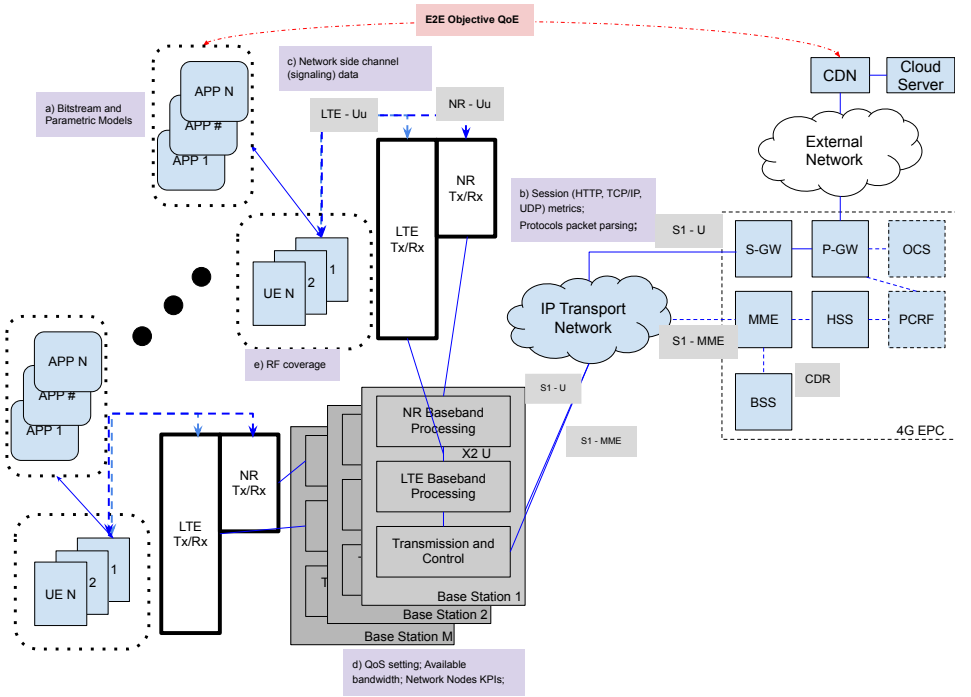


Fig. 1.1. Key data collection points for QoE modeling in a 5G NSA architecture (created by the author)

As shown in Figure 1.1, the 5G NSA network consists of multiple base stations, numbering from several thousand to tens of thousands, depending on the network scale. These base stations collectively serve hundreds of thousands or even millions of user devices. User data generated by various applications is transmitted to and from the UE via the LTE and NR air interfaces. The LTE and NR air interfaces, labeled Uu, handle both data-payload delivery and the control of UE

behavior when connected to the network. Base stations are connected to the core network elements through the IP transport network.

End-to-end (E2E) data transmission involves interaction between the application on the UE and the server or CDN at the remote end. This data delivery is managed using IP stack protocols such as HTTP, TCP/IP, and UDP. For real-time video streaming, a variety of media-layer protocols are available. RTMP and WebRTC are selected as representative examples of TCP- and UDP-based transport. Beyond the IP, the user data payload is encapsulated in the 5G and 4G protocol stack. Figure 1.2 illustrates the transmission of media protocols (RTMP over TCP and WebRTC over UDP) within the 5G NSA network.

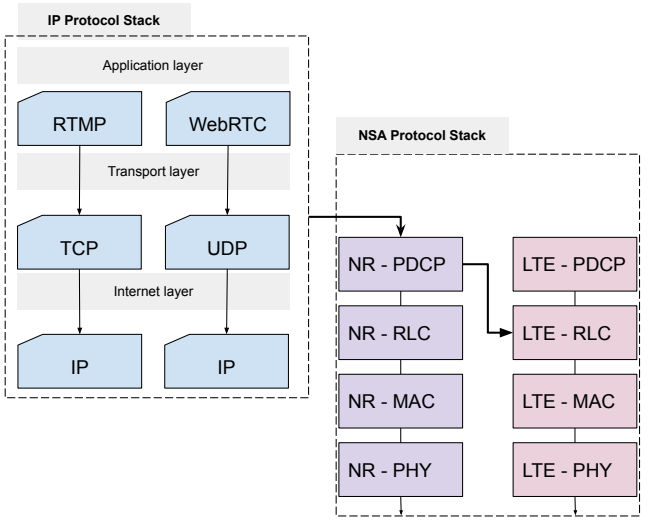


Fig. 1.2. Data delivery using the IP protocol stack (e.g., TCP-based RTMP and UDP-based WebRTC) over the 5G NSA cellular network protocol stack (created by the author)

The 3GPP specification TS 37.340 (3rd Generation Partnership Project, 2019) defines Option 3x, the most commonly used scenario in 5G NSA. Although the payload data is exchanged between the server and the user application, it must be transferred to the UE and the base station via the cellular network protocols. First, the payload data is encapsulated in the Packet Data Convergence Protocol (PDCP) on the NR side, then it is distributed across the lower layers in NR and LTE, including the Radio Link Control (RLC), Medium Access Control (MAC), and Physical (PHY) layers. Additionally, the signaling and control information for each client and session is transmitted in parallel.

1.3. Quality of Experience Definition and Available Mobile Network Data

The EU Qualinet QoE definition (Brunnström et al., 2013) highlights the multidimensional nature of QoE, emphasizing that both technical performance and user-specific factors must be considered in the assessment of user satisfaction. When the QoE definition is viewed together with the cellular network architecture and protocol stacks shown in Figures 1.1 and 1.2, it becomes clear that a significant knowledge gap remains in accurately assessing customer satisfaction. This gap arises because the data collected within the network provides only limited insight into the user's QoE, which depends on multiple factors such as the service, content, device, application, and usage context.

Thus, this section discusses how the depth of QoE insight depends on the ability to obtain and analyze data, which affects the accuracy of QoE. To illustrate this relationship, Figure 1.3 shows the inherent trade-off faced by MNOs. This trade-off involves balancing the availability of QoE-related data with the precision of the resulting QoE evaluations. Subjective evaluations provide the most precise QoE assessments, whereas MNOs are often restricted to indirect methods because of data availability and accessibility limitations.

1.3.1. Personal Quality of Experience

The QoE Modeling Pyramid shown in Figure 1.3 illustrates that personal QoE, positioned at the top, provides the most accurate representation. This concept aligns with the ITU definition (International Telecommunication Union, 2008) of QoE as "the overall acceptability of an application or service, as perceived subjectively by the end-user". However, because such a direct evaluation is often unattainable, it becomes necessary to rely on derived metrics.

1.3.2. Subjective Quality of Experience

To quantify subjective QoE, a mapping from perceived experience to a numerical value is necessary. Various methods have been proposed for this purpose, and a comprehensive survey of QoE assessment approaches for network services is provided in (Yang et al., 2018). These methods include binary classification (e.g., "acceptable" vs. "unacceptable") and paired-comparison tests, where users evaluate two stimuli and decide which one is superior. The canonical speech-oriented standard is ITU-T P.800 (08/1996), which defines Absolute Category Rating (ACR), Degradation Category Rating (DCR), and Comparison Category Rating (CCR)

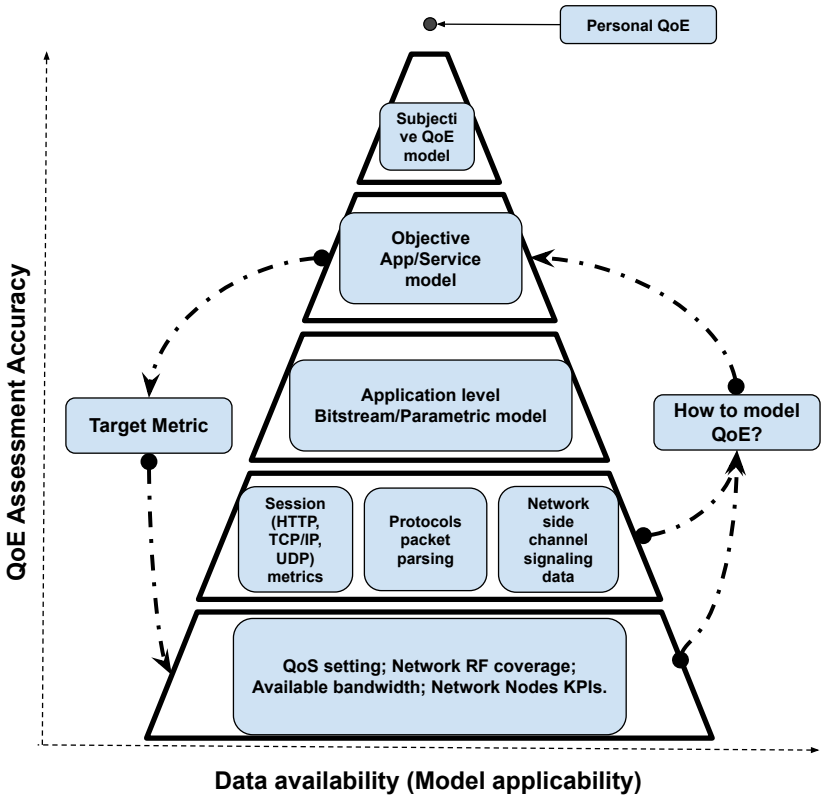


Fig. 1.3. Hierarchical view of QoE modeling approaches in terms of data availability, model applicability, and assessment accuracy (created by the author)

procedures for speech and transmission quality (International Telecommunication Union [ITU-T], 1996). For audiovisual services, the latest ITU-T P.910 (10/2023) extends these concepts, specifying *single-stimulus* tests (ACR, SAMVIQ), *double-stimulus* methods (DSIS, DSCQS), and *continuous-quality* viewing (ITU-T, 1999). Crowdsourced variants (e.g., ITU-T P.808) now complement lab tests, widening participant pools at the cost of stricter screening and outlier control (Naderi & Cutler, 2020).

All methods ultimately yield a five-point *Mean Opinion Score* (MOS). MOS is standardized in the ITU-T recommendations and divides subjective QoE perception into five levels, as shown in Table 1.1.

The MOS can be applied to various types of media content, including voice (Rämö & Toukomaa, 2015), video streaming (Liu et al., 2015), 360° video (Anwar et al., 2020), and online gaming (Metzger et al., 2022).

Table 1.1. Mean Opinion Score (MOS)

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Although subjective tests provide the most accurate representation of user experience, they are expensive and impractical for real-time network management. Consequently, MNOs supplement or replace them with objective or crowdsourced estimation methods calibrated to P.800/P.910 MOS benchmarks (Le Maguer et al., 2024).

1.3.3. Objective Quality of Experience

Although subjective QoE metrics such as MOS are invaluable in capturing users' perceived experiences, they have several notable drawbacks. First, subjective measures such as MOS are heavily dependent on human judgment, making them susceptible to bias and variability. Second, subjective evaluations are typically labor-intensive and time-consuming to conduct, as they involve user surveys and testing panels. This often makes them impractical for large-scale or real-time assessments. Due to these limitations, there is a need for objective QoE metrics that can provide a more consistent, efficient, and detailed evaluation.

These metrics utilize audio, image, and video features to estimate quality. Based on the amount of available source information, they are classified as Full Reference (FR), Reduced Reference, and No-Reference metrics (Kougioumtzidis et al., 2022). FR metrics have access to both the reference and the outcome sequences for direct comparison. In terms of human perception accuracy, FR metrics that perform a frame-by-frame examination between the source and the affected sequence produce more accurate results, compared to RR or NRM. Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) are examples of such metrics. More advanced metrics, such as Video Quality Model (VQM) (Pinson & Wolf, 2004) and the Video Multimethod Assessment Fusion (VMAF) (Li et al., 2016), provide more sophisticated video quality evaluations.

VMAF evaluates the visual quality of compressed video content, considering factors such as compression artifacts, sharpness, and color accuracy. VMAF uses a machine learning-based approach that combines multiple assessment methods, including SSIM, PSNR, and others, to provide a comprehensive and accurate as-

sessment of video quality. Considering various visual characteristics, VMAF aims to align its quality predictions with human perception, making it a valuable tool for video codec development, video optimization, and quality control in streaming services. According to (Rassool, 2017), VMAF is specifically designed to correlate strongly with subjective MOS scores. Using machine learning techniques, a large sample of MOS scores was used as ground truth to train a quality estimation model. VMAF can be used successfully not only for the evaluation of HTTP adaptive streaming (HAS) using the reliable TCP protocol but also for real-time streaming through unreliable UDP (García et al., 2019).

For the audio domain, Perceptual Objective Listening Quality Assessment (Beerends et al., 2013) is an industry-adapted and standardized model. It assesses audio quality using perceptual and signal-based parameters. A joint audio and video objective evaluation model, combining features from both domains, is proposed in (Min et al., 2020)

Objective end-to-end QoE assessment is possible only at the two communication endpoints: the UE application and the corresponding application server. Only these points have access to the original media stream and can directly observe impairments introduced during transmission. Therefore, objective FR metrics require access to the original bitstream and the application multimedia layer.

In operational mobile networks, this prerequisite is rarely satisfied. Most traffic originates from OTT services whose servers are outside the control of the MNO. The operator therefore lacks both the reference at the server side and the decoded payload within the UE application. Under these constraints, the MNO must rely on indirect evidence. Quality estimation is derived from transport- and network-layer observations such as bitstream statistics, TCP/UDP flow dynamics, protocol-layer counters, radio-access-network performance indicators, cell load, coverage, and estimated user location. Using these secondary measurements, the operator can model, evaluate, predict, and even forecast user QoE in the absence of direct access to the underlying media.

1.3.4. Bitstream and Parametric Models

Given the difficulty of accessing original media layer data, bitstream-level and parametric models are extensively studied for HAS. They provide a practical alternative to FR methods. HAS is the dominant delivery format for large content delivery networks such as YouTube, Netflix, and Twitch (Roettgers, 2013a). These models rely on key quality indicators (KQIs) such as initial buffering delay, stalling events, delivered resolution, and smoothness (quality switches) (Seufert et al., 2014; Tang et al., 2018). Among these, initial delay and stalling have been

identified as the most influential factors (Seufert et al., 2014). A comprehensive taxonomy of HAS QoE influence factors, reproduced from (Seufert et al., 2014), is shown in Table 1.2.

Table 1.2. Taxonomy of factors influencing HAS QoE (adapted from (Seufert et al., 2014))

Dimension	Subcategory	Representative factors
Perceptual	Waiting times	Initial delay, stalling frequency, stalling duration
	Video adaptation	Switching frequency, amplitude, and time spent on each layer
	Video quality	Spatial resolution, frame rate, image quality
	Context factors	Device, content, usage
Technical	Server side	Video codec, segment size
	Client side	Video player, adaptation logic, buffering, QoS monitoring
	Adaptation logic	Bitrate estimation, buffer occupancy control
	Entity interaction	Multiple HAS clients, interaction with TCP or other apps

Inferring video streaming QoE occurs at the application layer of the network protocol stack. This analysis benefits MNOs only if they have access to application or media layer information. Such data is collected by third-party OTT service providers. The recent standardization of Server and Network Assisted DASH (SAND) (Barakabitze et al., 2019) marks the beginning of cooperative device-network streaming. However, the adoption of such technologies in commercial networks remains limited. Without collaboration between OTT providers and MNOs, the latter cannot use this data effectively.

These studies are important because they investigate the relationship between user satisfaction or annoyance and KQIs, often assessed using MOS. This has led to the development of parametric QoE evaluation models such as CQM (Tran et al., 2021), the ITU-T standardized P.1203 (Robitza et al., 2018), and the incorporation of machine-learning-based approaches (Barman & Martini, 2019).

P.1203 is an ITU-T parametric model that predicts per-segment and aggregate MOS for HAS. The ITU-T recommendation defines four modes of operation that trade estimation accuracy against the amount of data available to the evaluator (Robitza et al., 2018):

- Mode 0 (metadata-based). Only streaming metadata is required: initial loading delay, stalling events, codec, average bitrate, spatial resolution, and frame rate. No packet or bitstream inspection is necessary, making this mode suitable for analytics platforms that rely exclusively on player-side logs or server manifests.

- Mode 1 (packet header-based). In addition to Mode 0 inputs, the evaluator extracts frame lengths and I-/non-I frame flags from transport headers (e.g., MPEG-TS over TCP or QUIC). No payload parsing is performed, yet the extra structural information improves detection of quality switches and playback freezes.

- Mode 2 (parser-based, light). Roughly 2% of the video bitstream is parsed, yielding frame boundaries and more precise codec parameters. This mode requires partial payload access but remains lightweight enough for real-time QoE probes.

- Mode 3 (parser-based, full). The complete elementary stream is parsed; all frame types, quantization parameters, and GOP structures are extracted, enabling the most accurate estimation at the cost of higher processing complexity.

For each segment, P.1203 computes three sub-scores: video quality, audiovisual synchrony, and playback fluency. These scores are then combined into an overall MOS using a perceptual integration function.

The CQM proposed by (Tran et al., 2021) takes a different approach. It accumulates time-varying impairments on-the-fly using only application-level KQIs obtainable from the DASH player or from transport logs. The required input parameters include:

- Segment bitrate and download time (used to derive effective throughput),

- Resolution layer and quality-switch amplitude,

- Stalling frequency and duration,

- Start-up delay.

CQM maintains a running quality score that decays with every stall or downward switch and recovers with sustained high-quality playback. This dynamic scoring model captures the recency and duration effects observed in subjective evaluations.

Since no bitstream or packet header data is required, CQM is especially useful when only manifest files and player events are accessible.

Both P.1203 and CQM are well-suited to downlink adaptive streaming scenarios, where the client receives segmented content and where the required metadata or playback logs are typically available. However, they are not directly applicable to uplink real-time video (e.g., live conferencing), where the encoder operates on-device, segments are not used, and transport behavior is driven by sender-side congestion control rather than adaptive downloading.

While application-level data forms the foundation for developing and evaluating QoE models, MNOs face significant challenges in accessing this data due to constraints from servers or the UE. Although these models offer valuable insights, there is a pressing need to integrate QoE evaluation more closely with network data. This integration raises privacy concerns as it involves handling sensitive user information. Additionally, reliance solely on application-level data, without considering signal quality measurements and network element load status, limits MNOs' ability to accurately predict QoE across different geographic regions and times, particularly given the fluctuating network load depending on the time of day. To address these challenges and enhance customer QoE insights, academia, network equipment manufacturers, and MNOs are actively researching and developing methods to leverage available data more effectively.

1.3.5. Quality of Experience Estimation Based on Session and Transport Metrics

QoE can be estimated based on metrics derived from transport-layer protocols (e.g., TCP/IP, QUIC, UDP) and session-level indicators (e.g., HTTPS) that reflect overall network performance. This approach has been applied to estimate QoE for web browsing, VoIP, online gaming, and most extensively in the context of HAS. The primary network traffic metrics used as input for QoE estimation are Round-Trip Time (RTT), packet loss, throughput, and jitter. These metrics affect the initial page load or time-to-first-frame, the congestion window or frame loss, media resolution, and the QoS of real-time services, respectively.

While network metrics such as RTT, packet loss, throughput, and jitter allow the modeling of QoE, widespread encryption (TLS 1.3, QUIC, SRTP) has made this task far more challenging for operators. Classical monitoring based on packet parsing and flow signatures has become infeasible. The operators are now limited to packet headers, timing, size distributions, and statistical flow behavior. Developing robust QoE models under these constraints remains an active area of research.

Work by Gutterman et al. (2019) demonstrated that video and audio chunks can be distinguished solely from IP-header fields in encrypted HAS flows. Statistics of the identified chunks were used as input features for a machine-learning classifier that predicted buffer state (low, high), playback phase (increase, steady, decay, stall), and video resolution, achieving high accuracy on a YouTube dataset. Mangla et al. (2019) proposed eMIMIC, which models HAS sessions using packet-header features to estimate average bitrate and rebuffering ratio. The system achieved 89.6% accuracy in classifying buffer occupancy and 85.7% in classifying the bitrate of recent chunks for both VoD and live streams. More recent works leverage sequence-learning models. (Zhang et al., 2024) used a Transformer architecture trained on time series of RTT, loss, and jitter extracted from encrypted UDP traffic to predict future real-time video quality, achieving mean-squared errors below one VMAF point. Shen et al. (2020) introduced DeepQoE, a CNN model that relies solely on upstream RTT samples. Despite this minimal feature set, it estimates start-up delay, rebuffering events, and resolution changes with accuracy comparable to feature-rich baselines while substantially reducing online complexity.

Collectively, these studies show that even under encryption, carefully selected header-level and timing features, when coupled with ML models, can provide reliable estimates of perceptual QoE metrics such as start-up delay, stall probability, buffer level, and delivered resolution.

1.3.6. Radio Access Network Metrics and Parameters

Application-level data are well-suited for the development and evaluation of QoE models. IP packets and transport-layer segments flow through the MNO network, encapsulated in the 3GPP technology stack, as shown in Figure 1.1. However, MNOs face challenges in accessing or parsing this data due to encryption and restrictions imposed by user equipment and application servers. Although the application-level QoE model provides valuable insights, transferring the QoE state to the network is not a common practice. This transfer raises privacy concerns, as it may involve sensitive user information. Furthermore, relying on application-level data, without accounting for radio signal quality and network element load, limits the ability of MNOs to predict QoE accurately across varying locations and time periods.

Given the challenges of directly assessing QoE, MNOs can rely on indirect data sources. These sources include network data flows, side-channel information, QoE forecasts derived from equipment-related KPIs, and RF signal quality across the network coverage area. A fundamental aspect of QoE evaluation for MNOs involves understanding the types of data that can be collected, the locations for data

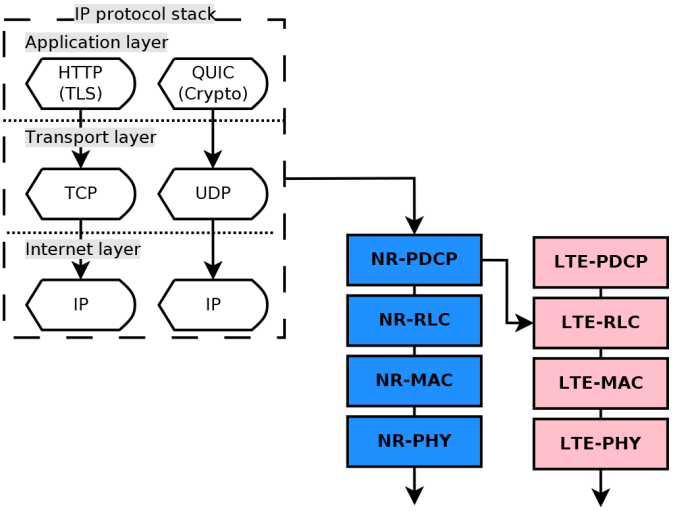


Fig. 1.4. Data delivery via the IP stack (TCP/IP or UDP/QUIC) mapped onto the 5G NSA protocol layers, combining NR and LTE domains (created by the author)

collection, the timing, and the methods to be used. As highlighted by Barakovic et al. (Baraković & Skorin-Kapov, 2013), the following four monitoring challenges are crucial for the research community to address: (1) identifying the data to collect; (2) determining the collection points; (3) deciding the timing of collection; and (4) establishing the collection methods. Understanding the cellular network architecture, its interfaces (Fig. 1.1), and the protocol stack is key to addressing these challenges.

A particularly data-rich interface in mobile networks is the air interface, which connects the UE to the RAN. In LTE and 5G networks, this interface, known as Uu, is responsible for transmitting both payload and signaling and control information. The Uu interface not only transfers payload data encapsulated at the application layer to lower-layer protocols, such as TCP/IP, but also transmits information such as RF channel quality, downlink and uplink buffer status, and more. Additionally, it handles the transfer of LTE and NR signaling messages for various operations, including scheduling, call setup and termination, carrier aggregation, handovers, and control information. Given the comprehensive nature of data available through the Uu interface, RF parameters extracted from it are often used as indirect estimators of QoE.

Collecting payload and signaling data between the UE and RAN provides valuable insights. Beyond standard signaling and payload, MNOs can request UEs to perform network measurements and report the findings. The 3GPP Minimization of Drive Tests (MDT) initiative enables the collection of signal strength and quality

data, along with global navigation satellite system (GNSS) coordinates. However, processing and analyzing the vast amount of data from the Uu interface is essential to obtain meaningful insights. When combined with UE signaling and positioning data, such as MDT reports containing GNSS coordinates, this analysis enables the assessment of QoE across different network regions.

In addition to detailed RAN measurements and probing data, MNOs can infer QoE from basic performance metrics of network nodes, such as RF coverage quality, QoS configurations, and other KPIs. Signaling metrics, such as RRC setup time, scheduling patterns, handover frequency, and radio link failures, are critical for QoE evaluation. For instance, long RRC setup times or failed attempts delay service start and degrade user experience. Sparse scheduling and frequent handovers can disrupt data flow. This is particularly relevant for services such as high-bitrate video-on-demand, where performance may suffer due to delays in carrier aggregation setup and release (Chmieliauskas & Paulikas, 2021). Similarly, real-time uplink streaming is sensitive to frequent or failed handovers (Chmieliauskas & Paulikas, 2025). Integrating signaling data with transport-layer metrics enables more accurate and service-aware QoE modeling.

At the most fundamental level, RF coverage and signal quality determine raw capacity and reliability available to users. Metrics such as Reference Signal Received Power (RSRP) and Signal-to-Interference-plus-Noise Ratio (SINR) reflect coverage strength and interference conditions. Poor signal quality results in lower throughput and higher error rates, directly degrading QoE. For example, video streaming may switch to a lower resolution or buffer more frequently. The capacity of a mobile network is typically expressed in bits per second per square kilometer (bit/s/ $^{\mathrm{km2}}$) and can be modeled as follows:

$$T_{\text{area}} = B \cdot D_{\text{cell}} \cdot \eta_{\text{spec}}, \tag{1.1}$$

where T_{area} is the area throughput (bit/s/km^2), B is the bandwidth (Hz), D_{cell} is the cell density (cells/km^2), and η_{spec} is the spectral efficiency (bit/s/Hz/cell).

For 3GPP-based cellular networks, the maximum data rate under ideal conditions can be estimated using the following approximate formula (3rd Generation Partnership Project, 2025a):

$$R = 10^{-6} \sum_{j=1}^{J} \left(v_{\text{Layers}}^{(j)} \, Q_m^{(j)} \, f^{(j)} \, R_{\max} \, \frac{N_{\text{PRB}}^{BW(j),\mu} \times 12}{T_s^{\mu}} \, (1 - OH^{(j)}) \right), \tag{1.2}$$

where R denotes the data rate in Mbps, J is the number of aggregated component carriers, $v_{\text{layers}}^{(j)}$ is the maximum number of MIMO layers, $Q_m^{(j)}$ is the modulation

order, $f^{(j)}$ is a scaling factor, $R_{\max} \approx 0.926$ is the maximum coding rate, $N_{\mathrm{PRB}}^{\mathrm{BW}(j),\mu}$ is the maximum number of physical resource blocks for carrier bandwidth $\mathrm{BW}(j)$ and numerology μ, T_{s}^{μ} is the OFDM symbol duration, and $OH^{(j)}$ is the overhead factor (e.g., 8% for UL in Sub-6 GHz frequency bands). The factor 12 represents the number of subcarriers in one PRB.

In NSA deployments, the total throughput is the sum of NR and LTE components (3rd Generation Partnership Project, 2025a):

$$R = 10^{-3} \sum_{j=1}^{J} TBS^{(j)}, \tag{1.3}$$

where $TBS^{(j)}$ denotes the Transport Block Size (TBS) in bits. The TBS is determined by the available resources, the Modulation and Coding Scheme (MCS), and other parameters defined in the 3GPP specifications.

The same formula (1.2) applies to real networks but considers actual RF and scheduling conditions instead of theoretical peak capacity. Achievable throughput in 5G NR NSA depends on multiple factors, including:

- Channel Quality: Estimated from measured SINR, channel quality determines MCS selection by the base station scheduler. Higher SINR enables higher-order modulation (e.g., 64-QAM, 256-QAM) and more efficient coding, resulting in larger transport block sizes per PRB and higher bit rates.

- Allocated Bandwidth: The number of PRBs assigned to a UE directly affects its bit rate, which depends on the total system bandwidth and how it is shared among users.

- Uplink Power Limitation: In the uplink direction, the number of PRBs can be limited by the UE's power capabilities. The UE is generally limited to a maximum transmission power (e.g., 21 dBm).

In real-time streaming scenarios, single UE throughput is important not only for achieving peak capacity under ideal conditions but also for ensuring consistent throughput to maintain stability (Chmieliauskas & Paulikas, 2025). Spectral efficiency can be improved by dynamically optimizing antenna orientation and tilt, which enhances signal quality and positively impacts QoE (Chmieliauskas et al., 2023). Thus, optimizing and modeling coverage and capacity is essential for both understanding and improving QoE-related metrics.

Network load and congestion represent additional critical factors. Even with a strong signal, if the cell is congested (due to many users or limited capacity), each

user's share of bandwidth and scheduling time decreases. High PRB utilization correlates with lower average user throughput (Chmieliauskas & Guršnys, 2019). This results in slower downloads and more frequent stalling. Advanced network features and optimal parameter settings, such as MIMO and scheduling strategies, also affect network throughput and, consequently, QoE. The study on MIMO configurations (Chmieliauskas, 2021) demonstrated that enabling closed-loop MIMO (TM4 in LTE) notably increases user data rates. Therefore, proper configuration of features such as MIMO, carrier aggregation, or scheduling algorithms can enhance or diminish QoE.

1.3.7. Crowdsourcing as a Supplementary Data Collection Method

The large subscriber base of mobile networks makes users well-suited for crowdsourced network evaluation. According to ITU-T P.912, crowdsourced network measurements involve outsourcing measurement tasks to participants who collect network related data (Hoßfeld et al., 2020). Crowdsourced network measurements not only gather application-level KQIs but also capture network-related indicators observed by the UE. These indicators include RAT, RSRP, and RSRQ, among others. Unlike application usage metrics that are typically only available to application developers, crowdsourcing provides insights into the QoS and QoE for the MNOs. From a research perspective, crowdsourcing provides access to millions of measurements without the need to perform individual tests (Raida et al., 2019). From the user's perspective, crowdsourced data collection can be classified into two categories: participant-initiated and automated measurements. Participant-initiated measurements can be exemplified by the Ookla Speedtest application, where the user actively performs download, upload, and latency measurements. In contrast, automated measurements are executed without user intervention and are typically performed in the background (Hoßfeld et al., 2020). Companies such as Umlaut and Tutela provide crowdsourcing services through partnerships with mobile applications. As users interact with these applications, throughput and latency measurements are periodically collected, and the results, along with coverage information, are reported. Implementing a crowdsourcing solution enables MNOs to obtain insights into user experience metrics such as coverage, throughput, and latency. This allows them to identify areas where QoE may be suboptimal and focus on RAN optimization. Data can be collected from multiple MNOs for comparison, facilitating inter-operator benchmarking. The key advantages of crowdsourcing include:

- Scale and coverage: Data are collected from real users across wide geo-

graphic areas, including indoor and rural locations where operators may not conduct drive tests.

- Real user experience: Measurements reflect device performance under typical conditions, with everyday handheld smartphones running various applications.

- Vendor-agnostic: Crowdsourcing captures multi-operator, multi-device data, enabling unbiased benchmarking.

As highlighted by (Seufert et al., 2021), the use of crowdsourcing requires caution. Crowdsourcing involves collecting large volumes of data from uncontrolled measurements performed by users. The density, number, and accuracy of these measurements can vary significantly due to the lack of control over the measurement environment.

Table 1.3 summarizes QoE assessment approaches and the types of mobile network data each approach relies on.

Table 1.3. Overview of QoE assessment approaches and the mobile-network data on which they rely

Approach	Data source(s)	Key metrics and features	Strengths	Limitations and challenges
Personal QoE	User self-report	Satisfaction, annoyance	Accurate ground truth	Intrusive; not scalable
Subjective QoE	Lab tests, crowdsourcing	MOS, ACR, DCR	Standardized; perceptual	Costly; slow; small scale
Objective QoE	Reference or processed media	SSIM, PSNR, VMAF, POLQA	Repeatable; no user panel	Requires payload access
Bitstream / parametric	Player logs, traces, manifests	Start-up delay, bitrate, stalling	Lightweight; suitable for HAS	Needs app-level visibility
Transport statistics	Encrypted headers, flow statistics	RTT, jitter, throughput, handshake time	TLS/QUIC-tolerant; scalable	Indirect QoE; model training needed
RAN metrics and KPIs	Radio interfaces, MDT, cell KPIs	RSRP, SINR, PRB load, handovers, RRC time	Operator-native; geo-time coverage	Context dependent; load sensitive
Crowd	Speed-test apps, background agents	Speed, latency, coverage, RAT, GNSS	Large scale; indoor/rural coverage	Uncontrolled; device bias

1.4. Big Data Platforms for Quality of Experience Analytics and Optimization

Once the available data types and their potential insights are identified, the next step is to collect and prepare them for analytics. Modern mobile networks generate massive volumes of QoE-related data, ranging from packet traces to signaling logs, requiring a big data approach for ingestion and analysis. The platform typically employs distributed processing (e.g., Hadoop, Spark, or cloud-native platforms) to manage large-scale datasets. QoE evaluation often involves correlating events across layers and time. Such correlation can result in billions of records per day in large networks. A distributed approach enables parallel computation, e.g., computing QoE indicators per cell or training ML models on terabytes of data. ML at scale enables training QoE prediction models on large sets of real network data. The availability of diverse data (transport metrics, signaling, user location, device type, etc.) enables the construction of multi-factor QoE models. However, data consistency across vendors and layers remains a key challenge. NWDAF was only recently standardized by 3GPP. Before standardization, data interfaces and collection types depended on vendor implementations. Therefore, building a fully compliant, multi-vendor platform for a future-proof O-RAN (Polese et al., 2023) architecture remains an open challenge. This platform must address connection interfaces, data collection, and xApp/rApp control loops to support efficient QoE slicing and AI-driven network optimization (Yeh et al., 2023). This vision is fully aligned with the 3GPP study on AI/ML for NG-RAN and the Release-18 work item on AI-enabled RAN management, which highlights data-driven control loops as a cornerstone of next-generation networks.

1.5. Service Types and Recognition

Each generation of cellular networks has introduced new use cases and application scenarios. Each step expands the spectrum of QoS/QoE requirements that operators must recognize and manage. Beginning with voice-centric services in 2G GSM, the evolution through mobile broadband in 4G LTE now extends toward advanced services envisioned for 6G. Figure 1.5, adapted from (International Telecommunication Union Radiocommunication Sector, 2023), illustrates the envisioned service types and usage scenarios for IMT-2030. Because each scenario imposes distinct latency, reliability, and throughput expectations, operators must first know which scenario a given flow belongs to before they can measure or guarantee its QoE. The diagram includes six primary usage scenarios. Three are

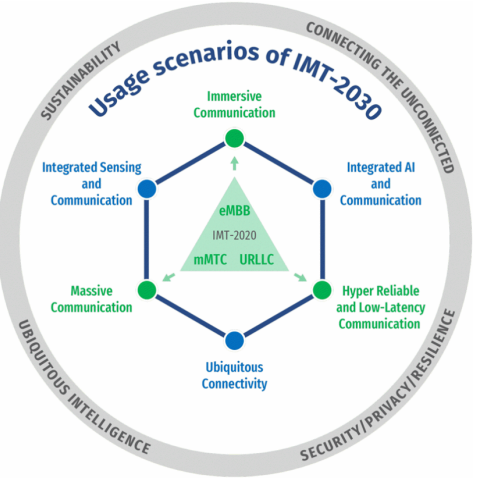


Fig. 1.5. Wheel diagram of usage scenarios and overarching aspects of IMT-2030 (International Telecommunication Union Radiocommunication Sector, 2023)

extensions from IMT-2020: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC). The other three highlight new 6G ambitions. Each usage scenario category encompasses a multitude of user applications, ranging from messaging and media streaming to remote control and teleoperation. Identifying the specific service type is a prerequisite for proper QoE evaluation and assurance. Every application has its own specific requirements in terms of latency, reliability, and throughput to meet the associated QoE demands. Consequently, each service type imposes distinct QoS constraints and requires distinct QoE evaluation methods. 3GPP defines groups of services and their corresponding QoS requirements, each associated with standardized QoS values. These values determine how network resources are allocated and prioritized. Such QoS profiles can be used to ensure QoE in network slicing scenarios that are tailored to the specific needs of different customer segments. The basic QoS requirements are defined in (3rd Generation Partnership Project, 2025b), in Table 5.7.4-1, which maps standardized 5G QoS Identifiers (5QIs) to a set of QoS characteristics. According to *3GPP TS 23.501* Table 5.7.4-1, each 5QI is associated with a fixed combination of parameters. Table 1.4 summarizes representative service types and their corresponding standardized 5QI mappings, highlighting how QoS parameters translate into application-specific QoE evaluation metrics.

These examples illustrate how, once a flow is recognized, it can be mapped to an appropriate 5QI and thus to the correct QoS. By enforcing these mappings in the RAN scheduler and core network policies, operators can deliver differentiated QoS

Table 1.4. Examples of service types, their standardized 5QI mappings, and commonly used QoE evaluation metrics.

Service type	Typical 5QI / QoS	QoE evaluation focus
Voice (call)	5QI 1; GBR; PDB 100 ms; PER 10^{-2}	Mean Opinion Score (MOS); call setup delay
Live video (e.g., VR)	5QI 2; GBR; PDB 150 ms; PER 10^{-3}	Frame stalling; motion-to-photon latency
Real-time gaming	5QI 3; Delay-critical GBR; PDB 50 ms; PER 10^{-3}	End-to-end latency; jitter; gaming MOS
Buffered / VoD video	5QI 4; Non-GBR; PDB 300 ms	Rebuffering ratio; average bitrate; SSIM
Web / FTP download	5QI 6; Non-GBR; PDB 300 ms	Page-load time; throughput percentiles
IMS signaling	5QI 5; Non-GBR; Priority 1; PER 10^{-6}	Session-setup delay; success rate
Mission-critical URLLC	5QI 70; Delay-critical GBR; PDB 10–20 ms; PER 10^{-5}	Reliability; outage probability; latency budget

guarantees that directly impact the QoE for each service class. This is especially relevant for 5G network slicing (Zhang, 2019), where multiple logical networks share physical infrastructure but require performance isolation. To enable dynamic network slicing and ensure QoE guarantees, MNOs must be able to accurately recognize the applications or services in use. Beyond infrastructure expansion, current efforts focus on improving visibility into user experience, which depends on recognizing and predicting service demands. For MNOs, service recognition and QoE assurance are challenging due to heterogeneous data sources and a lack of visibility into encrypted traffic patterns. Recognizing the type of application used by the client can be particularly challenging due to end-to-end encryption. Techniques such as port-based classification and DPI were widely used in the past; however, their accuracy has declined with the rise of encrypted traffic (Velan et al., 2015). To overcome these limitations, classical ML methods (Wang et al., 2019) and deep learning (Rezaei & Liu, 2019b) methods have been explored. Several studies have recommended deep learning as the preferred method for identifying encrypted traffic (Aceto et al., 2019; Lotfollahi et al., 2020). For time-series datasets, such as those used in this research, CNNs have been identified as particularly effective (Ismail Fawaz et al., 2019; Wang et al., 2017). CNN-based traffic classification has been demonstrated to accurately identify applications (Chmieliauskas & Paulikas, 2022) and has shown the capability in recognizing video traffic among other service types (Chmieliauskas & Paulikas, 2023).

1.6. Conclusions of the First Chapter and Formulation of the Dissertation Tasks

Based on the analytical review of QoE in cellular networks, the following conclusions are drawn.

1. QoE is a multidimensional concept shaped not only by network performance, but also by service type, content, device, application behavior, and usage context.
2. In OTT service scenarios, MNOs lack direct access to end-user QoE and reference media streams. Therefore, operator-side QoE evaluation must rely on indirect indicators derived from transport/session and RAN-layer measurements.
3. Service-type recognition is a prerequisite for operator-side QoE inference. Due to widespread encryption, MNOs must rely on header-level, timing-based, and traffic shape features combined with learning-based models.
4. Downlink HAS is supported by mature bitstream-based and parametric QoE models, whereas uplink real-time streaming still lacks operator-side QoE models.
5. Accurate operator-side QoE estimation requires the fusion of heterogeneous cross-layer data, including transport/session metrics such as RTT, loss, jitter, and throughput, together with RAN indicators such as coverage, scheduling, load, and mobility events.
6. RAN coverage and capacity are the foundation of achievable QoE. Predictive QoE models, together with NWDAF/O-RAN data-driven control loops, are needed to support proactive QoE-oriented RAN optimization.

As a result of the performed literature survey, the following research hypotheses are formulated:

H1. Objective QoE inference for video-streaming services is feasible at the operator side using only MNO-observable RAN measurements and IP transport metrics, without access to application payload or server-side reference streams.

H2. Encrypted video-streaming traffic can be recognized from aggregated 1 s byte-count traces, without deep packet inspection, by applying learning-based classification methods.

H3. For uplink real-time video streaming, objective QoE, expressed by VMAF, can be predicted from operator-observable path loss, MCS, and allocated PRBs with practically useful accuracy for network monitoring and control.

In line with the above conclusions, the dissertation addresses the following tasks.

1. Analyze the end-to-end data delivery path in 4G LTE and 5G NR networks, identifying which radio, transport, and application layer metrics are observable to the MNO.

2. Collect and classify MNO-accessible data sources (radio KPIs, transport protocol metrics, and media-layer data) and establish their relationship to QoE definitions for video streaming services.

3. Develop a lightweight and encryption-resilient algorithm for recognizing video streaming services based solely on network-side data.

4. Design and validate a model that maps RAN and IP metrics to objective QoE estimates for uplink real-time video streaming.

5. Develop and evaluate machine learning and reinforcement learning models to predict RAN capacity and support QoE-driven RAN parameter adaptation.

2

Used Service Recognition in Cellular Networks

For MNOs, knowing which services their subscribers use is essential for QoE evaluation and management, including traffic prioritization, throttling, and differentiated pricing. Each service imposes different requirements on delay, bandwidth, and scheduling period. Therefore, ensuring the desired QoE first requires reliable service recognition. However, the growing use of end-to-end encryption has reduced the effectiveness of service recognition. In this section, two feasibility studies are presented: (i) application-level classification of encrypted mobile traffic (Chmieliauskas & Paulikas, 2022), and (ii) a downlink video-on-demand recognition method that relies solely on the shape of the bitstream (Chmieliauskas & Paulikas, 2023). The goal is to test the concept of application recognition based solely on the bitstream shape as transferred over the mobile network.

2.1. Encrypted Traffic Recognition

End-to-end data transmission is an interaction between the application on the UE and a remote server or CDN. Typically, CDNs are owned by third-party OTT service providers and not by MNOs. Data delivery is handled using IP stack protocols such as HTTP, TCP, or UDP. From the MNO's perspective, it is treated as payload and delivered on a best-effort basis, regardless of the service type. MNOs could benefit from recognizing service types: they could prioritize, block, throttle, or apply different pricing models for specific traffic classes. Due to encryption, traditional traffic recognition approaches such as port-based identification and DPI became less effective (Papadogiannaki & Ioannidis, 2021; Rezaei & Liu, 2019a). To overcome this limitation, researchers have turned to conventional machine learning techniques (Velan et al., 2015) and, more recently, to deep learning models (Wang et al., 2019). Current and future mobile networks should be capable of recognizing user services and predicting both present and forthcoming demand for their delivery. On the MNO side, service recognition and QoE assurance are challenging due to heterogeneous data sources, varying temporal granularity of collected data, large data volumes, and stochastic user behavior. The downloaded and uploaded bytes can be considered a time series. Classification of time-series data is extensively covered in the literature (Bagnall et al., 2017; Ismail Fawaz et al., 2019; Wang et al., 2017). From a practical perspective, downlink/uplink buffer occupancy is a simple metric that is always available to MNOs and does not require complex parsing of TCP/IP or QUIC packets or inter-packet metrics calculation. Moreover, using only the amount of transferred bytes respects customer privacy, since no content parsing is involved.

2.2. Adaptive Streaming over Mobile Networks

Among various services used over mobile networks, HAS is particularly important due to its high bandwidth demand and sensitivity to fluctuating network conditions.

In mobile 4G and 5G networks, the wireless channel represents the last link in data, video, and voice delivery. In the wireless channel, the achievable throughput and delay fluctuate rapidly due to (a) the changing quality of the radio frequency (RF) signal, (b) the usage of the shared medium by other clients, and (c) the unknown delay and bandwidth demand of the current user. An unstable wireless environment makes it difficult to guarantee the QoS requirements for the services used. HAS was developed to mitigate rapid bandwidth fluctuations in the stochastic RF environment.

HAS adjusts video quality parameters to match the current network conditions (Seufert et al., 2014). HAS requires the video to be available in different bitrate versions, that is, in multiple quality levels, and divided into small chunks, each containing a few seconds of playback. The video stream receiver (also known as the client) monitors the available network bandwidth, buffer status, or both. The next portion of the video is requested at an optimal bitrate to prevent stalling (i.e., the discontinuation of playback due to insufficient data in buffers). Besides preventing stalling, HAS also maximizes utilization of the available network bandwidth. The ability of HAS to adjust video quality to available bandwidth has made it a popular choice among providers of OTT video services such as YouTube, Netflix, Twitch, and others (Roettgers, 2013a).

In 4G LTE and 5G NR networks, the individual user's bandwidth depends on the current cell load or the physical resource blocks (PRBs) available in the time-frequency grid (Capozzi et al., 2012) and the quality of the RF signal. To maintain stable video playback, users should always have a higher bandwidth than at least the average required by the video codec for a specific resolution, plus transmission-related overhead. Every video chunk should be completely downloaded before the previous chunk's playback time expires; otherwise, stalling occurs.

An example of this situation is shown in Figure 2.1.

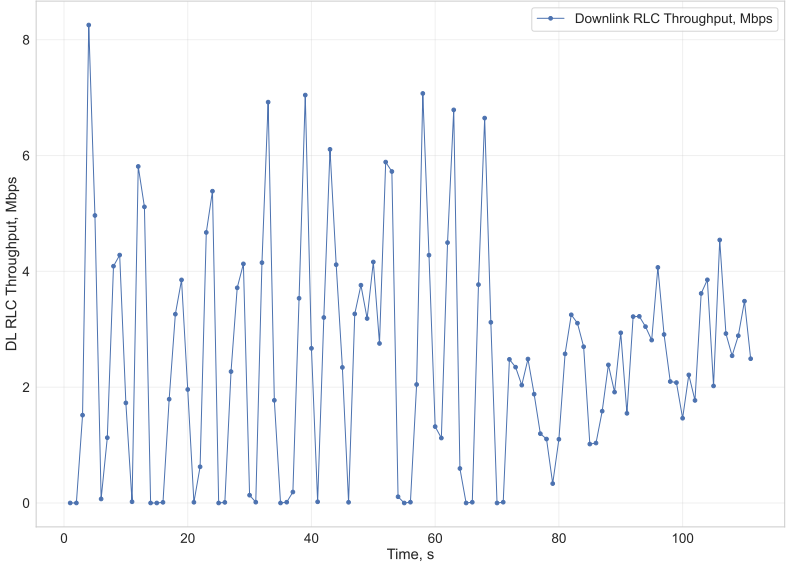


Fig. 2.1. Measured downlink bitrate at the base station during a stalled video stream (created by the author)

It shows the downlink bitrate aggregated over one-second intervals. The measurements were taken during video streaming at the base station side of the mobile network. During the first 70 seconds, the LTE network provided sufficient bandwidth to download video chunks at regular intervals. After the 70th second, the download rate became too low to sustain the periodic delivery of video chunks. The UE was moving toward a region of lower SINR within the LTE cell. As a result of reduced SINR, modulation and coding efficiency decreased, requiring more spectrum resources to transfer the same volume of data. Eventually, all available PRB resources were fully utilized by the video stream. After a few seconds of limited bandwidth caused by the combination of PRB saturation and low SINR, the playback buffer was depleted, and stalling occurred. HAS operates in a periodic pattern of downloading video chunks, followed by idle periods while the chunks are played back. If a base station were aware of the chunk size and download interval, its scheduler could periodically reserve the necessary time–frequency resources for that user. Therefore, recognizing the video service at the LTE/NR base-station side could help prevent video stalling and reduce QoE degradation for the user.

2.3. Deep Learning for Time Series Classification

Previous studies propose using TCP/IP packet header or packet interarrival time information to recognize services and applications (Alan & Kaur, 2016; Conti et al., 2015; Lopez-Martin et al., 2017). However, this method adds complexity related to data preprocessing and requires knowledge of TCP/IP stack protocols. Instead, a simpler approach is proposed that uses only the number of bytes transferred and received per second. The bitstream is treated as a time series, reducing the problem to time series classification. Time series classification has been extensively documented in numerous studies (Bagnall et al., 2017; Ismail Fawaz et al., 2019; Wang et al., 2017). In practical scenarios, downlink and uplink buffer occupancy is a simple metric that is always available to the MNO and does not require complex processing of TCP/IP or QUIC packets, nor the calculation of packet-level metrics. Moreover, because only the volume of transferred bytes is recorded, customer privacy is maintained without the need for content-level parsing. Several authors have recommended using deep learning, among other machine learning methods, for identifying encrypted network traffic (Aceto et al., 2019; Lotfollahi et al., 2020). The authors of (Casas et al., 2022) proposed using CNN for QoE inference during the use of various Internet applications. Based on related literature, CNN was chosen to be used in this study.

2.4. Proposed Method for Encrypted Traffic Recognition

This section describes the method for differentiating video streams from other mobile applications based solely on the bitstream shape. A CNN was trained for this task using our collected dataset. The dataset consists of bitstreams from popular Android applications, including Facebook, Chrome, YouTube, and others.

2.4.1. Collected Data Description

The assembled dataset consists of 3257 hours of 1-second resolution time-series data collected from real mobile network users. Acquiring data from actual users, rather than using synthetic bot-generated data, provides a more precise representation compared with other research (Conti et al., 2015; Taylor et al., 2017). The raw dataset includes the following features: the name of the application displayed on the UE screen, the number of bytes transferred for both download and upload by the network module, and a timestamp. An Android application was used to collect the names of the foreground applications and network buffer information from several volunteers. The model is trained using UE-side measurements. The same data are also available at the mobile network's base station and can be used for inference. Using only the number of bytes downloaded and uploaded makes it possible for MNOs to use this model without directly violating privacy or incurring the high computational costs of parsing encrypted TCP/IP packets.

2.4.2. Dataset Generation

To prepare the training dataset and retain temporal information, the collected data are divided into windows of 60, 300, and 600 s, with overlaps of 59, 299, and 599 s, respectively. A 60 s window corresponds to short-form viewing and captures short-term HAS download/idle cycles. Additionally, 60 s is consistent with the typical length of short online videos (Bynder, 2024). In contrast, 300 s and 600 s windows correspond to session-scale behavior and align with typical 5–10 min network management intervals. Using multiple window lengths enables the model to learn and validate recognition across both short-term and long-term temporal patterns. Each data window is assigned a label indicating the application that is active on the smartphone at that time. Windows containing multiple applications are discarded. The windowing-based data augmentation technique generated a large number of training samples: over 9 million windows of 60 s, more than 5 million of 300 s, and about 3 million of 600 s duration. From this dataset, a balanced number of samples representing each application can be selected.

2.4.3. Dataset Categories Description

To gain insight into the usage patterns of the collected data, it was analyzed in terms of duration of use and statistical properties. The applications with the longest usage duration are presented in Table 2.1.

Table 2.1. Applications ranked by usage duration

Application on Screen	Duration Percentage (%)	60 s Sample Count
Start Screen	28.63%	2,672,802
Facebook	13.87%	1,294,841
Chrome	11.26%	1,051,053
Google	9.62%	898,295
YouTube	6.59%	614,714
9GAG	3.15%	293,852
Android Auto	2.77%	258,592
Messenger	2.55%	238,133
Outlook	2.30%	214,883

Over 28% of the samples are generated during inactive phone usage and have the Start Screen label. Even when the smartphone is not in use, applications continue to generate background data transfer. The next most frequently used application is Facebook. Ranked third and fourth are Chrome and the Google browser. The list continues with YouTube, a video streaming platform, followed by 9GAG, which is an image-downloading application.

A few representative examples of the most frequently used applications were selected and visualized. Figure 2.2 shows five-minute data samples illustrating downlink and uplink bitstreams during the use of YouTube, Facebook, Chrome, Outlook, Messenger, and 9GAG. Analysis of random samples shows that different applications show distinct bitstream patterns. For example, the Chrome, Facebook, and 9GAG apps show periods of high data transfer and intervals of minimal network activity. These patterns likely correspond to periods of active engagement and content downloading, followed by intervals of passive viewing or reading of previously downloaded material. The HAS pattern can be clearly observed in the YouTube application, as video segments are both downloaded and played. The email application Outlook uses both downloading and uploading simultaneously to receive and send emails. The Messenger app shows a pattern of constant downlink and uplink activity during voice or video calls.

Random samples provide an initial understanding of bitstream patterns but do not represent the majority of cases. To compare a large number of data samples systematically, statistical features must be used. For this purpose, the library of

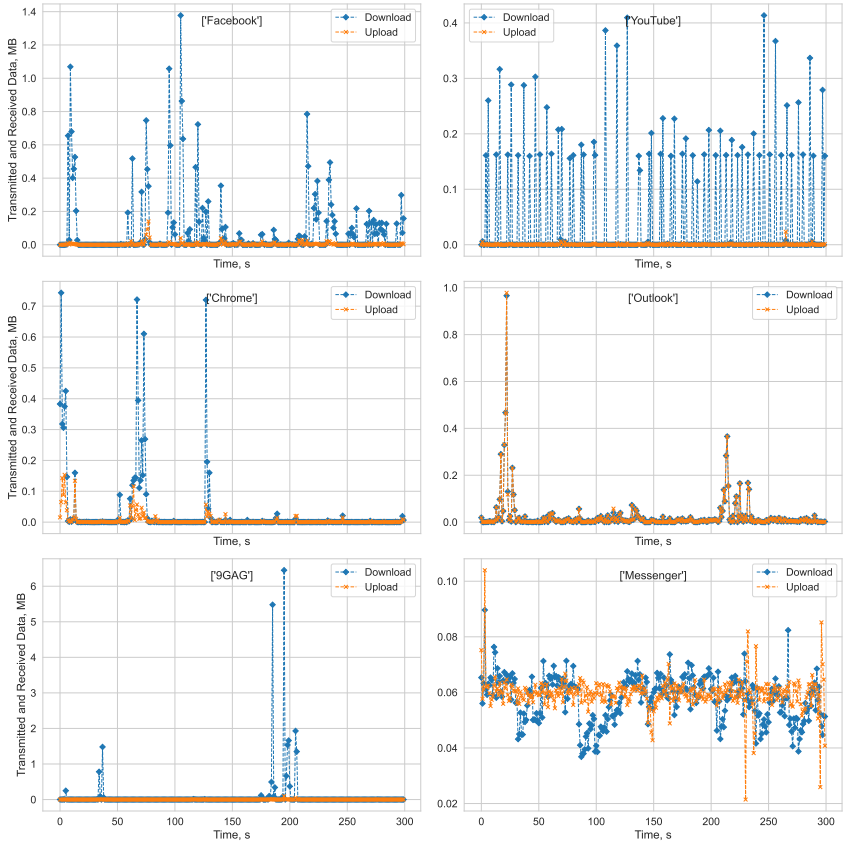


Fig. 2.2. Random samples of bitstreams representing different applications (created by the author)

time series feature generation (Barandas et al., 2020) was applied to extract key statistical properties from the collected data. In the analysis, the focus was on two simple yet informative metrics: the mean and standard deviation of downloaded and uploaded bytes. Figure 2.3 shows the kernel density estimate (KDE) of the mean and standard deviation calculated from the generated samples, grouped by application. For clearer visualization, a logarithmic scale was applied to the x-axis. It can be observed that the mean and standard deviation of the samples exhibit a high degree of overlap among most applications. However, some applications show distinctive characteristics. For example, the Messenger app shows higher uplink traffic, which is likely associated with video or voice calls. Similarly, YouTube shows a larger volume of downlink traffic and a higher standard deviation with multiple peaks, which can be attributed to video streams of different resolu-

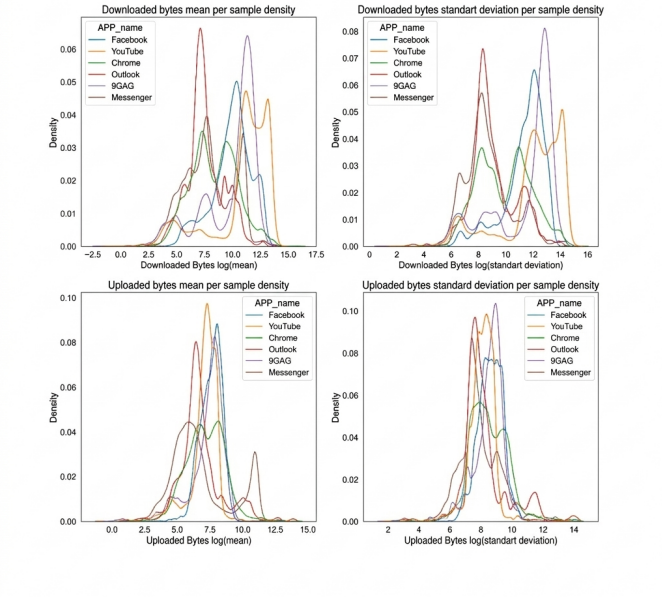


Fig. 2.3. Mean and standard deviation of downlink and uplink bitstreams, grouped by application (created by the author)

tions. Due to the substantial overlap in these statistical features, relying solely on the mean and standard deviation is insufficient to reliably distinguish among all applications. Thus, it becomes necessary to incorporate time-series shape analysis. To address this, a CNN was employed to classify applications based on the time-series shape of the received and transmitted byte patterns. In addition to the general classification, a focused comparison of YouTube samples was conducted, representing the HAS video service, against all other applications.

Figure 2.4 illustrates the KDE of the mean and standard deviation for these two categories. It can be observed that video samples tend to exhibit both a higher mean bitrate and a higher standard deviation compared to other services. Additionally, YouTube exhibits multiple peaks, which can be attributed to video streams of different resolutions. Another notable aspect of the video bitstream is its fundamental frequency. Figure 2.5 shows the cumulative distribution function (CDF) of the downlink bitrate during adaptive video streaming.

By analyzing the bitstream as a signal, a fundamental period of approximately 10 seconds was observed. This corresponds to the delivery period of video chunks or the duration of each chunk's playback.

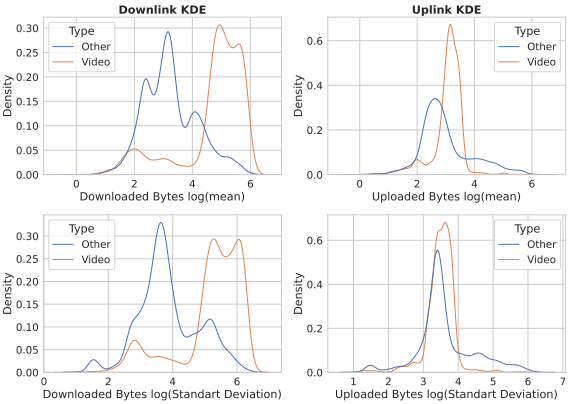


Fig. 2.4. Mean and standard deviation of downlink and uplink traffic samples for video streaming and other applications (created by the author)

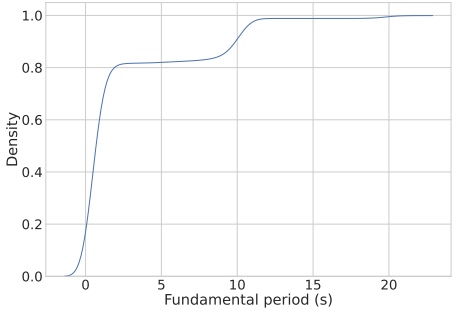


Fig. 2.5. Cumulative distribution function (CDF) of the fundamental period of the downlink bitrate during adaptive video streaming (created by the author)

Recognizing video streaming traffic allows the mobile network base station to estimate additional parameters such as the average bitrate, chunk size, and fundamental frequency. This information can be used to configure the scheduler for optimized video service delivery and improved QoE. In summary, although the statistical features of mean and standard deviation provide some level of differentiation between video and other services, a more reliable distinction requires analyzing the time-series shape. For this reason, a CNN model was used to classify applications based on the byte patterns over time.

2.4.4. Convolutional Neural Network for Video Stream Recognition

CNNs are neural networks that employ convolution in at least one of their layers, rather than using general matrix multiplication (Goodfellow et al., 2016) and are

optimized to handle data in array form (LeCun et al., 2015). The dataset consists of one-dimensional (1D) arrays when only the downlink is considered, and two-dimensional (2D) arrays when both downlink and uplink are included. For this reason, a CNN was used to recognize video streams and classify traffic using only the received and transmitted bytes. The architecture of the proposed neural network is shown in Figure 2.6.

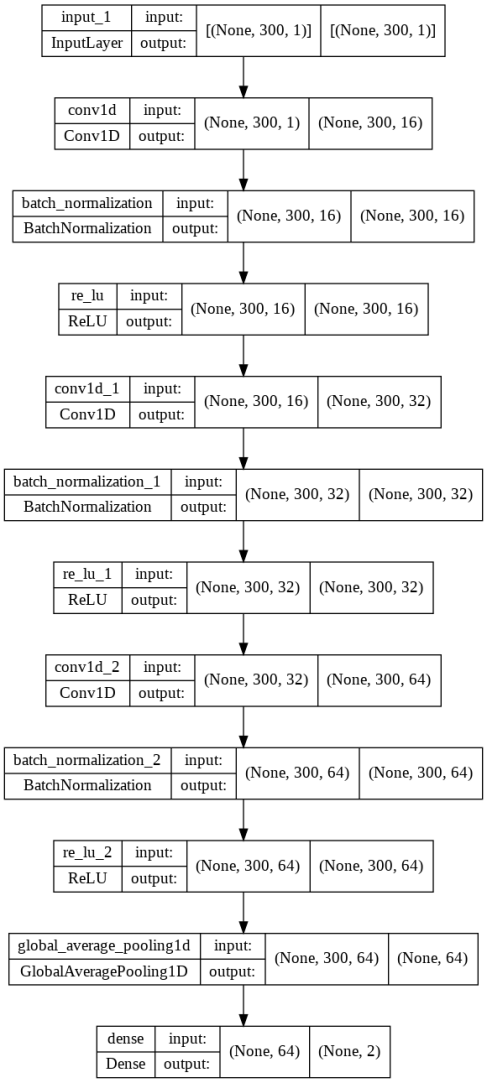


Fig. 2.6. Architecture of the convolutional neural network used for traffic classification (created by the author)

The Keras library was used to implement the design and training of the CNN model (Chollet et al., 2015). The deep learning architecture was chosen specifically to evaluate the feasibility of traffic classification and video stream recognition, without engaging in detailed algorithm evaluation or parameter optimization. The CNN takes as input a vector of one or two rows, representing either downlink-only or both downlink and uplink bytes, with each input consisting of 60, 300, or 600 numerical values. The input vector undergoes convolution, followed by batch normalization (Ioffe & Szegedy, 2015). Nonlinear Rectified Linear Unit (ReLU) functions are then applied to the normalized data. The process of convolution, normalization, and ReLU is repeated three times using multiple layers. After the last ReLU function, Global Average Pooling is applied. Finally, the output layer produces the probability distribution over video and non-video application classes. The deep learning model is trained using the Adam Optimizer and the sparse categorical cross-entropy loss function. Training progress is evaluated using sparse categorical accuracy. After training, the model is evaluated against new data. The preparation of the dataset has a high memory requirement, and the model training demands GPU resources. The Google Cloud Platform was used to deliver high-performance and flexible computing resources.

2.5. Performance Evaluation

This section describes the testing methods, evaluation metrics, and results for both traffic classification and video stream recognition based on the bitstream shape. The traffic classification, video stream recognition, and application recognition models are trained using samples of 60, 300, and 600 seconds. For traffic classification by application task, randomly selected samples are used to represent each application. For video stream recognition, the samples are divided into two categories: video and non-video. Although the HAS service can be used within a web browser or other applications, the video category includes only samples from the YouTube app. For any sample sequence length, 100,000 samples were selected to represent video and 100,000 samples from all other applications. Selecting an equal number of samples from each category helps avoid training with imbalanced data. The model is trained and evaluated using different sequence lengths by considering three input configurations:

1. Downlink bytes only;

2. Downlink followed by uplink bytes combined into a single 1D sequence;

3. A 2D sequence where downlink and uplink are fed simultaneously as separate channels.

2.5.1. Evaluation Metrics

To evaluate the performance of the classification, several evaluation metrics were used. First, the accuracy was calculated using Equation (2.1). Then, the precision and recall were determined using Equation (2.2).

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}. \tag{2.1}$$

$$P = \frac{T_p}{T_p + F_p}, \quad R = \frac{T_p}{T_p + F_n}. \tag{2.2}$$

Accuracy is calculated as the number of correct classification predictions divided by the total number of predictions. Precision P is defined as the ratio of true positives T_p to the sum of true positives and false positives ($T_p + F_p$). Recall R is defined as the ratio of true positives T_p to the sum of true positives and false negatives ($T_p + F_n$). Furthermore, the misclassification rate was calculated for each of the most used applications according to Equation (2.3). The misclassification rate shows the number of instances that are incorrectly classified out of the total instances. In this case, it represents the number of instances where the YouTube application, identified as a video streaming service, was incorrectly classified as another application. Similarly, it includes cases where other popular applications (e.g., Facebook or the Chrome browser) were misclassified as video streaming.

$$\text{Misclassification rate} = \frac{F_n + F_p}{T_n + T_p + F_n + F_p}. \tag{2.3}$$

Using the misclassification rate for each of the most frequently used applications, both the frequency of correct and incorrect video recognition was assessed, and it was identified, which applications are most frequently misclassified by the model.

2.5.2. Evaluation Results: Application Recognition Based on Bitstream Shape in Mobile Networks

The evaluation was performed on input sequences of equal length using randomly selected unseen data from the collected dataset. The classification results using the

CNN model trained on the 60, 300, and 600 s sequences are presented in Table 2.2.

Table 2.2. Precision and recall values for application classification using different input sequence lengths

App / Sample Time	Precision			Recall		
	60 s	300 s	600 s	60 s	300 s	600 s
Facebook	0.517	0.522	0.615	0.100	0.379	0.666
YouTube	0.976	0.978	0.970	0.515	0.587	0.718
Chrome	0.250	0.278	0.284	0.278	0.507	0.744
Outlook	0.256	0.354	0.626	0.022	0.688	0.255
9GAG	0.632	0.774	0.815	0.423	0.473	0.536
Messenger	0.267	0.772	0.769	0.908	0.225	0.378

The normalized confusion matrix among the six application categories used for classification is presented in Figure 2.7.

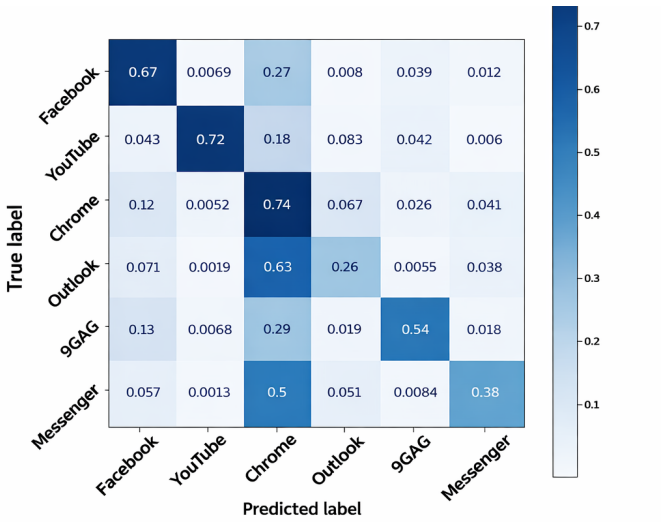


Fig. 2.7. Normalized confusion matrix among six applications (created by the author)

The confusion matrix illustrates the performance of the trained model. It shows how applications from the test dataset were misclassified. The y-axis represents the actual application, while the x-axis represents the predicted application. The values in the matrix show the proportion of times each application was predicted for a given true application.

In Figure 2.7, per-class recall ranges from 0.26 to 0.74 (mean $\approx$ 0.55), which

is above the level and confirms that the bitstream shape is informative. The main errors are systematic confusions, where Outlook and Messenger are predicted as Chrome, likely due to overlapping browser-like traffic patterns. YouTube is rarely confused with other applications, suggesting low video misclassification. Given the dominant share and high bandwidth demand of video traffic, this motivates the binary task of video versus non-video recognition.

2.5.3. Evaluation Results: Video Stream Recognition Based on Bitstream Shape in Mobile Networks

Due to the importance of video streaming, as introduced in the previous chapter, it was selected for a detailed evaluation. Video streaming constitutes the largest share of Internet traffic and requires high QoS, particularly in terms of bandwidth and delay sensitivity.

The model is evaluated using input sequences of identical duration to those in training, but drawn from completely new, unseen samples. Samples are chosen at random from the gathered dataset in equal proportions between the video category and all other applications. The classification accuracy achieved by the trained CNN on 60, 300, and 600 s sequences is shown in Table 2.3.

Table 2.3. Classification accuracy of the CNN model for video stream recognition

Sequence Length	Downlink Only	Downlink + Uplink (1D)	Downlink + Uplink (2D)
60 s	0.772	0.814	0.837
300 s	0.859	0.902	0.896
600 s	0.815	0.900	0.929

Precision and recall values for distinguishing between video streams and other applications, obtained using the CNN to classify 60, 300, and 600 s sequences, are shown in Tables 2.4 and 2.5.

Table 2.4. Precision values by input type and sequence length for CNN-based video stream classification

Input type and class	60 s	300 s	600 s
Downlink only – Video	0.902	0.991	0.986
Downlink only – Other app	0.769	0.830	0.875
1D (DL+UL) – Video	0.974	0.969	0.983
1D (DL+UL) – Other app	0.752	0.840	0.879
2D (DL+UL) – Video	0.965	0.982	0.997
2D (DL+UL) – Other app	0.770	0.834	0.908

Table 2.5. Recall values by input type and sequence length for CNN-based video stream classification

Input type and class	60 s	300 s	600 s
Downlink only – Video	0.726	0.799	0.860
Downlink only – Other app	0.921	0.993	0.988
1D (DL+UL) – Video	0.679	0.816	0.985
1D (DL+UL) – Other app	0.982	0.974	0.866
2D (DL+UL) – Video	0.711	0.806	0.900
2D (DL+UL) – Other app	0.974	0.985	0.998

To enhance the robustness of the evaluation and make it more representative of real mobile network usage, the model was evaluated using new data with the same, naturally unbalanced proportions of applications as observed in the collected dataset. The number of samples from each application and their corresponding misclassification rates are shown in Figure 2.8.

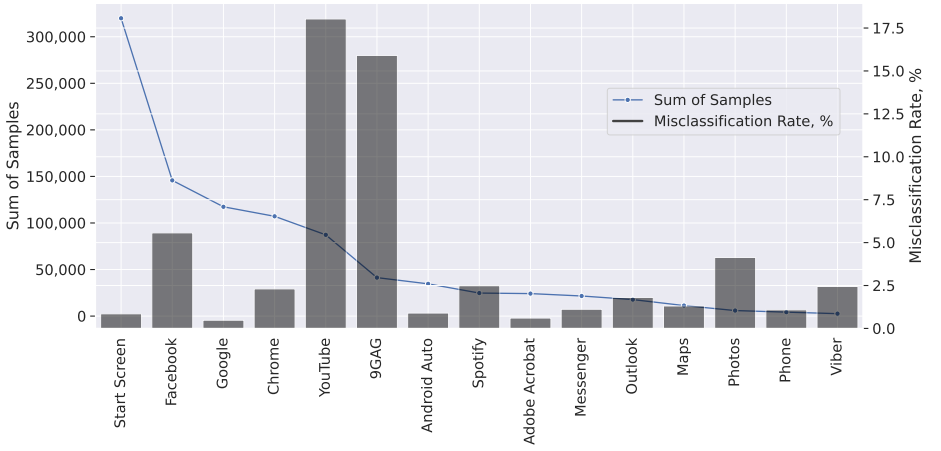


Fig. 2.8. Misclassification rates by application: instances where video traffic is classified as other applications, and vice versa (created by the author)

A comparison is presented between the proposed method and other techniques used for identifying service classes or specific applications in network traffic. The comparison is based on various criteria, such as the purpose of classification or recognition, the algorithm type, the method of dataset collection or generation, and the complexity of data preprocessing and required networking domain knowledge. The performance of the methods is compared in terms of accuracy, precision, and recall, and the results are presented in Table 2.6. The proposed method simplifies network traffic classification by eliminating the need for TCP/IP packet

Table 2.6. Comparison of the proposed method with other traffic classification and application recognition techniques

Work	Alg.	Data	Inputs/prep	Acc	P	R
(Zheng et al., 2022)	Trans.	ISCX VPN–nonVPN	Packet files; requires TCP/IP knowledge	0.993	0.988	0.988
(Yao et al., 2019)	LSTM	ISCX VPN–nonVPN	Flow reconstruction; requires TCP/IP	0.912	—	—
(Lotfollahi et al., 2020)	Autoen.	ISCX VPN–nonVPN	Header and payload parsing; requires TCP/IP	0.980	0.990	0.990
(Lopez-Martin et al., 2017)	CNN	RedIRIS	IP-flow headers; requires TCP/IP	0.961	0.952	0.961
Proposed method	CNN	Authors' dataset	Bitstream shape only; no TCP/IP required	0.929	0.997	0.900

parsing and session flow construction. Furthermore, the proposed method does not require knowledge of the TCP/IP protocol stack and its implementation in actual network nodes. The comparison shows that, despite using simple bitstream input data, the proposed method achieves accuracy, precision, and recall comparable to other methods. It is worth noting that other related studies use datasets containing more detailed information from within network protocols, including source and destination ports, packet payload size, TCP window size, interarrival time, and packet direction. In contrast, the proposed method only evaluates the shape of bitstreams generated by different services on a second-level granularity, without inspecting the content of packets. If the goal is not only to recognize encrypted video streaming but also to guarantee QoS and QoE at the mobile network base station, data availability must be considered. The scheduler of LTE and NR nodes is implemented in lower layers, such as MAC, which are not aware of TCP/IP packet content and treat it as payload only. Therefore, a method for identifying video streaming services based solely on the shape of the generated bitstream enables the use of simple and accessible data without raising privacy concerns.

Based on the evaluation results presented in Table 2.3, it was observed that the precision of recognizing video streams from other applications is relatively low for input sequences of 60 s, but reaches approximately 90% for input sequences of 300 s and longer. Also, using downlink and uplink data gives better performance than relying only on downlink bitstreams. The precision and recall results in Tables 2.4 and 2.5 also suggest that a duration of 60s is too short for recognizing that a video streaming service is in use. As shown in Figure 2.5, the fundamental frequency of HAS is around 10 s, so only a few chunks are downloaded and acknowledged during the one-minute interval. Based on the misclassification rate results

of the 300s input sequences shown in Figure 2.8, YouTube is the top application in terms of usage time in our test dataset, and it is correctly recognized in 85% of predictions. Misclassifications mainly occur during the times when HAS was not streamed, but the application start page was browsed. Other applications with high misclassification rates are 9GAG and Facebook. Both applications, in addition to picture browsing, have embedded video players. In this work, misclassification between video and non-video services mostly occurs due to the capabilities of the applications, such as when the video application is only being browsed, or when another application is using a video streaming service. The results obtained show that distinguishing between video streaming and other types of applications is feasible, but the dataset collection should be improved to avoid times when the application with the video streaming capability is in use, but no video is being streamed.

2.6. Conclusions of the Second Chapter

This chapter studied service recognition for QoE management. The low-complexity, privacy-preserving approach that uses only per-second downlink and uplink byte counts, treated as a time series, was evaluated. A CNN-based classifier was trained on real-user data to recognize HAS video traffic and distinguish among common applications.

1. The results confirm that service recognition in mobile traffic is feasible without packet-content inspection or protocol-level feature extraction. This is particularly relevant for MNO-side QoE management, where port-based and DPI-based methods become less effective under encryption.

2. The obtained results show that the shape of the encrypted bitstream carries sufficient discriminative information to support service recognition from simple second-level downlink and uplink traffic-volume measurements.

3. The proposed CNN-based method achieved video-versus-other classification accuracy above 0.90 for 300–600 s windows, demonstrating that HAS traffic can be identified with practically useful accuracy from low-complexity traffic measurements.

3

Evaluation of Video-Streaming Quality of Experience in Cellular Networks

This chapter explores QoE-related challenges and evaluation techniques for uplink and downlink video streaming. It presents detailed insights based on two studies. First, it analyzes OTT video streaming methods, emphasizing their impact on network load and the associated operator costs due to physical resource block (PRB) utilization across varying RF quality scenarios (Chmieliauskas & Paulikas, 2021). Second, it evaluates real-time uplink video streaming over protocols such as RTMP and WebRTC, focusing on QoE degradation under varying network conditions. The study provides datasets consisting of video captures, network traces, and QoE metrics designed for regression analysis and QoE prediction (Chmieliauskas & Paulikas, 2025; Chmieliauskas & Paulikas, 2024).

3.1. Video Delivery and Downlink Network Load Analysis

The bandwidth and latency requirements of video delivery are very high. Downlink video delivery, especially for high-resolution content, is the main contributor

to radio network load and congestion. Moreover, user expectations regarding the QoE of OTT video services are particularly high. Viewers begin abandoning a video if it takes more than 2 s to start, with each additional 1 s delay increasing the abandonment rate by 5.8% (Krishnan & Sitaraman, 2013). About 10% of viewers abandoned a video when the pre-roll advertisement lasted for 15 s (Nam et al., 2014). On one hand, video content from third-party providers generates a significant share of the network load for MNOs; on the other hand, poor QoE can lead to customer dissatisfaction, complaints, and churn.

For an MNO to evaluate and ensure high QoE during downlink video streaming, it is essential to understand the OTT delivery methods, underlying transport protocols, and video codecs. Such knowledge helps to relate service configurations to the equivalent bit rate as measured in terms of LTE network load. The amount of scheduled PRBs within the LTE time–frequency resource grid can be directly related to the MNO's cost of delivering such traffic.

3.1.1. Application and Transport Layer Protocols for Video Delivery

Video streaming relies on different application and transport layer protocols. The choice of protocols, methods, and codecs often depends on the CDN provider. Examples of such protocols include HTTP or HTTPS, SPDY, and QUIC. Table 3.1 summarizes the main protocol stack options used for OTT HTTP adaptive video streaming. It compares HTTP/1.1, HTTP/2, and HTTP/3 (QUIC) in terms of their application protocols, transport layers, security mechanisms, and typical usage in commercial video platforms.

HTTP over TCP is the most commonly used protocol for OTT video delivery. An alternative combination, QUIC over UDP, is used by the leading video platform YouTube, accounting for 27.1% of global traffic (Sandvine, 2021). QUIC was introduced in (Roskind, 2013), described in greater detail in (Carlucci et al., 2015; Langley et al., 2017), and compared with TCP in (Bhat et al., 2017).

3.1.2. Video Delivery Methods

Current Internet video delivery platforms primarily rely on three protocols: HTTP Progressive Download (HPD) and two major implementations of HAS: Dynamic Adaptive Streaming over HTTP (DASH) and HTTP Live Streaming (HLS). In HPD, a video file is downloaded as a regular file using HTTP from a web server. The client can start playback while the video is downloading. However, HPD play-

Table 3.1. OTT adaptive streaming protocol stacks

Stack	App. layer	Transp.	Security	OTT usage/notes
HTTP/1.1	HLS, MPEG-DASH	TCP	TLS 1.2/1.3	On-demand dominant; CDN/cache-friendly; TCP head-of-line blocking.
HTTP/2	HLS, MPEG-DASH	TCP	TLS 1.2/1.3	Multiplexed streams over one TCP connection; still affected by TCP-level head-of-line blocking on loss; used by Netflix and Prime Video.
HTTP/3	HLS, MPEG-DASH, LL-HLS, LL-DASH	QUIC (UDP)	QUIC with TLS 1.3	No TCP head-of-line blocking; per-stream loss isolation; 0-RTT/1-RTT; connection migration; widely used by YouTube and Meta, with Netflix rollout.

back may be interrupted under conditions of low bandwidth or high packet loss. This results in playback rebuffering or stalling. Additionally, HPD downloads the entire video file at the maximum available rate and stores it locally. If the user exits early, the data downloaded but unviewed data is wasted (Xinggong et al., 2019). HAS addresses these issues by adapting video delivery to current network conditions (Seufert et al., 2014) (Seufert et al., 2014). Today, most CDNs (e.g., YouTube, Netflix, Twitch) employ proprietary or standardized HAS implementations (Pires & Simon, 2014; Roettgers, 2013b). HAS requires that the video be available in multiple bit rates (i.e., in different quality levels or representations) and divided into small segments, each containing a few seconds of playtime. The client measures the current bandwidth and/or buffer status and requests the next video segment at an appropriate bit rate, ensuring that stalling (i.e., interruptions due to empty playout buffers) is avoided and the available bandwidth is efficiently utilized (Seufert et al., 2014). Figure 3.1 illustrates the principle of adaptive HTTP streaming, where multiple video representations with different bit rates are stored on the server and the client dynamically switches between them depending on the estimated network throughput. The main solutions are HTTP Live Streaming (HLS), introduced by Apple in 2009 (“HTTP Live Streaming Overview”, 2016), and MPEG-DASH, standardized in 2012 (International Organization for Standardization, 2012). Video delivery method selection depends on the combination of CDN capabilities and the user’s browser or app environment (e.g., iOS app, Android browser, desktop browser). Adaptive streaming enables dynamic bit rate adjustment, typically by switching between video resolutions encoded with different codec profiles. This capability is crucial for efficient video delivery and maintaining high QoE, as it depends on the flexibility and performance of the video codec.

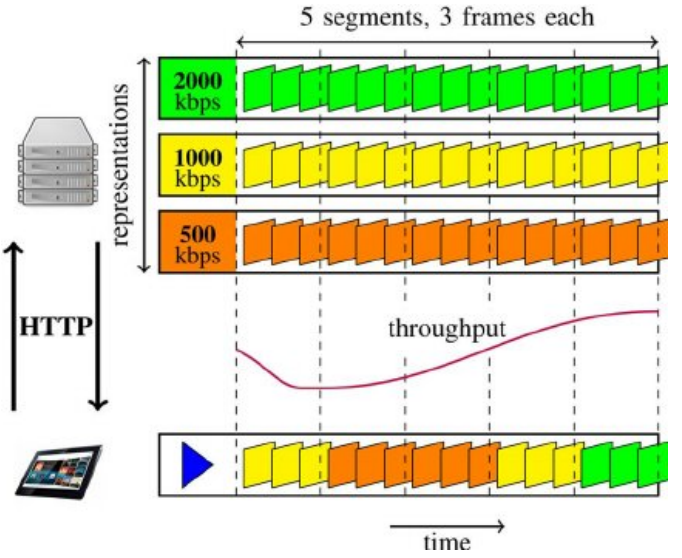


Fig. 3.1. Principle of adaptive HTTP streaming (Seufert et al., 2014)

3.1.3. Video Codecs and Resolutions

OTT video content can be delivered using several different codecs. From the perspective of bandwidth usage and user experience, codecs differ in their supported resolutions and bit rates. This section highlights the most commonly used codecs on the web: H.264/AVC, H.265/HEVC, VP8, VP9, AV1, and H.266. The Advanced Video Coding (AVC) standard, part of the MPEG-4 specification suite, is also defined as ITU-T H.264 and MPEG-4 Part 10. This remains the most widely used video codec to date. Its relatively low encoding and decoding complexity remains a practical advantage (Layek et al., 2017). The High Efficiency Video Coding (HEVC) codec is standardized as ITU-T H.265 and MPEG-H Part 2. HEVC was designed to support efficient encoding and decoding of video at high and ultra-high resolutions. Theoretically, HEVC can reduce file sizes by up to 50% compared to AVC while maintaining comparable image quality (MDN contributors, 2026). The AOMedia Video 1 (AV1) codec is an open format designed by the Alliance for Open Media specifically for Internet video. It achieves higher compression efficiency than VP9 and H.265/HEVC, and up to 50% higher than AVC. However, support for AV1 is still limited across devices, and it requires significant computational resources from end-user equipment. The Video Processor 8 (VP8) codec was initially created by On2 Technologies. In terms of quality and compression rate, VP8 is comparable to AVC(MDN contributors, 2026). Video Processor 9 (VP9) is the successor to the older VP8 standard developed by Google. Like VP8,

VP9 is entirely open and royalty-free. Its encoding and decoding performance is comparable to or slightly faster than that of AVC, but with better quality. VP9's encoded video quality is comparable to that of HEVC at similar bit rates (MDN contributors, 2026). Versatile Video Coding (VVC) (Wieckowski et al., 2021), standardized as H.266 in 2020 by the Joint Video Experts Team (JVET), is the successor to HEVC. VVVC aims to reduce bitrate by up to 50% compared to HEVC while maintaining the same subjective quality, making 4K/8K VoD delivery and high-frame-rate content (e.g., 120 fps sports or VR) more efficient and cost-effective. The standard introduces flexible block partitioning, advanced intra/inter prediction, and dedicated tools for screen content and 360° video. Although commercial deployments are still in their early stages major codec vendors and some streaming trials (e.g., Fraunhofer VVC test channels) have demonstrated real-world gains, positioning VVC as a key option for future high-resolution VoD libraries. Even with efficient codecs and resolution adaptation, video delivery places demanding requirements on mobile networks, including OFDM-based LTE and NR. Table 3.2 provides indicative bit-rate ranges for typical resolutions used in HTTP adaptive streaming.

Table 3.2. Typical video bit rates (Mbit/s) for HTTP adaptive streaming

Codec	360p	480p	720p	1080p	4 K	8 K
H.264/AVC	1.0	2.5	5.0	8.0	20	75
H.265/HEVC	0.6	1.3	2.5	4.0	10	40
VP9	0.7	1.4	3.0	4.0	12	45
AV1	0.5	0.8	2.0	3.0	9	30
H.266/VVC	0.3	0.5	0.8	1.5	8	20

3.1.4. Video Stream Delivery Over Cellular Networks

This subsection analyzes how HTTP adaptive video streams are transported over OFDMA-based cellular networks, with a focus on LTE employing carrier aggregation. Packet-level captures are combined with drive-test radio measurements to link application-layer chunking with MAC and PHY resource utilization and resulting QoE outcomes.

To evaluate video delivery over LTE, field measurements were performed using dedicated mobile network drive test equipment. This setup allows for real-time monitoring of radio conditions, throughput, handovers, and other key network metrics while moving through various coverage areas. Wireshark was also used to track the flow of data packets and compare network protocols. Typical

HTTP adaptive streaming transmission consists of two phases: initial buffering and video playback. In adaptive streaming, data chunks, which carry video, are relatively small compared to a large file download. As the RTT in commercial LTE networks is relatively high, the download throughput may be limited by TCP performance and may not fully utilize the available radio link bandwidth. This leads to longer LTE or NR channel occupancy time.

Figure 3.2 shows the bit rate of a 720p video stream obtained from an IP packet capture.

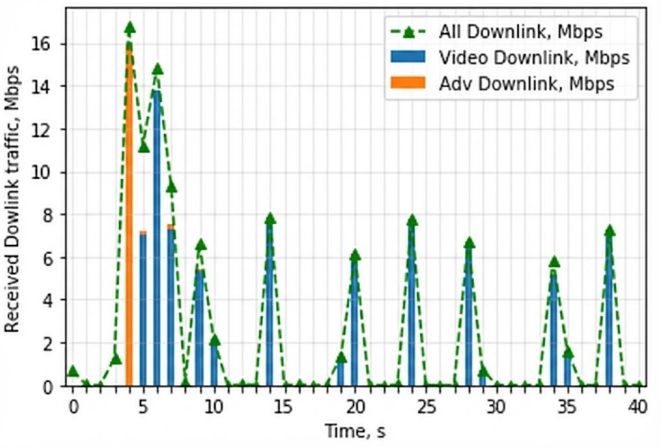


Fig. 3.2. Downlink bit rate aggregated over 1 s intervals, showing total received traffic, pre-roll advertisement traffic, and video stream chunks delivery (created by the author)

The plot distinguishes the bit rate contributions of the pre-roll advertisement and the main video stream. At a one-second aggregation level, the initial buffering phase shows longer transmission bursts and higher bit rates. From the 4th to the 7th second, 28 Mbit of data was downloaded during the initial buffering phase. Subsequently, every 5 s, new chunks of approximately 7.5 Mbit were downloaded to maintain continuous video playback. This example illustrates a typical non-real-time stream (with approximately 30 seconds of delay), similar to on-demand playback, where a large initial buffer is followed by periodic downloading of video chunks. In Figure 3.3, the same data rate is analyzed from the perspective of a multi-carrier LTE network. It shows the downlink MAC throughput for 720p video streaming, obtained from LTE drive-test measurements. In LTE, at the MAC layer, data is distributed among multiple component carriers (CCs). Each video chunk is transmitted using carrier aggregation. Due to the small size of each chunk, the amount of data carried in the secondary cells (SCCs) is relatively small compared

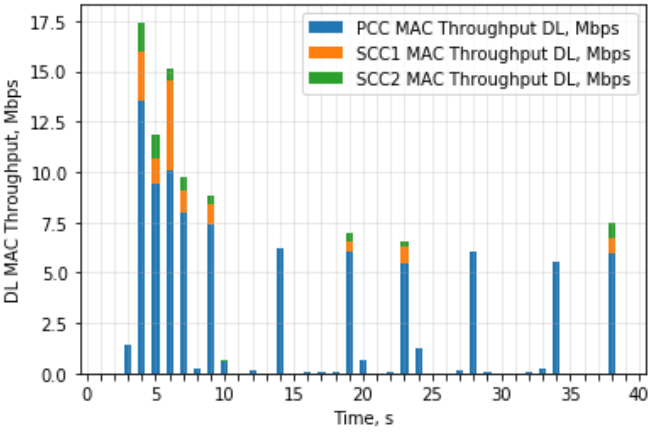


Fig. 3.3. LTE downlink MAC throughput aggregated over 1 s intervals, illustrating limited secondary-cell contribution despite three-carrier operation during video chunk downloads (created by the author)

to the primary cell (PCC). In the test video, the first SCC carried 12% and the second SCC 6% of the MAC-layer data. Because each chunk requires completion of the carrier activation procedure, efficiency losses occur in multi-carrier LTE networks when available bandwidth exceeds the video bit-rate demand. Typical video streaming has high bit-rate requirements, which results in greater utilization of LTE or NR time–frequency resources. Video resolutions and codecs impose different bit-rate requirements for maintaining good QoE during streaming. To avoid stalling and maintain the target resolution, the available bit rate must consistently exceed the stream's average bit rate. Figure 3.4 shows an example of a VP9-encoded 1080p video stream over an LTE cell. As the user moved into an area with lower SINR, all available PRB resources were allocated to maintain the video stream. After several seconds of reduced bandwidth, caused by PRB exhaustion and low CQI, the playback buffer was depleted, resulting in stalling.

Table 3.3 shows the bit-rate requirements of the test video streamed at 1080p resolution. The average downlink bit rate varied from 2.5 Mbit/s using the AV1 codec to 6 Mbit/s with legacy AVC. Uplink requirements differ between codecs, as larger downlink chunks require more acknowledgments. There is also a clear difference in uplink load between QUIC and TCP. TCP guarantees the reliability of downlink packets, but at the cost of higher uplink utilization. Cellular network coverage is typically uplink-limited, especially in LTE and 5G TDD configurations. Using QUIC may offer advantages in uplink resource efficiency and improve effective network coverage for video delivery.

The downlink bit-rate requirements for the most commonly used codec, H.264,

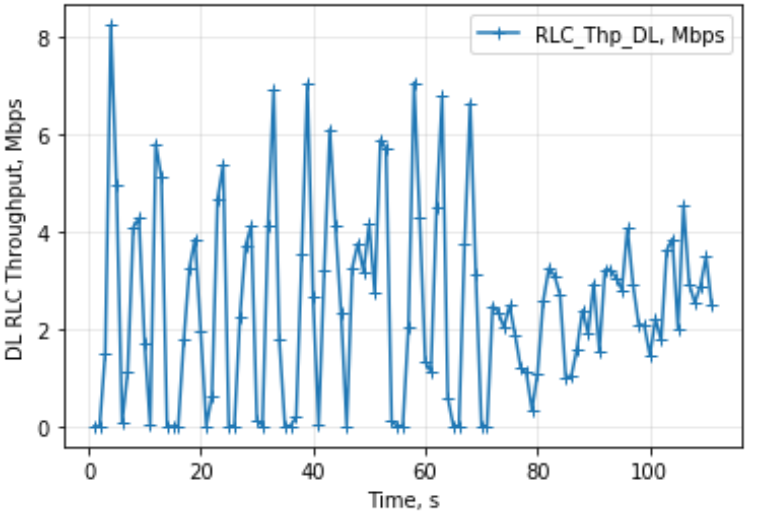


Fig. 3.4. Downlink bit rate aggregated over 1 s intervals, showing the transition from sufficient to insufficient bandwidth for smooth video streaming (created by the author)

Table 3.3. Measured average bit rate by codec (1080p)

Video stream	RLC DL (Mbit/s)	RLC UL (Mbit/s)
IE–TLSv1.2/TCP–AVC–1080p	6.04	0.20
Chrome–QUIC/UDP–VP9–1080p	4.15	0.03
Firefox–TLSv1.2/TCP–VP9–1080p	4.08	0.07
Chrome–QUIC/UDP–AV1–1080p	2.84	0.02
Firefox–TLSv1.2/TCP–AV1–1080p	2.58	0.05

range from approximately 2.5 Mbit/s for 480p, 5–6.5 Mbit/s for 720p, 8–10 Mbit/s for 1080p, and up to 85 Mbit/s for 4K video streaming. To evaluate the cost of video streaming for mobile network operators, average bit rate requirements can be measured and compared to PRB utilization. The LTE UE was locked to an LTE 1800 cell with 20 MHz bandwidth (100 PRBs). During the test, a constant bit rate was generated to represent average bit-rate requirements for typical video streams (ranging from 1 Mbit/s to 40 Mbit/s). Figure 3.5 shows the relationship between SINR and the minimum number of PRBs required for successful video delivery. The values were obtained from field measurements conducted on three different mobile operators using various LTE equipment vendors. When SINR is low, even a low-bitrate video encoded with an efficient codec can consume a large number of LTE PRBs. For example, with a 5 Mbit/s bit-rate requirement, a 20 MHz LTE cell can accommodate only two users streaming video simultaneously. Based on mo-

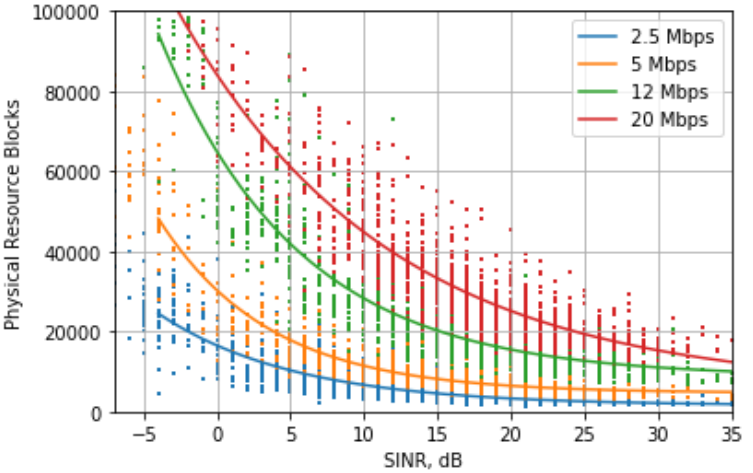


Fig. 3.5. LTE Physical Resource Block requirements for sustaining target bit rates in video streaming (created by the author)

bile network data, 5% of CQI reports are below 5 and 10% below 6, corresponding to SINR values below 3 dB. This indicates that many video streams are delivered to users in poor coverage areas, resulting in high network resource consumption and an increased risk of stalling.

3.2. Uplink Real-Time Video Streaming Quality of Experience

With the advent of 5.5G technology, the paradigm in cellular networks is shifting, placing equal emphasis on uplink video streaming as on downlink, recognizing its growing importance in applications ranging from content creation to teleoperation. However, until now, the major improvements in 5G NR, such as the use of massive MIMO, high carrier bandwidth, higher-order modulations, and carrier aggregation (CA), are primarily targeted toward enhancing the downlink direction. Commercial cellular networks efficiently employ spatial multiplexing gains of massive MIMO in the downlink direction, typically aggregating four layers to a single UE. However, most UEs are capable of only one spatial multiplexing layer in the uplink direction. The large bandwidth in C-band NR is used in a Time Division Duplex (TDD) scenario, typically with a 4:1 subframe configuration, where the majority of the transmission time is in the downlink direction. Although IMT-2020 specifies a user-experienced data rate requirement of 50 Mbps in the uplink

direction (Shafi et al., 2017), commercial networks are more limited due to coverage and capacity constraints. Coverage is a key consideration for operators as it directly impacts service quality, capital expenditures (CAPEX), and operational expenditures (OPEX) in commercial deployments. Uplink coverage is often found to be the bottleneck in these deployments (Lin, 2022). These statements are supported by 5G field measurements. Figure 3.6 presents the measured downlink and uplink throughputs on a commercial 5G cell without traffic load. The measurements show that the uplink throughput rapidly degrades as the signal strength decreases, while the downlink throughput remains high.

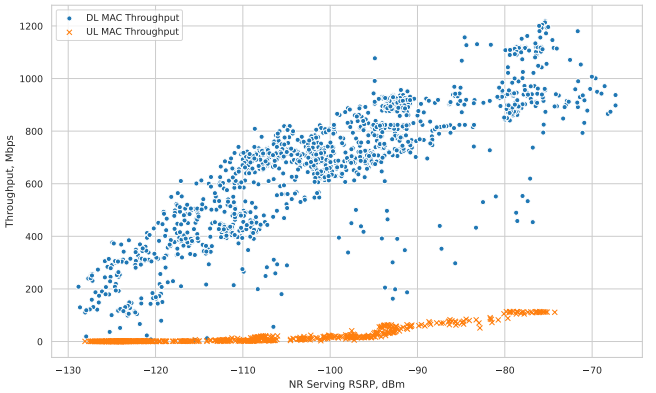


Fig. 3.6. Downlink and Uplink Throughput of a 5G Cell. Measured in an empty C-band cell with 100 MHz bandwidth (created by the author)

Another key advantage of downlink video streaming over uplink live streaming in achieving high QoE is the use of HAS. Although the adaptation techniques for downlink Video on Demand (VoD) and live streaming are well developed (Kua et al., 2017), the same cannot be said for uplink video stream adaptations. Video bitrate and resolution control for real-time multimedia streaming remain relatively underexplored. This disparity, combined with the downlink-focused designs of 5G NR and the limited maturity of video streaming adaptations, poses additional challenges in ensuring high QoE for real-time multimedia streaming from the perspective of MNOs. These challenges become even more complex by the dynamics of a wireless channel, which constitutes the critical last mile in the delivery of data, video streaming, and voice services.

As new applications such as 5G-connected teleoperation for Industry 4.0 and smart cities emerge and gain popularity, there is an increasing focus on uplink

video streaming. This marks a shift from the traditional emphasis on downlink-oriented services and networks. To address video delivery challenges in limited-scale private 5G networks, network slicing and Mobile Edge Computing (MEC) (Wijethilaka & Liyanage, 2021) can be used. However, this approach is suitable only for specific use cases. For broader applications, commercial networks must be able to provide a similar level of service on a public scale. This includes offering network slices that guarantee a certain QoE level. These performance indicators must be met not only by private networks, but also by commercial and public networks.

Given these constraints and the growing need to ensure and assess QoE in uplink video streaming, a method of uplink video streaming QoE evaluation in cellular networks is proposed. This investigation utilizes data available to MNOs to understand and improve the uplink real-time video streaming experience.

3.2.1. Video Stream Ingestion Protocols

To meet the growing personal and industrial demand for live video streaming, various protocols are developed and optimized for low-latency applications. RTMP and WebRTC are the most widely used protocols on the public Internet, while emerging protocols such as MoQ and SRT are gaining traction. Table 3.4 summa-

Table 3.4. Comparison of video ingestion protocols

Protocol	Trans.	Adapt.	Adopt.	Latency
RTMP	TCP	TCP CC	High	1–5 s*
WebRTC	UDP	GCC	Moderate	<1 s
SRT	UDP	ARQ	Growing	1–2 s*
MoQ	QUIC	QUIC CC	Low	<1 s

rizes the main characteristics of several commonly used video ingestion protocols. Protocols are compared with respect to their transport layer protocol (Trans.), congestion control or adaptation mechanism (Adapt.), industry adoption level (Adopt.), and typical latency range (Latency). Latency values marked with an asterisk (*) indicate that sub-second latency is achievable under optimized conditions, such as reduced buffering or retransmission windows.

RTMP (Incorporated, 2012) is a well-established protocol originally developed by Macromedia (now Adobe). Historically, it became the de facto standard for video contribution over the public Internet. It uses the reliable TCP transport and traditional congestion control algorithms such as Reno or CUBIC. Under typical configurations, latency ranges from 1 s to 5 s depending on network conditions

and buffering settings. However, when a precisely tuned sub-second latency can be achieved. While newer protocols such as SRT and WebRTC offer advanced features for ultra-low-latency or highly resilient streaming, RTMP remains a practical choice for many real-time streaming workflows due to its broad compatibility, long-standing adoption, and inherent reliability of TCP.

WebRTC is a protocol suite designed for low-latency, browser-based, peer-to-peer multimedia exchanges (Alvestrand, 2021). WebRTC is widely adopted due to its native in-browser support, open-source nature, and backing from major companies such as Apple, Google, Microsoft, and Mozilla. Additionally, 3GPP includes WebRTC as part of the 5G Real-time Media Communication Architecture (European Telecommunications Standards Institute, 2024), and designates it as the foundation for cellular streaming congestion control in teleoperation (Johansson et al., 2018). WebRTC employs UDP as its transport layer and utilizes Google Congestion Control (GCC) (Holmer et al., 2016), a hybrid algorithm that combines delay-based and loss-based metrics to adjust the sending bitrate. The delay-based controller measures variations in arrival and departure times, feeding these observations into a Kalman filter and an "overuse" detector that flags network queue buildup. If overuse is detected, the sending rate is rapidly reduced (e.g., via multiplicative decrease). Otherwise, the rate increases either multiplicatively or additively, depending on how "close" the algorithm determines it is to the available network capacity. The loss-based controller further refines the rate by monitoring packet loss ratios: a small loss (<2%) prompts a gradual rate increase, while a loss above 10% triggers a significant rate reduction. While bandwidth adaptability can alleviate congestion, in rapidly changing RF environments, it may cause excessive QoE degradation.

Secure Reliable Transport (SRT) is a UDP-based protocol designed for secure, low-latency video delivery. Originally developed by Haivision, SRT is built on top of UDP and employs mechanisms such as Automatic Repeat Request (ARQ), packet acknowledgments, and end-to-end latency management. SRT supports Forward Error Correction (FEC) and selective packet retransmission via ARQ, offering the flexibility to use either mechanism or both simultaneously. This adaptability makes SRT adaptable for use cases requiring either ultra-low latency or maximum reliability (Sharabayko et al., 2021). Although SRT lacks native support in web browsers, it is widely adopted in broadcast and professional media workflows.

Media over QUIC (MoQ) is an emerging protocol under active development by the Internet Engineering Task Force (IETF) that aims to enable low-latency and scalable media streaming over QUIC (Curley et al., 2024). MoQ uses QUIC's native congestion control for efficient network adaptation. MoQ enhances network

adaptation further through priority and grouping mechanisms. Each track or subgroup is prioritized, ensuring that higher-priority content is transmitted first during network congestion. Although MoQ remains in draft form and subject to change, its long-term goal is to unify ingestion and delivery protocols into a single QUIC-based solution. This approach seeks to streamline media pipelines and minimize end-to-end latency for interactive real-time applications.

This study focuses on RTMP and WebRTC as they offer well-established, widely adopted solutions while representing two distinct transport paradigms: TCP and UDP. RTMP is widely recognized for its extensive adoption, inherent reliability due to TCP, and compatibility with consumer-grade streaming tools, making it a strong baseline for reliability-centric ingestion. WebRTC is supported within the 3GPP ecosystem and serves as a foundation for Ericsson's SCReAM (Johansson et al., 2018) congestion control, which is designed to achieve sub-second latency and support interactive applications. By selecting these two protocols, a reliability-focused approach (RTMP) and a low-latency model (WebRTC) are captured, aligning with prevalent industry practices. Although the analysis focuses on these protocols, the same measurement methods and QoE assessment methodology, particularly in media-layer QoE evaluation, can be readily applied to emerging or alternative protocols, such as SRT, MoQ, and other low-latency architectures. This broad applicability ensures that this study's findings and approaches remain relevant to both current and future uplink video streaming scenarios.

3.2.2. Selection of Cellular Network Data for QoE Assessment

Understanding the architecture and interfaces of the cellular network, as described in Section 1.2., together with the protocol stack, is key to evaluating QoE from the point of view of the MNO. A particularly data-rich interface is the air interface, which connects the UE to the RAN. In LTE and 5G networks, this interface, known as "Uu", is responsible for transmitting both payload and signaling/control information. The Uu interface not only transfers payload data encapsulated at the application layer to lower-layer protocols such as TCP/IP but also transmits information such as RF channel quality, downlink and uplink buffer status, and more. Additionally, it handles the transfer of LTE and NR signaling messages for various operations, including scheduling, call setup and termination, carrier aggregation, handovers, and control information.

Collecting payload and signaling data between the UE and RAN provides valuable insights. Beyond standard signaling and payload, MNOs can request UEs to perform network measurements and report the findings. The 3GPP's Minimization of Drive Test (MDT) initiative enables the collection of signal strength, qual-

ity data, and global navigation satellite system (GNSS) coordinates. However, processing and analyzing the vast data from the Uu interface is essential for obtaining meaningful insights. When combined with UE signaling and positioning data, such as MDT reports with GNSS coordinates, this analysis provides valuable insights into QoE across different network regions.

Although 3GPP has initiated the standardization of data collection through the Network Data Analytics Function (NWDAF), it has yet to achieve commercial maturity. Collected payload data, potentially rich in KQI, often encounter encryption barriers, making direct evaluation challenging. Techniques such as TCP/IP packet parsing and packet inter-arrival metrics are employed, but these methods struggle with end-to-end encryption.

Given the comprehensive nature of the Uu interface data, the study focuses on RF parameters as QoE estimators in the uplink real-time video streaming scenario. Furthermore, video media layer information and IP packet capture (pcap) are incorporated to enhance the understanding of protocol behavior during live video streaming.

3.2.3. System Architecture for Data Collection

To establish an objective model for live streaming video QoE, a data collection architecture was developed that integrates actual network elements with streaming servers hosted on a public cloud platform.

As illustrated in Figure 3.7, the architecture enables the assessment of objective FR video quality metrics, including VMAF, PSNR, and SSIM. Using this setup, the metrics are evaluated for uplink live video streaming over a commercial cellular network. Network measurements are collected in parallel to analyze the impact of varying network conditions on QoE. The first component in the data collection architecture is video streaming and recording. To assess video quality degradation using FR methods, both the original and impaired versions of the video content are preserved. Impairments occur due to transmission over the cellular network or transcoding to lower resolutions and bit rates.

A 4K webcam was mounted on the dashboard of a car to capture the video feed. This camera connects to OBS Studio, an open-source video recording and live-streaming platform. OBS Studio manages streaming parameters and embeds timestamps for subsequent quality and latency analysis. The used streaming protocols were OBS Studio for RTMP streams and an adapted version (RTC Studio) for WebRTC streaming.

In this configuration, the zero-latency setting was enabled, generating only I- and P-frames (with B-frames disabled) and discarding delayed frames. The

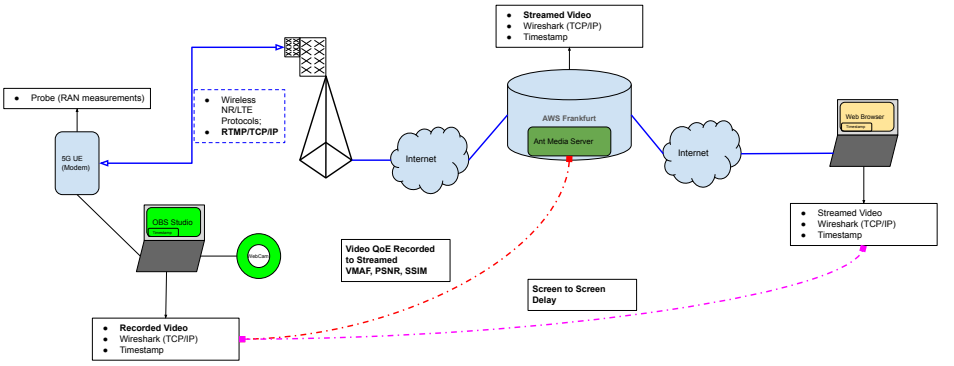


Fig. 3.7. RAN measurement collection architecture for uplink real-time video QoE evaluation (created by the author)

video recorded locally via OBS Studio serves as the baseline for objective QoE comparisons.

Real-time uplink video streaming is investigated under diverse environmental and network conditions, addressing critical requirements for teleoperation and smart city architectures. To capture realistic behaviors, tens of hours of drive tests were conducted across various environments, including urban districts, suburban neighborhoods, and highways, as shown in Figure 3.8. This multi-environment approach allowed for observing a broad range of RF conditions (e.g., path loss and SINR), varying levels of cell overlap during handovers, and the effects of high mobility on service continuity. Before conducting the drive tests, 15-minute aggregated network statistics were analyzed across the operator's entire network to assess typical uplink resource usage. Moreover, selected base stations' high-temporal-resolution traces with 1-second granularity were analyzed. Unlike the downlink, which often experiences congestion, uplink PRB overload is rare and remains relatively stable over daily cycles. Therefore, the dataset focuses on coverage-limited scenarios, where SINR and path loss are the primary factors affecting QoE. Consequently, our drive tests focus on coverage-limited behavior (dominated by path loss and SINR constraints) and mobility-induced impairments. Since uplink performance is more often constrained by coverage (power limitations and path loss) rather than congestion, the UE was deliberately locked to single cells in certain tests and gradually moved away from the site. This approach covers a wide range of signal strengths and SINR values. Real-time services, such as teleoperation, are highly sensitive to mobility events, such as handovers. Drive-test routes were selected to cross multiple cell boundaries, enabling the capture of interactions among handovers, latency spikes, and throughput drops under real-world conditions. The live video stream is transmitted to the Ant Media Server

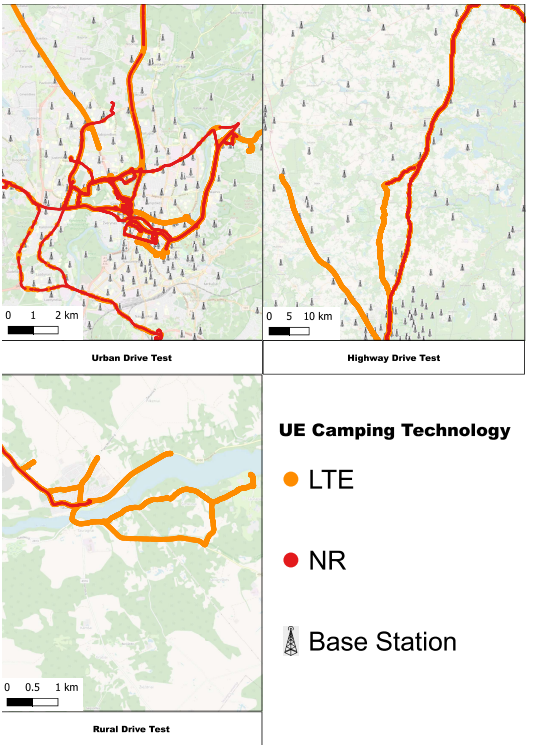


Fig. 3.8. Drive-test routes across urban, rural, and highway environments used for uplink video streaming measurements (created by the author)

located within the Amazon Web Services (AWS) cloud infrastructure. The streaming configuration is optimized for low latency, ensuring that most delays are the result of network conditions rather than video encoding.

In this setup, the streaming server not only facilitates video delivery to viewers but also archives the video in AWS cloud storage. The archived video, which has undergone network transmission impairments, serves as a distorted reference for FR quality evaluation.

The overall data delivery path consists of multiple interconnected elements. Video feeds are captured via OBS Studio and encapsulated into RTMP/TCP and WebRTC/UDP transport protocols. The setup consists of a PC running OBS Studio, connected via USB to a Huawei P40 Pro UE, which functions as a modem. This UE selectively uses various network technologies, including LTE and NR, and specific frequency bands such as 800 MHz (LTE Band 20) and 1800 MHz (LTE Band 3), or specific cells. The video data is transmitted to the cellular network base station through the LTE and/or NR air interfaces. Tests are conducted

by locking the UE onto specific frequency bands such as L800 at 10MHz and LTE1800 at 20MHz, either to a single cell or by allowing the UE to move freely in an NSA scenario. This approach captures a broad spectrum of SINR, Reference Signal Received Power RSRP, and other network parameter values. Beyond the base station, IP packets are routed by IP RAN routers to the core network, as shown in Figure 1.1. Subsequently, the MNO network interfaces with the external network infrastructure, ultimately connecting to an AWS server. Although this segment of the data path is not monitored during this research, it is assumed that wired connections are stable, with any fluctuations reflecting real commercial network conditions. To ensure data alignment, IP packets are tracked using packet capture on the local PC and AWS server.

Furthermore, as illustrated in Figure 3.7, the setup includes a wired connection to another PC for stream monitoring and screen-to-screen latency measurements in selected scenarios.

In addition to evaluating objective QoE, an equally important aspect of this research is conducting comprehensive network measurements. These measurements capture similar types of data that MNOs can obtain. MNOs can conduct similar drive tests to analyze related parameters in a given geographical area. Additionally, crowdsourcing UE Measurement Reports (MR) collection by cell site traces or results with accurate MDT geolocation can be used. Furthermore, low-granularity measurements can be collected from the eNodeB and gNodeB performance statistics.

A designated computer functions as a dual-purpose node, locally recording the video to establish a reference for the original stream. Additionally, it runs Wireshark, a network protocol analyzer, to capture IP packets from the live stream. This approach allows for gathering comprehensive data on network traffic, which is important for correlating video quality with network behavior.

To enhance the network data collection, this computer is connected to a UE that also serves a dual purpose, acting as a modem and conducting comprehensive cellular network measurements. Commercial drive-test software is used on this setup to record a wide array of RAN parameters. These parameters include, but are not limited to, signal strength, signal quality, allocated PRBs, path loss, instantaneous throughput, bit error rate, and other vital RF metrics. The insights gained from these RAN measurements are the basis for understanding the network's behavior and performance under diverse environmental conditions and user scenarios. This depth of information is crucial for MNOs in the context of QoE evaluation of real-time video streaming services.

Collectively, the data from IP packet captures and RAN measurements form a comprehensive dataset that mirrors the type of information MNOs typically gather

and analyze. Consequently, this research is not only relevant for academic purposes but also directly applicable to real-world network operations and optimizations.

3.2.4. Screen-to-Screen Delay

This research evaluates screen-to-screen delay across various network environments. It is achieved by using an auxiliary computer connected to the Internet via a reliable wired connection. Figure 3.9 illustrates the method for screen-to-screen delay assessment.

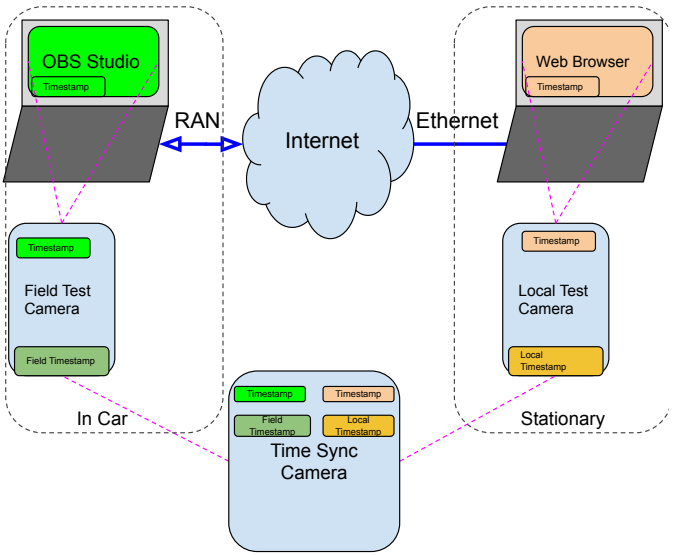


Fig. 3.9. Screen-to-Screen Delay Assessment Architecture (created by the author)

The used approach aligns with the methodology of Uitto et al. (Uitto & Heikkinen, 2022) in certain aspects but differs by evaluating a range of coverage scenarios rather than a fixed location.

Along with the live feed from the webcam, a real-time clock is displayed within the OBS Studio user interface. Although the clock runs continuously with 1 ms granularity, the camera operates at 30 fps, with a frame period of approximately 33 ms. The camera frame rate introduces variation in end-to-end latency evaluation. The latency tests use actual video feeds encoded at Variable Bit Rate (VBR). This ensures that the test scenarios closely mimic the load conditions typical of video streams, as opposed to a scenario where only a timer is transmitted, which could result in lower bit rates due to VBR.

The video stream, including an on-screen clock, is transmitted from a vehicle under various RF coverage conditions. It is expected that different RF coverage levels and traffic loads from other commercial cellular network users will affect latency. It is assumed that the latency from the wired connections between the RAN base station, the next hop to the AWS server, and the receiving PC remain consistent and reflect real network behavior.

The video receiver is connected via a reliable Ethernet connection. A web browser and the WebRTC protocol are used to play the video stream. Clock differences are used to evaluate overall screen-to-screen latency.

Clock synchronization across all devices is essential. One method involves placing the transmitting and receiving PCs in the same location and capturing their screens with a third camera. However, this method is only feasible if both parties are colocated.

For the drive-test scenario, additional cameras were used with internal clocks and time-stamps on the recorded videos. A third camera records two smartphones with time-stamped videos to determine a constant time offset. Determining the time offset between the two cameras allows us to record original and received time stamps and calculate the latency caused by network transmission.

$$\Delta T_{\text{internal}} = T_F - T_L. \tag{3.1}$$

The constant time difference between the two cameras is defined in (3.1). The control camera T_C captures both the field camera T_F and the local clock T_L simultaneously. This allows for determining the constant time offset of internal clocks, $\Delta T_{\text{internal}}$.

The network latency, denoted as T_{lat}, is calculated using the original timestamp T_{original} and the received timestamp T_{received}, as shown in (3.2):

$$T_{\text{lat}} = (T_{\text{received}} - \Delta T_{\text{internal}}) - T_{\text{original}}, \tag{3.2}$$

where $\Delta T_{\text{internal}}$ is subtracted from T_{received} to adjust for any internal clock differences before comparing it with T_{original} to obtain the network latency.

To analyze latency, computer vision and Optical Character Recognition (OCR) techniques are used. Initially, video files are split into individual frames using FFmpeg, and then Tesseract OCR (Kay, 2007) is applied to extract timestamp values from each frame.

The proposed screen-to-screen latency evaluation method has multiple limitations and should be used only as a reference for expected screen-to-screen delays. The first limitation is the video frame capture interval of 33 ms, which may not

capture precise latency measurements. The second limitation is the unpredictable nature of Internet routing, which can introduce variability. Finally, internal clock shifts may occur during the measurement periods, potentially affecting accuracy.

3.3. Dataset Collection and Video Quality Assessment

The real-time uplink video streaming collection architecture described in 3.2.3. is used to collect a comprehensive dataset. After collecting network-related measurements and storing video streams, FR video quality assessment is conducted.

To achieve this, VMAF, SSIM, and PSNR are used as objective QoE metrics. This approach directly evaluates the presentation of the media stream. Unlike bitstream-based or parametric evaluations, such as those proposed by (Robitza et al., 2018), this method is expected to yield results that reflect human perception more closely. The core components of the evaluation pipeline include:

- Original video: The original video was recorded directly on a PC inside the vehicle. It served as a benchmark for evaluating the quality of the transmitted video streams.
- Impaired video: To assess the impact of network transmission on video quality, videos transmitted via RTMP or WebRTC over 4G and 5G networks were compared against the original recordings.
- Video Multimethod Assessment Fusion (VMAF): VMAF was used as the primary metric in the QoE assessment. It integrates multiple quality indicators to align with human visual perception. In addition, the Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) were incorporated to provide insights into the structural integrity and noise levels of the video content.
- Supporting measures – detailed video file parameters: Several supplementary metrics were analyzed, including:
  - Frame type: Identifies the type of each frame, including intra-coded (I), predicted (P), and bidirectionally predicted (B) frames.
  - Packet decoding timestamp: Indicates when each frame is decoded and assists in evaluating playback timing and synchronization.
  - Packet size: Represents the data size per frame, providing information on the amount of data transmitted.

– Frame duration: Shows how long each frame is displayed, allowing detection of frame losses or playback stalls.

Objective FR evaluation of video streams affected by network transmission is challenging due to the frame-by-frame evaluation methods of VMAF, PSNR, and SSIM. To conduct an objective FR evaluation, each frame must be time-aligned. This challenge is particularly pronounced for WebRTC, as noted by (García et al., 2019). The following subsubsection details our approach to addressing frame-to-frame misalignment in both RTMP and WebRTC protocols.

3.3.1. Objective Evaluation of RTMP Stream

RTMP, which relies on TCP for reliable transmission, ensures that the data received by the server matches the data sent by the client without loss. Video recording may begin before or after the stream starts, with a potential delay of up to a few seconds. An on-screen clock and OCR technology (Kay, 2007) are used to determine the start of each video segment. However, on-screen clock values may vary. Timestamps may appear either at the start or end of a video frame. Additionally, issues with frame duplication may occur due to encoding stalls or repeated frames. Therefore, direct on-screen clock evaluation to determine the required frame shift for QoE assessment cannot be fully automated. The size of each frame was analyzed using ffprobe. Frames with identical sizes were identified and their positions were determined in recorded and streamed videos. Using this information, the required frame shift was calculated for the FR QoE assessment. This approach allows for large-scale automated assessment.

Figure 3.10 shows the comparison between the packet sizes of streamed and locally recorded video frames. The reliable TCP-based RTMP protocol ensures that each successfully transmitted packet maintains the same size, except for a visible dropout in the streamed packet around the 600^{th} frame due to network conditions. By comparing the sizes of the first streamed and recorded frames, the videos are aligned from the beginning, ensuring that the presentation timestamps remain synchronized throughout the video file.

3.3.2. Objective Evaluation of WebRTC Stream

WebRTC uses UDP and incorporates native adaptive mechanisms to manage the streaming bitrate. It dynamically adjusts to changing network conditions, resulting in variable frame rates throughout the video session. WebRTC's inherent adaptability supports real-time communication but poses challenges for frame alignment

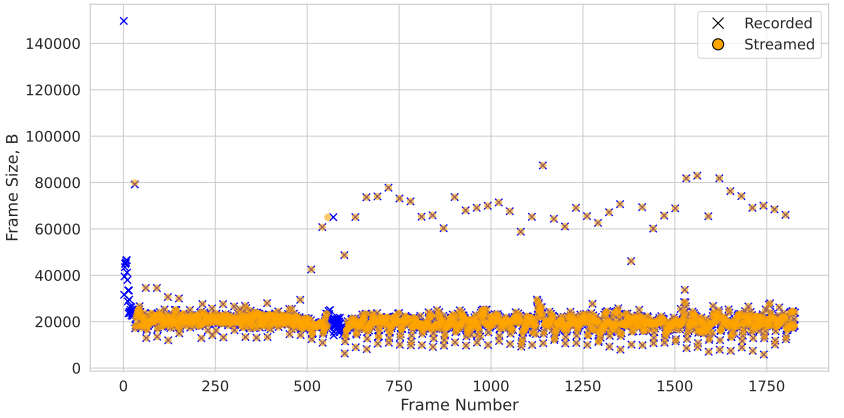


Fig. 3.10. Comparison of video frame sizes for locally recorded and RTMP-transmitted streams (created by the author)

in FR video quality assessment methods. Video recording may begin before or after the stream starts. To accurately determine the start of each video segment, an on-screen clock and OCR are used.

A key challenge in aligning frames for FR methods in WebRTC streams is the variability in frame rates, as shown in Figure 3.11.

This variability is most pronounced during the initial stages of streaming as the system adapts to network conditions. As shown in Figure 3.11, the Presentation Timestamp (PTS) of each subsequent frame increases sharply. Additionally, the duration of each frame exceeds 0.033 ms for a 30-frame-per-second (fps) video. Eventually, the frame duration stabilizes at approximately 0.033 ms, but packet loss can cause it to rise again. Such fluctuations complicate frame alignment, as the frame rate can dynamically change throughout the video session.

To address the challenges posed by frame rate variability in WebRTC streaming, the following strategies are used:

- Normalize Video Playback: One approach is to standardize video playback at a consistent frame rate, such as 30 fps, using frame interpolation.
- Convert Video Frames into Images: Converting video frames into individual images enables a more detailed analysis of each frame's content.
- Focus on Midpoint for Synchronization: Given the initial variability in frame rate, videos are aligned on the midpoint of the stream. This method assumes

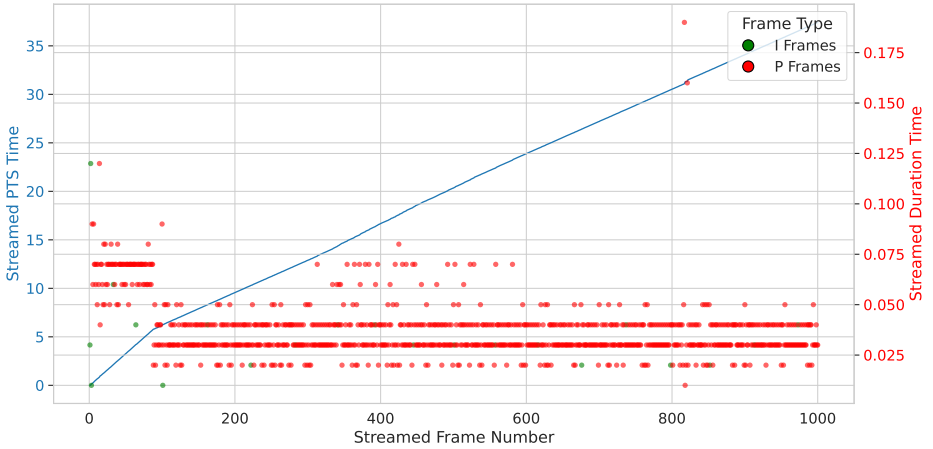


Fig. 3.11. Presentation timestamp and frame duration in a WebRTC streamed video (created by the author)

that by the midpoint, any initial frame rate adjustments have stabilized, providing a more reliable basis for comparison.

- Evaluate QoE on Converted 30 FPS Video: After determining the timing difference between locally recorded and WebRTC-streamed video files, the locally recorded original video is compared with its 30 FPS-converted version.

These strategies collectively mitigate the challenges posed by the variable frame rates in WebRTC streams, ensuring more precise alignment for FR video quality assessment.

3.3.3. Distribution of Dataset Samples

The dataset collected by (Baena et al., 2023) has a similar structure but focuses on downlink VOD delivery. VOD is already well-researched and widely used by billions of cellular network users. In contrast, data on real-time uplink video streaming was collected in realistic urban and rural environments. The dataset is organized into folders, each containing real-time uplink video streams, network measurements, and QoE results. The dataset overview is presented in Table 3.5. The contents are as follows:

- Recorded Video File: The locally recorded original video file serves as a reference for FR evaluation.

- Streamed Video File: Video files are streamed over 5G or 4G networks and stored in the AWS cloud.
- Objective QoE Metrics: Files contain PSNR, SSIM, and VMAF values for every frame. For example, a 30 s video file provides 900 QoE measurements.
- Video Info: Files detail streamed and recorded video parameters, such as PTS, frame duration, frame size, and picture type.
- PCAP Packet Data: Pcap files are collected by Wireshark on a PC where the original video is recorded. In some, cases additional PCAP files are also collected from the AWS server. These files capture data from before the video stream session is established until the stream ends.
- RAN Measurements: A CSV file that contains 4G and/or 5G radio network measurements sampled every second. It includes over a thousand RAN parameters, such as signal strength, signal quality, allocated PRBs, path loss, instantaneous throughput, bit error rate, and other RF metrics.

Table 3.5 presents the key parameters and results of the video streaming tests conducted under different network conditions.

The Resolution (Res.) column specifies the video resolution used during each test. Bitrate (Br.) indicates the streaming and recording rate in megabits per second (Mbps). In these experiments, videos were streamed and recorded at identical resolutions and bitrates to isolate the impact of network conditions. For reference, additional tests were performed at reduced resolutions and bitrates to evaluate the effect of encoding. The Network column denotes the cellular network configuration, specifying whether the UE was locked to a particular frequency band or cell, or operated in an unlocked mode. Protocol (Prot.) identifies the streaming protocol. The Duration (Dur., s) column lists the approximate length of each sample video in seconds, while Number of Samples (N.) indicates how many test samples belong to each category. Finally, the Value (V.) column shows the average VMAF score obtained from each video streaming test. Figure 3.12 presents a violin graph that illustrates the evaluated QoE values using VMAF. The VMAF method assigns a dimensionless score between 0 and 100 to each video frame, where 100 represents quality identical to the reference video. To align with the network measurement periodicity, which is averaged to one second, the average VMAF score per second is also computed.

Violin plots illustrate the density of VMAF scores derived from these one-second samples. Results indicate that the TCP-based RTMP protocol typically

Table 3.5. Video streaming parameters

Res.	Br.	Network	Prot.	Dur., s	N.	V
720p/720p	5	L8 B	RTMP	~30 s	93	88.02
1080p/360p	9/1	L18 C	RTMP	~10 s	118	32.29
1080p/480p	9/2.5	L18 C	RTMP	~10 s	136	42.47
1080p/720p	9/5	L18 C	RTMP	~10 s	103	77.93
1080p/1080p	9	L18 C	RTMP	~10 s	19	82.91
1080p/1080p	9	Unlocked	RTMP	~10 s	30	84.17
1080p/480p	9/2.5	Unlocked	RTMP	~30 s	67	62.07
1080p/720p	9/5	Unlocked	RTMP	~30 s	51	76.89
1080p/1080p	9/9	Unlocked	RTMP	~30 s	28	88.39
720p/720p	5/5	Unlocked	RTC	~60 s	42	76.16
1080p/1080p	9/9	Unlocked	RTC	~60 s	22	33.25
720p/720p	5/5	Unlocked	RTMP	~60 s	38	97.85
720p/720p	5/5	Unlocked	RTMP	~60 s	41	97.62
720p/720p	5/5	Unlocked	RTC	~60 s	26	75.43
720p/720p	3/3	L18	RTMP	~60 s	33	97.53
720p/720p	3/3	L18	RTC	~60 s	16	61.95
720p/720p	3/3	Ether	RTMP	~60 s	6	98.10
720p/720p	3/3	L18N35 B	RTMP	~60 s	6	91.41
720p/720p	3/3	L18N35 B	RTC	~60 s	8	76.85
720p/720p	3/3	Ether	RTC	~60 s	10	89.95
720p/720p	3/3	L8 B	RTC	~60 s	16	80.98
720p/720p	3/3	L18N35 B	RTC	~60 s	19	59.52
720p/720p	3/3	L8 B	RTMP	~60 s	20	98.00
720p/720p	3/3	L18N35 B	RTMP	~60 s	33	97.80

yields either near-perfect scores or very low values, reflecting sharp quality transitions caused by lost frames. In contrast, WebRTC-based evaluations display a more dispersed distribution, influenced by congestion control mechanisms and adaptive frame rate reductions.

Figure 3.13 illustrates the detailed results of a sample RTMP stream, providing an overview of QoE metrics and network performance indicators. Figure 3.13 presents the QoE scores evaluated using the VMAF FR model. The first subplot displays the VMAF scores, providing an objective measure of video quality from the end user's perspective. The next subplot presents the streamed video presentation timestamps and the duration of each video frame. If the duration exceeds 0.033 seconds, it indicates frame loss and stalling. The subsequent chart is derived from PCAP evaluation, focusing on throughput and round-trip time (RTT). An analysis of the bitrate reveals that the requested video bitrate is not achieved,

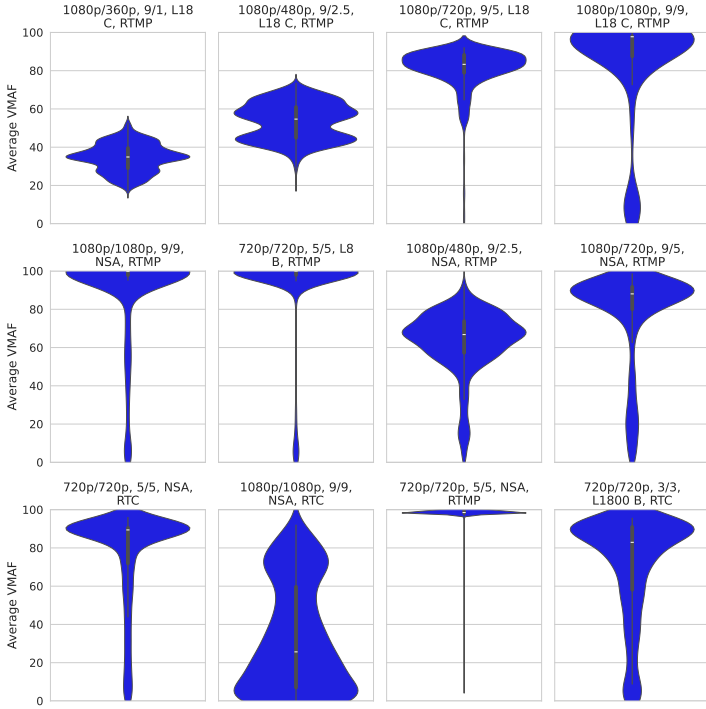


Fig. 3.12. Distribution of collected VMAF scores (created by the author)

suggesting a limitation in TCP performance over a cellular network. The last four subplots present RAN measurements in the NSA scenario, including path loss, allocated PRB numbers, and MCS values.

3.4. Screen-to-Screen Delay Results

End-to-end latency is evaluated in four static scenarios using a similar approach to (Uitto & Heikkinen, 2021) and in four drive-test scenarios using the method described in Section 3.2.4. Table 3.6 summarizes the results. The latency values in Table 3.6 are observed using Tesseract OCR (Kay, 2007) and are sampled for each video frame. To avoid outliers resulting from OCR misrecognition, the Interquartile Range (IQR) method is applied. This method filters out outliers from OCR misrecognition and eliminates artificially inflated screen-to-screen delays caused by repeated lost frames. Values in Table 3.6 present typical screen-to-screen delays.

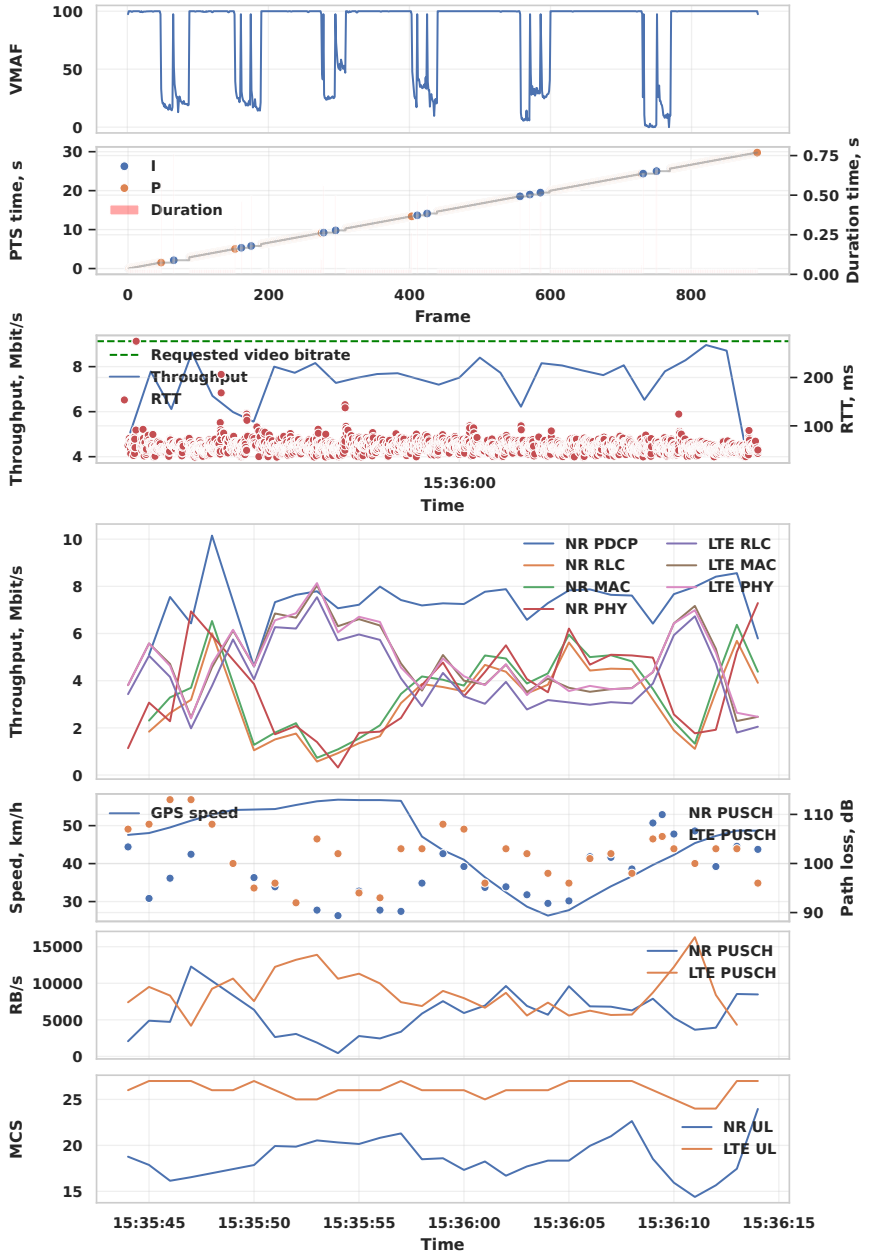


Fig. 3.13. Example of collected QoE and network measurements (created by the author)

Table 3.6. Latency (s)

Test type	Network	Protocol	Mean	Std.
Static	Ether	RTMP	0.1542	0.0233
Static	L18N35 B	RTMP	0.2439	0.0423
Static	L18N35 B	RTC	0.2299	0.0296
Static	Ether	RTC	0.1934	0.0242
Drive test	L8 B	RTC	0.1462	0.0248
Drive test	L18N35 B	RTC	0.1501	0.0232
Drive test	L8 B	RTMP	0.1701	0.0240
Drive test	L18N35 B	RTMP	0.1839	0.0285

The IQR is calculated as:

$$\text{IQR} = Q_3 - Q_1, \tag{3.3}$$

where Q_1 and Q_3 are the first and third quartiles, respectively.

The lower and upper bounds for filtering outliers are defined as:

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR}, \tag{3.4}$$

$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR}. \tag{3.5}$$

E2E latency evaluation method using cameras and timestamps can introduce deviations due to the video sampling rate. The system captures the timestamp either at the beginning or the end of the 0.033-second duration frame. Additionally, in the drive-test scenario, the internal clock of the cameras may experience deviations. Considering these limitations, the presented values are for rough estimation only. A significant factor contributing to screen-to-screen latency is frame loss. If a video frame is lost, the previous frame remains displayed until the subsequent frame arrives.

Figure 3.14 shows the per-frame VMAF score together with screen-to-screen delay and rendered frame duration for an example WebRTC uplink stream. In this trace, most delay samples are around 0.1–0.3 s, with occasional spikes up to approximately 0.6 s. Figure 3.14 shows cases where the video scene is static (the view does not change). The temporal impairments (freezes) may not be reflected by FR metrics such as VMAF because repeated frames can remain perceptually similar to the reference. In contrast, freeze events consistent with lost or late frames are indicated by simultaneous increases in screen-to-screen delay and rendered frame duration. Screen-to-screen delay and frame duration show a Pearson correlation of 0.8. Screen-to-screen delay variation due to lost video frames can be

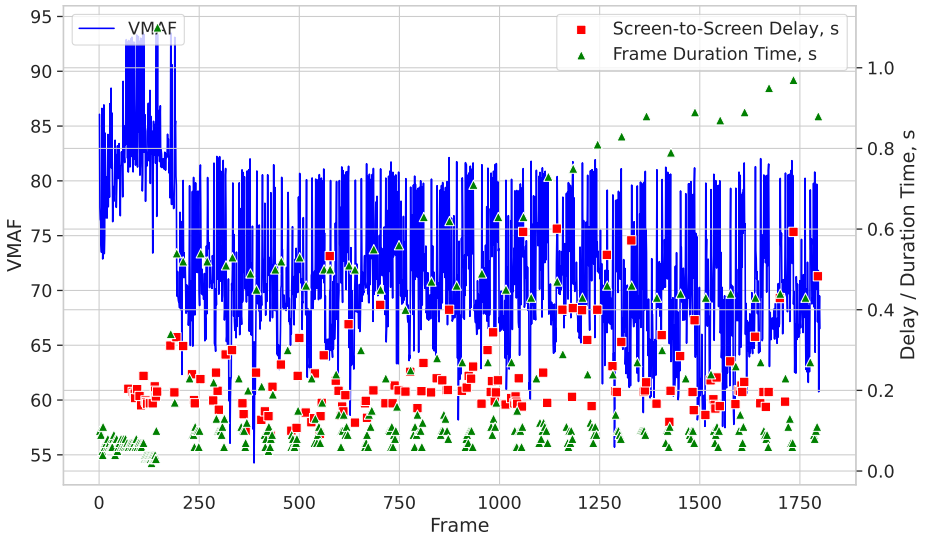


Fig. 3.14. Screen-to-Screen Delay, Frame Duration, and VMAF Score (created by the author)

inferred if the frame timestamps are available. To relate application-level delay to network-path dynamics, a round-trip time (RTT) is also estimated from the PCAP recorded during the same session. Because the video transport is UDP-based and does not expose a TCP-style RTT, periodically exchanged SSH messages are used as an application-layer probe to approximate the network-path RTT. The SSH RTT is measured periodically and varies only between 50–60 ms (source PC to AWS), indicating a relatively stable network-path RTT during the experiment. This RTT shows no meaningful correlation with screen-to-screen delay or frame duration, suggesting that the dominant delay variations originate from media-layer buffering and frame loss effects rather than from RTT fluctuations. Compared to TCP-based RTMP in the same setup, WebRTC yields approximately 20 ms lower screen-to-screen delay, which is beneficial for latency-sensitive applications. However, given the overall delay level in our scenario, this reduction is modest; depending on the application, the higher stability of TCP-based RTMP may be preferable. In this trace, screen-to-screen delay is typically on the order of 0.1–0.3 s, with occasional spikes up to approximately 0.6 s during freeze events. These magnitudes support aligning objective video-quality samples with RAN measurements at a 1 s sampling resolution, since the dominant delay variations occur within sub-second time scales.

Objective FR metrics such as VMAF quantify perceptual similarity but are

less sensitive to temporal impairments (e.g., freezes) in static scenes. Therefore, latency and continuity should be monitored separately using timing information from the media or transport layers. Such cross-layer monitoring is also required to verify compliance with end-to-end tele-operation delay budgets of 100–200 ms for UL video, even when the FR quality score remains high (International Telecommunication Union, 2022).

3.5. Evaluation of Uplink Video Streaming in 4G and 5G Cellular Networks

This section explores the relationship between LTE and NR RF conditions and their impact on QoE. It starts by examining how video transcoding affects VMAF scores. This is followed by an analysis of the achievable bit rates in Orthogonal Frequency Division Multiple Access (OFDMA) and Single Carrier Frequency Division Multiple Access (SC-FDMA) systems to identify the factors influencing QoE.

Detailed evaluations are performed in rural and urban scenarios to assess how RF measurements can be utilized to estimate and enhance QoE in different network environments. The evaluation includes the use of WebRTC and RTMP protocols, different coverage levels, and handover scenarios.

The first test is designed to evaluate the impact of video transcoding on average VMAF scores in a wide range of RSRP, SINR, and path-loss ranges. An LTE 1800 MHz cell is selected with a 20 MHz bandwidth and the UE is locked to the cell. During the driving test, different target bitrates are streamed (1 Mbps, 2.5 Mbps, 5 Mbps, and 9 Mbps) and resolutions (360p, 480p, 720p, and 1080p, respectively) using the RTMP protocol. The video QoE results are presented in Figures 3.15 and 3.16. Each point represents the average value per second. Transcoding to a lower bitrate and resolution results in a lower VMAF score. The overall impact of transcoding is well-studied, and a linear degradation can be observed. Transcoding from an unimpaired 1080p to 720p results in a degradation of approximately 20 VMAF points. Transcoding to 480p leads to a degradation of

Table 3.7. Statistical comparison of video resolutions

Resolution	Average VMAF	STD of VMAF
1080p vs 360p	34.40	7.24
1080p vs 480p	53.26	9.64
1080p vs 720p	80.54	11.75
1080p vs 1080p	86.23	26.44

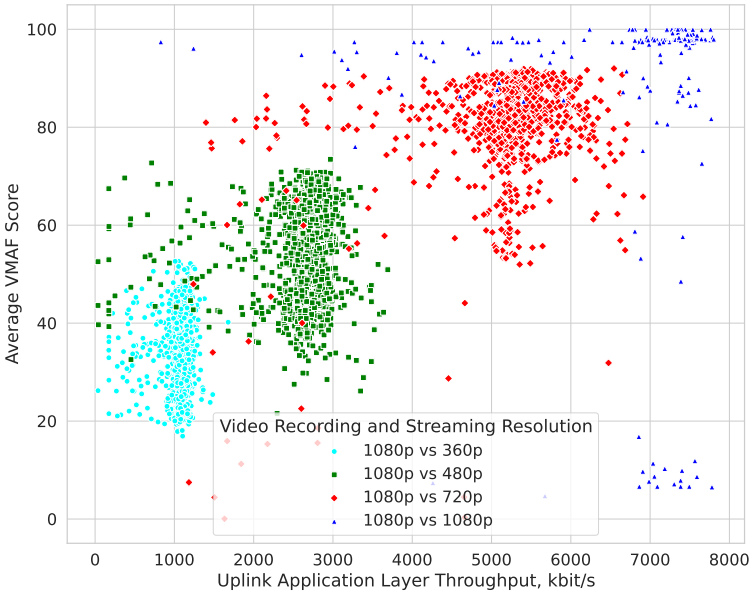


Fig. 3.15. Impact of transcoding (created by the author)

about 45 VMAF points, while transcoding an unimpaired 1080p reference video to 360p causes a degradation of approximately 65 VMAF points. TCP packet and video frame losses are observed for a 720p, 5 Mbps video stream due to insufficient bitrate. As shown in Figures 3.15 and 3.16, the 1080p stream with a 9 Mbps bitrate faces limitations due to TCP interactions during real-time video streaming over the radio network. These limitations arise from a long RTT caused by the cellular base station scheduler and the long distance to the server. Consequently, a significant number of video frames are lost, leading to a low QoE score.

3.6. Analysis and Modeling of Video Based on RF Measurements

3.6.1. 5G New Radio Uplink Throughput

The fundamental mechanisms governing achievable uplink throughput are channel quality modulation, bandwidth allocation, and power limitations. These are detailed in Section 1.3.6. Consequently, the throughput analysis in this study utilizes the definition in (1.2), adapted to reflect instantaneous RF and scheduling conditions rather than theoretical peak capacity.

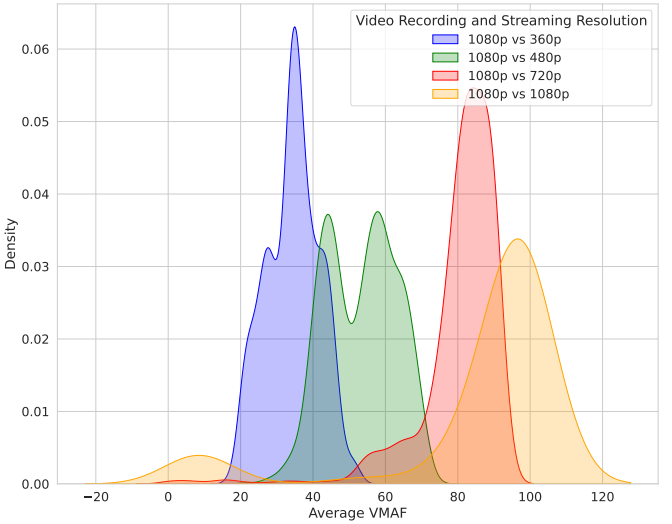


Fig. 3.16. VMAF Probability Density Function for different resolutions (created by the author)

3.6.2. Modeling of Video Quality in a Rural Scenario

A test is designed to evaluate the VMAF score in a rural scenario. The UE is restricted to a typical rural LTE800 cell using a 10 MHz bandwidth and perform multiple drive tests from near the base station to a distance of 4 km. The base station has a low uplink load, which is typical for most LTE FDD cells, so no limitations are expected due to PRBs.

In the test, 720p video recording and streaming is performed. This test evaluates only the effect of video transmission across the network, excluding any effects of video transcoding.

Figure 3.17 shows the relationship between the achieved uplink throughput and the average VMAF score.

Achievable throughput closer to the required video bitrate leads to a higher VMAF score. However, the relationship is not strictly linear due to the nature of the video presentation. In scenarios where the view is static, high objective FR QoE results can still be attained despite packet or frame losses caused by low throughput. This is illustrated by the blue samples, where the vehicle is stationary. Using machine learning (ML) techniques, non-linear regression can be performed to approximate the relationship between throughput and VMAF score. Knowing the achieved UL throughput and the required video bitrate allows for accurate modeling of experienced QoE.

However, the exact required bitrate is unknown to the MNO, as the operator is

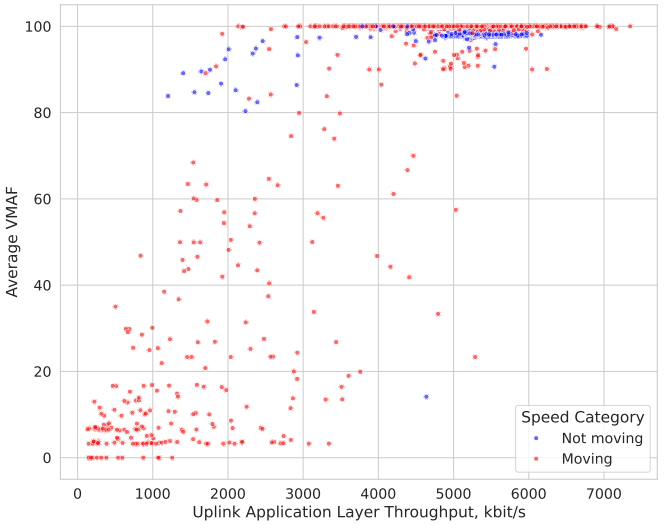


Fig. 3.17. Relationship between the achieved uplink throughput and the average VMAF (created by the author)

not aware of which application is used in a specific area, by a specific customer at any given moment. Still, network performance statistics collect other Uu interface-related measurements that can be used as inputs to model achievable bitrate and QoE.

Figure 3.18 shows the relationship between the main components of uplink throughput.

As the distance from the base station increases, path loss also increases, leading to RF channel degradation. To maintain the target bitrate, the UE is allocated more PRBs to transmit the same amount of data. This continues until a turning point is reached where the path loss is too high for the UE to reach the base station with the same power per subcarrier. At this point, the scheduler reduces the number of PRBs and lowers the MCS. Consequently, the data rate becomes insufficient to sustain the desired real-time streaming.

Although a much longer list of RAN measurements and other specific metrics is collected from the MNO side, by applying knowledge of OFDM modulation in a 3GPP environment, feature selection for regression can be performed.

To verify the relationship, a Pearson correlation coefficient matrix is used, which is presented in Figure 3.19.

Although the relationship in this multivariate regression is expected to be non-linear, the Pearson correlation coefficient reveals the relationship between VMAF and uplink throughput. The throughput components known to MNOs show that

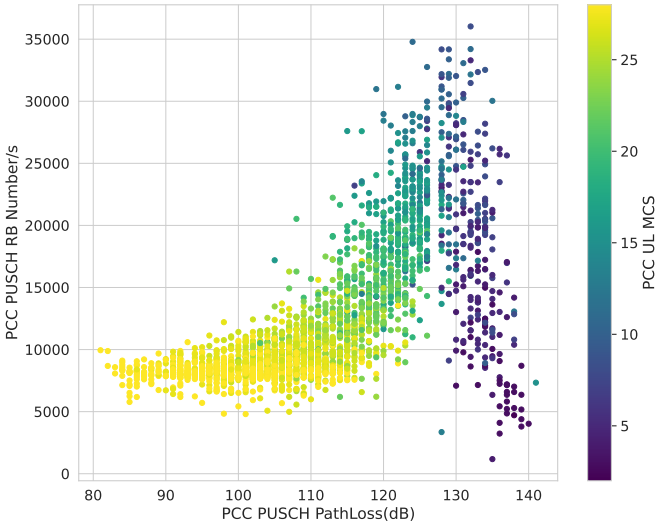


Fig. 3.18. Relationship between path loss, PRB Allocation, and MCS for a constant bitrate (created by the author)

MCS has a strong positive correlation, while path loss has a negative correlation. The relationship with PRBs is more complex, as it can be limited by a lack of power or high utilization by other customers.

Next, to model and forecast objective QoE values expressed as VMAF, various regression techniques are applied using the Scikit-learn library (Pedregosa et al., 2011). Non-deep-learning-based techniques are chosen to preserve explainability. The Mean Squared Error (MSE) (3.6) and the R^2 (3.7) score are chosen as quality metrics.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{3.6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{3.7}$$

The results are presented in Table 3.8. The Support Vector Regression (SVR) with a polynomial kernel of degree 4 shows the best performance, with an MSE of 0.2067 and a score of 0.8852 for UL throughput prediction, and an MSE of 124.27 and an R^2 score of 0.7985 for VMAF. The polynomial regression of degree 6 and SVR with a Radial Basis Function (RBF) kernel also performs well. Linear regression techniques show relatively lower R^2 scores, indicating less accuracy in modeling the relationship. This analysis demonstrates that higher-degree polynomial regression models can better capture the non-linear relationships in the

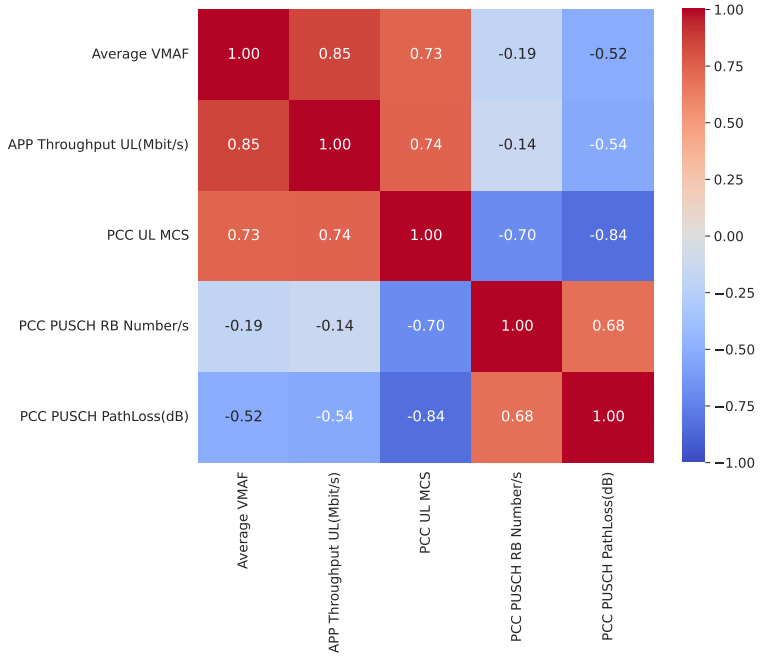


Fig. 3.19. Pearson correlation matrix between VMAF and various uplink-throughput components (created by the author)

data, providing more accurate QoE predictions. However, even with polynomial regression of degree four, an almost 0.8 R^2 score value can be achieved.

3.6.3. Analysis of Video Quality in an Urban Scenario

The next scenario evaluates the VMAF score in an urban environment. In an urban environment, cellular networks are deployed more densely than in a rural scenario. The inter-site distance is lower, which reduces the path loss. Additionally, most base stations operate with 5G NSA connections on 700 MHz FDD or 3.5 GHz TDD bands. With lower path loss, the UE can transmit using 4G LTE and 5G NR simultaneously.

In the test, 720p video recording and streaming is performed. This test evaluates only the effect of video transmission across the network, excluding any effects of video transcoding.

Figure 3.20 shows the relationship between the achieved uplink throughput and the average VMAF score. In this scenario, two protocols are tested: RTMP and WebRTC. In urban scenarios, the RTMP protocol always delivers video streams

Table 3.8. Performance of regression models for UL throughput (UL) and VMAF (V)

Model	MSE (UL)	R^2 (UL)	MSE (V)	R^2 (V)
SVR Poly. (deg 4)	0.2067	0.8852	124.27	0.7985
Poly. Reg. (deg 6)	0.2073	0.8849	128.42	0.7918
SVR RBF (C=100, γ=0.1)	0.2095	0.8837	124.61	0.7980
Poly. Reg. (deg 4)	0.2099	0.8834	124.27	0.7985
Poly. Reg. (deg 5)	0.2129	0.8818	135.38	0.7805
Poly. Reg. (deg 3)	0.2230	0.8761	125.90	0.7959
SVR Poly. (deg 3)	0.2285	0.8731	125.74	0.7961
Poly. Reg. (deg 2)	0.2314	0.8715	119.88	0.8057
SVR RBF (C=1000, γ=1)	0.2480	0.8623	155.67	0.7476
Lin. Reg.	0.3114	0.8271	154.85	0.7490
Ridge Reg.	0.3114	0.8271	154.85	0.7490
Bayesian Reg.	0.3114	0.8271	154.85	0.7490
Elastic Net	0.3124	0.8265	154.85	0.7490
Lasso Reg.	0.3128	0.8263	154.85	0.7490
SVR Lin.	0.3188	0.8229	178.34	0.7109
SVR Poly. (deg 2)	0.3291	0.8172	128.46	0.7917
GLM Poisson	0.3629	0.7985	200.24	0.6754
Dec. Tree	0.3705	0.7943	229.51	0.6279
LARS	0.4994	0.7227	154.85	0.7490

with high VMAF scores at the cost of higher throughput. On the other hand, WebRTC uses native adaptation to network conditions, which reduces bandwidth utilization but results in a significantly lower VMAF score. The plot reveals a mismatch in video stream adaptation over the cellular network.

Table 3.9. Comparison of RAN and RF parameters for uplink throughput and VMAF

Metric	NM (WebRTC)	M (WebRTC)	NM (RTMP)	M (RTMP)
Avg. VMAF	87.93	57.96	98.45	99.28
App UL Thp (kb/s)	3333.48	2294.29	4828.94	5146.38
App DL Thp (kb/s)	19.27	19.00	79.73	80.56
PCC RSRP (dBm)	-89.32	-81.17	-82.32	-83.44
PCC UL MCS	25.59	25.53	26.60	25.30
PCC Path Loss (dB)	107.64	99.06	98.91	101.38

RTMP achieves a higher VMAF score compared to WebRTC, but at the cost of using more bandwidth, even though the other RF network metrics are comparable. As RTMP consistently achieves a near-perfect score, the evaluation focuses on the

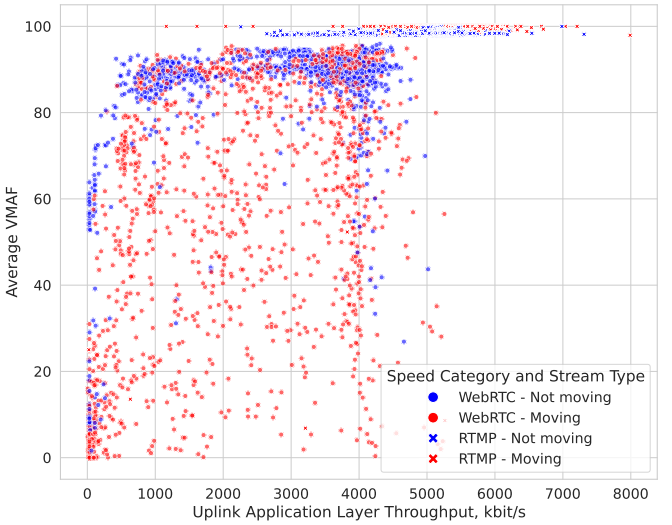


Fig. 3.20. Relationship between the achieved uplink throughput and the average VMAF (created by the author)

WebRTC protocol. As shown in Figure 3.20, the distribution is highly scattered, and there is a low correlation between RF network conditions and video performance. This is due to the native WebRTC congestion control algorithm, which first evaluates network conditions and then gradually increases the video frame rate and bit rate.

Rapid fluctuations in LTE and NR radio conditions, along with packet loss and increased latency, negatively impact the bitrate, causing a significant delay in throughput reaching the required level. This is illustrated in Figure 3.21, where the color indicates handover events. The top color-coded line indicates changes in the NSA (NR) serving cell, while the bottom color-coded line represents LTE handovers.

Whenever a handover or cell change occurs (as shown by the color transitions in the top/bottom lines), the uplink throughput experiences a sharp drop because WebRTC's congestion control (GCC) significantly lowers the sending bitrate, reducing the frame rate and degrading QoE. After the handover, throughput gradually increases, but GCC may take several seconds to restore the video bitrate. This delay extends periods of degraded quality, as reflected by dips in the VMAF plot.

On the other hand, TCP-based RTMP demonstrates greater robustness during handovers. As shown in Figure 3.22, a more stable performance is observed during LTE handovers in terms of VMAF and uplink throughput.

During the handover period, an increase in RTT can be observed from the col-

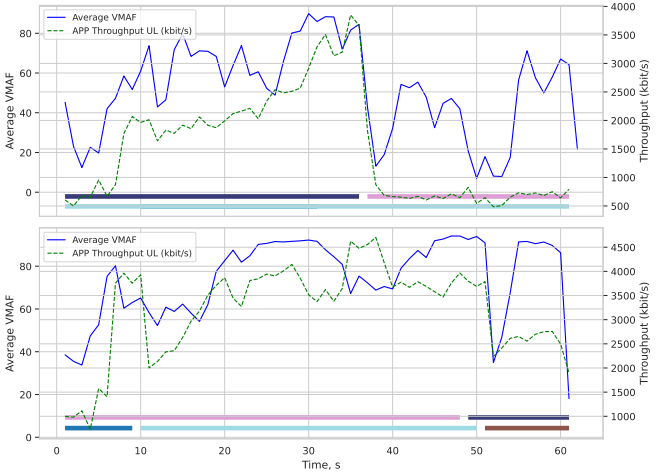


Fig. 3.21. Impact of NSA cell change and LTE handover on VMAF and uplink throughput during WebRTC streaming (created by the author)

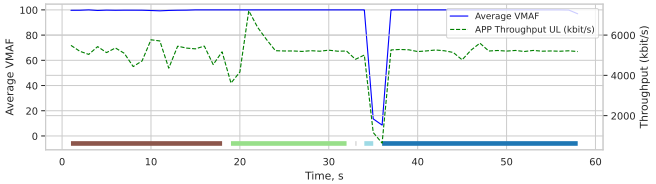


Fig. 3.22. Impact of LTE handover on VMAF and uplink throughput during RTMP streaming (created by the author)

lected TCP pcap traces. This is presented in Figure 3.23. If the handover duration is not excessively long, the subsequent increase in throughput helps recover packet losses. If the handover duration and RTT increase significantly, video frame loss occurs, but throughput recovers immediately, avoiding prolonged VMAF degradation.

3.7. Analysis and Modeling of Video Quality Based on RF Measurements Discussion

This study conducts a comprehensive evaluation of uplink real-time video streaming QoE over 4G and 5G cellular networks. By collecting extensive datasets that include objective QoE metrics, network measurements, and RF parameters, the

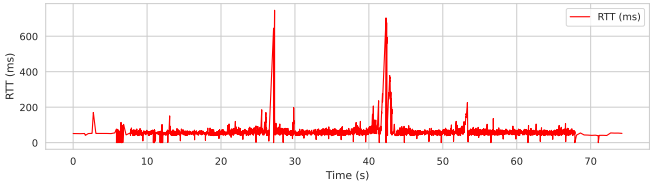


Fig. 3.23. Impact of LTE handover on RTT during RTMP streaming (created by the author)

impact of various factors on video streaming quality is analyzed. The key findings from these experiments are discussed below.

3.7.1. Impact of Video Transcoding on Quality

The initial experiments focus on understanding how video transcoding affects the VMAF score under varying network conditions. A linear degradation is observed in VMAF scores corresponding to reductions in video resolution and bitrate due to transcoding. Specifically, transcoding from an unimpaired 1080p to 720p results in a degradation of approximately 20 VMAF points, while transcoding to 480p and 360p leads to degradations of around 45 and 65 VMAF points, respectively. These results are aligned with the effects of transcoding on video quality.

Moreover, when attempting to stream 1080p video at 9 Mbps, limitations are encountered due to TCP interactions over the cellular network. The high bitrate demand cannot be consistently met, leading to packet losses and a significant drop in QoE. This highlights the challenges of streaming high-resolution video over cellular networks, where achieving and maintaining high throughput is difficult due to network variability.

3.7.2. Factors Influencing Quality in Rural Scenarios

In rural environments, path-loss and MCS values play significant roles in determining the achievable uplink throughput and, consequently, the QoE. As the distance from the base station increases, the higher path loss leads to a reduction in signal quality. To compensate, the network scheduler initially allocates more PRBs to maintain the desired bitrate. However, beyond a certain point, the UE's power limitations prevent further increases in PRB allocation, forcing the scheduler to reduce the MCS and, ultimately, the achievable throughput.

The correlation analysis shows a strong positive correlation between VMAF scores and MCS values, and a negative correlation with path loss. The number of

allocated PRBs has a more complex relationship with QoE due to its dependency on both UE power capabilities and network load. Regression models, particularly polynomial regression of degree four, demonstrate reasonable accuracy in predicting QoE based on network measurements, achieving an R^2 score of approximately 0.8.

These findings suggest that in rural scenarios, where network coverage and capacity are limited, RF parameters such as path loss and MCS can serve as reliable indicators of potential QoE degradation. MNOs can leverage this insight to optimize network configurations and proactively manage resources to improve user experience.

3.7.3. Quality in Urban Scenarios and Protocol Comparison

In urban environments, the performance of TCP-based RTMP and UDP-based WebRTC protocols are compared. The results show that RTMP consistently achieves higher VMAF scores, close to the maximum, due to its reliable transmission over TCP. However, this comes at the cost of higher bandwidth consumption. In contrast, WebRTC exhibits significantly lower VMAF scores despite similar RF conditions.

The lower performance of WebRTC is attributed to its native congestion control mechanisms, which adjust the video bitrate based on perceived network conditions. Rapid fluctuations in RF conditions and packet losses in the cellular network mislead the congestion control algorithms, causing them to reduce the bitrate excessively and react slowly to improvements. Additionally, the variability and latency introduced by frequent LTE handovers and NR cell changes further impact WebRTC's ability to maintain a stable bitrate.

The analysis also reveals that RTMP streams are more robust to handovers, maintaining higher throughput and quickly recovering from temporary disruptions. The inherent reliability of TCP allows for retransmissions and better handling of packet losses, which is beneficial in the dynamic conditions of cellular networks.

3.7.4. Impact of Handovers and Network Dynamics

Handovers are a critical factor affecting uplink video streaming QoE. In the experiments of this study, LTE handovers and NR cell changes introduce delays and throughput drops, particularly impacting WebRTC streams. The slow growth in throughput following a handover event leads to prolonged periods of reduced video quality. For RTMP streams, although handovers cause temporary increases in RTT and potential frame losses, the throughput recovered more swiftly due to

TCP's congestion control and retransmission mechanisms.

These observations underscore the importance of optimizing handover procedures and minimizing their impact on real-time applications. Enhancements in handover algorithms, such as predictive handovers or make-before-break strategies, can mitigate the negative effects on QoE.

3.7.5. Screen-to-Screen Latency Analysis

The latency measurements indicate that both RTMP and WebRTC protocols exhibit screen-to-screen delays below 200 ms in most scenarios, which are within acceptable limits for many real-time applications. However, WebRTC shows slightly lower latency compared to RTMP, owing to its use of UDP and avoidance of TCP's acknowledgment overhead. Despite this advantage, the overall latency difference is minimal, and the stability offered by RTMP might outweigh the marginal latency benefits of WebRTC in certain use cases.

Frame losses are identified as a significant contributor to increased latency and QoE degradation. Lost frames result in prolonged frame durations and longer screen-to-screen delays. This effect is more pronounced in WebRTC streams due to their susceptibility to packet losses without retransmission mechanisms.

3.7.6. Implications for Network Operators and Application Developers

The results deliver several implications for both MNOs and application developers:

- Protocol Selection: The choice of streaming protocol has a substantial impact on QoE. While UDP-based protocols like WebRTC can offer lower latency, they may struggle with maintaining video quality in variable network conditions. TCP-based protocols like RTMP provide greater stability and robustness, which is crucial for ensuring high QoE in uplink video streaming.

- Adaptive Streaming Mechanisms: The native adaptation algorithms in WebRTC are found to be suboptimal in rapidly changing cellular environments. There is a need for improved adaptation strategies that more accurately respond to network conditions, possibly by incorporating additional network feedback or predictive models.

- Network Optimization: MNOs can utilize RF measurements such as path loss, MCS, and PRB allocations to predict and enhance QoE. By proactively

managing network resources and optimizing handover procedures, operators can mitigate QoE degradation, especially in areas with poor coverage or high mobility.

- QoE Modeling: This analysis demonstrates the feasibility of modeling QoE based on network measurements accessible to MNOs. This capability allows for real-time QoE monitoring and the potential for network-side interventions to improve user experience without requiring access to application-layer data.

- 5G Advanced Features: Network Slicing, Edge Computing, URLLC, and AI-based optimization: The investigation evaluates commercially operational large-scale networks, but the application of the findings is not limited to the currently available network functionalities. The results benefit from optimal use of advanced 5G features to improve uplink real-time streaming reliability and QoE.

The application of Network Slicing, as a logical network that provides specific capabilities and characteristics with dedicated resources for particular services or customer groups, can significantly improve QoE. However, due to limited spectrum availability, end-to-end Network Slicing must be optimally configured by balancing the trade-off between maintaining expected QoE levels and allocating network resources efficiently. The method proposed enables the precise calculation of network resources for slicing. The results emphasize critical factors, including PRB requirements, coverage-path loss constraints, the importance of avoiding handovers, and the need to manage latency effectively when configuring static or dynamic network slices with QoE guarantees. By understanding the specific network requirements of various real-time streaming services, operators can implement either static resource allocation or dynamic slicing optimized with AI-based techniques. Furthermore, the analisys confirms the importance of minimizing the distance between the user and network servers to reduce end-to-end latency. Placing streaming servers closer to end-users, for example, through Edge Computing, can significantly reduce latency and improve throughput stability. This approach improves the performance of protocols like WebRTC by mitigating the impact of network variability over long distances.

The commercial adaptation of Network Slicing and Edge Computing, combined with URLLC features designed for low latency but optimized for lower bandwidth usage, can support highly efficient real-time video streaming with tens-of-milliseconds-level delay control over the URLLC sidelink. In general, the proposed method allows MNOs to offer QoE as a Service, allowing efficient resource

allocation and optimized network performance for real-time streaming applications.

3.8. Conclusions of the Third Chapter

This chapter examined downlink OTT video delivery and uplink real-time video streaming over commercial LTE and 5G NR networks from the perspective of QoE evaluation using operator-relevant measurements. The main conclusions are as follows.

1. Downlink video delivery is a major source of radio-network load, and its resource cost depends strongly on codec efficiency and RF conditions. More efficient codecs, such as VP9 and AV1, reduce the required bit rate compared with AVC, whereas poor SINR substantially increases PRB consumption and raises the probability of stalling. Carrier aggregation provides only limited benefit for short HAS chunk transfers when the secondary carriers carry only a small share of the traffic.

2. In the evaluated LTE and NR scenarios, uplink real-time video streaming was constrained by coverage, path loss, and mobility. Uplink throughput degraded rapidly as signal conditions worsened, confirming that uplink coverage remains a practical bottleneck for real-time video services in commercial cellular networks.

3. Objective full-reference metrics such as VMAF, SSIM, and PSNR can be combined with packet captures, frame-level media information, and RAN measurements to evaluate uplink video QoE in commercial cellular networks.

4. Protocol choice significantly affects uplink streaming QoE. RTMP was more robust and stable under handovers and varying RF conditions, whereas WebRTC provided slightly lower latency but exhibited stronger QoE degradation under radio dynamics.

5. In coverage-limited uplink scenarios, operator-visible indicators, such as path loss, MCS, PRB allocation, and achieved uplink throughput, showed clear relationships with VMAF. Using these features, non-linear regression models predicted VMAF with an R^2 close to 0.80, demonstrating the feasibility of operator-side QoE estimation without access to the application payload.

4

Data-Driven Methods for Quality of Experience Enhancement

At the lowest level of user experience, QoE depends on baseline cellular-network metrics such as throughput, latency, SINR, network load, and the selection of optimal 3GPP parameters.

Meeting the demand for a high-quality user experience makes cellular-network coverage planning increasingly complex. The evolution of mobile networks toward the next generation introduces new frequency bands, heterogeneous cell sizes, and additional planning dimensions such as beam-pattern selection for 5G massive-MIMO cells (Agiwal et al., 2016). Cellular-network planning must therefore consider numerous parameters, including antenna radiation pattern or beam type, electrical and mechanical tilt, azimuth, antenna height, and reference-signal transmit power.

The multitude of possible RF and algorithm-parameter settings must be applied across large geographical areas with diverse propagation conditions while reflecting the time-varying demands of MNO users. Both temporal and spatial environmental characteristics, such as peak-hour traffic load, seasonal mobility patterns, and user velocity, must be considered.

This section presents methods that address these challenges. First, it describes network-load time-series forecasting (Chmieliauskas & Guršnys, 2019), which captures temporal traffic patterns for proactive QoE management. Next, it discusses the application of reinforcement learning to the search for optimal network geometry (Chmieliauskas et al., 2023), aimed at improving RF conditions. Finally, it examines specific MIMO parameter settings that maximize achievable user throughput (Chmieliauskas, 2021).

4.1. Cell Traffic Growth and Congestion Forecasting

Temporal variations within cells and across the network are among the most important factors in planning and predicting performance, which ultimately impact the QoE of a cellular network. Both LTE and NR are based on OFDM, so similar load estimation and throughput estimation methods are applicable. Here, the proposed methods for traffic and congestion forecasting are described, which can be applied to predict both network load and QoE. They consist of three main parts:

- Forecast of total traffic per cell per day using the `fbProphet` algorithm.
- Daily traffic trend forecasting over 24-hour periods using the `fbProphet` algorithm.
- Analysis of the relationship between data traffic volume and LTE cell-load status.

4.1.1. Total Daily Traffic Forecasting

In LTE FDD networks, the main cause of congestion is downlink traffic. For example, across 100 LTE sites used in this study, downlink traffic was 11 times higher than uplink traffic. Due to the asymmetry between downlink and uplink traffic, only the downlink direction will be considered for forecasting and congestion evaluation.

Cellular network traffic forecasting method. Cellular traffic prediction has been surveyed in (Jiang, 2022), where Prophet (Taylor & Letham, 2018) is identified as one of the well-defined forecasting methods. Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fitted with yearly, weekly, and daily seasonality, as well as holiday effects. It works best with time series that have strong seasonal effects and multiple seasons

of historical data. LTE daily traffic data perfectly fits the description of the Prophet time series forecasting algorithm. LTE data traffic forecasting and evaluation were performed on 100 randomly selected sites of the live LTE network. Each site may have a different number of LTE cells depending on the current network plan. Selected sites contain 141 LTE Band 3 (LTE 1800MHz) cells and 237 LTE Band 20 (LTE 800MHz) cells. The selected sites make up a significant proportion of the live network and represent diverse geographical areas and traffic patterns (e.g., city hotspots, suburban areas with many fixed wireless clients, and wide rural coverage sites with low traffic). Such random selection provides insight into forecasting performance across different scenarios. The aim of the experiment is to forecast future traffic with comparable or better accuracy to other forecasting methods. For example, the goal is to achieve forecasting accuracy comparable to the holiday traffic forecast results described in (Xu et al., 2018), where 84% of the data points had an error of less than 20% relative to the actual value. This accuracy for a 15-day-ahead holiday forecast is generally accepted by the industry (Xu et al., 2018). MNOs store daily aggregated performance indicator data for the long term and data with a higher granularity (e.g., 1-hour granularity stored for 30 days) for the short term. Daily forecasting was performed using 180 days of historical data. The six-month period was split into five months for training and one month for forecast testing. In the 180-day period, there is sufficient data to observe weekly, monthly, and seasonal trends. A 28-day forecast provides sufficient time for operators to act if cell congestion is predicted. In the experiment, the forecasted 28 days were compared to real traffic, and forecast accuracy was evaluated using various forecast performance metrics. In the second step, using the available 30 days of hourly data, the traffic trend over a 24-hour period was defined. From the long-term forecast of total daily traffic and the defined hourly trend, the traffic volume for each hour of the day can be predicted. The forecasted hourly traffic was compared to the spectral efficiency of the LTE cell to evaluate whether the load threshold was reached. By combining all the steps described above, future congestion in LTE cells during busy hours can be detected. The overall procedure for LTE cell-traffic forecasting and congestion evaluation is summarized in Algorithm 1.

Algorithm 1 LTE Cell Traffic Forecasting and Congestion Evaluation

Require: Historical daily LTE traffic per cell (180 days), hourly traffic per cell (30 days)

Ensure: Forecasted congestion status for LTE cells

for each LTE cell **do**

Step 1: Total Daily Traffic Forecast

    Collect daily LTE traffic data for 180 days

    Split data: first 150 days for training, last 30 days for testing

    Train Prophet forecasting model using training data

    Forecast daily traffic for the next 28 days

    Evaluate forecast accuracy using MAPE, R^2, and explained variance

Step 2: Hourly Traffic Trend

    Collect hourly LTE traffic data for the last 30 days

    Using hourly data, calculate hourly traffic proportions (daily traffic distribution)

    Identify the busy hour for the cell

Step 3: Hourly Traffic Forecast and Congestion Evaluation

    for each forecasted day **do**

        for each hour **do**

            Compute forecasted hourly traffic volume:

                Hourly volume = (Forecasted daily traffic) × (Hourly traffic proportion)

            Determine cell-specific spectral efficiency from historical PRB utilization

            Calculate PRB utilization for forecasted traffic using spectral efficiency

            Check if PRB utilization exceeds predefined thresholds (e.g., 80% for LTE1800 MHz, 60% for LTE800 MHz)

            if PRB utilization exceeds the threshold **then**

                Mark cell as congested for this hour

            else

                Mark cell as non-congested for this hour

            end if

        end for

    end for

end for

Step 4: Output

Provide hourly congestion status for each LTE cell for the forecasted 28-day period

LTE traffic forecasting results. In Figure 4.1, an example of real and forecasted traffic is shown. The forecasted traffic exhibits a similar decreasing trend and similar short-term spikes as the real traffic.

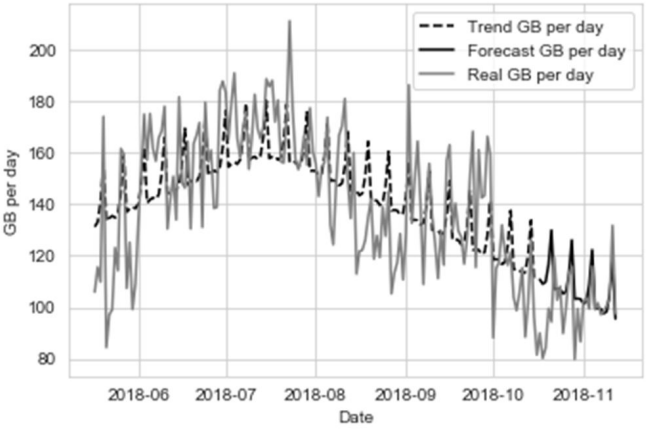


Fig. 4.1. Daily cell-traffic forecast example (created by the author)

The forecasting accuracy was evaluated using the explained variance score (4.1), the R^2 score (4.2), and the Mean Absolute Percentage Error (MAPE) (4.3).

$$explained\_variance(y, \hat{y}) \;=\; 1 - \frac{\mathrm{Var}\{\, y - \hat{y} \,\}}{\mathrm{Var}\{\, y \,\}}, \tag{4.1}$$

$$R^2(y, \hat{y}) \;=\; 1 - \frac{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \bar{y})^2}, \quad \text{where } \bar{y} = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} y_i, \tag{4.2}$$

$$\text{MAPE}(y, \hat{y}) \;=\; \frac{100\%}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \frac{|y_i - \hat{y}_i|}{y_i}. \tag{4.3}$$

Here, y denotes the actual traffic volume (bits), and $\hat{y}$ denotes the forecasted traffic volume (bits). Performance metrics were calculated utilizing `scikit-learn` (Pedregosa et al., 2011), a comprehensive Python toolkit for statistical modeling and data analysis.

Table 4.1 shows the MAPE, explained-variance score, and R^2 of the overall

forecast, obtained by comparing the 28-day predictions with the actual daily traffic of every LTE cell.

Table 4.1. LTE traffic forecasting accuracy: overall result

MAPE	Explained variance score	R^2 score
44.22%	0.90007	0.8980

Figure 4.2 shows the MAPE, and Figure 4.3 shows the explained-variance score and the R^2 score for all cells.

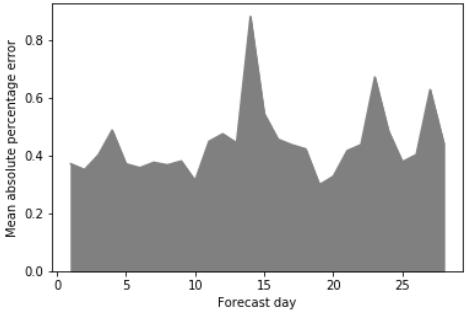


Fig. 4.2. Mean absolute percentage error (MAPE) versus forecast day (created by the author)

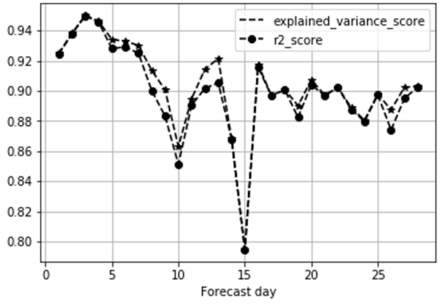


Fig. 4.3. R^2 score and explained-variance score versus forecast day (created by the author)

Figure 4.4 shows a scatter plot of forecasting accuracy (MAPE) versus the carried traffic volume for each cell.

By evaluating the forecasting results for each cell, the forecast MAPE can be compared with the total carried traffic volume. From Figure 4.4, it is clear that cells that carry more traffic are usually more predictable. If traffic in a cell is

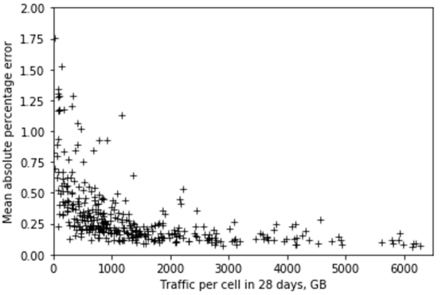


Fig. 4.4. Mean absolute percentage error of forecasted traffic versus transferred traffic per cell (created by the author)

low, any individual peak or heavy data user can cause large daily variations in total traffic. The traffic distribution of such cells exhibits more randomness, which makes accurate forecasting more difficult.

4.1.2. Hourly Traffic Trend

MNOs usually store daily performance indicator values for the long term. However, to evaluate congestion, shorter periods of peak usage (e.g., 1 hour) should be considered.

LTE Traffic Forecasting Hourly Method. Using `fbProphet`, the peak hour and the proportion of daily traffic volume can be obtained for every cell. Using the proportion of daily traffic, the hourly traffic volume can be estimated and later checked whether it reaches the capacity threshold for that cell. To obtain the hourly traffic trend, 30 days of hourly traffic data were used.

LTE Traffic Forecasting Hourly Results. Figure 4.5 shows real hourly traffic data for an LTE cell and the trend modeled using `fbProphet`. The traffic volume trend, representing the proportion of total daily traffic, accurately follows the real traffic pattern but smooths out random peaks and drops, allowing forecasting of hourly traffic volume.

Using the `fbProphet` algorithm, a traffic trend can be created for each cell. The traffic trend reveals the busy hour for each cell and the proportion of busy-hour traffic relative to the total daily traffic. By combining total daily and hourly traffic forecasting, it is possible to predict the future traffic for each LTE cell on an hourly basis.

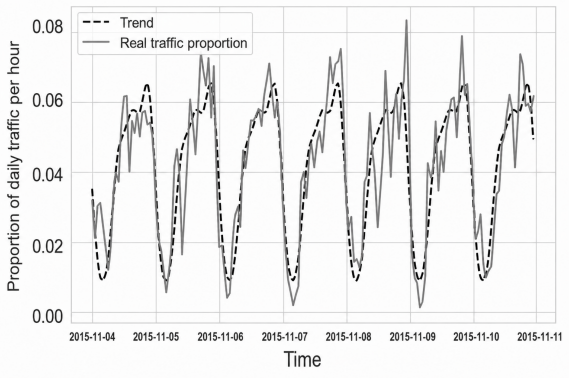


Fig. 4.5. One week of LTE cell traffic with one-hour granularity compared to the traffic trend modeled using the `fbProphet` algorithm (created by the author)

4.1.3. Cell Capacity Evaluation

The hourly traffic forecasts obtained in the previous section are now used to determine whether an LTE cell can sustain the expected load without degrading QoE due to congestion or spectrum contention. LTE downlink throughput perceived by a single user is determined by three main components:

1. Scheduled physical-resource blocks (PRBs) allocated by the eNodeB scheduler (Capozzi et al., 2012).

2. SINR, reported by the UE as a Channel Quality Indicator (CQI), which the eNodeB maps to the highest modulation-and-coding scheme (MCS) that satisfies the 10% BLER target (Kawser et al., 2012).

3. MIMO layer usage, which provides further spectral-efficiency gains through spatial multiplexing (3rd Generation Partnership Project, 2009).

MIMO-layer usage and CQI depend on user location, overall interference levels, and UE receiver quality. The LTE eNodeB selects the MCS based on UE-reported downlink channel quality, using Table 7.1.7.1-1 (Modulation and TBS index) (3rd Generation Partnership Project, 2009). Given the MCS index and the number of scheduled PRBs, the transport block size (TBS) for each Transmission Time Interval (TTI) can be obtained from Table 7.1.7.2.1-1 (Transport block size) (3rd Generation Partnership Project, 2009). An LTE cell with 20 MHz bandwidth can schedule a maximum of 100 PRBs for a single UE. If 100 PRBs are scheduled for the UE using MCS 28, approximately 75.376 Mbit/s per MIMO stream, or about 150 Mbit/s for a 2×2 MIMO 64-QAM UE, can be achieved.

If the same 100 PRBs are scheduled with MCS 0 and transmit diversity, only 2.792 Mbit/s can be achieved. Depending on channel quality, LTE throughput for the same amount of time–frequency resources can vary by a factor of 54. This also shows that transferring the same amount of data can require very different amounts of time-frequency radio resources. In a live LTE network, users experience different radio conditions, and a direct relationship between individual user throughput and the total traffic carried per cell cannot be defined.

The number of resources that can be scheduled to a single user depends on the cell load and scheduler activity. A threshold can be defined to evaluate each cell on an hourly basis, determining whether it provides limited QoE based on available performance indicators. This criterion can be the average downlink experienced user throughput. The average UE throughput, R_{UE}, is defined by Equation (4.4):

$$R_{\mathrm{UE}} = \frac{V_{\mathrm{PDCP}}}{T_{\mathrm{eff}}}. \tag{4.4}$$

In this expression, V_{PDCP} corresponds to the total PDCP Service Data Unit (SDU) volume transferred per cell during the reporting period, excluding the data in the last Transmission Time Intervals (TTIs). The denominator, T_{eff}, represents the effective transmission time over the same period (3rd Generation Partnership Project, 2010).

Measurements are aggregated for all users in a cell. Figure 4.6 shows the relationship between the average user throughput and the cell traffic volume. When average user throughput is used as the evaluation criterion, measurements from live network cells show that the correlation between cell traffic volume and average user throughput is very low. Cell congestion or low average user throughput can be evaluated based on PRB load. The relationship between PRB utilization and user throughput is explained in (Salo & Zacarías, 2017). Figure 4.7 shows the relationship between PRB load and average user throughput for a set of live network test cells. A clear linear correlation is observed. As the PRB load approaches 100%, the average user throughput decreases, and its variation is reduced. The PRB load threshold, above which a cell is considered congested, is defined based on commercial requirements. An LTE 800MHz cell with 10 MHz bandwidth is considered loaded when PRB utilization exceeds 60% (20 PRBs unused), while an LTE 1800 MHz cell with 20 MHz bandwidth is considered congested when PRB utilization exceeds 80% (again, 20 PRBs unused). From the relationship between PRB load and traffic, the spectral efficiency can be calculated. Using the spectral efficiency, it is possible to calculate how much traffic the cell can carry before reaching the PRB load threshold. The spectral efficiency, η_{SE}, is determined by Equation (4.5):

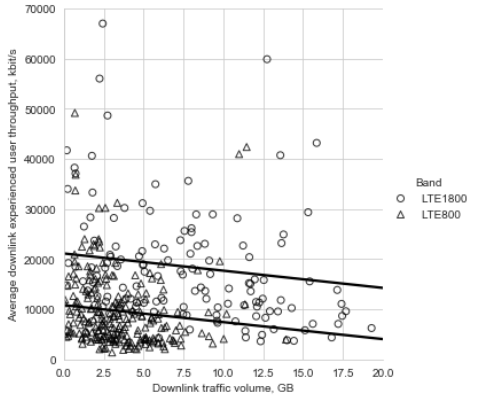


Fig. 4.6. Traffic volume versus average user throughput during the busy hour for the test LTE cells. Each dot represents one cell (created by the author)

$$\eta_{\mathrm{SE}} = \frac{V_{\mathrm{data}}}{U_{\mathrm{PRB}} \cdot B \cdot T_{\mathrm{meas}}}. \tag{4.5}$$

In this derivation, η_{SE} is expressed in bits/s/Hz. The numerator, V_{data}, denotes the total traffic volume in bits carried by the cell. The denominator comprises the product of the average PRB utilization (U_{PRB}), the LTE cell bandwidth (B) in Hz, and the measurement period (T_{meas}), which represents the duration of the data-collection interval. Knowing the spectral efficiency allows for calculating the amount of traffic that can be carried until the PRB threshold is reached. In the industry, an average network-level approach is typically used to evaluate and classify cells. However, as discussed earlier, every cell has a unique user and traffic distribution, meaning that different amounts of spectrum resources are utilized for the same data volume. It is, therefore, proposed to evaluate each cell individually, as this approach provides a more accurate estimate of the traffic that can be carried before the high-load threshold is reached. Using regression, the spectral efficiency is calculated for each cell and the amount of traffic is determined, which it can carry before reaching the PRB-load threshold.

Fig. 4.8 shows LTE cells with different spectral efficiencies.

This subsection evaluated the applicability of the `fbProphet` time-series forecasting algorithm for LTE networks. Although this approach is widely used for Internet services, it is not commonly applied to cellular network performance evaluation. Using the `fbProphet` algorithm, daily cell traffic can be forecast with R^2 and explained-variance scores of approximately 0.9. The fbProphet algorithm can forecast daily LTE traffic and support the calculation of future traffic for each hour. The forecasted traffic can be compared with the spectral efficiency (bits/s/Hz) of

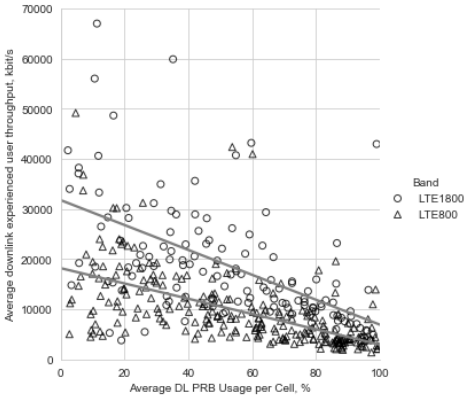


Fig. 4.7. Average user throughput versus PRB utilization during the busy hour (created by the author)

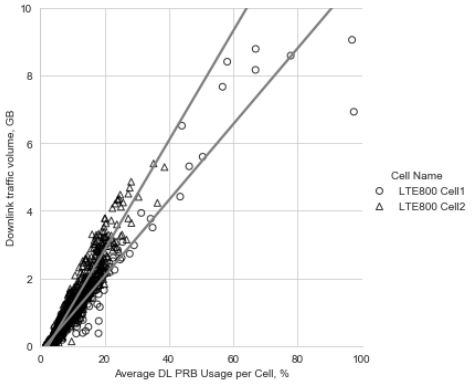


Fig. 4.8. Carried traffic volume versus PRB utilization for two LTE (created by the author)

each cell, allowing the corresponding future PRB load to be estimated. Knowing the future LTE cell load can help mobile network operators maintain an acceptable user experience while avoiding unnecessary over-provisioning.

4.2. Coverage-and-Capacity as the Foundation of User-Perceived Quality

To meet the demand for a high-quality user experience, planning the coverage of cellular networks is becoming increasingly complex. The evolution of mobile networks toward the next generation introduces new frequency bands, heterogeneous cell sizes, and additional planning dimensions such as beam-pattern selection for

5G massive-MIMO cells (Agiwal et al., 2016). Cellular-network planning must also account for numerous parameters, including the antenna propagation pattern or beam type, antenna tilt, azimuth, antenna height, and reference-signal transmit power. Cellular network planning aims to maximize network capacity while ensuring satisfactory coverage. This objective is commonly referred to as coverage-and-capacity optimization (CCO) (Buenestado et al., 2016).

The capacity component of a cellular network, also known as area throughput (Van Chien & Björnson, 2017), is mathematically expressed as:

$$R_{\text{area}} = B \cdot \rho_{\text{cell}} \cdot \eta_{\text{SE}}. \tag{4.6}$$

In this formulation, R_{area} denotes the area throughput (measured in bits/s/km^2), B represents the available bandwidth in Hz, ρ_{cell} is the cell density (cells/km^2), and η_{SE} corresponds to the spectral efficiency (bits/s/Hz/cell). Based on Equation (4.6), one primary approach to enhancing mobile-network capacity is to increase the system bandwidth, B. This can be achieved by adding higher-frequency cells within the base-station service area. A second approach is to increase cell density by deploying additional base stations, which can be costly due to active and passive hardware expenses and site-leasing fees. As a result, MNOs typically aim for only moderate increases in cell density and bandwidth usage to control costs. However, improving the SE of the cellular network is an effective way to enhance network performance without incurring additional costs.

During network planning and optimization, MNOs aim to maximize SINR across the cellular network by tuning network parameters such as downtilt, transmitted reference signal power, and antenna height. However, this task is challenging due to the large number of parameters involved, and the complexity grows exponentially as additional variables are added. As a result, the problem of cellular-network planning is NP-hard (Nguyen & Liu, 2019). Given the size and complexity of cellular networks, both academia and industry are actively researching and applying machine learning techniques for network optimization and control (Abedin et al., 2022; Sun et al., 2019). Furthermore, future 6G networks are envisioned to incorporate AI as an integral component (Letaief et al., 2019; Tong & Zhu, 2021).

The related literature proposes using reinforcement learning (RL) techniques to optimize capacity and coverage. The work by (Dreifuerst et al., 2021) uses Bayesian optimization and a deterministic policy-gradient algorithm in a multi-cell network to tune antenna downtilt and transmit power in each sector. (Tekgul et al., 2022) proposes optimizing antenna downtilt and the vertical and horizontal half-power beamwidths with the aim of improving three metrics: (i) the sum–log

rate of the user equipment (UE) for each base station (BS), (ii) the outage probability, and (iii) the SINR distribution across all UEs. This is achieved by employing differential evolution (DE) with Bayesian optimization (BO). In the work by (Liu et al., 2022), a multi-agent reinforcement-learning (MARL) method is proposed for optimizing antenna downtilt and the vertical and horizontal half-power beamwidths.

In contrast to most related works, the optimal selection of antenna direction (azimuth) is investigated. This increases the number of degrees of freedom while enabling more precise control of network coverage. This approach offers several benefits, including simplicity, explainability, and reduced workload for planning engineers. In addition, it reduces the likelihood of errors during network planning and optimization. To this end, the NP-hard problem of finding the optimal antenna azimuth is addressed. A Q-learning-based method (Sutton & Barto, 2018) is used to maximize SINR in a given area. To obtain SINR values for various antenna configurations, channel-propagation simulations are performed using the MATLAB-based QuaDRiGa model (Jaeckel et al., 2014). The RL agent is implemented using the Gym environment (Brockman et al., 2016).

4.2.1. System model for area coverage optimization

This subsection presents the system model for optimizing coverage over a target area in a cellular network comprising B base stations, as shown in Figure 4.9. The main objective is to maximize SINR by using learning-based antenna-azimuth selection.

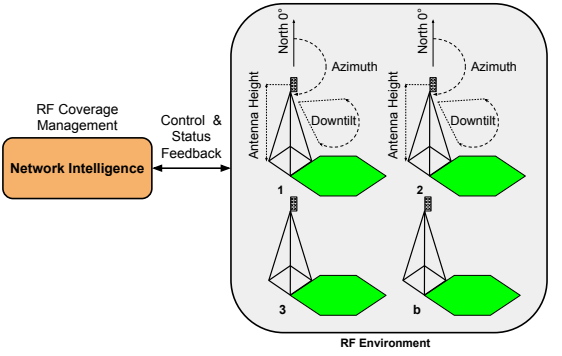


Fig. 4.9. System model for learning-based RF coverage optimization (created by the author)

A specific geographical area of interest is considered, where a total of B base stations are deployed. Each base station has an adjustable azimuth angle to optimize coverage. The network topology is represented by a collection of base stations positioned at predetermined coordinates within the area. The actual deployment scheme and inter-base station distances are determined based on real-world scenarios and practical considerations.

To improve area coverage and SINR, it is proposed to select the antenna azimuth at each base station. This technique allows adaptive adjustment of each antenna's azimuth angle to optimize coverage and minimize interference. By intelligently selecting the azimuth angle, a base station can focus the transmission energy toward the desired coverage area while mitigating interference from other antennas. Each antenna can be adjusted by turning $-20°$, maintaining its current direction ($0°$), or turning $+20°$. For example, if four base stations are considered and each can select from 18 possible azimuth values, the total number of combinations is $18^4 = 104{,}976$.

The 3GPP 3D UMa (Urban Macrocell) LOS channel model (Mondal et al., 2015) is employed to capture the propagation characteristics in the target environment. This model is particularly suitable for scenarios with a high density of base stations, such as urban areas, and accurately accounts for LOS conditions. By utilizing this channel model, it is possible to precisely evaluate the received power from each base station at any location within the geographical area of interest. This allows for the assessment of performance and analysis of the effectiveness of antenna azimuth selection under realistic propagation conditions. Using the 3GPP 3D UMa LOS channel model, coverage predictions are performed for every possible combination of $20°$ azimuth adjustments. Subsequently, a power map for each transmitter is generated, as shown in Figure 4.10. As the optimization metric, the SINR values across the area grid are aggregated into a single mean SINR value using the power maps of all transmitters. To begin, the maximum power value is extracted in each grid location $(i, j), i \in \{1, \cdots, N\}, j \in \{1, \cdots, M\}$ from the power maps of each transmitter, denoted as $P_{\text{max}}^{(i,j)}$, as

$$P_{\text{max}}^{(i,j)} = \max\left(P_{\text{rx, b}}^{(i,j)}\right), \quad b \in \{1, \cdots, B\}, \tag{4.7}$$

where $P_{\text{rx, b}}^{(i,j)}$ represents the power value at grid location (i, j) with respect to the bth transmitter.

Subsequently, the total power is calculated in each grid point, represented as

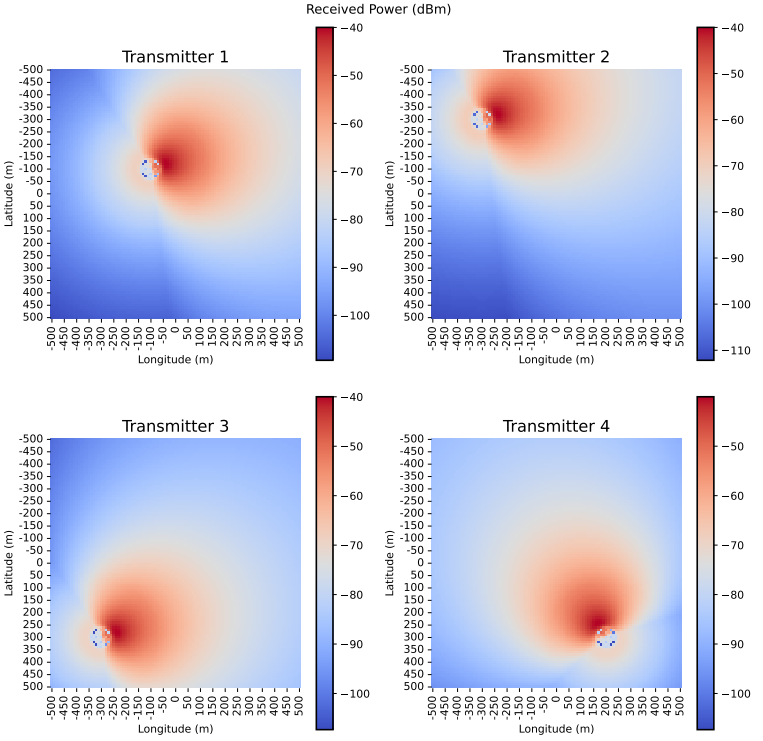


Fig. 4.10. Power map of each base station (transmitter) (created by the author)

$P_{\text{total}}^{(i,j)}$, by summing the powers of all transmitters as

$$P_{\text{total}}^{(i,j)} = \sum_{b=1}^{B} P_{\text{rx},b}^{(i,j)}, \tag{4.8}$$

where $P_{\text{max}}^{(i,j)}$ represents the signal strength from the strongest transmitter at grid location (i, j), while the remaining transmitters contribute to interference. Therefore, the interference power, $P_{\text{interference}}^{(i,j)}$, can be determined from (4.7) and (4.8) as

$$P_{\text{interference}}^{(i,j)} = P_{\text{total}}^{(i,j)} - P_{\text{max}}^{(i,j)}. \tag{4.9}$$

By ignoring the noise level, the mean value of SINR across the area can be obtained using

$$\text{SINR} = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{P_{\text{max}}^{(i,j)}}{P_{\text{interference}}^{(i,j)}}. \tag{4.10}$$

In this way, a single scalar value is derived representing the average SINR over the entire area, which is utilized as a reward in Q-learning.

To generate the simulation dataset, the 3GPP 3D UMa LOS channel model is employed. During dataset generation, area-coverage optimization is modeled by generating RF power maps using the QuaDRiGa simulator, which is implemented in MATLAB. The mean SINR serves as the reward metric and constitutes the main objective function of the optimization process. A coverage simulation is conducted using the parameters given in Table 4.2. These parameters are typical for antennas and transmitters used in LTE cells. The generated dataset is then used in a Q-learning-inspired optimization algorithm.

Table 4.2. Simulation parameters

Parameter	Value
Antenna configuration	
Antenna model	3GPP-3D
Number of vertical elements	8
Number of horizontal elements	1
Center frequency	1800 MHz
Electric downtilt angle	8 deg.
Antenna height	20 m
Transmit power	18.1 dBm
Coverage prediction	
Grid sample distance	10 m
Area size	1000 m × 1000 m
Receiver height	1.5 m
Prediction model	3GPP_3D_UMa_LOS

4.2.2. Reinforcement Learning Problem Formulation

The application of Q-learning is explored to address the antenna-azimuth optimization problem. Q-learning is a value-based RL algorithm that learns optimal actions to take in each state of the environment. It learns by estimating the value of the actions based on the expected reward from that state, without requiring a model of the environment. The general RL-based optimization approach for cellular-network planning is shown in Figure 4.11, and the process is summarized as follows:

1. *Initialization*: The procedure begins by initializing a set of human-planned

parameters or assigning random values from the allowed parameter space.

2. *RF Prediction*: The RF prediction model is utilized to gather information about the coverage grid, which includes signal strength, SINR, and QoE indicators such as throughput, for each geographical grid polygon.

3. *Reward Calculation*: The output of the planning model is converted into a scalar reward value. This reward value serves as a feedback signal for the RL agent to evaluate the performance of the current parameter configuration.

4. *RL Agent Update*: Based on the reward, the RL agent updates the parameters to improve future rewards. Its decision-making process aims to maximize the expected reward.

The process continues until the specified stopping condition is met, such as the number of steps or a slowdown in the growth of the reward.

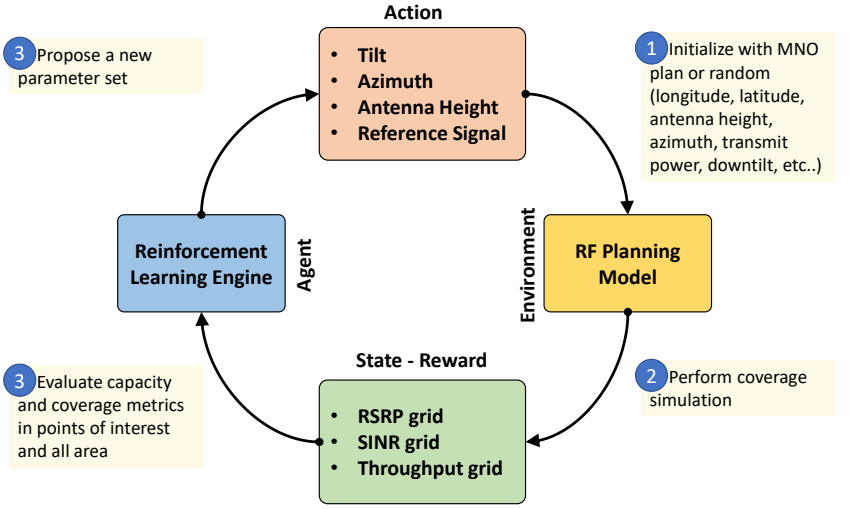


Fig. 4.11. General reinforcement learning for cellular network planning (created by the author)

4.2.3. Learning Task for Antenna Direction Optimization

The problem is formulated to investigate the applicability of an RL agent in a simplified antenna-optimization environment. The reward function is defined as the SINR value in (4.10), and configurations are limited to the antenna direction

of each base station. Although the azimuth of each antenna could be selected from the full 360-degree range, it is restricted to a step size of 20°. This results in 18 options per antenna, leading to a total search complexity of $O(18^B)$ for B base stations. Thus, exhaustive search becomes infeasible as the number of antennas increases.

The following components are defined for the RL formulation of the problem:

- **Environment (E)** In this formulation, the environment E corresponds to the RF coverage map of the considered area.

- **States (S)** Let S represent the set of states within the environment E. Each state $s \in S$ represents a specific configuration of the antenna azimuth of each base station in a given area.

- **Actions (A)** Let A denote the set of actions within the environment E. Each action $a \in A$ corresponds to a potential change in the antenna azimuth. Each action corresponds to rotating an antenna clockwise by 20°, counter-clockwise by 20°, or keeping it unchanged.

- **Rewards (R)** The reward function $R(s, a)$ defines the immediate feedback received by the agent after taking action a in state s. The reward is the average SINR in (4.10) in a given area.

- **Q-Function (Q)** The Q-function, denoted as $Q(s, a)$, represents the expected reward obtained by taking action a in state s and following the optimal policy thereafter. It can be defined recursively as

$$Q(s, a) = R(s, a) + \gamma \cdot \max_{a'} Q(s', a'), \tag{4.11}$$

  where s' is the resulting state after taking action a in state s, and $\gamma \in [0, 1]$ is the discount factor that balances the importance of immediate and future rewards. Cellular networks can turn an antenna immediately to any state, without the necessity to continuously follow turns. Due to it, the immediate reward for being in the state (azimuth configuration) is the same as cumulative SINR improvements during the movement of antenna azimuth. Thus, $\gamma = 0$ is chosen.

- **Policy.** The policy, denoted π, represents the agent's strategy for selecting actions in different states. The ϵ-greedy policy is used based on the Q-values. The ϵ-greedy policy selects the action with the highest Q-value with

a probability of $1 - \epsilon$ and selects a random action with a probability of ϵ to encourage exploration.

The policy can be defined as

$$\pi(s) = \begin{cases} \arg\max_a Q(s, a), & \text{with probability } 1 - \epsilon, \\ \text{random action}, & \text{with probability } \epsilon. \end{cases} \tag{4.12}$$

where the parameter ϵ controls the exploration rate. This case initially uses a small ϵ value to make random choices of the new state. Later, reduce it to follow ϵ-greedy policy by choosing an action that leads to the state with the highest known reward.

4.2.4. Q-Learning-Inspired Antenna Direction Optimization Algorithm

Algorithm 2 (lines 1–18) describes the process of finding the optimal antenna azimuth. The algorithm starts by initializing a Q-table with zeros for all state-action pairs (line 1) and setting the state s to a random azimuth configuration (line 2). In this case, state s represents the current azimuths of all antennas, and the algorithm's actions involve rotating any antenna by 20° or keeping the same azimuth. The algorithm then proceeds through a series of episodes, executing each allowed action and observing the new state s_{new} and the reward r associated with that state (lines 4–15). During exploration (lines 6–8), an action is chosen randomly with probability ϵ. During exploitation (lines 9–10), the action with the highest Q-value in the current state s is selected. For each episode, the algorithm executes each allowed action and observes the resulting new state s_{new} and the reward r associated with that state (lines 11–13).

The Q-table is updated with the observed rewards for each state-action pair (line 14). After executing all allowed actions within an episode, the algorithm selects a new state s_{new} based on the exploration-exploitation trade-off. Specifically, it chooses either a random state or a new state with the highest expected reward based on the Q-table (line 16). The current state s is then updated to the newly selected state s_{new} (line 17). After the algorithm has run all episodes, it outputs the azimuth configuration that maximizes the expected reward according to the learned Q-table (line 3).

Algorithm 2 Q-learning-inspired azimuth optimization

Initialize: Q-table with zeros for all state-action pairs
Initialize: State s using random azimuth configuration
Output: The azimuth configuration that maximizes the expected reward according to the learned Q-table
for $e \leftarrow 1$ to max_episodes **do**
    for $a \leftarrow 1$ to allowed_actions **do**
        if random number $< \epsilon$ **then** ▷ *Exploration*
            Select a random action a
        else ▷ *Exploitation*
            Select action a with highest Q-value in state s
        end if
        Execute action a
        Observe new state s_{new}
        Observe reward r for being in new state s_{new}
        Update Q-table with r for each new state s_{new}
    end for
    Select new state s_{new} with $\arg\max Q(s_{new}, a)$ ▷ *Exploitation*
    Update state: $s \leftarrow s_{new}$
end for

4.2.5. Capacity and Coverage Optimization Results

We test the Q-learning-inspired azimuth optimization algorithm in areas with three and four base stations to assess whether RL methods can be applied to cellular-network planning. Figure 4.12 shows the test results for three and four base stations. Multiple tests were performed to observe the convergence behavior toward the optimal SINR. The upper part of Figure 4.12 shows the average SINR value of three base stations, while the lower part shows four base stations. The x-axis represents the action index, corresponding to azimuth-adjustment steps. Vertical dashed lines distinguish the learning epochs. Fast convergence toward a high SINR is observed, with a clear dependence on the randomly chosen initial state.

The prepared dataset is also evaluated using an exhaustive search, allowing us to compare the Q-learning solution with the global optimum. In most cases, the algorithm reaches the global optimum within six to seven episodes. This is significantly more efficient than performing a full exhaustive search or a random search procedure.

To illustrate the azimuth selection process, an example is given of the maximum SINR obtained in each learning episode in Table 4.3. Table 4.3 summarizes

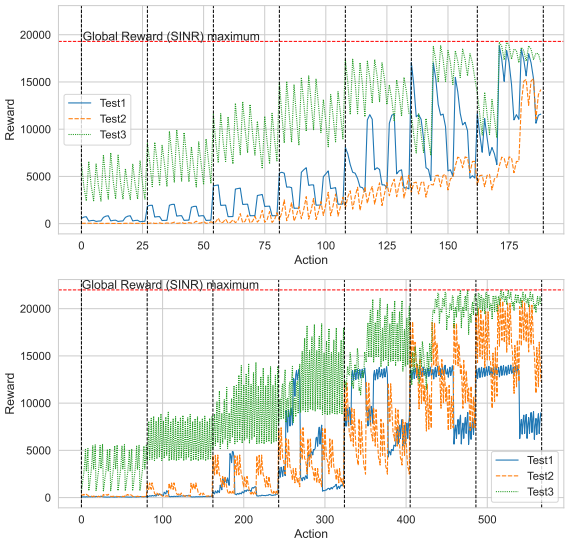


Fig. 4.12. Q-learning-inspired azimuth optimization results for 3 base stations (top) and 4 base stations (bottom) (created by the author)

the azimuth configuration of the three base station tests. It lists the episode number, the current antenna configuration, the selected action, the corresponding average SINR, and the expected SINR after applying the action.

Table 4.3. Q-learning-inspired azimuth optimization results

Episode	Current State	Action	Current SINR	New SINR
1	0, 120, 0	0, 0, -20	4504	7496
2	0, 120, 340	0, 0, -20	7496	9940
3	0, 120, 320	0, 0, -20	9940	12913
4	0, 120, 300	0, 0, -20	12913	15642
5	0, 120, 280	-20, 0, -20	15642	17411
6	340, 120, 260	0, 0, -20	17411	18841
7	340, 120, 240	0, 0, -20	18841	19295

Furthermore, the cumulative distribution function (CDF) of the SINR across all grid cells is presented in Figure 4.13. It is observed that each episode improves not only the average SINR but also the overall SINR distribution across the area. The probability of achieving higher SINR values increases consistently with each learning episode. For example, the probability $\Pr(\text{SINR} \leq 20, \text{dB})$ decreases from approximately 0.8 in Episode 1 to 0.5 in Episode 6, approaching the distribution obtained at the optimal azimuth configuration.

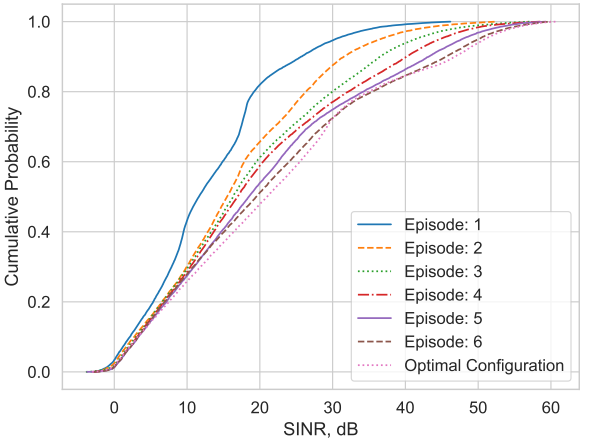


Fig. 4.13. CDF of SINR values in each 10×10 m grid (created by the author)

The results demonstrate that the Q-learning-inspired azimuth optimization algorithm can be effectively applied to cellular-network planning. The azimuth selection process is illustrated in Figure 4.14, showing an example of the maximum SINR obtained in each learning episode. Episode 1 begins with a random azimuth configuration. In subsequent episodes, the antennas follow greedy actions that adjust their azimuths to increase SINR. By Episode 6, the algorithm converges to an azimuth configuration that is close to optimal for the given base-station layout, resulting in high SINR coverage across most of the area.

The proposed algorithm consistently adjusts antenna azimuths to increase SINR. The algorithm is tested with three and four base stations using a 20° step size. In these cases, there is a complexity of $O(18^n)$, where n is the number of states and rewards for 3 or 4 base stations, resulting in 5,832 and 104,976 choices, respectively. In each episode, the algorithm explores all combinations of actions with a complexity of $O(m^n)$, where m is the number of degrees of freedom, such as azimuth, downtilt, and transmit power, and n number of base stations. As additional actions and base stations are considered, the underlying optimization problem becomes NP-hard due to the exponentially expanding search space. Moreover, it is observed that the convergence speed strongly depends on the initial state, and consequently on its associated reward.

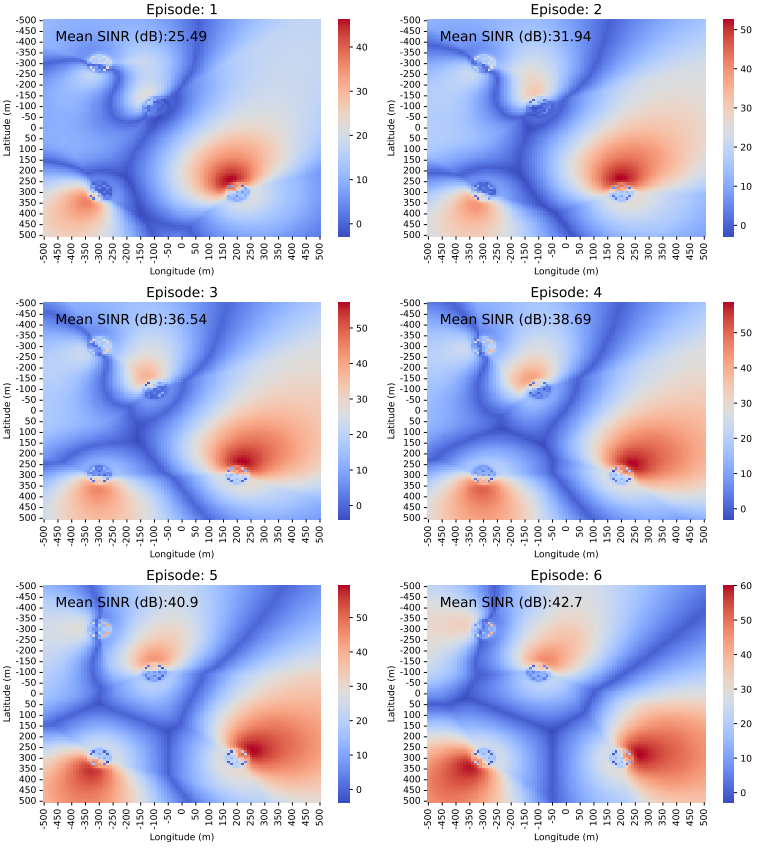


Fig. 4.14. Example of Q-learning actions leading to SINR improvement (created by the author)

4.3. Case Study: Enhancing Quality via Optimized Parameter Selection

Mobile data traffic continues to grow rapidly, placing increasing pressure on MNOs to enhance network capacity and user QoE. One critical approach to improving cellular network performance involves selecting optimal RAN parameters from the broad set defined by 3GPP standards. Specifically, MIMO configurations, such as closed-loop (TM4) and open-loop (TM3), significantly influence user throughput and spectral efficiency.

This subsection presents a case study comparing closed-loop MIMO with open-loop MIMO in a commercial LTE FDD network. Performance was evaluated across three realistic scenarios: high-speed mobility, diverse SINR conditions at

pedestrian speeds, and large-scale deployments encompassing various cell configurations (2T2R and 4T4R). The findings demonstrate that selecting optimal MIMO parameters yields measurable QoE improvements and provides practical guidance for MNOs seeking to maximize network performance and user satisfaction.

4.3.1. Single Cell Capacity Components

Overall, LTE network capacity can be viewed as the aggregate of individual single-cell capacities. The capacity of a single LTE cell, defined as the achievable downlink throughput in its coverage area, depends on three main factors:

- The number of scheduled PRBs (Capozzi et al., 2012);
- The SINR, reported as the CQI, which determines the highest MCS meeting the 10
- MIMO performance, which increases spectral efficiency through the use of multiple antennas.

4.3.2. Single-User Transmission Modes

MIMO systems employ multiple transmitting antennas (Tx) to send signals on the same frequency to multiple receiving antennas (Rx). If multiple antennas are located within the same UE, the configuration is referred to as Single-User (SU) MIMO. In contrast, when a base station equipped with multiple antennas communicates simultaneously with several UEs using the same time–frequency resources, the configuration is known as Multi-User (MU) MIMO (Zheng et al., 2015). In practice, downlink SU MIMO is typically implemented using two or four Tx antennas at the base station. UEs are equipped with two or four Rx antennas, and the penetration of 4-Rx devices reaches approximately 20With multiple Tx and Rx antennas, two major operations are utilized in SU-MIMO: transmit diversity and spatial multiplexing. Spatial multiplexing enhances the UE's spectral efficiency at the cell center, while transmit diversity improves the reliability of UEs located at the cell edge (Liu et al., 2012). LTE supports transmission modes from 1 to 10, as defined in the standard (European Telecommunications Standards Institute, 2014). In practice, the most common options are TM3 (open-loop spatial multiplexing with cyclic delay diversity) and TM4 (closed-loop spatial multiplexing). Figure 4.15 shows the LTE physical layer block diagram for the 2×2 MIMO scenario, which is the most common configuration in commercial networks. The main difference between TM3 and TM4 lies in the type of feedback reported by the UE.

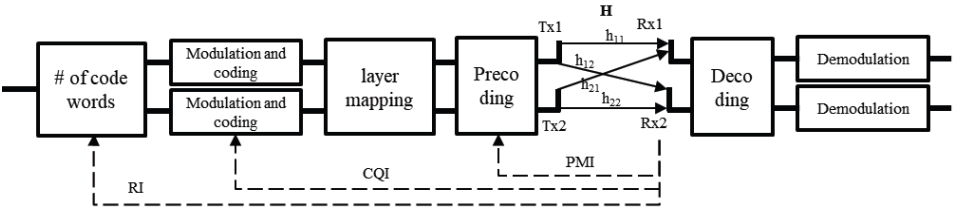


Fig. 4.15. MIMO channel model (2-Tx 2-Rx). PMI is reported only in closed-loop spatial multiplexing mode

In the open-loop transmission mode, the UE reports the Rank Indicator (RI) and CQI values but does not report the Precoding Matrix Indicator (PMI). In TM4 (closed-loop), the UE reports RI, CQI, and PMI values. On the transmitter side, in TM3, the signal is transmitted from each antenna with a specific delay (cyclic delay diversity), thereby artificially creating frequency diversity. This mode requires reduced feedback from the UE and is typically used when channel information is unavailable or rapidly changing, such as for UEs moving at high velocity (Schulz, 2015). TM3 can deliver better performance not only for fast-moving UEs but also when PMI reports are unreliable or when UEs have limited demodulation capabilities. For UEs in static scenarios or moving at walking speed, TM4 provides higher downlink throughput compared to TM3. In high-SINR conditions, TM3 and TM4 are expected to provide similar throughput, as the UE can decode two codewords. At lower SINR levels, TM4 is expected to have an advantage, provided that the channel matrix is accurately reported. The gains of TM4 increase when the number of Tx antennas is larger. In an LTE cell with two transmit antennas, the UE may select a preferred precoding matrix from a codebook with four options; with four transmit antennas, the codebook expands to sixteen options. The precoding process for TM3 and TM4 is described in detail in (European Telecommunications Standards Institute, 2017). For performance evaluation in live mobile networks, the following aspects should be considered: UE PMI reporting capabilities, the proportion of 2-Tx and 4-Tx cells, and typical UE mobility patterns.

4.3.3. Evaluation Scenarios for Transmission Modes

LTE network performance was compared between TM3 and TM4 transmission modes. Three different scenarios were constructed to reflect customer experience under various coverage levels and mobility conditions. These testing scenarios aim to cover different cases, including slow- and fast-moving UEs, varying SINR levels, and large-scale deployments. In all scenarios, the main performance metric was the achievable downlink throughput per user. The first scenario was designed

to measure the achievable downlink throughput at driving speeds. Drive test measurements are widely used for MNO benchmarking. It is important to evaluate the impact of closed-loop MIMO on UEs moving at high speed. It is expected that closed-loop MIMO performs worse in drive tests compared to open-loop due to rapid channel variations and outdated PMI reports (Adhikari, 2011; Schulz, 2015). The drive test cluster consisted of six base stations. These base stations include 24 4T4R and 46 2T2R cells. The test scenario was designed to resemble typical commercial network conditions. The test UE was not locked to a specific frequency or cell; it camped freely on available carriers according to network mobility settings. To evaluate the achievable downlink throughput under TM3 and TM4, large file downloads were conducted using the test UE. To minimize the impact of fluctuating traffic from other clients, the tests were performed on the same cluster under live network conditions, at the same weekday and hour. For a single UE, lower downlink throughput is expected with closed-loop transmission mode due to PMI becoming outdated at high speeds. However, in a live network, where most clients are stationary, it is expected that less PRB will be consumed for the same amount of downlink data, allowing a larger share of PRBs to be allocated to the test UE. To demonstrate the absence of negative impact in the drive test scenario, the UE's speed ranged from 60 to 90 km/h. The second scenario was designed to compare downlink throughput under varying Reference Signal Received Power (RSRP) and SINR levels at walking speed. Throughput measurement samples collected over a wide SINR range can highlight the benefits of closed-loop MIMO for customers in different coverage areas. For the walk test scenario, a test UE capable of locking to a specific cell was used. The UE was locked to a single E-UTRA operating band 3 cell with 20 MHz bandwidth. Without cell locking, the UE could perform handovers to neighboring cells, complicating the collection of measurements across a wide RSRP/SINR range. By locking the UE to a single cell, measurements were collected starting from the cell center and moving toward the cell edge. A comparison between MIMO TM3 and TM4 was performed in a live network environment with other users sharing the cell. This could lead to large fluctuations in the time-frequency resources allocated to the test UE. To mitigate the impact of varying cell loads during measurement, the achievable throughput was normalized to estimate the full cell capacity, R_{norm}, using Equation (4.13):

$$R_{\text{norm}} = R_{\text{RLC}} \cdot \frac{N_{\text{avail}}}{N_{\text{sched}}}, \tag{4.13}$$

where R_{RLC} denotes the measured Radio Link Control (RLC) downlink throughput. The scaling factor is derived from the ratio of the total available Physical Resource Blocks (N_{avail}) to the number of PRBs actually scheduled to the UE (N_{sched})

during the observation window. For the specific case of a 20 MHz bandwidth cell containing 100 PRBs in the frequency domain and a sampling interval of 1000 ms, N_{avail} equals 100,000 PRBs.

All measurements were collected using specialized LTE network measurement equipment across a wide SINR range. The third scenario is the large-scale rollout. The large-scale rollout shows the impact of parameter settings on the overall MNO customer base. Although it does not provide a fine-grained evaluation of TM4 versus TM3 for individual UEs, it offers a statistical view of their impact on overall network performance and capacity.

The average user downlink throughput, denoted as R_{avg}, is calculated using Equation (4.14):

$$R_{\text{avg}} = \frac{V_{\text{PDCP}}}{T_{\text{eff}}}. \tag{4.14}$$

In this formulation, V_{PDCP} represents the total Packet Data Convergence Protocol (PDCP) data volume transferred, explicitly excluding the data transmitted during the last Transmission Time Interval (TTI). Similarly, the denominator T_{eff} corresponds to the total effective transmission time, also excluding the duration of the final TTI to ensure accurate throughput estimation.

4.3.4. Results of Transmission Modes

LTE network performance in terms of downlink user throughput was compared between MIMO TM3 and TM4 configurations across the previously described scenarios. The drive test results show improvements in the average values and all percentile ranges. Table 4.4 shows an increase in the average throughput in the cluster from 105 Mbps to 128 Mbps after switching from TM3 to TM4.

Table 4.4. Measured downlink throughput in the drive test

MIMO mode	Measured downlink throughput, Mbps			
	Mean	Median	25^{th} percentile	75^{th} percentile
TM3	105	96	61	142
TM4	128	110	66	188
Increase, %	22%	15%	8%	32%

Figure 4.16 shows the probability density function (PDF) of the measured downlink throughput. The shift toward higher throughput ranges is visible. Measured downlink throughput increased even for fast-moving UEs, contrary to expectations. A single fast-moving UE in an empty cell is typically expected to achieve

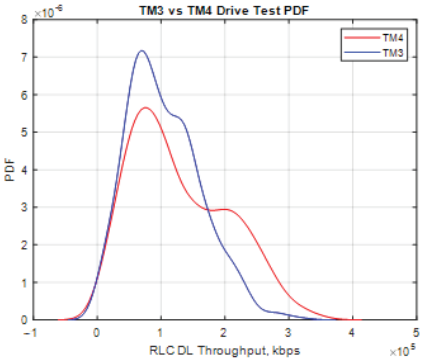


Fig. 4.16. PDF function of throughput measurements during drive test (created by the author)

higher throughput with open-loop MIMO; however, in the live network, closed-loop MIMO outperforms open-loop even for fast-moving UEs. The throughput increase is attributed to both improved MIMO utilization and the availability of additional PRBs, as stationary clients require fewer PRBs to transfer the same amount of data. Figure 4.17 shows the scatter plot of the measurements with an approximating curve.

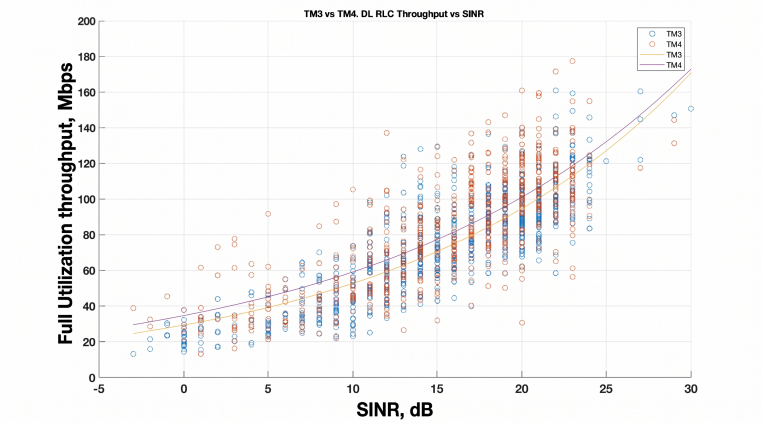


Fig. 4.17. Calculated full utilization throughput at different SINR levels (created by the author)

The presented results were obtained from the walk test scenario, where measurements were made at low speed across a wide SINR range. The closed-loop MIMO configuration demonstrates its advantage, improving downlink throughput compared to open-loop for slow-moving UEs across both low and high SINR

regions. The last results were obtained from the large-scale comparison. It includes thousands of 2T2R and 4T4R cells and hundreds of thousands of UEs. The commercial network comparison included both 2Rx and 4Rx UEs moving at various speeds. This comparison highlights the average throughput differences for a specific commercial network, with similar behavior expected across other LTE MNOs. To mitigate the impact of traffic fluctuations, one week with TM3 enabled was compared to one week with TM4 enabled. Table 4.5 summarizes the daily average user downlink throughput improvements for cells with 2T2R and 4T4R configurations. Even for legacy cells with two transmit antennas, large-scale TM4 activation shows improvements. The average user downlink throughput increased by 13.5% on 2T2R cells. As expected, the gains were even larger for 4T4R cells after switching to TM4 from TM3. The average user downlink throughput increased by 29% on 4T4R cells. The large-scale rollout demonstrates the benefits of closed-loop MIMO selection, improving both customer experience and MNO network capacity.

4.4. Generalized Network-Side Quality of Experience Framework: Recognition, Evaluation, and Optimization

Based on the individual techniques and methods described in the previous chapters and sections, a generalized network-side QoE framework can be proposed as an integrated system. It involves three major components: recognition, evaluation, and optimization within a user-centric QoE scope. Its implementation additionally requires a clear understanding of the following prerequisites: (a) network capabilities and limitations; (b) services and applications, including service types and use cases; (c) QoE scope, including applicable metrics and evaluation methods; and

Table 4.5. Daily downlink-throughput improvement of TM4 over TM3

Day	Improvement (%)	
	2T2R cells	4T4R cells
Mon	9.9	26.8
Tue	12.9	30.1
Wed	13.3	29.6
Thu	13.2	28.5
Fri	19.4	30.7
Sat	13.3	30.5
Sun	12.5	29.6

(d) data accessibility, including relevant protocols, interfaces, and data collection methods. As a result, a practical MNO workflow must: (i) evaluate the network

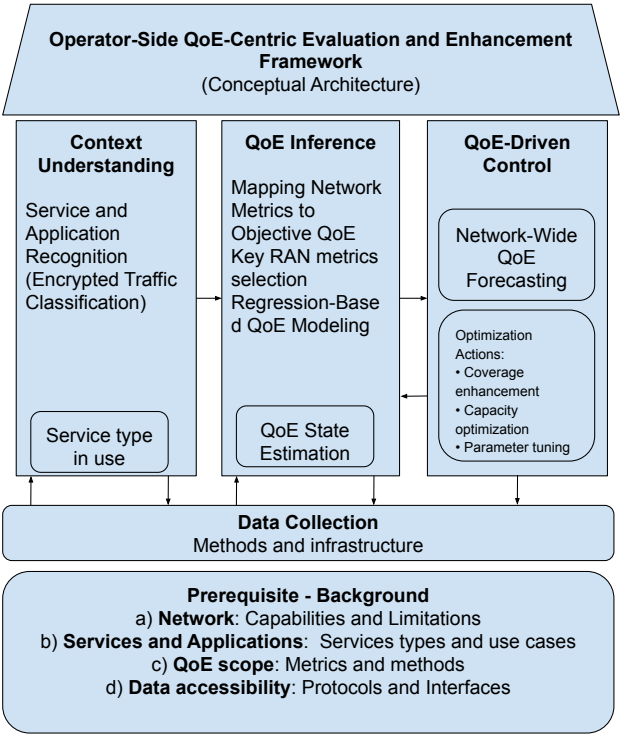


Fig. 4.18. Conceptual architecture of the proposed operator-side QoE-centric evaluation and enhancement framework (created by the author)

infrastructure and application types in order to relate data flows and observation points to network and QoE evaluation methods; (ii) identify the service category and traffic direction; (iii) translate multi-layer network observations into an objective QoE estimate; and (iv) use QoE as an optimization target for network planning and parameter control. The conceptual structure of this framework is illustrated in Figure 4.18. This dissertation contributes to all components of the proposed operator-side QoE-centric framework that integrates three previously separate activities into a single, actionable pipeline: service recognition, QoE evaluation, and QoE-oriented optimization.

4.5. Conclusions of the Fourth Chapter

This chapter investigates QoE-oriented enhancement of LTE networks through traffic forecasting, coverage-and-capacity optimization, and MIMO-parameter tuning. The main conclusions are summarized below.

1. The fbProphet algorithm was shown to be applicable to LTE performance data. Using 180 days of cell-traffic measurements, the proposed two-stage procedure predicts busy-hour traffic and cell load up to 28 days ahead, achieving R^2 scores of approximately 0.9 and enabling proactive congestion detection.

2. A Q-learning-based algorithm for antenna-azimuth selection was developed. Despite the large search space, the algorithm converged within a few episodes to near-optimal azimuths, improving both mean SINR and its distribution.

3. A case study on MIMO transmission-mode selection showed that closed-loop MIMO consistently outperforms open-loop MIMO. Field tests and large-scale rollout data demonstrate downlink throughput gains of 20–30%, confirming that MIMO tuning increases QoE and network capacity.

4. A generalized network-side QoE framework was proposed to consolidate service recognition, empirical QoE evaluation, and enhancement. The framework integrates multi-layer network observations into an objective QoE-oriented workflow and provides a conceptual foundation for QoE-driven network management.

General Conclusions

1. Operator-side QoE evaluation for video streaming is feasible using MNO-observable cross-layer data (RAN and IP transport metrics) without requiring packet-payload inspection or server-side references.

2. Data collected from operational mobile networks enabled a generalized network side QoE framework unifying service recognition, objective QoE evaluation, and QoE-driven enhancement.

3. The proposed CNN-based method demonstrates that HAS video traffic can be identified with practically useful accuracy from low-complexity measurements, achieving a video-versus-other classification accuracy above 0.90 for 300–600 s windows using only aggregated 1 s byte-count sequences.

4. In coverage-limited uplink scenarios, operator visible indicators such as path loss, MCS, PRB allocation, and uplink throughput shows clear relationships with VMAF. Non-linear regression predicts VMAF with $R^2 \approx 0.80$, demonstrating feasible operator-side QoE estimation without payload access.

5. Proactive network management is feasible by combining ML-based KPI prediction with RL-based CCO. The proposed Prophet-based method predicts busy-hour traffic and cell load up to 28 days ahead ($R^2 \approx 0.9$), while Q-learning-based antenna-azimuth selection converges in a few episodes to near-optimal settings which improves mean SINR and its spatial distribution.

References

3rd Generation Partnership Project. (2009, October). *Lte; evolved universal terrestrial radio access (e-utra); physical layer procedures* (Technical Specification No. 3GPP TS 36.213 V8.8.0, Release 8). https://www.etsi.org/deliver/etsi_ts/136200_136299/136213/08.08.00_60/ts_136213v080800p.pdf

3rd Generation Partnership Project. (2010, July). *Umts; lte; telecommunication management; key performance indicators (kpi) for evolved universal terrestrial radio access network (e-utran): Definitions* (Technical Specification No. 3GPP TS 32.450 V9.1.0, Release 9). https://www.etsi.org/deliver/etsi_ts/132400_132499/132450/09.01.00_60/ts_132450v090100p.pdf

3rd Generation Partnership Project. (2019, May). *Universal mobile telecommunications system (umts); lte; 5g; nr; multi-connectivity; overall description; stage-2* (Technical Specification No. 3GPP TS 37.340 V15.5.0, Release 15). Retrieved July 14, 2022, from https://www.etsi.org/deliver/etsi_ts/137300_137399/137340/15.05.00_60/ts_137340v150500p.pdf

3rd Generation Partnership Project. (2025a, January). *5G; NR; User Equipment (UE) radio access capabilities* (tech. rep. No. TS 38.306, version 18.4.0, Release 18). https://www.etsi.org/deliver/etsi_ts/138300_138399/138306/18.04.00_60/ts_138306v180400p.pdf

3rd Generation Partnership Project. (2025b, April). *5G; System architecture for the 5G System (5GS)* (tech. rep. No. TS 23.501, version 18.9.0, Release 18). https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/18.09.00_60/ts_123501v180900p.pdf

Abedin, S. F., Mahmood, A., Tran, N. H., Han, Z., & Gidlund, M. (2022). Elastic O-RAN slicing for industrial monitoring and control: A distributed matching game and

deep reinforcement learning approach. *IEEE Transactions on Vehicular Technology*, *71*(10), 10808–10822. https://doi.org/10.1109/TVT.2022.3188217

Aceto, G., Ciuonzo, D., Montieri, A., & Pescapé, A. (2019). Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges. *IEEE Transactions on Network and Service Management*, *16*(2), 445–458. https://doi.org/10.1109/TNSM.2019.2899085

Adhikari, S. (2011). Downlink transmission mode selection and switching algorithm for lte. *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*, 1–10. https://doi.org/10.1109/COMSNETS.2011.5716434

Agiwal, M., Roy, A., & Saxena, N. (2016). Next generation 5G wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *18*(3), 1617–1655. https://doi.org/10.1109/COMST.2016.2532458

Alan, H. F., & Kaur, J. (2016). Can android applications be identified using only tcp/ip headers of their launch time traffic? *Proceedings of the 9th ACM conference on security & privacy in wireless and mobile networks*, 61–66. https://doi.org/10.1145/2939918.2939929

Alvestrand, H. T. (2021, January). Overview: Real-Time Protocols for Browser-Based Applications. https://doi.org/10.17487/RFC8825

Anwar, M. S., Wang, J., Khan, W., Ullah, A., Ahmad, S., & Fei, Z. (2020). Subjective qoe of 360-degree virtual reality videos and machine learning predictions. *IEEE Access*, *8*, 148084–148099. https://doi.org/10.1109/ACCESS.2020.3015556

Baena, C., Penaherrera-Pulla, O., Camacho, L., Barco, R., & Fortes, S. (2023). Video streaming and cloud gaming services over 4g and 5g: A complete network and service metrics dataset. *IEEE Communications Magazine*. https://doi.org/10.1109/MCOM.005.2200719

Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. (2017). The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data mining and knowledge discovery*, *31*(3), 606–660. https://doi.org/10.1007/s10618-016-0483-9

Barakabitze, A. A., Barman, N., Ahmad, A., Zadtootaghaj, S., Sun, L., Martini, M. G., & Atzori, L. (2019). Qoe management of multimedia streaming services in future networks: A tutorial and survey. *IEEE Communications Surveys & Tutorials*, *22*(1), 526–565. https://doi.org/10.1109/COMST.2019.2958784

Baraković, S., & Skorin-Kapov, L. (2013). Survey and challenges of QoE management issues in wireless networks. *Journal of Computer Networks and Communications*, *2013*. https://doi.org/10.1155/2013/165146

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., & Gamboa, H. (2020). Tsfel: Time series feature extraction library. *SoftwareX*, *11*, 100456. https://doi.org/10.1016/j.softx.2020.100456

Barman, N., & Martini, M. G. (2019). QoE modeling for HTTP adaptive video Streaming–A survey and open challenges. *IEEE Access*, *7*, 30831–30859. https://doi.org/10.1109/ACCESS.2019.2901778

Beerends, J. G., Schmidmer, C., Berger, J., Obermann, M., Ullmann, R., Pomy, J., & Keyhl, M. (2013). Perceptual objective listening quality assessment (POLQA), the third-generation ITU-T standard for end-to-end speech quality measurement, part i—temporal alignment. *Journal of the Audio Engineering Society*, *61*(6), 366–384. https://secure.aes.org/forum/pubs/journal/?elib=16829

Bhat, D., Rizk, A., & Zink, M. (2017). Not so QUIC: A performance study of DASH over QUIC. *Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 13–18. https://doi.org/10.1145/3083165.3083175

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). OpenAI Gym. Retrieved April 27, 2026, from https://arxiv.org/abs/1606.01540

Brunnström, K., Beker, S. A., de Moor, K., Dooms, A., Egger, S., Garcia, M.-N., Hossfeld, T., Jumisko-Pyykkö, S., Keimel, C., & Larabi, M.-C. (2013, March). *Qualinet white paper on definitions of quality of experience* [Output from the 5th Qualinet meeting, Novi Sad, March 12, 2013]. Retrieved November 17, 2023, from https://www.qualinet.eu/resources/qualinet-white-paper/

Buenestado, V., Toril, M., Luna-Ramírez, S., Ruiz-Avilés, J. M., & Mendo, A. (2016). Self-tuning of remote electrical tilts based on call traces for coverage and capacity optimization in LTE. *IEEE Transactions on Vehicular Technology*, *66*(5), 4315–4326. https://doi.org/10.1109/TVT.2016.2605380

Bynder. (2024, January). *Average video length predicted to decline to 65 seconds in 2024* [Press release]. Bynder. Retrieved February 7, 2026, from https://www.bynder.com/en/press-media/bynder-reveals-video-length-set-to-decline/

Capozzi, F., Piro, G., Grieco, L. A., Boggia, G., & Camarda, P. (2012). Downlink packet scheduling in lte cellular networks: Key design issues and a survey. *IEEE communications surveys & tutorials*, *15*(2), 678–700. https://doi.org/10.1109/SURV.2012.060912.00100

Carlucci, G., De Cicco, L., & Mascolo, S. (2015). HTTP over UDP: An experimental investigation of quic. *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, 609–614. https://doi.org/10.1145/2695664.2695706

Casas, P., Seufert, M., Wassermann, S., Gardlo, B., Wehner, N., & Schatz, R. (2022). Deepcrypt-deep learning for qoe monitoring and fingerprinting of user actions in adaptive video streaming. *2022 IEEE 8th International Conference on Network Softwarization (NetSoft)*, 259–263. https://doi.org/10.1109/NetSoft54395.2022.9844113

Chollet, F., et al. (2015). *Keras*. https://keras.io

Conti, M., Mancini, L. V., Spolaor, R., & Verde, N. V. (2015). Analyzing android encrypted network traffic to identify user actions. *IEEE Transactions on Information Forensics and Security*, *11*(1), 114–125. https://doi.org/10.1109/TIFS.2015.2478741

Curley, L., Pugin, K., Nandakumar, S., Vasiliev, V., & Swett, I. (2024, October). *Media over QUIC Transport* (Internet-Draft No. draft-ietf-moq-transport-07) (Work in Progress). Internet Engineering Task Force. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-ietf-moq-transport/07/

Dreifuerst, R. M., Daulton, S., Qian, Y., Varkey, P., Balandat, M., Kasturia, S., Tomar, A., Yazdan, A., Ponnampalam, V., & Heath, R. W. (2021). Optimizing coverage and capacity in cellular networks using machine learning. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8138–8142. https://doi.org/10.1109/ICASSP39728.2021.9414155

Ericsson. (2025, November). *Ericsson mobility report: November 2025* (tech. rep. No. EAB-25:009321 Uen Rev B). Ericsson. SE-164 80 Stockholm, Sweden. Retrieved February 10, 2026, from https://www.ericsson.com/4aca6f/assets/local/reports-papers/mobility-report/documents/2025/ericsson-mobility-report-november-2025.pdf

European Telecommunications Standards Institute. (2014, October). *Lte; evolved universal terrestrial radio access (e-utra); physical layer procedures* (Technical Specification No. ETSI TS 136 213 V12.3.0 (2014-10)). https://www.etsi.org/deliver/etsi_ts/136200_136299/136213/12.03.00_60/ts_136213v120300p.pdf

European Telecommunications Standards Institute. (2017, April). *Lte; evolved universal terrestrial radio access (e-utra); physical channels and modulation* (Technical Specification No. ETSI TS 136 211 V14.2.0 (2017-04)). https://www.etsi.org/deliver/etsi_ts/136200_136299/136211/14.02.00_60/ts_136211v140200p.pdf

European Telecommunications Standards Institute. (2024, October). *5G; 5G Real-time Media Communication Architecture (Stage 2)* (tech. rep. No. TS 26.506, version 18.4.0, Release 18). https://www.etsi.org/deliver/etsi_ts/126500_126599/126506/18.04.00_60/ts_126506v180400p.pdf

García, B., López-Fernández, L., Gortázar, F., & Gallego, M. (2019). Practical evaluation of vmaf perceptual video quality for webrtc applications. *Electronics*, *8*(8), 854. https://doi.org/10.3390/electronics8080854

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org

Gutterman, C., Guo, K., Arora, S., Wang, X., Wu, L., Katz-Bassett, E., & Zussman, G. (2019). Requet: Real-time qoe detection for encrypted youtube traffic. *Proceedings of the 10th ACM Multimedia Systems Conference*, 48–59. https://doi.org/10.1145/3304109.3306226

Holmer, S., Lundin, H., Carlucci, G., De Cicco, L., & Mascolo, S. (2016, July). *A google congestion control algorithm for rtc* (Internet-Draft No. draft-ietf-rmcat-gcc-02) (Work in progress). Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-ietf-rmcat-gcc/02/

Hoßfeld, T., Wunderer, S., Beyer, A., Hall, A., Schwind, A., Gassner, C., Guillemin, F., Wamser, F., Wascinski, K., Hirth, M., Seufert, M., Casas, P., Tran-Gia, P., Robitza, W., Wascinski, W., & Ben Houidi, Z. (2020). White paper on crowdsourced network and QoE measurements – definitions, use cases and challenges. *arXiv*. https://arxiv.org/abs/2006.16896

Http live streaming overview. (2016, March). Apple Inc. https://developer.apple.com/streaming/

Incorporated, A. S. (2012). Real-time messaging protocol (rtmp) specification version 1.0 [Accessed: 2025-01-27]. https://rtmp.veriskope.com/pdf/rtmp_specification_1.0.pdf

International Organization for Standardization. (2012, April). *Information technology—dynamic adaptive streaming over http (dash)—part 1: Media presentation description and segment formats* (Standard No. ISO/IEC 23009-1:2012). https://www.iso.org/standard/57623.html

International Telecommunication Union. (1996, August). *Methods for subjective determination of transmission quality* (Recommendation No. ITU-T P.800). International Telecommunication Union. Retrieved April 27, 2026, from https://www.itu.int/rec/t-rec-p.800-199608-i

International Telecommunication Union. (1999, September). *Subjective video quality assessment methods for multimedia applications* (Recommendation No. ITU-T P.910). International Telecommunication Union. Retrieved April 27, 2026, from https://www.itu.int/rec/t-rec-p.910

International Telecommunication Union. (2008, July). *Vocabulary for performance and quality of service: Amendment 2, new definitions for inclusion in recommendation itu-t p.10/g.100* (ITU-T Recommendation No. P.10/G.100 (2006) Amd. 2). https://www.itu.int/rec/T-REC-P.10-200807-S!Amd2/en

International Telecommunication Union. (2022, June). *Quality of experience (qoe) requirements for real-time multimedia services over 5g networks* (ITU-T Technical Report No. GSTR-5GQoE). https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-QOS-2022-1-PDF-E.pdf

International Telecommunication Union Radiocommunication Sector. (2023, November). *Framework and overall objectives of the future development of IMT for 2030 and beyond* (Recommendation No. ITU-R M.2160-0). https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2160-0-202311-I!!PDF-E.pdf

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd international conference on machine learning* (pp. 448–456, Vol. 37). PMLR. Retrieved April 27, 2026, from https://proceedings.mlr.press/v37/ioffe15.html

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: A review. *Data mining and knowledge discovery*, *33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Jaeckel, S., Raschkowski, L., Börner, K., & Thiele, L. (2014). QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials. *IEEE transactions on antennas and propagation*, *62*(6), 3242–3256. https://doi.org/10.1109/TAP.2014.2310220

Jiang, W. (2022). Cellular traffic prediction with machine learning: A survey. *Expert Systems with Applications*, *201*, 117163. https://doi.org/10.1016/j.eswa.2022.117163

Johansson, I., Dadhich, S., Bodin, U., & Jönsson, T. (2018). Adaptive video with scream over lte for remote-operated working machines. *Wireless Communications and Mobile Computing*, *2018*(1), 3142496. https://doi.org/10.1155/2018/3142496

Kawser, M. T., Hamid, N. I. B., Hasan, M. N., Alam, M. S., & Rahman, M. M. (2012). Downlink SNR to CQI mapping for different multiple antenna techniques in LTE. *International Journal of Information and Electronics Engineering*, *2*(5), 757–760. https://doi.org/10.7763/IJIEE.2012.V2.201

Kay, A. (2007). Tesseract: An open-source optical character recognition engine. *Linux Journal*, *2007*(159), 2. https://doi.org/10.5555/1288165.1288167

Kougioumtzidis, G., Poulkov, V., Zaharis, Z. D., & Lazaridis, P. I. (2022). A survey on multimedia services qoe assessment and machine learning-based prediction. *IEEE Access*, *10*, 19507–19538. https://doi.org/10.1109/ACCESS.2022.3149592

Krishnan, S. S., & Sitaraman, R. K. (2013). Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. *IEEE/ACM Transactions on Networking*, *21*(6), 2001–2014. https://doi.org/10.1109/TNET.2013.2281542

Kua, J., Armitage, G., & Branch, P. (2017). A survey of rate adaptation techniques for dynamic adaptive streaming over http. *IEEE Communications Surveys & Tutorials*, *19*(3), 1842–1866. https://doi.org/10.1109/COMST.2017.2685630

Langley, A., Riddoch, A., Wilk, A., Vicente, A., Krasic, C., Zhang, D., Yang, F., Kouranov, F., Swett, I., Iyengar, J., et al. (2017). The quic transport protocol: Design and internet-scale deployment. *Proceedings of the conference of the ACM special interest group on data communication*, 183–196. https://doi.org/10.1145/3098822.3098842

Layek, M. A., et al. (2017). Performance analysis of h.264, h.265, vp9 and av1 video encoders. *2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. https://doi.org/10.1109/APNOMS.2017.8094162

Le Maguer, S., King, S., & Harte, N. (2024). The limits of the mean opinion score for speech synthesis evaluation. *Computer Speech & Language*, *84*, 101577. https://doi.org/10.1016/j.csl.2023.101577

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Zhang, Y.-J. A. (2019). The roadmap to 6G: AI empowered wireless networks. *IEEE Communications Magazine*, *57*(8), 84–90. https://doi.org/10.1109/MCOM.2019.1900271

Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., & Manohara, M. (2016, June 6). *Toward a practical perceptual video quality metric*. Netflix Technology Blog. Retrieved April 27, 2026, from https://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html

Lin, X. (2022). An overview of 5g advanced evolution in 3gpp release 18. *IEEE Communications Standards Magazine*, *6*(3), 77–83. https://doi.org/10.1109/MCOMSTD.0001.2200001

Liu, G., Huang, Y., Chen, Z., Liu, L., Wang, Q., & Li, N. (2020). 5g deployment: Standalone vs. non-standalone from the operator perspective. *IEEE Communications Magazine*, *58*(11), 83–89. https://doi.org/10.1109/MCOM.001.2000230

Liu, L., Chen, R., Geirhofer, S., Sayana, K., Shi, Z., & Zhou, Y. (2012). Downlink mimo in lte-advanced: Su-mimo vs. mu-mimo. *IEEE Communications Magazine*, *50*(2), 140–147. https://doi.org/10.1109/MCOM.2012.6146493

Liu, X., Chuai, G., Wang, X., Xu, Z., Gao, W., Zhang, K., Liu, Q., Maimaiti, S., & Zuo, P. (2022). QoE-driven antenna tuning in cellular networks with cooperative multi-agent reinforcement learning. *IEEE Transactions on Mobile Computing*. https://doi.org/10.1109/TMC.2022.3230711

Liu, Y., Dey, S., Ulupinar, F., Luby, M., & Mao, Y. (2015). Deriving and validating user experience model for dash video streaming. *IEEE Transactions on Broadcasting*, *61*(4), 651–665. https://doi.org/10.1109/TBC.2015.2460611

Lopez-Martin, M., Carro, B., Sanchez-Esguevillas, A., & Lloret, J. (2017). Network traffic classifier with convolutional and recurrent neural networks for internet of things. *IEEE access*, *5*, 18042–18050. https://doi.org/10.1109/ACCESS.2017.2747560

Lotfollahi, M., Jafari Siavoshani, M., Shirali Hossein Zade, R., & Saberian, M. (2020). Deep packet: A novel approach for encrypted traffic classification using deep learning. *Soft Computing*, *24*(3), 1999–2012. https://doi.org/10.1007/s00500-019-04030-2

Mangla, T., Halepovic, E., Ammar, M., & Zegura, E. (2019). Using session modeling to estimate http-based video qoe metrics from encrypted network traffic. *IEEE Transactions on Network and Service Management*, *16*(3), 1086–1099. https://doi.org/10.1109/TNSM.2019.2924942

MDN contributors. (2026, January). *Web video codec guide*. Retrieved May 19, 2025, from https://developer.mozilla.org/en-US/docs/Web/Media/Guides/Formats/Video_codecs

Metzger, F., Geißler, S., Grigorjew, A., Loh, F., Moldovan, C., Seufert, M., & Hoßfeld, T. (2022). An introduction to online video game qos and qoe influencing factors. *IEEE Communications Surveys & Tutorials*, *24*(3), 1894–1925. https://doi.org/10.1109/COMST.2022.3177251

Min, X., Zhai, G., Zhou, J., Farias, M. C., & Bovik, A. C. (2020). Study of subjective and objective quality assessment of audio-visual signals. *IEEE Transactions on Image Processing*, *29*, 6054–6068. https://doi.org/10.1109/TIP.2020.2988148

Mondal, B., Thomas, T. A., Visotsky, E., Vook, F. W., Ghosh, A., Nam, Y.-H., Li, Y., Zhang, J., Zhang, M., Luo, Q., et al. (2015). 3D channel model in 3GPP. *IEEE Communications Magazine*, *53*(3), 16–23. https://doi.org/10.1109/MCOM.2015.7060514

Naderi, B., & Cutler, R. (2020). An open source implementation of ITU-T recommendation P.808 with validation. Retrieved April 27, 2026, from https://arxiv.org/abs/2005.08138

Nam, H., Kim, K., Schulzrinne, H., & Calin, D. (2014). YouSlow: A performance analysis tool for adaptive bitrate video streaming [Demo/Poster]. *Proceedings of SIGCOMM 2014*. https://doi.org/10.1145/2619239.2631433

Nguyen, N.-T., & Liu, B.-H. (2019). The mobile sensor deployment problem and the target coverage problem in mobile wireless sensor networks are NP-hard. *IEEE Systems Journal*, *13*(2), 1312–1315. https://doi.org/10.1109/JSYST.2018.2828879

Papadogiannaki, E., & Ioannidis, S. (2021). A survey on encrypted network traffic analysis applications, techniques, and countermeasures. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35. https://doi.org/10.1145/3457904

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. Retrieved April 27, 2026, from https://jmlr.org/papers/v12/pedregosa11a.html

Pinson, M. H., & Wolf, S. (2004). A new standardized method for objectively measuring video quality. *IEEE Transactions on broadcasting*, *50*(3), 312–322. https://doi.org/10.1109/TBC.2004.834028

Pires, K., & Simon, G. (2014). DASH in Twitch: Adaptive bitrate streaming in live game streaming platforms. *Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming*, 13–18. https://doi.org/10.1145/2676652.2676657

Polese, M., Bonati, L., D'oro, S., Basagni, S., & Melodia, T. (2023). Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges. *IEEE Communications Surveys & Tutorials*, *25*(2), 1376–1411. https://doi.org/10.1109/COMST.2023.3239220

Raida, V., Svoboda, P., Lerch, M., & Rupp, M. (2019). Crowdsensed performance benchmarking of mobile networks. *IEEE Access*, *7*, 154899–154911. https://doi.org/10.1109/ACCESS.2019.2949051

Rämö, A., & Toukomaa, H. (2015). Subjective quality evaluation of the 3gpp evs codec. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5157–5161. https://doi.org/10.1109/ICASSP.2015.7178954

Rassool, R. (2017). Vmaf reproducibility: Validating a perceptual practical video quality metric. *2017 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB)*, 1–2. https://doi.org/10.1109/BMSB.2017.7986143

Rezaei, S., & Liu, X. (2019a). Deep learning for encrypted traffic classification: An overview. *IEEE Commun. Mag.*, *57*(5), 76–81. https://doi.org/10.1109/MCOM.2019.1800819

Rezaei, S., & Liu, X. (2019b). Deep learning for encrypted traffic classification: An overview. *IEEE communications magazine*, *57*(5), 76–81. https://doi.org/10.1109/MCOM.2019.1800819

Robitza, W., Göring, S., Raake, A., Lindegren, D., Heikkilä, G., Gustafsson, J., List, P., Feiten, B., Wüstenhagen, U., Garcia, M.-N., et al. (2018). Http adaptive streaming qoe estimation with itu-t rec. p. 1203: Open databases and software. *Proceedings of the 9th ACM Multimedia Systems Conference*, 466–471. https://doi.org/10.1145/3204949.3208124

Roettgers, J. (2013a, March 13). *Don't touch that dial: How YouTube is bringing adaptive streaming to mobile, TVs* [Originally published by GigaOM]. GigaOM. Retrieved March 10, 2021, from https://finance.yahoo.com/news/don-t-touch-dial-youtube-224155787.html

Roettgers, J. (2013b, March 13). *Don't touch that dial: How YouTube is bringing adaptive streaming to mobile, TVs* [Originally published by GigaOM]. GigaOM. Retrieved May 19, 2025, from https://finance.yahoo.com/news/don-t-touch-dial-youtube-224155787.html

Roskind, J. (2013, November). *QUIC: Quick UDP internet connections: Multiplexed stream transport over UDP* (IETF-88 TSV Area Presentation). Internet Engineering Task Force. Retrieved April 27, 2026, from https://www.ietf.org/proceedings/88/slides/slides-88-tsvarea-10.pdf

Salo, J., & Zacarías, E. (2017). Analysis of LTE radio load and user throughput. *International Journal of Computer Networks and Communications*, *9*(6), 33–45. https://doi.org/10.5121/ijcnc.2017.9603

Sandvine. (2021, May). *The mobile internet phenomena report* (Report). Sandvine. Retrieved May 19, 2025, from https://www.applogicnetworks.com/hubfs/Sandvine_Redesign_2019/Downloads/2021/Phenomena/MIPR%20Q1%202021%2020210510.pdf

Schatz, R., Schwarzmann, S., Zinner, T., Dobrijevic, O., Liotou, E., Pocta, P., Barakovic, S., Barakovic Husic, J., & Skorin-Kapov, L. (2018). Qoe management for future networks. *Autonomous Control for a Reliable Internet of Services: Methods, Models, Approaches, Techniques, Algorithms, and Tools*, 49–80. https://doi.org/10.1007/978-3-319-90415-3_3

Schulz, B. (2015, July). *Lte transmission modes and beamforming* (White Paper No. 1MA186-2e). Rohde & Schwarz. https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/dl_application/application_notes/1ma186/1MA186_2e_LTE_TMs_and_beamforming.pdf

Seufert, A., Wamser, F., Wunderer, S., Hall, A., & Hoßfeld, T. (2021). Trust but verify: Crowdsourced mobile network measurements and statistical validity measures. *2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit)*, 443–448. https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482428

Seufert, M., Egger, S., Slanina, M., Zinner, T., Hoßfeld, T., & Tran-Gia, P. (2014). A survey on quality of experience of http adaptive streaming. *IEEE Communications Surveys & Tutorials*, *17*(1), 469–492. https://doi.org/10.1109/COMST.2014.2360940

Shafi, M., Molisch, A. F., Smith, P. J., Haustein, T., Zhu, P., De Silva, P., Tufvesson, F., Benjebbour, A., & Wunder, G. (2017). 5g: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE journal on selected areas in communications*, *35*(6), 1201–1221. https://doi.org/10.1109/JSAC.2017.2692307

Sharabayko, M., Sharabayko, M., Dube, J., Kim, J., & Kim, J. (2021, September). *The SRT Protocol* (Internet-Draft No. draft-sharabayko-srt-01) (Work in Progress). Internet Engineering Task Force. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-sharabayko-srt/01/

Shen, M., Zhang, J., Xu, K., Zhu, L., Liu, J., & Du, X. (2020). Deepqoe: Real-time measurement of video qoe from encrypted traffic with deep learning. *2020 IEEE/ACM*

28th International Symposium on Quality of Service (IWQoS), 1–10. https://doi.org/10.1109/IWQoS49365.2020.9212897

Sun, Y., Peng, M., Zhou, Y., Huang, Y., & Mao, S. (2019). Application of machine learning in wireless networks: Key techniques and open issues. *IEEE Communications Surveys & Tutorials*, *21*(4), 3072–3108. https://doi.org/10.1109/COMST.2019.2924243

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. Retrieved April 27, 2026, from https://incompleteideas.net/book/the-book-2nd.html

Tang, S., Qin, X., & Wei, G. (2018). Network-based video quality assessment for encrypted http adaptive streaming. *IEEE Access*, *6*, 56246–56257. https://doi.org/10.1109/ACCESS.2018.2872932

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Taylor, V. F., Spolaor, R., Conti, M., & Martinovic, I. (2017). Robust smartphone app identification via encrypted network traffic analysis. *IEEE Transactions on Information Forensics and Security*, *13*(1), 63–78. https://doi.org/10.1109/TIFS.2017.2737970

Tekgul, E., Novlan, T., Akoum, S., & Andrews, J. G. (2022). Uplink-downlink joint antenna optimization in cellular systems with sample-efficient learning. *IEEE Global Communications Conference*, 6499–6504. https://doi.org/10.1109/GLOBECOM48099.2022.10001578

Tong, W., & Zhu, P. (Eds.). (2021). *6G: The next horizon: From connected people and things to connected intelligence*. Cambridge University Press. Retrieved April 27, 2026, from https://www.cambridge.org/core/books/6g-the-next-horizon/B12A6013018F4FA14441DA47EEECCC54

Tran, H. T., Ngoc, N. P., Hoßfeld, T., Seufert, M., & Thang, T. C. (2021). Cumulative quality modeling for http adaptive streaming. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *17*(1), 1–24. https://doi.org/10.1145/3423421

Uitto, M., & Heikkinen, A. (2021). Evaluation of live video streaming performance for low latency use cases in 5g. *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 431–436. https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482605

Uitto, M., & Heikkinen, A. (2022). Evaluating 5g uplink performance in low latency video streaming. *2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 393–398. https://doi.org/10.1109/EuCNC/6GSummit54941.2022.9815703

Van Chien, T., & Björnson, E. (2017). Massive MIMO communications. *5G mobile communications*, 77–116. https://doi.org/10.1007/978-3-319-34208-5_4

Velan, P., Čermák, M., Čeleda, P., & Drašar, M. (2015). A survey of methods for encrypted traffic classification and analysis. *International Journal of Network Management*, *25*(5), 355–374. https://doi.org/10.1002/nem.1901

Wang, P., Chen, X., Ye, F., & Sun, Z. (2019). A survey of techniques for mobile service encrypted traffic classification using deep learning. *IEEE Access*, *7*, 54024–54033. https://doi.org/10.1109/ACCESS.2019.2912896

Wang, Z., Yan, W., & Oates, T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. *2017 International joint conference on neural networks (IJCNN)*, 1578–1585. https://doi.org/10.1109/IJCNN.2017.7966039

Wieckowski, A., Lehmann, C., Bross, B., Marpe, D., Biatek, T., Raulet, M., & Le Feuvre, J. (2021). A complete end to end open source toolchain for the versatile video coding (vvc) standard. *Proceedings of the 29th ACM International Conference on Multimedia*, 3795–3798. https://doi.org/10.1145/3474085.3478320

Wijethilaka, S., & Liyanage, M. (2021). Survey on network slicing for internet of things realization in 5g networks. *IEEE Communications Surveys & Tutorials*, *23*(2), 957–994. https://doi.org/10.1109/COMST.2021.3067807

Xinggong, Z., Lan, X. I. E., & Zongming, G. U. O. (2019). Quality assessment and measurement for internet video streaming. *ZTE Communications*. https://doi.org/10.12142/ZTECOM.201901003

Xu, M., Wang, Q., & Lin, Q. (2018). Hybrid holiday traffic predictions in cellular networks. *Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 1–6. https://doi.org/10.1109/NOMS.2018.8406291

Yang, M., Wang, S., Calheiros, R. N., & Yang, F. (2018). Survey on qoe assessment approach for network service. *Ieee Access*, *6*, 48374–48390. https://doi.org/10.1109/ACCESS.2018.2867253

Yao, H., Liu, C., Zhang, P., Wu, S., Jiang, C., & Yu, S. (2019). Identification of encrypted traffic through attention mechanism based long short term memory. *IEEE transactions on big data*, *8*(1), 241–252. https://doi.org/10.1109/TBDATA.2019.2940675

Yeh, S.-P., Bhattacharya, S., Sharma, R., & Moustafa, H. (2023). Deep learning for intelligent and automated network slicing in 5g open ran (oran) deployment. *IEEE Open Journal of the Communications Society*, *5*, 64–70. https://doi.org/10.1109/OJCOMS.2023.3337854

Zhang, S. (2019). An overview of network slicing for 5g. *IEEE Wireless Communications*, *26*(3), 111–117. https://doi.org/10.1109/MWC.2019.1800234

Zhang, Y., Cheng, S., Guo, Z., & Zhang, X. (2024). Inferring video streaming quality of real-time communication inside network. *IEEE Transactions on Circuits and Systems for Video Technology*. https://doi.org/10.1109/TCSVT.2024.3375604

Zheng, K., Zhao, L., Mei, J., Shao, B., Xiang, W., & Hanzo, L. (2015). Survey of large-scale mimo systems. *IEEE Communications Surveys & Tutorials*, *17*(3), 1738–1760. https://doi.org/10.1109/COMST.2015.2425294

Zheng, W., Zhong, J., Zhang, Q., & Zhao, G. (2022). Mtt: An efficient model for encrypted network traffic classification using multi-task transformer. *Applied Intelligence*, *52*(9), 10741–10756. https://doi.org/10.1007/s10489-021-03032-8

List of Scientific Publications by the Author on the Topic of the Dissertation

Papers in the Reviewed Scientific Journals

Chmieliauskas, D. (2021). Fdd lte mimo closed-loop vs open-loop performance evaluation in commercial network. *Mokslas–Lietuvos ateitis/Science–Future of Lithuania*, *13*. https://doi.org/10.3846/mla.2021.15176

Chmieliauskas, D., & Paulikas, Š. (2023). Video stream recognition using bitstream shape for mobile network qoe. *Sensors*, *23*(5), 2548. https://doi.org/10.3390/s23052548

Chmieliauskas, D., & Paulikas, Š. (2025). Evaluation of uplink video streaming qoe in 4g and 5g cellular networks using real-world measurements. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3554340

Papers in Other Editions

Chmieliauskas, D., & Guršnys, D. (2019). Lte cell traffic grow and congestion forecasting. *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 1–5. https://doi.org/10.1109/eStream.2019.8732145

Chmieliauskas, D., & Paulikas, Š. (2021). Ott video over mobile networks: Operators cost and challenges. *2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 1–5. https://doi.org/10.1109/eStream53087.2021.9431414

Chmieliauskas, D., & Paulikas, Š. (2022). Application recognition based on bit-stream shape in mobile networks. *2022 IEEE Open Conference of Electrical, Electronic*

and Information Sciences (eStream), 1–4. https://doi.org/10.1109/eStream56157.2022.9781681

Chmieliauskas, D., Mahmood, A., Thar, K., Paulikas, Š., & Gidlund, M. (2023). Q-learning inspired method for antenna azimuth selection in cellular networks. *2023 Workshop on Microwave Theory and Technology in Wireless Communications (MTTW)*, 7–12. https://doi.org/10.1109/MTTW59774.2023.10320055

Chmieliauskas, D., & Paulikas, Š. (2024). Real-time uplink video streaming over 4g and 5g: Qoe and network metrics dataset. *2024 IEEE 11th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 1–6. https://doi.org/10.1109/AIEEE62837.2024.10586711

Summary in Lithuanian

Įvadas

Problemos formulavimas

Mobiliojo ryšio duomenų srautas kasmet auga daugiau nei 20 %, o vaizdo turinys sudaro 76 % visų perduodamų duomenų (Ericsson, 2025). Nepaisant 5G plėtros, aukštos patirties kokybės (angl. *Quality of Experience* – QoE) užtikrinimas mobiliojo ryšio operatoriams (angl. *Mobile Network Operators* – MNO) išlieka reikšmingu iššūkiu. Tai ypač aktualu vaizdo transliacijų paslaugoms dėl didelių pralaidumo reikalavimų ir jautrumo tinklo delsai. Šioje disertacijoje nagrinėjama QoE vertinimo ir valdymo problema, siūlomas sisteminis modelis, kuriame atsižvelgiama į šiuolaikinių tinklų sudėtingumą. Tikslui pasiekti būtina spręsti kelias tarpusavyje susijusias problemas. Pirmoji iš jų yra ribotas paslaugų matomumas dėl srauto šifravimo. Nustačius paslaugos tipą, reikia spręsti kitą uždavinį – vertinti vaizdo transliacijų QoE. Radijo kanalas yra paskutinė perdavimo grandis duomenų, vaizdo ir balso paslaugoms. Vartotojui pasiekiama sparta ir delsa kinta dėl: (a) kintančios radijo signalo kokybės, (b) konkurencijos dėl bendros perdavimo terpės su kitais vartotojais ir (c) nežinomų esamos sesijos delsos bei pralaidumo poreikių.

Stochastinė RF aplinka ir atsitiktinis vartotojų prieigos prie tinklo pobūdis lemia sunkiai prognozuojamą paslaugų kokybę (QoS), o tai dar labiau apsunkina vartotojo patirties kokybės (QoE) vertinimą. Priimamojo ryšio vaizdo transliavimas yra plačiai ištirtas, tačiau didėjanti siunčiamojo ryšio svarba, kurią lemia tiesioginio turinio kūrimo ir nuotolinio valdymo taikymo scenarijų plėtra, reikalauja specializuotų vertinimo modelių. Galiausiai aukšta QoE priklauso nuo tinklo aprėpties ir veikimo efektyvumo. Todėl šiam tikslui pasiekti būtina pereiti nuo įrangos veikimo parametrų derinimo prie į vartotoją orientuoto,

QoE grindžiamo tinklo valdymo.

Disertacija priskirtina elektros ir elektronikos inžinerijos mokslo krypčiai, tiksliau – ryšių inžinerijos, signalų apdorojimo ir tinklų sistemų valdymo sandūrai. Tyrime sprendžiami uždaviniai, susiję su patikimu paslaugų atpažinimu šifruotame sraute, QoE vertinimu ir prognozavimu siunčiamojo ryšio realaus laiko vaizdo perdavimo atvejais, taip pat radijo prieigos tinklo (angl. *Radio Access Network* – RAN) konfigūracijos ir aprėpties gerinimu, siekiant maksimalios QoE stochastinėmis radijo ryšio sąlygomis.

Darbo aktualumas

Mobiliojo ryšio tinklų paslaugų strateginė svarba sparčiai auga. Tai lemia visur paplitęs daugialypės terpės turinio transliavimas pramogų tikslais ir pramoniniai taikymai, tokie kaip tiesioginė žurnalistika, nuomonės formuotojų turinio kūrimas bei nuotolinio valdymo sprendimai išmaniojo miesto infrastruktūroje. Visais šiais atvejais būtinas patikimas QoE vertinimas. Akademinė bendruomenė ir telekomunikacijų pramonė aktyviai kuria ateities korinio ryšio sistemas bei tobulina 5G ir vėlesnių kartų tinklus. Šios inovacijos apima tinklo skaidymą (angl. *network slicing*), atvirojo radijo prieigos tinklo (angl. O-RAN) architektūrą su išmaniuoju RAN valdikliu (angl. RIC), skirtu į QoE orientuotai optimizacijai, taip pat tinklo duomenų analizės funkciją (angl. NWDAF), skirtą QoE valdymui.

Tyrimo objektas

Vaizdo srautinio perdavimo paslaugų objektyvios patirties kokybės vertinimas ir valdymas iš operatoriaus perspektyvos 4G LTE ir 5G NR judriojo ryšio tinkluose.

Darbo tikslas

Sukurti ir eksperimentiškai įvertinti tinklo pusėje veikiančią sistemą, kuri leistų mobiliojo ryšio operatoriams atpažinti naudojamas paslaugas, vertinti ir prognozuoti vaizdo srautinio perdavimo paslaugų kokybę (QoE) bei vykdyti proaktyvų jos valdymą.

Darbo uždaviniai

Darbo tikslui pasiekti sprendžiami šie uždaviniai:

1. Išanalizuoti ištisinį duomenų perdavimo kelią 4G LTE ir 5G NR tinkluose ir nustatyti, kurios radijo, transporto ir taikomojo lygmenų metrikos yra prieinamos mobiliojo ryšio operatoriui.

2. Surinkti ir suklasifikuoti mobiliojo ryšio operatoriui prieinamus duomenų šaltinius (radijo ryšio KPI, transporto protokolų metrikas ir medijos lygmens duomenis) bei nustatyti jų ryšį su vaizdo srautinio perdavimo paslaugų QoE apibrėžimais.

3. Sukurti mažų skaičiavimo sąnaudų ir šifravimui atsparų vaizdo srautinio perdavimo paslaugų atpažinimo algoritmą, pagrįstą vien tik iš tinklo pusės prieinamais duomenimis.

4. Suprojektuoti ir validuoti modelį, transformuojantį RAN ir IP metrikas į objektyvius QoE įverčius, skirtus vertinti realaus laiko siunčiamosios krypties vaizdo srautinio perdavimo paslaugų kokybę.

5. Sukurti ir įvertinti mašininio mokymosi ir skatinamojo mokymosi modelius, skirtus RAN talpai prognozuoti ir QoE pagrįstam RAN parametrų valdymui palaikyti.

Tyrimų metodika

Tyrimo metodika apima sisteminę 4G LTE ir 5G NR tinklų architektūrų bei vaizdo srautinio perdavimo paslaugų QoE vertinimo metodų literatūros analizę. Hipotezėms patikrinti suprojektuotos ir įgyvendintos duomenų surinkimo sistemos, apimančios komercinius judriojo ryšio tinklus, vartotojo įrangos ir tinklo dalies matavimus bei etaloninį (angl. *full-reference*) medijos kokybės vertinimą. Plataus masto lauko matavimai realiuose LTE ir 5G tinkluose atlikti siekiant atpažinti paslaugas, taikyti objektyvios QoE regresiją, prognozuoti srautą ir įgyvendinti į QoE orientuotą RAN tobulinimą. Taip pat atlikti duomenų inžinerijos darbai, taikant specializuotas programines priemones ir modeliavimo aplinkas. Tyrime naudotos „Android“ platformos duomenų rinkimo priemonės, „Wireshark“ / PCAP analizė, „OBS Studio“ ir „RTC Studio“ vaizdo srautams generuoti, „FFmpeg“ / „ffprobe“ bei OCR metodais pagrįstas kadrų sinchronizavimas, „Python“ aplinkoje įgyvendinti mašininio mokymosi metodai, naudojant „Keras“, „scikit-learn“ ir „Prophet“, taip pat „MATLAB“ vykdytas „QuaDRiGa“ modeliavimas ir „Gym“ pagrindu sukurta skatinamojo mokymosi aplinka.

Darbo mokslinis naujumas

Disertacijoje pateikiamas mokslinis indėlis:

1. Sukurta unifikuota operatoriaus pusėje veikianti 4G LTE ir 5G NR tinklų sistema, skirta objektyviajai patirties kokybei vertinti, remiantis RAN ir IP transporto lygmenų metrikomis.

2. Pasiūlytas mažų skaičiavimo sąnaudų duomenų srauto klasifikavimo metodas, pagrįstas konvoliuciniu neuroniniu tinklu, kuris analizuoja bitų srauto formą kaip laiko eilutę ir leidžia atpažinti vaizdo srautinio perdavimo paslaugas, pasiekdamas sudėtingiems DPI (angl. *Deep Packet Inspection*) metodams artimą tikslumą.

3. Sukurti ir validuoti regresijos modeliai, leidžiantys prognozuoti objektyvius QoE įverčius realaus laiko siunčiamosios krypties vaizdo srautinio perdavimo paslaugoms, remiantis stebimomis radijo sąsajos metrikomis.

4. Pasiūlyti tinklo valdymo metodai, tokie kaip „Prophet“ modeliu pagrįstas srauto ir perkrovos prognozavimo algoritmas bei skatinamojo mokymosi pagrindu sukurtas dinaminio antenos azimuto parinkimo algoritmas.

Darbo rezultatų praktinė reikšmė

Nustatytas formalizuotas ryšys tarp RAN metrikų ir objektyviosios patirties kokybės (QoE) realaus laiko vaizdo paslaugoms, įgalinantis automatizuotą, į QoE orientuotą tinklo valdymą eksploatuojamuose LTE, 5G NR ir ateities mobiliojo ryšio tinkluose. Be to, sukurtas duomenų srauto analize pagrįstas metodas, apdorojantis tik 1 s trukmės bitų sekas ir leidžiantis atpažinti paslaugas mobiliojo ryšio tinklo operatoriaus (MNO) pusėje, išlaikant mažas skaičiavimo sąnaudas bei užtikrinant vartotojų privatumo apsaugą. Siūlomi modeliai yra apmokyti ir validuoti naudojant didelės apimties matavimų duomenis, surinktus iš realiai veikiančių LTE ir 5G tinklų, o ne remiantis simuliacijomis. Duomenų rinkiniai ir programinis kodas yra viešai prieinami tolesniems moksliniams tyrimams.

Ginamieji teiginiai

1. Realaus laiko siunčiamosios krypties vaizdo srautinio perdavimo objektyvioji patirties kokybė, išreikšta VMAF rodikliu, gali būti prognozuojama remiantis mobiliojo ryšio operatoriui prieinamais sklidimo slopinimo (angl. *path loss*), moduliacijos ir kodavimo schemos (MCS) bei priskirtų fizinių išteklių blokų (PRB) rodikliais, pasiekiant didesnes nei 0,79 R^2 reikšmes.

2. Vaizdo srautinio perdavimo duomenų srautas gali būti identifikuojamas naudojant agreguotas 1 s bitų sekas ir sąsūkos neuroninį tinklą (CNN) be giliosios paketų analizės (DPI), pasiekiant didesnį nei 0,9 klasifikavimo tikslumą.

3. Piko valandų tinklo ląstelių duomenų srautas ir perkrova gali būti prognozuojami iki 28 dienų laikotarpiui, taikant modifikuotą dviejų pakopų „Prophet“ pagrįstą metodą, kurio prognozavimo tikslumas pasiekia 0,89 R^2.

4. Vaizdo srautinio perdavimo paslaugų atpažinimas ir objektyvios patirties kokybės prognozavimas gali būti atliekami naudojant operatoriaus pusėje veikiančią, į QoE orientuotą sistemą, pagrįstą vien tik mobiliojo ryšio operatoriui prieinamais RAN matavimais ir IP transporto metrikomis.

Darbo rezultatų aprobavimas

Pagrindiniai disertacijos rezultatai paskelbti aštuoniuose moksliniuose straipsniuose: du straipsniai – užsienio mokslo žurnaluose, įtrauktuose į *Clarivate Analytics Web of Science Core Collection* duomenų bazę, vienas – nacionaliniame recenzuojamame žurnale, o likę penki – tarptautinių konferencijų pranešimų medžiagoje.

- eStream 2019: *Open Conference of Electrical, Electronic and Information Sciences*, 2019 m., Vilniuje, Lietuvoje.
- eStream 2021: *IEEE Open Conference of Electrical, Electronic and Information Sciences*, 2021 m., Vilniuje, Lietuvoje.
- eStream 2022: *IEEE Open Conference of Electrical, Electronic and Information Sciences*, 2022 m., Vilniuje, Lietuvoje.

– MTTW 2023: *Workshop on Microwave Theory and Techniques in Wireless Communications*, 2023 m., Rygoje, Latvijoje.
– AIEEE 2024: *IEEE 11th Workshop on Advances in Information, Electronic and Electrical Engineering*, 2024 m., Valmieroje, Latvijoje.

Disertacijos struktūra

Disertaciją sudaro keturi pagrindiniai skyriai, bendrosios išvados, literatūros sąrašas, autoriaus publikacijų sąrašas ir santrauka lietuvių kalba. Disertacijos apimtis – 152 puslapių, kuriuose 54 paveikslai, 26 formulės, 24 lentelės, 142 šaltiniai.

Padėka

Tam vieninteliam žmogui. Ačiū už kantrybę.

1. Bendrasis įvadas į patirties kokybės vertinimą judriojo ryšio tinkluose

Mobiliojo ryšio duomenų srautas pasauliniu mastu nuolat auga, o vaizdo turinys sudaro didžiausią jo dalį (apie 76%) (Ericsson, 2025). Nors 5G technologijų diegimas didina tinklo pralaidą, operatoriai susiduria su iššūkiu užtikrinti aukštą paslaugų kokybę aplinkoje, kurioje dominuoja šifruotas srautas ir trečiųjų šalių OTT paslaugos.

Šioje disertacijoje QoE apibrėžiama remiantis EU Qualinet bendruomenės formuluote (Brunnström ir kt., 2013), kurioje teigiama, kad tai vartotojo pasitenkinimo arba susierzinimo laipsnis, kurį lemia lūkesčių, susijusių su paslauga, patenkinimas. Operatoriams, neturintiems tiesioginės prieigos prie vartotojų QoE įverčių, tenka uždavinys susieti objektyvius tinklo rodiklius su suvokiama paslaugos kokybe.

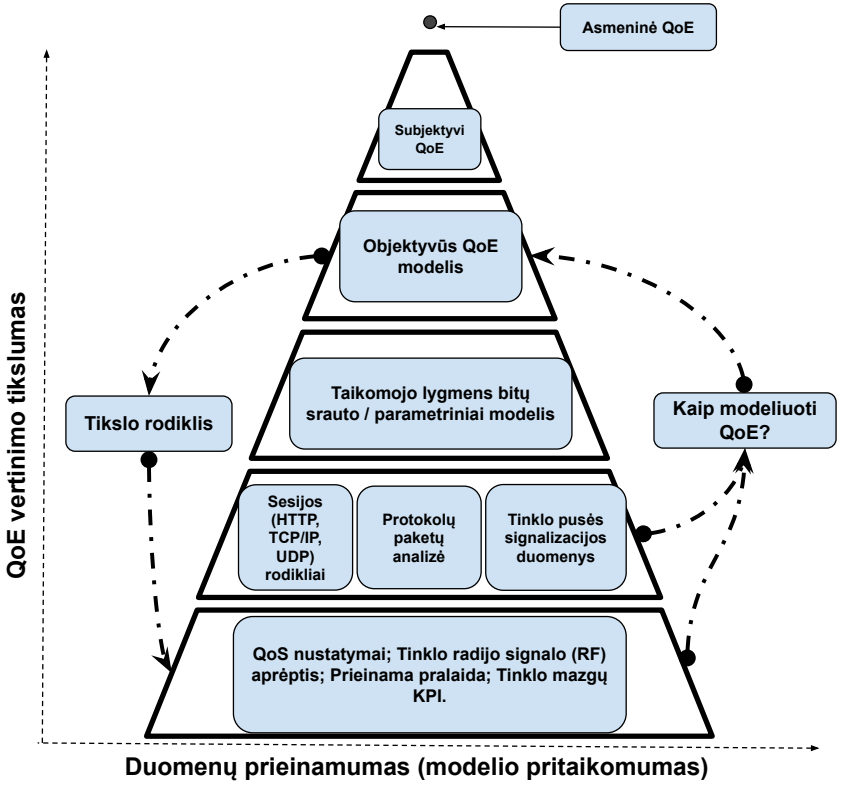
Norint tiksliai modeliuoti, vertinti ir prognozuoti QoE, būtina suprasti duomenų perdavimo kelią judriojo ryšio tinklo architektūroje ir nustatyti taškus, kuriuose operatorius gali rinkti reikšmingą informaciją. Todėl pirmasis žingsnis, vertinant QoE iš MNO perspektyvos, yra nustatyti, kokius duomenis galima surinkti, kur jie generuojami ir kokio detalumo lygiu jie pasiekiami. Šiuolaikiniai 4G/5G tinklai dažnai diegiami remiantis 5G NSA (angl. *Non-Standalone*) architektūra. Vartotojo įrenginys per LTE ir NR Uu sąsajas jungiasi prie bazinių stočių, kurios per IP transporto tinklą sujungtos su branduolio tinklu bei turinio pristatymo tinklais ar debesų serveriais. Duomenų srautas nuo taikomosios programos iki serverio perduodamas naudojant IP (HTTP, TCP, UDP, QUIC) ir 3GPP (PDCP, RLC, MAC, PHY) protokolus. MNO gali rinkti šiuos duomenis:

- taikomosios programos arba bitų srauto lygmens duomenis (jei jie pasiekiami bendradarbiaujant su OTT paslaugų teikėjais),
- transporto ir sesijos lygmens rodiklius (paketų praradimą, duomenų perdavimo spartą, vėlinimą ir vėlinimo svyravimą),
- signalizacijos ir RAN KPI (RSRP, SINR, PRB panaudojimo lygį, persijungimus (angl. *handover*), RRC ryšio užmezgimo trukmę),

- radijo signalo aprėpties ir geolokacijos duomenis (pvz., MDT ataskaitas su GNSS koordinatėmis).

Šie įvairių lygmenų duomenų šaltiniai leidžia MNO netiesiogiai modeliuoti vartotojo QoE, net jei originalus medijos srautas ir subjektyvūs vartotojų vertinimai nėra prieinami.

QoE gali būti vertinama keliais lygmenimis, kurių kiekvienas remiasi skirtingo detalumo ir prieinamumo duomenimis. Šį kompromisą tarp vertinimo tikslumo ir duomenų prieinamumo iliustruoja QoE modelių piramidė (1 pav.). Piramidės viršuje yra asmeninė QoE – individualus vartotojo suvokimas, kuris ITU-T dokumentuose apibrėžiamas kaip bendras paslaugos priimtinumas, subjektyviai vertinamas galutinio vartotojo (International Telecommunication Union, 2008). Praktikoje šis lygmuo operatoriui beveik niekada nėra tiesiogiai prieinamas. Subjektyvi QoE matuojama standartizuotomis testavimo procedūromis (ACR, DCR, CCR, SAMVIQ, DSCQS ir kt.), apibrėžtomis ITU-T P.800 ir P.910 rekomendacijose (ITU-T, 1996, 1999). Testų rezultatai dažniausiai išreiškiami vidutiniu nuomonės balu MOS (nuo 1 iki 5), kuris naudojamas kaip etalonas kitiems modeliams. Tačiau tokie testai yra brangūs, lėti ir netaikomi realiuoju laiku. Objektyvūs QoE



S1.1 pav. Hierarchinė QoE modeliavimo metodų struktūra pagal duomenų prieinamumą, modelių pritaikomumą ir vertinimo tikslumą

metodai remiasi signalo savybėmis ir siekia prognozuoti MOS be tiesioginių vartotojų apklausų. Priklausomai nuo prieigos prie etaloninio signalo, išskiriami trys objektyvių QoE metodų tipai: visiško etalono (pvz., PSNR ir SSIM), dalinio etalono ir be etalono metodai (Kougioumtzidis ir kt., 2022). Šis lygmuo pasirinktas kaip pagrindinis disertacijos tyrimų objektas. Sudėtingesni modeliai, tokie kaip VQM ir VMAF, sujungia kelis požymius ir apmokomi naudojant MOS duomenų rinkinius, siekiant pasiekti aukštą koreliaciją su subjektyviais vartotojų vertinimais (Li ir kt., 2016; Pinson & Wolf, 2004; Rassool, 2017). Objektyvūs modeliai teoriškai reikalauja prieigos prie originalaus medijos srauto ir jo iškraipytos versijos. Dėl OTT paslaugų ir srauto šifravimo MNO paprastai neturi nei etaloninio srauto, nei dekoduoto turinio, todėl etaloniniai metodai realiuose tinkluose yra riboto pritaikomumo. Dėl šios priežasties operatoriai yra priversti remtis netiesioginiais tinklo lygmens požymiais. Operatoriai QoE vertinti naudoja radijo parametrus, pvz., RSRP, SINR, PRB užimtumą, siuntimo galią ir moduliacijos schemą. Be to, norint tinkamai taikyti QoE modelius, būtina atpažinti naudojamos paslaugos tipą (pvz., VoIP, vaizdo transliaciją ar žaidimą). Srauto šifravimas riboja tradicinių metodų, tokių kaip IP ir TCP prievadų analizė bei giluminė paketų analizė, taikymą, todėl pereinama prie mašininio mokymosi metodų.

Apibendrinant, analitinė QoE vertinimo judriojo ryšio tinkluose apžvalga leidžia suformuluoti šias pagrindines išvadas:

1. Patirties kokybė (QoE) yra daugiamatė sąvoka, kurią formuoja ne tik tinklo našumas, bet ir paslaugos tipas, turinys, įrenginys, taikomosios programos elgsena bei naudojimo kontekstas.

2. OTT paslaugų scenarijuose judriojo ryšio tinklų operatoriai neturi tiesioginės prieigos prie galutinio naudotojo QoE ar etaloninių medijos srautų. Todėl operatoriaus lygmens QoE vertinimas privalo remtis netiesioginiais rodikliais, gaunamais iš transporto, sesijos ir RAN lygmenų matavimų.

3. Paslaugos tipo atpažinimas yra būtina sąlyga operatoriui atliekant QoE vertinimą. Dėl plačiai taikomo šifravimo MNO privalo kliautis antraščių lygmens, laiku grįstais (angl. *timing-based*) ir srauto formos požymiais, derinamais su mašininiu mokymusi grįstais modeliais.

4. Siekiant tiksliai įvertinti operatoriaus lygmens QoE, būtinas duomenų integravimas, apimantis transporto ir sesijos rodiklius, tokius kaip vėlinimas, paketų praradimas, vėlinimo svyravimas ir pralaida, kartu su RAN rodikliais, tokiais kaip aprėptis, apkrova ir mobilumo įvykiai.

5. RAN aprėptis ir talpa sudaro QoE pagrindą. QoE modeliai, kartu su NWDAF ir O-RAN duomenimis grįstais valdymo ciklais, yra būtini siekiant užtikrinti proaktyvų, į QoE orientuotą RAN optimizavimą.

2. Naudojamų paslaugų atpažinimas judriojo ryšio tinkluose

Judriojo ryšio operatoriams būtina žinoti, kokiomis paslaugomis naudojasi jų klientai, nes ši informacija reikalinga QoE vertinti ir valdyti, apimant srauto prioritetų nustatymą, pralaidos ribojimą ir diferencijuotą paslaugų kainodarą. Norint užtikrinti pageidaujamą QoE lygį, pirmiausia reikia patikimai atpažinti naudojamos paslaugos tipą. Dėl plačiai taikomo srauto šifravimo tradiciniai paslaugų atpažinimo metodai tapo neefektyvūs (Papadogiannaki & Ioannidis, 2021; Rezaei & Liu, 2019a). Todėl tyrėjai pradėjo taikyti klasikinius mašininio mokymosi (Velan ir kt., 2015) ir giliojo mokymosi modelius (Wang ir kt., 2019). Šiame skyriuje pristatomi du tyrimai: (i) taikomųjų programų šifruoto judriojo ryšio srauto klasifikavimas (Chmieliauskas & Paulikas, 2022) ir (ii) vaizdo srauto atpažinimo metodas, paremtas bitų srauto formos požymiais (Chmieliauskas & Paulikas, 2023).

Tyrimui surinktas 3257 valandų trukmės duomenų rinkinys iš realių judriojo ryšio tinklų naudotojų, kurio laiko skiriamoji geba yra 1 sekundė. Duomenų rinkinyje fiksuojamas naudojamos programos pavadinimas bei priimtų ir išsiųstų bitų kiekis. Statistinė analizė parodė, kad vien tik vidurkio ir standartinio nuokrypio nepakanka patikimai atskirti programas dėl didelio jų reikšmių persidengimo. Tačiau vizualinė analizė rodo, kad skirtingos programos (pvz., *YouTube*, kuri pasižymi periodiniu vaizdo perdavimo srautu, ir *Messenger*, kuri skambučio metu generuoja nuolatinį srautą) pasižymi skirtinga bitų srauto forma. Laiko eilučių duomenims klasifikuoi pasirinktas konvoliucinis neuroninis tinklas. Šio tinlo architektūra suprojektuota priimti vienmačius (tik priimtų bitų) arba dvimačius (priimtų ir išsiųstų bitų) vektorius. Modelis apmokytas spręsti du pagrindinius uždavinius: (i) daugiaklasę programų klasifikaciją (*Facebook*, *YouTube*, *Chrome*, *Outlook*, *9GAG*, *Messenger*) ir (ii) dvejetainę klasifikaciją, atskiriant vaizdo srautą (tik *YouTube*) nuo visų kitų programų.

Atliekant dvejetainę klasifikaciją, atskiriant vaizdo srautą nuo kitų programų, pasiektas didelis klasifikavimo tikslumas. Ilgesnės sekos (300 s ir 600 s) reikšmingai pagerino klasifikavimo rezultatus, pateiktus S2.1 lentelėje. Geriausi rezultatai gauti naudojant 600 s sekas ir abi kryptis (DL+UL) 2D formatu: pasiektas 0,929 tikslumo rodiklis. Modelis taip pat įvertintas naudojant naują, nesubalansuotą duomenų rinkinį, kuriame programos pasiskirsčiusios pagal realų jų naudojimo laiką. Rezultatai rodo, kad *YouTube* pavyzdžiai teisingai atpažįstami maždaug 85 % atvejų, o klaidingos klasifikacijos dažniausiai susijusios su situacijomis, kai *YouTube* programa naudojama tik naršymui, bet ne aktyviam vaizdo transliavimui. Kitos programos, tokios kaip *9GAG* ir *Facebook*, pasižymi palyginti didele klaidingų priskyrimų vaizdo kategorijai dalimi, nes jose integruoti vaizdo grotuvai generuoja adaptyvaus vaizdo transliavimo srautus.

Lyginant su metodais, reikalaujančiais sudėtingos duomenų analizės (pvz., TCP/IP

S2.1 lentelė. Modelio tikslumo priklausomybė nuo sekos trukmės ir įėjimo tipo, atskiriant vaizdo srautą nuo kitų programų

Sekos trukmė	Tik priimti baitai	Priimti + išsiųsti (1D)	Priimti + išsiųsti (2D)
60 s	0,772	0,814	0,837
300 s	0,859	0,902	0,896
600 s	0,815	0,900	0,929

S2.2 lentelė. Siūlomo metodo palyginimas su kitais srauto klasifikavimo ir programų atpažinimo būdais

Darbas	Metai	Alg.	Duomenys	Įėjimai / paruošimas	Acc	P	R
(Zheng ir kt., 2022)	2022	Trans.	ISCX VPN–nonVPN	Paketų (pcap) failai; reikalingos TCP/IP žinios	0,993	0,988	0,988
(Yao ir kt., 2019)	2019	LSTM	ISCX VPN–nonVPN	Tinklo srautų rekonstravimas; reikalingos TCP/IP žinios	0,912	–	–
(Lotfollahi ir kt., 2020)	2020	Autoen.	ISCX VPN–nonVPN	Antraščių ir naudingosios apkrovos analizė; reikalingos TCP/IP žinios	0,980	0,990	0,990
(Lopez-Martin ir kt., 2017)	2017	CNN	RedIRIS	IP srautų antraštės; reikalingos TCP/IP žinios	0,961	0,952	0,961
Siūlomas metodas	2023	CNN	Autorių duomenų rinkinys	Tik bitų srauto forma; TCP/IP žinios nereikalingos	0,929	0,997	0,900

paketų antraščių analizės ar srautų rekonstravimo), matyti, kad siūlomas konvoliucinis neuronų tinklo modelis, naudodamas tik bitų srauto formos požymius, pasiekia palyginamus rezultatus (S2.2 lentelę). Tai patvirtina metodo praktiškumą: jam pasitelkti nereikia specifinių žinių apie TCP/IP protokolus ir gali būti lengvai pritaikomas bazinės stoties MAC lygmenyje.

3. Vaizdo srautų patirties kokybės vertinimas judriojo ryšio tinkluose

Šiame skyriuje nagrinėjama patirties kokybės priklausomybė nuo radijo prieigos tinklo rodiklių, perduodant vaizdo srautus per komercinius LTE ir NR tinklus. Analizuojami du esminiai scenarijai: (1) OTT vaizdo perdavimas priėmimo kryptimi (Chmieliauskas & Paulikas, 2021), kuris sudaro didžiąją dalį judriojo ryšio duomenų srauto, ir (2) realaus laiko vaizdo perdavimas naudojant RTMP arba WebRTC protokolus (Chmieliauskas & Paulikas, 2025).

Didžioji dalis OTT vaizdo paslaugų naudoja HTTP adaptyvųjį srautinį perdavimą. Vaizdo turinys suskaidomas į kelių sekundžių segmentus, parengtus skirtingais bitų srauto spartos lygiais. Taip sumažinama vaizdo atkūrimo strigimo rizika, o vaizdo kokybė automatiškai pritaikoma prie tinklo sąlygų. Pagrindinių vaizdo kodekų analizė parodė, kaip kodeko pasirinkimas tiesiogiai lemia RAN apkrovą ir operatoriaus radijo išteklių sąnaudas. Taip pat įvertintos HAS perdavimo savybės LTE tinklo bazinėse stotyse. Dėl santykinai

didelės dvipusės delsos ir periodinio srauto pobūdžio TCP ne visada išnaudoja visą prieinamą radijo sąsajos pralaidumą, todėl radijo ištekliai užimami ilgiau, nei būtų idealiomis sąlygomis. Vaizdo srauto radijo išteklių sąnaudos vertinamos pagal fizinių išteklių blokų poreikį. Eksperimentas su 20 MHz LTE 1800 MHz kanalu (100 PRB) parodė, kad esant žemam SINR net ir 5 Mbit/s srautas gali sunaudoti didelę dalį ląstelės išteklių ir apriboti galimybes gauti vaizdo srautą. Esant žemam SINR (CQI < 5–6), 20 MHz LTE ląstelė gali aptarnauti tik kelis naudotojus, vienu metu gaunančius apie 5 Mbit/s vaizdo srautą. CQI statistika rodo, kad reikšminga seansų dalis patenka į „blogos aprėpties“ zoną, todėl didėja PRB sąnaudos ir vaizdo atkūrimo strigimo rizika. 1080p raiškos vaizdui naudojant AV1 arba VP9 kodekus dažniausiai pakanka 2,5–4,5 Mbit/s spartos, o naudojant AVC kodeką pakanka apie 6 Mbit/s, todėl efektyvesni kodekai mažina RAN apkrovą.

5G NR technologiniai patobulinimai (pvz., *massive* MIMO, platesnės dažnių juostos ir TDD konfigūracijos su 4:1 DL/UL santykiu) daugiausia skirti gaunamosios krypties pralaidumui didinti. Be to, adaptyvaus srautinio perdavimo technologija nepritaikyta realaus laiko vaizdui perduoti siunčiamąja kryptimi. Analizei pasirinkti RTMP ir WebRTC protokolai, reprezentuojantys dvi skirtingas paradigmas: patikimumą užtikrinantį perdavimą TCP protokolo pagrindu ir mažos delsos perdavimą UDP protokolu su adaptyvia spartos kontrole.

Sukurta viso perdavimo kelio matavimų architektūra, kurioje realaus laiko 4K kameros vaizdas, naudojant *OBS Studio* (RTMP) ir *RTC Studio* (WebRTC) perduodamas per komercinį 4G/5G tinklą į *Ant Media Server*, veikiantį AWS debesų infrastruktūroje. Kartu buvo renkami ir šie duomenys: lokaliai įrašytas referencinis vaizdo failas, per tinklą perduotas ir AWS serveryje įrašytas vaizdo srautas, IP paketų PCAP įrašai, RAN matavimai (RSRP, SINR, sklidimo nuostoliai, PRB skaičius, MCS, momentinis pralaidumas ir kt.) bei GNSS koordinatės. Vaizdo kokybė vertinama pilnosios atskaitos kokybės metrikomis VMAF, SSIM ir PSNR, papildomai analizuojant kadrų tipus, medijos lygmens laiko žymas, kadrų trukmes ir dydžius. Taip gaunama kiekvienam kadrui apskaičiuota QoE laiko seka, kuri siejama su kas 1 s agreguotais RAN matavimais. Kadangi WebRTC adaptacija lemia kintamą kadrų dažnį ir galimą kadrų praradimą, sukurta speciali kadrų sinchronizavimo metodika. Surinktas duomenų rinkinys apima keliasdešimt valandų realaus vaizdo įkėlimo bandymų, atliktų miesto, priemiesčių ir automagistralių trasose LTE 800 MHz, LTE 1800 MHz ir NSA (LTE+NR) tinkluose, naudojant įvairias bitų srauto spartas ir raiškas (360p–1080p) bei RTMP ir WebRTC protokolus. LTE 800 MHz dažnių juostoje buvo tirtas 720p realaus laiko vaizdo srautas. Nustatyta, kad VMAF yra glaudžiai susijęs su pasiekiamu pralaidumu: didėjant pralaidumui ir artėjant prie reikiamo bitų srauto, QoE gerėja. Rezultatai pateikti S3.1 lentelėje.

Išanalizavus pralaidumo komponentus (MCS, PRB skaičių ir sklidimo nuostolius), nustatyta, kad MCS stipriai teigiamai koreliuoja su VMAF, o sklidimo nuostoliai su VMAF koreliuoja neigiamai. PRB skaičiaus įtaka yra sudėtingesnė, nes ją riboja UE perdavimo galia ir kitų naudotojų apkrova. Pritaikius regresijos modelius (polinominę regresiją ir SVR), parodyta, kad vien iš operatoriui prieinamų rodiklių (sklidimo nuostolių, MCS, PRB ir pralaidumo) galima prognozuoti VMAF kai $R^2 \approx 0{,}8$, o siunčiamosios krypties pralaidumą – kai $R^2 \approx 0{,}89$. Tai rodo, kad QoE galima patikimai modeliuoti tinklo pusėje, neturint prieigos prie taikomosios programos duomenų.

S3.1 lentelė. Regresijos modeliai siunčiamosios krypties pralaidai (UL) ir VMAF (V)

Modelis	MSE (UL)	R^2 (UL)	MSE (V)	R^2 (V)
Polinominė SVR (4 laipsnio)	0,2067	0,8852	124,27	0,7985
Polinominė regresija (6 laipsnio)	0,2073	0,8849	128,42	0,7918
SVR (RBF, C=100, γ=0,1)	0,2095	0,8837	124,61	0,7980
Polinominė regresija (4 laipsnio)	0,2099	0,8834	124,27	0,7985
Polinominė regresija (5 laipsnio)	0,2129	0,8818	135,38	0,7805
Polinominė regresija (3 laipsnio)	0,2230	0,8761	125,90	0,7959
Polinominė SVR (3 laipsnio)	0,2285	0,8731	125,74	0,7961
Polinominė regresija (2 laipsnio)	0,2314	0,8715	119,88	0,8057
SVR (RBF, C=1000, γ=1)	0,2480	0,8623	155,67	0,7476
Tiesinė regresija	0,3114	0,8271	154,85	0,7490
Keterinė regresija	0,3114	0,8271	154,85	0,7490
Bajesinė regresija	0,3114	0,8271	154,85	0,7490
Tampriojo tinklo regresija	0,3124	0,8265	154,85	0,7490
LASSO regresija	0,3128	0,8263	154,85	0,7490
Tiesinė SVR	0,3188	0,8229	178,34	0,7109
Polinominė SVR (2 laipsnio)	0,3291	0,8172	128,46	0,7917
Puasono GLM	0,3629	0,7985	200,24	0,6754
Sprendimų medis	0,3705	0,7943	229,51	0,6279
LARS regresija	0,4994	0,7227	154,85	0,7490

Miesto NSA (LTE+NR) scenarijuje buvo palyginti RTMP ir WebRTC rezultatai esant didesniam ląstelių tankiui ir dažnesniems persijungimo tarp ląstelių (angl. *handover*) įvykiams. Nustatyta, kad RTMP beveik visuose bandymuose pasiekia labai aukštą VMAF įvertį (daugiau kaip 95), tačiau tam reikia didesnio bitų srauto. WebRTC, dėl judriojo ryšio tinklams nepritaikytos adaptacijos, sumažina bitų srautą, todėl vidutinis VMAF įvertis yra gerokai mažesnis. Persijungimo tarp ląstelių įvykiai WebRTC atveju sukelia staigų bitų srauto kritimą ir ilgalaikį kokybės prastėjimą. RTMP atveju TCP protokolas geba greitai atkurti pralaidumą.

4. Duomenimis grįsti patirties kokybės gerinimo metodai

Šiame skyriuje nagrinėjami praktiniai metodai, leidžiantys gerinti naudotojo patirties kokybę komerciniuose LTE ir 5G NR tinkluose, remiantis RAN rodikliais ir 3GPP parametrų parinkimu. QoE glaudžiai siejasi su pagrindiniais tinklo rodikliais, tokiais kaip pralaidumas, delsa, SINR ir tinklo apkrova. Skyriuje pristatomi trys vienas kitą papildantys QoE gerinimo metodai: (i) tinklo apkrovos prognozavimas taikant laiko eilučių analizę (Chmieliauskas & Gu̇šnys, 2019), (ii) aprėpties ir talpos optimizavimas taikant skatinamąjį mokymąsi (Chmieliauskas ir kt., 2023) ir (iii) MIMO parametrų parinkimas, didinantis naudotojui pasiekiamą pralaidumą (Chmieliauskas, 2021).

Laikiniai duomenų srauto svyravimai ląstelės ir viso tinklo mastu yra svarbus veiks-

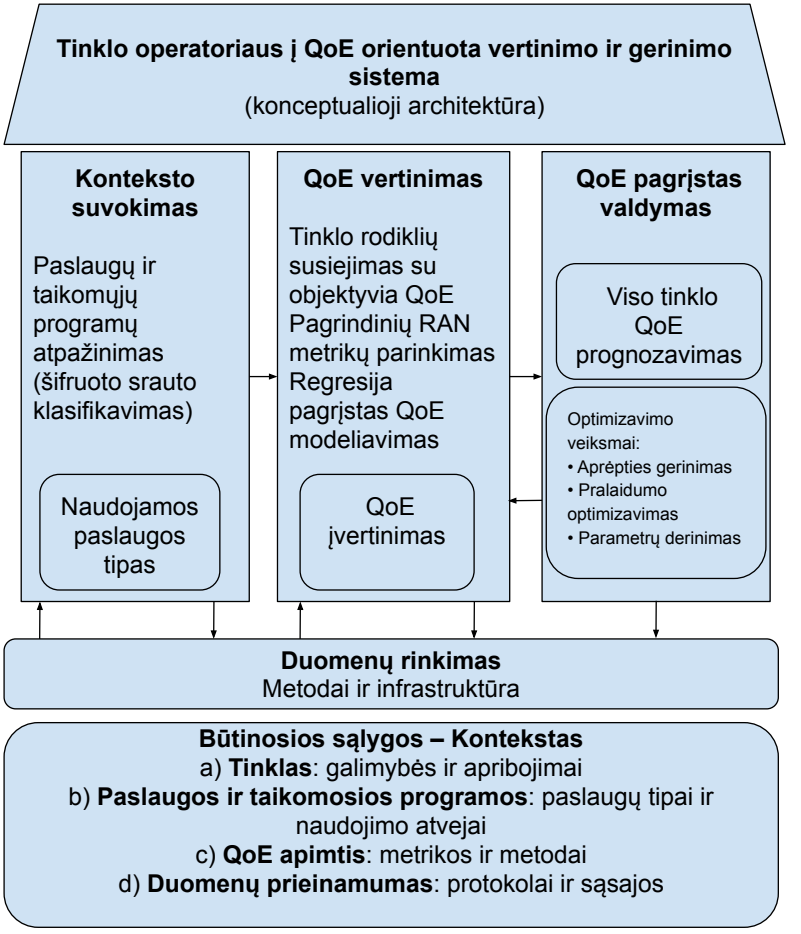
nys planuojant tinklo našumą ir prognozuojant galimą QoE prastėjimą. Siūloma dviejų pakopų procedūra, leidžianti prognozuoti paros srautą ir įvertinti piko valandos apkrovą. Pirmajame etape kiekvienai LTE ląstelei prognozuojamas bendras paros duomenų srautas, taikant laiko eilučių metodą fbProphet. Prognozavimui naudojami 180 dienų istoriniai duomenys: 150 dienų skiriama mokymui, o 30 dienų – testavimui. Analizuojama 100 atsitiktinai parinktų komercinio tinklo stočių imtis (378 LTE ląstelės: 141 B3 ir 237 B20), apimanti miesto, priemiesčių ir kaimo teritorijas. Prognozės tikslumas vertinamas pagal R^2 ir vidutinės absoliučios procentinės paklaidos (MAPE) rodiklius. Rezultatai rodo, kad pagal R^2 prognozės tikslumas siekia apie 0,9, tačiau vidutinė absoliuti procentinė paklaida yra didesnė mažo srauto ląstelėse dėl atsitiktinių naudotojų aktyvumo svyravimų. Didelio srauto ląstelės pasižymi stabilesniais ir geriau prognozuojamais srautais. Antrajame etape iš 30 dienų valandinės statistikos sudaromas paros eigos profilis, nustatoma piko valanda ir jos dalis bendrame paros sraute. Tai leidžia prognozuoti piko valandos apkrovą ir įvertinti, ar ląstelė artėja prie radijo išteklių ribos. Prognozuota apkrova toliau siejama su ląstelės talpa, vertinant PRB panaudojimą. Kadangi pralaidumas priklauso nuo SINR, MCS ir MIMO sluoksnių, siūloma naudoti PRB apkrovą kaip pagrindinį perkrovos rodiklį ir kiekvienai ląstelei įvertinti spektrinį efektyvumą (bit/s/Hz). Tai leidžia apskaičiuoti likutinę ląstelės talpą ir prognozuoti galimą QoE prastėjimą dėl perkrovos.

Aprėpties ir talpos optimizavimo tikslas – maksimaliai padidinti tinklo talpą, išlaikant pakankamą aprėptį. Tinklo talpa teritorijoje gali būti išreiškiama vertinant šiuos parametrus: B – juostos plotį, CD – ląstelių tankį ir SE – spektrinį efektyvumą. Talpą galima padidinti plečiant juostą, didinant ląstelių tankį arba gerinant spektrinį efektyvumą. Pirmieji du būdai dažnai reikalauja didelių investicijų, todėl svarbu gerinti spektrinį efektyvumą optimizuojant RAN parametrus. Siūloma skatinamuoju mokymusi paremta metodika, orientuota į bazinių stočių antenų azimutų parinkimą. Uždavinys formuluojamas kaip nepolinominio sudėtingumo paieška didelėje konfigūracijų erdvėje, todėl, didėjant stočių skaičiui, didėjant stočių skaičiui, išsamioji paieška tampa nepraktiška. Aplinka apibrėžiama kaip aprėpties žemėlapis, o būsena atitinka konkrečią stočių azimutų konfigūraciją. Veiksmai modeliuoja diskretųjį azimuto keitimą (pvz., $-20°$, $0°$, $+20°$), o atlygis parenkamas kaip vidutinis SINR. Konfigūracijoms įvertinti naudojamas kanalo ir aprėpties modeliavimui skirtas 3GPP 3D UMa LOS modelis. Skatinamojo mokymosi agentas iteratyviai ieško būsenų, kurios suteikia didesnį atlygį. Trijų ir keturių stočių scenarijų rezultatai rodo spartų konvergavimą link didesnio SINR ir tolygesnį SINR pasiskirstymą teritorijoje, taip pat leidžia reikšmingai sumažinti iteracijų skaičių, palyginti su išsamia paieška.

LTE FDD technologijoje lyginami du dažniausiai komerciniuose tinkluose naudojami MIMO perdavimo režimai: TM3 (angl. *open-loop spatial multiplexing*) ir TM4 (angl. *closed-loop spatial multiplexing*). Pagrindinis skirtumas yra atgalinio ryšio informacija: TM3 atveju UE pateikia RI ir CQI, o veikiant TM4 režimu papildomai pateikia ir PMI, leidžiantį bazinei stočiai pagal kanalo savybes tiksliau parinkti išankstinį kodavimą (angl. *precoding*). Vertinimas atliktas trimis scenarijais: naudotojui judant dideliu greičiu, naudotojui judant lėtai plačiame SINR diapazone bei atliekant didelio masto tinklo konfigūracijų palyginimą (apimantį tūkstančius 2T2R ir 4T4R ląstelių). Atliekant matavimus, pralaidumas normalizuotas pagal suplanuotų PRB skaičių ir paverstas *visiško išteklių panaudojimo* reikšme – taip sumažinta momentinės tinklo apkrovos įtaka. Rezultatai ro-

do nuoseklią TM4 naudą: išmatuotas vidutinis pralaidumas padidėja nuo 105 Mbps iki 128 Mbps, o atliekant didelio masto palyginimą nustatyta, kad vidutinis naudotojo pralaidumas išauga apie 13–14 % 2T2R ląstelėse ir apie 29–30 % 4T4R ląstelėse.

Remiantis ankstesniuose skyriuose ir poskyriuose aprašytomis technologijomis bei metodais, siūloma apibendrinta tinklo lygmens QoE vertinimo struktūra. Ją sudaro trys pagrindiniai komponentai: atpažinimas, vertinimas ir gerinimas. Norint įgyvendinti šią sistemą, būtina aiškiai suprasti tokius aspektus: a) tinklo galimybes ir apribojimus; b) paslaugas ir taikomąsias programas, įskaitant paslaugų tipus bei naudojimo atvejus; c) QoE apimtį, apibrėžtą taikytinais rodikliais ir vertinimo metodais; d) duomenų prieinamumą, įskaitant atitinkamus protokolus, sąsajas ir duomenų rinkimo metodus.



S4.1 pav. Operatoriaus pusės į QoE orientuota vertinimo ir gerinimo sistema

Bendrosios išvados

1. Vaizdo srautinio perdavimo patirties kokybė (QoE) operatoriaus pusėje vertinama integruojant jam prieinamus kelių lygmenų duomenis (pvz., RAN ir IP transporto rodiklius), nenaudojant paketų naudingosios apkrovos analizės ar etaloninių srautų serverio pusėje.

2. Remiantis iš eksploatuojamų mobiliojo ryšio tinklų surinktais duomenimis, sukurta apibendrinta tinklo pusės QoE vertinimo sistema, apimanti paslaugų atpažinimą, objektyvų patirties kokybės vertinimą ir juo grįstą kokybės gerinimą.

3. Pasiūlytas CNN grįstas metodas rodo, kad vaizdo srautas gali būti identifikuojamas praktiškai pritaikomu tikslumu, naudojant mažo sudėtingumo algoritmą – pasiekiant didesnį nei 0,90 vaizdo ir kito srauto klasifikavimo tikslumą 300–600 s laiko langams, naudojant tik 1 s intervalu agreguotas priimtų ir išsiųstų duomenų kiekio sekas.

4. Ribotos aprėpties aukštynkrypčio ryšio scenarijuose nustatyta aiški sąsaja tarp VMAF rodiklio ir operatoriaus pusės indikatorių (signalo slopinimo, MCS, PRB priskyrimo bei pasiekto pralaidumo). Remiantis šiais požymiais, netiesinės regresijos modeliai prognozavo VMAF reikšmes (determinacijos koeficientas R^2 artimas 0,80). Tai patvirtina, kad operatoriaus pusės QoE vertinimas galimas neturint prieigos prie taikomosios programos naudingosios apkrovos.

5. Proaktyvus tinklo valdymas įgyvendinamas derinant ML grįstą KPI prognozavimą su skatinamuoju mokymusi grįstu aprėpties ir talpos optimizavimu. „Prophet“ grįstas metodas leidžia prognozuoti piko valandų srautą ir tinklo ląstelių apkrovą iki 28 dienų į priekį ($R^2 \approx 0,9$), o „Q-learning“ grįstas antenos azimuto parinkimas per kelis epizodus konverguoja į beveik optimalias reikšmes, pagerindamas vidutinį SINR ir jo erdvinį pasiskirstymą.

Darius CHMIELIAUSKAS

QUALITY OF EXPERIENCE EVALUATION AND ENHANCEMENT
OF VIDEO STREAMING SERVICES IN CELLULAR NETWORKS

Doctoral Dissertation

Technological Sciences,
Electrical and Electronic Engineering (T 001)

VAIZDO PERDAVIMO PASLAUGŲ PATIRTIES KOKYBĖS
VERTINIMAS IR GERINIMAS JUDRIOJO RYŠIO TINKLUOSE

Daktaro disertacija

Technologijos mokslai,
Elektros ir elektronikos inžinerija (T 001)

Lietuvių kalbos redaktorė Deimantė Grigaitė
Anglų kalbos redaktorė Jūratė Griškėnaitė

2026 04 30. 14 sp. l. Tiražas 20 egz.
Leidinio el. versija https://doi.org/10.20334/2026-024-M
Vilniaus Gedimino technikos universitetas
Saulėtekio al. 11, 10223 Vilnius
Spausdino UAB „Ciklonas“,
Žirmūnų g. 68, 09124 Vilnius